Demographic Classification with Local Binary Patterns

Zhiguang Yang and Haizhou Ai

Department of Computer Science and Technology,
Tsinghua University, Beijing 100084, China
ahz@mail.tsinghua.edu.cn

Abstract. LBP (Local Binary Pattern) as an image operator is used to extract LBPH (LBP histogram) features for texture description. In this paper, we present a novel method to use LBPH feature in ordinary binary classification problem. Given a restricted local patch, the Chi square distance between the extracted LBPH and a reference histogram is used as a measure of confidence belonging to the reference class, and an optimal reference histogram is obtained by iteratively optimization; real AdaBoost algorithm is used to learn a sequence of best local features iteratively and combine them into a strong classifier. The experiments on age, gender and ethnicity classification demonstrate its effectiveness.

Keywords: real AdaBoost, LBPH, demographic classification.

1 Introduction

LBP operator was introduced by Ojala in 1996 [14] for texture classification, later it was used for face recognition [1], facial expression recognition [2] and so on. In the early stage, the image is divided into several equal sized windows and represented as the combination of LBPH features from all windows; the classification methods can be nearest neighbors [1] [14] or linear programming [2]. Later, the LBPH features with various sizes and locations are used as weak classifiers, from which JS Boosting is used to learn a strong face recognition classifier by Huang et al [6]. In this paper, we use LBP histogram for demographic classification, which is age, gender and ethnicity classification, by face texture. Age is classified into three periods: child, youth and oldness; ethnicity is classified into Asian and non-Asian. We treat demographic classification as an ordinary binary classification problem, regarding age as a composition of two binary classifications. Our method is integrated with face detection [5] to form an automatic system.

Human face represents a variety of information, such as identity, age, gender, ethnicity, expression and so on. Specifically for one person, gender and ethnicity always remain the same as his or her identity does, while his or her age changes. As a result, it is reasonable that age classification is even harder. Recent works on demographic classification can be divided into two main approaches. One is pattern classification by face texture, such as decision tree [3], SVM [4][9],

S.-W. Lee and S.Z. Li (Eds.): ICB 2007, LNCS 4642, pp. 464–473, 2007.

real AdaBoost [15] and so on; the other focuses on both shape and texture information, such as dealing with 2D shape and wrinkled texture in [7] and 3D structure in [8]. In this paper, we treat LBPH as a distribution description of given local texture patch, and partition the feature domain of Chi square distance between this histogram and a reference histogram to form a LUT (look up table) based weak classifier. Within the set of all promising local patches, real AdaBoost is used to select the best ones and construct a strong classifier.

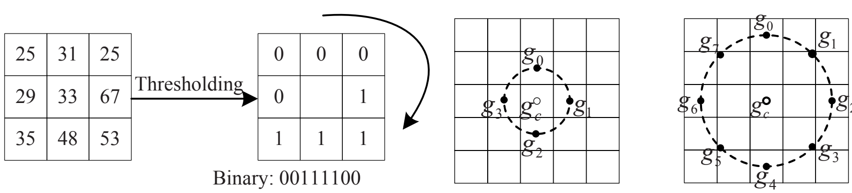
This paper is organized as follows: the following section describes the LBP histogram feature, real AdaBoost algorithm and the organization of age classification; the experiment on a large snapshot image database is shown in section 3; and section 4 gives the conclusion finally.

2 Demographic Classification

We combine face detection [5] and our demographic classification algorithm into an automatic system. After face detection, we detect six facial feature points, four eye corners and two mouth corners; and normalize the face texture by affine transformation based on the correspondence between the six points and the frontal template shape. In the normalized face texture, real AdaBoost selects a sequence of local LBPH features, and combines them into a strong binary classifier. The gender, ethnicity and age classifiers are trained respectively.

