

# Real-Time Facial Analysis in Still Images and Videos for Gender and Age Estimation

Lionel Prevost<sup>1(✉)</sup>, Philippe Phothisane<sup>2</sup>, and Erwan Bigorgne<sup>2</sup>

<sup>1</sup> University of the French West Indies,  
BP 250, 97157 Pointe à Pitre, Guadeloupe  
lionel.prevost@univ-ag.fr

<sup>2</sup> Eikeo, 11 Rue Léon Jouhaux, 75010 Paris, France  
erwan.bigorgne@eikeo.com

**Abstract.** Research has recently focused on human age and gender estimation because they are useful cues in many applications such as human-machine interaction, soft biometrics and demographic studies.

In this paper, we propose a real time face tracking framework that includes a sequential estimation of people's gender then age. Local binary patterns histograms extracted from facial images. A single gender estimator and several gender-specific age estimators are trained using a boosting scheme. Their decisions are combined to output a gender and an age in years.

The whole process is thoroughly tested on state-of art databases and video sets. Results on the popular FG-NET database are comparable to human perception (overall 70 % correct responses within 5 years tolerance and almost 90 % within 10 years tolerance). The age and gender estimators combined with the face tracker provide real-time estimations at 21 frames per second.

**Keywords:** Face analysis · LBP · Boosting · Gender estimation · Age estimation

## 1 Introduction

Humans can glean a wide variety of information from a face image, including identity, age, gender, and ethnicity. Despite the broad exploration of person identification from face images, there is only a limited amount of research on how to automatically and accurately estimate demographic information contained in face images such as age or gender.

Gender identification is a cognitive process learned and consolidated throughout childhood. It finally becomes mature in teenage years [1] Children can make good guesses but also use many social stereotypes such as facial features, hair type, clothes or interests. Cropped faces without these external features are generally enough for adult operators to classify properly men and women. The goal of an automatic gender estimator is to match adult human accuracy on cropped face image. As social stereotypes are too variable, they are not considered by most existing methods.

Age identification is also learned throughout life experiences. It is easier to guess someone's age when his(her) age range and ethnicity are seen frequently [2]. One's appearance age may be altered by his (her) individual growth pattern, general health,

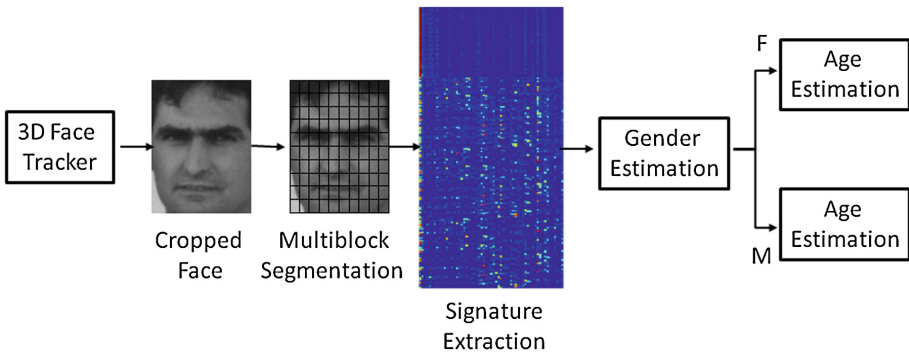
ethnicity, gender, etc. All these parameters should be considered to determine someone's age with accuracy. But most of the time, facial images provide enough information about a subject to be able to estimate his(her) age range.

Automatic age and gender estimators can be used in many different applications such as human-computer interaction, security control, demographic segmentation for marketing studies, etc. Research teams report good performances on databases well spread in the FAR (Facial Analysis and Recognition) community, such as FERET or LFW [3].

In this article, we introduce a real-time algorithm for age and gender estimation. It is implemented in a real-time 3D face tracker derived from the one detailed in [4] and provide age and gender estimation (Fig. 1). The main contributions of the paper are:

- The estimation of gender using one binary boosted classifier based on fast multi-block local binary pattern (MB-LBP) histogram comparisons.
- The estimation of age using several gender specific age classifiers (using the same features) combined with a weighted sum rule to output an age.
- The comparison of human perception vs. automatic estimation of age on two benchmark databases.
- The study of our estimators in videos using 3D face tracking. Such experiments on real data appear rarely in publications focusing on age and gender estimation.

In Sect. 2, previous methods are introduced, providing state of the art performances on age and gender estimation. Section 3 describes the feature extraction process. Section 4 details boosting training and decision process for gender and age estimation. Section 5 presents various experiments on common databases and video sequences to validate our approach. Comparisons with state of the art are provided too. Finally, Sect. 6 concludes and adds some prospects.



**Fig. 1.** Gender then age estimation process flowchart. The apparent blocks are the 64 (8 rows, 8 columns) regions of interest. The signature is the  $256 \times 59$  2-uniform LBP normalized histograms matrix.

