Multi-view Gender Classification Using Local Binary Patterns and Support Vector Machines

Hui-Cheng Lian and Bao-Liang Lu

Department of Computer Science and Engineering, Shanghai Jiao Tong University,
800 Dong Chuan Rd., Shanghai 200240, China
{lianhc, blu}@cs.sjtu.edu.cn

Abstract. In this paper, we present a novel approach to multi-view gender classification considering both shape and texture information to represent facial image. The face area is divided into small regions, from which local binary pattern(LBP) histograms are extracted and concatenated into a single vector efficiently representing the facial image. The classification is performed by using support vector machines(SVMs), which had been shown to be superior to traditional pattern classifiers in gender classification problem. The experiments clearly show the superiority of the proposed method over support gray faces on the CAS-PEAL face database and a highest correct classification rate of 96.75% is obtained. In addition, the simplicity of the proposed method leads to very fast feature extraction, and the regional histograms and global description of the face allow for multi-view gender classification.

1 Introduction

Gender classification is one of the most challenging problems for face recognition researchers. Similar to any pattern classification problems, two key points for gender classification are feature extraction and pattern classification. From the view of feature extraction, the most simple method is to use gray-scale or color pixel vectors as features [1]. Another kind of methods comes from the theory of subspace transformation such as PCA, ICA and LDA, which project faces into a low-dimensional space and then recognize them [2]. This kind of method has been shown not very robust to variations of face orientation. The third kind of methods is using texture information such as winkle and complexion [3]. The last kind of methods is combining the facial feature detection with wavelet transform to extract the local facial feature for classification [4, 5], such as the analysis of facial winkles and shapes.

Traditional pattern classifiers such as k-nearest-neighbor, Fisher linear discriminant, neural network and SVMs are often employed to gender classification. SVMs seem to be superior to all other classifiers [1]. Moghaddam and Yang developed an appearance-based method to classify gender from facial images using nonlinear SVMs and compared their performance with traditional classifiers including Fisher linear discriminant, nearest-neighbor, RBF networks and large ensemble-RBF classifiers [1]. However, they focused their study on very

J. Wang et al. (Eds.): ISNN 2006, LNCS 3972, pp. 202–209, 2006.

low-resolution masked images, in which only the main frontal facial regions are visible and almost completely excluded hair information.

Recently a powerful way of texture description called local binary patterns (LBP) has been proposed for texture classification [6], face detection [7], and face recognition [8, 9]. An exciting recognition rate of 97.9% has been yielded via using this kind of texture description on the FERET FA/FB image sets. Considering the efficiency and effectiveness of LBP for face representation, we propose a method that combines LBPs and SVMs for gender classification. We evaluate the proposed method on the CAS-PEAL face database with comparison to [1]. The experimental results indicate that LBP-based methods can do well on both low degree faces and 30 degree faces.

2 Feature Extraction with Local Binary Patterns

The original LBP operator, introduced by Ojala *et al.* [6], is a powerful way of texture description. 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. Then the histogram of the labels can be used as a texture descriptor. The basic LBP operator is illustrated in Fig. 1 (a).

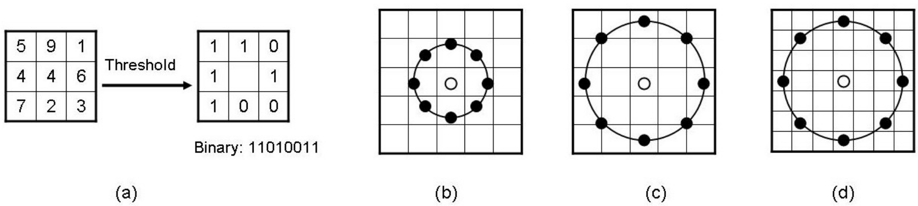


Fig. 1. The basic LBP operator and three examples of extended LBPs. (a) The basic LBP operator. (b) The circular (8,1) neighborhood. (c) The circular (8,2) neighborhood. (d) The circular (8,3) neighborhood.

