Chapter 4
Local Representation of Facial Features

Joni-Kristian Kämäräinen, Abdenour Hadid, and Matti Pietikäinen

The aim of this chapter is to give a comprehensive overview of different facial representations and in particular describe local facial features.

4.1 Introduction

Developing face recognition systems involves two crucial issues: facial representation and classifier design [47, 101]. The aim of facial representation is to derive a set of features from the raw face images which minimizes the intra-class variations (i.e., within face instances of a same individual) and maximizes the extra-class variations (i.e., between face images of different individuals). Obviously, if inadequate facial representations are adopted, even the most sophisticated classifiers fail to accomplish the face recognition task. Therefore, it is important to carefully decide on what facial representation to adopt when designing face recognition systems. Ideally, the facial feature representation should: (i) discriminate different individuals well while tolerating within-class variations; (ii) be easily extracted from the raw face images in order to allow fast processing; and (iii) lie in a low dimensional space (short vector length) in order to avoid a computationally expensive classifier. Naturally, it is

J.-K. Kämäräinen (✉)
Machine Vision and Pattern Recognition Laboratory, Lappeenranta University of Technology, Lappeenranta, Finland
e-mail: Joni.Kamarainen@lut.fi

A. Hadid · M. Pietikäinen
Machine Vision Group, Dept. of Electrical and Information Engineering, University of Oulu, Oulu, Finland

A. Hadid
e-mail: hadid@ee.oulu.fi

M. Pietikäinen
e-mail: mkp@ee.oulu.fi

S.Z. Li, A.K. Jain (eds.), *Handbook of Face Recognition*,
DOI 10.1007/978-0-85729-932-1_4,

not easy to find features which meet all these criteria because of the large variability in facial appearances due to different imaging factors such as scale, orientation, pose, facial expressions, lighting conditions, aging, presence of glasses, etc. These considerations are important for the other subtasks in face biometrics: detection, localization and registration, and verification, and thus, a key issue in face recognition is finding efficient facial feature representations.

Numerous methods have been proposed in literature for representing facial images for recognition purposes. The earliest attempts, such as Kanade's work in early 70s [41], are based on representing faces in terms of geometrical relationships, such as distances and angles, between the facial landmarks (eyes, mouth etc.). Later, appearance based techniques have been proposed. These methods generally consider a face as a 2D array of pixels and aim at deriving descriptors for face appearance without explicit use of face geometry. Following these lines, different holistic methods such as Principal Component Analysis (PCA) [82], Linear Discriminant Analysis (LDA) [21] and the more recent 2D PCA [92] have been widely studied. Lately local descriptors have gained an increasing attention due to their robustness to challenges such as pose and illumination changes. Among these descriptors are Gabor filters and Local Binary Patterns [2] which are shown to be very successful in encoding facial appearance.

4.1.1 Structure and Scope of the Chapter

The aim of this chapter is to give a comprehensive overview of different facial representations and in particular describe local facial features. Section 4.2 discusses the major methods which have been proposed in literature. Then, more detailed descriptions of two widely used approaches, namely local binary patterns and Gabor filters, are presented in Sects. 4.3 and 4.4, respectively. Section 4.5 discusses related issues and promising directions. Finally, concluding remarks are drawn in Sect. 4.6.

The methods discussed in this chapter can be applied to detection and recognition of faces or face parts (landmarks). Face parts are also referred to as facial features, but we use the terms feature and facial feature interchangeably for any features extracted from the face area. We specifically discuss local binary patterns in the context of face recognition and Gabor features in the context of face part detection, but they can be used in the both tasks. Furthermore, the feature extraction methods are discussed from the face image processing point of view and other face description methods are available for the modeling purposes, such as the active shape models and morphable model described in the following chapters. These novel modeling methods can also be applied to face recognition without explicit feature extraction and classification as discussed in this chapter.

4.2 Review of Facial Feature Representations

We first justify and restrict the scope of this chapter to generic features which do not require optimization or learning stages and then proceed to the actual review.

Zhao et al. [101] divide face recognition algorithms into (i) *appearance-based* (holistic), (ii) *feature-based*, and (iii) *hybrid* approaches. This taxonomy is widely accepted and also applies to face detection, localization and verification algorithms [33]. This chapter specifically focuses on the feature-based and hybrid methods which utilize representations of local face parts. Zhao et al. further divide the feature-based and hybrid approaches into: (1) *generic methods based on generic image processing features*, such as edges, lines, curves, etc.; (2) *feature-template-based methods* that are used to detect specific facial features such as eyes, nostrils, etc.; and (3) *structural matching methods* that take into consideration geometrical constraints on the features. From the feature extraction point of view, the holistic approach and the feature-template-based methods are equivalent. They both learn a scanning window template or templates to represent and detect faces or facial parts. The most popular solutions are Viola–Jones detector [85] and PCA or LDA computed subspace-templates (Eigenfaces or Fisherfaces) [9] and their seminal works. These methods can be effective, but we do not include the Haar-cascades produced by the Viola-Jones method or subspace templates produced by the PCA and LDA to this chapter since they are not generic features. They should be considered as learned statistical or algorithmic detectors themselves. Subspace methods are discussed in Chap. 3 and Viola–Jones type boosted detectors in Chap. 11. The Haar-like features used by the Viola-Jones detector, however, are generic features for facial feature representation. The structural matching methods are not in the scope either since they too involve the learning stage for a "constellation model" which captures information about spatial relationships between local features. Typical examples are active shape models, discussed in Chap. 4, and the Elastic Bunch Graph Matching (EBGM) [89]. The generic low level features used by these methods, however, belong to this chapter.

The selection of features for a proper facial feature representation is actually similar to the feature selection and extraction task occurring in the most computer vision and image analysis applications. But what features are the most suitable for face biometrics? The best results have been achieved by concatenating and learning person specific features computed from several local areas, for example, from fixed area (Fig. 4.1(a)) or varying area regions (Fig. 4.1(b)) which can be regular or feature-driven, or simply at specific locations with no strictly defined spatial extent (Fig. 4.1(c)). As already mentioned, implementations based on the subspace approach [11] and the boosted Haar-like features [103] for face detection and recognition exist, but they are not included here due to their need of task-specific learning.

Computer vision and image processing literature contains numerous features and feature extraction methods. In face biometrics, however, certain features retain their popularity and continuously succeed to producing state-of-the-art results for various benchmarks. Widely adopted are features constructed from responses of Gabor filters on various orientations and scales. More recent, and particularly successful,

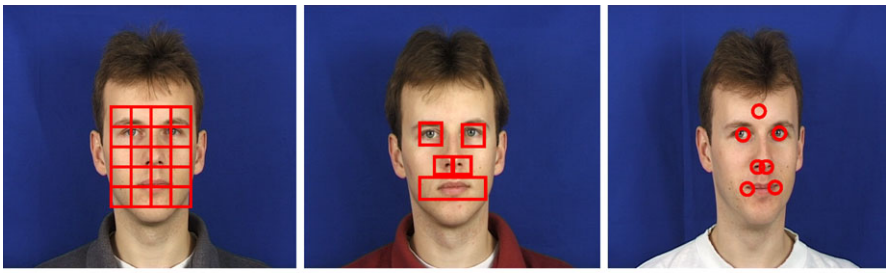

Fig. 4.1 Facial feature computation from **a** a regular grid of fixed size regions, **b** irregular variable size regions (feature-driven) and **c** around central feature locations

are local binary pattern (LBP) features. In order to verify their status and to spot new trends, we reviewed the recently published feature-intense articles in the top tier forums of computer vision and face biometrics. A short summary of the review is presented in Table 4.1. We draw the following conclusions: (1) Gabor filters and other similar "local oriented frequency approaches" are still a popular choice and produce state-of-the-art results in face detection and recognition; (2) a new feature appears in the literature: the SIFT descriptor which is popular in visual object categorization and baseline matching; (3) gray-level patch remains as a popular choice as well despite of its extreme simplicity; and finally (4) success of LBP in biometrics promotes other similar algorithmically constructed features. An interesting work is the method by Xu et al. [90], which uses several different kind of features on different processing levels in their hierarchical system.

The most popular region features, modular PCA, LBP and Gabor magnitudes, were compared for face recognition in [103]. The LBP and Gabor features produced good results and were generally recommended. In Table 4.1, we classify many features, such as complex and smooth wavelets, steerable filters and difference of Gaussians, to Gabor-based methods, because there is no fundamental difference between them and properly utilized they should lead to equally good results. Similarly, SIFT, LBP and Daugman's phase descriptor have similar characteristics. The flexibility of LBP features, however, makes them more suitable and preferable for face biometrics. The flexibility, appearing as various intuitive parameterizations and extensions to the standard LBP are further discussed in Sect. 4.3. The Haar-like features seem to succeed for the boosting approaches, but as a generic method for face biometrics there is no clear evidence for their success. Their accuracy to locate different facial landmarks have been studied in [11] and recently, other kind of features, such as anisotropic Gaussian [60] or constructed features [87], have succeeded in the boosting scheme.

