Gabor Volume based Local Binary Pattern for Face Representation and Recognition

Zhen Lei[1]   Shengcai Liao[1]   Ran He[1]   Matti Pietikäinen[2]   Stan Z. Li[1]

[1]Center for Biometrics and Security Research & National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences,
95 Zhongguancun Donglu, Beijing 100190, China.
{zlei,scliao,rhe,szli}@cbsr.ia.ac.cn
[2]Machine Vision Group, University of Oulu, PL4500, FI-90014 Oulun yliopisto, Finland.
mkp@ee.oulu.fi

Abstract

This paper presents a novel face representation and recognition approach. The face image is first decomposed by multi-scale and multi-orientation Gabor filters and local binary pattern (LBP) analysis is then applied on the derived Gabor magnitude responses. Different from [9], the present method not only describes the neighboring relationship in spatial domain, but also exploit those between different scales (frequency) and orientations. Specifically, we first reformulate the Gabor magnitude responses as a 3rd-order volume and then apply LBP analysis on three orthogonal planes of the Gabor volume, named GV-LBP-TOP in short, in a hope to encode sufficient information for face representation. Further, a computationally effective version, E-GV-LBP, is proposed to depict the neighboring changes in spatial, frequency and orientation domains simultaneously. Finally, the weighted histogram intersection metric is utilized to measure the dissimilarity of faces. Experimental results on FERET and FRGC ver 2.0 databases show the significant advantages of the proposed method.

1. Introduction

Face recognition has attracted much attention due to its potential values for applications as well as theoretical challenges. The face images are usually affected by different expressions, poses, occlusions and illuminations which result in the differences of faces from the same person usually larger than those of faces from different persons. Therefore, how to extract robust and discriminant features which make the intra-person faces compact and enlarge the margin among different persons becomes a critical and difficult problem in face recognition.

Up to now, many representation approaches have been introduced, including subspace based holistic features and local appearance features. In the former category, typical ones include the well known Principal Component Analysis (PCA), Linear Discriminate Analysis (LDA) [2], Independent Component Analysis (ICA) [3] etc. PCA provides an optimal linear transformation from the original image space to an orthogonal eigenspace with reduced dimensionality in sense of the least mean square reconstruction error. LDA seeks to find a linear transformation by maximizing the ratio of between-class variance and within-class variance. ICA is a generalization of PCA, which is sensitive to the high-order relationships among the image pixels.

Local appearance features, as opposed to holistic features such as PCA, LDA and ICA, have certain advantages in that they are more stable to global changes such as illumination, and inaccurate alignment. Gabor wavelets [8, 5] are used to capture the local structure corresponding to spatial frequency (scale), spatial localization, and orientation selectivity which are proved to be discriminative and robust to illumination and expression changes. Local binary pattern (LBP) [1] which describes the neighboring changes around the central point, is a simple yet effective way to represent faces. It is invariant to monotone transformation and hence is somewhat robust to illumination changes.

Recently, Zhang et al. [9] combine the Gabor and LBP descriptors, applying LBP analysis on the Gabor magnitude faces and propose a local Gabor based binary pattern histogram sequence (LGBPHS) to represent face and have significantly improved the face recognition performance. However, in their method, they just describe neighboring changes in the spatial domain of a face image, but ignore the counterparts among different scales and orientations. In fact, there may exist some complementary information during neighboring scales and orientations helpful for face recognition. Therefore, in this paper, we propose

a novel face representation method that not only describes the neighboring relationship in the spatial domain, but also depicts the neighboring changes during different scales and orientations. First, the Gabor face images are formulated as a 3rd-order Gabor volume. Then we applied LBP operator on three orthogonal planes of Gabor volume respectively, named GV-LBP-TOP in short. In this way, we encode the neighboring changes both in spatial space and during different types of Gabor faces. Zhao et al. [10] have proposed a similar method LBP-TOP and applied it for the face expression analysis. The difference is that their method is taken in the spatial and temporal domains of the video sequence, whereas ours is conducted on the Gabor face volume to explore the neighboring relationship in spatial, frequency and orientation domains. Moreover, in order to reduce the computational complexity, we further propose an effective GV-LBP (E-GV-LBP) descriptor that describes the neighboring changes according to the central point in spatial, scale and orientation domains simultaneously for face representation. After that, the local histograms are extracted and the weighted histogram intersection metric is finally used as a dissimilarity measure for face recognition.

