Pyramid Multi-Level Features for Facial Demographic Estimation

Salah Eddine Bekhouche [1,4,*]

[1]Laboratory of LESIA, University of Biskra, Algeria
* *Corresponding author:*
EMAIL: salah@bekhouche.com TEL: +213661705459

Abdelkrim Ouafi [1]

[1]Laboratory of LESIA, University of Biskra, Algeria
EMAIL: a.ouafi@univ-biskra.dz

Fadi Dornaika [2,3]

[2] University of the Basque Country, Spain
[3] IKERBASQUE, Basque Foundation for Science, Spain
EMAIL: *fadi.dornaika@ehu.es* TEL: 0034 943018034

Abdelmalik Taleb-Ahmed [4]

[4]Laboratory of LAMIH, UMR CNRS 8201 UVHC
University of Valenciennes, France
EMAIL: taleb@univ-valenciennes.fr

Abdenour Hadid [5]

[5]Center for Machine Vision Research, University of Oulu, Finland
EMAIL: hadid@ee.oulu.fi

Pyramid Multi-Level Features for Facial Demographic Estimation

SE. Bekhouche [1,4,*], A. Ouafi [1], F.Dornaika [2,3], A. Taleb-Ahmed [4] and A. Hadid[5]

[1] Laboratory of LESIA, University of Biskra, Algeria

[2] University of the Basque Country, Spain

[3] IKERBASQUE, Basque Foundation for Science, Spain

[4] Laboratory of LAMIH, UMR CNRS 8201 UVHC, University of Valenciennes, France

[5] Center for Machine Vision Research, University of Oulu, Finland

Abstract

We present a novel learning system for human demographic estimation in which the ethnicity, gender and age attributes are estimated from facial images. The proposed approach consists of the following three main stages: 1) face alignment and preprocessing; 2) constructing a Pyramid Multi-Level face representation from which the local features are extracted from the blocks of the whole pyramid; 3) feeding the obtained features to an hierarchical estimator having three layers. Due to the fact that ethnicity is by far the easiest attribute to estimate, the adopted hierarchy is as follows. The first layer predicts ethnicity of the input face. Based on that prediction, the second layer estimates the gender using the corresponding gender classifier. Based on the predicted ethnicity and gender, the age is finally estimated using the corresponding regressor.

Experiments are conducted on five public databases (MORPH II, PAL, IoG, LFW and FERET) and another two challenge databases (Apparent age; Smile and Gender) of the 2016 ChaLearn Looking at People and Faces of the World Challenge and Workshop. These experiments show stable and good results. We present many comparisons against state-of-the-art methods. We also provide a study about cross-database evaluation. We quantitatively measure the performance drop in age estimation and in gender classification when the ethnicity attribute is misclassified.

Keywords : Demographic estimation, Classification, Age, Gender, Ethnicity.

1 Introduction

The human face is crucial for the identity of persons, because it contains much information about personal characteristics. Thus, the face image is important for most biometrics systems. The face image provides lots of useful information, including the person's identity, gender, ethnicity, age, emotional expression, etc. Recently, several applications that exploit demographic attributes have emerged. These applications include access control (Jain, Dass, & Nandakumar, 2004), re-identification in surveillance videos (Park & Jain, 2010), integrity of face images in social media (Correa, Hinsley, & de Ziga, 2010), intelligent advertising, and human-computer interaction, law enforcement (Kumar, Belhumeur, & Nayar, 2008).

In this paper, we propose an automatic facial demographic estimation system that is composed of three parts: face alignment, feature extraction/selection and demographic attributes estimation.

The purpose of face alignment is to localize faces in images, rectify the 2D or 3D pose of each face and crop the region of interest. This preprocessing stage is important since the subsequent stages depend on it and since it can affect the final performance of the system. The processing stage can be challenging since it should overcome many variations that may appear in the face image.

Feature extraction and selection stage extracts the face features. These features are extracted either by holistic method or by a local method. The extracted features are then selected using a supervised feature selection method in order to omit possible irrelevant features.

The last stage which is called the demographic estimation where we firstly classify the ethnicity, the gender then we estimate the age.

Our main contributions include:

- The use of specific hierarchical order for demographic estimation which leads to good performances as a hierarchical estimator.
- The advantages of Pyramid Multi-Level face representation which gives most of distinctive features.

The rest of the paper is organized as follows: In section 2 we summarize the existing techniques of facial demographic estimation. We introduce our approach in section 3. The experimental results are given in section 4. In section 5 we present the conclusion and some perspectives.

2 Related Works

Although there is growing interest in facial demographic classification, most of the previous researches focused on a single demographic attribute. There are few researches on all demographic attributes (age, gender and ethnicity) such as (Yang & Ai, 2007a; Hadid & Pietikinen, 2013; Han & Jain, 2014; Guo & Mu, 2014; Han, Otto, Liu, & Jain, 2015). We summarize these works as well as their performances in Table 1.

From a general perspective, the demographic estimation approaches can be categorized based on the face image representation and the estimation algorithm. We

Table 1: A comparison of demographic estimation approaches.

Approach	Strength	Weakness	Database	Performance Age (MAE)	Age group (%)	Gender (%)	Ethnicity (
LBP + Real AdaBoost[a]	❏Suited for real-time applications	❏Poor performance	FERET	-	92.1%	93.3%	-
			PIE	-	87.5%	91.1%	93.2%
Manifold Learning[b]	❏Ability to work on different face analysis tasks	❏High computational cost	Private	-	83.1%	96.8%	100.0%
BIF+rKCCA+SVM[c]	❏Good performance	❏High computational cost	MORPH II	3.9	-	98.5%	99.0%
BIF+SVM[d]	❏Achieving good and stable results in different databases ❏Suited for real-time applications	❏Commercial software for face and eye detection	FG-NET	3.8	-	-	-
			MORPH II	3.6	-	97.6%	99.1%
			PCSO	4.1	-	97.1%	98.7%
			LFW	7.8	-	96.8%	-
			FERET	-		94.0%	90.0%
Ours	❏Ability to work on different face analysis tasks ❏Achieving good and stable results in different databases ❏Suited for real-time applications	❏Vulnerable to profile faces and wild poses.	MORPH II	3.5	-	98.9%	99.3%
			PAL	5.0	-	98.2%	99.3%
			IoG	-	68.2%	90.0%	-
			LFW	-	-	98.3%	-
			FERET	-	-	99.0%	95.6%
			ChaLearn2016	0.29[e]	-	90.0%	-

[a](Yang & Ai, 2007b)
[b](Hadid & Pietikinen, 2013)
[c](Guo & Mu, 2014)
[d](Han et al., 2015)
[e]Evaluation using epsilon (Escalera et al., 2016)

will give a brief review of the existing approaches by grouping them based on face image representation similar to (Han et al., 2015).

The anthropometry-based approach mainly depends on measurements and distances of different facial landmarks. Kwon and Lobo (Kwon & da Vitoria Lobo, 1994) published the earliest paper on age classification based on facial images by computing distance ratios to distinguish babies from others. Gilani *et al.* (Gilani, Shafait, & Mian, 2013) extracted 3D Euclidean and Geodesic distances between biologically significant facial landmarks then determined the relative importance of these landmarks by minimal Redundancy Maximal Relevance (mRMR) algorithm. To classify gender they used Linear Discriminant Analysis (LDA). In (Gunay & Nabiyev, 2007), the authors designed a neural network to locate facial features and calculate several geometric ratios and differences which are used for estimating the age from facial image. The anthropometry-based approaches might be useful for babies, children, and young adults, but this approach is impractical for adults. For the adults, the skin appearance is the main source of information about ethnicity, gender, and age.