## 2 Previous Work

### 2.1 Gender Recognition

Most published methods use face cues for gender recognition. The first attempts of automatic gender estimation started in the early 90's with the SEXNET [5]. This method used a two-layer neural network trained to classify  $30 \times 30$  facial images. Tests were done on 90 images (45 males, 45 females) and obtained a 8.68 % error rate. At the same time, Cottrell and Metcalfe [6] used 160 images  $64 \times 64$  (10 males, 10 females). Images were reduced to 40 components vectors and used to train a single layer neural network. This experiment provided a perfect recognition rate on the training database.

Recently, other methods have also used gait cues to gather more information on targeted subjects [7, 8]. As our human interaction applications are aimed at being used at close range, only faces are visible. The focus here is set on methods using only face images. Many recent papers report results on the FERET database. Moghaddam and Yang [9] achieve an overall 3.38 % error rate using a support vector machine with a RBF kernel on low resolution images. Baluja and Rowley [10] report comparable results using simple pixel comparisons on  $20 \times 20$  face images. This feature extraction process is very interesting because it is not time consuming. Other studies published results on more unconstrained databases, as image sets downloaded from the web. Shan's gender estimator [11] applies SVMs to LBP histograms on 7,443 images of the LFW database, obtaining a 5.19 % error rate. Shakhnarovich et al. [12], Gao and Ai [13], Kumar et al. [14] experimented on non publicly available databases. In [12] authors use an adaBoost on Haar filters outputs applied to  $30 \times 30$  images. Kumar et al. [14] obtain an 8.62 % error rate on a 1,954 images database (1,087 males, 867 females) with SVM comparable to those seen in [9].

Most studies focus on still image databases using  $k$ -fold cross-validation, and few provide cross-database results. Makinen and Raisamo [15] provide a deep comparison of some state-of-art classifiers on gender estimation (Fig. 2). These classifiers are trained on the FERET database and evaluated on a homemade "internet" database. Authors show that the mean accuracy is good (from 80 % to 90 %) and does not vary significantly from one classifier to another. Overall, only a few experiments were conducted on video sequences [16].

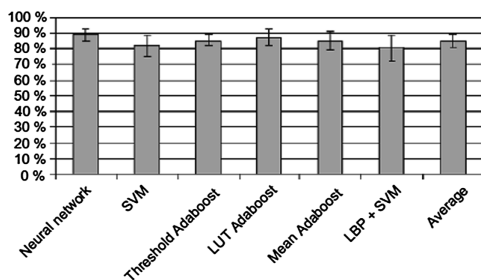


Fig. 2. Comparison of state of art methods for gender classification (reprinted from [15]).

## 2.2 Age Estimation

Estimating an age means to automatically assign an age to the current subject, whether in years or as an age interval. It is the reverse action of age modeling [17]. There appear to be several definitions of “age” described in [18].

- The actual age is the real age of an individual.
- The perceived age is gauged by another person.
- The appearance age, given by the person’s image.
- The estimated age is given by a computer.

Age estimation can be seen as two different problems. The first is a regression problem where the estimator has to predict someone’s age as closely as possible with a year precision. The second aims at classifying a face image into one of several bins. As an example, Gao and Ai [13] use a linear discriminant analysis on Gabor wavelets and classify images into 4 bins (“baby”, “child”, “adult”, “old”). Recently, Guo et al. [19] studied both questions using bio-inspired features (BIF) and achieved a 4.77 years accuracy on the FG-NET database. Thukral et al. [20] report a mean absolute error (MAE) of 6.2 years on the whole FG-NET database using geometric features and relevance vector machines. Luu et al. report a MAE of 4.37 years on a subset of FG-NET using active appearance models and support vector machine regressors [21] and 4.12 years on the complete set with a contourlet appearance model [22]. Others report results on the whole database, using methods such as RUN (Regressor on Uncertain Nonnegative labels) [23, 24]. In [25], Guo et al. report 4.69 years accuracy on the non-publicly available Yamaha Gender and Age (YGA) database. As we can see, many studies report results on the FG-NET database which is publicly available (<http://www.fgnet.rsunit.com>). Most of them use a “Leave-One-Subject-Out” evaluation scheme and their MAE varies between 4 and 6 years.

## 2.3 Discussion

We decide to perform LBP histogram bin comparisons instead of the simplistic pixel comparison proposed in [10] because this feature extraction process is one the fastest in the literature (see Sect. 3). Then, for same reasons, we use an AdaBoost scheme [11, 29] to perform sequential gender then age comparison (see Sect. 5). We will study our method’s behavior and compare it to state of art methods on standard image databases and unconstrained videos (see Sect. 5).

## 3 Multi-scale Block Local Binary Pattern Histograms

First, we perform 2D face detection and 3D face alignment using a real-time face tracker and pose estimator described in [4]. Its precision allows us to track facial features accurately. Eyes coordinates are used to extract a cropped face from the source image and normalize to 128×96 pixels. Pose estimation also provides valuable information and allows to reject images of face too far from a frontal pose. This rejection is only used on our live video tests. For still image databases, no rejection is applied.

Then, we compute Multi-scale Block LBP (MBLBP) and histograms for MBLBP. The following subsections describe this feature extraction process thoroughly.

