

Available online at <http://www.mecspress.net/ijem>

Combining Local Binary Patterns and Visual Attention for Face Recognition

Zhiyong Gao^{a,*}, Haihua Liu^a, Xinhao Chen^a

^a College of Biomedical Engineering, South-Central University for Nationalities, Wuhan, China

Abstract

Effectiveness of local binary pattern (LBP) for face recognition has been proven. But the weight of weighted LBP is difficult to determine. In this paper, we proposed a biologically plausible approach to set the weight automatically. Combining LBP and visual attention, a weight map can be constructed by summing over the saliency map. The weight map outlines salient information in the image and helpful for recognition. Experimental results show that the presented method is efficient and effective.

Index Terms: face recognition; local binary pattern; visual attention; saliency map

© 2011 Published by MECS Publisher. Selection and/or peer review under responsibility of the Research Association of Modern Education and Computer Science.

1. Introduction

Face recognition is an interesting and challenging problem, and impacts important applications in many areas. An important issue in face recognition is feature extraction. There are two common approaches to extract facial features: geometric feature-based methods and appearance-based methods [1, 2]. Geometric features present the shape and locations of facial components, which are extracted to form a feature vector that represents the face geometry. However, geometric properties are not reliable for face recognition because rich information contained in the facial texture or appearance is discarded. The appearance-based approach, such as PCA (2D PCA, KPCA), LDA, Gabor-wavelet, and Local Binary Pattern (LBP) based methods, are applied to either the whole-face or specific face-regions to extract the appearance changes of the face. Due to their superior performance, the major works on feature extraction methods have focused on using appearance-based approach.

Among appearance-based methods, using LBP for face description is one of the most effective methods. The LBP operator was originally designed for texture description and recently has been introduced to represent faces in facial images analysis [3-5]. The main advantages of LBP are its invariance to illumination

* Corresponding author.

E-mail address: zhiyonggao@mail.scuec.edu.cn

changes and computational efficiency. It has proven to be highly discriminative and has been widely used in many applications.

In application, it has been demonstrated that weighted LBP outperforms non-weighted LBP both in processing time and recognition rate [3]. Therefore, weighted LBP is used more widely [4, 5]. Appropriate weight for different windows is important for face recognition. But the weights of different regions are difficult to find. Shan et al. designed the weight set empirically based on observation [4]. Ahonen et al. set the weights by the recognition rate of each window. Lei et al. using Fisher criterion to compute the weight sets [5]. These methods are either arbitrary [4] or can't represent the shape and texture information in original image [3, 5]. Weighted LBP is introduced to differentiate the useful information of each region, especially the different edges and corners, so it is more plausible to set the weight of each region according to its shape and texture information.

In this work, we introduce saliency map of visual attention to approximate the shape information of each window and compute corresponding weight. Saliency map of visual attention model topographically codes for local conspicuity over the entire image, only locations which locally stand out from their surround can persist [6]. Different shapes and textures information in the image can then be outlined clearly. Therefore, although saliency map is usually used to locate focus of attention, it is obvious that it also can be used to represent the shape information of the entire image, which is related to the different weights of windows in LBP.

2. Local Binary Patterns

A. LBP Operators

The LBP operator labels every pixels of an image by thresholding a certain neighborhood of each pixel with the center pixel value and transforming the result into a binary number. The binary numbers of each pixel are named as Local Binary Patterns and contain information about the distribution of the local patterns, such as edges, spots and flat areas, over the whole image on a pixel-level.

The size of neighborhood of LBP can be different to capture the information of different scale structures. Usually, circular neighborhoods and bilinear interpolation are used to allow any radius and number of pixels in the neighborhood, which can be represented as the notation (P, R). The notation means that P sampling points evenly spaced on a circle of radius of R. Denoting the binary number of pixel at (x, y) as $B(x, y) = (b_{p-1}, \dots, b_0)$, where $b_i \in \{0, 1\}$, the decimal code can be calculated:

$$LBP_{P,R}(x, y) = \sum_{i=0}^{P-1} b_i 2^i \quad (1)$$

For efficient face representation, the face image is divided into several regions to produce regional information by counting the frequencies of each value of LBP codes in each region, i.e. LBP histograms are calculated for each region to generate spatially enhanced histogram. Finally, the regional LBP histograms are concatenated into a single vector. As a result, the face is represented by the LBP and the shape of the face is recovered by the concatenation of different local histograms.

An important extension of LBP operator is uniform LBP. A LBP is called uniform if and only 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 (0 transition), 00000001 (1 transition) and 00000010 (2 transitions) are uniform patterns, while 00001010 (4 transitions) and 00101010 (6 transitions) are not uniform ones. It is observed that uniform patterns account for most patterns in texture images. In practice, each uniform LBP is cast into a unique bin of the histogram and all remaining non-uniform LBP are accumulated into a single bin, which is often denoted as $LBP_{P,R}^{u2}$.

