

Gender recognition based on face image using reinforced local binary patterns

ISSN 1751-9632

Received on 18th March 2016

Revised 1st March 2017

Accepted on 10th March 2017

E-First on 1st June 2017

doi: 10.1049/iet-cvi.2016.0087

www.ietdl.org

Ithaya Rani Panner Selvam¹✉, Muneeswaran Karuppiyah¹

¹Department of Computer Science and Engineering, Mepco Schlenk Engineering College, Mepco Nagar, Amathur (Post), Sivakasi 626 005, India

✉ E-mail: muhilrani@gmail.com

Abstract: Gender recognition is a challenging and innovative research topic in the present sophisticated world of visual technology. This study proposes a system which can identify the gender based on face image. For finding the location of the face region, each input image is divided into overlapping blocks and Gabor features are extracted with different scale and orientations. Generate the enhanced feature, concatenate mean, standard deviation and skewness of Gabor features which are obtained from each block. For detecting face region, this feature is passed to ensemble classifier. To recognise the gender, reinforced local binary patterns are used to extract the facial local features. Adaboost algorithm is used to select and classify the discriminative features such as male or female. The authors' experimental results on Labeled Faces in the Wild (LFW), FERET and Gallagher databases for face detection using Gabor features achieve 98, 98.5 and 96.5% accuracy, respectively. Moreover, the reinforced local binary patterns achieve the accuracy for gender classification as 97.08, 98.5 and 94.21% on the LFW, FERET and Gallagher databases, respectively. Both are achieving improved performance compared with other standard methodologies described in the literature.

1 Introduction

Gender identity is a vital attribute for human living in social life, the identification of which has drawn a lot of attention in the research area [1]. Several techniques have been proposed like hand shape, iris, face and gait for gender recognition. However, majority of techniques are based on facial information. Recognising gender based on facial images has gained much interest in the computer vision and machine learning community. Different features around the mouth, eyes and cheeks can be robustly indicative of either male or female face [2]. The presence of moustaches and beards helps more the classification of male's images. Gender identification is an essential task since lots of social interactions and services depend on it. The responses of people tend to differ based on their gender. Human behaviours also greatly rely on the gender of a person with whom he/she intend to interact. Gender identification finds its application in psychology, security (gender-based restricted access), collecting demographic information in public places, marketing research, counting the number of women entering a retail store, vision-based human monitoring and so on. The rest of the paper is organised as follows. Section 2 briefly analyses several related works. Section 3 describes the proposed work. Section 4 shows the experimental setup and results. Section 5 offers the conclusion and direction for future work.

2 Related work

Techniques related to the gender recognition have been examined extensively from the past previous decades onwards. The most straightforward method is to use the grey-scale pixels (raw pixels) on the down-sampled images (12×21 pixels) from FERET database with support vector machine (SVM) classification for gender identification proposed by Moghaddam and Yang [3]. This method is very simple but fails to be effective due to variations of poses, expressions and illuminations. Hence in this area, feature extraction plays an important role in improving accuracy and efficiency of the object identification process. Generally there are two approaches for feature extraction. They are geometric-based and appearance-based methods. Geometric-based feature extraction method reported in [4–6] extract the shape and location information of major components such as the eyes, eyebrow, mouth

and chin which can represent the position and/or appearance of these facial components. Brunelli and Poggio [4] used a set of 16 geometric features per image to train two competing networks with the radial basis function, one network for male and the other for female, and the rate of classification was 79% on 168 testing image set. Lahoucine *et al.* [5] presented a facial surface by two types of facial curves such as radials and levels. These curves are considered as geometric features that capture local facial shape and to learn the most relevant curves using adaptive boosting for face recognition and gender classification. In [6], the authors used a precise patch histogram (PPH) based on active appearance model to align the face images. The images were modelled by the patches around the coordinates of certain landmarks for gender classification. However, this method is sensitive to variations of facial poses. However, the geometric-based feature methods usually require accurate and reliable detection of facial feature which are hard to accommodate in many environmental situations.

The appearance-based method described in [7, 8] elaborates the usage of whole face image for generating the feature vector. A discriminative face feature extracted for automatic recognition of gender is proposed in [7] using multi-scale independent component analysis texture pattern. Though this method produced high detection rate, it consumes more time and memory. Dynamic motion features which are extracted from motion capture data using principal component analysis (PCA) are proposed in [8] for gender recognition. These features are further refined in time and spatial domains by exploiting gait phase cycles and significant body part indicators obtained from analysing non-pathological gait kinematics and then classification is performed using support vector machine with a radial basis function.

Recently, a more generic local appearance-based approach has been proposed [9–15], where instead of considering the whole face at once this approach divide the input face image into blocks, without considering any salient region, from which facial features are extracted. This method becomes much popular due to its robustness in environmental change and also it is independent from the location of facial components. The work proposed in [9] for gender classification employs PCA and fuzzy clustering technique, respectively, for feature extraction and classification steps. The work reported in [10] uses lips movements for classifying gender.

This method obtained gender-specific information from the way in which a person moves their lips during speech. Furthermore, it indicated that the lip dynamics during speech provided greater gender discriminative information than simply lip appearance. However, this method does not give better performance based on the time. Hong *et al.* [11] presented a feature named self-similarity of gradients (GSS) which captures pairwise statistics of localised gradient distributions from each facial region for gender identification. However, their results are not satisfactory. The work in [12] reported a classification scheme named dropout-SVM and it combines different feature extraction methods such as local binary pattern (LBP) and four patch LBP (FPLBP) for estimating age and gender of unfiltered faces.

A local Gabor binary pattern extracted the facial texture and depth features proposed in [13] for gender classification. Feature vectors are formed by combining depth and texture features and then histogram bins are computed. This approach is robust to noise and variations of illumination. The work in [14] proposed the face image that is divided into non-overlapping regions. Features are extracted from each region using LBP different scale and are fused. Then the selection of the discriminative features using mutual information for gender classification is used. Though this method produced high detection rate, it consumes more time and memory. Gender classification on real-life faces using the Labeled Faces in the Wild (LFW) database (7443 still images) is addressed in [15]. The selected 500 boosted LBP features around the eye region achieved only moderate performance.

The appearance-based method suffers less from issues of initialisation and tracking the error and they are believed to contain more information than geometric-based method on the relative positions of a finite set of facial features. In spite of many algorithms for appearance-based methods that have been explored in the literature for gender recognition, there is scope for the exploration of new algorithm to improve performance. Our proposed algorithm addresses the exploration of reinforced LBP (RLBP) to improve the results. The location of the face regions is found with the help of Gabor wavelet (GW). The proposed features deliver the high recognition rate in less time.

3 Proposed work

The proposed work describes the gender recognition based on face images. Hence, there is a need to estimate the location and the scale of the face region in each input image. Initially, the input image is divided into overlapping blocks. Features are extracted from each block using GW with selected different scale and orientations. Generate the enhanced feature, concatenate mean, standard deviation and skewness of Gabor features which are obtained from each block. These Gabor features are passed to Adaboost classifier with the well-trained signature of face and non-face images (i.e. background) for detecting the location of the face region.

To recognise gender, the facial local features are extracted using RLBP. Initially, the face image is divided into non-overlapping blocks. The proposed RLBP encodes the information in a block of pixels. The block is again divided into overlapping sub-blocks. More specifically, three kinds of RLBP features are computed. Within a sub-block, the difference between the row sums of an overlapping window is calculated. Similarly, the difference between the column sums of overlapping window and the difference between the diagonal sums of overlapping window are computed. The windows within a sub-block have same size. Adaboost algorithm is used to select and classify the discriminative RLBP features such as male and female. The RLBP method is robust to illumination variation and produced the high recognition rate. The schematic diagram for detection of human face and identification of gender is shown in Fig. 1, where the functions of various stages are self-explanatory. For conducting and evaluating the work, LFW, FERET and Gallagher databases are used. The outcome of the proposed method makes it a good choice for real-time system combining face detection and gender identification.

