

RESEARCH

Open Access



A study on different experimental configurations for age, race, and gender estimation problems

Pierluigi Carcagni^{*}, Marco Del Coco, Dario Cazzato, Marco Leo and Cosimo Distante

Abstract

This paper presents a detailed study about different algorithmic configurations for estimating soft biometric traits. In particular, a recently introduced common framework is the starting point of the study: it includes an initial facial detection, the subsequent facial traits description, the data reduction step, and the final classification step. The algorithmic configurations are featured by different descriptors and different strategies to build the training dataset and to scale the data in input to the classifier. Experimental proofs have been carried out on both publicly available datasets and image sequences specifically acquired in order to evaluate the performance even under real-world conditions, i.e., in the presence of scaling and rotation.

Keywords: LBP, CLBP, HOG, SWLD, Digital signage, Audience measurement

1 Introduction

Biometrics have attracted much interest in the last decade, for the high number of applications they can have in industrial and academic fields [1].

Recently, soft biometrics have been clearly defined in [2] as a set of traits providing information about an individual, even though the lack of distinctiveness and permanence does not allow to authenticate individuals. These traits can be either continuous (e.g., height and weight) or discrete (e.g., gender, eye color, ethnicity, etc.).

Soft biometrics have widespread use for security, digital signage, domotics (the study of the realization of an intelligent home environment), home rehabilitation, artificial intelligence, and so on. Even socially assistive technologies are a new and emerging field where these solutions could considerably improve the overall human-machine interaction level (for example with autistic individuals [3], as well as with people with dementia [4] and, generally, for elderly care [5]).

In the last few decades, computer vision, as well as other information science fields, have largely investigated the problem of the automatic estimation of the main soft

biometric traits by means of mathematical models and ad hoc coding of the visual images. In particular, the automatic estimation of gender, race, and age from facial images are among the most investigated issues, but there is still a lot of open challenges especially for race and age. The extraction of this kind of information is not trivial due to the ambiguity related to the anatomy of each individual and his lifestyle. In particular, in race recognition, the somatic traits of some population could be not well defined: for example, one person may exhibit some features more than another one. Similar considerations apply to age estimation, where the appearance of biological age could be very different from the chronological one.

The recent efforts led to a shared and effective framework that includes an initial facial detection, a subsequent description of facial traits, a data reduction step (optional), and a final classification step. The effectiveness of this framework has been already experimentally proved in [6] where some key aspects (like the data reduction strategy) were definitively clarified. Nevertheless, in the literature, there is no comprehensive study that investigates which methods and techniques are best suited to address the involved algorithmic steps, i.e., to give evidence of the best configuration of the aforementioned framework in terms of facial descriptors, learning strategies, and numerical scale of the data in input to

^{*}Correspondence: pierluigi.carcagni@cnr.it
Institute of Applied Sciences and Intelligent Systems, National Research Council of Italy c/o Dhitech scarl, Via Monteroni, 73100 Lecce, Italy

the classifier. Moreover, the performance of the framework under real world conditions (even in the presence of scaling and rotation of the head) has not been fully addressed. To fill this gap, in this paper, a detailed study about different algorithmic configurations for estimating soft biometric traits is presented. First, some of the most used facial descriptors for gender, race, or age estimation were selected, and then, their performances, by using each of them into the considered framework, were compared. This experimental phase was carried out on publicly available face datasets, considering different working strategies (i.e., the balancing/unbalancing of the training dataset and the numerical range of the input measurements). Then, the best configuration was thoroughly tested, for all the involved descriptors, on ad hoc image sequences acquired in a real environment and also considering changes of the head in both position and scale.

The considered feature descriptors were those that, from a detailed study of the literature (reported in next session), demonstrated to be the best suited for encoding the biometric traits both in terms of accuracy and computational efficiency. In particular, this paper investigates local binary pattern (LBP) [7, 8], compound local binary pattern (CLBP) [9] (both oriented to texture analysis), histogram of oriented gradient (HOG) [10] (designed for shape analysis), and one descriptor that represents a trade-off between the two aspects: Weber local descriptor (WLD) [11]. They were evaluated in a consolidated framework where data reduction is achieved by Fisher linear discriminant analysis (LDA) [12] and then support vector machines (SVMs) are used both for classification (race and gender) and regression (for age estimation). SVM [13, 14] is one of the most exploited method in discrete biometric estimation problems whose superiority has been well proved in preliminary studies about the performance of different classifiers [15, 16].

Briefly, the main contributions of this paper are as follows:

1. A survey of existing methods and techniques addressing soft biometric issues, in order to discover the most affordable framework;
2. An exploration of various configurations about the components of the framework, in terms of learning strategies and numerical scale of data in input to the classifier, in order to discover the best way to exploit it in soft biometric contexts;
3. The evaluation of the performances of the aforementioned framework when using the different descriptors also in real-world conditions (even in the presence of scaling and head rotation).

The remainder of the paper is organized as follows: Section 2 discusses the existing literature for the

considered biometric issues; in Section 3, the framework for age, gender, and race classification is explained in detail while in Section 4, the experimental setup and results are detailed and discussed. Finally, conclusions and future developments are discussed in Section 5.

2 Related works

The study of algorithms investigating soft biometric traits has attracted considerable interest in the last two decades. In the following subsections, the most relevant solutions for each specific addressed problem (such as gender, race, and age) are reported.

2.1 Gender

Automatic gender recognition is the most debated soft biometric issue and, over the years, many approaches have been proposed in literature. Three exhaustive surveys can be found in [17, 18] and [19]. Early studies mainly exploited geometrical-based methods; Brunelli and Poggio [20] computed 18 distances between face key points and then they used them to train a hyper basis function network classifier, whereas 22 normalized vertical and horizontal fiducial distances starting from 40 manually extracted facial points were used by Fellous in [21]. This approach has been recently reassessed by Mozzafari et al. [22] who used the aspect ratio of face ellipse as geometric feature to reinforce appearance information.

In the early 2000s, appearance-based methods were introduced in their basic form, i.e., by evaluating the pixel intensity values. With these approaches, the amount of useless data is a crucial issue that Abdi et al. [23] managed by a principal components analysis (PCA), whereas Tamura et al. [24] investigated gender estimation capabilities from pixel intensity on small-sized facial images. Conversely, the most interesting progress was made with the introduction of the LBP descriptor [7] to encode information about facial appearance. LBP, in combination with SVM, was used by Lian and Lu [25] for multi-view gender classification. It was also exploited by Yang and Ai [26] to classify age, gender, and ethnicity. LBP descriptor was employed in [27] in combination with intensity and shape features to get a multi-scale fusion approach, while Ylioinas et al. [28] combined it with contrast information in order to achieve a more robust classification. To extract the most discriminative LBP features, Shan [29] proposed an AdaBoost selection method and many other variants have also been proposed to get more informative features, e.g., local Gabor binary mapping pattern [30, 31] and local directional pattern [32].

Besides LBP descriptor, other appearance-based solutions have also been investigated. In fact, the use of HOG was proposed in [33] and [34], whereas the spatial Weber local descriptor (SWLD) was proposed in [35] and active

shape models (ASM) were exploited by Gao and Ai in [36]; Saatci and Town [37] used some features extracted by trained active appearance models (AAM) [38] and supplied them to SVM classifiers arranged into a cascade structure in order to optimize overall recognition performance.

Scale invariant feature transform (SIFT) presents some advantages in terms of invariance to image scaling, translation, and rotation; accordingly, the use of SIFT does not require preprocessing stage, like accurate face alignment. Demirkus et al. [39] exploited these characteristics using a Markovian classification model, and Wang et al. [40] extracted SIFT descriptors at regular image grid points and combined them with global shape contexts of the face, adopting AdaBoost for classification.

The desire to imitate nature recently led researches to adopt Gabor filter to copy the cortex view system. In [41], 2-D Gabor wavelets at different scales and eight orientations were extracted, selected by AdaBoost and classified by a Fuzzy SVM that indicated the degree of one's person face belonged to female/male faces. Scalzo et al. [42] extracted a large set of features using Gabor and Laplace filters which were fused by genetic algorithms. Gabor filters, also inspired the modeling of biological inspired features (BIF) employed by Meyers and Wolf [43] for face processing. Finally, Guo et al. in [34] demonstrated that the performance of the gender recognition is affected significantly by human age.

2.2 Race

In the last few years, the race estimation problem has been addressed through several approaches. In [44], the skin color, the forehead area, and the lip color were used in order to estimate ethnicity, whereas in [45] kernel class-dependent feature analysis was used in combination with skin color features to tackle the problem of ethnicity estimation on large scale face databases. PCA was introduced in [46] to perform two-class (Myanmar vs. non-Myanmar) race classification and in [47] to classify, using a feed forward neural network, three (Arabic, Asian, and Caucasian) races. The problem of classifying people into Asian and non-Asian was addressed in [48]. The authors used PCA for features generation and independent component analysis for features extraction. Final classification was performed by means of a combination of different SVM classifiers. In [49], Gabor features were extracted from facial parts taking into account displacement issues, and SVM was then employed as classifier. A method, referred to as local gradient Gabor pattern, was proposed in [50]. It is based on the use of gradient information on Gabor-transformed images, demonstrating its effectiveness with images having variations in expression and illumination. In [51], faces were represented using elastic graphs labelled with 2-D Gabor wavelet features,

and the system was trained from examples to classify faces on the basis of sex, race, and expression using linear discriminant analysis. In [52], the problem of large-scale ethnicity estimation under variations of gender and age was addressed. Biologically inspired features, with manifold learning, were used to represent faces for ethnicity estimation demonstrating their robustness and high ethnicity classification accuracies within same-gender and same-age groups. An interesting analysis of age estimation performance under variations across race and gender was proposed in [53–55], whereas in [56], a joint estimation framework capable to deal with the mutual influence of gender, race, and age was introduced.

In [57], a fusion scheme for ethnicity estimation that combines Haar wavelet-based global features and uniform LBP-based local features was presented. The dimensionality of the obtained vector of features was reduced by means of PCA, and a K-nearest neighbor classifier was employed as final step. LBP histogram (LBPH) features were extracted in [26] to estimate Asian and non-Asian ethnicity by means of AdaBoost classifiers. LBP was also used in [58] and then fused with Weber local descriptors through concatenation to produce a more powerful set of features to be supplied to a minimum distance classifier for the final race estimation. WLD histograms were employed in [59] demonstrating that they outperform PCA-based holistic methods. Race estimation was also addressed in [60] where a LDA representation was developed under the Bayesian statistical decision framework: the task was formulated as a two-category classification problem to classify between Asian or non-Asian subjects. Finally, in [61], two of the above problems (the estimation of gender and race) were faced using facial imagery. In particular, the examination of the 3-D meshes of the face, without any associated texture or photographic information, was performed.

2.3 Age

Age classification is the most complex issue among those treated in this work. Age estimation is, in fact, an intrinsically continuous soft biometric that requires regression techniques to be accomplished. Nevertheless, it is often treated as a discrete problem by means of an age grouping allowing the use of classical classification methods. Some comprehensive surveys about automatic age estimation can be found in [62] and [63], where complete overviews are given via a description of the problem from a generic point of view and successively with well-structured presentations of the most significant works. The earliest works of age estimation were based on the cranio-facial development theory [64]. A mathematical model describing the shape of the human cranio was the key of this theory, firstly exploited by Kwon and da Vitoria Lobo [65]. Successively, Farkas [66] exploited

the face anthropometry, the science of measuring human face size and proportions: 57 facial key points were used to measure distances of interest and a model for age evaluation was provided. Unfortunately, these approaches were proved to be unreliable for the evaluation of age in adult people but sufficiently discriminative just in the growing age.

A more interesting approach was proposed by Lanitis et al. [67] who introduced an aging function (valid for younger ages) related to the features vector of raw AAM parameters [38]. In a similar way, Liu et al. [68] proposed, in recent years, the use of AAM in association with PCA and an SVM classifier experiencing interesting results.

An innovative prospective was instead given by Geng et al. [69] who used an EM-like iterative learning algorithm to model the incomplete aging pattern extracted by observing the same individual at different ages. A similar approach, but oriented to find a common pattern among different individuals, was proposed in [70] and [71].

Another major trend, in the aging-related facial features extraction, focuses on the use of an appearance model that can be handled as a global or local features extraction issue. The first attempts in this research line were in [72, 73] where texture and shape features were jointly exploited to increase the descriptor robustness in order to estimate the human ages through a multiple-group classification scheme with 5-year intervals taking also advantage from the gender knowledge (since the aging patterns are different for males and females). Successively, the interest in appearance-based descriptors arose exponentially and LBP was used for appearance features extraction in an automatic age estimation system proposed in [74], whereas some variants were proposed and tested in [75, 76]. Gabor features were also tried for age estimation purposes [77] demonstrating their discriminative power. BIF [78, 79] in aging problems were also explored by Guo et al. in [80] and Lian et al. in [81].

Recently, many other approaches were investigated. Lou et al. in [82] tested the use of HAAR features and SVM for a human-robot interaction application based on the age estimation. A learning-based encoding was proposed by Alnajjar et al. in [83] showing that an image encoding based on the learning feedback can improve the overall estimation accuracy. One of the most recent works exploited the information derived from the dynamic of facial landmarks in video sequences [84].

An approach to gender and age classification problems from video sequences, which encodes and exploits the correlation between the face images through manifold learning, was proposed in [85]. Finally, in [86], a study of the influence of demographic appearance in face recognition was proposed: gender (male and female),

race/ethnicity (Black, White, and Hispanic), and age group (18–30, 30–50, and 50–70 years old) were taken into account comparing six different solutions.

