ORIGINAL PAPER

Glasses detection on real images based on robust alignment

Alberto Fernández[1] · Rodrigo García[1] · Rubén Usamentiaga[2] · Rubén Casado[3]

Abstract Automatic glasses detection on real face images is a challenging problem due to different appearance variations. Nevertheless, glasses detection on face images has not been thoroughly investigated. In this paper, an innovative algorithm for automatic glasses detection based on Robust Local Binary Pattern and robust alignment is proposed. Firstly, images are preprocessed and normalized in order to deal with scale and rotation. Secondly, eye glasses region is detected considering that the nosepiece of the glasses is usually placed at the same level as the center of the eyes in both height and width. Thirdly, Robust Local Binary Pattern is built to describe the eyes region, and finally, support vector machine is used to classify the LBP features. This algorithm can be applied as the first step of a glasses removal algorithm due to its robustness and speed. The proposed algorithm has been tested over the Labeled Faces in the Wild database showing a 98.65 % recognition rate. Influences of the resolution, the alignment of the normalized images and the number of divisions in the LBP operator are also investigated.

✉ Alberto Fernández
alberto.fernandez@fundacionctic.org

Rodrigo García
rodrigo.garcia@fundacionctic.org

Rubén Usamentiaga
rusamentiaga@uniovi.es

Rubén Casado
ruben.casado@treelogic.com

[1] Fundación CTIC (Technological Center), Technological Scientific Park of Gijón, 33204 Gijón, Asturias, Spain

[2] Department of Computer Science and Engineering, University of Oviedo, Campus de Viesques, 33204 Gijón, Asturias, Spain

[3] Treelogic, Technological Scientific Park of Asturias, 33428 Llanera, Asturias, Spain

1 Introduction

Automatic face recognition is an attractive research issue due to its great potential in real-life applications. However, robust face recognition is not a trivial task and occlusions (like wearing glasses) can decrease the performance of an automatic face recognition system. In some cases, certain glasses can identify a person or in other cases, act like soft biometrics for person identification (matrix glasses, U2 glasses). Automatic gender recognition suffers from the same problems as face recognition. However, there is certain appearance difference between men glasses and women glasses, which could help to distinguish a person's gender. Even, glasses appearance between a young person and an older one could improve an automatic age classification algorithm rather than decrease its performance. Many security systems require the use of glasses, their use is mandatory, and intelligent advertising and marketing could also take advantage of this. Last but not least, glasses can provide valuable information. For instance, glasses could reveal the activity a person is doing at a given time (safety glasses, sport glasses, snow glasses, sunglasses, underwater glasses, diving mask, reading glasses, eyeglasses, Google glasses, etc).

Automatic recognition of eyeglasses on facial images needs a thorough study. Different types of glasses on real images can be seen in Fig. 1. The inherent appearance variation on glasses frames, which are even not clearly distinguishable (Fig. 1d), makes it a challenging problem. Further challenges are the different types of glasses: sport glasses (Fig. (1a, b, h) or sunglasses (Fig. 1g, i), different sizes (Fig. 1e, f), reflections and glare (Fig. 1c) or rotation of the face image (Fig. 1d).

(a) **(b)** **(c)**

(d) **(e)** **(f)**

(g) **(h)** **(i)**

Fig. 1 Glasses can convey information about the user's context. Different variations in glasses appearance can be seen. These images are obtained from LFW database

Summarizing, the main areas that can benefit from an automatic eyeglasses recognition on facial images would be the use of safety glasses in security systems, intelligent advertising and marketing, soft biometrics for person identification, face recognition, gender classification, age classification, hints of activity detection based on type of glasses, Google glasses recognition and first step of automatic glasses removal.

The main contributions are summarized as follows. Firstly, the problem of eyeglasses detection needs an up-to-date review. Main studies on glasses detection are summarized and analyzed. Secondly, a LBP-based algorithm is developed to cope with the glasses detection problem and a simple yet robust algorithm is proposed in order to deal with face alignment normalization. Lastly, (1) influences of resolution of the image, (2) alignment of faces, (3) number of divisions of the LBP operator and (4) a performance comparison between LBP and Robust LBP are analyzed.

2 Related work

2.1 Glasses detection, extraction and removal

Facial image analysis has stimulated increasing interest, including tasks as face detection, face recognition, facial expression analysis, demographic classification and other related topics. Although there are less research papers focusing on glasses detection, there are some studies in this regard (a short description of the results of main studies on glasses detection can be seen in Table 1).

In [1], six measures for the likelihood of glasses in different regions for detecting the presence of glasses are introduced. Furthermore, a combination of them turns out to consistently improve the performance of the individual measures. Experiments showed that measure from the region containing the nosepiece of the glasses is the most powerful criterion for glasses recognition. In another work [3], glasses detection and extraction are proposed. Detection is carried out using edge information within a small area defined between the eyes. Detection is similar to the method proposed in [1]. As mentioned in [1], the nosepiece is one of the most common features existing on all of the glasses. Extraction is achieved with a deformable contour, combining edge and geometrical features. In [11], glasses detection and removal for face recognition using Bayes rules are proposed. Experiments within a face image database showed that this method is effective and has a better performance than that proposed in [3]. Extraction is realized using Bayes rules that incorporate the features around each pixel and the prior knowledge on glasses features that were learnt and stored in a database. Glasses removal is achieved with an adaptive median filter conducted in the points classified as glasses.

A method that makes use of the 3D features obtained by a trinocular stereo vision system is proposed in [6] to perform glasses frame detection with 3D Hough transform. It is based on the fact that the rims of a pair of glasses lie on the same plane in 3D space. This method requires more cameras and computational time than other 2D image-based methods. In [7], a framework for classifying facial attributes is presented. Similarity is measured by the reconstruction error motivated by the Eigenface-based methods, and thus, this paper extends the Eigenface method by representing each attribute by using a 'facial attribute-specific subspace (FASS)'.

Wavelet feature-based boosting methods can also be applied to glasses recognition [5]. This method was originated from the work on face detection [9]. The glasses detectors use a variation of the original boosting algorithm called real AdaBoost [12]. Authors suggest that the nosepiece of the glasses frame is an important clue for glasses detection. An algorithm for eyeglasses detection, localization and removal is proposed in [8]. An eye region detector, trained offline, is used to approximately locate the region of the eyes and thus the region of the eyeglasses. The location of the eyeglasses pattern is defined as twice as big as the detected eye area. Afterward, eyeglasses localization is done with a Markov chain Monte Carlo (MCMC) method. A Delaunay triangulation-based method is proposed to locate glasses from a front-up face image [13]. They decompose