3.1 Face detection

3.1.1 Gabor-based feature extraction: Initially, the input image (I) is divided into overlapping blocks which is represented as B . Typically each block ' B ' is the size of 30*30 pixels. The number of blocks are $(r - \text{block_height} - 1) * (c - \text{block_width} - 1)$ where r and c are the number of rows and columns of the input image, respectively, and block_height and block_width are the size of the block height and block width, respectively. The reason for the selection of block size of 30*30 pixels is to match against the training images of 30*30 pixels

$$I_B(a, b) = I(a : a + \text{block_height} - 1, b : b + \text{block_width} - 1) \\ 1 \leq a \leq r - \text{block_height} - 1; 1 \leq b \leq c - \text{block_width} - 1 \quad (1)$$

The local appearance features are extracted using GW in each block. GW offers the best simultaneous localisation of spatial and frequency information. It has been widely applied in image processing tasks such as edge detection, invariant object recognition and compression [16]. Gabor filter is a band-pass filter and has been widely used in many applications specially capturing the texture information, which is combined with other feature extraction techniques like moments and spectral distribution has exhibited a very good classification rate [16]. The two-dimensional Gabor filter is adopted and it can be mathematically expressed as

$$\psi(a, b, \sigma, \theta) = \exp\left(-\frac{(A^2 + Y^2 B^2)}{2\sigma^2}\right) \cos\left(\frac{2\pi}{\lambda} A\right) \\ A = a \cos \theta + b \sin \theta; B = -a \sin \theta + b \cos \theta \quad (2)$$

Here θ is the orientation, σ is the effective width (scale), λ is the wavelength and Y is the aspect ratio. The GW procedure by its convolution with a Gabor filter Ψ is defined as

$$GF_{B,s,o}(a, b) = I_B(a, b) * \Psi(a, b, s, o) \\ \text{where } se \sum_{aa=1}^{rr} \sum_{bb=1}^{cc} oe \Phi \quad (3)$$

where $*$ is the symbol for convolution. The B th block within I is convoluted with the Gabor filters as in (3), resulting in a series of Gabor filtered images with features such as bars and edges usefully emphasised for better finding the location of the face region. Accordingly, each pixel within a block ' B ' is delineated as $(|\Sigma| * |\Phi|)$ -dimensional features. Each Gabor (G) filtered images are converted into statistical mean (M), standard deviation (S) and skewness (Sk) of Gabor features ($GMSSk$). All statistical features are concatenated into single feature vector as follows:

$$G_M = \frac{1}{rr * cc} \left(\sum_{aa=1}^{rr} \sum_{bb=1}^{cc} GF_{B,s,o}(aa, bb) \right) \quad (4)$$

$$G_S = \left(\sum_{aa=1}^{rr} \sum_{bb=1}^{cc} (GF_{B,s,o}(aa, bb) - G_M)^2 / m * n \right)^{1/2} \quad (5)$$

$$G_{Sk} = \left(\frac{1}{rr * cc} \sum_{aa=1}^{rr} \sum_{bb=1}^{cc} (GF_{B,s,o}(aa, bb) - G_M)^3 \right) / (G_S)^3 \quad (6)$$

where $1 \leq s \leq \sigma; 1 \leq o \leq \theta$

$$GMSSk_B = [G_M \cup G_S \cup G_{Sk}] \quad (7)$$

where rr and cc are the size of rows and columns of the block, respectively. As a result, each block contains 96-dimensional features ($\sigma * \theta * 3$).

3.1.2 Classification: Boost algorithm as proposed in [17] for object detection exhibits a low false positive rate. The benefit of Boost algorithm is to reduce the redundancies of the high-dimensional feature space and computational expenses. It is

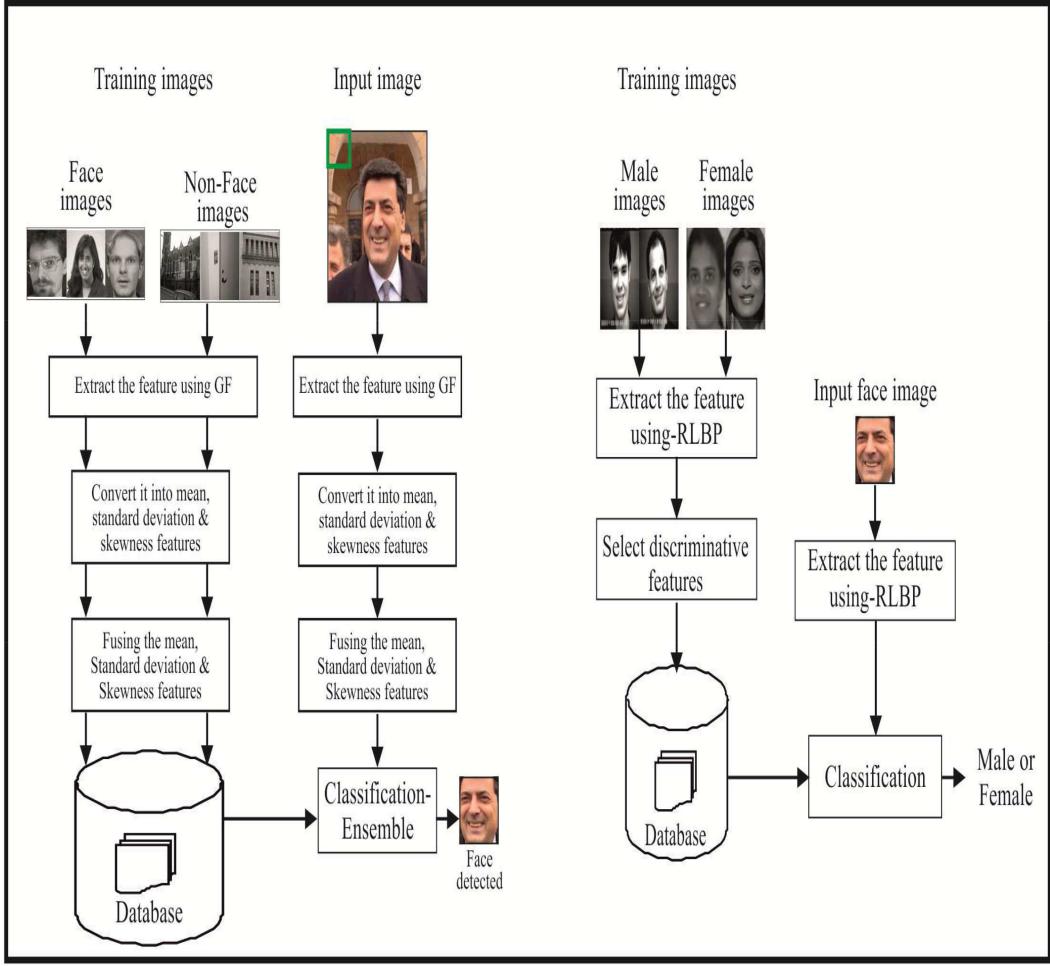


Fig. 1 Schematic diagram for detecting the face and identifying the gender where green colour box represents block, respectively

sensitive to noisy data and outliers. Some of the boost classifiers such as RealAdaboost [17] (RA), GentleAdaboost [18] (GA) and ModestAdaboost [19] (MA) are the variants of the boost algorithms which are deployed in the proposed work to detect the faces. Initially the training set is formed in the database (DB) using the GMSSk signatures from the collection of training images having both face and non-face images. The detection of human face is described in the algorithm as shown in Fig. 2.

3.2 Gender recognition

The location of the face region is used for gender recognition. A novel descriptor is proposed for gender recognition analysis, i.e. RLBP which outperforms state-of-the-art existing methods. The RLBP is a reinforced spatial representation of local binary descriptors and it represents its local texture. The proposed RLBP descriptor is simple and computationally efficient and it shows significantly improved performance for gender recognition task. The existing LBP [15] method labels the pixels of an image by thresholding a 3×3 neighbourhood of each pixel with the centre value and considering the results as a binary number (see Fig. 3a for an illustration). Then the histogram of the labels can be used as a texture descriptor. This method achieved only average performance.