3 System overview

In this section, the common framework used to address different soft biometric issues (gender, age, and ethnicity) is introduced: it is able to automatically extract soft biometrics information from generic video sequences through a complex pipeline of operational blocks. The pipeline is schematically depicted in Fig. 1: the first step is devoted to the detection and normalization of faces in the running image. The outcomes of the first step are the input of the following step, aimed at extracting a new representation of facial regions, depending on the selected descriptor. In particular LBP, HOG, SWLD, and CLBP are considered in this paper.

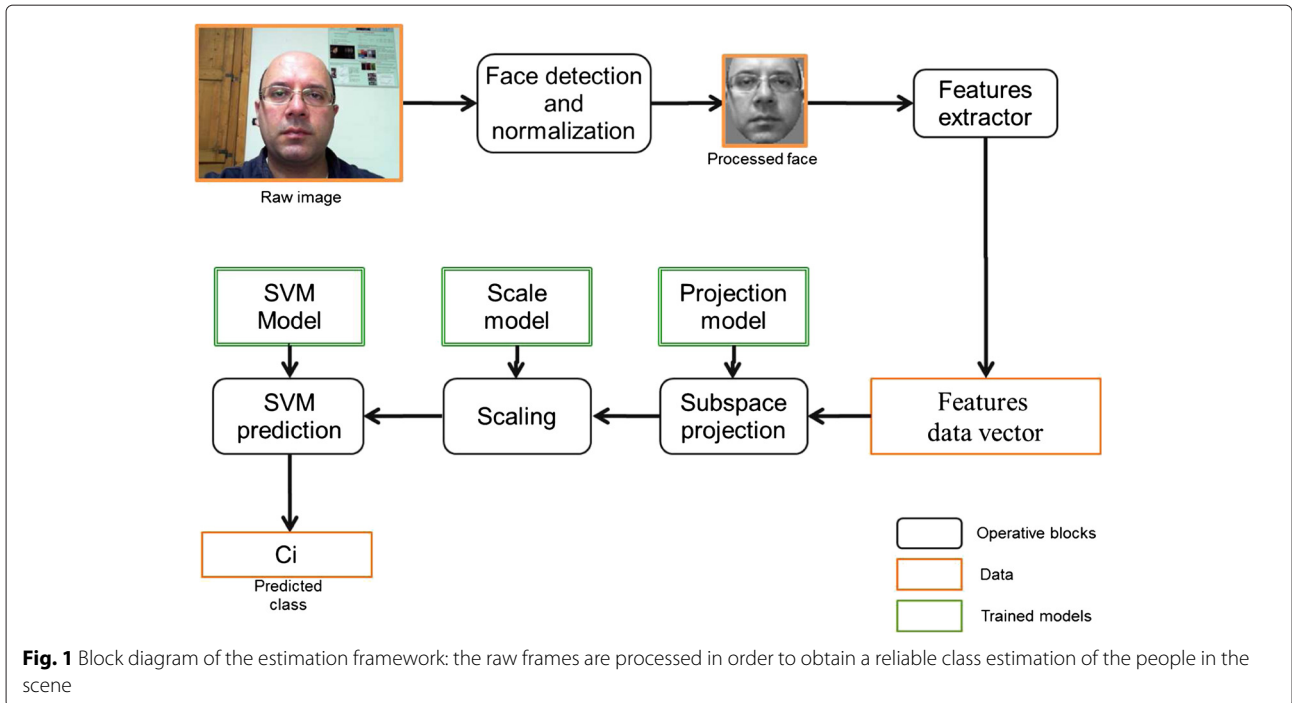
The set of extracted features is then projected (through LDA) on a more suited subspace where it is possible to include most of the initial discriminative information in a reduced number of variables. Finally, the retained variables are given as input to the learned SVM classifier that assures efficiency both in terms of accuracy and computational load.

The following subsections will detail each step giving information, also, about the family of comparing descriptors.

3.1 Face detection and normalization

In this step, human faces are detected in the input images and then a normalization operation is done. Normalization is a fundamental preprocessing step since the subsequent algorithms work better if they can evaluate input faces with predefined size and pose. The face detection is performed by means of the well-known Viola-Jones [87] approach (which combines increasingly more complex classifiers in a cascade) based detectors for face and head-shoulder (which helps to handle low resolution images) detection. Whenever a face is detected, the face normalization is carried out as follows: the system, at first, fits an ellipse to the face blob (exploiting facial features color models) in order to rotate it to a vertical position and hence a Viola-Jones-based eye detector searches the eyes. Finally, eye positions, if detected, provide a measure to crop and scale the frontal face candidate to a standard size of 65×59 pixels. The above face registration procedure is schematized in Fig. 2.

The registered face is then modeled using different features (average color using red-green normalized color space and considering just the center of the estimated face container; eyes patterns; whole face pattern) in order to re-detect it, for tracking purposes, in the subsequent frames [88]. Finally, it is given as input to the successive step for the features extraction.



3.2 Features extraction

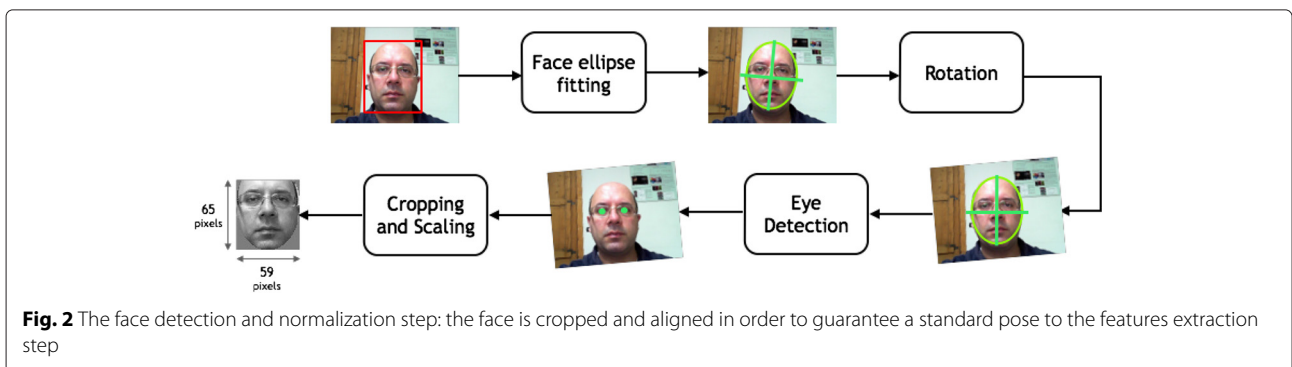
After the face detection and normalization step, a suitable family of descriptors has to be chosen in order to build a vector of features able to encode distinctive local structures in the facial region. The selected descriptors should be highly distinctive, i.e., with low probability of mismatch but, at the same time, their computational load and sensibility to the noise, such as changes in illumination, scaling, rotation, skew, etc., have to be kept low. The choice of the best family of descriptors is not straightforward: many different approaches have been introduced and effectively used in different academic fields, but the best strategy in the context of soft biometric estimations does not emerge from the literature. For this reason, in this paper, four different data descriptors have been considered and experimentally compared. For

each of them, a short description is given in the following subsections.

3.2.1 LBP

LBP is a local feature widely used for texture description in pattern recognition. The original LBP assigns a label to each pixel. Each pixel is used as a threshold and compared with its neighbor. If the neighbor is higher, it takes value 1; otherwise it takes value 0. Finally, the thresholded neighbor pixel values are concatenated and considered as a binary number that becomes the label for the central pixel. A graphical representation is shown in Fig. 3a and defined as follows:

$$LBP_{P,R}(x_c) = \sum_{p=0}^{P-1} u(x_p - x_c) 2^p \tag{1}$$



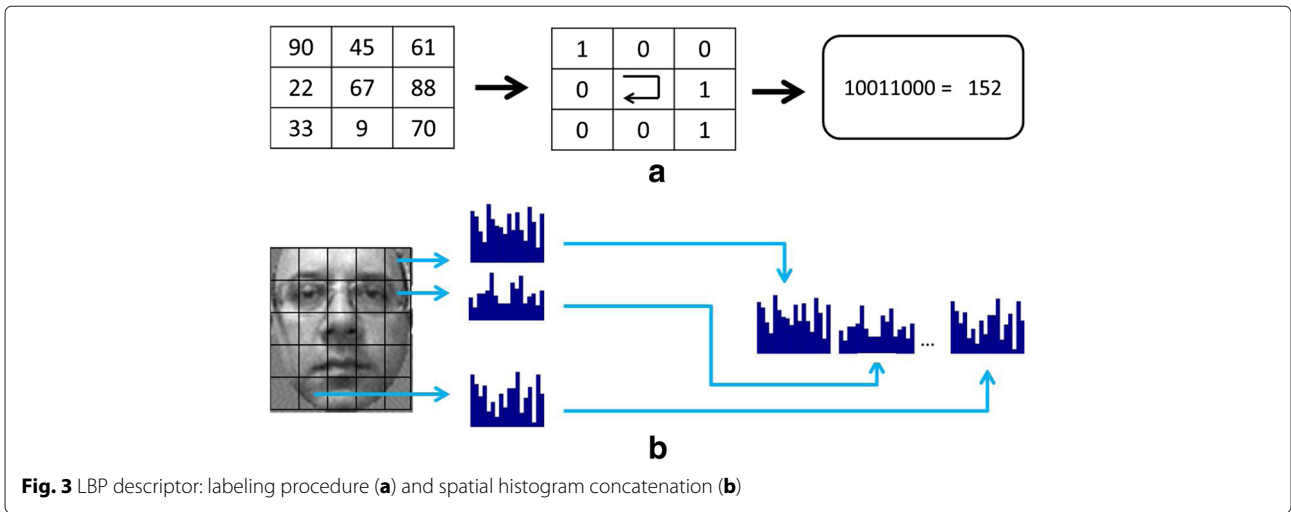


Fig. 3 LBP descriptor: labeling procedure (a) and spatial histogram concatenation (b)

where $u(y)$ is the step function, x_p the neighbor pixels, x_c the central pixel, P the number of neighbors, and R the radius. To account for the spatial information, the image is then divided into sub-regions. LBP is applied to each sub-region, and an histogram of L bins is generated from the pixel labels. Then, the histograms of different regions are concatenated in a single higher dimensional histogram, as represented in Fig. 3b.

3.2.2 CLBP

CLBP is an extension of the LBP descriptor designed essentially for facial expression recognition [9]. Compared to LBP, CLBP encodes the sign of the differences between the center pixel and its P neighbors as well as the magnitude of the differences. This involves a code length of $2P$ -bit: P for the signs and P for the magnitudes (Fig. 4). That is,

$$CLBP_{P,R}(x_c) = \sum_{p=0}^{P-1} u(x_p, x_c) 2^p \tag{2}$$

$$u(x_p, x_c) = \begin{cases} 00 & x_p - x_c < 0, |x_p - x_c| \leq M_{avg} \\ 01 & x_p - x_c < 0, |x_p - x_c| > M_{avg} \\ 10 & x_p - x_c \geq 0, |x_p - x_c| \leq M_{avg} \\ 11 & \text{otherwise} \end{cases} \tag{3}$$

where M_{avg} is the average magnitude of the difference between the center and the neighbors values.

As an example, considering a neighborhood 3×3 , the CLBP encodes a vector of 16 bits long provided by 8 bits for the magnitude and 8 bits for the sign patterns. In particular, the two patterns are built by concatenating the corresponding values of the bit sequence (1, 2, 5, 6, ..., $2P-3$, $2P-2$) and (3, 4, 7, 8, ..., $2P-1$, $2P$) of the original CLBP code. In this way, two images are obtained from the original one, by replacing each pixel with the two sub-CLBP patterns. A feature representation of the original image is then obtained by concatenating the histograms of the two images. In order to account for spatial information, the image is firstly divided into sub-regions and then CLBP is applied to each of them, similarly to the spatial LBP operator.

3.2.3 HOG

Local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions. HOG is a well-known feature descriptor based on the accumulation of gradient directions over the pixel of a small spatial region referred to as a “cell” and in the

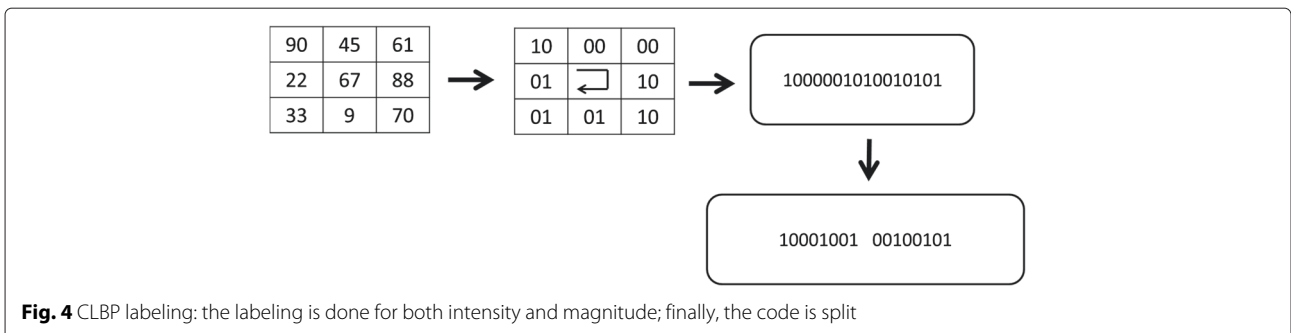


Fig. 4 CLBP labeling: the labeling is done for both intensity and magnitude; finally, the code is split

consequent construction of a 1-D histogram. Although HOG has many precursors, it has been used in its mature form in scale invariant features transformation [89] and widely analyzed in human detection by Dalal and Triggs [10]. This method is based on evaluating well-normalized local histograms of image gradient orientations in a dense grid. Let L be the image to analyze. The image is divided into cells (Fig. 5a) of size $N \times N$ pixels, and the orientation θ of each pixel $x = (x_x, x_y)$ is computed (Fig. 5b) by means of the following relationship:

$$\theta(x) = \tan^{-1} \frac{L(x_x, x_y + 1) - L(x_x, x_y - 1)}{L(x_x + 1, x_y) - L(x_x - 1, x_y)} \quad (4)$$

The orientations are accumulated in an histogram of a predetermined number of bins (Fig. 5c, d). Finally, histograms of each cell are concatenated in a single spatial HOG histogram (Fig. 5e). In order to achieve a better invariance to noise, it is also useful to contrast-normalize the local responses before using them. This can be done by accumulating a measurement of local histogram energy over larger spatial regions, named blocks, and using the results to normalize all of the cells in the block. The normalized descriptor blocks will represent the HOG descriptor.