2.1 LBPH Feature

Basic LBP operator [14] is a computational efficient operator. Taking each pixel as a threshold, the operator transferred its 3×3 neighborhood into a 8-bit binary code, as shown in Fig.1(a). Later in [10], LBP operator is extended that an arbitrary number of bilinear interpolated pixels on a circle with arbitrary size are used as neighbor pixels, instead of its 3×3 neighborhood, as shown in the Fig.1(b).



(a) Basic LBP operator (b) Extended LBP operator

Fig. 1. LBP operators

Another extension in [10] to the basic LBP operator is the so called *uniform* LBP that is found to be the fundamental property of local image texture. A LBP is called uniform if there are no more than two 0/1 or 1/0 bitwise transitions in its binary code, being considered as a circular code. The extended LBP operator is donated as $LBP_{P,R}^{u2}$, where P is the number of bilinear interpolated pixels, R is the radius of the neighbor circle and $u2$ stands for uniform criterion.

It is reported in [10] that, the contribution of uniform pattern to $LBP_{8,1}^{u2}$ and $LBP_{16,2}^{u2}$ is about 87.2% and 70.7% respectively. That is to say, the uniform patterns take a majority percentage of all patterns. As a result, each uniform pattern is given a unique label and all other minorities are given a mutual label in histogram calculation. In this paper, $LBP_{8,2}^{u2}$ is used, and all the LBP feature value is quantified into $[1, 59]$ by uniform criterion.

Given a local texture patch, LBPH is used to summarize all pixels (except pixels on the border of the image) with each pixel represented by the quantified LBP label.

2.2 Construct a Weak Classifier

Suppose S and M are two different histograms, the Chi square distance can be defined as:

$$\chi^2(S, M) = \sum_{i=1}^{n} \frac{(S_i - M_i)^2}{S_i + M_i}, \tag{1}$$

where n is the number of elements in the histogram ($n = 59$ in this paper).

Chi square distance is an effective measurement of similarity between a pair of histograms, hence it is suitable for nearest neighbor [10] and intra/extra based classification [6]. However, find a pair of similar samples doesn't make sense in most binary classification problem.

In this section, we want to find an optimal template histogram M as the reference template for every positive or negative sample to calculate a Chi square distance for ordering. With the hypothesis of Gaussian distribution of positive and negative samples respectively, the template M is expected to be near to one cluster and far from the other, in order that positive and negative samples can be discriminated successfully by Chi square distance.

Via Chi square distance, the separability of each LBPH feature could be measured as Fisher's discriminant ratio:

$$FDR = \frac{(\mu_+ - \mu_-)^2}{\sigma_+^2 + \sigma_-^2}, \tag{2}$$

where μ and σ donate the mean value and variance value of samples' Chi square distance to the template M. In order to find the optimal template histogram, we first initialize M as the mean value of positive histograms, and then use steep decent method to find an optimal solution.

Given sample set $S = (h_1, y_1), \ldots, (h_m, y_m)$, where h_i is the LBPH feature, and $y_i = \pm 1$ is the class label. μ_y and σ_y of positive and negative samples illustrated by $y = 1$ and $y = -1$ respectively can be written as:

$$\mu_y = \frac{1}{N_y} \sum_{k=1, y_k = y}^{m} \chi^2(h_k, M), \tag{3}$$

$$\sigma_y^2 = \frac{1}{N_y} \sum_{k=1, y_k=y}^{m} (\chi^2(h_k, M) - \mu_y)^2. \tag{4}$$

For the n-elements template histogram, whose items sum up to 1, the previous $n-1$ items are supposed to be independent, and their partial derivative of μ_y and σ_y can be written as:

$$\frac{\partial \mu_y}{\partial M_i} = \frac{1}{N_y} \sum_{k=1, y_k=y}^{m} \frac{-4h_k^2(k)}{(h_k(i) + M_i)^2} + 1, \tag{5}$$

$$\frac{\partial \sigma_y^2}{\partial M_i} = \frac{2}{N_y} \sum_{k=1, y_k=y}^{m} (\chi^2(h_k, M) - \mu_y)(\frac{-4h_k^2(k)}{(h_k(i) + M_i)^2} + 1 - \frac{\partial \mu_y}{\partial M_i}). \tag{6}$$