The appearance-based approach like in (Lanitis, Taylor, & Cootes, 2002; Geng, Zhou, Zhang, Li, & Dai, 2006; Geng, Zhou, & Smith-Miles, 2007; Fu & Huang, 2008; Luu, Seshadri, Savvides, Bui, & Suen, 2011) used shape and texture facial features. The statistical model of face shape and texture which is called Active Appearance Model (AAM) is the most used in this sort of approaches.

Due to their significant performance improvement in facial recognition domain, deep learning approaches have been recently proposed for facial demographic estimation (e.g. (Huerta, Fernandez, Segura, Hernando, & Prati, 2015; Levi & Hassncer, 2015; Mansanet, Albiol, & Paredes, 2016; Ranjan et al., 2015)). Deep learning approaches claim to have the best performances in demographic classification. However, this claim cannot be always true.

Image-based approach is one of the most popular approaches for facial demographic estimation since a face image can be viewed as texture pattern. Many texture features have been used like Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG), Biologically Inspired Features (BIF), Binarized Statistical Image Features (BSIF) and Local Phase Quantization (LPQ) in demographic estimation works.

In (Yang & Ai, 2007a), the authors used LBP histogram as features. The steep decent method was applied to find an optimal reference template. Given a local patch, the Chi-square distance between a sample and the optimal reference template is used as a measure of confidence belonging to the reference class. This was applied for all demographic attributes (age, gender and ethnicity). For their experiments, the authors considered the FERET, PIE, and a snapshot databases. The LBP was also investigated in (Hadid & Pietikinen, 2013) for automatic demographic classification from human faces which includes age categorization, gender recognition and ethnicity classification. They focused on the LBP based spatiotemporal method as a baseline system for combining spatial and temporal information. Furthermore, the correlation between the face images through manifold learning has been exploited. LBP and its variants were also used by many works like in (Shan, 2010; Ylioinas, Hadid, & Pietikainen, 2012; Fazl-Ersi, Mousa-Pasandi, Laganire, & Awad, 2014; Eidinger, Enbar, & Hassner, 2014; Patel, Maheshwari, & Balasubramanian, 2015; Gunay & Nabiyev, 2008).

BIF and its variants are widely used in age estimation works such us (Mu, Guo, Fu, & Huang, 2009; Han & Jain, 2014; Guo & Mu, 2011; Chang, Chen, & Hung, 2011; Geng, Yin, & Zhou, 2013; Guo & Mu, 2013, 2014; Han et al., 2015). Guo and Mu (Guo & Mu, 2014) presented a framework that can estimate demographic attributes (including age, gender and ethnicity) jointly. They extracted BIF features which are projected by either Canonical Correlation Analysis (CCA), Partial Least Squares (PLS) or theirs variants. For the stage of estimation, they use SVM and SVR. They conducted their experiments on MORPH II database. Han *et al.* (Han et al., 2015) used selected BIF features to estimate the age, gender and ethnicity attributes. To further improve the performance they designed a quality assessment method to detect low-quality face images. They evaluate their approach on different databases (FG-NET, FERRET, MORPH II, PCSO and LFW), the performance was well compared to most demographic estimation approaches.

Some researches used multi-modal features like our approach. For instance, Fazl-Ersi *et al.* (Fazl-Ersi et al., 2014) proposed a framework for gender and age classification, where they combined LBP, Scale-Invariant Feature Transform (SIFT) and Color Histogram (CH) features. As well, Eidinger *et al.* (Eidinger et al., 2014) mixed LBP with one of its variants which is Four Patch (FPLBP) to do a facial estimation of two attributes (age and gender).

3 Proposed approach

We describe below our proposed approach for automatic demographic estimation using facial images. The proposed approach consists of three stages: (i) Face alignment; (ii) feature extraction and selection; (iii) demographic estimation. Figure 1 illustrates the general structure of our approach. In our approach, we take a different way to represent the face by using the proposed PML which preceded the image texture descriptor. Also, the new structure of the hierarchical demographic estimator which relies mainly on the results of ethnicity to obtain the gender information, and the two previous attributes (ethnicity and gender) to obtain the person's age.

This order was proved experimentally to give better results.

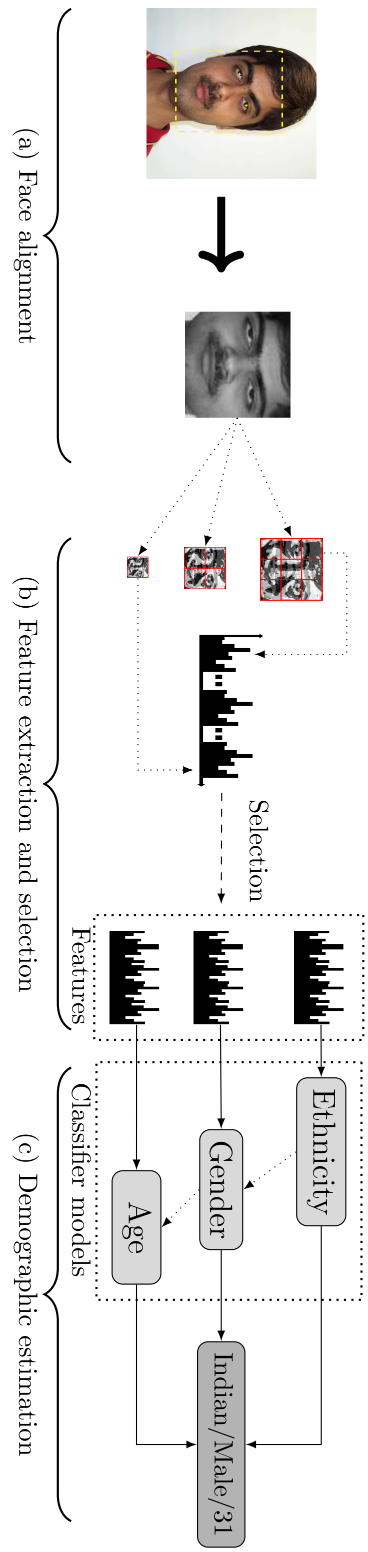


Figure 1: General structure of the proposed approach.

3.1 Face alignment

Face alignment is one of the most important stages in facial soft biometrics systems. The face preprocessing stage consists of three steps : (i) face and facial points detection, (ii) pose correction, and (iii) face region cropping. Figure 2 shows the face alignment stage.

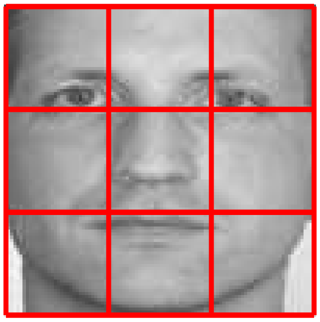
Figure 2: Face alignment stage.