### 3.1 Uniform Local Binary Patterns

LBP are commonly used local texture descriptors [26]. Their evolutions include Multi-resolution Histograms of Local Variation Patterns [27] and Multi-scale Block LBP (MB-LBP) [28] which inspired our method.

The original (scale-1) LBP operator labels the pixels of an image by thresholding the  $3 \times 3$ -neighborhood of each pixel with the center value  $p_0$  and considering the result as a binary string. For each  $p_i$ ,  $i = \{1, \dots, 8\}$  surrounding  $p_0$  in a circular fashion, the boolean  $b_i$  is defined as follows (Eq. 1).

$$b_i = \begin{cases} 1 & \text{if } p_i > p_0 \\ 0 & \text{if } p_i \leq p_0 \end{cases} \quad (1)$$

Using the 8 bits,  $LBP_{p_0}$  has 256 possible values (Eq. 2). The histogram of the labels can be used as a texture descriptor.

$$LBP_{p_0} = \sum_i 2^{i-1} \times b_i \quad (2)$$

In MB-LBP, the comparison operator between single pixels in LBP is simply replaced with comparison between average gray-values of sub-regions. Each sub-region is a square block containing neighboring pixels (or just one pixel particularly). The whole filter is composed of 9 blocks. We take the size  $k$  of the filter as a parameter, and  $k \times k$  denoting the scale of the MB-LBP operator. For instance, a scale-3 LBP centered on pixel  $p_0$  uses all the pixels in the  $9 \times 9$  region surrounding it. The reference region,  $r_0$  is a  $3 \times 3$  area around  $p_0$ . Each other  $r_i$ ,  $i = \{1, \dots, 8\}$  is another  $3 \times 3$  region encircling  $r_0$  defined as the sum of its pixel values  $p$ . Thus, the  $b_i$  are defined according to the sum of the pixel values  $p$  inside the  $r_i$  regions (Eq. 3).

$$b_i = \begin{cases} 1 & \text{if } \sum_{p \in r_i} p > \sum_{p \in r_0} p \\ 0 & \text{if } \sum_{p \in r_i} p \leq \sum_{p \in r_0} p \end{cases} \quad (3)$$

In our method, scale-1, 3, 5 and 9 LBP are computed before conversion into 2-uniform LBP. The  $k$ -uniform LBP are a subset of the original LBP. The criterion used is the number of circular bit transitions: a  $k$ -uniform LBP has  $k$  or less transitions. For instance, 11001111 and 00000001 have both 2 transitions and thus are 2-uniform LBP. According to [11] and [26], 2-uniform LBP provide the majority of seen patterns. There are 58 possible values of 2-uniform LBP, the remaining values are all set as non-uniform. We use a look-up table which directly transforms LBP values into (58+1) different values. We obtain in the end four  $128 \times 96$  2-uniform LBP maps for each input image (one for each scale).

### 3.2 Block Histograms

In order to add spatial information, in a similar fashion of [11] and [28], we divided the image into blocks of  $26 \times 20$  pixels. Then, we compute histograms of 2-uniform LBP values on 4 scales (1, 3, 5 and 9). Blocks are regularly distributed on 8 rows and 8 columns. In the end, we obtain  $8 \times 8 \times 4$  59-bin histograms. Each 59-bin histogram is normalized to obtain a unit vector. These  $256 \times 59$  matrix signatures are computed on each face image.

## 4 Boosting and Decision

The signatures are used to classify age and gender. Before training, the face databases (described in Sect. 5.1) are labeled, with the actual gender and a perceived age. Each image is mirrored to avoid asymmetrical bias during the learning process. The database thus doubles in size. The weak classifiers  $f(c, j_1, j_2)$ ,  $c = \{1, \dots, 59\}$ ,  $j_n = \{1, \dots, 256\}$ ,  $j_1 \neq j_2$  are simple comparisons of histogram components across blocks. For instance, the  $c^{th}$  histogram bin value  $h(c, j_1)$  block  $j_1$  is compared to every other  $c^{th}$  histogram bin value  $h(c, j_2)$  of block  $j_2$ ,  $j_2 \neq j_1$ .

- If  $h(c, j_1) > h(c, j_2)$ ,  $f(c, j_1, j_2) = 1$
- if  $h(c, j_1) \leq h(c, j_2)$ ,  $f(c, j_1, j_2) = 0$

There are  $C = 59 \times (256 \times 255) / 2 = 1,925,760$  weak classifiers in total. All these  $C$  weak classifiers are used to build our gender and age estimators.