Table 1 Results of main studies on glasses detection from face images

Jiang et al. [1]	Forerunner algorithm on detection of glasses in facial images. Results are reported similar in [2]
Dantcheva et al. [2]	Train: FERET database. Test performed on a testing set of images on FERET database. Detection rate: 87.7 %. True-positive rate: 7.17 %. False- positive rate: 5.66 %
Jing et al. [3]	Tested on a set of 419 people, 151 wearing glasses under different imaging conditions. Detection rate: 99.52 %, falsely detecting the presence of glasses in two facial images
Vaquero et al. [4]	1044 internet images were used for training (308 sunglasses, 318 eyeglasses, 418 no glasses). Tested on a surveillance video. Sunglasses (precision: 85 %, recall: 35 %), eyeglasses (precision: 90 %, recall: 8.9 %), no glasses (precision: 77 %, recall: 99.5 %)
Wu et al. [5]	3000 images (1500 wearing glasses and 1500 with no glasses) obtained from FERET database and world wide web were used for fivefold cross-validation. Best results were obtained using gabor boost features. Accuracy: 98.9 % (98.1 % with glasses and 99.7 % without glasses). They implemented an automatic glasses detection system whose input is real-life photographs and the overall correctness is 94.0 %
Wu et al. [6]	513 facial images of 19 people wearing three kinds of glasses and nine kinds of pose were used for training. Detection rate: ≈90 %
Shan et al. [7]	100 face images from Bern face database were used for training: 50 examples from five subjects wearing glasses were used to learn glasses-FASS, and 50 examples from five subjects no wearing glasses were used to learn non-glasses-FASS. Tested on 200 images from Bern face database with an average correct rate of 77 %
Wu et al. [8]	They train a detector for the eye region based on the face detector [9] with 12710 samples. They tested the detector with 1386 face images. Detection rate: 96 % and false alarm at 10^{-4}
Heo et al. [10]	Correct detection rate was 86.6 % for the subjects wearing eyeglasses. For the face images with no eyeglasses, 97.1 % true negative accuracy was achieved. False-positive and false-negative errors were 2.9 and 13.4 %, respectively. The database used in this experiment is comprised of thermal images from the database developed by the National Institute of Standards and Technology (NIST) and Equinox Corporation

the face shape using the Delaunay triangulation. In the first step of the algorithm, they perform a binarization of the face. This method can have some problems if glass frames have very low contrast. In addition, they do not obtain very good results if the glasses do not cover eyes. Data-fused image composed of visual and thermal ones is used in [10] to produce an integrated image for eyeglass removal in order to perform a robust face recognition. The eyeglass region in thermal face images can be represented by two ellipses.

A method for removing glasses from a human frontal face image is proposed [14]. Glasses region is automatically extracted using color and shape information. They perform glasses removal, but not glasses detection. Firstly, glasses region is automatically extracted using color and shape information, and then, a natural looking facial image without glasses is generated by means of recursive error compensation using PCA reconstruction. They carried out some experiments, and the recognition performance is improved by using glasses removal methods. In [4], authors trained nine Viola and Jones detectors, one for each facial attribute. Glasses type is performed. They distinguish three types: sunglasses, eyeglasses and the absence of them.

Soft biometrics (eyeglasses, for example) can be used for person identification [2]. They implement an algorithm deduced from [1], which performs edge detection on a pre-processed gray-level image.

2.2 Face alignment and normalization

Alignment step of a face recognition algorithm is often ignored or not detailed. In some cases, the alignment is done manually, where the positions of the eyes are manually labeled. In other cases, this step is ignored, under the assumption that the face detection algorithm will perform some kind of alignment. Many recognition algorithms depend on an accurate positioning of the face region into a canonical pose before the recognition step takes place. This alignment process can lead to improve recognition accuracy on real images, even for algorithms robust to misalignments [16]. In [17], authors propose a method for aligning images using poorly aligned examples of a class with no additional data generating a alignment machine for that object class. In [18], authors use a combination of unsupervised joint alignment with unsupervised feature learning in order to align images.

Some studies are done in relation to the alignment of the images. It was found that automatic alignment methods did not increase gender classification rate, while manual alignment increased the classification rate a little [19]. Therefore, they concluded that automatic alignment needs improving. This statement, valid at that time, should be reviewed, because automatic alignment methods have improved in both accuracy and robustness.

A face normalization algorithm based on eyes detection is proposed [20]. The algorithm detects the position of the pupils in the face image. After that, the algorithm normalizes the orientation, the scale and the gray scale of the face

image. Face region is adjusted based on the position of the mouth. This algorithm has the drawback it needs to locate the position of the pupils, and thereby, glasses and partial occlusions can alter the result of the normalization step. Another face normalization algorithm is proposed [21]. The eyes are located manually in the faces, and faces are rotated and aligned in the images so that each face has eyes in the same location. After the normalization step, images are scaled to a fixed size.

In this work, a face alignment and normalization algorithm is proposed. In order to get the angle of misalignment, a regression line is calculated based on four points of the eyes. There are some automatic facial feature points detector [22], [23], but the detector of facial landmarks used in this paper [15] takes advantage of relationship among some facial features of the face (eyes, nose and mouth) in terms of appearance and structure distribution. In this way, this algorithm can estimate the position of the eyes under occlusion (like wearing glasses or sunglasses).

3 Methodology

In Fig. 2, the main steps of the algorithm, which are discussed below, can be seen. First of all, this work starts with location of the face using the algorithm by Viola and Jones [9]. After the face is located in the image, some preprocessing is necessary in order to deal with pose, rotation, scale and inaccuracies of the located face. A face normalization algorithm is applied to get the region around the eyes. Afterward, LBP and Robust LBP are applied in order to get the feature sets. Finally, support vector machine (SVM) is applied on the classification step. SVM is applied to classify the extracted feature histograms over the normalized eyes glasses regions. The output of the SVM classifier is a two-class classification problem, i.e., glasses vs no glasses. Therefore, different type of glasses (sport glasses, sunglasses, safety glasses, reading glasses, etc.) belong to the same category. SVMs are a useful technique for data classification and have been proven useful in a number of pattern recognition tasks including face and facial recognition [24]. LIBSVM was used for the training and testing of SVMs [25].

3.1 Face alignment and normalization

3.1.1 Landmarks relationship to estimate eyes position

When the shape or intensity characteristics of the eyes cannot be reliably measured due to occlusion (like wearing glasses or sunglasses), the context characteristics are very useful for eye localization. This is because eyes in the face context usually have stable relationship with other facial features in terms of both appearance and structure distribution [26]. In this way,
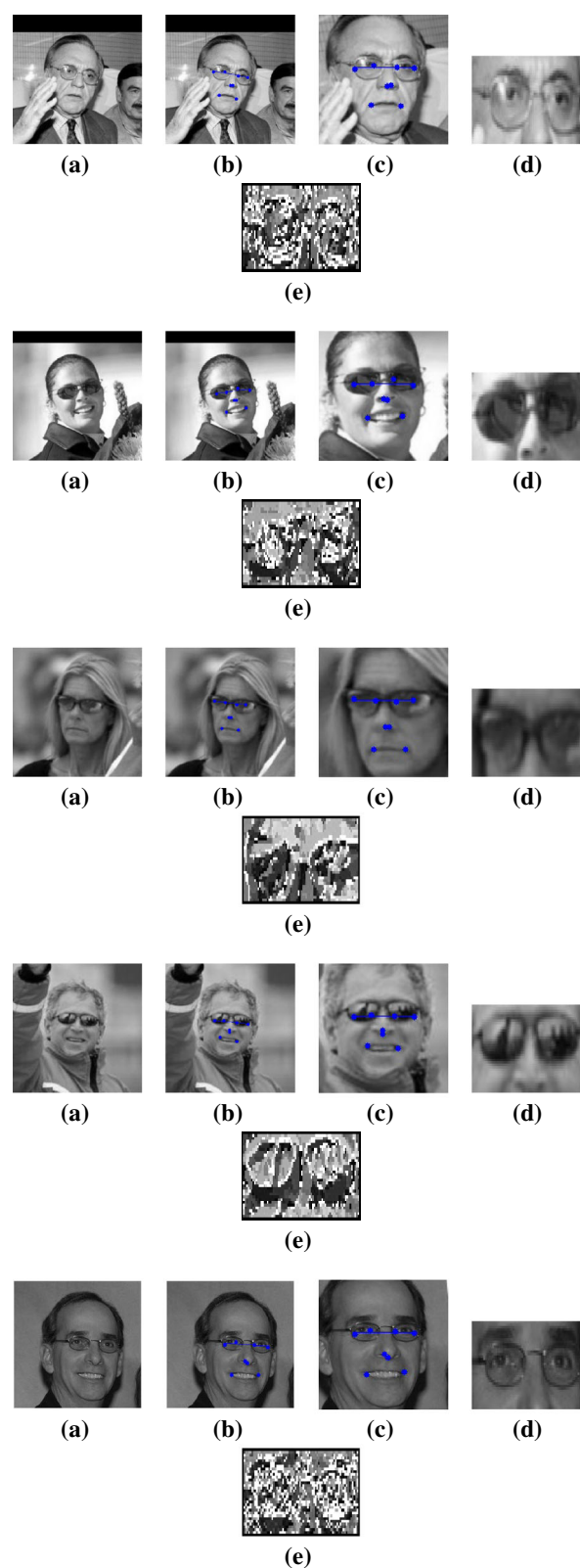