3.2.4 Spatial WLD

WLD [11] is a robust and powerful descriptor inspired to Weber's law. It is based on the fact that the human perception of a pattern depends not only on the amount of change in intensity of the stimulus but also on the original stimulus intensity. The proposed descriptor consists of two components: differential excitation (DE) and gradient orientation (OR).

Differential excitation allows to detect the local salient pattern by means of a ratio between the relative intensity difference of the current pixel against its neighbors. The computational approach is similar to that used for LBP (a complete description is given in [11]). Moreover, the OR of the single pixel is considered.

When DE and OR are computed for each pixel in the image, we construct a 2-D histogram of T columns and $M \times S$ rows where T is the number of orientations and $M \times S$ the number of bins for the DE quantization with the meaning of [11]. Figure 6 shows how 2-D histogram is mixed in such a way to obtain a 1-D histogram that is more suitable for successive operations.

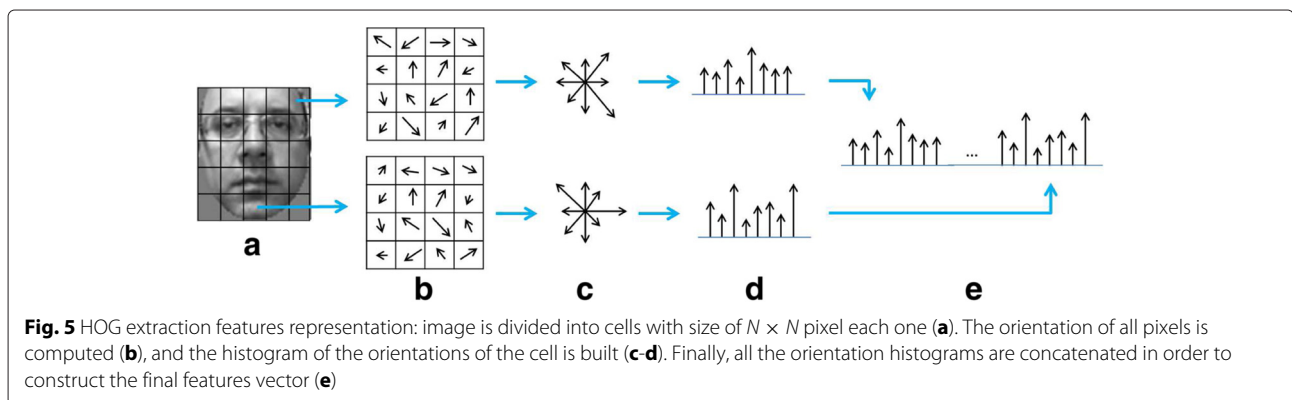
Furthermore, since for our purpose we had to account also for spatial information, we chose to use the SWLD approach [35] that, as for the spatial LBP, splits the image into sub-regions and computes an histogram for each of them. Finally, the histograms are concatenated in an ordered fashion.

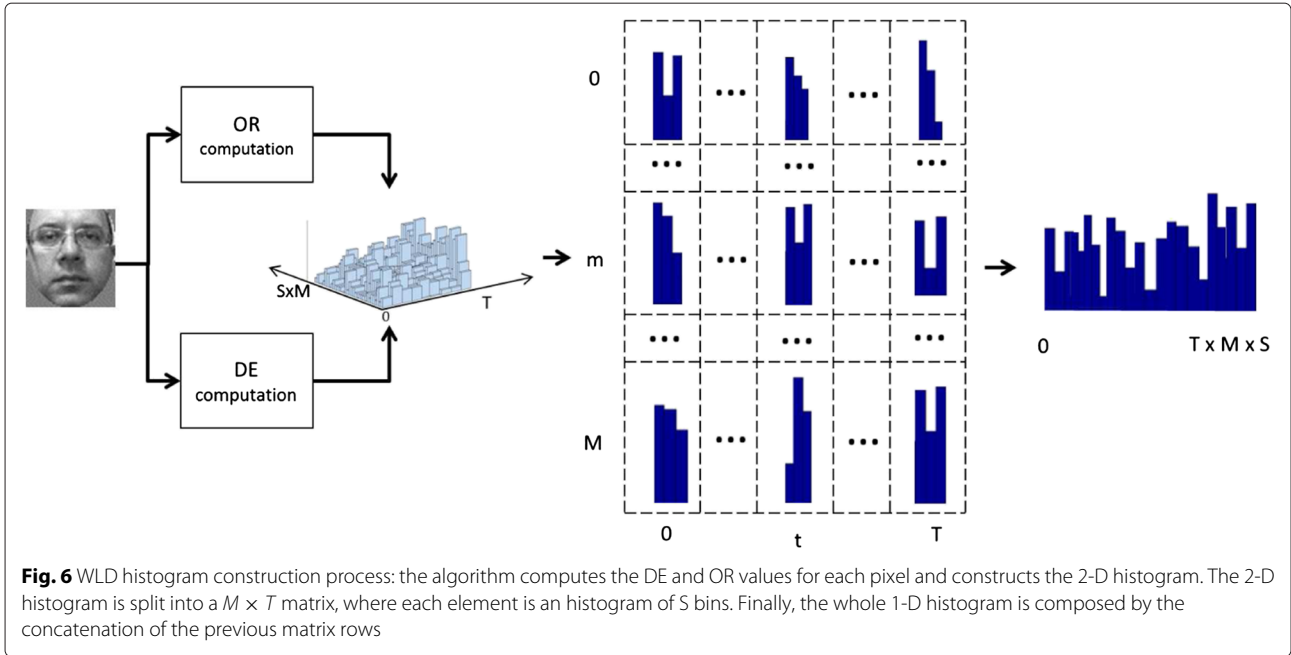
3.3 Features subspace projection

The number of used features for face description highly affects both the computational complexity and the accuracy of classification. Indeed, a reduced number of features allows SVM to use easier functions and to perform a better separation of clusters. In this work, LDA was considered to get a better data representation: this choice resulted from the high suitability experienced in traditional classification problems and in the preliminary experiments reported in [6].

3.3.1 LDA

LDA [12] is an alternative subprojection approach. In LDA, within-class and between-class scatters are used to formulate criteria for class separability. The optimizing criterion in LDA is the ratio of between-class scatter to the within-class scatter. The solution obtained by maximizing this criterion defines the axes of the transformed space. Moreover, in LDA analysis, the number of non-zero generalized eigenvalues, and so the upper-bound in eigenvectors numbers, is $c - 1$ where c represents the number of classes. Belhumeur et al. [12] demonstrated the robustness of this approach for face recognition and showed how LDA can sometimes be more discriminative. It is important to remark that the upper bound of eigenvectors for





LDA projection approach is a key factor also in terms of computational complexity that makes the workload of SVM lighter.

3.4 SVM classification

After data projection, in the proposed approach, gender, race, and age classifications were performed by using SVMs. In particular, three different SVM learning tasks were used: two-class support vector classification (SVC) for the gender problem; multi-class SVC for the race problem; and support vector regression (SVR) for the age estimation. All of these classification tasks can be biased by the data range (the j -th features of the data vector could take value in a different range with respect to the i -th one), and then, the opportunity of a scaling operation has to be analyzed in terms of classification performance. To analyze the scaling effect for soft biometric estimations, the following scaling approach was evaluated. Let $x_{i,MAX}$ and $x_{i,MIN}$ be the maximum and minimum values for the i -th features in the training data (a list of labeled features data vector) and $[y_{i,MAX}, y_{i,MIN}]$ the desired range after the scaling and N_f the total number of features. The equation for the scale becomes

$$y_i = \frac{(x_i - x_{i,MIN}) \cdot (y_{i,MAX} - y_{i,MIN})}{x_{i,MAX} - x_{i,MIN}} + y_{i,MIN} \quad (5)$$

where x_i is the initial value of the i -th feature and y_i the scaled one. Finally, the scaling model is constructed by the N_f pairs $\{x_{i,MAX}, x_{i,MIN}\}$ and the pair $\{y_{i,MAX}, y_{i,MIN}\}$.

With regard to SVC, the C -support vector classification (C -SVC) learning task implemented in the

well-known LIBSVM [90] was used. Given training vectors $x_i \in \mathbb{R}^n, i = 1, \dots, l$, in two classes, and a label vector $y \in \mathbb{R}^l$ such that $y_i \in \{1, -1\}$, C -SVC [13, 14] solves the following primal optimization problem:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^l \xi_i \\ \text{subject to} \quad & y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, i = 1, \dots, l, \end{aligned}$$

where $\phi(x_i)$ maps x_i into a higher-dimensional space and $C > 0$ is the regularization parameter. Due to the possible high dimensionality of the vector variable w , usually, the following dual problem is solved:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \\ \text{subject to} \quad & y^T \alpha = 0, \\ & 0 \leq \alpha_i \leq C, i = 1, \dots, l, \end{aligned}$$

where $e = [1, \dots, 1]^T$ is the vector of all ones, Q is an $l \times l$ positive semidefinite matrix, $Q_{ij} \equiv y_i y_j K(x_i, x_j)$, and $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$ is the kernel function. After the dual problem is solved, the next step is to compute the optimal w as follows:

$$w = \sum_{i=1}^l y_i \alpha_i \phi(x_i) \quad (6)$$

Finally, the decision function is

$$\text{sgn}(w^T \phi(x) + b) = \text{sgn} \left(\sum_{i=1}^l y_i \alpha_i K(x_i, x) + b \right) \quad (7)$$

Such an approach is suitable only for the two-class gender problems. For race classification, a multi-class approach was used, resorting to a “one-against-one” procedure [91]. Let k be the number of classes, then $k(k-1)/2$ classifiers are constructed where each one trains data from two classes. The final estimation is returned by a voting system among all the classifiers. Many other methods are available for multi-class SVM classification; in [92], a detailed comparison is given with the conclusion that “one-against-one” is a competitive approach.

With regard to SVR, instead, the ϵ -SVR learning task was used. Given training points $\{(\mathbf{x}_i, z_i), \dots, (\mathbf{x}_l, z_l)\}$ where $\mathbf{x}_i \in \mathbb{R}^n$ is a features vector and $z_i \in R$ is the target output, ϵ -SVR [13, 14] solves the following primal optimization problem:

$$\begin{aligned} \min_{\mathbf{w}, b, \xi, \xi^*} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^l \xi_i + C \sum_{i=1}^l \xi_i^*, \\ \text{subject to} \quad & \mathbf{w}^T \phi(\mathbf{x}_i) + b - z_i \leq \epsilon + \xi_i, \\ & z_i - \mathbf{w}^T \phi(\mathbf{x}_i) - b \leq \epsilon + \xi_i^*, \\ & \xi_i, \xi_i^* \geq 0, i = 1, \dots, l, \end{aligned}$$

with $C > 0$ and $\epsilon > 0$ as parameters. In this case, the dual problem is as follows:

$$\begin{aligned} \min_{\alpha, \alpha^*} \quad & \frac{1}{2} (\alpha - \alpha^*)^T Q (\alpha - \alpha^*) + \epsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l z_i (\alpha_i - \alpha_i^*) \\ \text{subject to} \quad & e^T (\alpha - \alpha^*) = 0, \\ & 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \dots, l, \end{aligned}$$

and the approximate solution function for the dual problem is as follows:

$$\sum_{i=1}^l (\alpha^* - \alpha) K(\mathbf{x}_i, \mathbf{x}) + b$$

Finally, in the SVM, the output model is reported.

4 Experimental setup and results

In this section, the experimental evaluation of the framework devoted to soft biometric estimations is reported. In particular, the abovementioned biasing factors were evaluated on different data benchmarks incorporating also real-world contexts with challenging operating conditions.

Three different system settings were considered:

- the family of descriptors used to represent the facial features;
- the composition of the training set in terms of representatives of the classes to be predicted;
- the scaling of the input variables in the features vector to be supplied to the classifier.

Successively, the best configurations for training and scaling were tested over two real-world sources of bias:

- the spatial resolution of the facial image to be analyzed;
- the head pose of the subject in the scene.

