

Live Stream Oriented Age and Gender Estimation using Boosted LBP Histograms Comparisons

Lionel Prevost¹, Philippe Phothisane² and Erwan Bigorgne²

¹LAMIA, University of the French West Indies and Guiana, Campus de Fouillole, BP 250, 97157 Pointe à Pitre, France

²Eikeo, 11 rue Léon Jouhaux, 75010 Paris, France

Keywords: Face Analysis, Boosting, Gender Estimation, Age Estimation.

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 or demographic statistics for marketing. Even though human perception of other people's age is often biased, attaining this kind of precision with an automatic estimator is still a difficult challenge. In this paper, we propose a real time face tracking framework that includes a sequential estimation of people's gender then age. A single gender estimator and several gender-specific age estimators are trained using a boosting scheme and their decisions are combined to output a gender and an age in years. We choose to train all these estimators using local binary patterns histograms extracted from still facial images. The whole process is thoroughly tested on state-of-art databases and video sets. Results on the popular FG-NET database show results comparable to human perception (overall 70% correct responses within 5 years tolerance and almost 90% within 10 years tolerance). The age and gender estimators can output decisions at 21 frames per second. Combined with the face tracker, they provide real-time estimations of age and gender.

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 (Wild et al., 2000). 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 (Anastasi and Rhodes, 2005). 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 (Huang et al., 2007).

In this article, we introduce our own method for age and gender estimation. They are implemented in a real-time 3D face tracker derived from the one detailed in (Phothisane et al., 2011) and provide age and gender estimation (figure 1). The main contributions of the paper are:

- The estimation of gender using one bi-class boosted classifier based on multi-block local binary pattern (MBLBP) histogram comparisons.
- The estimation of age using several gender

specific age classifiers combined with a weighted sum rule to output an age .

- The comparison of human perception vs. automatic estimation of age on two databases.
- The study of our estimators in videos using 3D face tracking. Such experiments on real data appear rarely in most publications focused on age and gender estimation.

In section 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 is provided too. Finally, section 6 concludes and adds some prospects.

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(Golomb et al., 1991). This method used a two-layer neural network trained to classify 30×30 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 Metclafe(Cottrell and Metclafe, 1990) used $160 \times 64 \times 64$ images (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(Li et al., 2008; Shan et al., 2008). 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 (Moghaddam and Yang, 2002), achieve an overall 3.38% error rate using a support vector machine with a Gaussian kernel on low resolution images. Baluja and Rowley(Baluja and Rowley, 2007) report comparable results using simply pixel comparisons in 20×20 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(Shan, 2012), applies SVMs to LBP histograms on 7,443 images of the LFW database, obtaining a 5.19% error rate. Shakhnarovich et al.(Shakhnarovich et al., 2002), Gao and Ai(Gao and Ai, 2009), Kumar et al.(Kumar et al., 2008) experimented on non publicly available databases. (Shakhnarovich et al., 2002) uses an AdaBoost on Haar filters outputs applied to 30×30 images. Kumar et al.(Kumar et al., 2008) obtain an 8.62% error rate on a 1,954 images database (1,087 males, 867 females) with SVM comparable to those seen in (Moghaddam and Yang, 2002). Most studies focus on still image databases using k-fold cross-validation, and few provide cross-database results. Makinen and Raisamo (Makinen and Raisamo, 2008) provide a deep comparison of some state-of-art classifiers (Neural Network, SVM and various AdaBoost) on gender estimation. These classifiers are trained on the FERET database and evaluated on a homemade "internet" database. Authors show that the mean accuracy is good (80% to 90%) and does not vary significantly from one classifier to another. Overall, only a few experiments were conducted on video sequences(Hadid and M., 2008).

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 (Ramanathan and Chellappa, 2006). There appear to be several definitions of "age" described in (Fu et al., 2010).