The input image is converted into a gray-scale image, after that we apply the cascade object detector that uses the Viola-Jones algorithm (Viola & Jones, 2001) to detect people's faces. The eyes of each detected face are detected using Ensemble of Regression Trees (ERT) algorithm (Kazemi & Sullivan, 2014) which is a good and very efficient algorithm for facial landmarks localization. The locations of the two eyes are used for the face pose rectification step. In order to correct the face pose, we apply a 2D similarity transform on the original face image. The rotation angle and the scale are computed from the locations of the two eyes. The adopted rotation makes the eyes horizontal. The global scale is used for normalizing the interocular distance to a fixed value d. To crop the region of interest (aligned face), a bounding box is centered on the new eyes locations and then stretched to the left and to the right by $d.k_{side}$, to top by $d.k_{top}$ and to bottom by $d.k_{bottom}$. Finally, in our case, k_{side}, k_{top}, k_{bottom} and d are chosen such that the final face image has a size of 100×100 pixels (S. Bekhouche, Ouafi, Taleb-Ahmed, Hadid, & Benlamoudi, 2014).

3.2 Feature extraction and selection

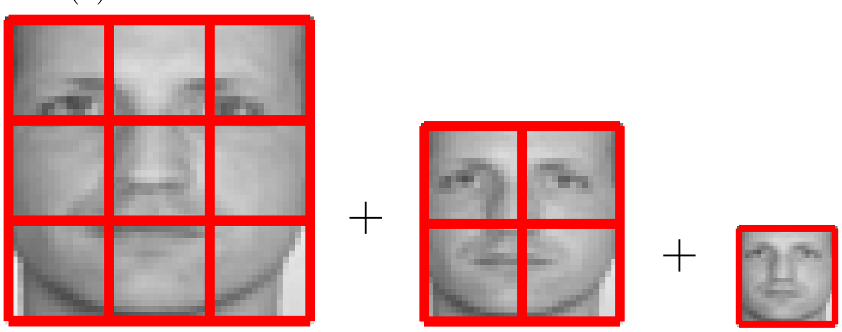
The most common face representation in computer vision is a regular grid of fixed size regions which we call it Multi-Blocks (MB) representation. MB face representation divides the image into n^2 blocks where n is the intended level of MB. Recently, a similar representation called Multi-Levels representation used in the age estimation and gender classification topics (Nguyen, Cho, & Park, 2014; S. E. Bekhouche, Ouafi, Benlamoudi, Taleb-Ahmed, & Hadid, 2015). ML face representation is a spatial pyramid representation which constructed by sorted series of Multi-Block (MB) representations. The ML face representation level n is constructed from level 1, 2 , ... n MB face representations. Inspired by ML and Pyramid-Based Multi-Scale (Wang, Chen, & Xu, 2011) representations, we present the Pyramid Multi-Level (PML) representation of the face image in order to extract the local texture features using different descriptors. Unlike, the ML representation, the PML representation adopts an explicit pyramid representation of the original image. This pyramid represents the image at different scales. For each such level or scale, a corresponding Multi-block representation is used. PML sub-blocks have the same size which determined by the image size and the chosen level. Figure 3 illustrates the difference between these face representations.

The main idea of the PML is to extract features from different division of each pyramid level. In other words, given n, w, h as PML level, original image width and original image height respectively. We assume that the original face ROI corresponds to level n. The pyramid levels (n, $n-1$, $n-2$, ..., 1) are built as follows. The image of level $n-1$ is obtained from image at level n by resizing it to ($w' = w \times \frac{(n-1)}{n}, h' =$



(a) `Multi-Block`

(b) `Multi-Level`



(c) `Pyramid Multi-Level`

Figure 3: Different face representations.

$h \times \frac{(n-1)}{n}$). At each level n, the obtained image is divided into n^2 blocks. Figure 4 illustrates the PML principle for a pyramid of four levels. This figure corresponds to the Local Phase Quantization (LPQ) descriptor.

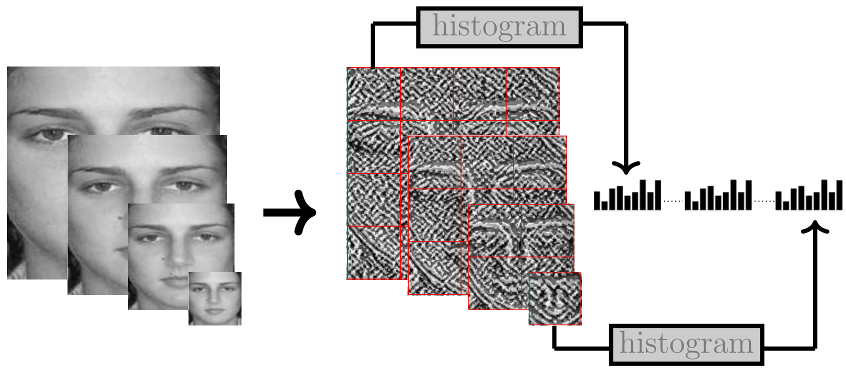


Figure 4: Pyramid Multi-Level Local Phase Quantization example of 4 levels.

In our work, we used two types of texture descriptors LPQ and BSIF.

- **Local Phase Quantization (LPQ)**: is a descriptor for texture classification which utilizes phase information computed locally in a window for every image position. The phases of the four low-frequency coefficients are decorrelated and uniformly quantized in an eight-dimensional space. A histogram of the resulting code words is created and used as features (Ojansivu & Heikkilä, 2008).

- **Binarized Statistical Image Features (BSIF)**: is a local image descriptor which computes a binary code for each pixel by linearly projecting local image patches onto a subspace, whose basis vectors are learnt from natural images via independent component analysis, and by binarizing the coordinates in this basis via thresholding (Kannala & Rahtu, 2012).

3.2.1 Feature Selection

For feature selection, we used a linear discriminant approach based on Fisher's score, which quantifies the discriminating power of features. This score is given by:

$$W_i = \frac{\sum_{j=1}^{C} N_j.(m_j - \bar{m})^2}{\sum_{j=1}^{C} N_j.\sigma_j^2} \tag{1}$$

where W_i is the score of feature i, C is the number of classes, N_j is the number of samples in class j, $\bar{m}$ is the feature mean, m_j and σ_j^2 are the mean and the variance of the class j in intended feature. The features are sorted according to the value of their scores, i.e., the sorting is carried out by descending order. We select the first 5% of the features for each demographic attribute since empirically this was found to be as a good choice for the trade-off efficiency-accuracy. This choice was made after we conducted the following experiment on PAL database using five-fold cross-validation. In this experiment, only the first four folds of the PAL database are used for training and the last fold for test. Table 2 summarizes the obtained results that justify our choice. In this table, we show the accuracy of estimating the three demographic attributes (age, gender, and ethnicity) as a function of features ratio (after Fisher score-based ranking). The features correspond to a 7 level PML (BSIF+LPQ) which has 280 blocks, each with 256 bins. Thus, we have 71680 features in total. In all experiments, the adopted PML has 7 levels.

The last row of the table illustrates the CPU time (in seconds) needed for training the model. This CPU time corresponds to the feature selection process and to the training of the demographic attribute classifiers. As can be seen, the choice of 5% can be considered as a good trade-off between accuracy and computational complexity. Indeed, if only 5% of the features are used the Mean Age Error (MAE) increases

0.3 years, the classification accuracy of gender drops by 1%, and that of ethnicity by 0.2%. At the same time, the CPU time of the training phase decreased from 189 s to 77 s.