As a result, the gradient of Fisher discriminate ratio can be calculated as follow

$$\frac{\partial FDR}{\partial M_i} = \frac{2(\mu_+ - \mu_-)}{\sigma_+ + \sigma_-}(\frac{\partial \mu_+}{\partial M_i} - \frac{\partial \mu_-}{\partial M_i}) - \frac{(\mu_+ - \mu_-)^2}{(\sigma_+ + \sigma_-)^2}(\frac{\partial \sigma_+^2}{\partial M_i} + \frac{\partial \sigma_-^2}{\partial M_i}). \tag{7}$$

By this means, an optimal M could be found by iterative search, and it is used as the reference template for the given local texture patch. We divide the main feature domain of Chi square distance from training samples to the reference template into 32 bins, and Fig.2 gives an example distribution of samples' Chi square distance to the reference histogram, which meets the prior hypothesis of Gaussian distribution well.

With respect to training samples' distribution on the Chi square distance, a LUT based weak classifier is used. Its output on each domain can be defined as:

$$\forall h \in H_i, f(h) = \frac{1}{2}(\frac{W_+^i + \varepsilon}{W_-^i + \varepsilon}), i = 1, \ldots, 32, \tag{8}$$

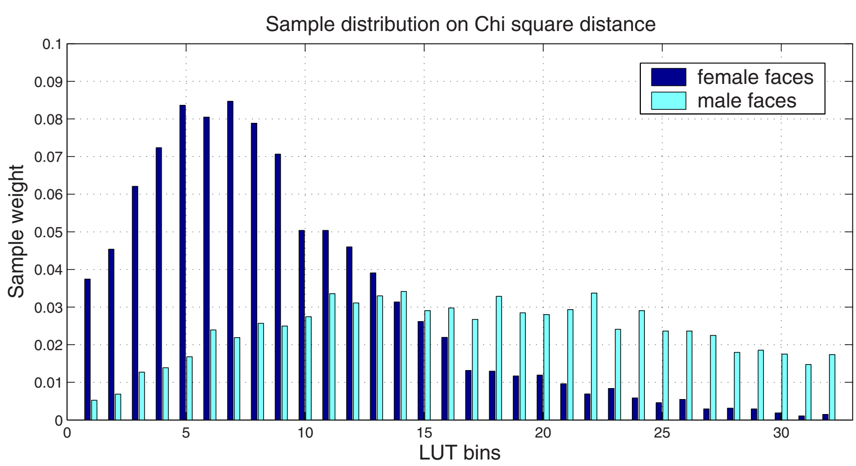


Fig. 2. Sample distribution on Chi square distance

where W_i is the sum weight of samples H_i on the i-th domain and ε is a small positive constant.

2.3 Real AdaBoost Learning

Real AdaBoost [12] is a statistical learning algorithm by maximizing classification margin iteratively. In each iteration t, one weak classifier is selected from a large hypothesis space; after the final iteration T, all the selected weak classifiers are combined together to construct a strong classifier. The training algorithm is shown in Fig. 3.

Given Training sample set $S = \{(x_1, y_1), \ldots, (x_m, y_m)\}$ where x_i is the face sample, $y_i = \pm 1$ is the class label. All promising local texture LBPH features L.

Initialize Sample weight $w_1(i) = \frac{1}{2N_{y_i}}$, $i = 1, \ldots, m$

Prepare Find a reference template histogram for each LBPH feature

For $t = 1, \ldots, T$

1. Under distribution w_t, select a weak classifier f_t to minimize the factor

$$Z = \sum_{j=1}^{32} \sqrt{W_+^j W_-^j}.$$

2. Update the weights:

$$w_{t+1}(i) = w_t(i) \frac{exp(-y_i f_t(x_i))}{U_t}, i = 1, \ldots, m,$$

where U_t is a normalization factor so that w_{t+1} is a p.d.f.

