ROBUST FACIAL EXPRESSION RECOGNITION USING LOCAL BINARY PATTERNS

Caifeng Shan, Shaogang Gong and Peter W. McOwan

Department of Computer Science, Queen Mary, University of London, London, E1 4NS, UK
{cfshan, sgg, pmco}@dcs.qmul.ac.uk

ABSTRACT

A novel low-computation discriminative feature space is introduced for facial expression recognition capable of robust performance over a rang of image resolutions. Our approach is based on the simple Local Binary Patterns (LBP) for representing salient micro-patterns of face images. Compared to Gabor wavelets, the LBP features can be extracted faster in a single scan through the raw image and lie in a lower dimensional space, whilst still retaining facial information efficiently. Template matching with weighted Chi square statistic and Support Vector Machine are adopted to classify facial expressions. Extensive experiments on the Cohn-Kanade Database illustrate that the LBP features are effective and efficient for facial expression discrimination. Additionally, experiments on face images with different resolutions show that the LBP features are robust to low-resolution images, which is critical in real-world applications where only low-resolution video input is available.

1. INTRODUCTION

Facial Expression is one of the most powerful, nature, and immediate means for human beings to communicate their emotions and intentions [14]. Due to its potential applications, automatic facial expression recognition has attracted much attention [2] over two decades. Though much progress has been made [3, 4, 5, 6], recognizing facial expression with a high accuracy remains to be difficult due to the complexity and variety of facial expressions.

Automatic facial expression recognition involves two vital aspects: facial feature representation and classifier design. Facial feature representation is to derive a set of features from original face images which minimizes within-class variations of expressions whilst maxmizes between-class variations. There are two main types of approaches to extract facial features: geometric feature-based methods and appearance-based methods [1]. Gabor-wavelet appearance features were demonstrated to be more effective than geometric features [7], and work better in real-world environments [6]. Although Gabor-wavelet representations have been widely adopted [7, 5, 6], it is computationally expensive to convolve face images with multi-banks of Gabor filters in order to extract multiscale and orientational coefficients.

In this paper, we introduce Local Binary Patterns (LBP) as novel low-cost discriminative features for facial expression recognition. LBP was proposed originally for texture analysis [8, 9]; recently Ahonen et al [10, 11] presented LBP based methods for face detection and recognition. Our motivation is that face images can be seen as a composition of micro-patterns which can be well described by LBP. A facial image is divided into a set of small regions from which LBP histograms are extracted and concatenated into a single, spatially enhanced feature histogram. The simple LBP features can be fast derived in a single scan through the raw image, whilst still retaining enough facial infomation in a compact representation. While regarding classifier design, many classifiers have been applied to expression recognition, such as neural network (NN) [7, 6], Support Vector Machine (SVM) [5, 12], Linear Discriminant Analysis (LDA) [3], and Bayesian network [4]. SVM is adopted here since it has been successfully applied to expression recognition [5, 12]. For comparison, we also used the simple template matching with weighted Chi square statistic [10]. Extensive experiements on the Cohn-Kanade database [13] clearly demonstrate that the LBP features are effective for facial expression recognition.

Most existing facial expression algorithms aim to recognize facial expressions from data collected in a highly controlled environment with high resolution frontal faces [6]. However, in real-world environments, face images are often in lower resolution. To our best knowledge, only Tian et al [14, 6] made attempts to recognize facial expressions with lower resolution recently. In this paper, we further evaluate the effectiveness of our LBP-based algorithm on different image resolutions. Comparative experiments show that the LBP features are robust for low-resolution face images.

2. LOCAL BINARY PATTERNS

The original LBP operator was introduced by Ojala et al [8]. The operator labels the pixels of an image by thresholding the 3×3 neighbourhood of each pixel with the center value and considering the result as a binary number (see left of

0-7803-9134-9/05/$20.00 ©2005 IEEE

Fig 1 for an illustration). Then the histogram of the labels can be used as a texture descriptor.

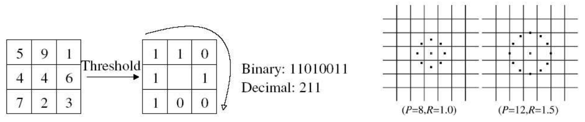

Fig. 1. Left: The basic LBP operator [10]. Right: Two examples of the extended LBP [9]: a circular $(8, 1)$ neighborhood, and a circular $(12, 1.5)$ neighbourhood.