Table 2: Accuracy of demographic estimation as a function of different features ratios. The last row depicts the CPU time (in seconds) of the training phase as a function of different features ratios.

Features ratio (%)	1%	5%	10%	20%	50%	60%	75%	100%
Age (MAE in years)	6.5	5.0	5.0	5.0	4.8	4.8	4.7	4.7
Gender (%)	93.7%	98.2%	98.5%	98.9%	99.0%	99.0%	99.1%	99.2%
Ethnicity (%)	95.0%	99.3%	99.4%	99.4%	99.5%	99.5%	99.5%	99.5%
CPU time (s)	49	77	91	108	134	142	165	189

Table 3 depicts the CPU time of the testing phase associated with one image. This CPU time corresponds to the estimation part. Besides the results associated with the training phase, the test results as well justify our choice of 5% of the selected features. Indeed, the CPU time of 7 ms per image enables the system to be used in real-time applications. Our test was done on a laptop with 4 core Intel Xeon Processor E3-1535M v5 (8M Cache, 2.90 GHz) and 64GB RAM.

Table 3: CPU time associated with the testing phase as a function of different features ratios.

Features ratio (%)	1%	5%	10%	20%	50%	60%	75%	100%
CPU time (ms)	3	7	14	23	49	58	71	92

We point out that Fisher Score is computed independently for each demographic attribute. The finally selected features are 5% for each demographic attribute. In the case of age, each class is an interval of 5 years without overlapping. We found that the separate use of selected subset of features (by the demographic attribute estimator) was found to be more accurate that using all selected subsets of features.

We apply a Min-Max scaling method for feature normalization on the subset of the training using the selected features. This method rescales the range of features between [0, 1]. The general formula of the Min-Max scaling is given by:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{2}$$

3.3 Demographic estimation

Gender and ethnicity estimation are classification problems requiring SVM classifier for each one. On the other hand, the age estimation task can be formulated as a regression or classification problem. We treat firstly ethnicity, gender then estimate the age based on this information. We choose ethnicity as root based on the fact that, according to our experiments, ethnicity attribute estimation has the highest accuracy among all three demographic attributes. In the case of the absence of ethnicity, gender will become the root of the demographic estimator. Figure 5 illustrates the adopted hierarchical demographic estimator.

The demographic estimation was performed using LIBLINEAR (Fan, Chang, Hsieh, Wang, & Lin, 2008) which is a package for large-scale regularized linear classification and regression. Tuning the best parameters of the classifier or the regressor was based on grid search strategy combined with cross-validation.

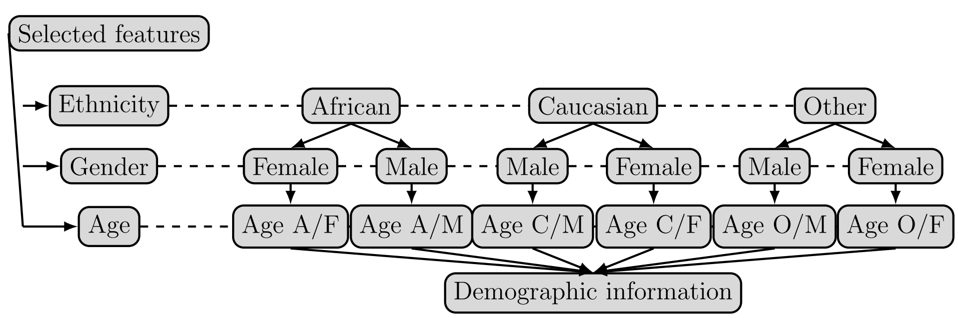


Figure 5: Hierarchical demographic estimation.

4 Experimental results

To evaluate the performance of the proposed approach, we used two databases MORPH II[1] and PAL[2] which are labeled with all demographic attributes. Moreover, we used IoG[3], LFW[4] and FERET[5] databases to test the approach in different scenarios and to do a cross-database evaluation.

The performance of ethnicity and gender classification is measured by the classification accuracy which is given by the proportion of correct predictions among the total number of cases examined. On the other hand, the age estimation performance is measured either by the accuracy in the case of age groups classification or the Mean Absolute Error (MAE), the Cumulative Score (CS) and the Standard Deviation (STD) of the age error in the case where the exact age is estimated. The mean absolute error is the average of the absolute errors between the ground-truth ages and the predicted ones. The MAE equation is given by:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|p_i - g_i| \quad (3)$$

where N, p_i, g_i are the total number of samples, the predicted age and the ground-truth age respectively.

The cumulative score reflects the percentage of tested cases where the age estimation error is less than a threshold. The CS equation is given by:

$$CS(l) = \frac{N_{e\leq l}}{N}\% \quad (4)$$

[1]http://www.faceaginggroup.com/morph/
[2]http://agingmind.utdallas.edu/download-stimuli/face-database/
[3]http://chenlab.ece.cornell.edu/people/Andy/ImagesOfGroups.html
[4]http://vis-www.cs.umass.edu/lfw/
[5]http://www.itl.nist.gov/iad/humanid/feret/feret˙master.html

where l, N and $N_{e\leq l}$ are threshold (years), the total number of samples and the number of samples on which the age estimation makes an absolute error no higher than the threshold.