Given a face image to be classified, LBP feature are extracted firstly, and different machine learning techniques, such as template matching, Support Vector Machines (SVM), Linear Discriminant Analysis (LDA)

and Neural Network, are used to recognize faces. Different classifiers can lead to different recognition results. In this work, we use template matching to classify face for its simplicity.

In practice, a template in the form of LBP feature vector is generated for each class of face images, a nearest-neighbor classifier is used to match the input image with the closest template. In training stage, the histograms of face images in a given class were averaged to form a template for this class. There are several dissimilarity measures for nearest-neighbor classifier, for example, Histogram intersection Log-likelihood statistic and Chi square statistic [3]. We choose the Chi square statistic as the dissimilarity measure for histograms:

$$\chi^2(S, M) = \sum_i \frac{(S_i - M_i)^2}{(S_i + M_i)} \quad (2)$$

where S and M are two LBP histograms to be compared, i is ith bin in histogram.

B. Weighted LBP

In fact, the LBP descriptor for face is composed by regional LBP histograms. Each element in the last feature corresponds to a certain region of the face. The psychophysical findings indicate that some facial features contain more useful information and play more important roles in human face recognition than other features. For example, facial features contributing to face appearance mainly lie in regions such as eye and mouth regions.

Therefore, it can be expected that some of the facial regions contribute more than others in recognition. Utilizing this assumption, weights can be set to corresponding window based on the importance of the information they contain. Then, the weighted dissimilarity measure (2) can be modified:

$$\chi_w^2(S, M) = \sum_{i,j} w_j \frac{(S_{i,j} - M_{i,j})^2}{(S_{i,j} + M_{i,j})} \quad (3)$$

in which S and M are the two histograms, i and j refer to ith bin in histogram corresponding to the jth region, w_j is the weight for region j.

To find the weight w_j , different procedures were adopted. There are two kinds of approaches to set the weight: directly appointing weights of each window [4] based on observation and setting the weights according to discriminative information of each window [3, 5]. Obviously, the former method is arbitrary and the later can't fully represent the discriminative information of each window. Essentially, the information in different regions may be different greatly. For recognition, the unique information of each class is very important. Although which kind of information is more important is unknown, conspicuity locations in the image are more easily to be focused and more important for recognition has been observed in physiological studies as well as human psychophysics [6-8]. Consequently, the information contained in conspicuity regions should be paid more attention. Especially, for face recognition using LBP, the salient location should be assigned great weight. In this way, the useful information can be evaluated by corresponding saliency and the weights w_j can be determined. To compute the conspicuity features, some researchers have proposed computational visual attention models [6, 8, 9].

3. Visual Attention and Saliency Map

Visual attention means the ability to process of selecting and gating visual information based on saliency in the image itself and on prior knowledge about scenes or a particular task. People have investigated how the concepts of human selective attention can be exploited for computational systems and many computational models have been presented to locate focus of attention. Most of them model attention as a hierarchy of operations including feature extraction, saliency map generation and attention location.

During the attention procedure, the generation of saliency map is crucial. The saliency map highlights the most important visual information to be processed further, which can be used to represent the location where useful information is contained in the image. If the saliency map is divided into several regions and then

summed over each region, then the conspicuity information contained in each region can be expressed, as can be used to set the weight sets of LBP.

A. Saliency Map

Our computational attention model is based on the implementation of Itti et al. [6]. The effectiveness of the model has been demonstrated. The output of the model is a pair of coordinates in the image corresponding to the most salient location [7, 8].

The model subsampled input image into Gaussian pyramid. Each pyramid level is decomposed into channels for red (R), green (G), blue (B), yellow (Y), intensity (I) and local orientation (O_θ). With the r , g and b being the red, green, and blue channels of the input image, the intensity image I is obtained as $I=(r+g+b)/3$. After r , g and b normalized by I , they are used to create four color channels R, G, B, Y: $R=r-(g+b)/2$, $G=g-(r+b)/2$, $B=b-(r+g)/2$, and $Y=(r+g)/2-r-g/2-b$ (negative values are set to zero). Local orientation channel O_θ is obtained I by applying oriented Gabor filters to the intensity pyramid I .