It is clear from all previously published surveys and from the recent state-of-the-art results that the three mentioned features pop up as very popular and successful: features based on Gabor filter responses, local binary patterns (LBPs) and Haar-like features. Since the Haar-like features are covered in Chap. 11, this chapter

Table 4.1 Feature-based methods for face detection and/or recognition. Papers utilizing LBP are numerous and therefore not included here but in Sect. 4.3

#	Ref.	Feature(s)	Comment
1	Zhang et al. [98]	"Local derivative pattern"	Similar to LBP
2	Kozakaya et al. [42]	Histogram of gradients (HOG)	Similar to SIFT
3	Zhang and Wang [94]	SIFT	
4	Su et al. [77]	Gabor	Reg. grid, magn. only
5	Pinto et al. [68]	Gabor, Patch	Magn. only, post-processing
6	Hua and Akbarzadeh [34]	Gradient descriptor in [88]	
7	Lee et al. [46]	Modular PCA	
8	Liu and Dai [53]	Wavelet	Similar to Gabor
9	McCool and Marcel [56]	DCT coeffs.	Similar to Gabor magn. histogram
10	Ashraf et al. [7]	Patch	
11	Ding and Martinez [19]	Patch and geometric	
12	Liang et al. [50]	Patch	
13	Meyers and Wolf [59]	Gabor	V1 type post-processing
14	Mian et al. [61]	3D descriptor and SIFT	
15	Xu et al. [90]	Patch, gradient (AAM) and geometric	Fusion over layers of processing
16	Yan et al. [91]	Haar based pattern (LAB)	Similar to LBP
17	Gökberk et al. [27]	Gabor	Magn. only, centroids
18	Shastri and Levine [75]	Non-negative sparse codebook	Similar to Gabor magn.
19	Zhang et al. [97]	Gabor	Daugman's phase code [18] (similar to SIFT)
20	Arca et al. [6]	Gabor	Magn. only, centroids
21	Bicego et al. [10]	SIFT	
22	Ekenel and Stiefelhagen [20]	DCT coeffs.	Similar to Gabor magn. histogram
23	Zhang and Jia [93]	Steerable filters	Similar to Gabor
24	Dalal and Triggs [14]	Histogram of gradients (HOG)	Similar to SIFT

introduces the remaining two and presents results from face recognition and facial feature localization experiments.

4.3 Local Binary Patterns

The use of local binary patterns in face analysis started in 2004 when a novel facial representation for face recognition was proposed [1, 2]. In this approach, the face

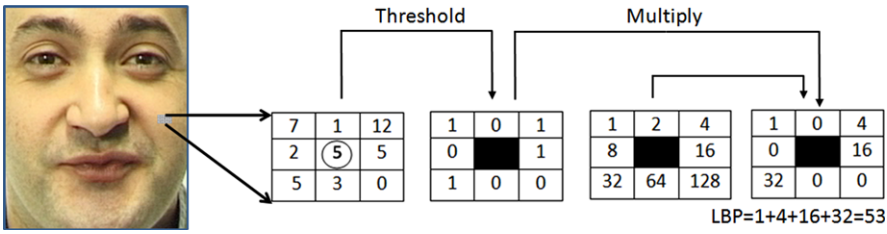

Fig. 4.2 The basic LBP operator

image is divided into several regions from which the LBP features are extracted and concatenated into an enhanced feature histogram which is used as a face descriptor. The approach has evolved to be a growing success and has been adopted and further developed by a large number of research groups and companies around the world. The LBP operator and its variants have been used not only in face recognition but also in various other face-related problems such as face detection, facial expression recognition, gender classification, age estimation and visual speech recognition. The success of LBP in face description is due to the discriminative power and computational simplicity of the operator, and its robustness to monotonic gray scale changes caused by, for example, illumination variations. The use of histograms as features also makes the LBP approach robust to face misalignment and pose variations. The Matlab code of the LBP operators can be found and freely downloaded from http://www.ee.oulu.fi/mvg/page/downloads.

4.3.1 Local Binary Patterns

4.3.1.1 LBP in the Spatial Domain

The LBP texture analysis operator, introduced by Ojala et al. [63, 64], is defined as a gray-scale invariant texture measure, derived from a general definition of texture in a local neighborhood. It is a powerful texture descriptor and among its properties in real-world applications are its discriminative power, computational simplicity and tolerance against monotonic gray-scale changes.

The original LBP operator forms labels for the image pixels by thresholding the 3×3 neighborhood with the center value and considering the result as a binary number. The histogram of these $2^8 = 256$ different labels can then be used as an image descriptor. See Fig. 4.2 for an illustration of the basic LBP operator. The operator has been extended to use neighborhoods of different sizes [64]. Using a circular neighborhood and bilinear interpolation at noninteger pixel coordinates allow any radius and number of sampling points. In the following, the notation (P, R) will be used for pixel neighborhoods which means P sampling points on a circle of radius R. See Fig. 4.3 for an example of circular neighborhoods.

Another extension to the original operator is the definition of so called *uniform patterns* [64]. This extension was inspired by the fact that some binary patterns

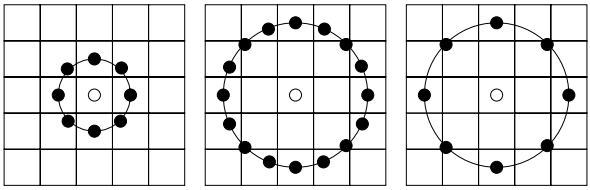

Fig. 4.3 Neighborhood set for different (P, R). The pixel values are bilinearly interpolated whenever the sampling point is not in the center of a pixel

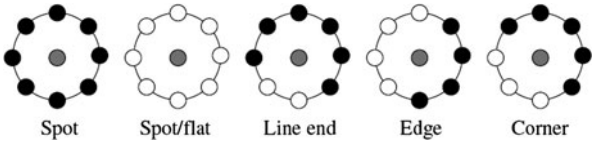

Fig. 4.4 Examples of texture primitives detected by LBP (white circles represent ones and black zeros)

occur more frequently than others in texture images. A local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is traversed circularly. For example, the patterns 00000000 (0 transitions), 01110000 (2 transitions) and 11001111 (2 transitions) are uniform whereas the patterns 11001001 (4 transitions) and 01010011 (6 transitions) are not. In the computation of the LBP labels, uniform patterns are used so that there is a separate label for each uniform pattern and all the non-uniform patterns are labeled with a single label. For example, when using $(8, R)$ neighborhood, there are a total of 256 patterns of which 58 are uniform thus yielding to the total of 59 different labels.

Ojala et al. noticed in their experiments with texture images that uniform patterns account for almost 90% of all patterns when using the $(8, 1)$ neighborhood and around 70% for the $(16, 2)$ neighborhood. We have found that 90.6% of the patterns in the $(8, 1)$ neighborhood and 85.2% of the patterns in the $(8, 2)$ neighborhood are uniform in the case of preprocessed FERET face images [67]. Each LBP code can be regarded as a micro-texton. Local primitives which are codified by these bins include different types of curved edges, spots, flat areas etc. as illustrated in Fig. 4.4.

We use the following notation for the LBP operator: $\mathrm{LBP}^{u2}_{P,R}$. The subscript denotes the operator in a (P, R) neighborhood. Superscript $u2$ stands for uniform patterns of maximum of 2 transitions and labeling all remaining patterns with a single label.

After the LBP labeled image $f_l(x, y)$ has been obtained, the LBP histogram can be defined as

$$H_i = \sum_{x,y} I\{f_l(x, y) = i\}, \quad i = 0, \ldots, n-1, \tag{4.1}$$

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

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

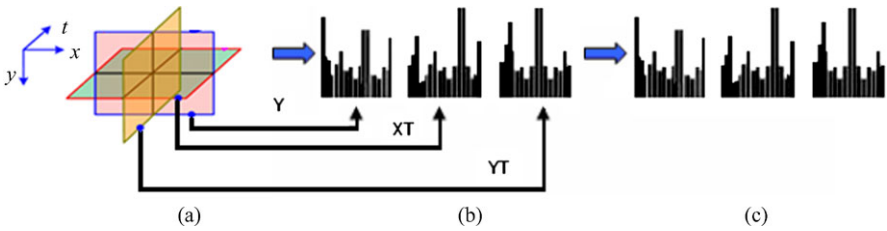

Fig. 4.5 **a** Three planes of dynamic texture; **b** LBP histograms of each plane; **c** Concatenated feature

When the image patches whose histograms are to be compared have different sizes, the histograms must be normalized to get a coherent description:

$$N_i = \frac{H_i}{\sum_{j=0}^{n-1} H_j}. \tag{4.2}$$

4.3.1.2 Spatiotemporal LBP

The original LBP operator was defined to only deal with the spatial information, but recently it has been extended to a spatiotemporal representation for dynamic texture (DT) analysis. This has yielded to so called Volume Local Binary Pattern operator (VLBP) [99]. The idea behind VLBP consists of looking at dynamic texture as a set of volumes in the (X,Y,T)-space where X and Y denote the spatial coordinates and T the frame index (time). The neighborhood of each pixel is thus defined in a three dimensional space. Then, similarly to LBP, volume textons can be defined and extracted into histograms. Therefore, VLBP combines motion and appearance into a dynamic texture description.

To make the VLBP computationally simple and easy to extend, the cooccurrences of the LBP on the three orthogonal planes (LBP-TOP) was introduced [99]. LBP-TOP consists of the three orthogonal planes: XY, XT and YT, and concatenating local binary pattern co-occurrence statistics in these three directions. The circular neighborhoods are generalized to elliptical sampling to fit to the space-time statistics. The LBP codes are extracted from the XY, XT and YT planes, denoted as XY-LBP, XT-LBP and YT-LBP, for all pixels, and statistics of the three different planes are concatenated into a single histogram. The procedure is shown in Fig. 4.5. In this representation, dynamic texture (DT) is encoded by XY-LBP, XT-LBP and YT-LBP.

Using equal radii for the time and spatial axes is not reasonable for dynamic textures [99] and therefore, in the XT and YT planes, different radii can be assigned to sample neighboring points in space and time. More generally, the radii in axes X, Y and T, and the number of neighboring points in the XY, XT and YT planes can also be different denoted by R_X, R_Y and R_T, P_{XY}, P_{XT} and P_{YT}. The corresponding feature is denoted as LBP-TOP$_{P_{XY},P_{XT},P_{YT},R_X,R_Y,R_T}$.