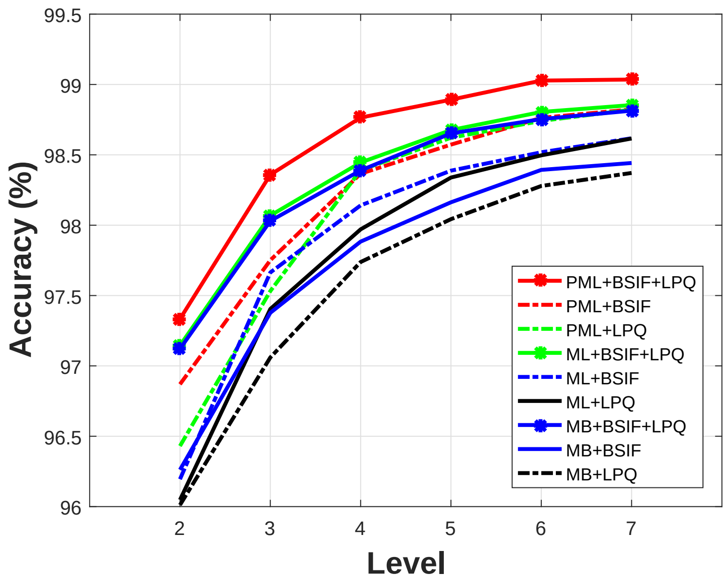
4.1 Single Database

4.1.1 MORPH (Album 2)

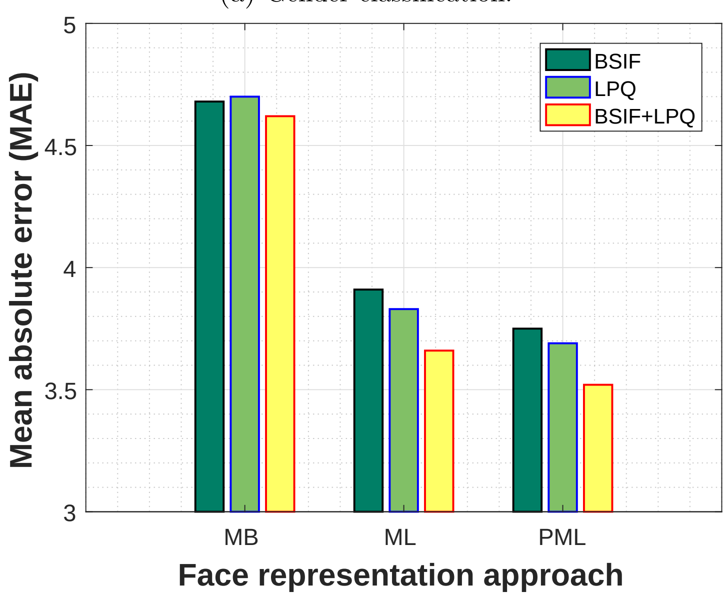
The MORPH (Album 2) database from the University of North Carolina Wilmington (Ricanek & Tesafaye, 2006) contains 55,000 unique images of 13,618 individuals (11,459 male and 2,159 female) in the age range from 16 to 77 years old. The average number of images per individual is 4. The MORPH (Album 2) database can be divided into three main ethnicities: African 42,589 images, European 10,559 images and other ethnicities 1,986 images. Following the BeFIT[6] recommendations, we use 5-fold cross-validation evaluation, the folds are selected in such a way to prevent algorithms from learning the identity of the persons in the training set by making sure that all images of individual subjects are only in one fold at a time. In addition to the distribution of age, gender and ethnicity distributions in the folds are similar to the distributions in the whole database.

Figure 6a presents the accuracy of gender classification obtained by different face representations (MB, ML and PML) and different texture descriptors (BSIF, LPQ and BSIF+LPQ). The accuracy is given as a function of the level. In the figure, the level is related to the face representation used. For example, when PML is used the level indicates the pyramid level. We can observe that the performance of the PML is better than MB and ML using the same descriptor. Moreover, in Figure 6b the performance of the PML is better than both MB and ML using 7 levels for each face representation.

[6] http://fipa.cs.kit.edu/425.php



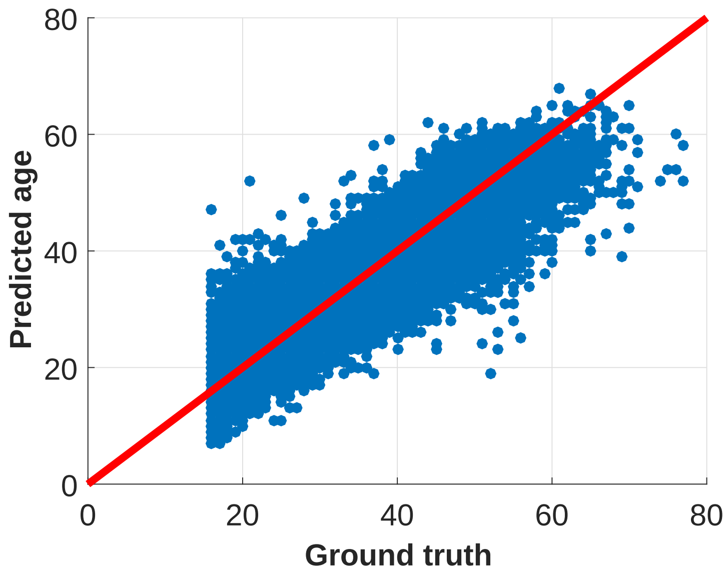
(a) Gender classification.



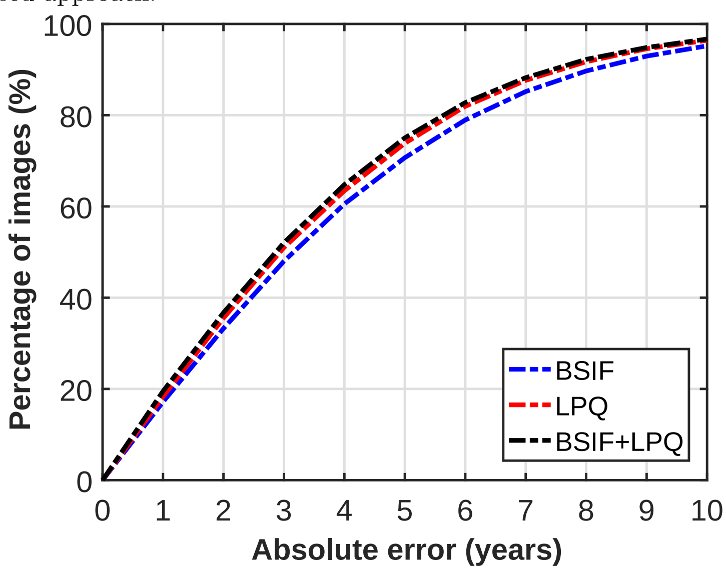
(b) Age estimation.

Figure 6: Comparison of different face representations on demographic estimation.

Table 4 illustrates the age estimation performance of the proposed approach as well as of that of some competing approaches. Figure 7 shows the performance of the proposed approach on MORPH II. Figure 7.(a) illustrates the correlation between the estimated ages and their ground-truth values. Figure 7.(b) illustrates the CS associated with three face descriptors.



(a) Correlations between True ages and Predicted ages by the proposed approach.



(b) Cumulative scores obtained by the proposed approach.

Figure 7: Age estimation results on Morph II database.

Table 4: A comparison of the proposed approach with other demographic estimation approaches on MORPH II database.

Publication	Approach	Performance		
		Age (MAE/STD)	Gender	Ethnicity
(Guo & Mu, 2011)	BIF*+KPLS†	4.2/NA	98.2%	98.9%
(Chang et al., 2011)	BIF*	6.1/NA	NA	NA
(Geng et al., 2013)	BIF*	4.8/NA	NA	NA
(Guo & Mu, 2013)	BIF*	4.0/NA	96.0%	98.9%
(Huerta et al., 2015)	CNN‡	3.9/NA	NA	NA
(Han et al., 2015)	DIF††	3.6/NA	97.6%	99.1%
Proposed approach	PML (BSIF+LPQ)+ SVM / SVR	**3.5/2.89**	**98.9%**	**99.3%**

* Biologically Inspired Features

† Kernel Partial Least Squares

‡ Convolutional Neural Networks

†† Demographic Informative Features

Figure 8 illustrates some examples for good and poor demographic estimation obtained by the proposed approach with MORPH II database. For age estimation, a poor estimation is declared for a test face if the absolute age error is equal to or greater than one year.