The rest of this paper is organized as follows. Section 2 briefly reviews the definition of Gabor filters and details the GV-LBP-TOP and E-GV-LBP representations based on the Gabor faces. Section 3 describes the details of weighted histogram distance metric and the process of face recognition. Experimental results and analysis are demonstrated in Section 4 and in Section 5, we conclude the paper.

2. GV-LBP-TOP and E-GV-LBP based face representation

2.1. Gabor faces

Gabor filters, which exhibit desirable characteristics of spatial locality and orientation selectively and are optimally localized in the space and frequency domains, have been extensively and successfully used in face recognition. The Gabor kernels we used are defined as follows:

$$\psi_{\mu,\nu} = \frac{k_{\mu,\nu}^2}{\sigma^2} \exp(-\frac{k_{\mu,\nu}^2 z^2}{2\sigma^2})[\exp(ik_{\mu,\nu}z) - \exp(-\frac{\sigma^2}{2})] \tag{1}$$

where μ and ν define the orientation and scale of the Gabor kernels respectively, $z = (x, y)$, and the wave vector $k_{\mu,\nu}$ is defined as:

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu} \tag{2}$$

where $k_\nu = k_{max}/f^\nu$, $k_{max} = \pi/2$, $f = \sqrt{2}$, $\phi_\mu = \pi\mu/8$. The Gabor kernels in (1) are all self-similar since they can be generated from one filter, the mother wavelet, by scaling and rotating via the wave vector $k_{\mu,\nu}$. Hence, a band of Gabor filters is generated by a set of various scales and rotations.

In this paper, we use Gabor kernels at five scales $\nu \in \{0, 1, 2, 3, 4\}$ and eight orientations $\mu \in \{0, 1, 2, 3, 4, 5, 6, 7\}$ with the parameter $\sigma = 2\pi$ [5] to derive the Gabor representation by convolving face images with corresponding Gabor kernels. For every image pixel we have totally 40 Gabor magnitude coefficients that means we can obtain 40 Gabor faces from a single input face image. These Gabor faces can then be resembled to form a 3rd-order Gabor tensor. Fig. 1 shows an example of a face image with its corresponding 3rd-order Gabor volume.
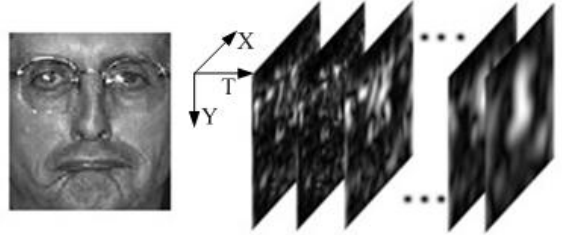

Figure 1. A face image and its corresponding 3rd-order Gabor volume.

2.2. Gabor volume based LBP on three orthogonal planes (GV-LBP-TOP)

Local Binary Patterns (LBP) is introduced as a powerful local descriptor for micro-features of images [1]. The LBP operator labels the pixels of an image by thresholding the 3×3-neighborhood of each pixel with the center value and considering the result as a binary number (or called LBP codes). An illustration of the basic LBP operator is shown in Fig.2.
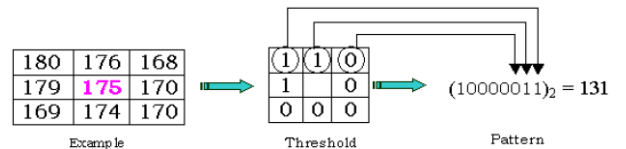

Figure 2. Calculation of LBP code from 3x3 subwindow.