Let us assume we are given an $X \times Y \times T$ dynamic texture ($x_c \in \{0, \ldots, X-1\}$, $y_c \in \{0, \ldots, Y-1\}$, $t_c \in \{0, \ldots, T-1\}$). A histogram of the DT can be defined as

$$H_{i,j} = \sum_{x,y,t} I\{f_j(x, y, t) = i\}, \quad i = 0, \ldots, n_j - 1; \ j = 0, 1, 2, \tag{4.3}$$

in which n_j is the number of different labels produced by the LBP operator in the jth plane ($j = 0 : XY$, $1 : XT$ and $2 : YT$) and $f_i(x, y, t)$ expresses the LBP code of central pixel (x, y, t) in the jth plane. Similarly to the original LBP, the histograms must be normalized to get a coherent description for comparing the DTs:

$$N_{i,j} = \frac{H_{i,j}}{\sum_{k=0}^{n_j - 1} H_{k,j}}. \tag{4.4}$$

4.3.1.3 Multi-Scale LBP

Noticing that LBP features calculated in a local 3×3 neighborhood cannot capture large-scale structures, multi-scale LBP has been proposed to overcome this limitation. A straightforward way of enlarging the spatial support area is to combine the information provided by N LBP operators with varying P and R values. This way, each pixel in an image gets N different LBP codes. The most accurate information would be obtained by using the joint distribution of these codes. However, such a distribution would be overwhelmingly sparse with any reasonable image size. Therefore, only the marginal distributions of the different operators are considered. Even though the LBP codes at different radii are not statistically independent in the typical case, using multi-resolution analysis often enhances the discriminative power of the resulting features. With most applications, this straightforward way of building a multi-scale LBP operator has resulted in very good accuracy.

An extension of multi-scale LBP operator is the multiscale block local binary pattern (MB-LBP) [51] which has gained popularity especially in facial image analysis. The key idea of MB-LBP is to compare average pixel values within small blocks instead of comparing pixel values. The operator always considers 8 neighbors, producing labels from 0 to 255. For instance, if the block size is 3×3 pixels, the corresponding MB-LBP operator compares the average gray value of the center block to the average values of the 8 neighboring blocks of the same size and the effective area of the operator is 9×9 pixels.

4.3.2 Face Description Using LBP

4.3.2.1 Description of Static Face Images

In the LBP approach for texture classification [64], the occurrences of the LBP codes in an image are collected into a histogram. The classification is then performed by

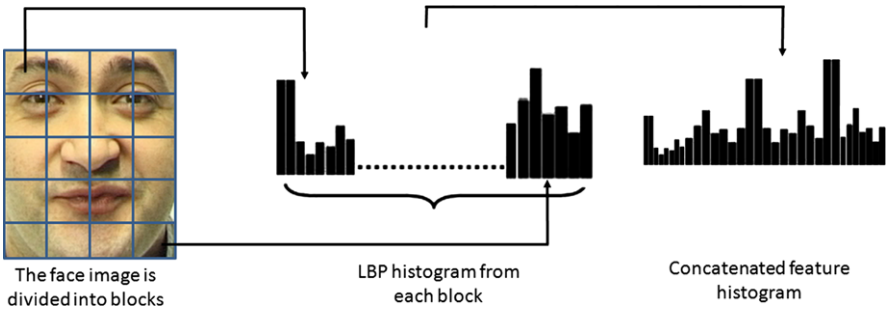

Fig. 4.6 Example of an LBP based facial representation

computing simple histogram similarities. However, considering a similar approach for facial image representation results in a loss of spatial information and therefore one should codify the texture information with their locations. One way to achieve this goal is to use the LBP texture descriptors to build several local descriptions of the face and combine them into a global description. Such local descriptions have gained interest lately which is understandable given the limitations of the holistic representations. These local feature based methods seem to be more robust against variations in pose or illumination than holistic methods.

The basic methodology for LBP based face description is as follows: The facial image is divided into local regions and LBP texture descriptors are extracted from the each region independently. The descriptors are then concatenated to a global face description, as shown in Fig. 4.6.

The basic histogram that is used to gather information about LBP codes in an image can be extended into a *spatially enhanced histogram* which encodes both the appearance and the spatial relations of facial regions. As the facial regions $R_0, R_1, \ldots, R_{m-1}$ have been determined, the spatially enhanced histogram is defined as

$$H_{i,j} = \sum_{x,y} I\{f_l(x,y) = i\} I\{(x,y) \in R_j\}, \quad i = 0, \ldots, n-1, \ j = 0, \ldots, m-1.$$

This histogram effectively has a description of the face on three different levels of locality: the LBP 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 regional histograms are concatenated to build a global description of the face. It should be noted that when using the histogram based methods the regions $R_0, R_1, \ldots, R_{m-1}$ do not need to be rectangular. Neither do they need to be of the same size or shape, and they do not necessarily have to cover the whole image. It is also possible to have partially overlapping regions.

This outlines the original LBP based facial representation [1, 2] that has been later adopted to various facial image analysis tasks [31, 45]. Figure 4.6 shows an example of an LBP based facial representation.

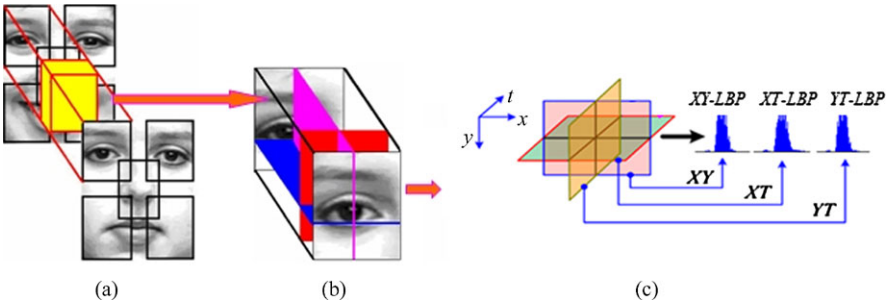

Fig. 4.7 Features in each block volume. **a** Block volumes; **b** LBP features from three orthogonal planes; **c** Concatenated features for one block volume with the appearance and motion

4.3.2.2 Description of Face Sequences

How can moving faces be efficiently represented? Psychophysical findings state that facial movements can provide valuable information to face analysis. Therefore, efficient facial representations should encode both appearance and motion. We thus describe an LBP based spatiotemporal representation for face analysis in videos using region-concatenated descriptors. Like in [2], an LBP description computed over a whole face sequence encodes only the occurrences of the micro-patterns without any indication about their locations. To overcome this effect, a representation in which the face image is divided into several overlapping blocks is used. The LBP-TOP histograms in each block are computed and concatenated into a single histogram, as illustrated in Fig. 4.7. All features extracted from the each volume are connected to represent the appearance and motion of the face in the sequence. The basic VLBP features could also be considered and extracted on the basis of region motion in the same way as the LBP-TOP features.

The LBP-TOP histograms in each block volume are computed and concatenated into a single histogram. All features extracted from each block volume are connected to represent the appearance and motion of the face. In this way, we effectively have a description of the face on three different levels of locality. The labels (bins) in the histogram contain information from three orthogonal planes, describing appearance and temporal information at the pixel level. The labels are summed over a small block to produce information on a regional level expressing the characteristics for the appearance and motion in specific locations, and all information from the regional level is concatenated to build a global description of the face sequence.

4.3.3 Face Recognition Using LBP Descriptors

This section describes the application of the LBP based face description to face recognition. Typically a nearest neighbor classification rule is used in the face recognition task. This is due to the fact that the number of training (gallery) images per

Fig. 4.8 **a** An example of a facial image divided into 7×7 windows. **b** The weights set for weighted χ^2 dissimilarity measure. The *black squares* indicate weight 0.0, *dark gray* 1.0, *light gray* 2.0 and *white* 4.0

subject is low, often only one. However, the idea of a spatially enhanced histogram can be exploited further when defining the distance measure for the classifier. An indigenous property of the proposed face description method is that each element in the enhanced histogram corresponds to a certain small area of the face. Based on the psychophysical findings, which indicate that some facial features (such as eyes) play a more important role in human face recognition than other features [101], it can be expected that some of the facial regions contribute more than others in terms of extra-personal variance. Utilizing this assumption the regions can be weighted based on the importance of the information they contain. Figure 4.8 shows an example of weighting different facial regions. The weighted Chi square distance can be defined as

$$\chi_w^2(\mathbf{x}, \xi) = \sum_{j,i} w_j \frac{(x_{i,j} - \xi_{i,j})^2}{x_{i,j} + \xi_{i,j}}, \tag{4.5}$$

in which $\mathbf{x}$ and ξ are the normalized enhanced histograms to be compared, indices i and j refer to ith bin corresponding to the jth local region and w_j is the weight for the region j.

In [1, 2, 4], Ahonen et al. performed a set of experiments on the FERET face images [67]. The results showed that the LBP approach yields higher face recognition rates than the control algorithms (PCA [82], Bayesian Intra/Extra-personal Classifier (BIC) [62] and Elastic Bunch Graph Matching EBGM [89]). To gain better understanding on whether the obtained recognition results are due to general idea of computing texture features from local facial regions or due to the discriminatory power of the local binary pattern operator, we also compared LBP to three other texture descriptors, namely the gray-level difference histogram, homogeneous texture descriptor [55] and an improved version of the texton histogram [83]. The details of these experiments can be found in [4]. The results confirmed the validity of the LBP approach and showed that the performance of LBP in face description exceeds that of other texture operators as shown in Table 4.2. We believe that the main explanation for the better performance over other texture descriptors is the tolerance to monotonic gray-scale changes. Additional advantages are the computational efficiency and avoidance of gray-scale normalization prior to the LBP operator.

Table 4.2 The recognition rates obtained using different texture descriptors for local facial regions. The first four columns show the recognition rates for the FERET test sets and the last three columns contain the mean recognition rate of the permutation test with a 95% confidence interval

Method	fb	fc	dup I	dup II	lower	mean	upper
Difference histogram	0.87	0.12	0.39	0.25	0.58	0.63	0.68
Homogeneous texture	0.86	0.04	0.37	0.21	0.58	0.62	0.68
Texton Histogram	0.97	0.28	0.59	0.42	0.71	0.76	0.80
LBP (nonweighted)	0.93	0.51	0.61	0.50	0.71	0.76	0.81

Fig. 4.9 Example of Gallery and probe images from the FRGC database, and their corresponding filtered images with Tan and Triggs' preprocessing chain [80]

Recently, Tan and Triggs developed a very effective preprocessing chain for face images and obtained excellent results using LBP-based face recognition for the FRGC database [80]. Since then, many others have adopted their preprocessing chain for applications dealing with severe illumination variations. Figure 4.9 shows an example of gallery and probe images from the FRGC database and the corresponding filtered images with the preprocessing method.