Fig. 2 Pipeline of the algorithm for glasses detection on facial images: **a** input image obtained from LFW database, **b** landmarks detection and regression line, **c** rotated face, **d** normalized eye glasses region and **e** Local Binary Pattern
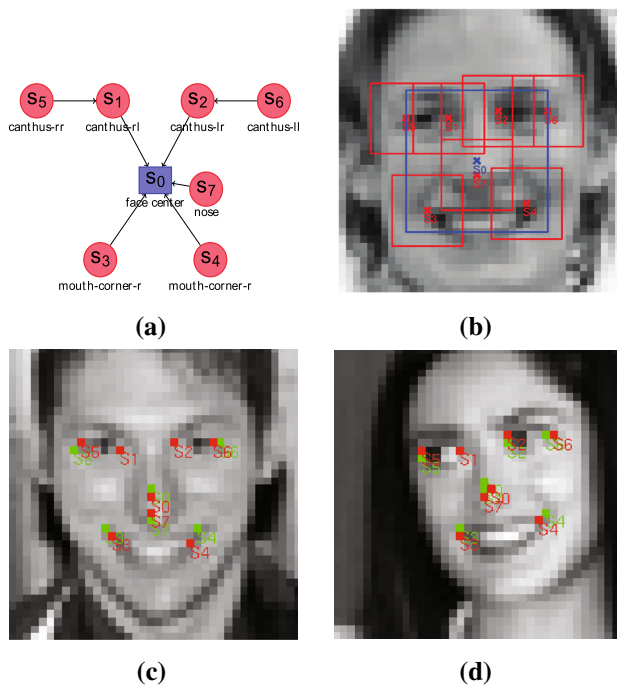
Fig. 3 Definition of **a** underlying graph for the landmark configuration and **b** the components of the detector. Sample images **c** and **d** where the estimated landmark positions have the mean normalized deviation equal to 10 %. The *green* and *red* points denote the manually annotated and estimated landmarks, respectively. These images are obtained from [15] (color figure online)

a detector of facial landmarks learned by structured output SVM [15] has been applied in order to detect the positions of the main points of the face and hence of the eyes.

The input of the classifier is a still image of fixed size containing a single face. The output of the classifier is estimated locations of a set of facial landmarks. A graph constraint with eight components is taken into account see Fig. 3a. Each facial feature is computed on a rectangular window see Fig. 3b. The accuracy of the classifier has been tested on LFW database [27]. The detector estimates around 97 % of the images with the mean normalized deviation between the estimated and the ground truth positions <10 %. In order to clarify this measure, two sample images (see Fig. 3c, d) are shown, where the estimated landmark positions have the mean normalized deviation equal to 10 %.

3.1.2 Procedure

In order to perform a robust, fully automatic ROI normalization a new algorithm is proposed in Algorithm 1. The input of the normalization algorithm is a still image. This image is resized to a certain size. Different tests are performed in order to know how this size affects the recognition rate (see Sect. 4). This algorithm is based on [21] but fully automatic, where:

- *eyes_distance_r* is the ratio of the distance from left outer eye to right outer eye in the resized image.
- *eye_line_r* is the ratio of the height above and below eyes.
- *size* is the size of the resized face image
  - *size.w* is the width of the resized image
  - *size.h* is the height of the resized image

Main steps are summarized as follows. Eight facial landmarks are obtained [15]. In order to get the angle of misalignment, a regression line is calculated based on four points: the canthi of the left and right eye, i.e., left inner eye, left outer eye, right inner eye and right outer eye (see Fig. 4). After that, faces are rotated and aligned in the images so that eyes are located in the same coordinates for all the images. The area around the eyes is calculated in source image. Once the coordinates for the corners of the eye glasses region in the rotated image are calculated, the resulting area is cropped from the rotated image.

Algorithm 1 Face ROI normalization

eyes_distance_r, eye_line_r, size

1: Eight facial landmarks are located. Linear regression is calculated based on the four landmarks of the eyes. Angle α of misalignment is calculated, and image is rotated based on α in order to align the source image
2: Euclidean distance d_0 between the eyes is calculated in the rotated image
3: Distance of the eyes in the resized image is calculated by $d_t = size.w * eyes\_distance\_r$
4: Ratio r is calculated by $r = d_0/d_t$
5: Width w_0 and height h_0 of the area around the eyes are calculated by $w_0 = r * size.w$ and $h_0 = r * size.h$
6: Coordinates for the corners of the face area in the rotated image are calculated by $x_l = x_e - w_0/2$, $y_t = y_e - h_0/eye\_line\_r$, $x_r = x_l + w_0$ and $y_b = y_t + h_0$, where x_l is x-coordinate of the left border, x_e is x-coordinate of the point in the halfway between the eyes, y_t is y-coordinate of the top border, y_e is y-coordinate of the eyes, x_r is x-coordinate of the right border, and y_b is y-coordinate of the bottom border. This ROI is cropped from the image
7: **return** ROI

3.2 Local Binary Pattern

The Local Binary Pattern operator is a type of feature used for classification. It has been found to be a powerful feature for texture classification. It was introduced in 1996 as a means of summarizing local gray-level structure [28]. The operator takes a local neighborhood around each pixel, thresholds the pixels of the neighborhood at the value of the central pixel and uses the resulting binary-valued image patch as a local image descriptor. It was originally defined for 3×3 neighborhoods, giving eight bit codes based on the eight pixels around the central one. Formally, the LBP operator takes the form:

Fig. 4 Linear regression based on four landmarks of the eyes
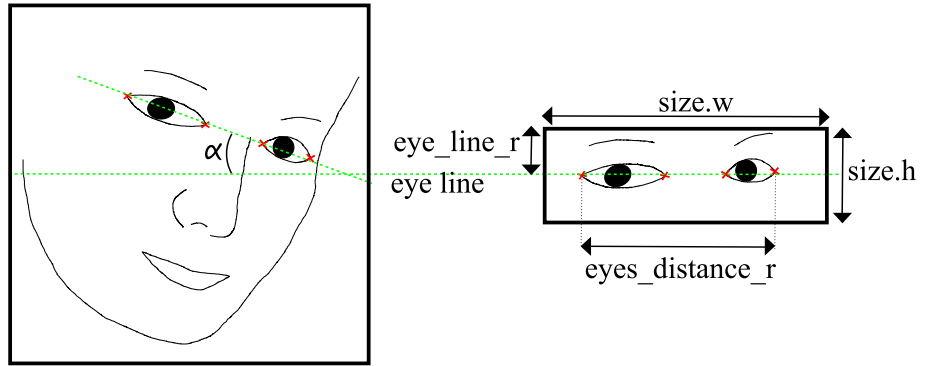
$$\text{LBP } (x_c, y_c) = \sum_{p=0}^{7} 2^p s(g_p - g_c) \tag{1}$$