Recently, the combination of Gabor and LBP has been proved to be an effective way for face recognition [9]. Inspired from this, we propose to explore discriminative information by describing the neighboring relationship not only in spatial domain, but also among different types of Gabor faces. Particularly, for a face image, the derived Gabor faces can be formulated as a 3rd-order volume as illustrated in Fig. 1, where the three axis X, Y, T denote the height, width of face image and different types of Gabor filters respectively. It can be seen that the method LGBP in [9] essentially applied LBP operator on XY plane. It is possible and natural to conduct the similar analysis in XT and YT planes to explore more sufficient and discriminative information for face representation. GV-LBP-TOP is originated from this idea. It first takes LBP analysis on the three orthogonal planes (XY, XT and YT) of Gabor face volume

respectively and then combines the results of these planes to represent faces.
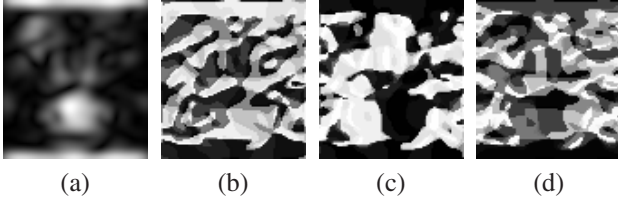

(a)   (b)   (c)   (d)

Figure 3. One Gabor face (a) and its GV-LBP-XY (b), GV-LBP-XT (c), GV-LBP-YT (d) results.

    Fig. 3 illustrates an example Gabor face and its corresponding GV-LBP codes on XY, XT and YT planes. It is easy to see the results from three planes are different and hence may supply complementary information helpful for face recognition. After that, three histograms corresponding to GV-LBP-XY, GV-LBP-XT and GV-LBP-YT codes are computed as

$$H_j(l) = \sum_{x,y} I(f_j(x,y) = l), l = 0, 1, \ldots, L_j - 1 \quad (3)$$

in which $I(.) \in \{0,1\}$ is an indication function of a boolean condition and $f_j(.)$ expresses the GV-LBP codes in j-th plane (j=0: XY; 1: XT; 2: YT), and L_j is the bin number of the j-th GV-LBP code.

    The GV-LBP-TOP histogram H is finally derived by concatenating these three histograms $H = [H_1, H_2, H_3]$ to represent the face that incorporates the spatial information and the co-occurrence statistics in Gabor frequency and orientation domains, and thus is more effective for face representation and recognition.

2.3. Effective GV-LBP

    The GV-LBP-TOP mentioned above is of somewhat high computational complexity. The length of histogram feature vector and the computational cost are three times than those of LGBPHS [9], so it is not very efficient in practical application. To address this problem, we propose an effective formulation of GV-LBP (E-GV-LBP) that encodes the information in spatial, frequency and orientation domains simultaneously and reduces the computational cost. Fig. 4 shows the definition of E-GV-LBP coding. For the centered point I_c, I_0 and I_4 are the orientation neighboring points; I_2 and I_6 are the scale neighboring ones; I_1, I_3, I_5 and I_7 are the neighboring points in spatial domains. Like LBP, all the values of these points surrounded are compared to the value of the centered point, thresholded into 0 or 1 and transformed into a value between 0 and 255 to form the E-GV-LBP value.