Chan et al. [12] considered multi-scale LBPs and derived new face descriptor from Linear Discriminant Analysis (LDA) of multi-scale local binary pattern histograms. The face image is first partitioned into several non-overlapping regions. In each region, multi-scale uniform LBP histograms are extracted and concatenated into a regional feature. The features are then projected on the LDA space to be used as a discriminative facial descriptor. The method was tested in face identification on the standard FERET database and in face verification on the XM2VTS database with very promising results.

Zhang et al. [95] considered the LBP methodology for face recognition and used AdaBoost learning algorithm for selecting an optimal set of local regions and their weights. This yielded to a smaller feature vector than that used in the original LBP approach [1]. However, no significant performance enhancement was obtained. Later, Huang et al. [36] proposed a variant of AdaBoost called JSBoost for selecting the optimal set of LBP features for face recognition.

In order to deal with strong illumination variations, Li et al. developed a very successful system combining near infrared (NIR) imaging with local binary pattern features and AdaBoost learning [49]. The invariance of LBP with respect to mono-

tonic gray level changes makes the features extracted from NIR images illumination invariant.

In [70], Rodriguez and Marcel proposed an approach based on adapted, client-specific LBP histograms for the face verification task. The method considers local histograms as probability distributions and computes a log-likelihood ratio instead of χ^2 similarity. A generic face model is considered as a collection of LBP histograms. Then, a client-specific model is obtained by an adaptation technique from the generic model under a probabilistic framework. The reported experimental results show that the proposed method yields good performance on two benchmark databases (XM2VTS and BANCA). Later, Ahonen and Pietikäinen [3] have further enhanced the face verification performance on the BANCA database by developing a novel method for estimating the local distributions of LBP labels. The method is based on kernel density estimation in xy-space, and it provides much better spatial accuracy than the block-based method of Rodriguez and Marcel [70].

4.3.4 LBP in Other Face-Related Problems

The LBP approach has also been adopted to several other face analysis tasks such as facial expression recognition [23, 74], gender recognition [78], age classification [86], face detection [30, 71, 91], iris recognition [79], head pose estimation [54] and 3D face recognition [48]. For instance, LBP is used in [35] with Active Shape Model (ASM) for localizing and representing facial key points since an accurate localization of such points of the face is crucial to many face analysis and synthesis problems. The local appearance of the key points in the facial images are modeled with an Extended version of Local Binary Patterns (ELBP). ELBP was proposed in order to encode not only the first derivation information of facial images but also the velocity of local variations. The experimental analysis showed that the combination ASM-ELBP enhances the face alignment accuracy compared to the original ASM method.

In [30], the authors devised another LBP based representation which is suitable for low-resolution images and has a short feature vector needed for fast processing. A specific aspect of this representation is the use of overlapping regions and a 4-neighborhood LBP operator ($\text{LBP}_{4,1}$) to avoid statistical unreliability due to long histograms computed over small regions. Additionally, the holistic description of a face was enhanced by including the global LBP histogram computed over the whole face image. The proposed representation performed well in the face detection problem.

Spatiotemporal LBP descriptors, especially LBP-TOP, have been successfully utilized in many video-based applications, for example, dynamic facial expression recognition [100], visual speech recognition [102] and gender recognition from videos [29]. They can effectively describe appearance, horizontal motion and vertical motion from the video sequence. LBP-TOP based approach was also extended to include multiresolution features which are computed from different sized blocks,

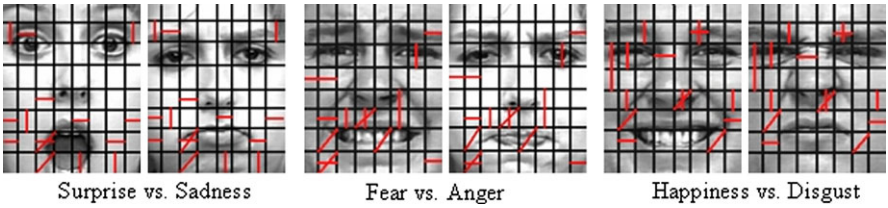

Fig. 4.10 Selected 15 slices for different facial expression pairs

different neighboring samplings and different sampling scales, and utilize AdaBoost to select the slice features for all the expression classes or every class pair, to improve the performance with short feature vectors. After that, on the basis of selected slices, the location and feature types of most discriminative features for every class pair are considered. Figure 4.10 shows the selected features for two expression pairs. They are different and specific depending on the expressions.

4.4 Gabor Features

4.4.1 Introduction

Methods using Gabor features have been particularly successful in biometrics. For example, Daugman's iris code [18] is The Method for iris recognition, Gabor features were used in the two best methods in the ICPR 2004 face recognition contest [57] and they are among the top performers in fingerprint matching [38], and so on. It is interesting, why feature extraction based on the Gabor's principle of simultaneous localization in the frequency and spatial domains [25], is so successful in many applications of computer vision and image processing. The same principle was independently found as an intuitive requirement for a "general picture processing operator" by Granlund [28], and later rigorously defined in 2D by Daugman [16].

As the well-known result in face recognition, Lades et al. developed a Gabor based system using dynamic link architecture (DLA) framework which recognizes faces by extracting a set of features ("Gabor jet") at each node of a rectangular grid over the face image [44]. Later, Wiskott et al. extended the approach and developed the well-known Gabor wavelet-based elastic bunch graph matching (EBGM) method to label and recognize faces [89]. In the EBGM algorithm, faces are represented as graphs with nodes positioned at fiducial points (such as the eyes and the tip of the nose) and edges labeled with distance vectors. Each node contains a set of Gabor wavelet coefficients, known as a jet. Thus, the geometry of the face is encoded by the edges while the local appearance is encoded by the jets. The identification of a face consists of determining among the constructed graphs the one which maximizes the graph similarity function.

In this section, we first explain the main properties of Gabor filters, then describe how image features can be constructed from filter responses, and finally, demonstrate how these features can accurately and efficiently represent and detect facial features. Note that, similarly to LBP, Gabor filters can be used to either detect face parts or whole face for recognition. In the previous sections, we explained the use of LBP for face appearance description. For completeness, we focus below on the use of Gabor filters for representing and detecting facial landmarks.

4.4.2 Gabor Filter

Gabor filter is Gabor function changed into the linear filter form, that is, a signal or an image can be convolved with the filter to produce a "response image". This process is similar to edge detection. Gabor features are formed by combining responses of several filters from a single or multiple spatial locations. Gabor function provides the minimal joint-uncertainty $\Delta t \times \Delta f$ simultaneously in the time (spatial) and frequency domains. In 1946, Dennis Gabor proved that: "*The signal which occupies the minimum area* $\Delta t \Delta f = \frac{1}{2}$ *is the modulation product of a harmonic oscillation*$^{(*)}$ *of any frequency with pulse of the form of a probability function*$^{(**)}$" [25]

$$\psi(t) = \underbrace{e^{-\alpha^2(t-t_0)^2}}_{(**)} \underbrace{e^{j2\pi f_0 t + \phi}}_{(*)}. \tag{4.6}$$

In (4.6), α is the sharpness (time duration and bandwidth) of the Gaussian, t_0 is the time shift defining the time location of the Gaussian, f_0 is the frequency of the harmonic oscillations (frequency location), and ϕ denotes the phase shift of the oscillation. The Gabor elementary function in (4.6) has a Fourier spectrum of analytical form

$$\Psi(f) = \sqrt{\frac{\pi}{\alpha^2}} e^{-(\frac{\pi}{\alpha})^2 (f-f_0)^2} e^{-j2\pi t_0 (f-f_0)+\phi}. \tag{4.7}$$

Two important findings can be seen in (4.6) and (4.7): Gabor function, or more precisely its magnitude, has the Gaussian form in the time domain and frequency domain; The Gaussian is located at t_0 in time and f_0 in frequency; If you increase the bandwidth α, the function will shrink in time (more accurate), but stretch in frequency (more inaccurate). These are the properties which help understand Gabor filter as a linear operator operating in time and frequency simultaneously. For the linear filter form, the function is typically simplified by centering it to origin ($t_0 = 0$) and removing the phase shift ($\phi = 0$).

Gabor's original idea was to synthesize signals using a set of these elementary functions. That research direction has lead to the theory of Gabor expansion (Gabor transform) [8] and more generally to the Gabor frame theory [22]. Feature extraction, however, is signal analysis. The development of the 2D Gabor elementary

functions began from Granlund in 1978, when he defined some fundamental properties and proposed the form of a general picture processing operator. The general picture processing operator had a form of the Gabor elementary function in two dimensions and it was derived directly from the needs of the image processing without a connection to Gabor's work [28]. It is noteworthy that Granlund addressed many properties, such as the octave spacing of the frequencies, that were reinvented later for the Gabor filters. Despite the original contribution of Granlund the most referred works are those conducted by Daugman [16, 17]. Daugman was the first who exclusively derived the uncertainty principle in two dimensions and showed the similarity between a structure based on the 2D Gabor functions and the organization and the characteristics of the mammalian visual system. Again, several simplifications are justifiable [39] and 2D Gabor function can be defined as

$$\begin{aligned} \psi(x, y) &= e^{-(\alpha^2 x'^2 + \beta^2 y'^2)} e^{j2\pi f_0 x'}, \\ x' &= x\cos\theta + y\sin\theta, \\ y' &= -x\sin\theta + y\cos\theta, \end{aligned} \tag{4.8}$$

where the new parameters are β for sharpness of the second Gaussian axis and θ for its orientation. In practice, the sharpness is connected to the frequency in order to make filters self-similar (Gabor wavelets) [39]. This is achieved by setting $\alpha = |f_0|/\gamma$ and $\beta = |f_0|/\eta$ and by normalizing the filter. Finally, the 2D Gabor filter in the spatial domain is

$$\begin{aligned} \psi(x, y) &= \frac{f^2}{\pi\gamma\eta} e^{-(\frac{f^2}{\gamma^2}x'^2 + \frac{f^2}{\eta^2}y'^2)} e^{j2\pi f x'}, \\ x' &= x\cos\theta + y\sin\theta, \\ y' &= -x\sin\theta + y\cos\theta, \end{aligned} \tag{4.9}$$

where f is the central frequency of the filter, θ the rotation angle of the Gaussian major axis and the plane wave, γ the sharpness along the major axis, and η the sharpness along the minor axis (perpendicular to the wave). In the given form, the aspect ratio of the Gaussian is η/γ. The normalized 2D Gabor filter function has an analytical form in the frequency domain

$$\begin{aligned} \Psi(u, v) &= e^{-\frac{\pi^2}{f^2}(\gamma^2(u'-f)^2 + \eta^2 v'^2)}, \\ u' &= u\cos\theta + v\sin\theta, \\ v' &= -u\sin\theta + v\cos\theta. \end{aligned} \tag{4.10}$$

The effects of the Gabor filter parameters, interpretable via the Fourier similarity theorem, are demonstrated in Fig. 4.11.