- 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(Gao and Ai, 2009) use a linear discriminant analysis on Gabor wavelets and classify images into 4 bins (baby, child, adult, old). Recently, Guo et al.(Guo et al., 2009b) studied both questions using bio-inspired features (BIF) and achieved a 4.77 years accuracy on the FG-NET database. Thukral et al. (Thukral et al., 2012) 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

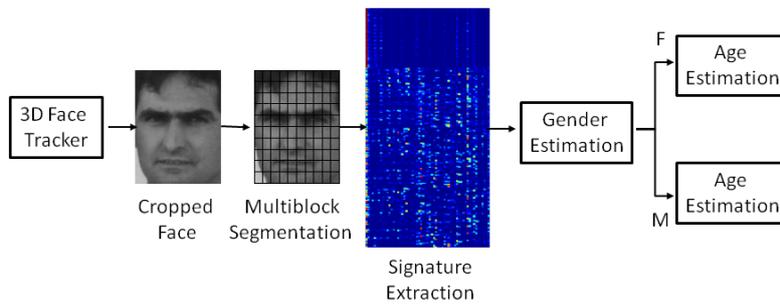


Figure 1: Gender then age estimation process flowchart. The apparent blocks are the 64 (8 rows, 8 columns) regions of interest. The signature is the 256×59 2-uniform LBP normalized histograms matrix.

a subset of FG-NET using active appearance models and support vector machine regressors (Luu et al., 2009) and 4.12 years on the complete set with a contourlet appearance model (Luu et al., 2011). Others report results on the whole database, using methods such as RUN (Regressor on Uncertain Nonnegative labels (Yan et al., 2007)) and (Lanitis et al., 2004). (Guo et al., 2009a) reports a 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.

We decided to perform LBP histogram bin comparisons instead of simplistic pixel comparison and to use an AdaBoost scheme to perform sequential gender then age comparison. We study our method's behavior and compare it to state of art methods on standard databases.

3 MULTI-SCALE BLOCK LOCAL BINARY PATTERN HISTOGRAMS

First, we perform 2D face detection and 3D face alignment using a real real-time face tracker and pose estimator described in (Phothisane et al., 2011). Its precision allows us to track facial features accurately. Eyes coordinates are used to extract a cropped face from the source image and normalize it to 128×96 . 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 process thoroughly.

3.1 Uniform Local Binary Patterns

LBP are commonly used local texture descriptors (Ojala et al., 2002). Their evolutions include Multi-resolution Histograms of Local Variation Patterns (Zhang et al., 2005) and Multi-scale Block LBP (Liao et al., 2007) (MBLBP) which inspired our method.

The original (scale-1) LBP operator labels the pixels of an image by thresholding the 3×3 -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 $b_i = 1$ if $p_i > p_0$ and $b_i = 0$ if $p_i \leq p_0$ is associated. Using the 8 bits, LBP_{p_0} has 256 possible values. The histogram of the labels can be used as a texture descriptor.

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

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 9×9 region surrounding it. The reference region, r_0 is a 3×3 area around p_0 . Each other $r_i, i = \{1, \dots, 8\}$ is another 3×3 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:

$$b_i = \begin{cases} 1 & \text{if } \sum_{p \in r_i} p > \sum_{p \in r_0} p \\ 0 & \text{if } \sum_{p \in r_i} p \leq \sum_{p \in r_0} p \end{cases} \quad (2)$$

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 (Shan, 2012) and (Ojala et al., 2002), 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 128×96 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 (Shan, 2012) and (Liao et al., 2007), we compute histograms of 2-uniform LBP values contained in subwindows of 26×20 pixels over the different layers. The subwindows are regularly distributed into 8 rows and 8 columns. In the end we obtain 256 59-bin histograms. Each 59-bin histogram is normalized to obtain a unit vector. These 59×256 matrix signatures are computed on each face image.

4 BOOSTING AND DECISION

The signatures are used to classify age and gender. Before any learning, the face database (described in section 5.1) is labeled, with 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 $f(c, j_1, j_2)$, $c \in \{1, \dots, 59\}$, $j_n \in \{1, \dots, 256\}$, $j_1 \neq j_2$ are simple comparisons of histogram components across subwindows. For instance, the c^{th} histogram bin value h_{c,j_1} of subwindow j_1 is compared to every other c^{th} histogram bin value h_{c,j_2} , of subwindow j_2 , $j_2 \neq j_1$. There are $C = \frac{(256 \times 255)}{2} \times 59 = 1,925,760$ weak classifiers. The C weak classifiers of each training image are used to build our gender and age estimators.