The most prominent limitation of the LBP operator is its small spatial support area. Features calculated in a local 3×3 neighborhood cannot capture large scale structure that may be the dominant features of some textures. Later the operator was extended to use neighborhoods of different size [6]. Using circular neighborhoods and bilinearly interpolating the pixel values allow any radius and number of pixels in the neighborhood. Examples of these kinds of extended LBP are shown in Fig. 1(b), (c), (d).

Another extension to the original LBP operator is to use so called uniform patterns [6]. A local binary pattern is defined 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, 00011110, and 10000011 are uniform patterns. Ojala *et al.* [6] noticed that in their experiments with texture images, uniform patterns account for a bit less than 90% of all patterns when using the (8,1) neighborhood and for 70% in (16,2) neighborhood.

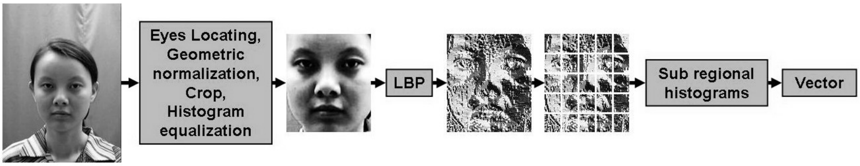


Fig. 2. Feature extraction. LBP transformation is done after locating eye positions from original images, geometric normalization, cropping and histogram normalization. The LBP face is divided into blocks and their histograms fitted together form a vector.

We use the notation $LBP^u_{P,R}$ for the uniform LBP operator. $LBP^u_{P,R}$ means using the LBP operator in a neighborhood of P sampling points on a circle of radius R. The superscript u stands for using uniform patterns and labelling all remaining patterns with a single label. The number of labels for a neighbourhood of 8 pixels is 256 for standard LBP and 59 for $LBP^u_{8,1}$.

A histogram of the labelled image $f_l(x, y)$ can be defined as

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

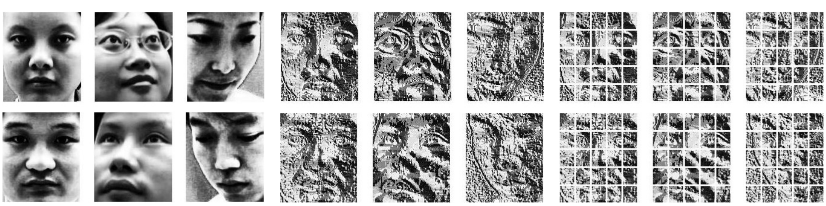
where n is the number of different labels produced by the LBP operator and

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

This histogram contains information about the distribution of the local micro-patterns over the whole image, such as edges, spots and flat areas. For efficient face representation, one should retain also spatial information. For this purpose, the image is divided into regions $R_0, R_1, ..., R_{m-1}$, as shown in Fig. 2 and the spatially enhanced histogram is defined as

$$H_{i,j} = \sum_{x,y} I\{f_l(x, y) = i\} I\{(x, y) \in R_j\}, \tag{1}$$

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



(a) (b) (c)

Fig. 3. Some of multi-view faces from the CAS-PEAL face database. (a) Preprocessed faces. (b) Corresponding LBP faces. (c) Corresponding 5×5 grid LBP faces.

The process of feature extraction for gender classification is illustrated in Fig. 3. An original image is processed by locating eye positions, geometric normalization, cropping and histogram normalization and a so called LBP face is obtained by performing LBP operator on this preprocessed facial image. $K \times K$ equal size blocks are divided from the LBP face with a grid on it and their histograms fitted together form a vector that will be fed into the gender classifier. Fig. 3 shows some of multi-view samples and their corresponding LBP faces from the CAS-PEAL face database.

3 Support Vector Machines

Support vector machine is a learning algorithm for pattern classification, regression and density estimation [10]. The basic training principle behind SVMs is finding the optimal linear hyperplane such that the expected classification error for unseen test samples is minimized. According to the structural risk minimization principle and VC dimension minimization principle [10], a linear SVM uses a systematic approach to find a linear function with the lowest capacity. For linearly nonseparable data, SVMs can nonlinearly map the input to a high-dimensional feature space where a linear hyperplane can be found.