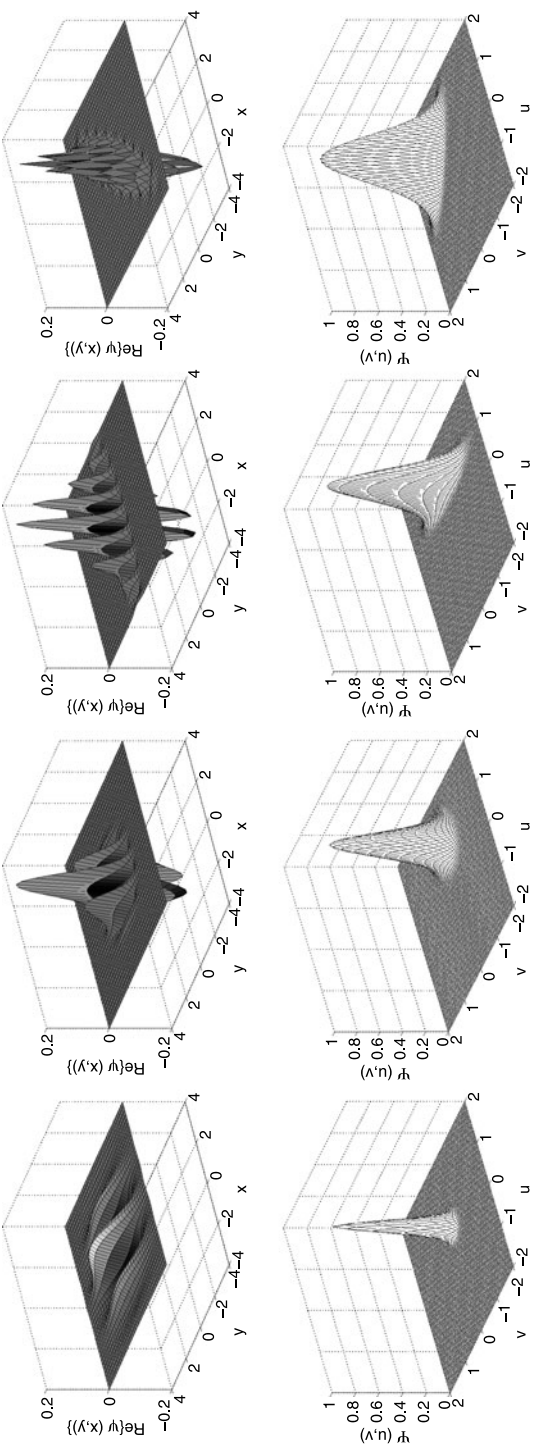

Fig. 4.11 2D Gabor filter functions with different values of the parameters f, θ, γ, and η in the space (*top*) and frequency domains (*bottom*). *Left*: ($f = 0.5$, $\theta = 0°$, $\gamma = 1.0$, $\eta = 1.0$); *middle-left*: ($f = 1.0$, $\theta = 0°$, $\gamma = 1.0$, $\eta = 1.0$); *middle-right*: ($f = 1.0$, $\theta = 0°$, $\gamma = 2.0$, $\eta = 0.5$); *right*: ($f = 1.0$, $\theta = 45°$, $\gamma = 2.0$, $\eta = 0.5$)

4.4.3 Constructing Gabor Features

Gabor features are constructed by convolution of an input image $\xi(x, y)$ with the filter in (4.9)

$$r_\xi(x, y; f, \theta) = \psi(x, y; f, \theta) * \xi(x, y)$$
$$= \int\int_{-\infty}^{\infty} \psi(x - x_\tau, y - y_\tau; f, \theta)\xi(x_\tau, y_\tau)\, dx_\tau\, dy_\tau. \quad (4.11)$$

The convolution produces a response image r_ξ of the same size. Only a single filter rarely succeeds but the response images are computed for a "bank" of filters tuned on various frequencies and orientations. The frequencies are typically drawn from the logarithmic scale similar to wavelets [15]:

$$f_k = c^{-k} f_{\max}, \quad \text{for } k = 0, \ldots, m - 1 \quad (4.12)$$

where $f_{\max}$ is the maximum frequency (the smallest scale) and c is the frequency scaling factor. Some useful values for c include $c = 2$ for octave spacing and $c = \sqrt{2}$ for half-octave spacing. The filter orientations are spaced uniformly

$$\theta_k = \frac{k2\pi}{n}, \quad k = \{0, \ldots, n - 1\}. \quad (4.13)$$

For real signals the responses on $[\pi, 2\pi[$ are complex conjugates of responses on $[0, \pi[$ and therefore only the responses for the half plane are needed:

$$\theta_k = \frac{k\pi}{n}, \quad k = \{0, \ldots, n - 1\}. \quad (4.14)$$

For a bank of Gabor filters, the responses computed at a single location (x_0, y_0) with the parameters drawn from (4.12) and (4.14) a feature matrix $\mathbf{G}$ can be constructed as

$$\mathbf{G} = \begin{pmatrix} r(x_0, y_0; f_0, \theta_0) & r(x_0, y_0; f_0, \theta_1) & \ldots & r(x_0, y_0; f_0, \theta_{n-1}) \\ r(x_0, y_0; f_1, \theta_0) & r(x_0, y_0; f_1, \theta_1) & \ldots & r(x_0, y_0; f_1, \theta_{n-1}) \\ \vdots & \vdots & \ddots & \vdots \\ r(x_0, y_0; f_{m-1}, \theta_0) & r(x_0, y_0; f_{m-1}, \theta_1) & \ldots & r(x_0, y_0; f_{m-1}, \theta_{n-1}) \end{pmatrix}. \quad (4.15)$$

In (4.15) the columns denote responses over different orientations and rows over different frequencies (scales). This structure is called as "simple Gabor feature space" formally defined in [43], later revised in [39] and utilized in face detection in [32]. A significant simplification made in the proposed feature space is the use of only one spatial location (x', y') to represent an object. The assumption is justified if the objects are simple or if they are distinguishable from each other in the feature space. This is not the case with, for example, the human face, but seems to hold between salient sub-parts, such as nostrils, eyes, mouth corners, etc. The filters in one location tuned to various frequencies and orientations span a sub-space whose accuracy

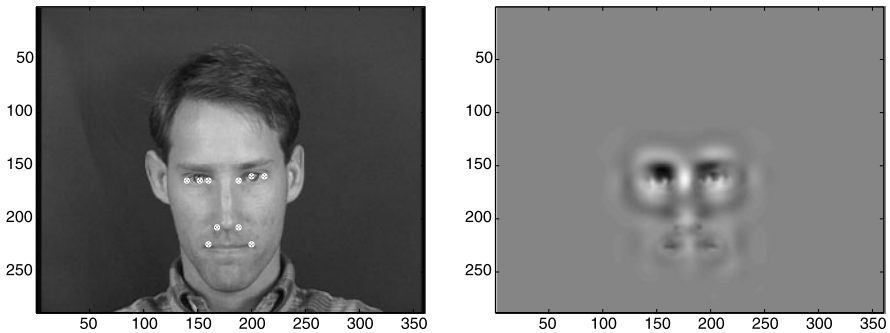

Fig. 4.12 Reconstruction from responses at 10 different locations (four orientations and five frequencies): **a** original; **b** reconstruction

decreases from the filter origin. This is demonstrated in Fig. 4.12 where an original face is reconstructed using filter responses from 10 locations.

Operations for rotation and scale invariant searches of objects can be defined as a column-wise circular shift of the response matrix corresponding to the rotation of the object around the location (x_0, y_0) and a row-wise shift corresponding to the scaling of an object by a factor c [43]. An illumination invariance can be achieved by normalizing the feature matrix [43].

4.4.4 Learning Facial Features

In principle, Gabor features can be used similarly to LBPs or any other local features. The filter responses are computed for various frequencies and orientations, and a descriptor formed from the responses inside one or multiple fixed-size windows as illustrated in Fig. 4.6. For example, Zou et al. [103] proposed a face recognition method using such region descriptor and reported state-of-the-art results for the FERET database: fb: 99.5%, fc: 99.5%, dup I: 85.0% and dup II: 79.5%. Gabor face descriptor is easy to implement, but for completeness, in this section we concentrate on local facial features and utilize the simple feature matrix to represent and learn them.