To carry out the aforementioned analysis, two different experimental phases were performed. The first one was

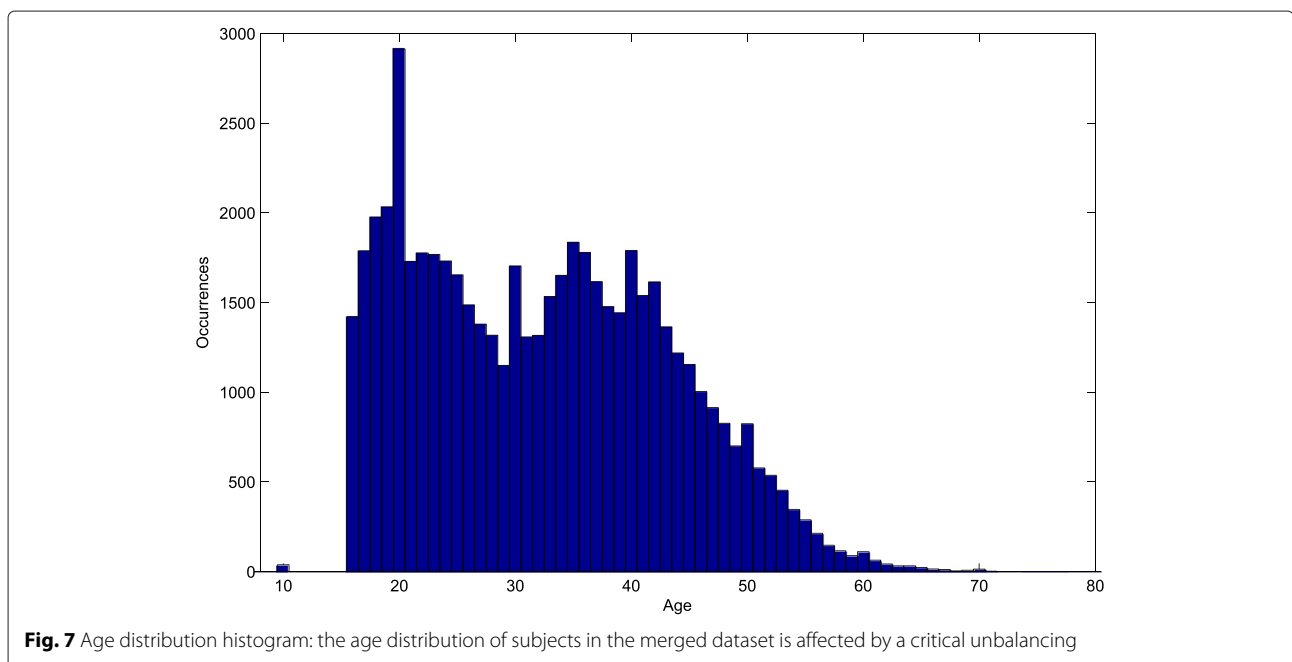


Fig. 7 Age distribution histogram: the age distribution of subjects in the merged dataset is affected by a critical unbalancing

carried out by using a large set of faces (more than 55,000) built by combining images available from different public datasets. The aim of this phase was to evaluate the performance of the soft biometrics estimation framework over images acquired under (quasi) ideal conditions (high resolution and frontal faces): in this way, the variability in classification performance when using balanced/unbalanced training sets, different descriptors, and data projection strategies, with and without data scaling, can be accurately evaluated. In the second phase, images acquired in a real environment were used instead: in this way, boundary conditions were relaxed by introducing changes in head pose (i.e., non-perfectly frontal face) and varying the distance from the camera (i.e., different spatial resolutions) with the aim to test the robustness of the different framework configurations under these two additional real-world biases. Both experimental phases are detailed in the following subsections.

4.1 Experimental phase 1

This experimental phase was carried out on two datasets built by properly merging two of the most representative publicly available repositories containing face images: the Morph [93] and Feret [94]. They consist of face images of people of different gender, race, and age, provided with complete annotation data. Moreover, both repositories present some replicated subject sampled in different dates (same person pictured in different years). The resolution of facial images is 200×240 , 400×480 , or 512×768 pixels.

Starting from the available images, a first experimental dataset was built without taking into account the balancing of the cardinality of each class. It consists of 46,669 male subjects and 9246 female subjects. The considered races were White, Black, Asian, and Hispanic with a cardinality respectively of 11,843, 41,334, 601, and 1889 subjects. Finally, the age distribution is represented in Fig. 7.

Moreover, two more balanced datasets were also built as subsets of the main one. The first one counts 8500 entries for each class of gender, whereas the second one is related to the race and counts 600 entries for each class (more balanced age group subsets cannot be built given the data distribution in the starting repositories).

The experimental results were obtained, on the aforementioned datasets, by means of an accurate cross validation procedure: the dataset under investigation was randomly split into k folds ($k = 5$) and, for each of the k validation steps, $k - 1$ folds were used for training, whereas the remaining fold was used for evaluating the estimation/validation capabilities. The k folds were built using a strategy to retain, in each of them, the same ratio of elements for the classes taken in the considered dataset: for example, in each fold, the ratio of male/female individuals was kept about to 1:10. The class separation for the

gender estimation task was straightforward. On the other hand, for race estimation, facial images were split into four categories: *White*, *Black*, *Asian*, and *Hispanic*. For the age issue, subjects were split into a number of categories equal to the number of the considered ages; the number of classes to be predicted was 42.

Face detection and normalization were then performed on each image of the selected $k-1$ training folds and then the vectors of features were extracted, properly projected

Table 1 Gender confusion tables

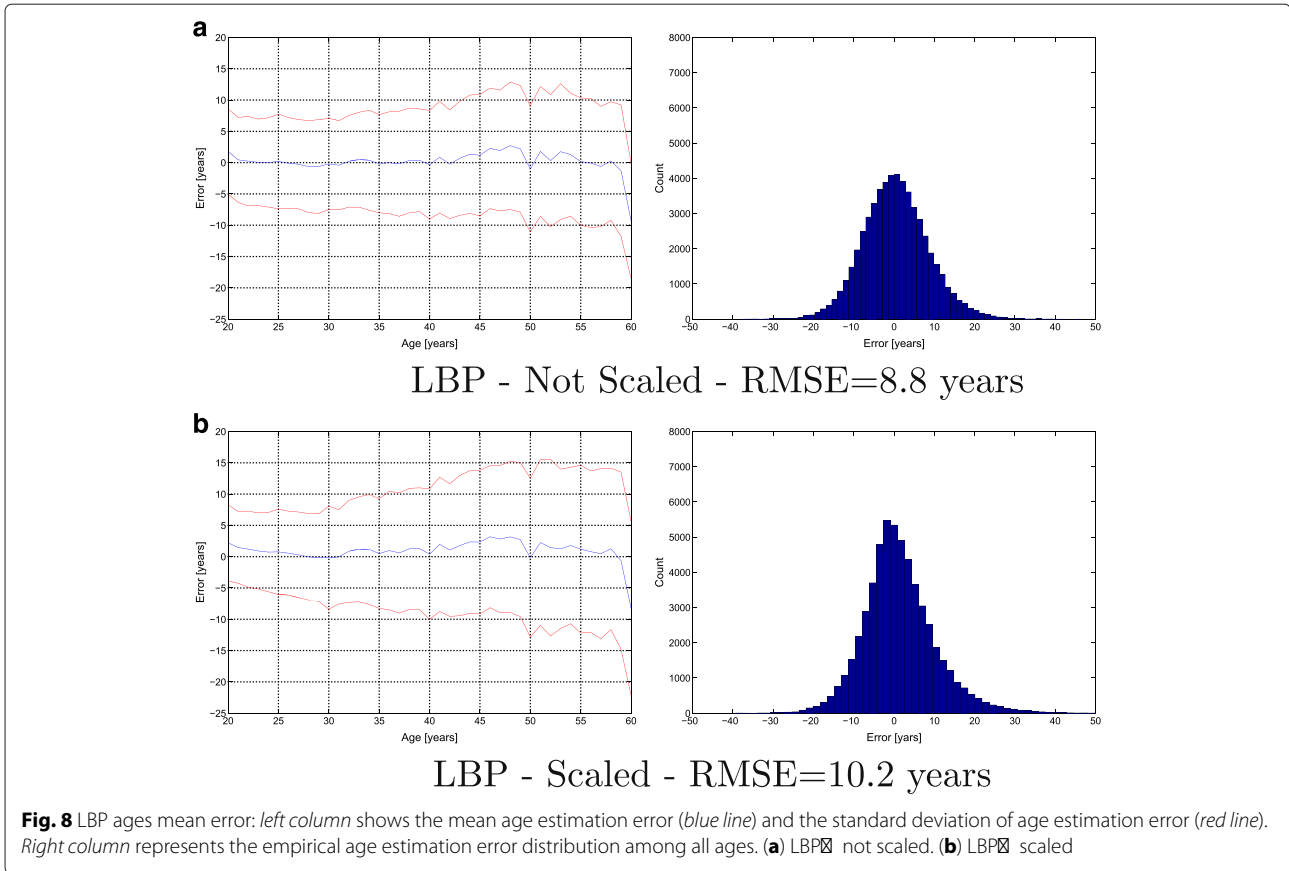
	(a) LBP \times unbalanced		(b) LBP \times balanced	
	M_T (%)	F_T (%)	M_T (%)	F_T (%)
M_P^N	97.9	2.1	M_P^N	93.8
F_P^N	14.1	85.9	F_P^N	5.2
M_P^S	98	2	M_P^S	73.5
F_P^S	14.7	85.3	F_P^S	28.4
TA^N		91.9	TA^N	94.3
TA^S		91.6	TA^S	72.5
	(c) HOG \times unbalanced		(d) HOG \times balanced	
	M_T (%)	F_T (%)	M_T (%)	F_T (%)
M_P^N	98.8	1.2	M_P^N	92.4
F_P^N	21.5	78.5	F_P^N	4.7
M_P^S	97.9	2.1	M_P^S	76.4
F_P^S	14.7	85.3	F_P^S	25.6
TA^N		88.6	TA^N	93.8
TA^S		91.6	TA^S	75.4
	(e) SWLD \times unbalanced		(f) SWLD \times balanced	
	M_T (%)	F_T (%)	M_T (%)	F_T (%)
M_P^N	97.7	2.3	M_P^N	92
F_P^N	17.8	82.2	F_P^N	7.5
M_P^S	97.7	2.3	M_P^S	83.2
F_P^S	17.7	82.3	F_P^S	20.4
TA^N		90	TA^N	92.3
TA^S		90	TA^S	81.4
	(g) CLBP \times unbalanced		(h) CLBP \times balanced	
	M_T (%)	F_T (%)	M_T (%)	F_T (%)
M_P^N	98.1	1.9	M_P^N	94.3
F_P^N	11.4	88.6	F_P^N	5.4
M_P^S	98.2	1.8	M_P^S	88
F_P^S	11.8	88.2	F_P^S	12.2
TA^N		93.4	TA^N	94.5
TA^S		93.2	TA^S	87.9

Each table presents the results for a specific descriptor for both scaled and non-scaled approaches. Moreover, the tables in the left column have been trained on unbalanced dataset; the tables in the right column over the balanced one. M and F are for male and female; suffix P and T indicate predicted and true value, respectively. TA is for total accuracy; superfix N and S refer to the non-scaled and scaled approach, respectively

Table 2 Race confusion tables

	(a) LBP _u unbalanced					(b) LBP _b balanced			
	W_T (%)	B_T (%)	A_T (%)	A_T (%)		W_T (%)	B_T (%)	A_T (%)	A_T (%)
W_P^N	95.7	2.1	0.5	1.7	W_P^N	62.8	5	10	22.2
B_P^N	0.8	98.8	0.1	0.3	B_P^N	4.8	81.8	3.6	9.8
A_P^N	22.6	10.3	57.3	9.8	A_P^N	10.2	5	70.8	14
H_P^N	31.9	14.2	2.7	51.2	H_P^N	22.4	10.4	14.4	52.8
W_P^S	96	2.2	0.3	1.5	W_P^S	64.6	4.8	9.4	21.1
B_P^S	0.8	98.8	0.1	0.2	B_P^S	5.8	79.8	3.4	1.1
A_P^S	22.1	11	55.9	11	A_P^S	10	4.40	71.6	14
H_P^S	32.5	15.9	2.2	49.4	H_P^S	21.2	9.4	12.6	56.8
TA^N		75.7			TA^N		67.1		
TA^S		75			TA^S		68.2		
	(c) HOG _u unbalanced					(d) HOG _b balanced			
	W_T (%)	B_T (%)	A_T (%)	A_T (%)		W_T (%)	B_T (%)	A_T (%)	A_T (%)
W_P^N	93.9	6.1	0	0	W_P^N	44.6	10	16.6	28.8
B_P^N	0.8	99.2	0	0	B_P^N	12.6	58.8	13.2	15.4
A_P^N	42.9	57.1	0	0	A_P^N	14.6	8.2	57	20.2
H_P^N	57.7	42.3	0	0	H_P^N	14.6	13	15.2	46
W_P^S	94.3	3.6	0.3	1.8	W_P^S	43.2	8.6	15.4	32.8
B_P^S	1	98.6	0	0.3	B_P^S	14.4	54	12.8	18.8
A_P^S	20	12.1	57.4	10.5	A_P^S	15.8	6.6	55.4	22.2
H_P^S	30.5	17.3	2	50.2	H_P^S	26.2	11.2	17	45.6
TA^N		48.3			TA^N		51.6		
TA^S		75.1			TA^S		49.5		
	(e) SWLD _u unbalanced					(f) SWLD _b balanced			
	W_T (%)	B_T (%)	A_T (%)	A_T (%)		W_T (%)	B_T (%)	A_T (%)	A_T (%)
W_P^N	94.6	3.4	0.4	1.6	W_P^N	51.2	9.8	14	25
B_P^N	1	98.5	0.1	0.4	B_P^N	12.6	58.8	9.6	19
A_P^N	23.5	10.8	56.9	8.8	A_P^N	15	6.2	61.8	17
H_P^N	37.8	19.9	2.4	39.9	H_P^N	22.4	17.6	16.4	43.6
W_P^S	94.3	3.4	0.4	1.9	W_P^S	50.4	8.2	13.2	28.2
B_P^S	1.1	98.5	0.1	0.3	B_P^S	13.8	54.2	9.4	22.6
A_P^S	25	10.6	53.2	11.2	A_P^S	14.6	6.4	60.4	18.6
H_P^S	36.6	20.2	2	41.2	H_P^S	24.4	15.6	15.2	44.8
TA^N		72.5			TA^N		53.8		
TA^S		71.8			TA^S		52.5		
	(g) CLBP _u unbalanced					(h) CLBP _b balanced			
	W_T (%)	B_T (%)	A_T (%)	A_T (%)		W_T (%)	B_T (%)	A_T (%)	A_T (%)
W_P^N	96.3	1.7	0.4	1.6	W_P^N	64.2	3.6	10.6	0
B_P^N	0.6	99.1	0.1	0.2	B_P^N	7.2	75	4.8	13
A_P^N	15.6	7.6	65.2	11.6	A_P^N	9.8	4.4	71.4	14.6
H_P^N	27.4	12.4	2.9	57.3	H_P^N	22.2	7.8	15.6	54.4
W_P^S	96.3	1.8	0.3	1.6	W_P^S	62.8	3.60	11.2	22.4
B_P^S	0.6	99.1	0.2	0.2	B_P^S	7.2	73.6	5	14.2
A_P^S	15.3	8.6	61.6	14.5	A_P^S	10.4	4	70	15.6
H_P^S	27.3	13.1	2.4	57.2	H_P^S	22.6	8.4	14.6	54.4
TA^N		79.4			TA^N		66.3		
TA^S		78.5			TA^S		65.2		