Given a labelled set of M training samples $(\mathbf{x}_i, y_i)$ where $\mathbf{x}_i \in R^N$ and y_i is the associated label ($y_i \in \{-1, 1\}$), a SVM classifier finds the optimal hyperplane that correctly separates the training data while maximizing the margin. The discriminant hyperplane is defined by:

$$f(\mathbf{x}) = \sum_{i=1}^{M} y_i \alpha_i \cdot k(\mathbf{x}, \mathbf{x}_i) + b$$

where $k(\cdot, \cdot)$ is a kernel function, b is a bias and the sign of $f(\mathbf{x})$ determines the class membership of $\mathbf{x}$. To construct an optimal hyperplane is equivalent to finding all the nonzero α_i and is formulated as a quadratic programming (QP) problem with constraints.

For a linear SVM, the kernel function is just a simple dot product in the input space while for a nonlinear SVM the kernel function projects the samples to a higher dimension feature space via a nonlinear mapping function:

$$\Phi : R^N \to F^M,$$

where $M \gg N$, and then constructs a hyperplane in F. By using Mercer's theorem [10], the projecting samples into the high-dimensional feature space can be replaced by a simpler kernel function satisfying the condition

$$k(\mathbf{x}, \mathbf{x}_i) = \Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}_i)$$

where Φ is the nonlinear projection function. Several kernel functions, such as, polynomials and radial basis functions, have been shown to satisfy Mercer's theorem and have been used successfully in nonlinear SVMs:

$$k(\mathbf{x}, \mathbf{x}_i) = ((\mathbf{x} \cdot \mathbf{x}_i) + 1)^d$$
$$k(\mathbf{x}, \mathbf{x}_i) = \exp(-\gamma|\mathbf{x} - \mathbf{x}_i|^2)$$

where d is the degree in a polynomial kernel and γ is the spread of a Gaussian cluster.

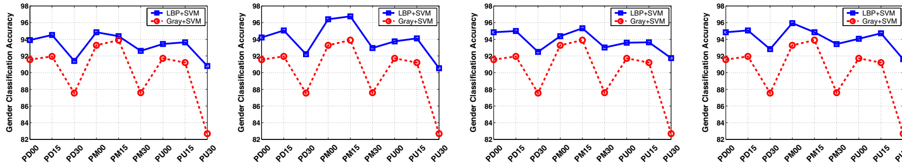
4 Experiments and Discussions

CAS-PEAL [11] is a large-scale face database that currently contains 14,384 pose images of 1,040 individuals. In this database there are thousands of samples with 9 different poses including looking up pose, looking middle pose and looking down pose with 0 degree, 15 degree and 30 degree, respectively. The eye coordinates are manually located. The image is scaled, lined up by the eye coordinates and cropped. The histogram of the image is equalized and the resolution of final image is 150×130 pixels. Table 1 shows the detailed contents about the database where 'Pxnn' gives all of the pose information about the data sets. The character 'P' represents pose variation. The 'x' (U, M, D) indicates the subject's pose (Up, Middle, Down). The 'nn' indicates the azimuth of the camera from which the image is obtained. Take the first row for example, it includes 1,040 images from 1,040 individuals (445 males and 595 females) and the first 400 images from 400 individuals (200 males and 200 females) are used as training samples and the rest 640 images from other 640 individuals are used as test samples. Consequently, the total number of training sample is 3,600 and the total number of test sample is 10,784.

Table 1. Description of training and test data sets from the CAS-PEAL face database for multi-view gender classification

set ID	Description	No. Male	No. Female	No. Total data	No. Training	No. Test
1	PD00	445	595	1,040	200×2	640
2	PD15	846	1,032	1,878	200×2	1,478
3	PD30	846	1,032	1,878	200×2	1,478
4	PM00	445	595	1,040	200×2	640
5	PM15	844	1,032	1,876	200×2	1,476
6	PM30	844	1,032	1,876	200×2	1,476
7	PU00	445	595	1,040	200×2	640
8	PU15	846	1,032	1,878	200×2	1,478
9	PU30	846	1,032	1,878	200×2	1,478
#	TOTAL	6,407	7,977	14,384	3,600	10,784