We assume an annotated training set of face images. The annotations are, for example, the centroids of selected facial landmarks (see Fig. 4.12(a)). Any classifier or pattern recognition method can be used to learn the facial representations from extracted Gabor features. A completely statistical approach, however, possess superior properties as compared to other methods [37]: the decision making has an interpretable basis from which the most probable option can be chosen and a within-class comparison can be performed using statistical hypothesis testing [66]. In the statistical approaches, a class is typically represented in terms of a class conditional probability density function (pdf) over feature space. It should be noted, that finding

a proper pdf estimate has a crucial impact on the success of the facial feature detection. Typically, the form of the pdf's is somehow restricted and the estimation is reduced to a problem of fitting the restricted model to the observed features. Often simple models such as a single Gaussian distribution (normal distributed random variable) can efficiently represent features but a more general model, such as a finite mixture model, must be used to approximate more complex pdf's. We adopt the method in [37] where Gaussian mixture models represent facial feature conditional pdf's given the Gabor feature matrix.

The multiresolution Gabor feature in a single location can be converted from the matrix in (4.15) to a feature vector

$$\mathbf{g} = \big[r(x_0, y_0; f_0, \theta_0)\; r(x_0, y_0; f_0, \theta_1)\; \ldots\; r(x_0, y_0; f_{m-1}, \theta_{n-1})\big]. \tag{4.16}$$

Since the feature vector is complex valued the complex Gaussian distribution function needs to be used,

$$\mathcal{N}^{\mathbb{C}}(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) = \frac{1}{\pi^D |\Sigma|} \exp\big[-(\boldsymbol{x} - \boldsymbol{\mu})^* \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\big], \tag{4.17}$$

where Σ denotes the covariance matrix. It should be noted that the pure complex form of the Gaussian in (4.17) provides computational stability in the parameter estimation as compared to a concatenation of real and imaginary parts to two real numbers as the dimensionality of the problem doubles in the latter case [66]. Now, a Gaussian mixture model (GMM) probability density function can be defined as a weighted sum of Gaussians

$$p(\boldsymbol{x}; \boldsymbol{\theta}) = \sum_{c=1}^{C} \alpha_c \mathcal{N}^{\mathbb{C}}(\boldsymbol{x}; \boldsymbol{\mu}_c, \Sigma_c), \tag{4.18}$$

where α_c is the weight of the cth component. The weight can be interpreted as *a priori* probability that a value of the random variable is generated by the cth source, and thus, $0 \leq \alpha_c \leq 1$ and $\sum_{c=1}^{C} \alpha_c = 1$. The Gaussian mixture model probability density function can be completely defined by the parameter list

$$\boldsymbol{\theta} = \{\alpha_1, \boldsymbol{\mu}_1, \Sigma_1, \ldots, \alpha_C, \boldsymbol{\mu}_C, \Sigma_C\}. \tag{4.19}$$

The main question remains how the parameters in (4.19) can be estimated from the given training data. The most popular estimation method is the expectation maximization (EM) algorithm, but the EM algorithm requires knowledge of the number of Gaussians, C, as an input parameter. The number is often unknown and this is a strong motivation to apply unsupervised methods, such as that of Figueiredo–Jain (FJ) [24] or the greedy EM algorithm [84]. Of the two unsupervised methods, the Figueiredo–Jain method provides more accurate results and its complex extension can be directly applied to pdf's of the complex feature vectors in (4.16) [66].

The probability distribution values, likelihoods, can be directly used to find the best or rank facial feature candidates [66]. It is even possible to reduce the search

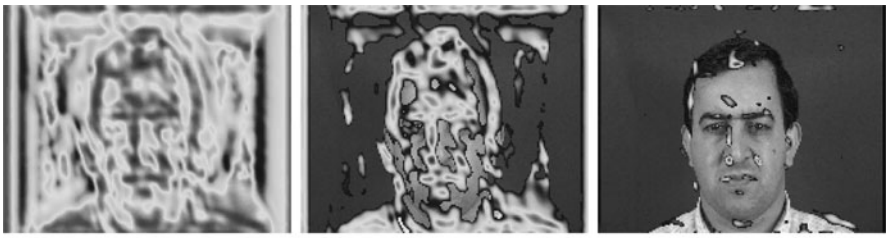

Fig. 4.13 Example of using density quantile of pdf values: **a** Pdf surface for the *left nostril* class; **b** Pdf values belonging to 0.5 density quantile; **c** Pdf values belonging to 0.05 density quantile [37]

Algorithm 4.1: Train facial feature classifier

1: **for all** Training images **do**
2: Align and normalize image to represent an object in a standard pose
3: Extract multiresolution Gabor features at given locations
4: Normalize the features
5: Store the features to the sample matrix P and their corresponding class labels to the target vector T
6: **end for**
7: With samples in P estimate class conditional pdf's for each class using Gaussian mixture models and FJ algorithm

space considerably by discarding image features beyond a requested score level, that is, density quantile [66]. In Fig. 4.13, the use of density quantile for reducing the search space is demonstrated; it is clear that the spatial area corresponding to the 0.05 (0.95 confidence) density quantile contains the correct image feature.

4.4.5 Detecting Facial Features

A supervised learning algorithm to extract simple Gabor features (multiresolution Gabor features) and to estimate the class conditional pdf's for the facial features is presented in Algorithm 4.1. Matlab functionality for efficient computation of the multiresolution Gabor features [76] and for the Gaussian mixture models and the FJ algorithm are publicly available [26]. In Algorithm 4.2, the main steps to extract the features from an image are shown.

Experiments Using the XM2VTS Face Database XM2VTS facial image database is a publicly available database for benchmarking face detection and recognition methods [58]. The frontal part of the database contains 600 training images and 560 test images of size 720×576 (width $\times$ height) pixels. For facial images ten

Algorithm 4.2: Extract K best face features of each class from an image I

```
Compute multiresolution Gabor features G(x, y; f_m, θ_n) for the whole image I(x, y)
for all Scale shifts do
    for all Rotation shifts do
        Shift Gabor features
        Normalize Gabor features
        Calculate confidence scores (pdf values) for all classes and for all (x, y)
        Update feature class confidence at each location
    end for
end for
Sort the features by their score for each class
Return the K best features of each facial feature class
```

specific regions (see Fig. 4.12(a)) have been shown to have favorable properties to act as keypoints [32]. A normalized distance between the eyes, 1.0, will be used as measure of image feature detection accuracy. The distance measure is demonstrated in Fig. 4.14(a).

Gabor parameters were experimentally selected by using a cross-validation procedure over the training and evaluation sets in the database: $n = 4$, $m = 6$, $k = \sqrt{3}$ and $f_{\text{high}} = 1/40$. Image features were extracted in a ranked order and a keypoint was considered to be correctly extracted if it was within a pre-set pixel distance limit from the correct location. Results with XM2VTS are presented in Fig. 4.14(b). The distances are scale normalized, so that the distance between centers of the eyes is 1.0 (see Fig. 4.14(a) for a demonstration). On average, 4 correct image features were included in the first 10 image features within distance limit 0.05, but as the number of features was increased to 100: over 9 for 0.05 and almost all features found for 0.10 and 0.20. It should be noted that accuracies of 0.10 and 0.20 are still very good for face registration and recognition. Increasing the number of image features over 100 (10 per class) did not improve the results anymore, but relaxing the distance limit to 0.10 almost perfect result were reached with only 10 first image features from each class. Typical detection results are demonstrated in Figs. 4.14(c)–(e).

Methods for accurate face and facial feature detection and localization based on the described Gabor representations have been proposed and reported to produce state-of-the-art detection accuracy for more difficult and realistic data sets (XM2VTS/non-frontal, BANCA and BioID) [32, 40].

4.5 Discussions on Local Features

A drawback of the LBP method, as well as of all local descriptors that apply vector quantization, is that they are not robust in the sense that a small change in the input

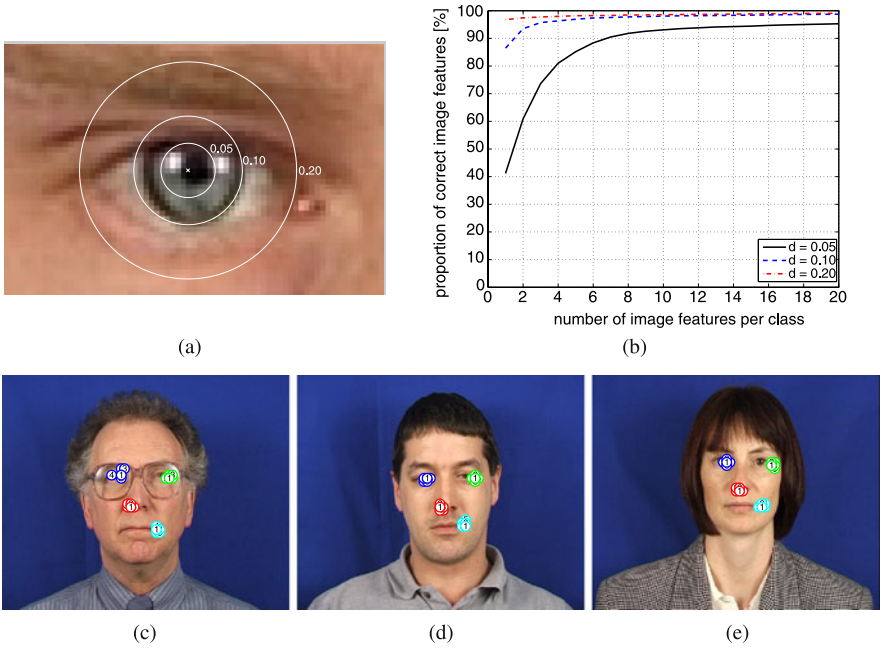

Fig. 4.14 **a** Demonstration of accuracy distance measure; **b** Performance for facial feature detection in XM2VTS test images; **c**, **d**, **e** Examples of extracted features (left eye center: *blue*, right eye outer corner: *green*, left nostril: *red*, right mouth corner: *cyan*, 5 best feature for each landmark numbered from 1 to 5) [37]

image would always cause a small change in the output. LBP may not work properly for noisy images or on flat image areas of constant gray level. Many variants of LBP have been proposed to improve its robustness. For instance, Tan and Triggs proposed a three-level operator called local ternary patterns for example, to deal with problems on flat image areas [80]. Liao et al. [52] introduced dominant local binary patterns which make use of the most frequently occurred patterns of LBP to improve the recognition accuracy compared to the original uniform patterns. Raja and Gong proposed sparse multiscale local binary patterns to better exploit the discriminative capacity of multiscale features available [69]. Inspired by LBP, higher order local derivative patterns (LDP) were proposed by Zhang et al., with applications in face recognition [98].