Hence, the proposed RLBP is an extension of LBP method for achieving good performance. Fig. 3b shows a worked out example of a sample block using the proposed method. The face image is divided into non-overlapping blocks. The block is again divided into overlapping sub-blocks. Each sub-block is of size 4×4 pixels. The difference between the rows sums of overlapping window size of 2×2 pixels within a sub-block is computed. As a result, the sub-block size is reduced from 4×4 pixels to 3×3 pixels. Then the average value of 3×3 pixels is calculated and it replaces the centre sub-block pixel value. Then compare the centre pixel of sub-block

with the neighbouring pixels within a sub-block into a binary code as in LBP feature which renders high recognition rate and is described in detail below.

3.2.1 RLBP-based feature extraction: The grey level of the face region F is divided equally into 42 (6 rows and 7 columns) non-overlapping blocks as proposed in [15]. The block is represented as $B(B = 1, 2, \dots, (r_1 \times c_1))$. Typically, each block is of size 12×12 pixels as given below

$$F_B(aa, bb) = F(aa:aa + bheight - 1, bb:bb + bwidth - 1) \\ \text{where } 1 \leq aa \leq r_1 \text{ increment } aa \text{ by } bheight; \\ 1 \leq bb \leq c_1 \text{ increment } bb \text{ by } bwidth \quad (8)$$

where r_1 and c_1 are the number of rows and columns in the face region, respectively, and $bheight$ and $bwidth$ are the size of the block height and block width, respectively. Each block is again divided into overlapping sub-blocks. It is represented as SB. Generally each sub-block is of size 4×4 pixels. The number of sub-blocks is $(m_1 - sb_height - 1) * (n_1 - sb_width - 1)$ where m_1 and n_1 are the number of rows and columns in each block, respectively, and sb_height and sb_width are the height and width of the sub-block, respectively, as given below

$$F_{1B,SB}(a_1, b_1) = F_B(a_1:a_1 + sb_height - 1, b_1:b_1 + sb_width - 1) \\ 1 \leq a_1 \leq m_1 - sb_height - 1; \quad 1 \leq b_1 \leq n_1 - sb_width - 1 \quad (9)$$

The RLBP is defined by a binary coding function to obtain the difference between the row sums of each overlapping window size (a_1, b_1) of 2×2 pixels within a sub-block (a_1, b_1) of 4×4 pixels as follows.

Input : Given an input image (I), Set scale $\sigma = 8$ and orientations $\theta = 4$, DB - training database

Output : F - face region

```

 $I(a,b) \leftarrow$  Resize the input image  $I$  into 110x110 pixels.

 $B \leftarrow 1$  // represent block

// where  $r$  and  $c$  are the rows and columns of input image respectively

// input image is divided into overlapping blocks

 $(block\_height, block\_width) = size(block)$  //each block size is of 30*30 pixels

Iterate 1 :  $a=1:r; b=1:c$ 

 $I_B(a,b)=I(a:a+block\_height-1, b:b+block\_width-1)$ 

// where  $rr$  and  $cc$  are the rows and columns of size of the block

 $GF_{B,s,o}(a,b) \leftarrow$  Extract GW feature  $(I_B(a,b)*\psi(a,b,s,o))$  where  $s \in \Sigma, o \in \Phi$ 

Iterate 2 :  $s = 1$  to  $\sigma$ ;  $o = 1$  to  $\theta$ 

 $G_M = \frac{1}{rr*cc} \left( \sum_{aa=1}^{rr} \sum_{bb=1}^{cc} GF_{B,s,o}(aa,bb) \right)$ 

 $G_S = \left( \sum_{aa=1}^{rr} \sum_{bb=1}^{cc} (GF_{B,s,o}(aa,bb) - G_M)^2 / rr*cc \right)^{1/2}$ 

 $G_{Sk} = \left( \frac{1}{rr*cc} \sum_{aa=m}^{n} \sum_{bb=n}^{n} (GF_{B,s,o}(aa,bb) - G_M)^3 \right) / (G_S)^3$ 

 $GMSSk_B = [G_M \cup G_S \cup G_{Sk}]$ 

```

End iterate 2

```

 $F \leftarrow$  Obtained the face region using boost classifier  $(GMSSk_B, DB)$ 

// pixel wise classification

 $B \leftarrow B + 1$ 

```

End iterate 1

Fig. 2 Algorithm for detection of face region

Let M represent the matrix defined as

$$M = F1_{B,SB}(a_1, b_1)$$

$$S(k_x, k_y) = \sum_{a_2=k_y}^{a_2+1} M(k_x, a_2) - \sum_{b_2=k_y}^{b_2+1} M(k_x + 1, b_2) \quad (10)$$

$$1 \leq k_x \leq sb_height - 1; \quad 1 \leq k_y \leq sb_width - 1$$

where $S(k_x, k_y)$ is the difference between the rows sum of overlapping window size of $2*2$ intensities in a sub-block 'SB' whose size is $4*4$ at pixel position (k_x, k_y) . The resultant value is available as $3*3$ matrices for each sub-block at pixel position (k_x, k_y) . From that result, replace the centre value of $3*3$ pixels by calculating the average of sub-blocks expressed as

$$S(2, 2) = \frac{1}{k_x * k_y} \sum_{i_2=1}^{k_x} \sum_{j_2=1}^{k_y} S(i_2, j_2) \quad (11)$$

where k_x and k_y are the number of rows and columns in the resulting sub-block ($3*3$ pixels), respectively. This centre value in the sub-block is compared with eight neighbour values as proposed in [15]. If the centre value is less than the neighbouring value then neighbouring pixel values set to 1; otherwise zero. An 8-bit code is coined, which represents the value in the decimal form at each

pixel position (a_1, b_1) in the block. The representations of these quantities are as follows.

The value by using the proposed method RLBP is obtained as (see (12)) and

$$T(i_1, j_1) = \begin{cases} 1, & S(i_1, j_1) \geq S(2, 2) \\ 0, & \text{else} \end{cases} \quad (13)$$

After obtaining the value using RLBP method for each pixel associated with a block in the face region, a 59-bin histogram as proposed in [15] is computed to extract the texture feature. The histograms of each block in the face region $RLBP_f$ can be defined as

$$RLBP_f = \sum_{a_1, b_1} I(RLBP(a_1, b_1) = L) \quad 1 \leq a_1 \leq m_1; \quad 1 \leq b_1 \leq n_1; \quad 1 \leq L \leq 59 \quad (14)$$

where L is the number of bins for the values produced by the RLBP operator

$$I(A) = \begin{cases} 1, & A \text{ is true} \\ 0, & A \text{ is false} \end{cases} \quad (15)$$