where in this case, p runs over the eight neighbors of the central pixel c, g_c and g_p are the gray-level values at c and p and

$$s(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

The LBP encoding process is illustrated in Fig. 5. The operator was extended to use neighborhoods of different sizes [29], thus making it feasible to deal with textures at different scales. This is denoted by (P, R) where P represents the number of sampling points and R represents the radius of the neighborhood. When the sampling points do not fall at integer positions, the intensity values are bilinearly interpolated. See Fig. 6 for an example of the circular $(8, 2)$ neighborhood. This implementation is called circular LBP (LBP$_{P,R}$):

$$\text{LBP}_{P,R} (x_c, y_c) = \sum_{p=0}^{P-1} 2^p s(g_p - g_c) \tag{3}$$

Another extension to the original operator [29] defined the so-called uniform patterns: An LBP is 'uniform' if it contains at most two bitwise transitions from 0 to 1 or vice versa when viewed as a circular bit string. For example, 00000000, 00011110 and 10000011 are uniform patterns. Uniformity is an important concept in the LBP methodology, representing primitive structural information such as edges and corners. Although only 58 of the 256 8-bit patterns are uniform, nearly 90 % of all observed image neighborhoods are uniform [29]. The following notation for the uniform LBP operator is used: LBP$_{P,R}^{u2}$. Because the gray value of the central pixel is used as threshold, LBP is sensitive to noise, especially in the near-uniform image regions. Another demerit of LBP is that many different structural patterns may have the same LBP code. In order to enhance the discriminative capability of the local structure, Completed LBP (CLBP)
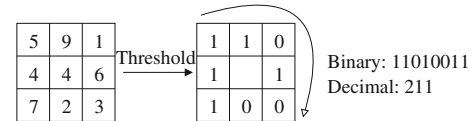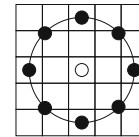

Fig. 5 The basic LBP operator

Fig. 6 The circular $(8, 2)$ neighborhood. The pixel values are bilinearly interpolated whenever the sampling point is not in the center of a pixel

was proposed [30]. Although CLBP solves some confusion of different patterns, not all of these patterns can be differentiated perfectly. Besides, CLBP is sensitive to noise since the value of a pixel is still used as a threshold directly.

In order to solve these difficulties, Completed Robust Local Binary Pattern (CRLBP) was proposed [31]. This descriptor is used to extract the region glasses information. Firstly, Average Local Gray Level (ALG) is calculated as follows:

$$\text{ALG} = \frac{\sum_{i=1}^{8} g_i + g}{9} \tag{4}$$

where g represents the gray value of the center pixel and g_i denotes the gray value of the neighbor pixel. ALG represents the average gray level of local texture, which is obviously more robust to noise than the gray value of the central pixel. ALG ignores the specific value of an individual pixel, while sometimes the specific information of the central pixel is needed. To make a balance between anti-noise and information of individual pixel, a Weighted Local Gray Level (WLG) is defined as follows:

$$\text{WLG} = \frac{\sum_{i=1}^{8} g_i + \alpha g}{8 + \alpha} \tag{5}$$

where α is a parameter set by user. Now the RLBP can be calculated as follows:

$$\text{RLBP}_{P,R} = \sum_{p=0}^{P-1} 2^p s(g_p - \text{WLG}_c) \qquad (6)$$

That is the same as:

$$\text{RLBP}_{P,R} = \sum_{p=0}^{P-1} 2^p s\left(g_p - \frac{\sum_{i=1}^{8} g_i + \alpha g_c}{8 + \alpha}\right) \qquad (7)$$

After labeling an image with the different LBP operators, a histogram of the labeled image $f_l(x, y)$ can be defined as:

$$H_i = \sum_{x,y} I\{f_l(x, y) = i\}, \quad i = 0, \ldots, n - 1 \qquad (8)$$

where n is the number of different labels produced by the LBP operator and:

$$I\{A\} = \begin{cases} 1, & \text{if } A = \text{ true} \\ 0, & \text{otherwise} \end{cases} \qquad (9)$$

For efficient face representation, feature extracted should retain also spatial information. Hence, the facial image is divided into m regions $\{R_0, \ldots, R_{m-1}\}$. In this way, the basic histogram defined above can be extended into a *spatially enhanced histogram* [32], which encodes both the appearance and the spatial relations of facial regions. The spatially enhanced histogram is defined as:

$$H_{i,j} = \sum_{x,y} I\{f_l(x, y) = i\} I\{(x, y) \in R_j\} \qquad (10)$$

where $i = 0, \ldots, n - 1$, $j = 0, \ldots, m - 1$. The histogram has a description of the normalized eye glasses region on three different levels of locality: The 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 [32]. In conclusion, the extracted feature histogram represents the local texture and global shape of the normalized eye glasses region.

4 Results

A series of experiments are conducted on LFW database [27]. LFW database is a database originally created to study the problem of unconstrained face recognition, which contains 13,233 face images of 5749 distinct subjects collected from the web. This database has also recently been used as benchmark for another attribute recognition algorithms [33], [34].

In this paper, the use of this database as benchmark for glasses recognition in real images is carried out. Three thousand face images (1500 wearing glasses, 1500 not wearing glasses) are considered in the experiment; see Fig. 1 for some examples. All experimental results were obtained using the commonly used fivefold cross-validation in order to use standardized practices [33].

In order to get a robust and computationally lightweight (efficient) algorithm, a good trade-off between recognition performance and feature vector length should be addressed. There are some parameters that can be chosen to optimize this algorithm: LBP operator, radius of the operator, number of neighbors, multiresolution analysis and number of divisions.

Uniform patterns produces 59 labels for a neighborhood of eight pixels and produces 256 labels for standard LBP. Additionally, for the 16 neighborhoods, the numbers are 243 and 65,536, respectively. In experiments with facial images [32], it was 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. Using uniform patterns instead of all the possible patterns has produced better recognition results in many applications [35]. By varying sampling radius R, LBP of different resolutions can be obtained, and thus, multiresolution analysis can be accomplished by combining the information provided by multiple operators varying (P, R) [36]. However, for most of the existing work, LBP with a fixed radius ($R = 2$) was applied. A small radius of the operator makes the information encoded in the histogram more local [37].

Preliminary tests carried out in this work showed that $\text{LBP}_{8,2}^{u2}$ gives better results than other combinations. Mean recognition rates achieved on these preliminary tests for the LBP indicated that multiresolution analysis is not necessary for glasses detection.