LBP has also inspired the development of new effective local face descriptors, such as the Weber Law Descriptor (WLD) containing differential excitation and orientation components [13] and the blur-invariant Local Phase Quantization (LPQ) descriptor [65]. The LPQ descriptor has received wide interest in blur-invariant face recognition [5]. LPQ is based on quantizing the Fourier transform phase in local neighborhoods. Similarly to the widely used LBP based face description, histograms of LPQ labels computed within local regions are also adopted as a face descriptor. The experiments showed that such LPQ descriptors are highly discriminative

and produce very promising face recognition results, outperforming LBP both with blurred and sharp images on CMU PIE and FRGC 1.0.4 datasets.

A current trend in the development of new effective local face image descriptors is to combine the strengths of complementary descriptors. From the beginning, the LBP operator was designed as a complementary measure of local image contrast. Applying LBP to Gabor-filtered face images, or using LBP and Gabor methods jointly, have provided excellent results in face recognition [81, 96]. For instance, Zhang et al. [96] proposed the extraction of LBP features from images obtained by filtering a facial image with 40 Gabor filters of different scales and orientations. Excellent results have been obtained on the all FERET sets. A downside of the method lies in the high dimensionality of the feature vector (LBP histogram) which is calculated from 40 Gabor images derived from each single original image. To overcome this problem of large feature dimensions, Shan et al. [73] presented a new extension using Fisher Discriminant Analysis (FDA), instead of the χ^2 (Chi-square), and histogram intersection, which have been previously used in [96]. The authors constructed an ensemble of piecewise FDA classifiers, each of which is built based one segment of the high-dimensional LBP histograms. Impressive results were reported on the FERET database. Other works have also successfully exploited the complementary of Gabor filters and LBP features by fusing the two set of features e.g. for age classification [86]. Combining ideas from Haar and LBP features have also given excellent results in accurate and illumination invariant face detection [71, 91].

Features based on Gabor filters are very versatile. By post-processing they can be transformed, for example, to binary descriptors of texture similar to LBPs. For example, in the Daugman's iris code the response phase is quantized to two bits (four quadratures in the complex plane) [18]. The Daugman's descriptor is very discriminative and its histograms were used in face recognition in [97]. Utilization of the phase information is important for discrimination, but many other efficient post-processing methods exist in the literature and they are used in human visual system oriented recognition methods [72]. Another important property of Gabor filters is that the original signal can be reconstructed. This property was employed in this chapter where we introduced the efficient facial feature descriptor based on Gabor features at a single location. Recently, the importance of phase information have been noticed and very good recognition results reported for features based on Gabor phase [96]. It is important to notice that the complex-valued response, including both magnitude and phase, is the most natural representation, and should be used in methods based on Gabor filters.

4.6 Conclusions

Finding efficient facial or facial feature representations is a key issue in developing robust face recognition systems. Many methods have been proposed for this purpose. Local feature based methods seem to be more robust against variations in pose or illumination than holistic methods. Especially methods based on Gabor filter

responses and local binary patterns have been particularly successful in face image processing.

Acknowledgements Abdenour Hadid and Matti Pietikäinen thank the Academy of Finland for the financial support.

References

1. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: Proc. of the ECCV (2004)
2. Ahonen, T., Hadid, A., Pietikäinen, M.: Face description with local binary patterns: Application to face recognition. IEEE Transactions on PAMI **28**(12) (2006)
3. Ahonen, T., Pietikäinen, M.: Pixelwise local binary pattern models of faces using kernel density estimation. In: Proc. of the International Conference on Biometrics (2009)
4. Ahonen, T., Pietikäinen, M., Hadid, A., Mäenpää, T.: Face recognition based on the appearance of local regions. In: Proc. of the ICPR (2004)
5. Ahonen, T., Rahtu, E., Ojansivu, V., Heikkilä, J.: Recognition of blurred faces using local phase quantization. In: Proc. of the ICPR (2008)
6. Arca, S., Campadelli, P., Lanzarotti, R.: A face recognition system based on automatically termined fiducial points. Pattern Recognit. **39**, 432–443 (2006)
7. Ashraf, A., Lucey, S., Chen, T.: Learning patch correspondences for improved viewpoint invariant face recognition. In: Proc. of the CVPR (2008)
8. Bastiaans, M.J.: Gabor's signal expansion and the Zak transform. Appl. Opt. **33**(23), 5241–5255 (1994)
9. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Trans. PAMI **19**(7) (1997)
10. Bicego, M., Lagorio, A., Grosso, E., Tistarelli, M.: On the use of SIFT features for face authentication. In: Proc. of the CVPR (2006)
11. Castrillón, M., Déniz, O., Hernández, D., Lorenzo, J.: A comparison of face and facial feature detectors based on the Viola–Jones general object detection framework. Mach. Vis. Appl. **22**, 481–494 (2011). doi:10.1007/s00138-010-0250-7
12. Chan, C.-H., Kittler, J., Messer, K.: Multi-scale local binary pattern histograms for face recognition. In: Proc. of the International Conference on Biometrics (2007)
13. Chen, J., Shan, S., He, C., Zhao, G., Pietikäinen, M., Chen, X., Gao, W.: WLD: A robust local image descriptor. IEEE Trans. Pattern Anal. Mach. Intell. **32**(9), 1705–1720 (2010)
14. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proc. of the CVPR (2005)
15. Daubechies, I.: The wavelet transform, time-frequency localization and signal analysis. IEEE Trans. Inf. Theory **36**(5) (1990)
16. Daugman, J.G.: Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. J. Opt. Soc. Am. A **2**(7), 1160–1169 (1985)
17. Daugman, J.G.: Complete discrete 2-D Gabor transform by neural networks for image analysis and compression. IEEE Trans. Acoust. Speech Signal Process. **36**(7) (1988)
18. Daugman, J.: High confidence visual recognition of persons by a test of statistical independence. IEEE Trans. PAMI **25**(9) (1993)
19. Ding, L., Martinez, A.: Precise detailed detection of faces and facial features. In: Proc. of the CVPR (2008)
20. Ekenel, H., Stiefelhagen, R.: Analysis of local appearance-based face recognition: Effects on feature selection and feature normalization. In: Proc. of the CVPR (2006)
21. Etemad, K., Chellappa, R.: Discriminant analysis for recognition of human face images. J. Opt. Soc. Am. A **14**, 1724–1733 (1997)

22. Feichtinger, H., Strohmer, T. (eds.): Gabor Analysis and Algorithms. Birkhäuser, Basel (1998)
23. Feng, X., Pietikäinen, M., Hadid, A.: Facial expression recognition with local binary patterns and linear programming. Pattern Recognit. Image Anal. **15**(2), 546–548 (2005)
24. Figueiredo, M., Jain, A.: Unsupervised learning of finite mixture models. IEEE Trans. PAMI **24**(3) (2002)
25. Gabor, D.: Theory of communication. J. Inst. Electr. Eng. **93**, 429–457 (1946)
26. GMMBayes Toolbox for Matlab. http://www2.it.lut.fi/project/gmmbayes
27. Gökberk, B., Irfanoglu, M., Akarun, L., Alpaydin, E.: Learning the best subset of local features for face recognition. Pattern Recognit. **40**, 1520–1532 (2007)
28. Granlund, G.H.: In search of a general picture processing operator. Comput. Graph. Image Process. **8**, 155–173 (1978)
29. Hadid, A., Pietikäinen, M.: Combining appearance and motion for face and gender recognition from videos. Pattern Recognit. **42**(11), 2818–2827 (2009)
30. Hadid, A., Pietikäinen, M., Ahonen, T.: A discriminative feature space for detecting and recognizing faces. In: Proc. of the CVPR (2004)
31. Hadid, A., Zhao, G., Ahonen, T., Pietikäinen, M.: Face analysis using local binary patterns. In: Mirmehdi, M., Xie, X., Suri, J. (eds.) Handbook of Texture Analysis, pp. 347–373. Imperial College Press, London (2008)
32. Hamouz, M., Kittler, J., Kamarainen, J.-K., Paalanen, P., Kalviainen, H., Matas, J.: Feature-based affine-invariant localization of faces. IEEE Trans. PAMI **27**(9) (2005)
33. Hjelmas, E., Low, B.K.: Face detection: A survey. Comput. Vis. Image Underst. **83**(3), 236–274 (2001)
34. Hua, G., Akbarzadeh, A.: A robust elastic and partial matching metric for face recognition. In: Proc. of the ICCV (2009)
35. Huang, X., Li, S.Z., Wang, Y.: Shape localization based on statistical method using extended local binary pattern. In: Proc. of the International Conference on Image and Graphics (2004)
36. Huang, X., Li, S., Wang, Y.: Jensen–Shannon boosting learning for object recognition. In: Proc. of the CVPR (2005)
37. Ilonen, J., Kamarainen, J.-K., Paalanen, P., Hamouz, M., Kittler, J., Kälviäinen, H.: Image feature localization by multiple hypothesis testing of Gabor features. IEEE Trans. Image Process. **17**(3) (2008)
38. Jain, A., Chen, Y., Demirkus, M.: Pores and ridges: Fingerprint matching using level 3 features. IEEE Trans. PAMI **29**(1) (2007)
39. Kamarainen, J.-K., Kyrki, V., Kälviäinen, H.: Invariance properties of Gabor filter based features—overview and applications. IEEE Trans. Image Process. **15**(5), 1088–1099 (2006)
40. Kamarainen, J.-K., Hamouz, M., Kittler, J., Paalanen, P., Ilonen, J., Drobchenko, A.: Object localisation using generative probability model for spatial constellation and local image features. In: Proc. of the ICCV Workshop on Non-Rigid Registration and Tracking Through Learning (2007)
41. Kanade, T.: Picture processing system by computer complex and recognition of human faces. Doctoral dissertation, Kyoto University (1973)
42. Kozakaya, T., Shibata, T., Yuasa, M., Yamaguchi, O.: Facial feature localization using weighted vector concentration approach. Image Vis. Comput. **28**, 772–780 (2010)
43. Kyrki, V., Kamarainen, J.-K., Kälviäinen, H.: Simple Gabor feature space for invariant object recognition. Pattern Recognit. Lett. **25**(3), 311–318 (2003)
44. Lades, M., Vorbrüggen, J.C., Buhmann, J., Lange, J., von der Malsburg, C., Würtz, R.P., Konen, W.: Distortion invariant object recognition in the dynamic link architecture. IEEE Trans. Comput. **42**, 300–311 (1993)
45. LBP bibliography. http://www.ee.oulu.fi/mvg/page/lbp_bibliography
46. Lee, P.-H., Hsu, G.-S., Hung, Y.-P.: Face verification and identification using facial trait code. In: Proc. of the CVPR (2009)
47. Li, S.Z., Jain, A.K. (eds.): Handbook of Face Recognition. Springer, New York (2005)