Feature maps are computed by applying center-surround operations on each channel of Gaussian pyramids. Let \odot denotes the center-surround differences between a center level c and a surround level s , the feature maps are constructed:

$$I(c, s) = |I(c) - I(s)| \quad (4)$$

$$RG(c, s) = |(R(c) - G(c)) \odot (R(s) - G(s))| \quad (5)$$

$$BY(c, s) = |(B(c) - Y(c)) \odot (B(s) - Y(s))| \quad (6)$$

$$O(c, s, \theta) = |O(c, \theta) \odot O(s, \theta)| \quad (7)$$

Normalized by a normalization operator $N(\cdot)$, feature maps are combined into conspicuity maps using across-scale addition (denoted by \oplus):

$$\bar{I} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c=4} N(I(c, s)) \quad (8)$$

$$\bar{C} = \bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c=4} [N(RG(c, s)) + N(BY(c, s))] \quad (9)$$

$$\bar{O} = \sum_{\theta \in \{0, 45, 90, 135\}} N \left[\bigoplus_{c=2}^4 \bigoplus_{s=c+3}^{c=4} N(O(c, s, \theta)) \right] \quad (10)$$

The three conspicuity maps \bar{I} , \bar{C} and \bar{O} are also normalized by $N(\cdot)$ and combined into a single saliency map S :

$$S = \frac{1}{3} [N(\bar{I}) + N(\bar{C}) + N(\bar{O})] \quad (11)$$

The location with the maximum of the saliency map represents the most salient part in the image and contains unique information. Subsequently, winner-take-all strategy and inhibition of return are adopt to locate the focus of attention and select the next most salient location.

B. Setting the weights of LBP

The purpose of the saliency map is to descript the saliency or conspicuity at every location in the visual field by a scalar quantity. The model mimic the processing procedure of primate early vision and select unique information in the scene to be further processed such as object search, object recognition and scene understanding.

Since the saliency of every location in the image is calculated, we can just sum over a region to get the whole saliency of this region S_j . In this way, the unique information of different regions can be extracted in terms of distinguishing between people. For simplicity, we directly set $w_j = S_j$. In dosing so, the saliency map can be looked as weight map for weighted LBP and can be used for face recognition. Fig. 1 shows an example of facial image and its saliency map.

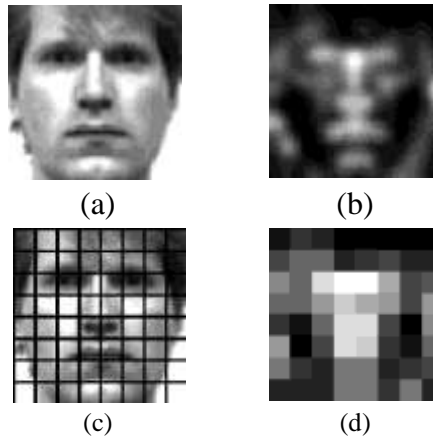


Fig. 1. (a) An example of a facial image; (b) its saliency map S ; (c) the image divided into 8×8 windows; (d) saliency map summing over regions, which can be treated as weights set for weighted LBP. The lighter the area, the bigger the weight is.

Fig. 1 demonstrates that saliency map can outline important regions seem to be important cues in face recognition such as eyes, nose and mouth. Moreover, edges related to the cheek and jaw are also highlighted, which may be helpful for recognition. At the same time, the changes of weights of different regions are not extremely abrupt so that the contribution of each region changes gradually.

4. Experimental Results

The Yale database is used to assess the recognition performance of weighted LBP with visual attention method. The Yale Face Database contains 165 grayscale images of 15 individuals. There are 11 images per subject. Each person has different facial expression or configuration: center-light, glasses, happy, left-light, no glasses, normal, right-light, sad, sleepy, surprised, and wink. All the face images are has fixed size of 100×100 pixels. Except resizing the face images to 64×64 pixels directly, no further preprocessing such as removing illumination changes and alignment of mouth is performed.

In the experiments, five samples of each person chose randomly are used to form the training set, and the remaining samples are used for testing. For each pair of parameters, recognition procedure was repeated twenty times and the average recognition results on the test sets are reported.

A. LBP Parameters

There are some parameters of LBP that can affect the performance greatly. The first is the types of LBP operator. In the preliminary experiments, uniform pattern with a neighborhood of 8 pixels and radius of 2, i.e. $LBP_{8,2}^u$, is the best tradeoff between recognition rate and processing time.

Another parameter is the division of the images into regions. A large number of small regions produces long feature vectors causing high memory consumption and slow classification, whereas using large regions causes more spatial information to be lost. The best result is yielded when the image is divided into 8×8 pixels regions (8×8 windows). The recognition rate is 85.39%. These parameters are fixed in the following experiments.