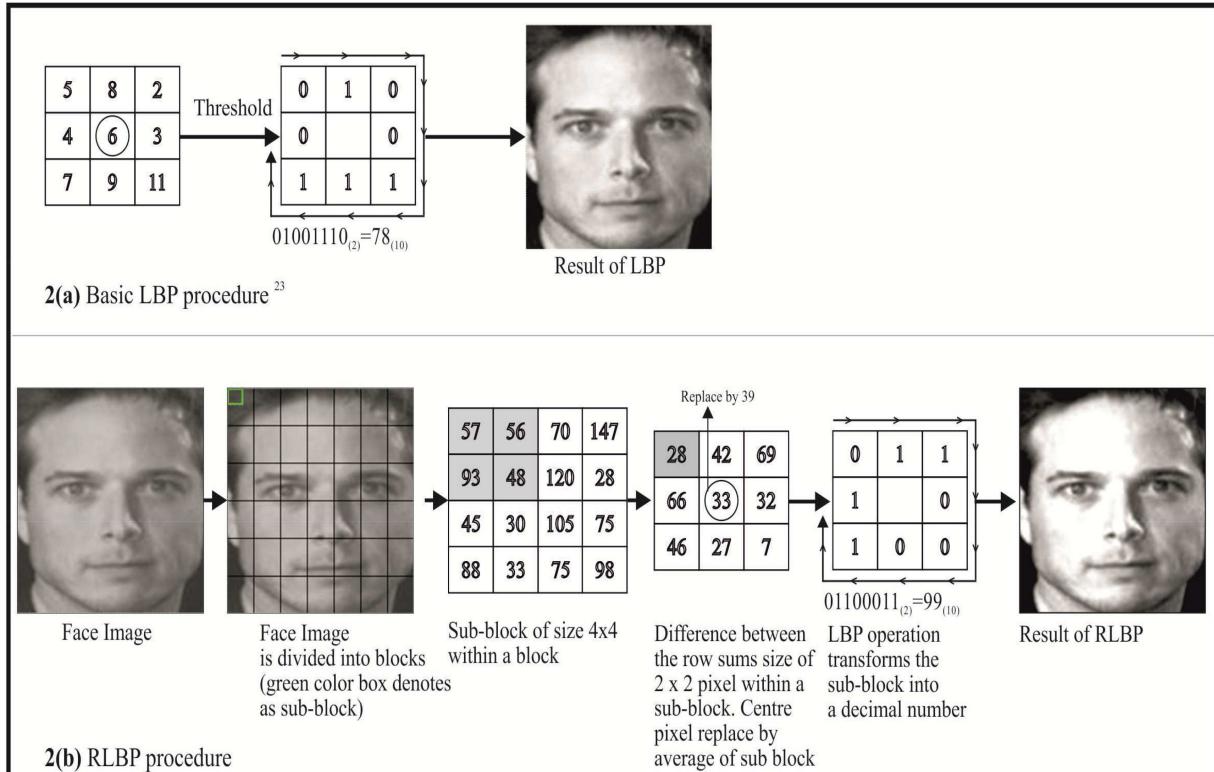


Fig. 3 LBP and RLBP procedures

Concatenate the $RLBP_f$ histograms which are extracted from all blocks in the face region into a single feature. As a result $(6 * 7 * L)2478$ RLBP histogram bins are obtained from the face region. The $RLBP_f$ histograms contain information about the description of the local micro-patterns such as edges, spots and flat areas over the whole image. Hence, it is used to statistically describe image characteristics. Similarly, we calculate difference between the column sums and diagonal sums values within a sub-block.

3.2.2 Learning discriminative RLBP histograms: During the feature extraction stage, each image contains 2478 RLBP histograms. To reduce the feature dimension and the redundant information, it is necessary to select the discriminative RLBP features. We propose to learn discriminative RLBP features for better gender recognition. AdaBoost provides a simple yet effective approach for stage wise learning of a non-linear classification function [20]. Hence, AdaBoost algorithm is adopted to learn the discriminative RLBP histogram bins. It is not only useful to select the discriminative features but also to train the classifier. AdaBoost learns a small number of weak classifiers whose performance is just better than random guessing, and boosts them iteratively into a strong classifier of higher accuracy. The process of AdaBoost maintains a distribution on the training samples. At each iteration, a weak classifier which minimises the weighted error rate is selected, and the distribution is updated to increase the weights of the misclassified samples to reduce the importance of the others. Similar to proposal in [20], the weak classifier $p_i(x)$ consists of a feature h_i which corresponds to a single RLBP histogram bin, a threshold θ_i and parity g_i indicating the direction of the inequality sign

$$p_i(x) = \begin{cases} 1, & \text{if } g_i h_i(x) \leq g_i \theta_i \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

$$RLBP(i_l, j_l) = \sum_{n_l=0}^7 T(i_l, j_l) 2^{n_l} \quad (12)$$

$$\text{where } 1 \leq i_l \leq sb_height - 1; \quad 1 \leq j_l \leq sb_width - 1, \quad i_l \neq 2 \text{ and } j_l \neq 2$$

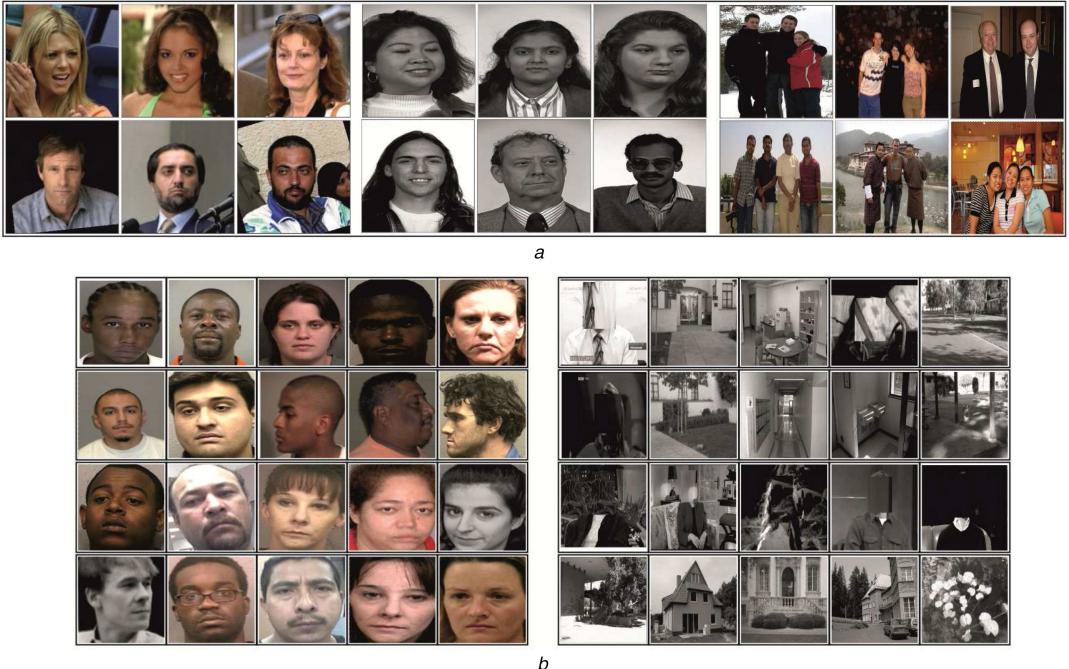


Fig. 4 Samples of database images and face and non-face images **(a)** Samples of database images, **(b)** Sample of face and non-face images

Table 1 Detection performances (%) obtained using four sets of orientation

Orientations	Recall		
	Training image resolutions		
	20 × 20	30 × 30	40 × 40
2	90	93	88
4	95	98	93
6	93	96	91
8	92	95	90

Bold values denote highest accuracy.

Table 2 Detection performances (%) obtained using four sets of scales

Scale	Recall		
	Training image resolutions		
	20 × 20	30 × 30	40 × 40
4	91	94	92
6	93	96	94
8	95	98	96
10	94	97	95

Bold values denote highest accuracy.

Each training images are scaled into three types of resolutions such as 20 × 20, 30 × 30 and 40 × 40.

4.1 Result for detection of face region

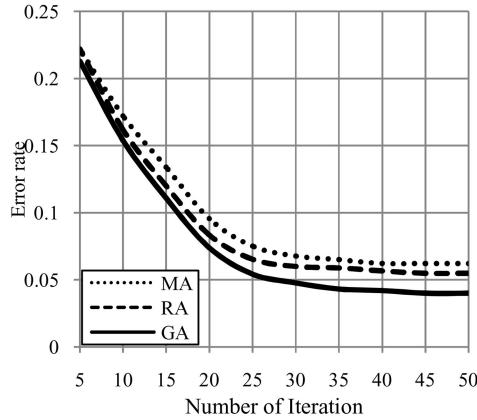
4.1.1 Performance of GW features: In our experiments, the orientation and scale of Gabor filters imposed on images are two key parameters that determine the effectiveness of the extracted texture features. Table 1 compares the detection results obtained using two, four, six and eight orientations of Gabor filters (the number of the scales is fixed at eight) using GA classifier using LFW database. The four sets of orientations are (90°, 0°), (-45°, 90°, 45°, 0°), (-45°, -22.5°, 0°, 22.5°, 45°, 90°) and (90°, 67.5°, 45°, 22.5°, 0°, -22.5°, -45°, -67.5°), correspondingly. An increase from four to six and eight orientations lowers the performance by 2 and 3%, respectively. However, eight orientations have better performances of 2% than two orientations performance at all resolutions. In addition, a decrease of the training image resolutions from 30 × 30 to 20 × 20 lowers the performance by 3%

in all different orientations. However, the 20 × 20 resolution training images have higher performance of 2% than 40 × 40 resolution training images. It can be seen that utilising the default four orientations has the highest accuracy of 98% among the four sets of orientation values in 30 × 30 resolutions of training images. This result suggests that Gabor filters require at least four orientations to capture most of the discrimination information.