4.1 Resolution and size of the normalized region

In many practical video surveillance applications, the faces acquired by outdoor cameras are of low resolution [38], so tests are carried out in order to establish the minimum resolution of the images. First of all, taken into account the set of selected images from LFW database, the average size of the facial regions is calculated to establish an approximate size of the facial region in the algorithm. The average distance from left outer eye to right outer eye is 63.86 pixels and a standard deviation of 5.39 pixels $\left(5.39 = \sqrt{\frac{\sum(63.86-x)^2}{(3000-1)}}\right)$. In order to decrease the amount of data to process and to avoid false-positive recognition, eye glasses region is selected taken into account only the eyes zone. A symmetrical region around the center of the eyes is selected considering that the nosepiece of the glasses is usually placed at the same level as the center of the eyes in both height and width. The height of the eye glasses region is heavily determined by the size of the

Fig. 7 Normalized eye glasses regions at different resolutions. Original image corresponds with Fig. 1e and normalized eyes glasses region matches with Fig. 13k

(a) 100 x 45 normalized eye glasses region (maximum size)

(b) 80 x 36 normalized eye glasses region

(c) 58 x 26 normalized eye glasses region

(d) 38 x 17 normalized eye glasses region

(f) 16 x 7 normalized eye glasses region (minimum size)
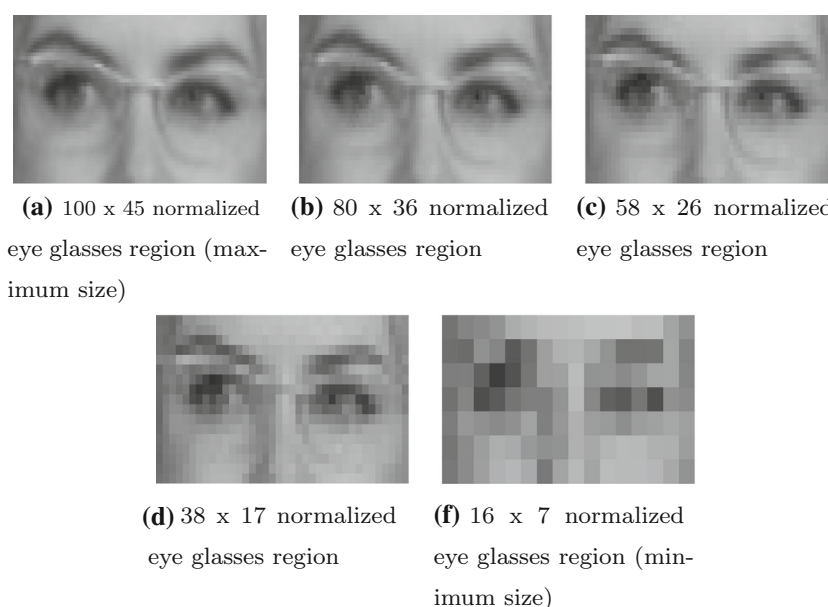
glasses. The width is determined taking into account the distance between the eyes and the sides of the glasses. In these experiments, the width of the normalized eye glasses region is 122 % bigger than the height of the normalized region (see Fig. 13 for some examples).

Since the average distance from left outer eye to right outer eye is 63.86 pixels, the maximum tested width for the normalized eye glasses region is 100 pixels and a height of 45 pixels. So, the size of the maximum normalized eye glasses region is 4500 pixels. The minimum tested width is 16 pixels and a height of seven pixels. So, the size of the minimum normalized eye glasses region is 112 pixels. See Fig. 7 for some examples. In Fig. 8, different tests are performed varying from the maximum to the minimum size of the normalized eye glasses region. It can be appreciated that the relationship between the size of normalized eye glasses region and the recognition rate follows a logarithmic distribution. This logarithmic distribution depends on the number of division of the LBP operator. Two examples can be seen in Fig. 9. $R^2 = \{0.9880, 0.9875, 0.9826, 0.9812, 0.9897, 0.9880, 0.9876, 0.9891, 0.9915, 0.9898\}$ correspond to the correlation coefficients obtained from the different LBP series from Fig. 8.

4.2 Number of divisions of the LBP operator

Since the LBP histogram is calculated over each rectangular region of the image, using a small number of divisions makes the feature vector shorter, but also means losing spatial information. In order to encode the spatial information, face image is usually divided into a grid of non-overlapping regions $\{R_0, \ldots, R_{m-1}\}$. So, the length of the feature vector becomes $B = 59 \times m$ using uniform patterns, in which m is the number of regions. The mean recognition rates for the
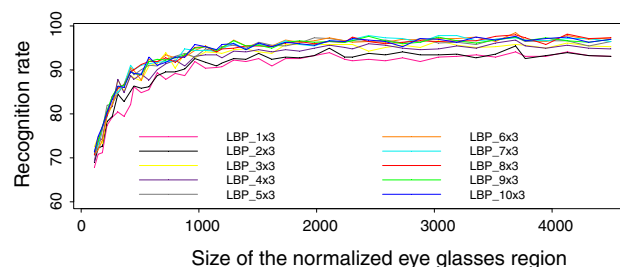

Fig. 8 Results varying from the maximum to the minimum size of the normalized eye glasses region
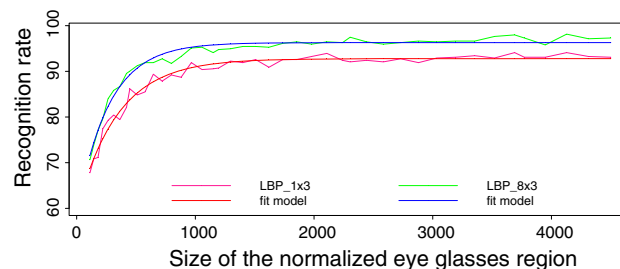

Fig. 9 Relationship between size of normalized eye glasses region and recognition rate follows a logarithmic distribution. Two examples are shown

$LBP_{8,2}^{u2}$ as a function of the number of regions is plotted in Figs. 10 and 11. When looking for the optimal number of regions, it is observed the changes in the number of regions may cause big difference in the length of the feature vector, but the performance is not necessarily affected significantly. However, it is noticed the divisions in the width of the image have a greater impact on the performance than the divisions in the height of the image. It makes sense due to the characteristics of the normalized eyes region.

	1	2	3	4	5	6	7	8	9	10
1	88.53	88.36	91.23	89.38	90.56	92.58	91.23	92.24	91.91	92.92
2	91.06	89.54	93.09	90.56	93.09	93.25	93.59	94.10	93.09	94.27
3	90.39	91.06	93.09	92.92	95.45	96.63	95.95	95.95	94.94	95.78
4	90.39	89.54	94.60	92.92	95.95	95.95	96.29	96.80	95.45	95.95
5	91.57	90.73	94.94	92.75	95.78	96.96	96.46	96.80	95.95	96.63
6	89.88	90.22	94.44	92.75	96.12	95.78	96.63	96.96	95.62	96.12
7	89.38	89.71	93.93	92.41	95.11	94.94	95.95	95.95	95.45	95.78
8	89.88	90.22	93.42	92.58	95.11	95.62	95.45	96.12	95.28	95.95
9	89.88	89.71	93.76	92.58	94.94	95.62	95.45	95.95	94.94	95.95
10	89.38	90.89	93.25	91.40	93.93	94.60	94.44	94.60	94.94	95.28

Fig. 10 Recognition rates as a function of the number of regions. No alignment of the images is carried out

	1	2	3	4	5	6	7	8	9	10
1	88.36	88.53	93.42	91.57	93.76	94.10	93.59	94.60	93.09	94.77
2	91.23	91.40	93.76	92.58	94.10	94.94	94.27	94.94	94.44	94.60
3	91.74	93.59	95.11	94.77	96.96	96.80	96.29	96.63	96.29	97.13
4	92.41	93.25	95.45	95.11	95.78	95.78	95.78	96.96	96.46	96.29
5	92.92	93.59	95.45	94.94	96.46	96.80	96.80	97.30	96.29	97.64
6	93.59	93.59	96.29	95.28	96.63	97.13	96.63	97.64	96.12	96.46
7	93.42	94.10	96.12	95.28	96.80	96.96	96.80	96.80	96.80	97.30
8	92.75	94.60	96.29	94.60	96.80	96.63	96.63	96.46	96.63	96.63
9	93.09	94.27	95.95	95.28	96.29	96.63	96.46	95.95	96.46	96.46
10	92.92	93.09	95.95	94.60	96.63	97.13	95.78	96.12	96.80	96.80

Fig. 11 Recognition rates as a function of the number of regions. Alignment of the images is carried out

4.3 Alignment of the images

A performance comparison between alignment of the faces and non alignment can be seen in Figs. 10 and 11. It can be seen that the recognition rate improves if the alignment step is performed, which is increased by an average of 1.55 %.