Each table presents the results for a specific descriptor for both scaled and non-scaled approaches. Moreover, the tables in the left column have been trained on the unbalanced dataset; the tables in the right column over the balanced one. White, Black, Asian, and Hispanic classes are referred respectively with W , B , A , and H . Subfix P and T indicate predicted and true value, respectively. TA is for total accuracy; superfix N and S refer to the non-scaled and scaled approaches, respectively



on the subspace defined by the LDA algorithm and finally supplied to the SVMs. After the training phase, the one-out fold was tested by using the available models. At the end of the iterative training/test process of the k fold method, the accuracy results were averaged.

The LBP algorithm¹ was spatially processed over a 5×5 grid, 8 neighbors, and 1 pixel radius; the CLBP algorithm was developed as reported in [9] and applied with the same parameters used for LBP. For the HOG operator, the *VLFeat library*² was used with standard parameters as suggested in [10] leading to a feature vector of 2016 elements; the WLD operator was implemented following the work in [35] and setting T , M , and S , respectively, to 8, 4, and 4 over a 4×4 grid of the image having a final vector of features of 2048 elements. SVM (for gender and race classification and for age regression) was used as implemented in the LIBSVM library [90], and the radial basis function (RBF) kernel, $K(\mathbf{x}, \mathbf{y}) = e^{-\gamma \|\mathbf{x} - \mathbf{y}\|^2}$, was used as suggested in [95] for non-linearly separable problems. Finally, the regularization parameter C was set to 1 and γ to $1/N_f$ where N_f is the number of features. In the case of scaling of the data in input to the SVM, the target range was [0,1] for each variable in the LDA vector of features.

The results of the experimental phase 1 for gender and race are reported in Tables 1 and 2, respectively.

For the age estimation, that is a regression problem, the pairs *predicted age/real age* of the whole k -fold procedure were considered and the average estimation error (blue line) and the standard deviations (red lines) are reported on the left side in Figs. 8, 9, 10, and 11. Let i be a specific age and N_i the corresponding number of entries; the age estimation error for the j -th entry of the age i is defined as:

$$e_{i,j} = \hat{i}_j - i_j \quad (8)$$

where \hat{i}_j is the predicted age and i_j is the age annotated in the dataset. Using the previous definition, it follows that the mean age estimation error for the age i can be defined as

$$\bar{e}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} e_{i,j} \quad (9)$$

and the error variance as

$$\sigma_i^2 = \frac{1}{N_i} \sum_{j=1}^{N_i} (e_{i,j} - \bar{e}_i)^2 \quad (10)$$

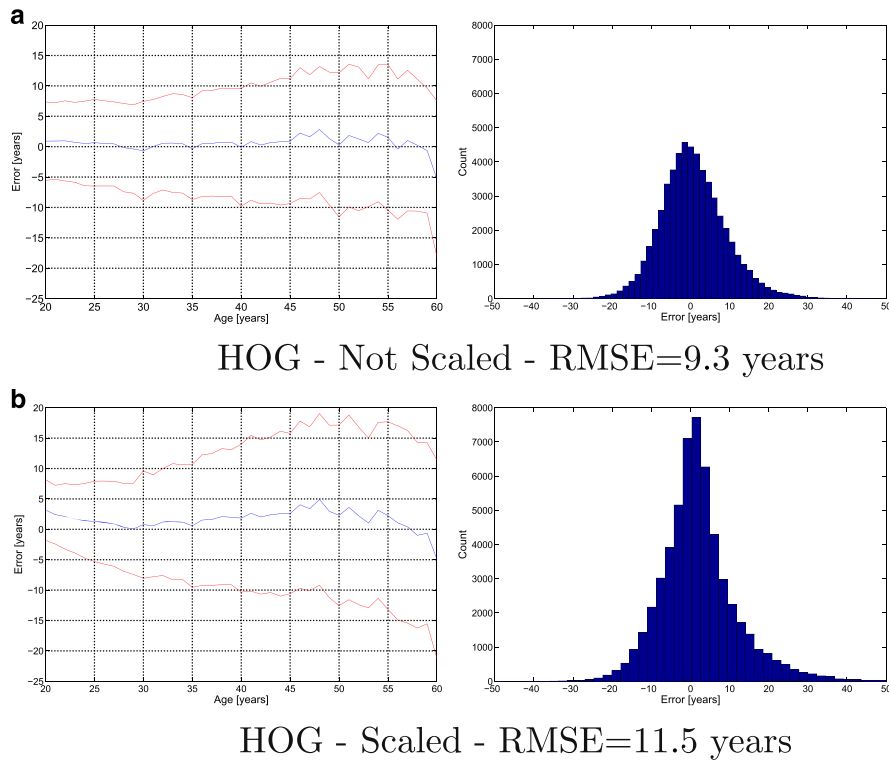


Fig. 9 HOG ages mean error: *left column* shows the mean age estimation error (*blue line*) and the standard deviation of age estimation error (*red line*). *Right column* represents the empirical age estimation error distribution among all ages. **(a)** HOG not scaled. **(b)** HOG scaled

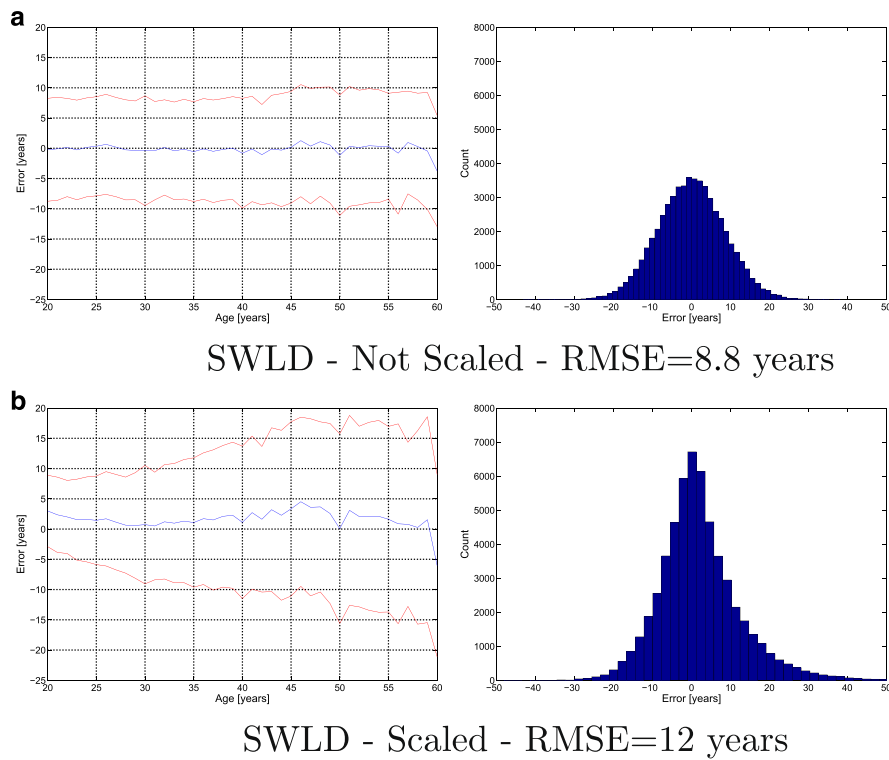
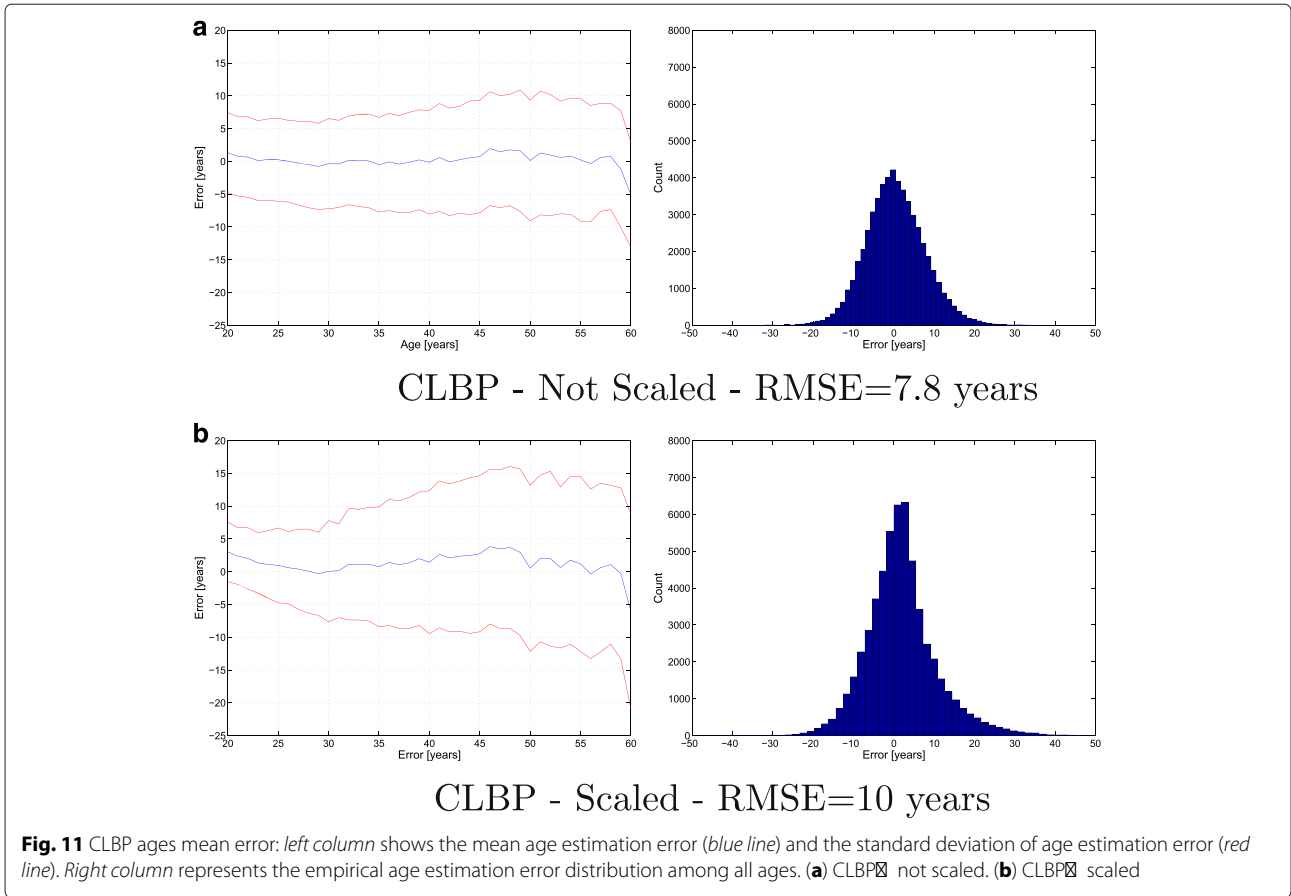


Fig. 10 SWLD ages mean error: *left column* shows the mean age estimation error (*blue line*) and the standard deviation of age estimation error (*red line*). *Right column* represents the empirical age estimation error distribution among all ages. **(a)** SWLD not scaled. **(b)** Scaled



Finally, for each age i , the RMS error ($RMSE_i$) was computed as

$$RMSE_i = \sqrt{\frac{1}{N_i} \sum_{j=1}^{N_i} (e_{ij})^2} \tag{11}$$

and the total RMS error (RMSE) as

$$RMSE = \frac{1}{N_a} \sum_{j=1}^{N_a} RMSE_j \tag{12}$$

where N_a is the number of ages between 20 and 60.

The histograms of the whole age estimation errors over all ages and samples were constructed and plotted on the right side of Figs. 8, 9, 10, and 11.

Results for gender and race were considered in terms of confusion tables where the rows are referred to the prediction (subscript P) and the columns to the true class (subscript T). In detail, the first two rows represent the prediction for the non-scaled case (superscript N), while the second two rows are referred to the scaled one (superscript S). The subtables (a), (c), (e), and (g) show the results for the unbalanced dataset; on the other hand, the subtables (b), (d), (f), and (h) refer to the results of the balanced

subset. The total accuracy (TA) for both the scaled and non-scaled cases is reported in the last two rows of each subtable.