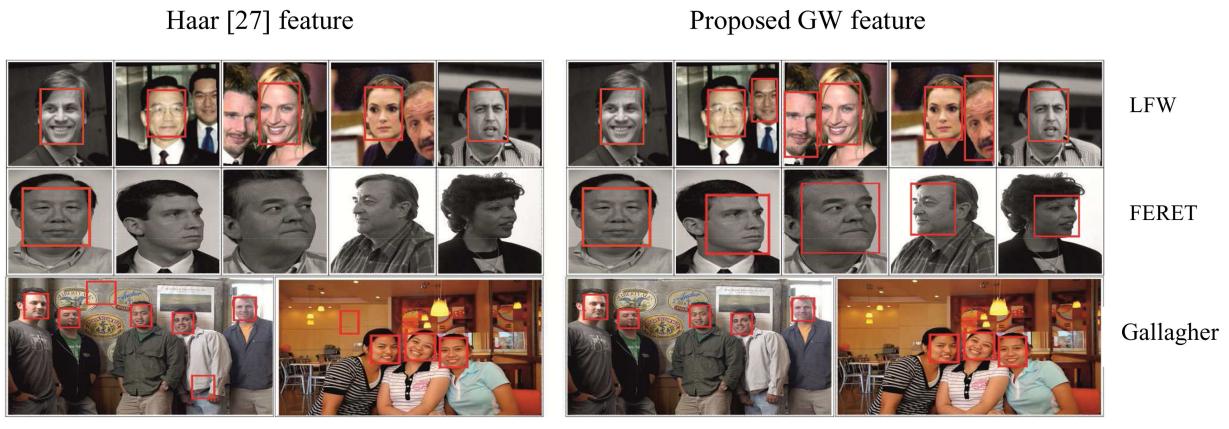
Table 2 compares the accuracies obtained using four, six, eight and ten scales of Gabor filters (the number of orientations of Gabor filters is fixed at four) in each resolution training images. The four sets of scales are composed of (5:2:11), (5:2:15), (5:2:19) and (5:2:23) pixels correspondingly. A decrease from eight to six and four scales reduces performance by 2 and 4%, respectively. However, the ten scales have higher performance of 1% than six scales performance at all resolutions of training images. Moreover, the 30 × 30 resolutions of the training images have higher performance of 2% than 40 × 40 resolutions of the training images in all orientations. The 40 × 40 resolution training images have a slight increase of 1% than 20 × 20 resolution training images. Similar to the comparison results for different orientations of Gabor filters in Table 1, using eight scales of Gabor filters deliver the highest accuracy of 98% among the four sets of scale values. In both Tables 1 and 2, 30 × 30 resolutions training images show the best result. Therefore, the training image resolutions 30 × 30 are a suitable size for detection of face. The results also confirm that the default four orientations and eight scales of Gabor filters are the optimal parameters for detecting the face.

4.1.2 Performance of classifier: The GW features are trained through variant Adaboost classifiers such as RA, GA and MA in 30 × 30 resolutions training images. They are compared for error checking with 50 boosting iterations as shown in Fig. 5a. The analysis of the GA returns error rate of 0.024 which is <0.055 and 0.063 error rates of RA and MA, respectively. Hence, GA is selected as the detection algorithm for our system. Due to space limitation, the face detection rate of 98% is displayed using LFW and 98.5 and 96.5% are obtained using FERET and Gallagher databases, respectively.

4.1.3 Performance comparison with the previous approaches for detection of face: The proposed work is carried out using a PC with Intel Core i5 @ 3.20 GHz, 8 GB RAM specification and the tool used was Matlab 2013a version for all the test cases. For comparison purpose, the existing Haar feature [27] and semi-modified census transform [28] were used. Though the Haar



a



b

Fig. 5 Error rates from three Adaboost algorithms and sample face detection results

(a) Error rates from three Adaboost algorithms in LFW database, (b) Sample face detection results in both databases

Table 3 Comparison of detection accuracy and time consumption with existing approaches

Databases	Feature	Recall, %	Precision, %	Execution time of features extraction, s	Execution time for detection of face, s
LFW	Haar [27]	94	91	0:030	0:0088
	SMCT [28]	93	90	0:053	0:0121
	GW	98	94	0:027	0:0078
FERET	Haar [27]	94.6	92	0:028	0:0084
	SMCT [28]	94	92	0:056	0:0135
	GW	98.5	95	0:021	0:0068
Gallagher	Haar [27]	92	89	2:21	0:056
	SMCT [28]	91	88	3:02	0:1
	GW	96.5	92	1:55	0:042

Bold values denote highest accuracy.

feature [27] is old one, it is a well-known method. Hence many authors [6, 12, 14, 15, 29] have recently used it for face detection. In their approach [27], they selected the salient 160,000 rectangle Haar features using AdaBoost classifier algorithm. In the proposed algorithm, 96 features are extracted using GW. The results are shown in Fig. 5b. The size of the bounding box is determined using the scale on the detected face. The Haar [27] feature did not detect the some profile view faces and different expressions on face. The results are shown in Table 3. We infer from that result, the proposed GW features achieved better detection rate and the execution time is lower than the existing methods. Moreover, it is robust against the different expression facial and pose variations.

4.2 Experimental results for gender recognition

To evaluate the gender recognition work, –ten-fold cross-validation testing scheme is adopted. In this technique, training is performed on nine splits and the performance is computed on the tenth split. This process was repeated ten times for each split in turn to be

omitted from the training process and calculated the average recognition rate in both the databases.

4.2.1 Performance analysis on different size of blocks and sub-blocks: Performance analysis are performed on different size of blocks and sub-blocks in various face image resolutions such as 60×70 , 66×77 , 72×84 , 78×91 and 84×98 . Each block size depends on the different resolutions of face images. There are different sub-blocks which are in different sizes such as 3×3 , 4×4 , 5×5 , 6×6 and 7×7 within a block. If the size of the sub-block is 3×3 , it follows the LBP. If the size is 4×4 , it is reduced to 3×3 by performing the difference between the row sums or column sums or diagonal sums of the overlapping window size of 2×2 pixels. Now, the resultant value is available as 3×3 which follows the LBP. Similarly, all other sub-blocks are also reduced to their corresponding sub-block of size 3×3 and the results are tabulated in Tables 4 and 5. From Tables 4 and 5, it may be understood that the face image resolution 72×84 and its corresponding 4×4 sub-block in row sums outperforms the different resolutions and sub-blocks

Table 4 Results of different sizes of blocks and sub-blocks in LFW database where M and F denote the male and female, respectively

Different sizes of sub-blocks within a block	Calculate difference between the row/column/diagonal sums of each overlapping window size within a sub-block										Recognition rate in %									
	Face images with different resolutions																			
	60×70		66×77		72×84		78×91		84×98		M	F	M	F	M	F	M	F	M	F
3*3 (LBP)	—	—	—	—	—	—	—	—	—	—	92	93	92	94	95	94	93	94	94	94
4*4	2*2	—	—	—	—	—	—	—	—	—	94	93	96	94	98	96	95	96	96	96
		row sums	row sums	row sums	94	95	93	94	96	94	95	95	95	94						
		column sums	column sums	column sums	92	93	91	92	93	95	94	95	95	93						
		diagonal sums	diagonal sums	diagonal sums	92	93	91	92	93	95	94	95	95	93						
5*5	3*3	—	—	—	—	—	—	—	—	—	92	94	92	94	93	92	95	94	92	95
		row sums	row sums	row sums	93	95	91	93	95	93	93	95	93	94						
		column sums	column sums	column sums	94	91	95	92	92	94	91	92	92	93						
		diagonal sums	diagonal sums	diagonal sums	94	91	95	92	92	94	91	92	92	93						
6*6	4*4	—	—	—	—	—	—	—	—	—	96	92	92	91	93	93	94	94	93	94
		row sums	row sums	row sums	94	95	96	93	94	94	95	94	95	93						
		column sums	column sums	column sums	93	96	94	96	93	95	91	93	94	95						
		diagonal sums	diagonal sums	diagonal sums	93	96	94	96	93	95	91	93	94	95						
7*7	5*5	—	—	—	—	—	—	—	—	—	95	93	92	93	92	93	94	92	93	93
		row sums	row sums	row sums	92	90	91	93	93	91	94	93	94	95						
		column sums	column sums	column sums	93	94	95	95	94	92	93	92	91	94						
		diagonal sums	diagonal sums	diagonal sums	93	94	95	95	94	92	93	92	91	94						

Bold values denote highest accuracy.