4.4 Varying α on the Robust Local Binary Pattern algorithm

A set of experiments is carried out to select the optimal parameter α (see Fig. 12). RLBP introduced a parameter α in order to make a balance of robustness and stability. It should be pointed out that RLBP performs more stably under complex illumination and viewpoint variations, since it extracts
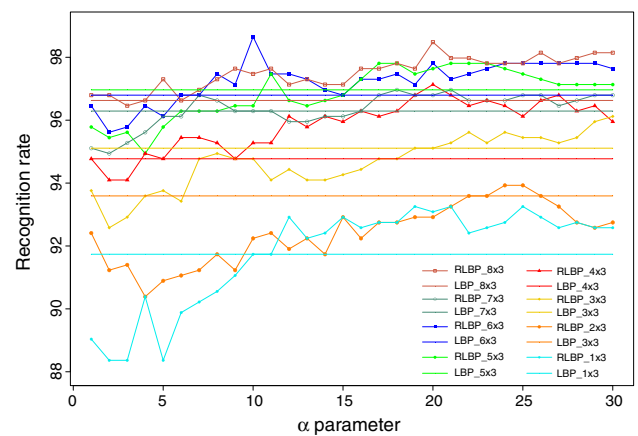

**Fig. 12** LBP and RLBP comparison varying α between [1–30]

the gray-level information of both local neighbor set and individual pixel. RLBP outperforms standard LBP.

5 Discussion

The proposed algorithm has been tested on real images over the Labeled Faces in the Wild database achieving a 98.65 % recognition rate (Fig. 13). Most algorithms are not tested with real-life photographs. In [5], the overall correctness is 94.0 % on real-life photographs. In [4], the system was tested on a surveillance video with real conditions and the best results are detecting eyeglasses with a precision of 90 % and a recall of 8.9 %.

As far as authors know, this is the first time this database is proposed for comparison of glasses detection. So, comparisons with our algorithm using this database are not possible. But in order to see that the proposed method advances the state of the art, the algorithm has also been tested on FERET database [39] because some of the previous methods were trained/tested with this database [2], [5]. A detection rate of 87.7 % was obtained by the algorithm proposed by Dantcheva et al. [2]. Wu et al. [5] achieved a remarkable accuracy of 98.9 % using wavelet features. It is worth mentioning that the algorithm proposed in [5] was trained/tested not only with FERET database images but also with other World Wide Web images. The results of our algorithm on FERET database achieved 99.89 and 99.83 % recognition rate with RBLP and LBP, respectively. Some examples can be seen in Figs. 14 and 15. In most images, the results are like the ones shown in Fig. 14, that is, the normalized eye glasses region are calculated on FERET face images based on the accurately position of the facial landmarks. In Fig. 15, even when some of the four landmarks of the eyes are not accurately detected, the normalized eye glasses region algorithm can operate robustly. This fact can be appreciated even better in Fig. 15g, h. The algorithm used for detecting the landmarks did not perform

Fig. 13 Normalized eye glasses region for all images from LFW database contained in this work
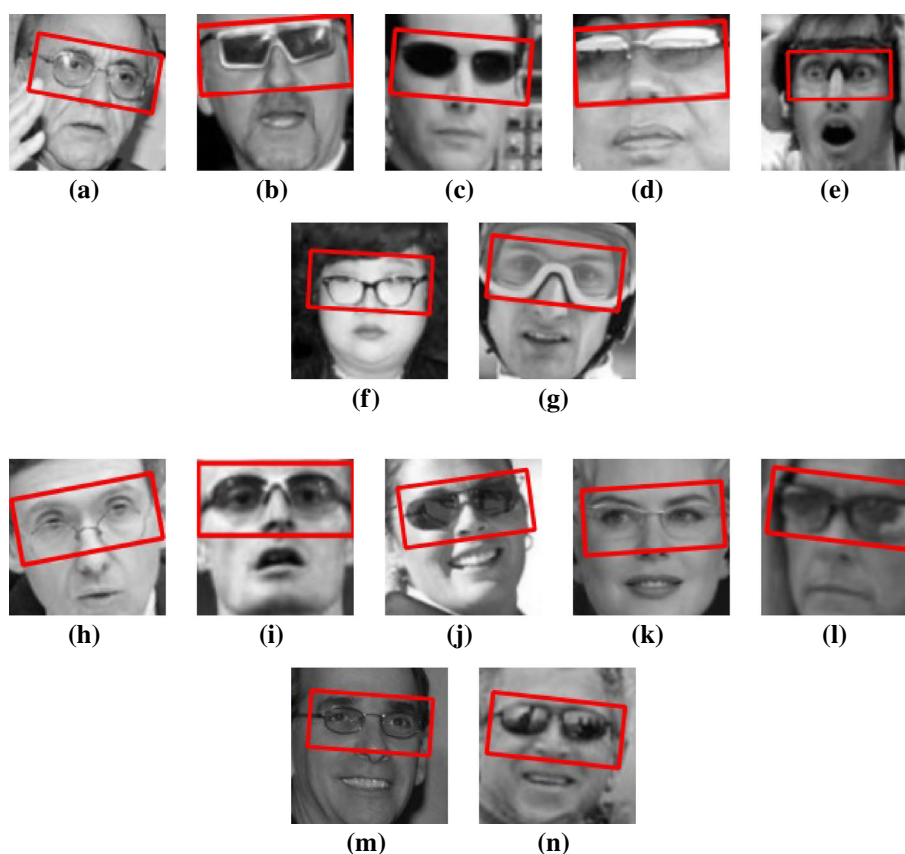
well probably because these images exhibit a profile view pose. Even so, due to the fact that the relationship among facial features could approximate the location of the eyes, the eye glasses normalization algorithm can operate quite well. The failure case from FERET database can be seen in Fig. 15i. So, in conclusion, the proposed glasses region normalization algorithm has the capabilities of detecting glasses region with different poses and under different conditions.

Furthermore, since a low-resolution face contains limited information [38], several tests are performed in order to know how the resolution of the face images affects recognition rate. A total of forty-three different resolutions of the normalized eye glasses region were studied in this research. The lower-resolution images were down-sampled form original images. It should also be noted that the relationship between the size of normalized eye glasses region and the recognition rate follows a logarithmic distribution. In other words, the LBP features perform robustly and stably over a useful range of low resolutions.

In case of alignment of the face images, an average recognition rate of 95.19 % is achieved (average value from Fig. 11). The best average performance is achieved with ten division in the width (average performance of 96.40 %) and seven division in the height (96.03 %). Additionally, if the choice is not to align the faces, an average recognition rate of 93.64 % is performed (average value from Fig. 10). The
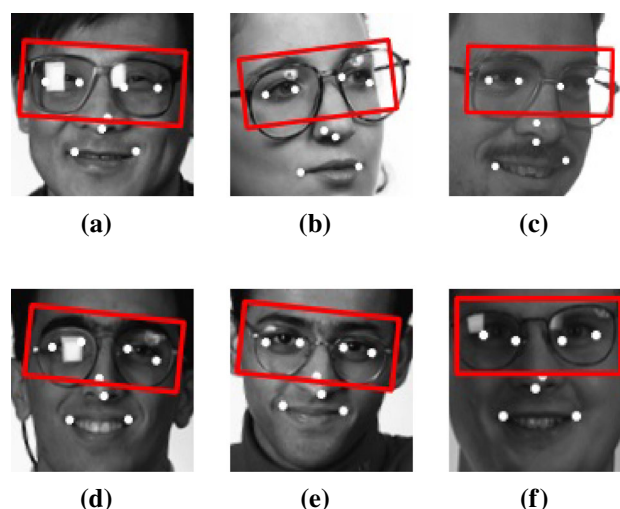

