Face Detection Using Improved LBP Under Bayesian Framework

Hongliang Jin, Qingshan Liu, Hanqing Lu, Xiaofeng Tong
National Lab of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, P.O.Box 2728, Beijing 100080, China
{hljin, qsliu, luhq,xftong}@nlpr.ia.ac.cn

Abstract

In this paper, we present a novel face detection approach using Improved Local Binary Patterns (ILBP) as facial representation. ILBP feature is an improvement of LBP feature that considers both local shape and texture information instead of raw grayscale information and it is robust to illumination variation. We model the face and non-face class using multivariable Gaussian Model and classify them under Bayesian framework. Extensive experiments show that the proposed method has an encouraging performance.

1. Introduction

Face detection is a key step of the automatic face analysis system. In recent years, many face detection methods have been published [6]. Among them, appearance-based approaches show the best performance in the face detection task, such as Adaboost approach by viola [9], sparse network of winnows (SNoW) [3], Neural Network [4], support vector machines (SVM) [7] and distribution-based approach [8]. Most appearance-based face detection approaches consist of three main steps: Facial feature selection, face model establishment and discriminative rules design. In this paper, we propose Improved LBP (ILBP) features, which are illumination invariant, as the facial features instead of the raw image appearance-based features. The objective of adopting such facial representation is that the ILBP feature is an excellent measure of the special structure of the local image texture and it is invariant against any monotonic transform of grayscale. Then we model the face and non-face class using the multivariable Gaussian distribution. Given ILBP features of an image patch, we could calculate the likelihood of face class and non-face class separately. Finally, the Bayesian decision rule is applied to decide the patch is a face or not.

This paper is organized as follows: Improved Local Binary Patterns (ILBP) as facial representation are introduced in Section 2. The face class and non-face class models, together with the classifier under Bayesian framework, are described in Section 3. Experiments are reported in section 4. Conclusion and our future work are given in Section 5.

2. The Improved Local Binary Patterns

LBP [1] is a grayscale irrelevant texture operator with powerful discrimination. With the neighborhood set *P* and a circle of radius *R*, we could compute the difference between the central pixel "g_c" and its neighborhood $\{g_0,\ldots g_{P-1}\}$ to generate the operator as $LBP_{P,R}$. In general, $LBP_{P,R}$ produces 2^P different values. Function (1) gives the computation of $LBP_{P,R}$ number.

$$LBP_{P,R} = \sum_{i=0}^{P-1} s(g_i - g_c)2^i.$$
$$s(x) = \begin{cases} 1, x > 0 \\ 0, x \le 0 \end{cases} \tag{1}$$

Pixels greater than the central pixel are mapped to 1, otherwise 0. Fig1 gives an example of computing LBP feature.

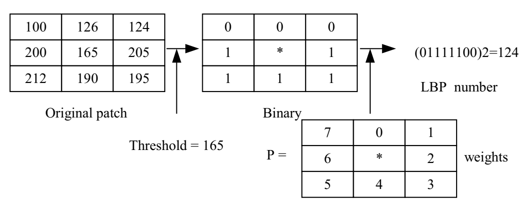

Figure 1. Example of Computing $LBP_{8,1}$

Though LBP features have great discriminative power, they miss the local structure under some certain circumstance. Taking $LBP_{8,1}$ for example, using function (1), we can only get $2^8 = 256$ of all the 511 patterns (all zeros and all ones are the same) from a 3*3 patch. This is because the central pixel "g_c" is set to 0 all through. In order to get all the representations of LBP, we should consider the effect of the central pixel. In most cases, the central point provides more information than its neighborhood, so we should give it the largest weight. In function (2), we take the mean of the 3*3 patch as the new threshold. The raw pixel value [0 255] then is

Proceedings of the Third International Conference on Image and Graphics (ICIG'04)
0-7695-2244-0/04 $20.00 © 2004 **IEEE**

mapped to [0 510] with local texture and shape information in. In the similar way, we can map the pixel value to [0 30] using $LBP_{4,1}$. The improved $LBP_{P,R}$ can be expressed as follows:

$$LBP_{P,R} = \sum_{i=0}^{P-1} s(g_i - m)2^i + s(g_c - m)2^P.$$

$$s(x) = \begin{cases} 1, x > 0 \\ 0, x \le 0 \end{cases} \tag{2}$$

$$m = \frac{1}{P+1}\left(\sum_{i=0}^{P-1} g_i + g_c\right)$$

Fig2 gives the mapping weights illustration of our Improved LBP in function 2.