Output The final strong classifier: $F(x) = sign(\sum_{t=1}^{T} f_t(x) - b)$,where b is a threshold whose default is 0. And the confidence of the classifier is defined as $\sum_{t=1}^{T} f_t(x) - b$.

Fig. 3. Real AdaBoost algorithm

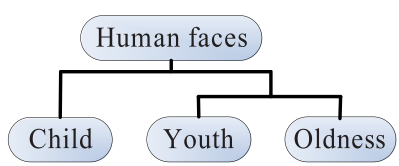


Fig. 4. Organization of age classification

There are also other variations of AdaBoost algorithms for multi-class classification, such as AdaBoost.MH [12]. However, an experiment study of age classification shows that, the three age periods don't share the same optimal separable features and the difficulty of classification differs as well. Child is the easiest to be discriminated, and the second one is oldness. As a result, a binary tree structure is used for age classification. That is, the first node is used to classify child and adult, the second one is used to discriminate old people from youth, as shown in Fig. 4.

3 Experimental Result

We carry out experiment on three different databases: FERET[11], PIE[13] and a snapshot database. All frontal images in FERET which account for 1196 individuals and 3540 images are used, and 68 individuals with 696 good illumination frontal PIE images are also used; the snapshot image database contains 9000 Chinese snapshot images. The three databases contain three kinds of expression, with glasses or without.

We use the snapshot database for five-fold cross validation on age and gender classification, FERET and PIE are used for independent database testing. When we do experiment on ethnicity classification, FERET and snapshot images are merged for cross validation, and PIE is left for testing.

All the faces are normalized automatically as described in section 2. To show the effectiveness of our approaches, the experimental result of the method (represented as LBPH* hereinafter) that takes the mean histogram of positive samples as the reference template histogram is also given in the three sub-sections for comparison. Meanwhile, Real AdaBoost with Haar like features is also used for comparison.

3.1 Gender Classification

There are 4696 male and 3737 female faces in the snapshot images (children are not used for gender classification), they are always upright and neutral without full beard or strange hairstyle. In each round of five fold cross validation experiment, 80% faces were used for training, and the others were used for testing. Besides cross validation, we also use all the snapshot images for training, and FERET and PIE database for testing. For a comparison, we additionally implement RBF-kernel SVM with ensemble face textures, in which PCA coefficients with 95% energy are used for dimension reduction. Tab.1 lists the experiment result of gender classification on cross validation and independent database verification.

SVM leads in the cross validation experiment, but is poor in generalization ability, as it suffers from improper representation by eigenvectos generated from independent database. LBPH is comparable to Haar like feature in the cross validation experiment, and slightly better in independent verification. Meanwhile, they are all much better than LBPH* feature.

3.2 Ethnicity Classification

In the experiment of ethnicity classification, snapshot database and FERET database are merged, that results in 11680 Asian characters and 1016 non-Asian

Table 1. Experiment result (error rate) on gender classification

	Cross validation	FERET	PIE
SVM	3.18%	19.4%	24.2%
Real AdaBoost(Haar)	3.58%	8%	12.3%
Real AdaBoost(LBPH*)	6.52%	11.9%	16.8%
Real AdaBoost(LBPH)	3.68%	6.7%	8.9%

characters. In each round of cross validation, a 80% number of characters were used for training, and the rest were used for testing, although some characters have more than one images in the integrated database. Tab.2 gives the comparison result of multiple methods on cross validation and independent database verification.

Table 2. Experiment result (error rate) on ethnicity classification

	Cross validation	PIE
Real AdaBoost(Haar)	2.98%	7.9%
Real AdaBoost(LBPH*)	7.11%	14.6%
Real AdaBoost(LBPH)	3.01%	6.8%

LBPH feature is again confirmed to be comparable to Haar like feature in the cross validation experiment, and slightly better in independent verification.