The limitation of the basic LBP opterator is its small 3×3 neighbourhood can not capture dominant features with large scale structures. Hence the operator was extended to use neighbourhood of different sizes [9]. Using circular neighbourhoods and bilinearly interpolating the pixel values allow any radius and number of pixles in the neighbourhood. Examples of the extended LBP are shown in right of Fig 1, where (P, R) denotes P sampling points on a circle of radius of R.

Further extension of LBP is to use uniform patterns [9]. A Local Binary Pattern is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular. For example, 00000000, 001110000 and 11100001 are uniform patterns. It is observed that uniform patterns account for nearly 90% of all patterns in the $(8, 1)$ neighbourhood and for about 70% in the $(16, 2)$ neighbourhood in texture images [9].

Here we adpot a LBP operator $LBP^{u2}_{P,R}$. The subscript represents using the operator in a (P, R) neighbourhood; the superscript $u2$ indicates using only uniform patterns and labelling all remaining patterns with a single label. A histogram of a labelled image $f_l(x, y)$ can be defined as

$$H_i = \sum_{x,y} I(f_l(x, y) = i), \qquad i = 0, \dots, n-1 \tag{1}$$

where n is the number of of different labels produced by the LBP operator and

$$I(A) = \begin{cases} 1 & A \text{ is true} \\ 0 & A \text{ is false} \end{cases} \tag{2}$$

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, feature extracted should retain also spatial information. Hence the face image is divided into m small regions $R_0, R_1, \dots, R_m$ (see left of Figure 2 for an example) and a spatially enhanced histogram is defined as

$$H_{i,j} = \sum_{x,y} I(f_l(x, y) = i) I((x, y) \in R_j) \tag{3}$$

where $i = 0, \dots, n-1, j = 0, \dots, m-1$.

In this histogram, the face is descripted on three different levels of locality: the labels for the histogram contain the pixel-level patterns, the labels are summed over a small region to produce information on a regional level and the regional histograms are concatenated to build a global description of the face [10]. The LBP histogram can be rapidly computed, and lies in a low dimensional space.

3. FACIAL EXPRESSION RECOGNITION

3.1. Template Matching

For expression recognition, we first adopt template matching for its simplicity. In training, the LBP histograms of face images in a given class are averaged to generate a histogram template for this class. In recognition, a nearest-neighbour classifier is adopted: the LBP histogram of the input image is matched with the closest template.

We select Chi square statistic (χ^2) as the dissimilarity measure for histogram. It is observed that facial features constributing to facial expressions mainly lie in some regions, such as eye area and mouth area, these regions contain more useful information for facial classification. Therefore, a weight can be set for each face region based on the importance of the information it contains (see right of Fig 2 for an illustration). Our weighted (χ^2) statistic is

$$\chi^2_w(\mathbf{S}, \mathbf{M}) = \sum_{i,j} w_j \frac{(S_{i,j} - M_{i,j})^2}{S_{i,j} + M_{i,j}} \tag{4}$$

where S and M are two LBP histograms, w_j is the weight for region j.

3.2. Support Vector Machine (SVM)

Support Vector Machine is a popular technique for classification [15]. SVM performs an implicit mapping of data into a higher dimensional feature space, where linear algebra and geometry can be used to separate data that is only separable with nonlinear rules in the input space.

Given a training set of labeled examples $T = \{(x_i, y_i), i = 1, \dots, l\}$ where $x_i \in R^n$ and $y_i \in \{1, -1\}$, the new test data x is classified by the following function:

$$f(x) = sgn(\sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b) \tag{5}$$

where α_i are Lagrange multipliers of a dual optimization problem, and $K(x_i, x)$ is a kernel function. Given a nonlinear mapping Φ that embeds input data into feature space, kernels have the form of $K(x_i, x_j) = \langle\Phi(x_i)\cdot\Phi(x_j)\rangle$. SVM finds a linear separating hyperplane with the maximal margin to separate the training data in feature space. b is the parameter of the optimal hyperplane.

SVM allows domain-specific selection of the kernel function. Though new kernels are being proposed, the most frequently used kernel function are the linear, polynomial, and RBF kernels. SVM makes binary decisions. Multi-class classification here is accomplished by a cascade of binary classifiers together with a voting scheme.