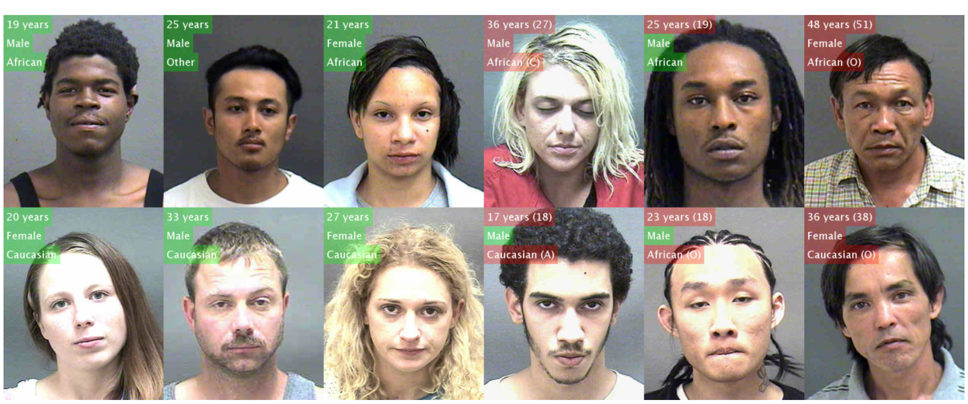
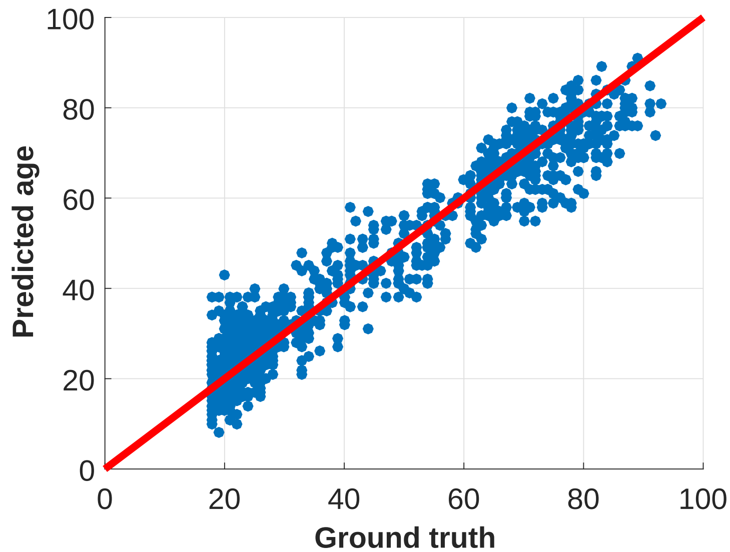


Figure 8: Examples of good and poor demographic estimation.

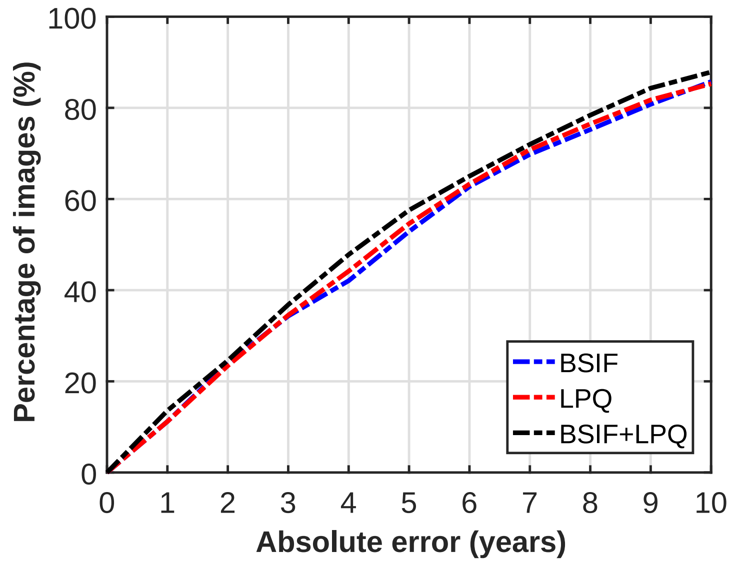
4.1.2 PAL

The Productive Aging Lab Face (PAL) database from the University of Texas at Dallas (Minear & Park, 2004) contains totally 1,046 frontal face images from different subjects (430 males and 616 females) in the age range from 18 to 93 years old. The PAL database can be divided into three main ethnicities: African-American subjects 208 images, Caucasian subjects 732 images and other subjects 106 images. The database contains faces having different expressions. For the evaluation of the approach, we conduct 5-fold cross-validation, our distribution of folds is selected based on age, gender and ethnicity.

Figure 9 shows the results of age estimation obtained with PAL dataset. The first row in the figure illustrates the correlation between the predicted ages and their ground truth values. The other row represents the cumulative score associated with three face descriptors.



(a) Correlations between True ages and Predicted ages by the proposed approach.



(b) Age estimation performance by the proposed approach.

Figure 9: Age estimation results on PAL database.

The comparison of our approach with the state-of-the-art approaches is shown in Table 5. The authors of (Sung Eun, Youn Joo, Sung Joo, Kang Ryoung, & Jaihie, 2010; Luu et al., 2011; Günay & Nabiyev, 2016) use part of the PAL database which corresponds to the frontal images without expression. Furthermore, they conduct K-fold cross-validation testing scheme for evaluating the performance of age estimation. In (Nguyen, Cho, Shin, Bang, & Park, 2014; S. Bekhouche et al., 2014), the authors use the whole frontal database with the same scheme to evaluate the performance of age estimation. For gender classification, Patel *et al.* (Patel et al., 2015) conduct 5-fold cross-validation without mentioning the number of images used.

Table 5: A comparison of the proposed approach with other demographic estimation approaches on PAI database.

Publication	Approach	Performance		
		Age (MAE/STD)	Gender	Ethnicity
(Sung Eun et al., 2010)	GHPF*+SVR	8.4/NA	NA	NA
(Luu et al., 2011)	CAM†+SVR	6.0/NA	NA	NA
(Nguyen, Cho, Shin, et al., 2014)	MLBP+GABOR+SVR	6.5/NA	NA	NA
(S. Bekhouche et al., 2014)	LBP+BSIF+SVR	6.2/NA	NA	NA
(Patel et al., 2015)	MQLBP‡+SVM	NA	95.5%	NA
(Günay & Nabiyev, 2016)	AAM+GABOR+LBP	5.4/NA	NA	NA
Proposed approach	PML(BSIF+LPQ)+SVM/SVR	**5.0/3.93**	**98.2%**	**99.3%**

* Gaussian High Pass Filter
† Contourlet Appearance Model
‡ Multi-Quantized Local Binary Patterns

Figure 10 shows some examples of good and poor demographic estimation in PAL database.

Figure 10: Examples of good and poor demographic estimation.

4.1.3 IoG

The Images of Groups (IoG) database (A. C. Gallagher & Chen, 2008) contains 5080 images with 28231 faces labeled with age and gender. There are seven age categories as follows : 0-2, 3-7, 8-12, 13-19, 20-36, 37-65, and 66+ years, roughly corresponding to different life stages. In some images, people are sitting, laying, or standing on elevated surfaces. People often have dark glasses, face occlusions, or unusual facial expressions. Note that in the case of IoG database, the demographic attributes are gender and age group. The hierarchical system of the proposed approach has the gender as a root.

We perform evaluations using the entire database with 5-fold cross-validation testing scheme. Table 6 illustrates the confusion matrix associated with age group classification. Table 7 illustrates the age estimation performance of the proposed approach as well as of that of some competing approaches. In this table, the column Age+ considers the estimated age group as correct if it corresponds to either the ground-truth age group or to its nearest neighbor age group. Figure 11 illustrates some examples for good and poor demographic estimation obtained by the proposed approach with IoG database.

Table 6: Confusion matrix of age classification on IoG database.