### 4.1 Gender Estimation

Gender identification is a bi-class segmentation and age estimation is a multi-class segmentation. For the gender estimator, a single strong classifier is built up, by sequentially selecting weak classifiers using an AdaBoost training scheme [29]. The gender strong classifier  $S_g$  combines weak classifiers outputs using Eq. (4).

$$S_g = \frac{\sum_{k=1}^K \log\left(\frac{1-e_k}{e_k}\right) \times f(c_k, j_{1k}, j_{2k})}{\sum_{k=1}^K \log\left(\frac{1-e_k}{e_k}\right)} \quad (4)$$

We define  $K$  as the number of training iterations and  $e_k$  as the weighted error on the training database after the  $k^{th}$  iteration. The output of the  $k^{th}$  selected weak classifier is  $f(c_k, j_{1k}, j_{2k})$ . Then, the decision is taken by thresholding  $S_g$ , with  $\delta$  being the decision threshold:

- if  $S_g > 0.5 + \delta$ , subject is a male.
- if  $S_g < 0.5 - \delta$ , subject is a female.

$S_g$  values within the  $[0.5 - \delta, 0.5 + \delta]$  interval are considered neutral. In our real-time implementation, we set  $\delta = 0.01$  and obtain good qualitative results in live demonstrations (see Sect. 5.3).

## 4.2 Age Estimation

According to the results provided by [25], age estimators provide better results after being trained on specific genders. This result is intuitive as male and female facial features are not altered by age in the same way. So, males and females are segregated in the training databases in order to build two gender specific age estimators. We build a complete age estimator by training several strong classifiers  $S_a$ ,  $a \in A = \{10, 15, 20, \dots, 50, 55\}$  which output a real value like the previous gender strong classifier. Each  $S_a$  is constructed using specific image selections and labels. The  $S_a$  learn to classify face images into two classes: those younger than  $a$ , and those older than  $a$ . The output of  $S_a$  is computed by using Eq. 4. Then, all the strong classifiers outputs are used to compute an over-the-ages score  $S_{age}$ . The age decision is made by finding the maximum value and associated age  $k \in \{10, 15, 20, \dots, 50, 55\}$  of  $S_{age}$ :

$$\operatorname{argmax}_k S_{age}(k) = \sum_{a \in A} (S_a - 0.5) \times \frac{1}{1 - e^{k-a}}$$

The sigmoid function is used to normalize the strong classifier outputs and to build a continuous over-the-ages score. Though simplistic, this decision is effective.

## 4.3 Real-Time Video Analysis

On still images databases used for validation, the age and gender estimations are made without any specific threshold. However, for sequential databases and the live application, the distribution of gender and age estimations associated with each target is important. Both age and gender estimators are implemented in our real-time facial analysis system, which provides accurate head pose estimation. The gender estimation is triggered when the face alignment is considered satisfactory enough, according to specific regression score thresholds on each facial landmark. According to the gender estimator's decision, the male or the female age estimator makes the age estimation.

To use the information provided by the face tracking, computed age and gender estimations are collected over the sequence. These estimations are recorded for each tracked target, building two one-dimensional votes distributions. Our observations made us choose to model these distributions with Gaussian mixture models, and to use the E-M algorithm to take decisions. After fitting the age model and the gender model, the most weighted Gaussian bell means are selected as the final decision.

For the video analysis, every frame is considered. The computation times of our C++ implementation were measured on Intel Core i7-2600 hardware: 42+/-1 ms for a MB-LBP matrix and 1.5+/-0.1 ms for 12,000 weak classifiers (age and gender). Added up, these single-threaded processes can be computed at 21 frames per second with one core while the other cores are dedicated to other tasks such as face tracking. The computational load can be lowered by reducing the MB-LBP signature generation frequency, as two consecutive frames are likely to be only slightly different.

## 5 Experiments

Our age and gender estimators are compared to other state of the art methods on still images databases used in the facial analysis community. The FG-NET database is used to test our method for age estimation, and LFW and FERET are used for gender recognition. Other experiments are conducted on video sequences. These databases are described in the next subsection.

### 5.1 Databases

**Gender.** Labeled Faces in the Wild (LFW) and FERET are commonly used database among the facial analysis community, particularly for face and gender recognition. We randomly select a subset of LFW to keep a balanced repartition of males and females. This final selection contains a total of 2,758 images. The FERET image selection contains the 1,696 images extracted from the *fa* and *fb* subsets. We provide results on these databases using 4-fold cross validation.

**Age.** The FG-NET database is an age estimation specific database. It contains 1002 images of 82 different people, with age labels going from 0 to 69 years. The “100” video (“*from 0 to 100 years in 150 s*”) is available on youtube. During this sequence, 101 people from 0 to 100 years old tell their age while facing the camera. The first frame corresponding to each subject was extracted and labeled accordingly. This “0 to 100” database is used to measure age estimation errors from human operators and automatic system.

**Age and Gender.** We built our own age and gender face database by using images collected from the web. The objective was to gather faces with a wide variety of pose, illumination and expression for a large number of people from various origins. It contains for now 5,814 images, including 3,366 males and 2,448 females. Ten human operators labeled these images with the age they perceived. In order to measure accuracy of human age perception, all these operators participated in a dedicated experiment described in Sect. 5.2. A test-only database was also collected in order to have a constant validation set, as the size of our training dataset aims to grow in size. Even though this method is inherently subject to bias, the error margin is measured in two experiments.