3.3 Age Classification

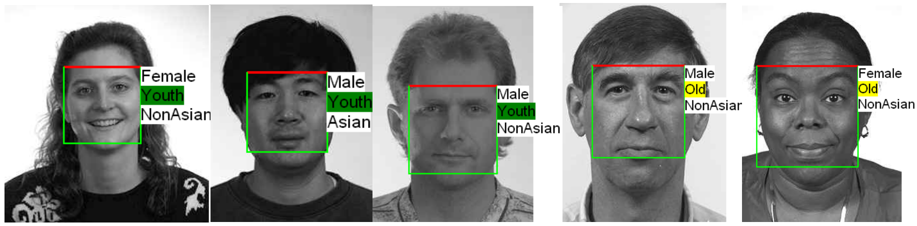
There are 567 children, 6951 youth and 1482 oldness in the snapshot images. We first train a children classifier to separate children from adult people, and then train an oldness classifier to separate oldness from youth. FERET and PIE databases are also used for testing. The experiment results are shown in Tab. 3.

Table 3. Experiment result (error rate) on age classification

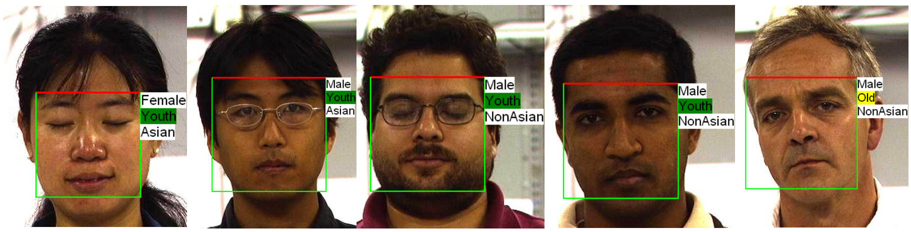
	Cross validation	FERET	PIE
Real AdaBoost(Haar)	10.8%	20.6%	23.3%
Real AdaBoost(LBPH*)	8.18%	18.9%	20.4%
Real AdaBoost(LBPH)	6.82%	7.88%	12.5%

The result corresponds with the above two experiments, and the result of LBPH is better than that of Haar like feature in cross validation and independent verification.

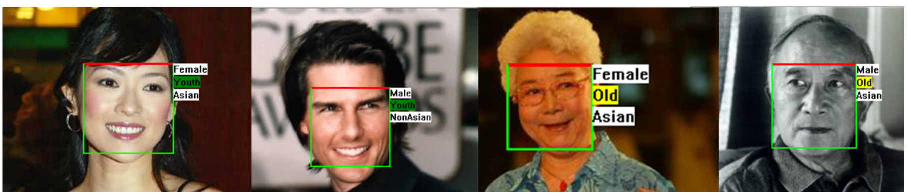
The result shown in the above three sub experiments proves effectiveness of LBPH feature. Via iterative optimization, LBPH makes much better result than LBPH*, and maintains a good generalization ability as well. The LBPH feature is confirmed to be comparable to the Haar like feature in gender and