7	0	1
6	8	2
5	4	3

ILBP81

X	0	X
3	4	1
X	2	X

ILBP41

Figure 2. Mapping weights for $ILBP_{8,1}$ and $ILBP_{4,1}$ ("X" stands for arbitrary pixel value)

In Fig3, we compare the performance of the original $LBP_{4,1}$ with our $ILBP_{4,1}$. Fig3.a is an image patch representing a corner microstructure and Fig3.b shows the result of original LBP according to function (1). Apparently we cannot get the right local structure using the original LBP because no pixel in Fig3.a is greater than $g_c (= 20 \ \ here)$. Using ILBP representation in function (2), the new threshold is set to 12 [62/5]. We can see that the local structure is obtained successfully after thresholding.

X	1	X
1	20	20
X	20	X

a. Original patch

X	0	X
0	*	0
X	0	X

b. Original LBP

X	0	X
0	1	1
X	1	X

c. improved LBP

Figure 3. Comparison of Original LBP and ILBP

In our experiments, we use the $ILBP_{8,1}$ and $ILBP_{4,1}$ features to extract facial features for face detection instead of raw grayscale features. All training face samples are cropped to the size m*n. We get the two ILBP images with the size (m-2)*(n-2) from each training sample without the border. Thus the length of facial feature is $N = 2*(\text{m - 2})*(\text{n - 2})$. Since the ILBP features are insensitive to the variation of illumination, there is no need to do image enhancement such as illumination equalization to remove the influence of light.

3. Face Class and Non-Face Class Modeling

The multivariate Gaussian model is one of the most important Bayesian tools, which we could use to establish the face class model and estimate the conditional probability density function (p.d.f.). However, it is difficult to model non-face class because anything that is not a face belongs to non-face class. Here, as in [4], we can use "bootstrap" strategy to gather non-face samples, which lie very close to the face class in the feature space. Then we can use these face-like non-face samples to represent the non-face class, for our objective is to learning the optimal hyperplane to separate the face and non-face class.

3.1 Multivariate Gaussian Model

Given a vector I of ILBP features, the conditional probability density function of the face class C_f can be modeled as fellows:

$$p(I \mid C_f) = \frac{\exp\{-\frac{1}{2}(I - M_f)^T \Sigma_f^{-1}(I - M_f)\}}{(2\pi)^{N/2}\left|\Sigma_f\right|^{1/2}} \tag{3}$$

where M_f and Σ_f are the mean and covariance matrix of face class C_f respectively.

Using principal component analysis (PCA) [5], covariance matrix Σ_f can be factorized as follows:

$$\begin{aligned} \Sigma_f &= \phi_f \Lambda_f \phi_f^T \\ \Lambda_f &= diag\{\lambda_1, \lambda_2 \cdots, \lambda_N\} \end{aligned} \tag{4}$$

where ϕ_f is an orthogonal eigenvector matrix. Λ_f is diagonal eigenvalue matrix with the eigenvalues in decreasing order.

The Mahalanobis distance $d_f(x)$ is sufficient to describing the likelihood of function (3):

$$\begin{aligned} d_f(x) &= (I - M_f)^T \Sigma_f^{-1}(I - M_f) \\ &= (I - M_f)^T(\phi_f \Lambda^{-1}\phi_f^T)(I - M_f) \\ &= u^T \Lambda^{-1} u \end{aligned} \tag{5}$$

$$where \ \ u = \phi_f^T (I - M_f)$$

However, computing likelihood from function (3) is time consuming due to the high-dimensionality. We can estimate the likelihood using only the first M projections $M << N$, which is principle subspace indicated by PCA. Therefore, we divide ϕ_f into two parts $\{\phi_1, \phi_2, \cdots \phi_M\}$ and $\{\phi_{M+1}, \phi_{M+2}, \cdots \phi_N\}$, together with $\{\lambda_1, \lambda_2 \cdots, \lambda_M\}$ and

$\{\lambda_{M+1}, \lambda_{M+2} \cdots, \lambda_N\}$ as in [5]. The remaining N-M eigenvalues can be approximated by their average:

$$\rho = \frac{1}{N-M}\sum_{i=M+1}^{N}\lambda_i \tag{6}$$

The form of the likelihood in (3) can be rewritten as the product of two independent Gaussian densities.

$$p(I \mid C_f) \approx \frac{\exp\{-\frac{1}{2}\sum_{i=1}^{M}\frac{u_i^2}{\lambda_i}\}}{(2\pi)^{M/2}\prod_{i=1}^{M}\lambda_i^{1/2}} \cdot \frac{\exp\{-\frac{1}{2\rho}\sum_{i=M+1}^{N}u_i^2\}}{(2\pi\rho)^{(N-M)/2}} \tag{7}$$