4. EXPERIMENTS

4.1. Data Set

The proposed algorithm were trained and tested on the Cohn-Kanade Facial Expression Database [13]. This database consists of 100 university students in age from 18 to 30 years, of which 65% were female, 15% were African-American, and 3% were Asian or Latino. Subjects were instructed to perform a series of 23 facial displays, six of which were based on descriptions of prototypic emotions (i.e., anger, disgust, fear, joy, sadness, and surprise). Image sequences from neutral to target display were digitized into 640×490 pixel arrays.

For our experiments, we selected 320 image sequences from the database for basic emotional expression recognition. The sequences come from 96 subjects, with 1 to 6 emotions per subject. For each sequence, the neutral face and three peak frames of each sequence were used. To evaluate generalization performance, a 10-fold cross-validation test scheme was adopted.

Similar to Tian [6], we normalized the faces to a fixed distance between the centers of the two eyes. The fixed distance was 55 pixels. It is observed that the width of a face is roughly two times of the distance, and the height is roughly three times. Hence, facial images of 110×150 pixels were cropped from orignial frames based on the two eyes location. No further alignment of facial features such as alignment of mouth [7] was performed in our algorithm. Due to LBP's gray-scale invariance, there was no attempt made to remove illumination changes in our algorithm.

Some parameters can be optimized for the LBP feature selection. The first one is the LBP operator, and another is the number of regions divided. We selected the $LBP_{8,2}^{u2}$ operator, which has 59 labels. We divided the 110×150 pixels facial images into 18×21 pixels regions, giving a good trade-off between recognition performance and feature vector length [10]. Thus the facial image was divided into 42(6×7) regions as shown in left of Fig 2.

4.2. Experiments: Template Matching

The weights we designed for template matching is shown in right of Fig 2. We set the weights empirically based on observation of facial features. 10-fold Cross-Validation experiments were carried out for both 6-class and 7-class (by including neutral face) expression recognition tasks. For the 6-class task, the generalization performance was 84.5% and for the 7-class task, it was 79.1%.

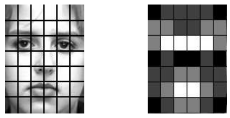

Fig. 2. Left: A face image divided into 6×7 sub-region. Right: The weights set for weighted dissimilarity measure. Black squares indicate weight 0.0, dark gray 1.0, light gray 2.0 and white 4.0.

Based on the tracked geometric facial feature (eyebrows, eyelids, and mouth), Cohen et al [4] adopted Bayesian network classifiers to classify 7-class expressions on the Cohn-Kanade database. The best performance of 73.2% was obtained by using Tree-Augmented-Naive Bayes (TAN) classifiers. Comparison in Table 1 illustrates that our simple template matching using LBP outperforms geometric features based TAN classifier. The experiments demonstrated that the LBP features are discriminative for facial expression recogniton.

Methods (Feature + Classifier)	Recogniton Results
LBP + Template Matching	79.1%
Geometric Feature + TAN [4]	73.2%

Table 1. Comparisons between the geometric features based TAN [4] and our LBP-based template matching.

4.3. Experiments: SVM

This experiment was designed to not only further verify the effectiveness of LBP features, but also compare our LBP-based algorithm to existing Gabor-wavelets-based methods. Here we used the SVM implementation in SPIDER [1]. The generalization performance with linear, polynomial and RBF kernels in 7-class recognition were 87.2%, 88.4% and 87.6% respectively.

	Linear	Polynomial	RBF
LBP+SVM	87.2%	88.4%	87.6%
Gabor+SVM [5]	84.8%	worse than linear/RBF	86.9%

Table 2. Comparisons between the Gabor-wavelets based SVM [5] and our LBP-based SVM.

Recently Bartlett et al [5] conducted similar experiments. They selected 313 image sequences from the Cohn-Kanade database. The sequences came from 90 subjects, with 1 to 6 emotions per subject. The facial images were converted into

[1] SPIDER can be downloaded from *http://www.kyb.tuebingen.mpg.de /bs/people/spider/index.html* freely.

a Gabor magnitude representation using a bank of 40 Gabor filters. Then they performed 10-fold cross-validation experiments using SVM with linear, polynomial, and RBF kernels. Linear and RBF kernels perfromed best, with recognition rates of 84.8% and 86.9% respectively; results of polynomial kernel were not given in their paper.