- if $h_{c,j_1} > h_{c,j_2}$, $f(c, j_1, j_2) = 1$
- if $h_{c,j_1} \leq h_{c,j_2}$, $f(c, j_1, j_2) = 0$

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 upon the selected weak classifiers using AdaBoost training scheme (Freund and Schapire, 1996). The gender strong classifier S_g combines weak classifiers

outputs using the following equation.

$$S_g = \frac{\sum_{k=1}^K \log\left(\frac{1-e_k}{e_k}\right) \times f(c_k, j_{1,k}, j_{2,k})}{\sum_{k=1}^K \log\left(\frac{1-e_k}{e_k}\right)} \quad (3)$$

K being the number of iterations computed by adaboost, and e_k representing the weighted error on the training database after the k^{th} iteration. $f(c_k, j_{1,k}, j_{2,k})$ is the output of the k^{th} selected weak classifier. Then, the decision is taken by thresholding S_g .

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

With δ being the decision threshold. 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.

4.2 Age Estimation

According to the results provided by (Guo et al., 2009a), 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 , $a \in \{10, 15, \dots, 50, 55\}$ 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 equation (3). 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 $k \in \{0, 1, \dots, 59, 60\}$ of $S_{age}(k)$:

$$\arg \max_k S_{age}(k) = \sum_{a \in A} (S_a - .5) \times \frac{1}{1 + e^{k-a}} \quad (4)$$

The sigmoid fonction in equation (4) 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 unidimensional 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 MBLBP 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 MBLBP 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 seconds") 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. The "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 section 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. In order to investigate the estimators's behaviour 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.

Gender Estimation. Experiments on gender estimation were done on the LFW and FERET databases. Four-fold cross validation tests were proceeded on both dababases separately, obtaining 90.7% accuracy on our subset of LFW and 93.4% of correct answers on the FERET database. The experiments described in 5.3 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 current automatic 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 figure 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 (Han et al., 2013) 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 (figure 3). This provides information about what kind of accuracy is achievable and reveals that the performance of the best age estimators are 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 (Guo et al., 2009b)’s BIF or (Yan et al., 2007)’s RUN. The best reported result is obtained by (Luu et al., 2011) with a mean absolute error (MAE) value of 4.12 years. Other reports detailed results on several age subsets as shown in table 1. We obtain results close to Guo’s performance (Guo et al., 2009b) with an 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+] subsets and the second best on the [20-39] subset. Cumulative scores on FG-NET are available in figure 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 figure 3. This is pure generalization as our estimators were trained on FG-NET and then tested on this database. Our estimator was not trained for age boundaries above 60 years. It explains its poor performance on this gener-

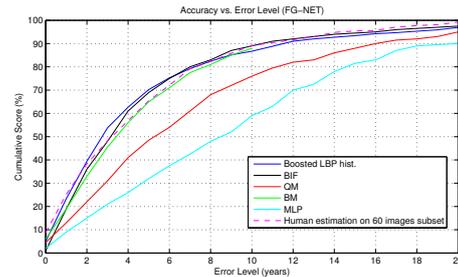


Figure 2: Human perception and estimation errors: comparison with state-of-art works (FGNET database).

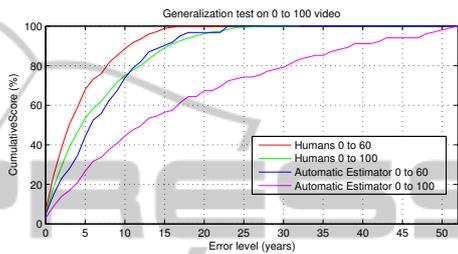


Figure 3: Human perception and estimation errors (“0 to 100” database).

alization 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. The cropped images and LBP histogram signatures used by both gender and age estimators are the same. 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, 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 δ . For instance, with $\delta = 0.01$, outputs within the [0.49 0.51] interval are discarded. The ROC-like graphs shown in figure 4 are plotted using the subsequent good and false response rates. This experiment uses two settings of regions of interest for the LBP histograms (figure 4.a). Both settings use regions of interest of the same size, $\frac{1}{5}$ of the original 128×96 crop. The first setting uses 25 (5×5) non-overlapping 26×20 pixels regions across the face crop and the second uses 64 (8×8) 26×20 pixels regions. The 8×8 blocks setting performs slightly better on the lateral illumination dataset (figure 4.b). The use of symme-

Table 1: Age estimation: results on the FG-NET database, using LOSO scheme test. Mean absolute error in years, on every age subsets. QM and MLP methods are described in (Lanitis et al., 2004).