Table 5 Results of different sizes of blocks and sub-blocks in FERET database where M and F denote the male and female, respectively

Different sizes of sub-blocks within a block	Calculate difference between the row/ column/diagonal sums of each overlapping window size within a sub-block										Recognition rate in %									
	Face images with different resolutions																			
	60×70		66×77		72×84		78×91		84×98		M	F	M	F	M	F	M	F	M	F
3*3 (LBP)	—	—	—	—	—	—	—	—	—	—	93	94	93	95	96	96	95	96	96	94
4*4	2*2	—	—	—	—	—	—	—	—	—	93	95	94	95	99	98	94	96	97	95
		row sums	row sums	row sums	95	92	93	93	95	97	95	95	96	93						
		column sums	column sums	column sums	92	95	94	95	94	96	93	94	94	94						
		diagonal sums	diagonal sums	diagonal sums	96	92	95	95	94	93	95	96	97	95						
5*5	3*3	—	—	—	—	—	—	—	—	—	95	95	96	92	94	95	95	94	95	96
		row sums	row sums	row sums	94	95	96	92	94	95	95	94	95	96						
		column sums	column sums	column sums	94	93	95	93	96	95	96	95	96	96						
		diagonal sums	diagonal sums	diagonal sums	95	97	97	94	95	96	96	95	94	95						
6*6	4*4	—	—	—	—	—	—	—	—	—	95	96	96	96	94	95	96	96	95	95
		row sums	row sums	row sums	94	96	96	96	94	95	96	94	95	95						
		column sums	column sums	column sums	95	95	94	95	96	96	97	94	95	97						
		diagonal sums	diagonal sums	diagonal sums	93	92	91	90	92	91	93	93	94	96						
7*7	5*5	—	—	—	—	—	—	—	—	—	95	93	93	92	92	90	94	91	92	93
		row sums	row sums	row sums	92	94	93	92	92	91	92	93	90	92						

Bold values denote highest accuracy.

in both databases. Therefore, the face image resolution 72×84 and its corresponding 4×4 sub-block in row sums used for further experiments.

4.2.2 Performance of discriminative features selection: Fig. 6 shows the minimum error performance of the boosted strong classifier as a function of the number of features selected. It is observed that, with the 500 selected RLBP histogram bins, the boosted strong classifier achieves the minimum error rate of 0.03, 0.05 and 0.06 in calculating the difference between the row sums, column sums and diagonal sums, respectively, using LFW, 0.015, 0.04 and 0.05 results obtained using FERET and 0.06, 0.065 and 0.072 using Gallagher databases. The difference between the row sums return better minimum error rate in discriminative feature selection than the column sums and diagonal sums in both databases.

Distribution of features – selected 500 features in LFW that are distributed among the 59 bins RLBP are shown in Fig. 7a. As it is observed, the selected features come from all the 59 bins, but some bins have more contribution to discriminating features. The spatial distribution of the 500 features selected is shown in Fig. 7b. It is observed that, for gender classification, discriminative RLBP features mainly distribute in the blocks around the mouth.

4.2.3 Performance of RLBP features: Fig. 8 compares the accuracies obtained using LBP and RLBP histograms examined in both the databases. Moreover, totally 65 (male: 38 female: 27) children images are available from both LFW and Gallagher databases. The average gender recognition rate has considerably improved to about 2, 3, 5 and 2.8% by reinforcing the LBP features in databases LFW, FERET, Gallagher and children images, respectively. The gender classification based on child face is difficult when compared with adult face, as the child is in a growing phase and has few face differences based on gender. As a result, we conclude that the RLBP features result in better gender recognition rate than LBP.

Fig. 9 shows the receiver operating characteristic (ROC) curves. The curve is plotted from the results obtained using LBP and RLBP features in both databases. Fig. 9 depicts the relationship between a precision and recall. It is observed that the proposed RLBP features deliver the best performance for recognition of gender. Mainly, they have good characteristics to classify the gender with high recognition rate in a little false positive number in both databases.

When consider the male images that have moustaches or beards with or without wearing glasses, the performance of gender recognition is achieved 100% accuracy in both databases. Moreover, age recognition is estimated as stated in [30]. The