Fig. 14 The normalized eye glasses region is calculated on FERET face images based on the accurate position of the facial landmarks

best average performance is achieved with eight divisions in the width (average performance of 95.54 %) and five divisions in the height (average peformance of 94.85 %). It can therefore be concluded that a large number of divisions can decrease the performance due to the nonalignment of the face images.
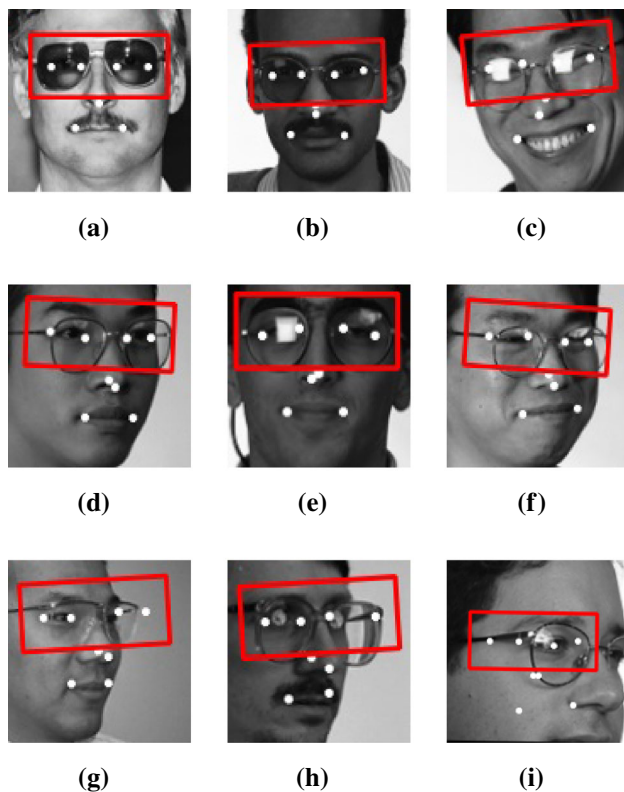
(a)   **(b)**   **(c)**

(d)   **(e)**   **(f)**

(g)   **(h)**   **(i)**

Fig. 15 The normalized eye glasses region algorithm can operate robustly under different image conditions from FERET database. The failure case is included (Fig. 15i)

Fig. 16 Relationship between size of the normalized eye glasses region and the execution time for both RLBP and LBP follows a liner distribution

6 Conclusions

Automatic glasses detection on real face images has many potential uses in security systems, intelligent advertising and marketing. In this study, glasses detection on real face images has been investigated. A normalized eye glasses region is proposed, and a comparison between the original LBP and RLBP is carried out varying α parameter. Influences of the resolution, the alignment of the normalized images and the number of divisions in the LBP operator are also carried out. Experimental results demonstrate that a simple yet efficient algorithm can obtain impressive classification accuracy achieving 98.65 % recognition rate on LFW database. This algorithm has been tested on FERET database too, achieving 99.89 % recognition rate. Experimental results also show that the proposed algorithm is robust under a wide range of lighting conditions, different poses and can deal with occlusion, that are very common with sunglasses, for example. Even when main features from the face are not correctly extracted, the algorithm can estimate glasses region quite robust.

Normalized eye glasses region described in this study can also be used to extract relevant information to deal with the problem of eye closeness detection. Eyes closeness detection from face images has wide applications like facial expression recognition, driver fatigue detection and so on. Closed eyes in the Wild (CEW) database has recently been released [41]. Future work includes the development of an architecture able to face with both problems simultaneously using LFW and CEW databases and new models and descriptors in order to get better image features to achieve better recognition performance without sacrificing simplicity. For example, Completed Hybrid Local Binary Pattern (CHLBP) has recently been proposed [42] to avoid tuning α parameter. A new comprehensive database to benchmark glasses detection and recognition will be proposed. To improve generalization, different databases will be combined and some checks will be performed [43] in order to build a realistic

A large number of LBP variations are designed to expand the scope of application, which offers better performance as well as improvement in the robustness of one or more aspects of the original LBP [40]. In this study, Robust Local Binary Pattern is applied and compared with Local Binary Pattern. RLBP outperforms LBP for almost all values of α in the range of (1–30). Small values of α did not increase the performance compared with original Local Binary Pattern. However, for bigger values, the performance is improved considerably. For instance, varying α from 10 to 30, the average performance can increase by 0,42 %.

Finally, it should also be noted that the relationship between size of the normalized eye glasses region and the execution time for both RLBP and LBP follows a linear distribution. $R^2 = \{0.9993, 0.9981\}$ values are obtained for RLBP and LBP, respectively, see Fig. 16. LBP is in average, 2.31 % times faster than RLBP. All steps and processes involved in the eye glasses region normalization operation (face detection, landmark detection, face rotation, eye glasses region cropping and resize to desired output) take an average of 30 ms. All experiments in this study are carried out on a 2500-MHz Intel running Windows.
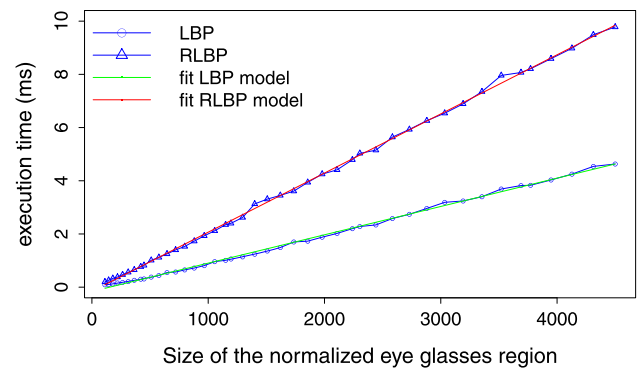
and, as far as possible, not biased database. Different types of glasses (sport glasses, sunglasses, safety glasses, reading glasses, etc) will be included in this database in order to test both detection and recognition.