$$E - GV - LBP = \sum_{p=0}^{7} 2^p S(I_p - I_c) \quad (4)$$

where $S(I_p - I_c)$ is a threshold function defined as

$$S(I_p - I_c) = \begin{cases} 1 & \text{if } I_p - I_c \geq 0 \\ 0 & \text{if } I_p - I_c < 0 \end{cases}$$
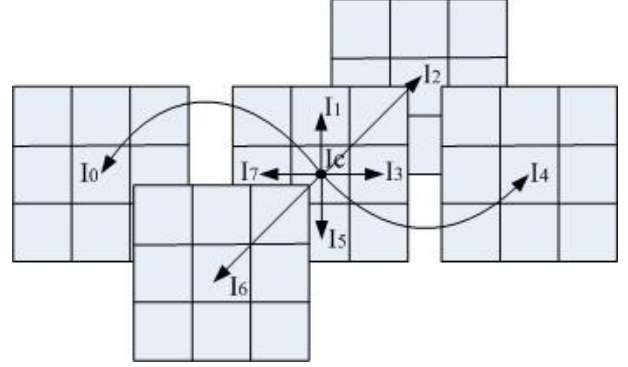

Figure 4. Formulation of E-GV-LBP.

    Fig. 5 demonstrates the 40 Gabor faces and their corresponding E-GV-LBP codes for an input face image. The histogram features are then computed based on the E-GV-LBP codes to provide a more reliable description as

$$H(l) = \sum_{x,y} I(f(x,y) = l), l = 0, 1, \ldots, L - 1 \quad (5)$$

where $I(.) \in \{0,1\}$ is an indication function of a boolean condition and $f(.)$ denotes the E-GV-LBP codes, and L is the bin value number of the E-GV-LBP codes.
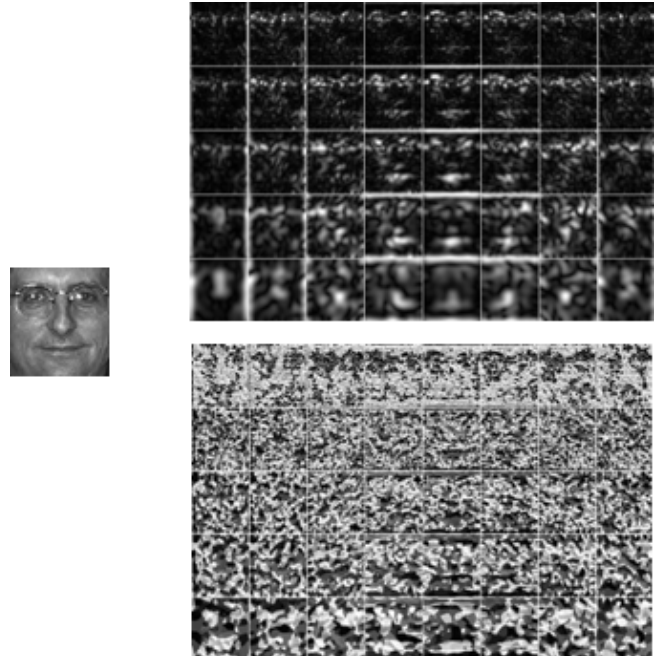

Figure 5. Gabor magnitude faces (top) and E-GV-LBP results (down) for one face image.

3. Weighted histogram intersection based Face recognition

The GV-LBP-TOP or E-GV-LBP histogram is utilized to represent faces and in face recognition phase, the histogram intersection defined in Equ. 6 is used as the dissimilarity to measure different face images.

$$D(H^1, H^2) = \sum_i min(h_i^1, h_i^2) \tag{6}$$

where H^1, H^2 are two histograms and h_i^1, h_i^2 denote the i-th bin value. Directly comparing the histograms based on the whole faces may lose the structure information of faces which is important for face recognition. One possible way is to partition the face into several blocks. The local histograms are first obtained based on different blocks and then concatenated into a histogram sequence to represent the whole face. In this way, we succeed to depict the face image at three levels. The GV-LBP-TOP or E-GV-LBP label contains information in spatial, Gabor frequency and orientation domains at pixel level. Local histogram expresses characteristic at regional level which is robust to alignment errors and finally, they are combined together as a global description for a face image to maintain both its accuracy and robustness. Pervious work has shown different regions of face make different contributions for the performance of recognition [9, 1], e.g., the areas nearby eyes and nose are more important than others. Therefore, it is sensible to assign different weights onto different blocks when measure the dissimilarity of two images.