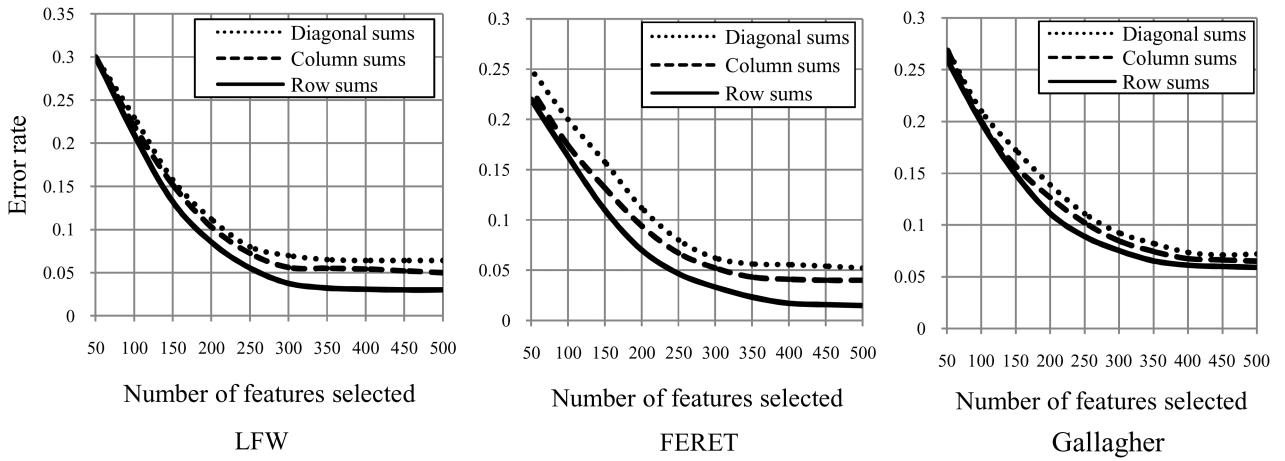


Fig. 6 Selection of discriminative features in both databases

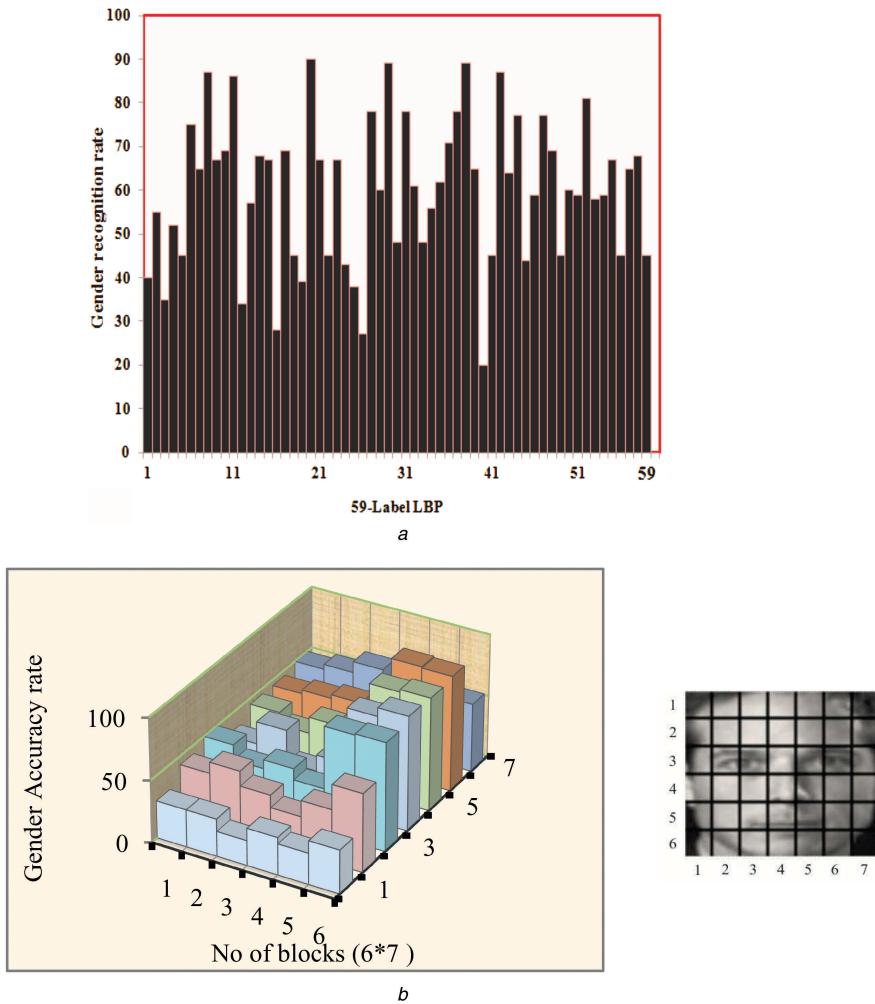


Fig. 7 Selected 500 features in LFW that are distributed among the 59 bins and spatial distribution of the 500 features selected (a) Selected 500 features that are distributed among 59 bins RLBP, (b) Spatial distribution of selected 500 features (an example face image that is divided into block is shown in the right side illustration)

performance of proposed work is achieved 82% accuracy for overall age group classification in the Gallagher's database. Age information is not provided in the LFW and FERET databases.

4.2.4 Performance comparison with the previous approaches for gender recognition: A comparison of the achieved recognition rate of the proposed RLBp features with the reported result in state-of-the-art methods is shown in Table 6. From that result, the proposed approach outperforms the existing three appearance-based methods in [15, 11, 12, 14, 31] and one geometric method

[6]. The recognition rate of the proposed approach is 2.27, 10.58, 3.85 and 8.48% higher than those used boosted LBP [15], PPH + LBP [6], HOG + LBP + GSS [11] and LBP + FPLBP [12], respectively, using LFW database. The proposed approach RLBp is 0.93% slightly lower than the existing approach and used fusion LBP features [14]; however this approach selecting the feature dimension is greater than RLBp method. Moreover, neural network based method [32, 33] have achieved lower performance than RLBp method. Generally, [32, 33] methods require extensive numbers of parameters to estimate. Here, the time complexity is increased in training phase. With few training samples, deep

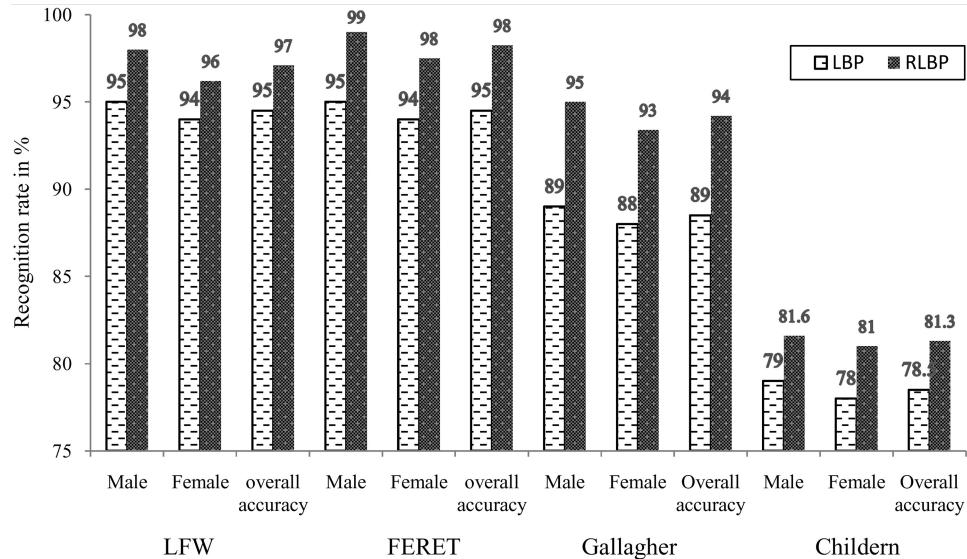


Fig. 8 Comparison of LBP and proposed RLBP features

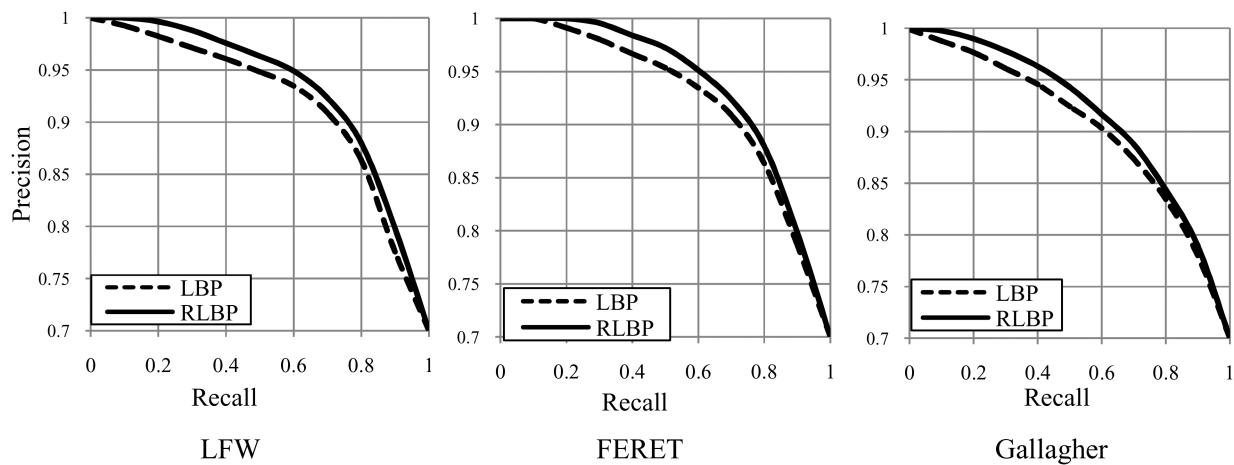


Fig. 9 ROC curve for gender recognition in both databases

Table 6 Comparisons between existing approaches and our proposed approach for gender identification where m and s denote minutes and seconds, respectively

Author	Database	Feature	Features selected	Classifier	Accuracy rate, %	Time, ms:s
Shih (2013)	LFW and FERET	PPH + bayesian	Geometric based 28 points	SVM	86.5	3:45
Shan (2012)	LFW	boosted LBP	500	SVM	94.81	2:15
Hong <i>et al.</i> (2014)	LFW	HOG + LBP + GSS	500	Adaboost	93.23	4:38
Eran <i>et al.</i> (2014)	LFW	LBP + FPLBP	—	Dropout-SVM	88.6	4:21
Juan and Claudio (2013)	LFW	fusion LBP	19,824	SVM	98.01	5:15
Torrisi <i>et al.</i> (2015)	LFW	CLBP	256	SVM	96.04	3:20
Jordi <i>et al.</i> (2016)	LFW	deep neural networks	—	—	96.25	3:45
	Gallagher				90.58	3:29
Gil and Tal (2015)	Adience benchmark convolutional neural networks		512 neurons	—	86.8	3:51
proposed	LFW	RLBP	500	Adaboost	97.08	2:14
	FERET				98.5	2:23
	Gallagher				94.21	2:42
	Children images				81.3	0:02

learning fails to learn the appropriate features. While considering the execution time, RLBP method is much lower than the existing methods. Therefore, the RLBP features have achieved the best recognition rate in less time among the several related works.