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

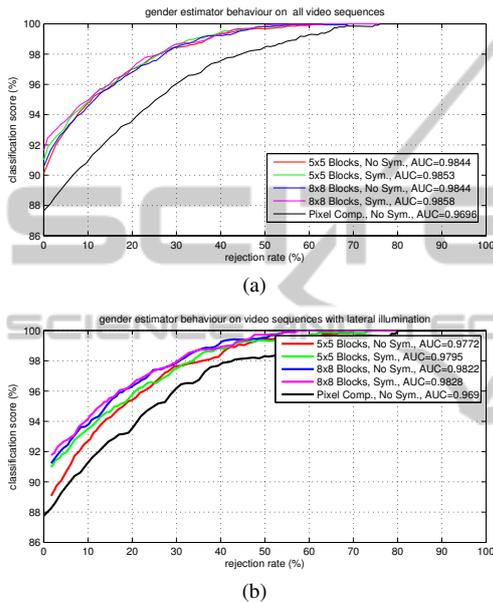


Figure 4: Gender estimator behavior and comparison with Baluja's estimator (Baluja and Rowley, 2007) for different test sets (a) and illuminations (b).

try refers to using each source frame's symmetric image, thus using two MBLBP 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, using 30×30 face images. It was used as a baseline for our latest experiments.

Age Estimation. Accordingly to the previous results, we choose to use the (8×8) blocks settings. Age estimation results over the video set are shown as cumulative scores in figure 5: 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 figure 3.

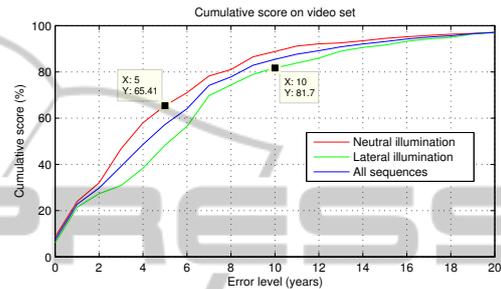


Figure 5: Age estimation: cumulative scores on the video set with neutral or with added lateral illumination.

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 (figures 6 and 7: tracking, pose, gender and age estimation).

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.

ACKNOWLEDGEMENTS

We provide our test videos including the ground truth



Figure 6: Results in outdoor conditions: face tracking, pose estimation, age and gender estimation (test video n°1).

measures on request. Please contact us by mail to receive our data. The authors gratefully acknowledge the contribution of the Agence Nationale de la Recherche (CIFRE N533/2009).



Figure 7: Results in outdoor conditions: face tracking, pose estimation, age and gender estimation (test video n°2).

REFERENCES

- Anastasi, J. and Rhodes, M. (2005). An own-age bias in face recognition for children and older adults. *Psy. Bul. & Rev.*, 12(6):1043–1047.
- Baluja, S. and Rowley, H. (2007). Boosting sex identification performance. *IJCV*, 71:111–119.