Thus, the weighted dissimilarity of different histogram sequences is formulated as:

$$D(\mathscr{H}^1, \mathscr{H}^2) = \sum_{i=1}^{n} w_i D(H_i^1, H_i^2) \tag{7}$$

where $\mathscr{H}^1, \mathscr{H}^2$ denote the two histogram sequences and w_i is the weight for the i-th local histogram pair H_i^1, H_i^2.

In this paper, we take the similar measure in [9] to set the weights for different blocks. For each block, we first compute the dissimilarity means m_i, m_e and variations σ_i^2, σ_e^2 based on the block, for intra (the same person) and extra (different persons) sample pairs respectively and then the weight for the block can be computed following the fisher criterion [4] as

$$w = \frac{(m_i - m_e)^2}{(\sigma_i^2 + \sigma_e^2)} \tag{8}$$

where m_i, σ_i^2 denote the mean and variation of intra sample pairs and m_e, σ_e^2 are those of extra sample ones.

Therefore, if the local histogram features are discriminative, where the means of intra and extra classes are far apart and the variances are small, the corresponding block will be assigned large weight. Otherwise, the weight will be small.

The algorithms of GV-LBP-TOP and E-GV-LBP for face representation and recognition are summarized as follows.

GV-LBP-TOP:

1. Compute Gabor face volume by convolving a face image with 40 Gabor filters.
2. Compute GV-LBP-XY, GV-LBP-XT, GV-LBP-YT codes based on XY, XT and YT planes of Gabor volume respectively.
3. Divide the face into several blocks and for each block, compute the local histogram H_{XY}, H_{XT}, H_{YT} respectively and concatenate them into one $H = [H_{XY}, H_{XT}, H_{YT}]$.
4. Concatenate the local histograms into a single histogram sequence and use the weighted histogram intersection defined in Equ. 7 to derive the dissimilarity score.

E-GV-LBP:

1. Compute Gabor faces by convolving a face image with different scales and orientations Gabor filters.
2. Compute effective GV-LBP code introduced in Sec. 2.3 on Gabor faces.
3. Divide the face into several blocks and for each block, compute the local histogram H of E-GV-LBP code.
4. Concatenate the local histograms into a single histogram sequence and use the weighted histogram intersection defined in Equ. 7 to derive the dissimilarity score.

4. Experiments

In this section, we investigate the performance of the proposed method, compared with other popular methods using the public face databases.

4.1. Data preparation

Two face databases, FERET [7] and FRGC ver 2.0 [6] are tested. All the images are rotated, scaled and cropped to 88×80 according to the provided eye positions succeeded by histogram equalization preprocessing. No further preprocessing is applied. For FERET database, the training set contains 731 images. In test phase, we use the gallery set containing 1196 images from 1196 subjects, and combine four provided probe sets together, totally including 2111 images to compose the probe set. So our test protocol should be more difficult than any of the four original protocols because we consider different factors (expression, illumination, aging etc.) together to evaluate the performance. For FRGC database, we select a subset from query set for experiment 4, which consists of still uncontrolled images including variations of illumination, expression, accessory and blurring. There are 316 subjects, each of which contains at least 10 images. We randomly select 10 images for

each subject to get a total $316 \times 10 = 3,160$ images. These images are randomly divided into three sets. The training set consists of 116 persons, with 10 images per person. The left 200 subjects are divided into gallery and probe sets. For each person, 2 images randomly selected from the 10 images compose the gallery set and the left 8 images compose the probe set. The persons in training set are disjoint with those in gallery and probe sets. Fig. 6 illustrates some cropped face examples of FERET and FRGC databases.