In the same way, we model the non-face class with the non-face samples selected by bootstrap strategy:

$$p(I \mid C_n) \approx \frac{\exp\{-\frac{1}{2}\sum_{i=1}^{M}\frac{v_i^2}{\lambda_i^{(n)}}\}}{(2\pi)^{M/2}\prod_{i=1}^{M}\lambda_i^{(n)1/2}} \cdot \frac{\exp\{-\frac{1}{2\varepsilon}\sum_{i=M+1}^{N}v_i^2\}}{(2\pi\varepsilon)^{(N-M)/2}} \tag{8}$$

where

$$\begin{aligned} d_n(x) &= (I-M_n)^T \textstyle\sum_n^{-1}(I-M_n) \\ &= (I-M_n)^T(\phi_n \Lambda^{-1}\phi_n^T)(I-M_n) \\ &= v^T \Lambda^{-1} v \end{aligned} \tag{9}$$

and

$$\begin{aligned} v &= \phi_n^T(I-M_n) \\ \varepsilon &= \frac{1}{N-M}\sum_{i=M+1}^{N}\lambda_i^{(n)} \end{aligned} \tag{10}$$

3.2 The Bayesian Classifier for face detection

After modeling the face class and non-face class, we can classify the unknown patch with Bayesian discriminating rules, i.e., maximizing the posteriori probability. Denote the posteriori probability of face class $P(C_f \mid I)$ and the non-face class $P(C_n \mid I)$ [2], and then we can discriminate a patch is face or not.

$$I \in \begin{cases} C_f & \text{if} \quad P(C_f \mid I) > P(C_n \mid I) \\ C_n & \text{otherwise} \end{cases} \tag{11}$$

according to Bayesian theorem:

$$\begin{aligned} P(C_f \mid I) &= P(I \mid C_f)P(C_f)/P(I) \\ P(C_n \mid I) &= P(I \mid C_n)P(C_n)/P(I) \end{aligned} \tag{12}$$

where $P(C_f)$, $P(C_n)$ are priori probabilities of face class C_f and non-face class C_n . $P(I)$ is the mixtures density function. Then we can get:

$$\begin{aligned} &P(I \mid C_f)P(C_f) > P(I \mid C_n)P(C_n) \Rightarrow \\ &\ln P(I \mid C_f) - \ln P(I \mid C_n) > \ln[P(C_n)/P(C_f)] \end{aligned} \tag{13}$$

Combining (7), (8) with (13) we can get finally:

$$I \in \begin{cases} C_f & \text{if} \quad \theta_f + \tau < \theta_n \\ C_n & \text{otherwise} \end{cases} \tag{14}$$

where

$$\begin{aligned} \theta_f &= \sum_{i=1}^{M}\frac{u_i^2}{\lambda_i} + \ln(\prod_{i=1}^{M}\lambda_i) + \frac{1}{\rho}\sum_{i=M+1}^{N}u_i^2 + (N-M)\ln\rho \\ \theta_n &= \sum_{i=1}^{M}\frac{v_i^2}{\lambda_i^{(n)}} + \ln(\prod_{i=1}^{M}\lambda_i^{(n)}) + \frac{1}{\varepsilon}\sum_{i=M+1}^{N}v_i^2 + (N-M)\ln\varepsilon \\ \tau &= 2\ln[\frac{P(C_n)}{P(C_f)}] \end{aligned} \tag{15}$$

Here τ is an empirical parameter that estimates the probability proportion of face class and non-face class. To further reduce the false alarm rate, another control parameter χ is introduced to the discriminating function as in [2]. The control parameters are empirically chosen.

$$I \in \begin{cases} C_f & \text{if} \quad \theta_f + \tau < \theta_n \quad \text{and} \quad \theta_f < \chi \\ C_n & \text{otherwise} \end{cases} \tag{16}$$

4 Experiments

In our experiments, we collect 4000 face samples for training. Most of training samples come from FERET database “fa”, ”fb”, ”rb”, ”rc” gallery, and some are from World Wide Web. The selected faces are all frontal or nearly frontal. But there are variation in image quality and lighting. All the faces are cropped and aligned to the same size $m = 19, n = 19$. Eyes’ positions of all samples are fixed in the same coordinates.

We use these 4000 face samples to build face class model, and selected randomly about 5000 non-face samples to build a course non-face class model first. After training the classifier, we use “bootstrap” strategy to refine the negative samples. In 400 scenery images with the size 320*240, we carry the bootstrap 4 times to gather about 10,000 negative samples in all. Afterwards, we use the refined non-face samples to train the final classifier.