P \ G	0-2	3-7	8-12	13-19	20-36	37-65	66+
0-2	712	170	10	2	9	0	0
3-7	177	959	304	57	21	11	0
8-12	2	36	53	14	8	1	0
13-19	0	13	23	40	10	2	0
20-36	52	375	439	1497	13852	3359	151
37-65	10	38	41	77	1138	3384	862
66+	1	4	2	5	10	60	240

P Predicted groups
G Ground-truth groups

Table 7: A comparison of the proposed approach with other demographic estimation approaches on

Publication	Approach	Pe
		Age
(A. Gallagher & Chen, 2009)	Appearance+Context+MLE	42.9%
(Shan, 2010)	boosted LBP+SVM	55.9%
(Ylioinas et al., 2012)	CLBP_M + SVM	51.7%
(Li, Liu, Liu, & Lu, 2012)	Gabor+PLO*	48.5%
(Han & Jain, 2014)	BIF+SVM	68.1%
(Eidinger et al., 2014)	LBP+FPLBP	66.6%
(Fazl-Ersi et al., 2014)	LBP+CH+SIFT	63.0%
(S. Bekhouche, Ouafi, Benlamoudi, Taleb-Ahmed, & Hadid, 2015)	ML-LBP+SVM	55.1%
(Mansanet et al., 2016)	Best Local-DNN†	NA
Proposed approach	PML(BSIF+LPQ)+SVM	**68.2**%

* Preserving Locality
† Deep Neural Network
‡ Multi-Quantized Local Binary Patterns

Figure 11: Examples of good and poor demographic estimation.

4.1.4 LFW

The Labeled Faces in the Wild (LFW) (Huang, Ramesh, Berg, & Learned-Miller, 2007) is a database of faces which contains 13,000 images of 1680 celebrities labeled with gender. For evaluating our approach, we follow the BeFIT benchmarking protocol by conducting a 5-fold cross-validation scheme. We do not use the hierarchical estimator in this case due to the absence of age and ethnicity attributes for this database. The comparison with the state-of-the-art approaches is reported in Table 8. Also, we include some images with their prediction of gender in Figure 12.

Table 8: A comparison of the proposed approach with other gender classification approaches on LFW database.

Publication	Approach	Performance Gender
(Dago-Casas, Gonzlez-Jimnez, Yu, & Alba-Castro, 2011)	LBP+PCA+SVM	97.2%
(Shan, 2012)	LBP+SVM	94.8%
(Tapia & Perez, 2013)	Fusion	98.0%
(Han & Jain, 2014)	BIF+SVM	95.4%
(Ren & Li, 2014)	SIFT+HOG+Gabor	98.0%
(Han et al., 2015)	DIF	94.0%
(Antipov, Berrani, & Dugelay, 2016)	CNN	97.3%
Proposed approach	PML(BSIF+LPQ)+SVM	**98.3%**

Figure 12: Examples of good and poor demographic estimation.

4.1.5 FERET

The FERET database is a collection of 11338 facial images collected by photographing 994 subjects at various angles, over the course of 15 sessions between 1993 and 1996. To evaluate our approach, we use only the fa subset (regular frontal image) and the fb subset (alternative frontal image) that contain 2722 images of 994 subjects. in this case, the demographic attributes are ethnicity and gender. We adopted a 5-fold cross-validation testing scheme in our experiments for FERET database. In table 9, we give a comparison between our approach and some state-of-the-art approaches. Figure 13 illustrates some examples for good and poor demographic estimation obtained by the proposed approach with FERET database.

Table 9: A comparison of the proposed approach with other demographic estimation approaches on FERET database.

Publication	Approach	Performance Gender	Performance Ethnicity
(Yang & Ai, 2007b)	LBP + Real AdaBoost	93.3%	NA
(Mkinen & Raisamo, 2008)	LBP + SVM	92.0%	NA
(Han et al., 2015)	DIF	96.8%	NA
(W. Zhang, Smith, Smith, & Farooq, 2016)	FVs+linear SVM	97.7%	NA
Proposed approach	PML(BSIF+LPQ)+SVM	**99.0%**	**95.6%**

Figure 13: Examples of good and poor demographic estimation.

4.1.6 Chalearn2016

2016 ChaLearn Looking at People and Faces of the World Challenge[7] ran three parallel quantitative challenge tracks on RGB face analysis. The first track was about apparent age estimation, the dataset used in this track contains 8,000 images labeled with the apparent age. Figure 14 shows some samples of the dataset. Each image has been labeled by multiple individuals, using a collaborative Facebook implementation and Amazon Mechanical Turk. The votes variance is used as a measure of the error for the predictions (Escalera et al., 2016). The mean is used as the ground truth age. For quantitative evaluation instead of MAE, the following score is used:

$$\epsilon = \frac{1}{N}\sum_{i=1}^{N} 1 - e^{-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}} \tag{5}$$

where N and x_i are the total number of samples and the predicted age, μ_i and σ_i are the mean and standard deviation of the human labels.

[7]http://gesture.chalearn.org/2016-looking-at-people-cvpr-challenge

Figure 14: Samples from Chalearn2016 dataset.

These are some features which make the dataset a challenge to researchers:

- Images with diverse backgrounds.
- Images taken in non-controlled environments.
- Non-labelled face bounding boxes, neither face landmarks.

Nearly 100 participants registered for this challenge, the top 8 participants are Deep learning-based approaches (shown in Table 10). OrangeLabs team(Antipov, Baccouche, Berrani, & Dugelay, 2016) which are the winners of the apparent age competition used the pretrained VGG-16 convolutional neural network (CNN) to train it on the huge IMDB-Wiki dataset (500k+ face images) for biological age estimation and then ne-tune it for apparent age estimation using the dataset of the challenge. They got an age error of 0.2411 in terms of the score ϵ. Neither the CPU time nor the time complexity was mentioned in their work. Both palm seu team(Huo et al., 2016) and cmp+ETH (Uricar, Timofte, Rothe, Matas, & Van Gool, 2016) teams used the pretrained VGG-16 CNN in their approaches.

We conducted our experiments using the same protocols as the contenders of the challenge did. Compared to OrangeLabs team approach the difference between our results and theirs is about 0.0463 in terms of the score ϵ. Also, according to the Table 10 our results come in the second position. The results showed that our approach can give better results even in uncontrolled complex conditions. Thus, it confirms that deep-learning-based approaches are not necessarily the best ones.

Table 10: Comparison of the results between the participating teams approaches and our approach in the apparent age-estimation competition.

Position	Team	Test Error
1	OrangeLabs	0.2411
2	palm seu	0.3214
3	cmp+ETH	0.3361
4	WYU CVL	0.3405
5	ITU SiMiT	0.3668
6	Bogazici	0.3740
7	MIPAL SNU	0.4569
8	DeepAge	0.4573
-	Our approach	0.2874

The third track was about smile and gender classification. They used a different subset of images from the FotW dataset to create this challenging dataset. This dataset is composed of a training set (6,171 images), a validation set (3,086 images) and a test set (8,505 images). Table 11 shows the images distribution of the dataset.

Table 11: The images distribution based on facial expression and gender in the dataset.