(a)

(b)

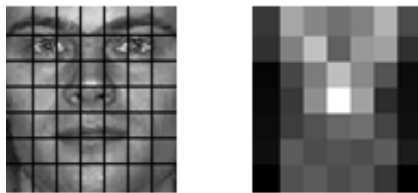Figure 6. Face examples of FERET (a) and FRGC (b) databases.

Figure 7. Face partition and their corresponding weights for E-GV-LBP.

4.2. Performance evaluation

For the LBP, LGBP, GV-LBP-TOP and E-GV-LBP methods, the face is divided into 7×7 blocks in the experiment and the weights of different blocks are fixed respectively using the FERET training set only for both tests on FERET and FRGC databases. Fig. 7 shows the face partition mode and the comparative weights of 7×7 blocks for E-GV-LBP. It can be seen that the regions of eyes and nose have more contributions for face recognition which are

consistent to our intuition. The recognition results are reported in terms of rank-1 recognition, the equal error rate (EER) and the verification rate (VR) when the false accept rate (FAR) is at 0.001.

Table 1. The performance of different methods on the FERET database

Methods	Rank-1	EER	VR (FAR=0.001)
PCA	0.6935	0.1597	0.6604
FLDA	0.7324	0.1697	0.6466
LBP [1]	0.7385	0.1133	0.6509
Gabor+FLDA	0.8574	0.0935	0.7489
LGBP [9]	0.8792	0.0639	0.7878
GV-LBP-TOP	0.8844	0.0586	0.8020
E-GV-LBP	**0.8873**	**0.0568**	**0.8072**

Table 2. The performance of different methods on the FRGC database

Methods	Rank-1	EER	VR (FAR=0.001)
PCA	0.6100	0.1721	0.2809
FLDA	0.7006	0.2027	0.3378
LBP [1]	0.6787	0.2283	0.3074
Gabor+FLDA	0.8319	0.1443	0.4884
LGBP [9]	0.8194	0.1291	0.5288
GV-LBP-TOP	0.8363	**0.1267**	**0.5453**
E-GV-LBP	**0.8381**	0.1315	0.5366

Table 1) and 2) exhibit the results of different methods on FERET and FRGC databases respectively, and Fig. 8 illustrates the corresponding ROC curves. For FRGC database, because of the comparatively low performance of some methods, we just plot the ROC curves of E-GV-LBP, GV-LBP-TOP and LGBP for better illustration. From the results, we can observe:

1. In almost all experiments, the proposed methods, GV-LBP-TOP and E-GV-LBP achieve better results than other popular methods which strongly prove the effectiveness of the proposed descriptors.

2. Regarding to the complexity and accuracy , E-GV-LBP is the best method for face representation and recognition.

3. In the case of uncontrolled illumination like FRGC database, the LBP method seems to fail, but the combination of Gabor and LBP can effectively reduce the affect of illumination and significantly improves the performance.

4. Unlike PCA, LDA etc. whose performances greatly depend on the training set, for LBP, LGBP, GV-LBP-TOP and E-GV-LBP methods, once the weight for every block is fixed, no further training procedure is
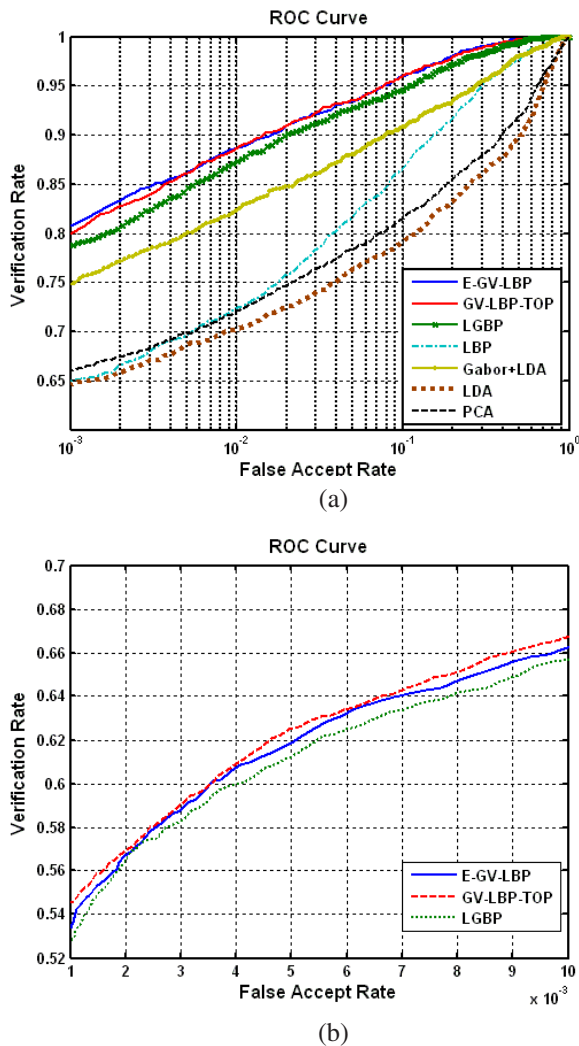
ROC Curve