- Cottrell, G. and Metcalfe, J. (1990). Empath: face, emotion, and gender recognition using holons. *Proc. of Adv. in NIPS*, 3:567–571.
- Freund, Y. and Schapire, H. (1996). Experiments with a new boosting algorithm. *In Int. Conf. on Machine Learning*, pages 148–156.
- Fu, Y., Guo, G., and Huang, T. (2010). Age synthesis and estimation via faces: A survey. *IEEE Trans. on PAMI*, 32(11):1955–1976.
- Gao, F. and Ai, H. (2009). Face age classification on consumer images with gabor feature and fuzzy lda method. *LNCS*, 5558:132–141.
- Golomb, B., Lawrence, D., and Sejnowski, T. (1991). Sexnet, a neural network identifies sex from human faces. *NIPS*, 3.
- Guo, G., Mu, G., Dyer, D., and T.S., H. (2009a). A study on automatic age estimation using a large database. *ICCV*, pages 1986–1991.
- Guo, G., Mu, G., Fu, Y., and Huang, T. (2009b). Human age estimation using bio inspired features. *CVPR*, pages 112–119.
- Hadid, A. and M., P. (2008). Combining motion and appearance for gender classification from video sequences. *ICPR*, pages 1–4.
- Han, H., Otto, C., and Jain, A. (2013). Age estimation from face images: Human vs. machine performance. *In Proc. ICB*, pages 4–7.
- Huang, G., Ramesh, M., Berg, T., and Learned-Miller, E. (2007). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. *Uni. of Massachusetts, Tech. Report 07-49*.
- Kumar, N., Belhumeur, P., and Nayar, S. (2008). Facetracer: A search engine for large collections of images with faces. *ECCV*, pages 340–353.
- Lanitis, A., Draganova, C., and Christodoulou, C. (2004). Comparing different classifiers for automatic age estimation. *IEEE Trans. on SMC-B*, 34(1):621–628.
- Li, X., Maybank, S., Yan, S., Tao, D., and D., X. (2008). Gait components and their application to gender recognition. *IEEE Trans. on SMC-C*, 38(2):145–155.
- Liao, S., Zhu, X., Lei, Z., Zhang, L., and Li, S. (2007). Learning multi-scale block local binary patterns for face recognition. *ICB*, pages 828–837.
- Luu, K., Ricanek, K., Bui, T., and Suen, C. (2009). Age estimation using active appearance models and support vector machine regression. *the IEEE Conf. on Biometrics: Theory, Applications, and Systems (BTAS)*, pages 1–5.
- Luu, K., Seshadri, K., Savvides, M., and Bui, T.D. and Suen, C. (2011). Contourlet appearance model for facial age estimation. *Int. Joint Conf. on Biometrics (IJCB)*, pages 1–8.
- Makinen, E. and Raisamo, R. (2008). An experimental comparison of gender classification methods. *Pattern Recognition Letters*, 29(10):1544–1556.
- Moghaddam, B. and Yang, M. (2002). Learning gender with support faces. *IEEE trans. on PAMI*, 24(5):707–711.
- Ojala, T., Pietikinen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. on PAMI*, 24:7:971–9.
- Phothisane, P., Bigorgne, E., Collot, L., and Prevost, L. (2011). A robust composite metric for head pose tracking using an accurate face model. *In Proc. FG*, pages 694–699.
- Ramanathan, N. and Chellappa, R. (2006). Face verification across age progression. *IEEE Trans. on Image Processing*, 15(11):3349–3361.
- Shakhnarovich, G., Viola, P., and Moghaddam, B. (2002). A unified learning framework for real time face detection and classification. *FG*, pages 14–21.
- Shan, C. (2012). Learning local binary patterns for gender classification on real-world face images. *Pattern Recognition Letters*, 33(4):431–437.
- Shan, C., Gong, S., and McOwan, P. (2008). Fusing gait and face cues for human gender recognition. *NVR*, 71(10-12):1931–1938.
- Thukral, P., Mitra, K., and Chellappa, R. (2012). A hierarchical approach for human age estimation. *ICASSP*, pages 1529–1532.
- Wild, H., Barrett, S., Spence, M. J., O’Toole, A., Cheng, Y., and Brooke, J. (2000). Recognition and sex categorization of adults’ and children’s faces: Examining performance in the absence of sex-stereotyped cues. *Jour. of Exp. Child Psychology*, 77:269–291.
- Yan, S., Wang, H., Tang, X., and T.S., H. (2007). Learning auto-structured regressor from uncertain nonnegative labels. *ICCV*, pages 1–8.
- Zhang, W., Shan, S., Zhang, H., Gao, W., and Chen, X. (2005). Multi-resolution histograms of local variation patterns (mhlvp) for robust face recognition. *Audio- and Video-Based Biometric Person Authentication*, pages 7:937–944.