5 Conclusion and future work

This paper investigated the benefits of gender recognition based on face images. The proposed method handled the face detection and

gender identification problems at the same time in such a way that it consumes less time and good performance compared with the existing methods. The following points are enhancing the results:

- The area of the human face is detected from different backgrounds images by applying GW for local feature extraction. The Gabor features exhibit strong characteristics of spatial locality, scale and orientation selectivity that is most suitable for face/non-face classification. It also has an excellent

performance in less time. Moreover, it is robust against the different facial expression and pose variations.

- Facial features are extracted using proposed RLBP that has strong discriminative ability to classify the gender and also delivered the high recognition rate in less time. It does not require any previous geometric or intensity normalisation of the face images.

The test results exhibit that the GW and RLBP features find their appropriate role for face detection and gender recognition, respectively, on the LFW, FERET and Gallagher databases. The current work for recognising the gender does not utilise temporal information and occlusion of the facial image, which will be addressed in future.

6 References

- [1] Santarcangelo, V., Farinella, G.M., Battiato, S.: ‘Gender recognition: methods, dataset and results’. Int. Workshop on Video Analytics for Audience Measurement, Torino, 2015, pp. 1–6
- [2] Toews, M., Arbel, T.: ‘Detection, localization and sex classification of faces from arbitrary viewpoints and under occlusion’, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2009, **31**, (9), pp. 1567–1581
- [3] Moghaddam, B., Yang, M.H.: ‘Learning gender with support faces’, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, (5), pp. 707–711
- [4] Brunelli, R., Poggio, T.: ‘HyperBF networks for gender classification’. Proc. DARPA Image Understanding Workshop, 1995, pp. 311–314
- [5] Lahoucine, B., Boulbaba, B.A., Mohamed, D.: ‘Boosting 3-D-geometric features for efficient face recognition and gender classification’, *IEEE Trans. Inf. Forensics Sec.*, 2012, **7**, (6), pp. 1766–1779
- [6] Shih, H.C.: ‘Robust gender classification using a precise patch histogram’, *Pattern Recognit.*, 2013, **46**, (2), pp. 519–528
- [7] Wu, M., Zhou, J., Sun, J.: ‘Multi-scale ICA texture pattern for gender recognition’, *IET Electron. Lett.*, 2012, **48**, (11), pp. 629–631
- [8] Jeffrey, B., Flora, D., Lochtefeld, F., et al.: ‘Improved gender classification using nonpathological gait kinematics in full-motion video’, *IEEE Trans. Hum.-Mach. Syst.*, 2015, **45**, (3), pp. 304–314
- [9] Hamid, H., Amin, Z., Avishan, N., et al.: ‘Gender classification based on fuzzy clustering and principal component analysis’, *IET Comput. Vis.*, 2016, **10**, (3), pp. 228–233
- [10] Darry, S., Adrian, P., Jianguo, Z.: ‘Gender classification via lips: static and dynamic features’, *IET Biometrics*, 2013, **2**, (1), pp. 28–34
- [11] Hong, L., Yuan, G., Can, W.: ‘Gender identification in unconstrained scenarios using self-similarity of gradients features’. IEEE ICIP, 2014
- [12] Eran, E., Roe, E., Tal, H.: ‘Age and gender estimation of unfiltered faces’, *IEEE Trans. Inf. Forensics Sec.*, 2014, **9**, (12), pp. 2170–2179
- [13] Moeini, A., Faez, K., Moeini, H.: ‘Real-world gender classification via local Gabor binary pattern and three-dimensional face reconstruction by generic elastic model’, *IET Image Process.*, 2015, **9**, (8), pp. 690–698
- [14] Juan, E., Claudio, A.: ‘Gender classification based on fusion of different spatial scale features selected by mutual information from histogram of LBP, intensity, and shape’, *IEEE Trans. Inf. Forensics Sec.*, 2013, **8**, (3), pp. 488–499
- [15] Shan, C.: ‘Learning local binary patterns for gender classification on real-world face images’, *Pattern Recognit. Lett.*, 2012, **33**, (4), pp. 431–437
- [16] Lee, T.S.: ‘Image representation using 2D Gabor wavelets’, *IEEE Trans. PAMI*, 1996, **18**, (10), pp. 959–971
- [17] Wu, B., Ai, H., Huang, C.: ‘Fast rotation invariant multi-view face detection based on RealAdaboost’. Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition, 2004, pp. 79–84
- [18] Friedman, J., Hastie, T., Tibshirani, R.: ‘Additive logistic regression: a statistical view of boosting’, *Ann. Stat.*, 2000, **28**, (2), pp. 337–407
- [19] Vezhnevets, A., Vezhnevets, V.: ‘Modest AdaBoost-teaching AdaBoost to generalize better’. Graphicon, Novosibirsk Akademgorodok, 2005
- [20] Viola, P., Jones, M.: ‘Rapid object detection using a boosted cascade of simple features’. Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2001, pp. 511–518
- [21] Huang, G., Ramesh, M., Berg, T., et al.: ‘Labeled faces in the wild: a database for studying face recognition in unconstrained environments’. Technical Report, University of Massachusetts, Amherst, 2007, pp. 07–49
- [22] Phillips, P.J., Moon, H., Rizvi, S.V., et al.: ‘The FERET evaluation methodology for face-recognition algorithms’, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000, **22**, (10), pp. 1090–1104
- [23] Gallagher, A.C., Chen, T.: ‘Understanding images of groups of people’. Proc. IEEE Conf. Computer Vision Pattern Recognition, 2009, pp. 256–263
- [24] Pablo, D., Daniel, G., Long, Y., et al.: ‘Single- and cross- database benchmarks for gender classification under unconstrained settings’. IEEE Int. Workshop on Benchmarking Facial Image Analysis Technologies, 2011
- [25] Edoardo, A., Marco, L.C., Marco, M.: ‘Probabilistic corner detection for facial feature extraction’. Image Analysis and Processing – ICIAP, 2009, vol. **5716**, pp. 461–470
- [26] Ngoc-Son, V., Alice, C.: ‘Face recognition with patterns of oriented edge magnitudes’. Computer Vision – ECCV, 2010, vol. **6311**, pp. 313–326
- [27] Viola, P., Jones, M.: ‘Robust real-time face detection’, *Int. J. Comput. Vis.*, 2004, pp. 137–154
- [28] Kyungjoong, J., Choi, J., Jang, G.: ‘Semi-local structure patterns for robust face detection’, *IEEE Signal Process. Lett.*, 2015, **22**, (9), pp. 1400–1403
- [29] Zhong, L., Liu, Q., Yang, P., et al.: ‘Learning multiscale active facial patches for expression analysis’, *IEEE Trans. Cybern.*, 2015, **45**, (8), pp. 1499–1510
- [30] Alnajar, F., Shan, C., Gevers, T., et al.: ‘Learning-based encoding with soft assignment for age estimation under unconstrained imaging conditions’, *Image Vis. Comput.*, 2012, **30**, (12), pp. 946–953
- [31] Torrisi, A., Farinella, G.M., Puglisi, G., et al.: ‘Selecting discriminative CLBP patterns for age estimation’. Int. Workshop on Video Analytics for Audience Measurement, Torino, 2015, pp. 1–6
- [32] Jordi, M., Alberto, A., Roberto, P.: ‘Local deep neural networks for gender recognition’, *Pattern Recognit. Lett.*, 2016, **70**, pp. 80–86
- [33] Gil, L., Tal, H.: ‘Age and gender classification using convolutional neural networks’. IEEE Workshop on Analysis and Modeling of Faces and Gestures, at the Conf. on Computer Vision and Pattern Recognition, 2015