48. Li, S., Zhao, C., Zhu, X., Lei, Z.: Learning to fuse 3D + 2D based face recognition at both feature and decision levels. In: Proc. of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (2005)
49. Li, S.Z., Chu, R., Liao, S., Zhang, L.: Illumination invariant face recognition using near-infrared images. IEEE Trans. PAMI **29**(4) (2007)
50. Liang, L., Xiao, R., Wen, F., Sun, J.: Face alignment via component-based discriminative search. In: Proc. of the ECCV (2008)
51. Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S.Z.: Learning multi-scale block local binary patterns for face recognition. In: Proc. of the International Conference on Biometrics (2007)
52. Liao, S., Law, M., Chung, A.: Dominant local binary patterns for texture classification. IEEE Trans. Image Process. **18**(5), 1107–1118 (2009)
53. Liu, C.-C., Dai, D.-Q.: Face recognition using dual-tree complex wavelet features. IEEE Trans. Image Process. 2593–2599 (2009)
54. Ma, B., Zhang, W., Shan, S., Chen, X., Gao, W.: Robust head pose estimation using LGBP. In: Proc. of the ICPR (2006)
55. Manjunath, B., Ohm, J.R., Vinod, V.V., Yamada, A.: Color and texture descriptors. IEEE Trans. Circuits Syst. Video Technol. **11**(6) (2001). Special Issue on MPEG-7
56. McCool, C., Marcel, S.: Parts-based face verification using local frequency bands. In: Proc. of the International Conference on Biometrics (2009)
57. Messer, K. et al.: Face authentication test on the BANCA database. In: Proc. of the ICPR (2004)
58. Messer, K., Matas, J., Kittler, J., Luettin, J., Maitre, G.: XM2VTSDB: The extended M2VTS database. In: Proc. of the International Conference on Audio and Video-based Biometric Person Authentication (1999)
59. Meyers, E., Wolf, L.: Using biologically inspired features for face processing. Int. J. Comput. Vis. **76**, 93–104 (2008)
60. Meynet, J., Popovici, V., Thiran, J.-P.: Face detection with boosted Gaussian features. Pattern Recognit. **40**, 2283–2291 (2007)
61. Mian, A., Bennamoun, M., Owens, R.: Keypoint detection and local feature matching for textured 3d face recognition. Int. J. Comput. Vis. **79**, 1–12 (2008)
62. Moghaddam, B., Nastar, C., Pentland, A.: A Bayesian similarity measure for direct image matching. In: Proc. of the ICPR (1996)
63. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. Pattern Recognit. **29**, 51–59 (1996)
64. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. PAMI **24** (2002)
65. Ojansivu, V., Heikkilä, J.: Blur insensitive texture classification using local phase quantization. In: Proc. of the International Conference on Image and Signal Processing (2008)
66. Paalanen, P., Kamarainen, J.-K., Ilonen, J., Kälviäinen, H.: Feature representation and discrimination based on Gaussian mixture model probability densities—practices and algorithms. Pattern Recognit. **39**(7), 1346–1358 (2006)
67. Phillips, P., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face-recognition algorithms. IEEE Trans. PAMI **22** (2000)
68. Pinto, N., DiCarlo, J., Cox, D.: How far can you get with a modern face recognition test set using only simple features? In: Proc. of the CVPR (2009)
69. Raja, Y., Gong, S.: Sparse multiscale local binary patterns. In: Proc. of the British Machine Vision Conference (2006)
70. Rodriguez, Y., Marcel, S.: Face authentication using adapted local binary pattern histograms. In: Proc. of the ECCV (2006)
71. Roy, A., Marcel, S.: Haar local binary pattern feature for fast illumination invariant face detection. In: Proc. of the British Machine Vision Conference (2009)
72. Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., Poggio, T.: Object recognition with cortex-like mechanisms. IEEE Trans. PAMI **29**(3) (2007)
73. Shan, S., Zhang, W., Su, Y., Chen, X., Gao, W.: Ensemble of piecewise FDA based on spatial histograms of local (Gabor) binary patterns for face recognition. In: Proc. of the ICPR (2006)

74. Shan, C., Gong, S., McOwan, P.: Facial expression recognition based on local binary patterns:a comprehensive study. Image Vis. Comput. **27**(6), 803–816 (2009)
75. Shastri, B., Levine, M.: Face recognition using localized features based on non-negative sparse coding. Mach. Vis. Appl. **18**, 107–122 (2007)
76. SimpleGabor Toolbox for Matlab. http://www2.it.lut.fi/project/simplegabor
77. Su, Y., Shan, S., Chen, X., Gao, W.: Hierarchical ensemble of global and local classifiers for face recognition. IEEE Trans. Image Process. **18**(8) (2009)
78. Sun, N., Zheng, W., Sun, C., Zou, C., Zhao, L.: Gender classification based on boosting local binary pattern. In: Proc. of the International Symposium on Neural Networks (2006)
79. Sun, Z., Tan, T., Qiu, X.: Graph matching iris image blocks with local binary pattern. In: Proc. of the International Conference on Biometrics (2006)
80. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: ICCV Workshop on Analysis and Modeling of Faces and Gestures (2007)
81. Tan, X., Triggs, B.: Fusing Gabor and LBP feature sets for kernel-based face recognition. In: Proc. of the ICCV Workshop on Analysis and Modeling of Faces and Gestures (2007)
82. Turk, M., Pentland, A.: Eigenfaces for recognition. J. Cogn. Neurosci. **3**, 71–86 (1991)
83. Varma, M., Zisserman, A.: Texture classification: Are filter banks necessary? In: Proc. of the CVPR (2003)
84. Verbeek, J.J., Vlassis, N., Kröse, B.: Efficient greedy learning of Gaussian mixture models. Neural Comput. **5**(2), 469–485 (2003)
85. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition, pp. 511–518 (2001)
86. Wang, J., Yau, W., Wang, H.: Age categorization via ECOC with fused Gabor and LBP features. In: Proc. of the IEEE Workshop on Applications of Computer Vision (2009)
87. Wang, P., Ji, Q.: Multi-view face and eye detection using discriminant features. Comput. Vis. Image Underst. **105**, 99–111 (2007)
88. Winder, S., Brown, M.: Learning local image descriptors. In: Proc. of the CVPR (2007)
89. Wiskott, L., Fellous, J.-M., Krüger, N., von der Malsburg, C.: Face recognition by elastic bunch graph matching. IEEE Trans. PAMI **19** (1997)
90. Xu, Z., Chen, H., Zhu, S.-C., Luo, J.: A hierarchical compositional model for face representation and sketching. IEEE Trans. PAMI **30**(6) (2008)
91. Yan, S., Shan, S., Chen, X., Gao, W.: Locally assembled binary (LAB) feature with feature-centric cascade for fast and accurate face detection. In: Proc. of the CVPR (2008)
92. Yang, J., Zhang, D., Frangi, A.F., yu Yang, J.: Two-dimensional PCA: A new approach to appearance-based face representation and recognition. IEEE Trans. PAMI **26** (2004)
93. Zhang, X., Jia, Y.: Face recognition with local steerable phase feature. Pattern Recognit. Lett. **27**, 1927–1933 (2006)
94. Zhang, G., Wang, Y.: Faceprint: Fusion of local features for 3d face recognition. In: Proc. of the International Conference on Biometrics (2009)
95. Zhang, G., Huang, X., Li, S.Z., Wang, Y., Wu, X.: Boosting local binary pattern LBP-based face recognition. In: Proc. of the Chinese Conference on Biometric Recognition (2004)
96. Zhang, W., Shan, S., Gao, W., Chen, X., Zhang, H.: Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition. In: Proc. of the ICCV (2005)
97. Zhang, B., Shan, S., Chen, X., Gao, W.: Histogram of Gabor phase patterns (HGPP): A novel object representation approach for face recognition. IEEE Trans. Image Process. **16**(1) (2007)
98. Zhang, B., Gao, Y., Zhao, S., Liu, J.: Local derivate pattern versus local binary pattern: Face recognition with high-order local pattern descriptor. Image Vis. Comput. **28**, 772–780 (2010)
99. Zhao, G., Pietikäinen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans. PAMI **29**(6) (2007)
100. Zhao, G., Pietikäinen, M.: Boosted multi-resolution spatiotemporal descriptors for facial expression recognition. Pattern Recognit. Lett. **30**, 1117–1127 (2009)

101. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. ACM Comput. Surv. **34**(4), 399–458 (2003)
102. Zhao, G., Barnard, M., Pietikäinen, M.: Lipreading with local spatiotemporal descriptors. IEEE Trans. Multimed. **11**(7) (2009)
103. Zou, J., Ji, Q., Nagy, G.: A comparative study of local matching approach for face recognition. IEEE Trans. Image Process. **16**(10) (2007)