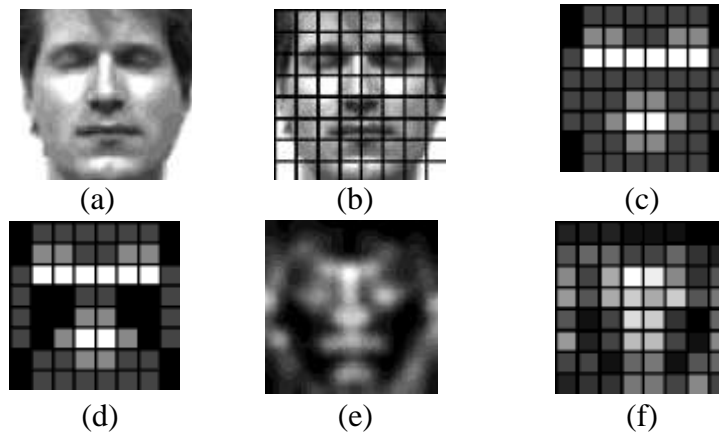


Fig. 2. (a) An example of a facial image; (b) the image divided into 8×8 windows; (c) The weights set for DA-weighted LBP. Black squares indicate weight 0, dark grey 1, light grey 2 and white 4; (d) The weights set for RR-weighted LBP. Also black squares indicate weight 0, dark grey 1, light grey 2 and white 4; (e) corresponding saliency map; (f) VA-weighted LBP, the intensity is proportional to the sum of saliency in the region.

B. Weighted LBP

There are three approaches to set the weights of LBP are compared in our experiment. They are directly assigning (denoted DA-weighted below), setting the weights according to recognition rate of each region (denoted RR-weighted below) and summing the saliency map of visual attention as weight map (denoted VA-weighted below).

The DA-weighted set the weight based on the observation. The weights of mouth and eyes regions are 4, 2, 1 and 0 are set to other regions as shown in Fig. 2(c). The RR-weighted set the weight according the recognition rates of corresponding windows. Then the windows whose rate lay below the 0.2 percentile of the rates got weight 0 and windows whose rate lay above the 0.6 and 0.7 percentile got weights 2.0 and 4.0, respectively. The other windows got weight 1.0. VA-weighted construct a weight map and sum over corresponding regions to get the weight and the weight is proportional to the sum of saliency in the region. Fig. 2 shows the different weight of the same image.

The classification accuracies are shown in Table I. As expected, VA-weighted LBP clearly outperforms the other weighted methods and non-weighted LBP. As has been noted, this is because the VA-weighted approach describes the salient information which is useful to discriminate different shapes and textures in the image, while other weighted methods can't capture these small changes so sensitively.

On the other hand, all weighted methods yield better results than non-weighted LBP. This is in accordance with other researcher's conclusions and shows that weighted LBP is more effective than non-weighted LBP for face recognition.

Table 1 The recognition rates of LBP with different weight setting methods

Method	Recognition Rate (%)
Non-weighted	85.39
DA-weighted	86.22
RR-weighted	87.78
VA-weighted	89.56

5. Conclusions

This paper presented a novel approach to set the weight set of weighted LBP automatically. Motivated by the observations in physiology and psychophysics, we combine LBP and visual attention to construct weight map. The weight map represents unique information contains in each region which is helpful for face recognition. Experimental results on the Yale database demonstrate that the proposed method is more efficient and effective than the two methods.

References

- [1].W. Zhao, R. Chellappa, A. Rosenfeld, P. Phillips, "Face recognition: A literature survey," *ACM Comput. Surv.* 2003,35 (4):399–458.
- [2].S. Li, A. Jain, "Handbook of face recognition", Springer,2005.
- [3].T. Ahonen, A. Hadid and M. Pietik änen, "Face recognition with local binary patterns", *European Conference on Computer Vision (ECCV)*, 2004:469-481.
- [4].C. Shang, S. Gong, and P. McOwan Ahonen, M. Pietik änen, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, 2009, 27: 803-816.
- [5].Z. Lei, S. Liao, M. Pietik änen and S. Li, "Face recognition by exploring information jointly in space, scale and orientation," *IEEE Transactions on Image Processing*, 2011, 20(1):247-256.
- [6].L. Itti, C. Koch, and E. Niebur, "A model of saliency based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligent*, 1998, 20(11):1254–1259.
- [7].S. Chikkerur, T. Serre, and T. Poggio, "Attentive processing improves object recognition," *MIT Technical Report* , 2009, MIT-CSAIL-TR-2009-046.
- [8].L. Elazary, L. Itti, "A Bayesian model for efficient visual search and recognition," *Vision Research*, 2010, 50(14): 1338-1352.
- [9].D. Walther, C. Koch, "Attention in hierarchical models of object recognition," *Progress in Brain Research*, 2007, Vol. 165, 57-78.