More specifically, Table 1 shows that, for *gender estimation*, the CLBP descriptor, using a balanced dataset and non-scaled values as input to the classifier, returned the best total classification accuracy (94.5 %). This trend was also confirmed if the average of correct estimations (taking into account all the possible combinations of data scaling and dataset balancing) corresponding to each descriptor is considered. In fact, the CLBP classification rate was of 92.2 % which overcame the equivalent measurements obtained through the comparing descriptors (HOG 87.3 %, SWLD 88.4 %, and LBP 87.6 %).

With regard to the opportunity to scale in the range [0,1], the features in input to the SVM and the experiments evidenced that non-scaled data better preserved the embedded information: the average correct estimations (taking into account all the possible descriptors and combinations of dataset balancing) were 92.3 % in the case of non-scaled input and 85.4 % in the case of scaled input.

Finally, concerning the opportunity to use a balanced dataset for training, experiments evidenced that balanced dataset led to an average correct estimation (taking into

account all the possible descriptors with no data scaling) of 93.7 % in the case of balanced dataset and 91.0 % in the case of unbalanced dataset.

To summarize, for gender estimation, the best configuration of the proposed framework can be assumed to use a CLBP classifier, learned on balanced dataset and working on non-scaled input data.

Race estimation problem is a very different problem due to the larger number of classes (four classes), the presence of borderline cases, and the critical unbalancing of subjects among classes (there are just 600 Asian people against the 42,000 Black subjects).

Using the same evaluation criteria reported above, Table 2 shows that, for *race estimation*, the CLBP descriptor, using an unbalanced dataset and non-scaled values as input to the classifier, returned the best total classification accuracy (79.4 %).

The average correct estimations (taking into account all the possible combinations of data scaling and dataset balancing) corresponding to each descriptor were LBP 71.5 %, HOG 56.1 %, SWLD 62.6 %, and CLBP 72.4 %.

Concerning the opportunity to scale in the range [0,1], the variables of the feature vectors and experiments evidenced that, taking into account all the possible

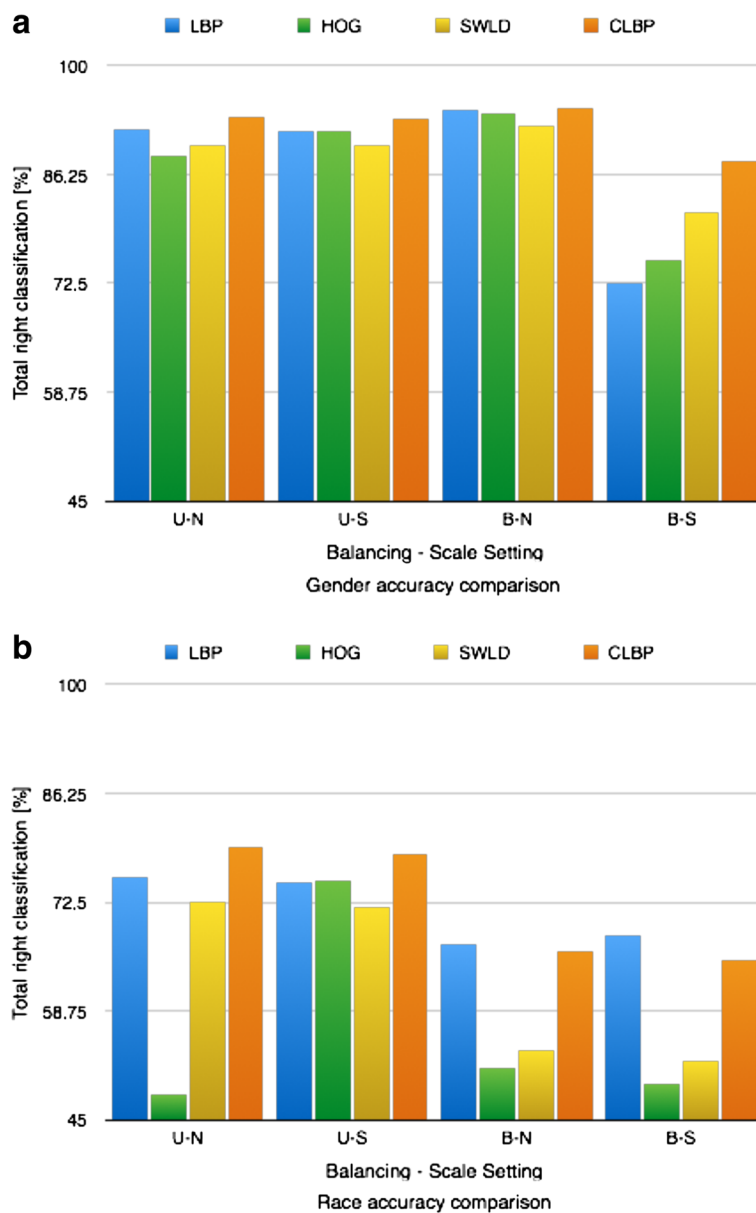


Fig. 12 Gender (a) and race accuracy (b) comparison: each group of columns represents (from left to right) the unbalanced non-scaled (U-N), the unbalanced scaled (U-S), the balanced non-scaled (B-N), and the balanced scaled (B-S) configuration

descriptors and combinations of dataset balancing, scaled data gave the best average correct estimations (67.0 %). However, an in-depth analysis of the confusion tables reveals as this result is mainly biased by the HOG descriptor, operating on the unbalanced dataset by means of non-scaled data, bad outcomes (48.3 %). With regard to all the other descriptors, the non-scaled data approach is not worse than the scaled one. Taking into account all the possible descriptors, the unbalanced dataset for training led to a better average correct estimation both in the case of scaled data (75.1 % for unbalanced case and 58.8 % for balanced case) and non-scaled data (69.0 % for unbalanced case and 59.7 % for balanced case).

The unbalanced dataset worked better because the number of Asiatic representatives is lower than the remaining classes, and therefore, in the case of the

balanced case, a low number of samples can be used for each class. In other words, if the number of representatives for each class of race is sufficiently large, the combination of CLBP, balanced dataset, and non-scaled input to the SVM seemed to be the best choice.

A compact view of the performance, for gender and race estimation, is presented in Fig. 12 by means of bar diagrams. In particular, Fig. 12a refers to gender estimation, whereas Fig. 12b refers to race estimation.

Finally, the *age estimation* was addressed as a regression problem and, in this case, having 42 classes to be predicted, the balanced dataset was not built (since there were not enough representatives in each class). The experimental results, reported in Figs. 8, 9, 10, and 11, demonstrated that the CLBP without data scaling was the best performing descriptor (RMSE = 7.8 years). The best



Fig. 13 Face examples: some pose examples for the dataset acquired in the lab. The first three columns are referred to the frontal pose, the second three columns to the rotated pose, and the last three columns to the far distance pose

average correct estimation (taking into account both combinations of data scaling/not scaling) was again obtained with CLBP (RMSE = 8.9 years) which overcame the other descriptors (LBP-RMSE = 9.5 years; HOG-RMSE = 10.4 years; SWLD-RMSE = 10.4 years). Concerning the opportunity to scale the data, experimental results demonstrated that non-scaled data led to a better average regression accuracy (RMSE = 8.7 years) with respect to the case in which data fed into SVR were scaled in the range [0,1] (RMSE = 10.9 years). It is interesting to observe that the error distribution in the age estimation problem (see plots on the right of Figs. 8, 9, 10, and 11) revealed a Gaussian trend (around the zero value) with a steep decrease, demonstrating that in the most of age classification cases, the estimations were quite correct.

4.2 Experimental phase 2

In this phase, experiments were carried out on facial images acquired in a real environment: in this way, boundary conditions were relaxed with respect to previous experimental phase. In particular, changes in head pose (i.e., non-perfectly frontal face) and distance from the camera (i.e., different spatial resolutions) were also considered as factors of classification/regression bias. A dataset of seven subjects (five males and two females, all belonging to the white race and age ranging from 23 to 48 years) was built on purpose with 900 images per subject: in 300 images, subjects were at a distance of 75 cm from the camera and they assumed a close frontal pose (this subset of images were labelled as I_{FC} , where the subscript stands for frontal-close); in other 300 images, subjects were at a distance of 75 cm from the camera but they performed a head rotation in the range $[10^\circ, 30^\circ]$ (this subset of images was labelled as I_{RC} , where the subscript stands for rotated-close). Finally, in the remaining 300 images, subjects were at a distance of 165 cm from the camera and they assume a close frontal pose (this subset of images were labelled as I_{FF} , where the subscript stands for frontal-far). In the case where the subjects were closer to the camera, the resolution of the facial patch is about $[240 \times 180]$ pixels, whereas in the case where the subjects were farther from the camera, their facial patch is of about $[100 \times 70]$ pixels. Some examples of the three aforementioned subsets of images are reported in Fig. 13 where columns 1–3 report frontal-close images, columns

Table 3 Gender estimation performance for the lab dataset

	LBP (%)	HOG (%)	SWLD (%)	CLBP (%)
I_{FC}	98.4	97.9	98.4	98.6
I_{RC}	95.7	85.3	92.1	97.9
I_{FF}	88.4	97.5	87.9	89.0

I_{FC} is referred to the frontal-close images, I_{RC} to the rotated-close images, and I_{FF} to the frontal-far images

Table 4 Race estimation performance rate for the lab dataset

	LBP (%)	HOG (%)	SWLD (%)	CLBP (%)
I_{FC}	97.7	93.3	89.2	96.1
I_{RC}	97.5	86.0	84.6	93.2
I_{FF}	97.1	92.4	89.9	89.1

I_{FC} is referred to the frontal-close images, I_{RC} to the rotated-close images, and I_{FF} to the frontal-far images

4–6 report rotated-close images, and columns 7–9 report frontal-far images.

Experiments were carried out as follows: the facial datasets built in the experimental phase 1 was used to set the projection (through LDA), classification/regression (through SVM), and scaling models. These models were then stored and exploited, without any intervention, during all the soft biometric estimations carried out in the experimental phase 2.

In this case, due to the dataset composition, the measurement of the error for age prediction was slightly modified as follows: let N_s the numbers of subjects and N_f the numbers of frames for each subject (given a pose configuration).

The age estimation was treated with an RMS age estimation error.

$$\text{RMSE}_j = \sqrt{\frac{1}{N_f} \sum_{i=1}^{N_f} (\hat{\theta}_{i,j} - \theta_j)^2} \quad (13)$$

where $\hat{\theta}_{i,j}$ is the age estimation for the i -th frame of the j -th subject and θ_j is the real age of the j -th subject. The mean RMS error (for all the subjects) is defined as

$$\text{RMSE} = \frac{1}{N_s} \sum_{j=1}^{N_s} \text{RMSE}_j \quad (14)$$

Finally, let $e_{i,j} = (\hat{\theta}_{i,j} - \theta_j)$ be the age estimation error for the i -th frame of the j -th subject and \bar{e}_j its mean among the N_f frames. The error variance of the j -th subject is as follows:

Table 5 Age estimation RMSE for the lab dataset

	LBP	HOG	SWLD	CLBP
I_{FC} RMSE	7.7	14.2	9.5	7.6
I_{RC} RMSE	8.3	16.6	9.5	7.9
I_{FF} RMSE	9.2	14.6	10.6	9.4
$I_{FC} \sigma^2$	12.2	15.9	19.3	12.9
$I_{RC} \sigma^2$	15.8	31.8	31.6	17.8
$I_{FF} \sigma^2$	9.1	21.5	26.7	14.9

I_{FC} is referred to the frontal-close images, I_{RC} to the rotated-close images, and I_{FF} to the frontal-far images

$$\sigma_j^2 = \frac{1}{N_f} \sum_{i=1}^{N_f} (e_{i,j} - \bar{e}_j)^2 \tag{15}$$

and the total mean variance is defined as

$$\sigma^2 = \frac{1}{N_s} \sum_{j=1}^{N_s} \sigma_j^2 \tag{16}$$

The experimental outcomes of this phase are reported in Tables 3, 4, and 5 and Figs. 14 and 15. In particular, tables report gender, race, and age estimation correctness

for all the descriptors and over all the three experimental conditions (frontal-close images I_{FC} , rotated-close images I_{RC} , frontal-far images I_{FF}). It is very important to highlight here that the best configurations (revealed by experimental phase 1) for balancing and scaling strategies were used and that, due to the lack of statistical significance (i.e., few subjects), the aim of the test on I_{FC} images was not to evaluate the robustness of the framework (as largely tested in the previous subsection) but to have a reference point in the evaluation of the influence of pose and resolution. From this point of view, tables have to