The recorded video dataset uses 16 videos of 8 people, including 6 males and 2 females. The face alignment was considered good enough in 2,086 frames. Every subject was asked to look at the camera then look at specific items in order to capture a wide range of face poses. The relative low number of sequences is compensated by the quantity of images (2,086). In order to investigate the estimators’ behavior towards asymmetric facial appearances, each subject was captured in two different illumination conditions: one in ambient lighting and the other with a supplemental lateral light source. The experimental results are provided in the next subsection.



## 5.2 Results on Still Images

For the following experiments, gender is estimated first and according to the estimator's decision, the age estimation uses either the male or the female features selection. Age estimation results are given in term of Mean Absolute Error (MAE). A preliminary study shows that AdaBoost training mainly selects weak classifiers comparing scale-3 ( $\approx 42\%$ ), then scale-5 ( $\approx 25\%$ ) and scale-7 ( $\approx 23\%$ ). On the other hand, scale-1 weak classifiers represent less than  $10\%$  of those selected. This result shows all the interest of using multi-scale LBP for these tasks.

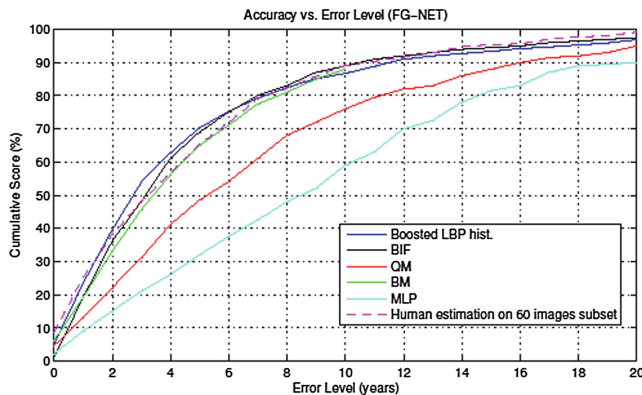
**Gender Estimation.** Experiments on gender estimation are done on the LFW and FERET databases. We proceed to four-fold cross validation tests on both databases separately, obtaining  $90.7\%$  accuracy on our subset of LFW and  $93.4\%$  of correct answers on the FERET database. This kind of experiments does not provide information about how the estimators behave when more than an image is provided per target, which is the case in video streams. This is why we proceeded to experiments using sequential data. They are described in Sect. 5.3 and measure the gender estimator's performance in video sequences.

**Human Age Perception Experiments.** Measuring human errors of age perception would help appreciate automatic age estimation results. Two distinct measurements are conducted. The first is done on a subset of the FG-NET database, and the other on the "0 to 100" dataset. Ten people participated in each experiment. The objective is to measure the accuracy of human age perception and compare it to state of art methods, including ours.

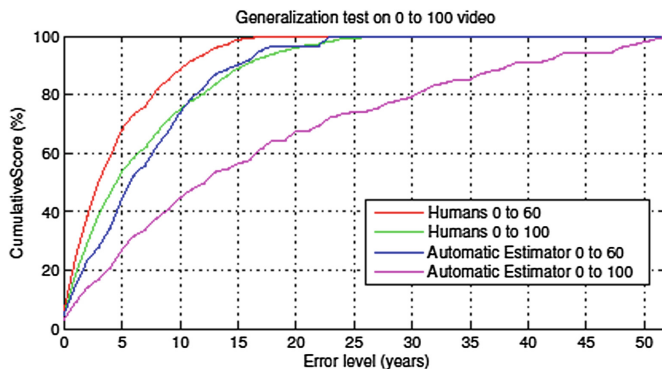
The first experiment uses an age-uniform selection of 60 clear FG-NET pictures of males and females from 0 to 69 years old. Human age perception errors are shown in Fig. 2. The MAE is 4.9 years with a standard deviation of 4.6 years. More than  $65\%$  of the errors are below 5 years and almost  $90\%$  of the errors are below 10 years. These values are close to the performance of the best age estimators published recently and not far from those reported by [30] on the whole FG-NET set (MAE=4.7 years).

For the second experiment, either all images (from 0 to 100 years old) or only a subset (from 0 to 60 years old) are considered. The human performance is similar to the one in the previous test using the FG-NET subset (Fig. 3). This provides information about what kind of accuracy is achievable and reveals that the performance of the best age estimators is actually close to human perception.

**Age Estimation.** To perform fair comparison with existing methods, a leave-one-subject-out (LOSO) scheme was used on the FG-NET database. Each strong age classifier was built with only 200 weak classifiers, as the LOSO scheme costs a lot of time. This comparison was conducted using many other age estimation methods, including [19]'s BIF or [23]'s RUN. As a large part of the database is focused on people younger than 20, we added the 5 and 10 years old strong classifiers to our initial set. The best reported result is obtained by [22] with a MAE of 4.12 years. Other report detailed results on several age subsets as shown in Table 1. We obtain results close to Guo's performance [19] with a MAE value of 4.94 years (0.04 year difference with human perception error). Our results are the best available on the [0-9] and [40+]



**Fig. 3.** Human perception and estimation errors: comparison with state-of-art works (FGNET database).



**Fig. 4.** Human perception errors and estimation errors (“0 to 100” database).

**Table 1.** Age estimation results on FG-NET database. Leave-One-Subject-Out training. Mean absolute error in year on every age subsets. QM and MLP methods are described in [24].