(a) Result on FERET images

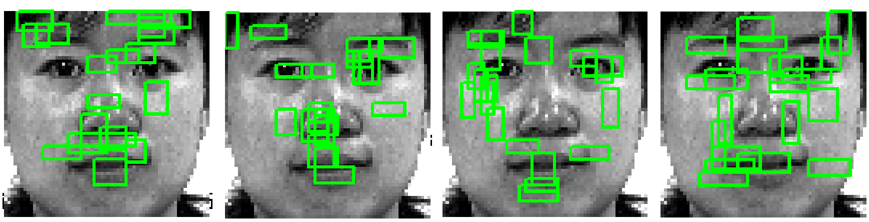


(b) Result on PIE images



(c) Result on www images

Fig. 5. LBP operators



(a) gender (b) ethnicity (c) child (d) oldness

Fig. 6. Example of boosted weak demographic classifiers

ethnicity classification, and much better in age classification. Meanwhile, each classifier used above is composed of less than 100 weak classifiers, which can be used for realtime demographic classification. Fig.5 gives some demographic classification results, and Fig.6 shows the first 20 boosted weak classifiers by three experiments respectively. As a semantic explanation, male faces are distinguished by the areas of pronounced brow bossing and longer philtrum; non-Asian faces are characterized by deeper eyehole, different shapes of nose and lip; children are featured by fat cheeks and tender skin; and oldness are recognized by wrinkles in multiple areas.

4 Conclusion

In this paper, we have proposed a new method to use LBPH feature for ordinary binary classification problems. For each LBPH feature, a sample's Chi Square distance to a reference template is used as a measurement of confidence for classification; the positive mean histogram is used as the initialization, and the steep decent method is used to find an optimal reference template. Real AdaBoost is used to train a strong classifier by composing a sequence of LBPH features. The experiments on gender, ethnicity and age classification prove our method's effectiveness. However, there is some work to be further studied. For example, it could be better if color image instead of gray level image is used for ethnicity classification. And LBP operator could be more effective if we find better parameters, or use a more efficient face texture normalization method.

Acknowledgments

This work is supported in part by National Science Foundation of China under grant No.60332010, No.60673107.

References

1. Ahonen, T., Hadid, A., Pietikainen, M.: Face recognition with local binary patterns. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3021, pp. 469–481. Springer, Heidelberg (2004)
2. Feng, X., Cui, J., Pietikainen, M., Hadid, A.: Real time facial expression recognition using local binary patterns and linear programming. In: MICAI, pp. 328–336 (November 2005)
3. Gutta, S., Wechsler, H., Phillips, P.J.: Gender and ethnic classification of face images. In: FG, pp. 194–199 (April 1998)
4. Hosoi, S., Takikawa, E., Kawade, M.: Ethnicity estimation with facial images. In: FGR, pp. 195–200 (May 2004)
5. Huang, C., Ai, H., Lao, S.: Boosting nested cascade detector for multi-view face detection. In: ICPR, pp. 415–418 (August 2004)
6. Huang, X., Li, S.Z., Wang, Y.: Jensen-shannon boosting learning for object recognition. In: CVPR, pp. 144–149 (June 2005)
7. Kwon, Y.H., da Vitoria Lobo, N.: Age classification from facial images. Computer Vision and Image Understanding 74(1), 1–21 (1999)
8. Lu, X., Chen, H., Jain, A.K.: Multimodal facial gender and ethnicity identification. In: ICB, pp. 554–561 (January 2006)
9. Moghaddam, B., Yang, M.-H.: Learning gender with support faces. IEEE Trans. Pattern Anal. Mach. Intell. 24(5), 707–711 (2002)
10. Ojala, T., Pietikainen, M., Maenpaa, M.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. 24(7), 971–987 (2002)

11. Phillips, P., Moon, H., Rizvi, S.A., Rauss, P.J.: The feret evaluation methodology for face-recognition algorithms. IEEE Trans. PAMI 22(10), 1090–1104 (2000)
12. Schapire, R.E., Singer, Y.: Improved boosting algorithms using confidence-rated predictions. Machine Learning 37(3), 297–336 (1999)
13. Sim, T., Baker, S., Bsat, M.: The cmu pose, illumination, and expression (pie) database. In: FGR, pp. 53–58 (May 2002)
14. Ojala, T., Pietikainen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distribution. Pattern Recognition 29(1), 51–59 (1996)
15. Wu, B., Ai, H., Huang, C.: Lut-based adaboost for gender classification. In: AVBPA, pp. 104–110 (June 2003)