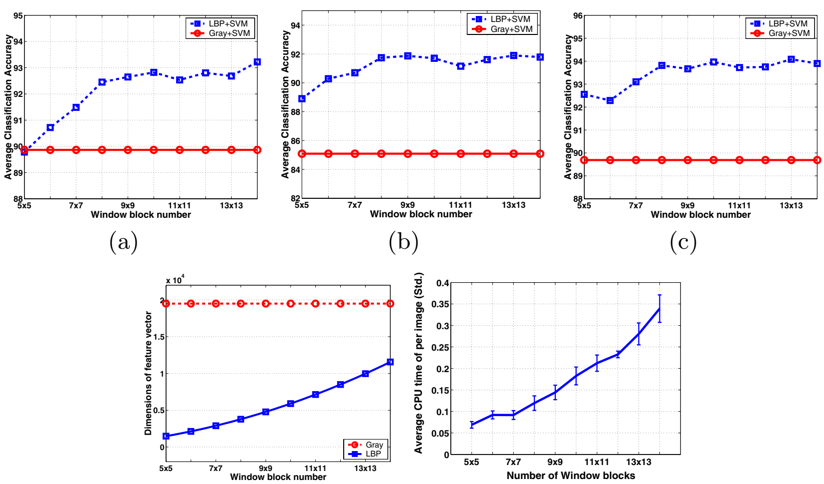
We use $LBP^{u}_{8,1}$ in all experiments and the 130×150 pixels image is divided into $K \times K$ equal size blocks, where K is ranged from 5 to 14. The LBP histograms of blocks are extracted and concatenated into a single, spatially enhanced feature histogram by using Equation (1). SVMs with linear kernel, polynomial kernel and RBF kernel are chosen to evaluate the performance of our



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

Fig. 4. Comparison of LBP method with the grey pixel approach using polynomial SVMs with degree 3 on the CAS-PEAL face database. (a) K=7. (b) K=8. (c) K=9. (d) K=10.

method. We also compare the proposed method with the existing approach [1]. The reasons why [1] is chosen are firstly this work of [1] is outstanding and secondly SVMs have been shown to be superior to traditional pattern classifiers for gender classification problem. The most important difference between our work and the existing approach [1] is that we use LBP histograms as feature extraction while the existing approach uses grey pixels. In addition to this difference, the simplicity of the proposed method allows for very fast feature extraction, and the regional histograms and global description of the face allow for efficiently multi-view gender classification. All of the SVMs in our experiments come from LibSVM [12] and all parameters are the default parameters of LibSVM except specially noting. To evaluate the efficiency of the LBP method, we firstly compare LBP method using K=7, 8, 9, 10 with the grey pixel approach the on data



(a) (b) (c)

(d) (e)

Fig. 5. Average classification accuracy comparison with three different kernels and the feature dimension comparison. (a) RBF kernel. (b) Linear kernel. (c) Polynomial kernel. (d) Feature dimension comparison. (e) CPU time (second) of LBP feature extraction.

sets of all degrees. The experimental results are shown in Fig. 4. From Fig. 4 we can see that the correct rates of LBP method are obviously higher than Moghaddam's method with a highest correct rate 96.75%. More particularly, on all 30 degree data sets (U, M, D), the grey pixel method almost fails with a lowest accuracy 82.68%, while LBP method can largely increase correct rates on these corresponding data sets, with a highest improvement from 82.68% to 91.75%.

To further show the performance, we compare LBP method with Moghaddam's method using three different SVM kernels and K ranged from 5 to 14 on average classification accuracy of all nine poses (see Fig. 5 (a), (b) and (c)). From them we can see that the average correct rate can even be highly improved from 89.69% to 94.08% meanwhile from Fig. 5 (d) we can see that the vector dimensions are much lower than that of the grey pixel approach. For example, when K is 5 or 6, the number of dimensions is 1,475($5 \times 5 \times 59$) or 2,124($6 \times 6 \times 59$), while the number of dimensions of the original grey pixel method is 19,500(150×130).