References

1. Jiang, X., Binkert, M., Bernard, A., Bunke, H.: Towards detection of glasses in facial images. Pattern Anal. Appl. **3**(1), 9–18 (2000)
2. Dantcheva, A., Velardo, C., Dangelo, A., Dugelay, J.-L.: Bag of soft biometrics for person identification. Multimed. Tool. Appl. **51**(2), 739–777 (2011)
3. Jing, Z., Mariani, R.: Glasses detection and extraction by deformable contour. In: Proceedings 15th International Conference on Pattern Recognition, 2000, vol. 2, pp. 933–936. IEEE (2000)
4. Vaquero, D.A., Feris, R.S., Tran, D., Brown, L., Hampapur, A., Turk, M.: Attribute-based people search in surveillance environments. In: Workshop on Applications of Computer Vision (WACV), 2009, pp. 1–8. IEEE (2009)
5. Wu, B., Ai, H., Liu, R.: Glasses detection by boosting simple wavelet features. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004, vol. 1, pp. 292–295. IEEE (2004)
6. Wu, H., Yoshikawa, G., Shioyama, T., Lao, T., Kawade, T.: Glasses frame detection with 3d hough transform. In: Proceedings 16th International Conference on Pattern Recognition, 2002, vol. 2, pp. 346–349. IEEE (20020
7. Shan, S., Gao, W., Lu, Y., Cao, B., Chen, X., Zhao, D., Zeng, W.: Unified framework for classifying facial images based on facial attribute-specific subspaces and minimum reconstruction error. In: Proc. of ACCV, vol. 2. Citeseer (2002)
8. Chenyu, W., Liu, C., Shum, H.-Y., Xy, Y.-Q., Zhang, Z.: Automatic eyeglasses removal from face images. IEEE Trans. Pattern Anal. Mach. Intell. **26**(3), 322–336 (2004)
9. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001, CVPR 2001, vol. 1, pp. I-511. IEEE (2001)
10. Heo, J., Kong, S.G., Abidi, B.R., Abidi, M.A.: Fusion of visual and thermal signatures with eyeglass removal for robust face recognition. In: Conference on Computer Vision and Pattern Recognition Workshop, 2004, CVPRW'04, pp. 122–122. IEEE (2004)
11. Jing, Z., Mariani, R., Wang, J.: Glasses detection for face recognition using bayes rules. In: Advances in Multimodal InterfacesICMI 2000, pp. 127–134. Springer, Berlin (2000)
12. Schapire, R.E., Singer, Y.: Improved boosting algorithms using confidence-rated predictions. Mach. Learn. **37**(3), 297–336 (1999)
13. Xiao, Y., Yan, H.: Extraction of glasses in human face images. In: Biometric Authentication, pp. 214–220. Springer, Berlin (2004)
14. Park, J.-S., Oh, Y.H., Ahn, S.C., Lee, S.-W.: Glasses removal from facial image using recursive error compensation. IEEE Trans. Pattern Anal. Mach. Intell. **27**(5), 805–811 (2005)
15. Uricár, M., Franc, V., Hlavác, V.: Detector of facial landmarks learned by the structured output svm. In: VISAPP12: Proceedings of the 7th International Conference on Computer Vision Theory and Applications, vol. 1, pp. 547–556 (2012)
16. Wolf, L., Hassner, T., Taigman, Y.: Similarity scores based on background samples. In: Computer Vision-ACCV 2009, pp. 88–97. Springer, Berlin (2010)
17. Huang, G.B., Jain, V., Learned-Miller, E.: Unsupervised joint alignment of complex images. In: IEEE 11th International Conference on Computer Vision, 2007, ICCV 2007, pp. 1–8. IEEE (2007)
18. Huang, G.B., Mattar, M., Lee, H., Learned-Miller, E.G.: Learning to align from scratch. In: Neural Information Processing Systems. Neural Information Processing Systems (NIPS), Lake Tahoe (2012)
19. Makinen, E., Raisamo, R.: Evaluation of gender classification methods with automatically detected and aligned faces. IEEE Trans. Pattern Anal. Mach. Intell. **30**(3), 541–547 (2008)
20. Li, G., Cai, X., Li, X., Liu, Y.: An efficient face normalization algorithm based on eyes detection. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, 2006, pp. 3843–3848. IEEE (2006)
21. Mäkinen, E., Raisamo, R.: An experimental comparison of gender classification methods. Pattern Recogn. Lett. **29**(10), 1544–1556 (2008)
22. Wu, H., Yokoyama, T., Pramadihanto, D., Yachida, M.: Face and facial feature extraction from color image. In: Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, 1996, pp. 345–350. IEEE (1996)
23. Wu, H., Inada, J., Shioyama, T., Chen, Q., Simada, T.: Automatic facial feature points detection with susan operator. In: Proceedings of the Scandinavian Conference on Image Analysis, pp. 257–263 (2001)
24. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2010, pp. 94–101. IEEE (2010)
25. Chang, C.-C., Lin, C.-J.: LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. **2**, 27:1–27:27 (2011)
26. Song, F., Tan, X., Chen, S., Zhou, Z.-H.: A literature survey on robust and efficient eye localization in real-life scenarios. Pattern Recogn. **46**(12), 3157–3173 (2013)
27. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. In: Technical Report 07–49. University of Massachusetts, Amherst (2007)
28. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern Recogn. **29**(1), 51–59 (1996)
29. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
30. Guo, Z., Zhang, L., Zhang, D.: A completed modeling of local binary pattern operator for texture classification. IEEE Trans. Image Proc. **19**(6), 1657–1663 (2010)
31. Zhao, Y., Jia, W., Hu, R.-X., Min, H.: Completed robust local binary pattern for texture classification. Neurocomputing **106**, 68–76 (2013)
32. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
33. Shan, C.: Learning local binary patterns for gender classification on real-world face images. Pattern Recogn. Lett. **33**(4), 431–437 (2012)
34. Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and simile classifiers for face verification. In: IEEE 12th International Conference on Computer Vision, 2009, pp. 365–372. IEEE (2009)
35. Pietikäinen, M., Hadid, A., Zhao, G., Ahonen, T.: Local binary patterns for still images. In: Computer Vision Using Local Binary Patterns, pp. 13–47. Springer, Berlin (2011)

36. Huang, D., Shan, C., Ardabilian, M., Wang, Y., Chen, L.: Facial image analysis based on local binary patterns: a survey. IEEE Trans. Sys. Man Cyber.-Part C **41**(6), 765–781 (2011)

37. Ahonen, T., Hadid, A., Pietikäinen, M.: Face recognition with local binary patterns. In: Computer Vision-eccv 2004, pp. 469–481. Springer, Berlin (2004)

38. Zou, W.W., Yuen, P.C., Chellappa, R.: Low-resolution face tracker robust to illumination variations. IEEE Trans. Image Proc. **22**(5), 1726–1739 (2013)

39. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The feret evaluation methodology for face-recognition algorithms. IEEE Trans. Pattern Anal. Mach. Intell. **22**(10), 1090–1104 (2000)

40. Huang, D., Shan, C., Ardabilian, M., Wang, Y., Chen, L.: Local binary patterns and its application to facial image analysis: a survey. IEEE Trans. Syst. Man Cyber. Part C Appl. Rev. **41**(6), 765–781 (2011)

41. Song, F., Tan, X., Liu, X., Chen, S.: Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients. Pattern Recogn. **47**(9), 2825–2838 (2014)

42. Yuan, J.-H., Huang, D.-S., Zhu, H.-D., Gan, Y.: Completed hybrid local binary pattern for texture classification. In: International Joint Conference on Neural Networks (IJCNN), 2014, pp. 2050–2057. IEEE (2014)

43. Torralba, A., Efros, A.A.: Unbiased look at dataset bias. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, pp. 1521–1528. IEEE (2011)

Rubén Usamentiaga Associate Professor in the Department of Computer Science and Engineering at the University of Oviedo. He received his M.S. and Ph.D. degrees in Computer Science from the University of Oviedo in 1999 and 2005, respectively. In recent years he has been working on several projects related to computer vision and industrial systems. His research interests include real-time imaging systems and thermographic applications for industrial processes.

Rubén Casado received a B.Sc. degree in Computer Science in 2005, a M.Sc. in Computing in 2008 and a PhD in Software Systems in 2013 from University of Oviedo, Spain. He has worked as a researcher and teaching assistant at the University of Oviedo, where he is currently a member of the Software Engineering Research Group. Currently he is the leader of the Big Data research program at Treelogic.

Alberto Fernández received his M.S. degree in Computer Science from University of Oviedo in 2009. Since 2007 he has been developing his professional career at the CTIC Foundation (Centre for the Development of Information and Communication Technologies in Asturias). He has been working in R&D projects. His research interests include artificial vision systems for both industrial and medical processes.

Rodrigo García received the MS degree in computer science from the University of Oviedo, Spain, in 2008. He worked for the R&D Department of CTIC Foundation until October 2014, when he founded his own company, UbiqWare. He also collaborates with the computer engineering area at the University of Oviedo where he is pursuing his PhD in the topic of QoS in distributed systems.