Comparisons summaried in Table 2 show that the LBP-based SVM algorithm outperforms the Gabor-wavelets-based SVM method. More crucially though, our advangtage lies at the simplicity of LBP histogram allows for very fast feature extraction, does not need complex analysis in extracting a large set of Gabor wavelet coefficients. Compared to high dimensionality of the Gabor resprentation $O(10^5)$ [12], the LBP features with short vector length $O(10^3)$ lies in a lower dimensional space, thus requires much less computational resource.

4.4. Experiments: Perfomance over Face Resolution

In many applications envolving facial expression recognition, the input facial images are often low-resolution. We further evaluated the LBP-based algorithm over a range of image resolutions, investigating its performance against low-resolution images. A total of six different resolutions of the facial region were studied (110×150, 55×75, 36×48, 27×37, 18×24, and 14×19 pixels). The lower resolution images were down-sampled from the original images. The recognition rates of 6-class expressions are summaried in Table 3.

	 110×150	 55×75	 36×48	 27×37	 18×24	 14×19
LBP	92.1%	88.6%	86.3%	83.5%	78.7%	75.8%
Gabor [6]	92.2%	91.6%	-	77.6%	-	68.2%

Table 3. Comparisons on image resolutions. The first row: our LBP-based algorithm; The second row: Gabor-wavelets based algorithm [6].

Recently Tian [6] explored the effects of different image resolutions for expression recognition on the Cohn-Kanade database. 375 image sequences were selected for six emotions recognition. A bank of 40 Gabor filters were applied to extract appearance feature. Tian used a three-layer neural network to recognize expressions. The recognition rates in [6] are shown in the second row of Table 3 [2]. We observed that the proposed LBP-based method is more effective for face images with low resolutions than Gabor-based methods. The LBP features perform robustly and stably over a useful range of low resolutions.

[2] In [6], the different resolutions of the head region are 144×192, 72×96, 36×48, 18×24 pixels, which are comparable to the resolutions of the face region 110×150, 55×75, 27×37, 14×17 in our experiments.

5. CONCLUSIONS

This paper presented a new method for facial expression recognition using Local Binary Patterns. Compared with Gabor wavelets, the simple LBP features save much computational resource whilst retaining facial information efficiently. Extensive experiments demonstrate that the LBP features are discriminative and robust over a range of facial image resolutions, which is critical in real-world applications where only low-resolution video input is available.

6. REFERENCES

[1] Y. Tian, T. Kanade, and J.F. Cohn, *Facial Expression Analysis*, Handbook of Face Recognition, Springer, October 2003.

[2] B. Fasel and J. Luettin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, vol. 36, pp. 259–275, 2003.

[3] M. J. Lyons, J. Budynek, and S. Akamatsu, "Automatic classification of single facial images," *IEEE PAMI*, 1999.

[4] I. Cohen, N. Sebe, Garg A., L. Chen, and T. Huang, "Facial expression recognition from video sequences: Temporal and static modeling," *CVIU*, vol. 91, pp. 160–187, 2003.

[5] M.S. Bartlett, G. Littlewort, I. Fasel, and R. Movellan, "Real time face detection and facial expression recognition: Development and application to human computer interaction," in *CVPR Workshop on CVPR for HCI*, 2003.

[6] Y. Tian, "Evaluation of face resolution for expression analysis," in *IEEE Workshop on Face Processing in Video*, 2004.

[7] Z. Zhang, M. J. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron," in *IEEE FG*, April 1998.

[8] T. Ojala, M Pietikinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distribution," *Pattern Recognition*, vol. 29, no. 1, 1996.

[9] T. Ojala, M. Pietikinen, and T. Menp, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE PAMI*, vol. 24, no. 7, July 2002.

[10] T. Ahonen, A. Hadid, and M. Pietikinen, "Face recognition with local binary patterns," in *ECCV*, 2004, pp. 469–481.

[11] A. Hadid, M. Pietikinen, and T. Ahonen, "A discriminative feature space for detecting and recognizing faces," in *IEEE CVPR*, June 2004, pp. 797–804.

[12] G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, and J. Movellan, "Dynamics of facial expression extracted automatically from video," in *IEEE Workshop on Face Processing in Video*, 2004.

[13] T. Kanade, J.F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *IEEE FG*, 2000, pp. 46–53.

[14] Y. Tian, L. Brown, A. Hampapur, S. Pankanti, A. Senior, and R. Bolle, "Real world real-time automatic recognition of facial expression," in *IEEE PETS*, Australia, March 2003.

[15] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.