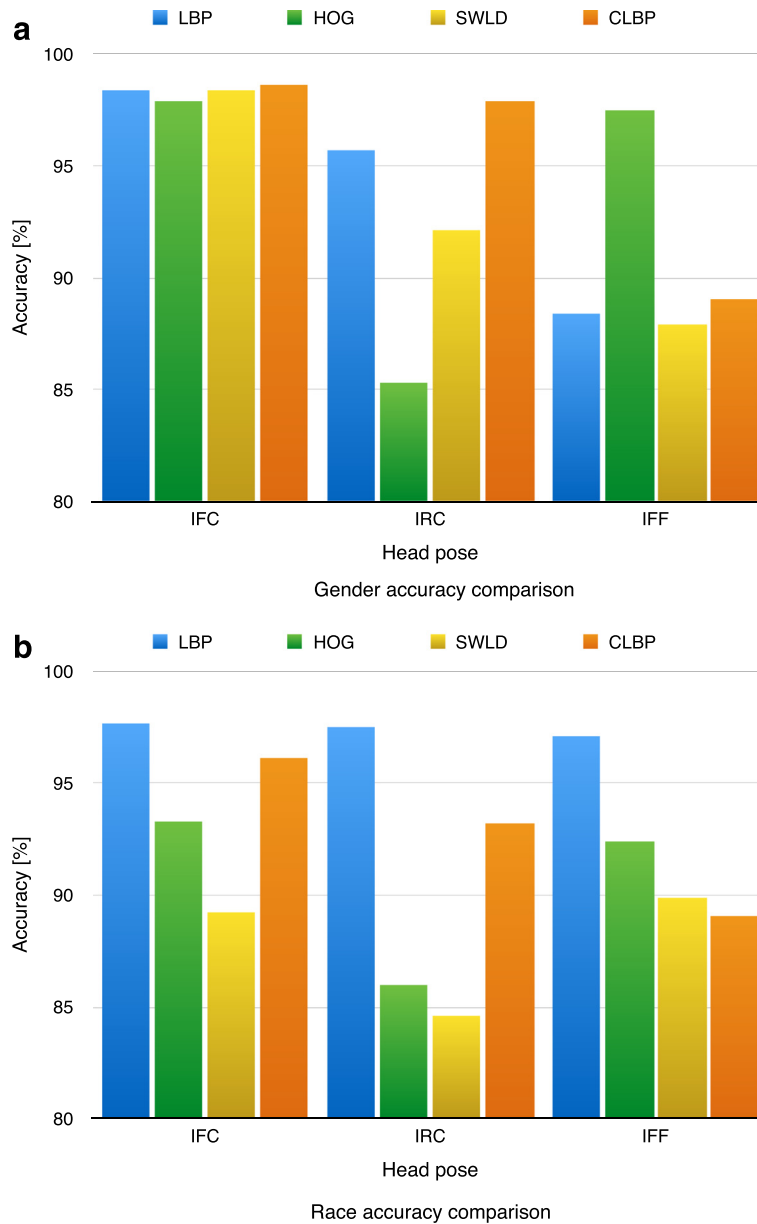


Fig. 14 Gender (a) and race accuracy (b) comparison for lab dataset: each group of columns is referred to (from left to right) the frontal-close images (I_{FC}), the rotated-close images (I_{RC}), and the frontal-far images (I_{FF}) configuration

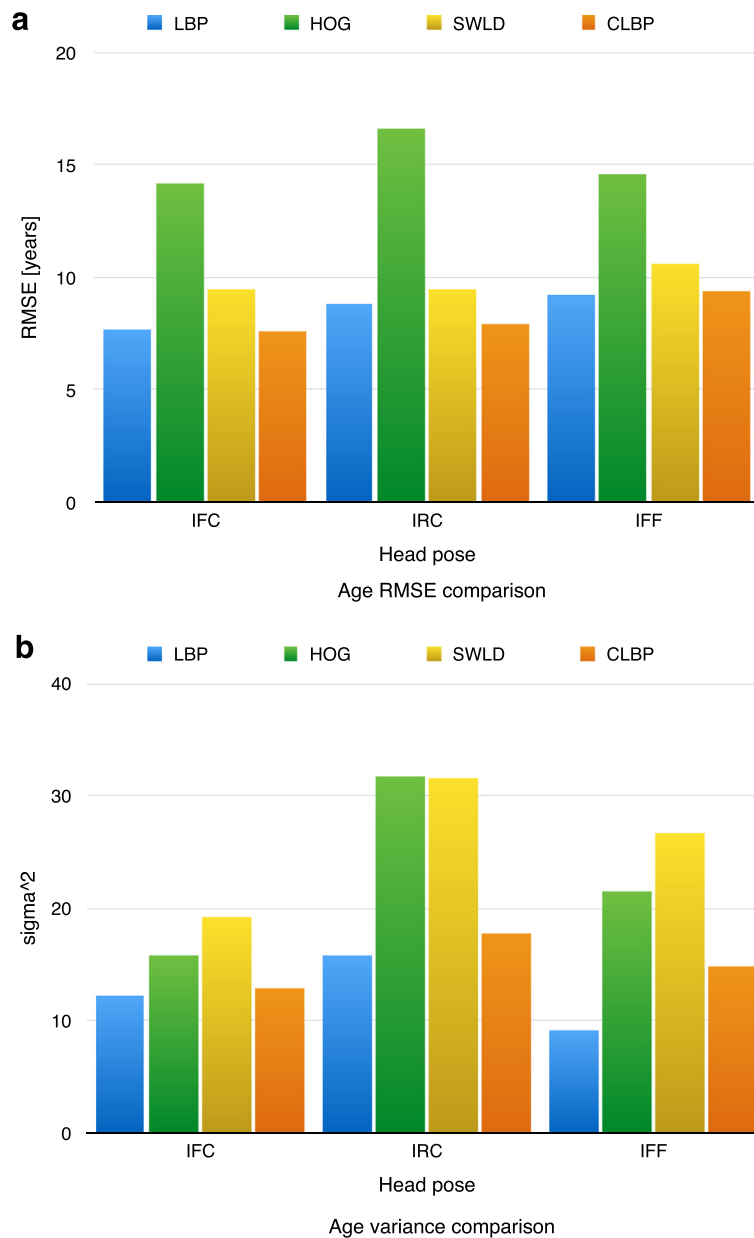


Fig. 15 Age RMSE (a) and variance (b) for lab dataset: each group of columns is referred to (from left to right) the frontal-close images (I_{FC}), the rotated-close images (I_{RC}), and the frontal-far images (I_{FF}) configuration

be exploited to understand how much the performance of the various configurations of descriptor, dataset balancing, and scaling strategies are influenced by pose and resolution changes. From the results, it is evident that the descriptor that is less sensitive to the image resolution is HOG (where a decrease of 0.4 % for gender, 0.9 % for race, and 0.4 years for age can be observed) and this is due to the fact that it is edge based (able to capture coarse texture information, as opposed to the other investigated methods able to represent finer structures). Unfortunately, the HOG descriptor experienced the maximum decrease in

performance in the case of changes in pose (12.6 % for gender, 7.3 % for race, and 2.4 years for age) where the most robust behavior was obtained by the CLBP (0.7 % for gender, 2.9 % for race, and 0.3 years for age) which, by using a more sophisticated encoding strategy, was able to handle variations in facial patterns.

5 Conclusions

In this paper, a detailed study about the performance of soft biometric estimations, in terms of accuracy as well as robustness under real-world conditions, has

been presented. Firstly, different facial descriptors (LBP, HOG, SWLD, CLBP), training datasets, and data scaling approaches have been extensively tested in a recently introduced common framework including a data reduction step (making use of linear discriminant analysis) and a final SVM classification step. Successively, robustness in unconstrained scenario, with scaling and rotation of the target with respect to the acquisition device, has been studied. Experimental evaluations have been carried out on benchmark data built from both publicly available datasets and from image sequences acquired in lab environment. Results highlighted that the best average performances have been reached by the CLBP operator in all the estimation problems. This work proved that the non-scaled features vector approach is preferable to the scaled one and that balanced class cardinalities during dataset building can give better performances. Moreover, tests have been performed over the lab dataset with relaxed boundary conditions which proved that the CLBP operator outperforms the other descriptors in terms of average estimation performance in the presence of scaling and rotation. Future works will aim at extending the lab dataset both in terms of number of subjects and pose and resolution variability, in order to have more significant statistics.

Endnotes

¹The on-line implementation, available at <http://www.bytefish.de>, was used.

²<http://www.vlfeat.org>.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work has been supported in part by the 2007–2013 NOP for Research and Competitiveness for the Convergence Regions (Calabria, Campania, Puglia and Sicilia) SARACEN with code PON04a3_00201 and in part by the PON Baitah, with code PON01_980.

Received: 31 October 2014 Accepted: 20 October 2015

Published online: 20 November 2015

References

1. T Martiriggiano, M Leo, T D'Erario, A Distanto, in *Innovations in Applied Artificial Intelligence*. Lecture Notes in Computer Science, ed. by M Ali, F Esposito. Face recognition by kernel independent component analysis, vol. 3533 (Springer, Berlin Heidelberg, 2005), pp. 55–58
2. AK Jain, SC Dass, K Nandakumar, in *Biometric Authentication*. Soft biometric traits for personal recognition systems (Springer, Berlin Heidelberg, 2004), pp. 731–738
3. P Carcagni, D Cazzato, M Del Coco, C Distanto, M Leo, in *Computer Vision-ECCV 2014 Workshops*. Visual interaction including biometrics information for a socially assistive robotic platform (Springer International Publishing, Gewerbestrasse 11 CH-6330 Cham (ZG)Switzerland, 2014), pp. 391–406
4. A Tapus, C Tapus, MJ Mataric, in *IEEE International Conference on Rehabilitation Robotics (ICORR)*. The use of socially assistive robots in the design of intelligent cognitive therapies for people with dementia (IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, California, 2009), pp. 924–929
5. R Bemelmans, GJ Gelderblom, P Jonker, L De Witte, Socially assistive robots in elderly care: a systematic review into effects and effectiveness. *J. Am. Med. Dir. Assoc.* **13**(2), 114–120 (2012)
6. Carcagni, M Del Coco, P Mazzeo, A Testa, C Distanto, in *Video Analytics for Audience Measurement*. Lecture Notes in Computer Science. Features descriptors for demographic estimation: a comparative study (Springer, International Publishing, Gewerbestrasse 11 CH-6330 Cham (ZG)Switzerland, 2014), pp. 66–85
7. T Ojala, M Pietikainen, D Harwood, A comparative study of texture measures with classification based on featured distributions. *Pattern Recognit.* **29**(1), 51–59 (1996)
8. T Ojala, M Pietikainen, T Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Anal. Mach. Intell.* *IEEE Trans.* **24**(7), 971–987 (2002)
9. F Ahmed, E Hossain, ASMH Bari, A Shihavuddin, in *Computational Intelligence and Informatics (CINTI), 2011 IEEE 12th International Symposium On*. Compound local binary pattern (CLBP) for robust facial expression recognition (IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, California, 2011), pp. 391–395
10. N Dalal, B Triggs, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Histograms of oriented gradients for human detection, vol. 1 (IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, California, 2005), pp. 886–893
11. J Chen, S Shan, C He, G Zhao, M Pietikainen, X Chen, W Gao, WLD: a robust local image descriptor. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(9), 1705–1720 (2010)
12. PN Belhumeur, JP Hespanha, D Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 711–720 (1997)
13. C Cortes, V Vapnik, Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
14. BE Boser, IM Guyon, VN Vapnik, in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. COLT **2**. A training algorithm for optimal margin classifiers (ACM, New York, NY, USA, 1992), pp. 144–152
15. Z Sun, G Bebis, X Yuan, SJ Louis, in *IEEE Workshop on Applications of Computer Vision*. Genetic feature subset selection for gender classification: a comparison study (IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, California, 2002), pp. 165–170
16. N Sun, W Zheng, C Sun, C Zou, L Zhao, ed. by J Wang, Z Yi, J Zurada, B-L Lu, and H Yin. *Advances in Neural Networks - ISNN. Lecture Notes in Computer Science*, vol. 3972 (Springer, Berlin Heidelberg, 2006), pp. 194–201
17. E Mäkinen, R Raisamo, An experimental comparison of gender classification methods. *Pattern Recognit. Lett.* **29**(10), 1544–1556 (2008)
18. M Sakarkaya, F Yanbol, Z Kurt, in *IEEE 16th International Conference on Intelligent Engineering Systems (INES)*. Comparison of several classification algorithms for gender recognition from face images (IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, California, 2012), pp. 97–101
19. C Ng, Y Tay, B-M Goi, in *PRICAI 2012: Trends in Artificial Intelligence*. Lecture Notes in Computer Science, ed. by P Anthony, M Ishizuka, and D Lukose. Recognizing human gender in computer vision: a survey, vol. 7458 (Springer, Berlin Heidelberg, 2012), pp. 335–346
20. R Brunelli, T Poggio, Face recognition: features versus templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(10), 1042–1052 (1993)
21. J-M Fellous, Gender discrimination and prediction on the basis of facial metric information. *Vis. Res.* **37**(14), 1961–1973 (1997)
22. S Mozaffari, H Behravan, R Akbari, in *20th International Conference on Pattern Recognition (ICPR)*. Gender classification using single frontal image per person: combination of appearance and geometric based features (IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, California, 2010), pp. 1192–1195
23. H Abdi, D Valentin, B Edelman, AJ O'Toole, More about the difference between men and women: evidence from linear neural networks and the principal-component approach. **24**, 539 (Pion Ltd, 1995)
24. S Tamura, H Kawai, H Mitsumoto, Male/female identification from 8 × 6 very low resolution face images by neural network. *Pattern Recognit.* **29**(2), 331–335 (1996)
25. H-C Lian, B-L Lu, ed. by J Wang, Z Yi, J Zurada, B-L Lu, and H Yin. *Advances in Neural Networks. Lecture Notes in Computer Science*, vol. 3972 (Springer, Berlin Heidelberg, 2006), pp. 202–209