Range	#image	MBLBP	BIF	RUN	QM	MLP
0-9	371	<b>2.42</b>	2.99	2.51	6.26	11.63
10-19	339	3.92	3.39	3.76	5.85	<b>3.33</b>
20-29	144	4.95	<b>4.30</b>	6.38	7.10	8.81
30-39	70	10.56	<b>8.24</b>	12.51	11.56	18.46
40-49	46	<b>14.59</b>	14.98	20.09	14.80	27.98
50-59	15	<b>18.45</b>	20.49	28.07	24.27	49.13
60-69	8	<b>27.62</b>	31.62	42.50	37.38	49.13
0-69	1002	4.94	<b>4.77</b>	5.78	5.57	10.39

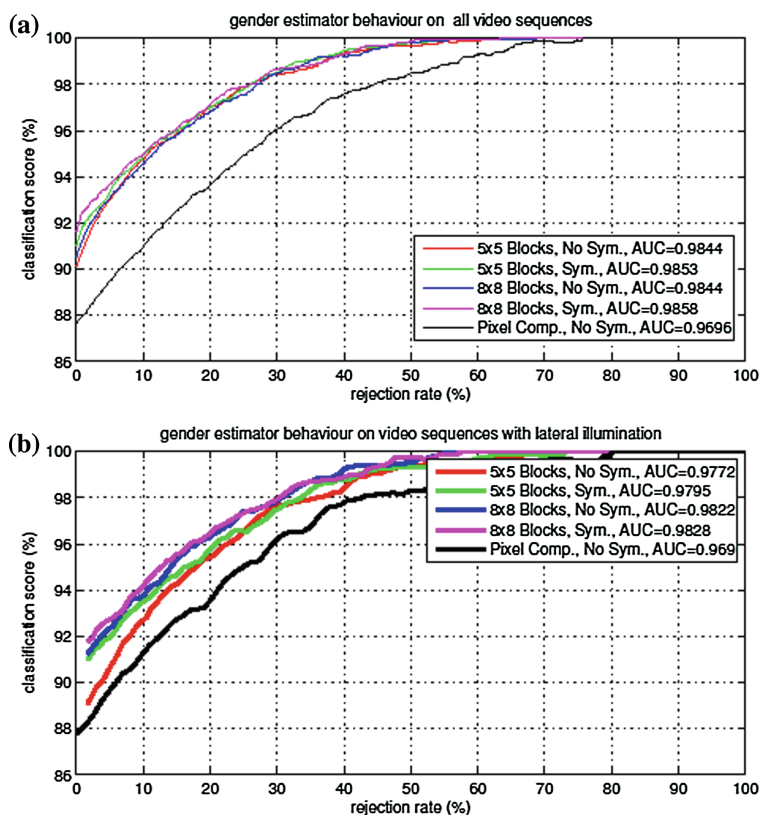
subsets and the second best on the [20-39] subset. Cumulative scores on FG-NET are available in Fig. 2. Table 1 compares mean errors for each age subset. The best age estimation methods (including ours) have results close to human perception on this database. Human perception errors and estimation errors on the “0 to 100” database are presented in Fig. 3. This is pure generalization as our estimators were trained on FG-NET and then tested on this new, different database. Our estimator was not trained for age boundaries above 60 years. It explains its poor performance on this generalization experiment over the full “0 to 100” database and its acceptable results (not far from human operators) on the “0 to 60” subset.

### 5.3 Results on Sequences

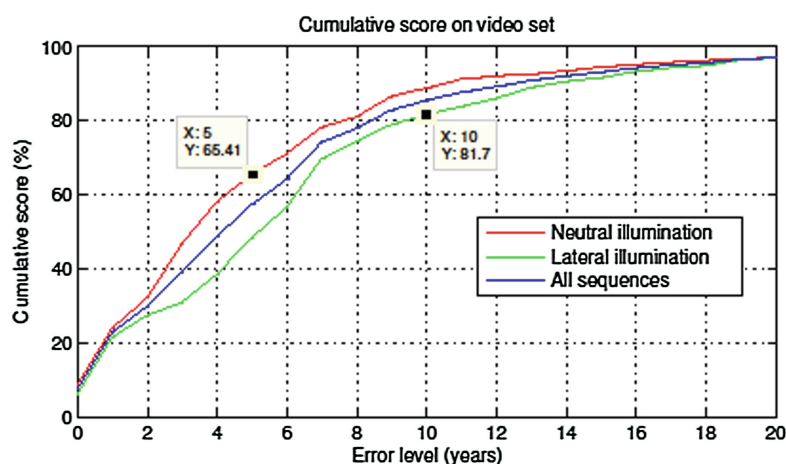
In the following experiments, the complete system is tested on our video dataset. Both gender and age estimators share the same cropped images and MB-LBP histogram signatures to predict their output. Each frame is computed to make a complete observation of the estimation outputs.