We evaluate the performance of our approach using two datasets. One comes from YaleB face dataset, which consists of 10 subjects. Each subject has 576 images in (9 poses and 64 illumination conditions). We select 234 images, the camera axis of which is less than 40 degrees azimuth and 50 degrees elevation. Thus we have 2340 images with pose and illumination variation. Another testset consists of images chosen from the MIT-CMU testset, in which there are 47 images containing 177 faces. We exclude some images from the original MIT-CMU testset in which there are faces with large rotation.

In order to get face patches at all scale, the pyramid strategy is adopted. The factor of downsample we use is 1.2. At each scale, not all the locations along vertical and horizontal will be scanned. Because our face detection arithmetic is not sensitive to one or two pixels shift, we can achieve a speedup by scanning the scaled images at intervals. Also this can decrease the false alarms in the images.

In YaleB testset, 2131 of 2340 faces are detected with only 52 false alarms and 167 faces are detected from 177 faces in the MIT-CMU testset with 47 false alarms. With the multi-scale scanning strategy, $3.173*10^8$ patches are approximately scanned in the YaleB testset and $1.38*10^7$ in the MIT-CMU testset. Comparing with the patches scanned in all, the false positive rate of our approach is $2.99*10^{-7}$ and the detection rate is more than 90%. Fig4 gives some examples of the result under various illumination conditions in the YaleB testset. Fig5 gives the first 10 eigenvectors of $ILBP_{4,1}$ features and $ILBP_{8,1}$ features in Fig6. Experiments demonstrate the proposed method has an encouraging performance.

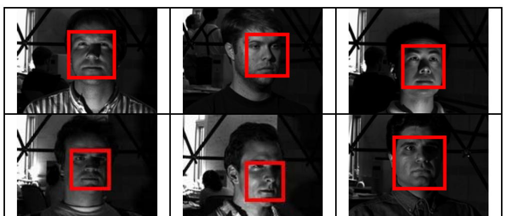

Figure 4. Experimental results in YaleB

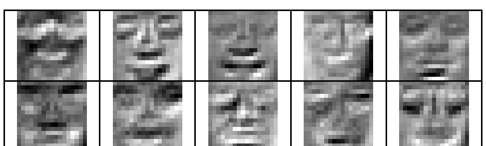

Figure 5. The first ten eigenvectors of ILBP41

Figure 6. The first ten eigenvectors of ILBP81

5 Conclusion and Future Work

In this paper, we propose a novel face detection method using ILBP features under Bayesian framework. Comparing with raw grayscale features, ILBP features are illumination invariant because it can well describe the local shape and texture information. We model the face and non-face class using the multivariable Gaussian distribution, and the Bayesian decision rule is applied to decide the patch is a face or not.

In the experiment, our approach achieves the detection rate of more than 90% with a reasonable false alarm rate under the CMU-MIT mixture dataset. In the YaleB dataset, though there exist various illuminations variations, the proposed method can also achieve a high detection rate with so few false alarms. Many false rejected ones in YaleB dataset are such images that have extreme light source, for the texture information has been lost.

In our ongoing work, we will consider to improve the computing efficiency and performance with other new classifiers and more ILBP features at different neighborhood, such as $ILBP_{8,2}$.

Acknowledgement

This work is supported by the NCFC under Grant No. 60135020 and 60121302, and the joint fund of NSFC-RGC under Grant No. 60318003.

References

[1] Ojala, T., Pietikainen, M., Maenpaa, 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

[2] C. Liu: "A Bayesian Discriminating Features Method for Face Detection", IEEE PAMI 2003 vol. 25, no. 6, pp. 725-740

[3] M.-H. Yang, D. Roth and N. Ahuja,: “A SNoW-based face detector”, Advances in Neural Information Processing Systems 12, pp.855-861, 2000.

[4] Henry A. Rowley, Shumeet Baluja, Takeo Kanade: “Neural Network-Based Face Detection” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, pp. 22–38, 1998.

[5] Baback M., “Probabilistic Visual Learning for Object Representation” PAMI1997 vlo.19 No.7

[6] Ming-Hsuan Yang, David Kriegman, and Narendra Ahuja. “Detecting faces in images: A survey.” Pattern Analysis & Machine Intelligence 24(1), January 2002, pp. 34-58

[7] E. Osuna, R. I'eund, and F. Girosi, "Training Support Vector Machines: An Application to Face Detection," IEEE CVPR, 1997.

[8] K.-K. Sung and T. Poggio, “Example-Based Learning for View-Based Human Face Detection,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 1, pp. 39-51, Jan. 1998.

[9] Paul Viola and Michael J. Jones. “Robust real-time object detection.” International Journal of Computer Vision, 2001