26. Z Yang, H Ai, ed. by S-W Lee, S Li. *Advances in Biometrics. Lecture Notes in Computer Science*, vol. 4642 (Springer, Berlin Heidelberg, 2007), pp. 464–473
27. LA Alexandre, Gender recognition: a multiscale decision fusion approach. *Pattern Recognit. Lett.* **31**(11), 1422–1427 (2010)
28. J Ylioinas, A Hadid, M Pietikäinen, ed. by A Heyden, F Kahl. *Image Analysis. Lecture Notes in Computer Science*, vol. 6688 (Springer, Berlin Heidelberg, 2011), pp. 676–686
29. C Shan, Learning local binary patterns for gender classification on real-world face images. *Pattern Recognit. Lett.* **33**(4), 431–437 (2012). Intelligent Multimedia Interactivity
30. T-X Wu, X-C Lian, B-L Lu, Multi-view gender classification using symmetry of facial images. *Neural Comput. Appl.* **21**(4), 661–669 (2012)
31. J Zheng, B-L Lu, A support vector machine classifier with automatic confidence and its application to gender classification. *Neurocomputing.* **74**(11), 1926–1935 (2011)
32. T Jabid, MH Kabir, O Chae, in *20th International Conference on Pattern Recognition (ICPR)*. Gender classification using local directional pattern (LDP) (IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, California, 2010), pp. 2162–2165
33. M Castrillón-Santana, J Lorenzo-Navarro, E Ramón-Balmaseda, in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Lecture Notes in Computer Science*, ed. by J Ruiz-Shulcloper, G Sanniti di Baja. Improving gender classification accuracy in the wild, vol. 8259 (Springer, Berlin Heidelberg, 2013), pp. 270–277
34. G Guo, CR Dyer, Y Fu, TS Huang, in *IEEE 12th International Conference on Computer Vision Workshops (ICCV)*. Is gender recognition affected by age? (IEEE, 2009), pp. 2032–2039
35. I Ullah, M Hussain, G Muhammad, H Aboalsamh, G Bebis, AM Mirza, in *19th International Conference on Systems, Signals and Image Processing (IWSSIP)*. Gender recognition from face images with local WLD descriptor (IEEE Signal processing Society, 445 Hoes Lane, Piscataway, N.J. 08854, U.S.A., 2012), pp. 417–420
36. W Gao, H Ai, ed. by M Tistarelli, M Nixon. *Advances in Biometrics. Lecture Notes in Computer Science*, vol. 5558 (Springer, Berlin Heidelberg, 2009), pp. 169–178
37. Y Saatici, C Town, in *7th International Conference on Automatic Face and Gesture Recognition (FGR)*. Cascaded classification of gender and facial expression using active appearance models (IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, California, 2006), pp. 393–398
38. TF Cootes, GJ Edwards, CJ Taylor, Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(6), 681–685 (2001)
39. M Demirkus, M Toews, JJ Clark, T Arbel, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Gender classification from unconstrained video sequences (IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, California, 2010), pp. 55–62
40. J-G Wang, J Li, W-Y Yau, E Sung, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Boosting dense sift descriptors and shape contexts of face images for gender recognition (IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, California, 2010), pp. 96–102
41. X Leng, Y Wang, in *15th IEEE International Conference on Image Processing (ICIP)*. Improving generalization for gender classification (IEEE Signal processing Society, 445 Hoes Lane, Piscataway, N.J. 08854, U.S.A., 2008), pp. 1656–1659
42. F Scalzo, G Bebis, M Nicolescu, L Loss, A Tavakkoli, in *19th International Conference on Pattern Recognition (ICPR)*. Feature fusion hierarchies for gender classification (IEEE Signal processing Society, 445 Hoes Lane, Piscataway, N.J. 08854, U.S.A., 2008), pp. 1–4
43. E Meyers, L Wolf, Using biologically inspired features for face processing. *Int. J. Comput. Vis.* **76**(1), 93–104 (2008)
44. SM Mansoor Roomi, SL Virasundarii, S Selvamegala, S Jeevanandhame, D Hariharasudhan, in *Third National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*. Race classification based on facial features (IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, California, 2011), pp. 54–57
45. Y Xie, K Luu, M Savvides, in *IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. A robust approach to facial ethnicity classification on large scale face databases (IEEE Signal processing Society, 445 Hoes Lane, Piscataway, N.J. 08854, U.S.A., 2012), pp. 143–149
46. HHK Tin, MM Sein, Race identification for face images. *ACEEE Int. J. Inform. Tech.* **1**(02) (2011)
47. A Awwad, A Ahmad, Salameh AW, Arabic race classification of face images. *Int. J. Comput. Technol.* **4**(2a1) (2013)
48. Y Ou, X Wu, H Qian, Y Xu, in *IEEE International Conference on Information Acquisition*. A real time race classification system (IEEE Signal processing Society, 445 Hoes Lane, Piscataway, N.J. 08854, U.S.A., 2005), p. 6
49. FS Manesh, M Ghahramani, YP Tan, in *11th International Conference on Control Automation Robotics Vision (ICARCV)*. Facial part displacement effect on template-based gender and ethnicity classification (IEEE Signal processing Society, 445 Hoes Lane, Piscataway, N.J. 08854, U.S.A., 2010), pp. 1644–1649
50. C Chen, A Ross, in *SPIE Defense, Security, and Sensing*. Local gradient gabor pattern (LGGP) with applications in face recognition, cross-spectral matching, and soft biometrics (International Society for Optics and Photonics, Bellingham, Washington USA, 2013), pp. 87120R–87120R
51. MJ Lyons, J Budynek, A Plante, S Akamatsu, in *Proceedings. Fourth IEEE International Conference on Automatic Face and Gesture Recognition*. Classifying facial attributes using a 2-d gabor wavelet representation and discriminant analysis (IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, California, 2000), pp. 202–207
52. G Guo, G Mu, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. A study of large-scale ethnicity estimation with gender and age variations (IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, California, 2010), pp. 79–86
53. G Guo, G Mu, Y Fu, C Dyer, T Huang, in *IEEE 12th International Conference on Computer Vision*. A study on automatic age estimation using a large database (IEEE Signal processing Society, 445 Hoes Lane, Piscataway, N.J. 08854, U.S.A., 2009), pp. 1986–1991
54. G Guo, G Mu, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Human age estimation: what is the influence across race and gender? (IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, California, 2010), pp. 71–78
55. G Guo, C Zhang, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. A study on cross-population age estimation (IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, California, 2014), pp. 4257–4263
56. G Guo, G Mu, A framework for joint estimation of age, gender and ethnicity on a large database. *Image Vis. Comput.* **32**(10), 761–770 (2014). Best of Automatic Face and Gesture Recognition 2013
57. SH Salah, H Du, N Al-Jawad, in *Proc. IEEE International Conference Signal Image Process*. Fusing local binary patterns with wavelet features for ethnicity identification, vol. 21 (IEEE Signal processing Society, 445 Hoes Lane, Piscataway, N.J. 08854, U.S.A., 2013), pp. 416–422
58. G Muhammad, M Hussain, F Alenezy, G Bebis, AM Mirza, H Aboalsamh, Race classification from face images using local descriptors. *Int. J. Artif. Intell. Tools.* **21**(05) (2012)
59. G Muhammad, M Hussain, F Alenezy, G Bebis, AM Mirza, H Aboalsamh, in *19th International Conference on Systems, Signals and Image Processing (IWSSIP)*. Race recognition from face images using weber local descriptor (IEEE Signal processing Society, 445 Hoes Lane, Piscataway, N.J. 08854, U.S.A., 2012), pp. 421–424
60. X Lu, AK Jain, in *Proc. SPIE Defense and Security*. Ethnicity identification from face images, (2004), pp. 114–123. International Society for Optics and Photonics
61. G Toderici, S O'Malley, G Passalis, T Theoharis, I Kakadiaris, Ethnicity- and gender-based subject retrieval using 3-d face-recognition techniques. *Int. J. Comput. Vis.* **89**(2–3), 382–391 (2010)
62. Y Fu, G Guo, TS Huang, Age synthesis and estimation via faces: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(11), 1955–1976 (2010)
63. M Singh, S Nagpal, R Singh, M Vatsa, On recognizing face images with weight and age variations. *IEEE Access.* **2**, 822–830 (2014)
64. TR Alley, *Social and Applied Aspects of Perceiving Faces*. (Psychology Press, Oxford, UK, 2013)
65. YH Kwon, N da Vitoria Lobo, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Age classification from facial images (IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, California, 1994), pp. 762–767

66. LG Farkas, *Anthropometry of the Head and Face*. (Raven Pr, New York, NY, 1994)
67. A Lanitis, CJ Taylor, TF Cootes, Toward automatic simulation of aging effects on face images. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(4), 442–455 (2002)
68. L Liu, J Liu, J Cheng, in *11th International Conference on Machine Learning and Applications (ICMLA)*. Age-group classification of facial images, vol. 1 (IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, California, 2012), pp. 693–696
69. X Geng, Z-H Zhou, K Smith-Miles, Automatic age estimation based on facial aging patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(12), 2234–2240 (2007)
70. HS Seung, DD Lee, The manifold ways of perception. *Science*. **290**(5500), 2268–2269 (2000)
71. Y Fu, Y Xu, TS Huang, in *IEEE International Conference on Multimedia and Expo*. Estimating human age by manifold analysis of face pictures and regression on aging features (IEEE Computer Society, 445 Hoes Lane, Piscataway, N.J. 08854, U.S.A., 2007), pp. 1383–1386
72. J Hayashi, M Yasumoto, H Ito, H Koshimizu, in *Seventh International Conference on Virtual Systems and Multimedia*. Method for estimating and modeling age and gender using facial image processing (IEEE Computer Society, 10662 Los Vaqueros Circle, Los Alamitos, California, 2001), pp. 439–448
73. J Hayashi, M Yasumoto, H Ito, Y Niwa, H Koshimizu, in *41st SICE Annual Conference*. Age and gender estimation from facial image processing, vol. 1 (IEEE Computer Society, 445 Hoes Lane, Piscataway, N.J. 08854, U.S.A., 2002), pp. 13–18
74. A Gunay, VV Nabiyev, in *23rd International Symposium on Computer and Information Sciences (ISCIS)*. Automatic age classification with LBP (IEEE Computer Society, 445 Hoes Lane, Piscataway, N.J. 08854, U.S.A., 2008), pp. 1–4
75. J Ylioinas, A Hadid, M Pietikainen, in *21st International Conference on Pattern Recognition (ICPR)*. Age classification in unconstrained conditions using LBP variants (IEEE Computer Society, 445 Hoes Lane, Piscataway, N.J. 08854, U.S.A., 2012), pp. 1257–1260
76. J Ylioinas, A Hadid, X Hong, M Pietikainen, in *International Conference on Image Analysis and Processing (ICIAP)*. Lecture Notes in Computer Science, ed. by A Petrosino. Age estimation using local binary pattern kernel density estimate, vol. 8156 (Springer, Berlin Heidelberg, 2013), pp. 141–150
77. F Gao, H Ai, in *Advances in Biometrics*. Lecture Notes in Computer Science, ed. by M Tistarelli, M Nixon. Face age classification on consumer images with gabor feature and fuzzy LDA method, vol. 5558 (Springer, Berlin Heidelberg, 2009), pp. 132–141
78. M Riesenhuber, T Poggio, Hierarchical models of object recognition in cortex. *Nat. Neurosci.* **2**(11), 1019–1025 (1999)
79. T Serre, L Wolf, S Bileschi, M Riesenhuber, T Poggio, Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(3), 411–426 (2007)
80. G Guo, G Mu, Y Fu, TS Huang, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Human age estimation using bio-inspired features (IEEE Computer Society, 445 Hoes Lane, Piscataway, N.J. 08854, U.S.A., 2009), pp. 112–119
81. H-C Lian, B-L Lu, E Takikawa, S Hosoi, ed. by L Wang, K Chen, and Y Ong. *Advances in Natural Computation. Lecture Notes in Computer Science*, vol. 3611 (Springer, Berlin Heidelberg, 2005), pp. 438–441
82. RC Luo, LW Chang, SC Chou, in *39th Annual Conference of the IEEE Industrial Electronics Society, (IECON)*. Human age classification using appearance images for human-robot interaction (IEEE Computer Society, 445 Hoes Lane, Piscataway, N.J. 08854, U.S.A., 2013), pp. 2426–2431
83. F Alnajar, C Shan, T Gevers, J-M Geusebroek, Learning-based encoding with soft assignment for age estimation under unconstrained imaging conditions. *Image Vis. Comput.* **30**(12), 946–953 (2012)
84. H Dibeklioglu, F Alnajar, A Ali Salah, T Gevers, Combining facial dynamics with appearance for age estimation. *IEEE Trans. Image Process.* **24**(6), 1928–1943 (2015)
85. A Hadid, M Pietikainen, Demographic classification from face videos using manifold learning. *Neurocomputing*. **100**(0), 197–205 (2013). Special issue: Behaviours in video
86. BF Klare, MJ Burge, JC Klontz, RW Vorder Bruegge, AK Jain, Face recognition performance: role of demographic information. *IEEE Trans. Inform. Forensics Secur.* **7**(6), 1789–1801 (2012)
87. P Viola, M Jones, Robust real-time face detection. *Int. J. Comput. Vis.* **57**(2), 137–154 (2004)
88. M Castrillón, O Déniz, C Guerra, M Hernández, Encara2: real-time detection of multiple faces at different resolutions in video streams. *J. Vis. Commun. Image Representation*. **18**(2), 130–140 (2007)
89. DG Lowe, Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**(2), 91–110 (2004)
90. C-C Chang, C-J Lin, LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**, 27–12727 (2011)
91. S Knerr, L Personnaz, G Dreyfus, in *Neurocomputing*. NATO ASI Series. Single-layer learning revisited: a stepwise procedure for building and training a neural network, vol. 68 (Springer, Berlin Heidelberg, 1990), pp. 41–50
92. C-W Hsu, C-J Lin, A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Netw.* **13**(2), 415–425 (2002)
93. Morph noncommercial face dataset. www.faceaginggroup.com/morph/
94. PJ Phillips, H Moon, SA Rizvi, PJ Rauss, The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(10), 1090–1104 (2000)
95. C-W Hsu, C-C Chang, C-J Lin, et al. A practical guide to support vector classification (2003). <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com