**Gender Estimation.** For the video set, our gender estimator outputs distributions of real values between 0 and 1. Using E-M algorithm [31], each video’s most weighted distributions’ mean is on the correct side of the 0.5 threshold. As the estimation score is centered on 0.5, we can study the estimator’s outputs distribution considering several rejection thresholds  $\delta$ . For instance, with  $\delta = 0.01$ , outputs within the [0.49 0.51] interval are discarded. The ROC-like graphs shown in Fig. 5 are plotted using the subsequent good and false response rates. This experiment uses two different block settings to compute the MB-LBP histograms (Fig. 5.a). Both settings use blocks of the same size, one fifth of the original 128×96 crop. The first setting uses 25 (5×5) non-overlapping 26×20 pixels blocks across the face crop and the second uses 64 (8×8) overlapping 26×20 pixels blocks. This setting (8×8 blocks) performs slightly better than the other on the lateral illumination dataset (Fig. 5.b). The use of symmetry refers to using each source frame’s symmetric image, thus using two MB-LBP histogram signatures for each computed frame. As the learning database itself was mirrored, using symmetry does not dramatically improve the results. The pixel comparisons method is an implementation of Baluja’s boosted gender estimator [10], using 30×30 face images. It was used as a baseline for our latest experiments. It seems more robust to illumination. Anyway, whatever the dataset, our method performs better as its classification rate is higher than 99 % even when rejecting one frame out of two (rejection rate≈50 %).

**Age Estimation.** Accordingly to previous results, we choose to use the  $8 \times 8$  blocks settings. Age estimation results over the video set are shown as cumulative scores in Fig. 6: 65.4 % of the estimations are within the 5 years threshold on the neutral illumination set. Despite having mirrored the database, we still obtain slightly better results on this dataset. These results are comparable to the human age perception experiment results on FG-NET or the “0 to 60” dataset seen in Fig. 4.



**Fig. 5.** Gender estimator behavior and comparison with Baluja's estimator [10] for different settings (a) and illumination (b).



**Fig. 6.** Age estimation: cumulative scores on the video set with neutral or with added lateral illumination.



Video: 1 Frame: 0117



Video: 2 Frame: 0728



Video: 1 Frame: 0190



Video: 2 Frame: 0822



Video: 1 Frame: 0352



Video: 2 Frame: 1052



Video: 1 Frame: 0375



Video: 2 Frame: 1068

**Fig. 7.** Results in outdoor conditions: face tracking, pose estimation, gender and age estimation. test video n°1 and test video n°2.

## 6 Conclusions and Perspectives

An alternative method for live stream oriented age and gender estimation is provided. It uses boosted comparisons over uniform LBP histograms based facial signatures. The system provides real-time estimations and is able to track several targets simultaneously. The system's performance is compared favorably to state of the art techniques of age and gender recognition on common databases. Other experimentations are conducted on video sequences and on live streams to show the accuracy of the whole process including face tracking, pose estimation, gender and age estimation (Fig. 7).

Apart from the inevitable database collection needed to improve our training sessions, many perspectives appear. The next natural evolution would be to change the face cropping for face warping using our 3D model to be more resilient to face orientation, instead of only rejecting extreme poses. The present final age and gender decision uses simple logic over the strong classifiers. It would be interesting to build a system able to decide from all the strong classifiers outputs. Another part of this industrial project is the design of a multi-target re-identification system across multiple camera streams.

**Acknowledgements.** We provide our test videos including the ground truth measures on request. Please contact us by mail to receive our data. The authors gratefully acknowledge the contribution of the Agence National de la Recherche (CIFRE N°533/2009).

## References

1. Wild, H.A., Barrett, S.E., Spence, M.J., O'Toole, A.J., Cheng, Y.D., Brooke, J.: Recognition and sex categorization of adults' and children's faces Examining performance in the absence of sex-stereotyped cues. *J. Exp. Child Psychol.* **77**(4), 269–291 (2000). Elsevier
2. Anastasi, J., Rhodes, M.: An own-age bias in face recognition for children and older adults. *Psychon. Bull. Rev.* **12**(6), 1043–1047 (2005). Springer
3. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. University of Massachusetts, Amherst, Technical Report 07–49 (2007)
4. Phothisane, P., Bigorgne, E., Collot, L., Prevost, L.: A robust composite metric for head pose tracking using an accurate face model. In: *IEEE International Conference on Automatic Face and Gesture Recognition (FG 2011)*, pp. 694–699. IEEE (2011)
5. Golomb, B., Lawrence, D., Sejnowski, T.: Sexnet, a neural network identifies sex from human faces. In: *NIPS-3 Proceedings of the 1990 Conference on Advances in Neural Information Processing Systems*, vol. 3, pp. 572–577 (1991)
6. Cottrell, G., Metcalfe, J.: Empath: face, emotion, and gender recognition using holons. In: *NIPS-3 Proceedings of the 1990 Conference on Advances in Neural Information Processing Systems*, vol. 3, pp. 567–571 (1990)
7. Li, X., Maybank, S., Yan, S., Tao, D., Dacheng, T.: Gait components and their application to gender recognition. *IEEE Trans. on SMC-B* **38**(2), 145–155 (2008). IEEE
8. Shan, C., Gong, S., McOwan, P.: Fusing gait and face cues for human gender recognition. *Neuro Comput.* **71**(10–12), 1931–1938 (2008). Elsevier