(a)

(b)

Figure 8. ROC curves on FERET (a) and FRGC (b) databases.

needed. Therefore, they are not troubled by generalization problem and the performances of these methods are comparatively stable on different databases.

5. Conclusions

This paper proposes two novel face representations. Different from LGBP [9], we first formulate Gabor faces as a 3rd-order Gabor volume and then applied LBP operators on three orthogonal planes (GV-LBP-TOP), encoding discriminative information not only in spatial domain, but also in Gabor frequency and orientation domains. In order to reduce the computational complexity, an effective GV-LBP (E-GV-LBP) descriptor is further proposed to describe the changes in spatial, frequency and orientation domains simultaneously. The experimental results prove the effectiveness of the proposed method. As in LBP [1], two extensions can be developed for the proposed method. First, we can adopt uniform pattern mechanism for better representa-

tion robust to noise, and second, we can extend the neighborhood size. Moreover, the length of histogram sequence feature of the proposed method is still a little long, therefore, we will also focus on the feature selection and reduction methods such as PCA, LDA, AdaBoost learning etc. to select the most discriminant features and remove the redundance in original histogram features to make the representation more effective and compact.

References

[1] T. Ahonen, A. Hadid, and M.Pietikainen. "Face description with local binary patterns:application to face recognition". *IEEE Trans. PAMI*, 28:2037–2041, 2006.

[2] P. Belhumeur, J. Hespanha, and D. Kriegman. "Eigenfaces vs. fisherfaces: recognition using class specific linear projection". *IEEE Trans. PAMI*, 19(7):711–720, 1997.

[3] P. Comon. "Independent component analysis - a new concept?" *Signal Processing*, 36:287–314, 1994.

[4] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification, Second Edition*. John Wiley and Sons, 2001.

[5] C. Liu and H. Wechsler. "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition". *IEEE Transactions on Image Processing*, 11(4):467–476, 2002.

[6] P. J. Phillips, P. J. Flynn, W. T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. J. Worek. "Overview of the face recognition grand challenge". In *CVPR (1)*, pages 947–954, 2005.

[7] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. "The FERET evaluation methodology for face-recognition algorithms". *IEEE Trans. PAMI*, 22(10):1090–1104, 2000.

[8] L. Wiskott, J. Fellous, N. Kruger, and C. V. Malsburg. "Face recognition by elastic bunch graph matching". *IEEE Trans. PAMI*, 19(7):775–779, 1997.

[9] W. C. Zhang, S. G. Shan, W. Gao, and H. M. Zhang. "Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition". In *Proceedings of IEEE International Conference on Computer Vision*, pages 786–791, 2005.

[10] G. Zhao and M. Pietikäinen. "Dynamic texture recognition using local binary patterns with an application to facial expressions". *IEEE Trans. PAMI*, 29(6):915–928, 2007.