For there is no time needed for using the grey face directly, we do not compare CPU time of LBP feature extraction method with grey face method, while compare the LBP method with the Gabor method which is a very widely used method and can be found in [3, 4, 11]. From Fig. 5(e) we can see that the average CPU time of per image of LBP methods are only 0.07 seconds and 0.34 seconds for $K = 5$ and $K = 14$, while under the same conditions (3GHz PC/Matlab 6.1) the CPU times of Gabor methods are 6.31 seconds and 9.71 seconds for 4 scales with 6 frequencies and 5 scales with 8 frequencies.

The reason why LBP method can do well on both low degree faces and 30 degree faces is that this kind of method can describe face on three different levels of locality: the labels for the histogram contain information about the patterns on a pixel-level, the labels are summed over a small region to produce information on a regional level, and the reginal histograms are concatenated to build a global description of the face. And the reason why the feature dimension can be largely reduced is that facial image is first divided into small regions from which LBP histograms are extracted and concatenated into a single, spatially enhanced feature histogram efficiently representing the facial image. The simplicity and efficiency of uniform LBP operator and the superiority of SVMs over traditional pattern classifiers lead to a rapid and precise multi-view gender classification.

5 Conclusions

We have introduced a new method for multi-view gender classification by combining powerful LBP-based facial description with support vector machines. The efficiency and simplicity of LBP allow for very fast feature extraction, and the regional and global descriptions allow for capturing multi-view information of faces. Meanwhile the high generalization ability of SVMs allows for learning and classifying gender from a large set of multi-view faces. The experimental results show that classification accuracies are highly improved and a highest correct rate is 96.75% and a highest average correct rate is 94.08% on the CAS-PEAL face database.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China via the grants NSFC 60375022 and NSFC 60473040. Portions of the research in this paper use the CAS-PEAL face database collected under the sponsor of the Chinese National Hi-Tech Program and ISVISION Tech. Co. Ltd.

References

1. Moghaddam, B.,Yang, M.H.: Learning Gender with Support Faces. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5) (2002) 707-711
2. Balci, K., Atalay, V.: PCA for Gender Estimation: Which Eigenvectors Contribute? In: 16 th International Conference on Pattern Recognition (ICPR'02) Vol. 3. Quebec City, QC, Canada, (2002) 363-366
3. Iga, R.: Gender and Age Estimation System from Face Images. In: SICE Annual Conference in Fukui. August 4-6(2003) 756-761
4. Hosoi, S., Takikawa, E., Kawade M.: Ethnicity Estimation with Facial Images. In: Sixth IEEE International Conference on Automatic Face and Gesture Recognition May 17-19, 2004 Seoul, Korea (2004) 195-200
5. Lian, H. C., Lu, B. L.: Gender Recognition Using a Min-Max Modular Support Vector Machine. In: Proc. ICNC'05-FSKD'05, LNCS 3611 (2005) 433-436
6. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (2002) 971-987
7. Jin, H.L, Liu, Q. S., Tong, X. F.: Face Detection Using Improved LBP Under Bayesian Framework. In: Proceedings of the Third International Conference on Image and Graphics (ICIG 04), Hong Kong, China, 18-20 Dec (2004) 306-309
8. Ahonen, T., Hadid, A., Pietikäinen, M.: Face Recognition with Local Binary Patterns. In: Proceedings of the European Conference on Computer Vision (2004) 469-481
9. Zhang, G.C.: Boosting Local Binary Pattern (LBP)-Based Face Recognition. In: Proceedings of the Sinobiometrics 2004, LNCS 3338 (2004) 179-186
10. Vapnik, V.: The Nature of Statistical Learning Theory. Springer, (1995)
11. Gao, W., Cao, B., Shan, S. G., *et al.*, The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations, technical report of JDL, (2004), avaible on http://www.jdl.ac.cn/~peal/peal_tr.pdf
12. Chang, C.C., Lin, C.J. LIBSVM : a Library for Support Vector Machines. http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.ps.gz