9. Moghaddam, B., Yang, M.: Learning gender with support faces. *IEEE Trans. on PAMI* **24** (5), 707–711 (2002). IEEE
10. Baluja, S., Rowley, H.: Boosting sex identification performance. *Int. J. Comput. Vis.* **71**, 111–119 (2007). Kluwer Academic Publishers Hingham
11. Shan, C.: Learning local binary patterns for gender classification on real-world face images. *Pattern Recogn. Lett.* **33**(4), 431–437 (2012). Elsevier
12. Shakhnarovich, G., Viola, P., Moghaddam, B.: A unified learning framework for real time face detection and classification. In: *IEEE International Conference on Automatic Face and Gesture Recognition (FG 2002)*, pp. 14–21. IEEE (2002)
13. Gao, F., Ai, H.: Face age classification on consumer images with gabor feature and fuzzy LDA method. In: Tistarelli, M., Nixon, M.S. (eds.) *ICB 2009*. LNCS, vol. 5558, pp. 132–141. Springer, Heidelberg (2009)
14. Kumar, N., Belhumeur, P.N., Nayar, S.K.: FaceTracer: a search engine for large collections of images with faces. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part IV*. LNCS, vol. 5305, pp. 340–353. Springer, Heidelberg (2008)
15. Makinen, E., Raisamo, R.: An experimental comparison of gender classification methods. *Pattern Recogn. Lett.* **29**(10), 1544–1556 (2008). Elsevier
16. Hadid, A., Pietikainen, M.: Combining motion and appearance for gender classification from video sequences. In: *International Conference on Pattern Recognition (ICPR 2008)*, pp. 1–4. IEEE (2008)
17. Ramanathan, N., Chellappa, R.: Face verification across age progression. *IEEE Trans. Image Proces.* **15**(11), 3349–3361 (2006). IEEE
18. Fu, Y., Guo, G., Huang, T.: Age synthesis and estimation via faces: A survey. *IEEE Trans. PAMI* **32**(11), 1955–1976 (2010). IEEE
19. Guo, G., Mu, G., Fu, Y., Huang, T.: Human age estimation using bio inspired features. In: *IEEE Conference on Computer Vision (CVPR 2009)*, pp. 112–119. IEEE (2009)
20. Thukral, P., Mitra, K., Chellappa, R.: A hierarchical approach for human age estimation. In: *IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP 2012)*, pp. 1529–1532. IEEE (2012)
21. Luu, K., Ricanek, K., Bui, T., Suen, C.: Age estimation using active appearance models and support vector machine regression. In: *IEEE Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, pp. 1–5. IEEE (2009)
22. Luu, K., Seshadri, K., Savvides, M., Bui, T.D., Suen, C.: Contourlet appearance model for facial age estimation. In: *IEEE International Joint Conference on Biometrics (IJCB 2011)*, pp. 1–8 (2011)
23. Yan, S., Wang, H., Tang, X., Huang, T.S.: Learning auto-structured regressor from uncertain nonnegative labels. In: *International Conference on Computer Vision (ICCV 2007)*, pp. 1–8 (2007)
24. Lanitis, A., Draganova, C., Christodoulou, C.: Comparing different classifiers for automatic age estimation. *IEEE Trans. on SMC-B* **34**(1), 621–628 (2004)
25. Guo, G., Mu, G., Dyer, D., T.S., H.: A study on automatic age estimation using a large database. In: *International Conference on Computer Vision (ICCV 2009)*, pp. 1986–1991 (2009)
26. Ojala, T., Pietikinen, M., Maenpää, T.: Multi-resolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. PAMI* **24**(7), 971–979 (2002)
27. Zhang, W., Shan, S., Zhang, H., Gao, W., Chen, X.: Multi-resolution histograms of local variation patterns (MHLVP) for robust face recognition. In: Kanade, T., Jain, A., Ratha, N. K. (eds.) *AVBPA 2005*. LNCS, vol. 3546, pp. 937–944. Springer, Heidelberg (2005)

28. Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S.Z.: Learning multi-scale block local binary patterns for face recognition. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 828–837. Springer, Heidelberg (2007)
29. Freund, Y., Schapire, H.: Experiments with a new boosting algorithm. In: International Conference on Machine Learning (ICML 1996), pp. 148–156. Morgan Kaufmann (1996)
30. Han, H., Otto, C., Jain, A.: Age estimation from face images: Human vs. machine performance. In: International Conference on Biometrics (ICB 2013), pp. 4–7. IEEE (2013)
31. Dempster, A.P., Laird, N.M., Rubin, D., et al.: Maximum Likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc. Ser. B (Methodological)* **39**(1), 1–38 (1977)