Attribute	Train	Val	Test
Male	2946	1691	4614
Female	3318	1361	3799
Not sure	93	34	92
Smile	2234	1969	4411
No smile	3937	1117	3849

Nearly 60 participants registered for this competition. The 7 teams that finally submitted predictions are listed in Table 12 together with the accuracy of each of the participating teams. The performance evaluation using this database give good results in both smile and gender classification. From the fact of the results of smiles classification (facial expression), we can say that our can be applied to other face-based learning tasks.

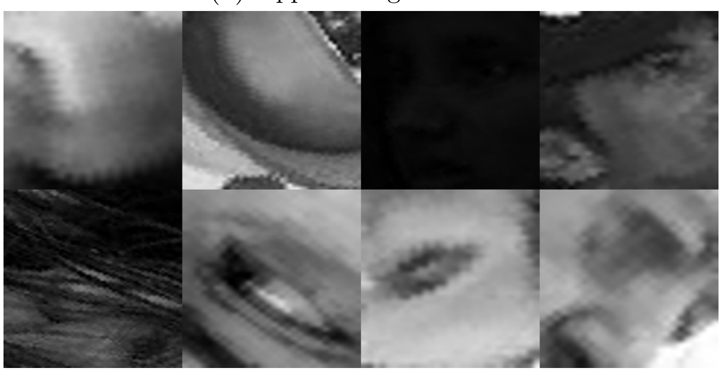
SIAT MMLAB team(K. Zhang, Tan, Li, & Qiao, 2016) won the challenge, they proposed a deep model composed of two CNNs GNet and SNet, for facial gender and smile classification tasks. Also, they used multi-task and general-to-specific fine-tuning scheme while training their models. Their proposed deep model gave results of 92.69% for gender and 85.83% for smile classifications. Our results came in the third position compared to the ranking table of the challenge. We get 90.36% and 82.15% for gender and smile classification respectively.

Table 12: Comparison of the results between the participating teams approaches and our approach in the gender/smiles classification competition.

Team	Gender	Smile	Mean
SIAT MMLAB	0.9269	0.8583	0.8926
IVA NLPR	0.9152	0.8252	0.8702
VISI.CRIM	0.9016	0.8212	0.8614
SMILELAB NEU	0.8999	0.8148	0.8574
Lets Face it!	0.8454	0.8439	0.8446
CMP+ETH	0.7465	0.7189	0.7327
SRI	0.5716	0.5853	0.5784
Our approach	0.9036	0.8215	0.8625

The databases used in the 2016 ChaLearn Looking at People and Faces of the World Challenge are considered as uncontrolled (In the wild) databases. Our system had some poor face alignment which affect the global performance of the proposed method. Figure 15 shows some poor face alignment.

(a) Apparent age dataset.



(b) Gender/Smiles dataset.

Figure 15: Poor face preprocessing on Chalearn2016 datasets.

4.2 Cross-database evaluation

In order to have a quantitative evaluation on multiple databases, we performed cross-database experiments on the demographic attributes.

4.2.1 Age estimation

For the age estimation task, we use two databases MORPH II and PAL since both provide the age, gender and ethnicity labels. This enables us to use the full hierarchical estimator (with three layers). The protocol adopted in this experiment is to use an entire database as the training subset and the other as the testing subset and vice-versa. The results are given in Table 13.

Table 13: The performance (MEA in years) of the cross-database evaluation in the case of age estimation.

Training \ Testing	PAL	MORPH II
PAL	/	6.1
MORPH II	5.5	/

We can observe that when MORPH II database is used as a training subset (PAL is used as a test dataset) the performance is better than the opposite case. This can be explained by the fact that MORPH II database is very rich and can better capture variation in the demographic attributes. Also, the performance on PAL deteriorated from 5.0 to 5.6. This is due to the fact that the two databases are different.

4.2.2 Gender classification

In the gender classification, we use all proposed databases to this end, we exclude the hierarchical estimation and use just the gender classifier. The protocol used in the task is the same which was used in the age estimation task. In table 14, we present the results. We can observe that whenever the IoG database is used in the training or testing subsets, the gender classification performance is not good compared to the other databases. This may be due to the distribution of IoG database which contains a lot of children images, unlike the others. Both IoG and LFW images are captured in the wild. Therefore, their results are quasi symmetric. In most cases, MORPH II and LFW have the best results which can be due to their richness in images.

Table 14: Cross-database evaluation in the case of gender classification.

Training \ Testing	PAL	MORPH II	IoG	LFW	FERET
PAL	/	87.6%	65.3%	82.1%	87.3%
MORPH II	89.0%	/	75.7%	90.3%	94.0%
IoG	75.6%	72.2%	/	79.1%	70.6%
LFW	84.5%	89.5%	81.0%	/	92.8%
FERET	91.1%	89.8%	71.5%	88.1%	/

4.2.3 Ethnicity classification

Like in age estimation and gender classifications tasks, we use the same protocol for ethnicity classification. We use three databases (MORPH II, PAL and FERET). The results are presented in Table 15. The results of cross-database evaluation are good given the fact that there is a difference in the distribution and in the images of these three databases.

Table 15: Cross-database evaluation in the case of ethnicity classification.

Training \ Testing	PAL	MORPH II	FERET
PAL	/	81.1%	91.6%
MORPH II	88.0%	/	87.7%
FERET	95.3%	85.2%	/

4.3 Other effect

In the proposed hierarchical estimator, the age is estimated after estimating the ethnicity and gender attributes. We have performed an experimental evaluation to investigate the effect of knowing one or more demographic attributes on the estimation of the other.

To this end, we split MORPH II database into six subsets that correspond to the six combinations gender-ethnicity. Then, a five-fold cross-validation is conducted in which the age of each test image is estimated using the corresponding SVR. Table 16 illustrates the MEA obtained by using the six SVRs. In this case, the ground-truth of ethnicity and gender attributes of the test image are used.

Table 16: Age estimation for different ethnicities and gender on MORPH II database. The ground-truth ethnicity and gender are used by the age estimator.

<table>
<tr><th>Gender</th><th>Ethnicity</th><th>Images</th><th colspan="2">Age</th></tr>
<tr><td rowspan="3">Male</td><td>Black</td><td>36804</td><td>3.2</td><td rowspan="6">3.4</td></tr>
<tr><td>Other</td><td>1843</td><td>3.5</td></tr>
<tr><td>White</td><td>7999</td><td>3.0</td></tr>
<tr><td rowspan="3">Female</td><td>Black</td><td>5758</td><td>4.4</td></tr>
<tr><td>Other</td><td>129</td><td>4.7</td></tr>
<tr><td>White</td><td>2601</td><td>3.9</td></tr>
</table>

According to table 4, the obtained MAE is 3.5 years. This corresponds to the use of the hierarchical estimate. However, when the ground-truth ethnicity and gender labels are used the same MAE becomes 3.4 years (see Table 16). Since we know the misclassified images in the ethnicity layer, it is possible to study the influence of these misclassified 408 images through the ethnicity classifier. To this end, we calculate the absolute age error for the case where the ground truth ethnicity is used for these 408 images. The corresponding MAE is 3.3 years. On the other side, where the full hierarchical system is used (i.e, the predicted ethnicity is propagated) the MAE associated with these 408 images is 5.2 years. Figure 16 shows the curve of the age error of the 408 misclassified images in both cases (ground-truth ethnicity and predicted ethnicity).

We also studied the effect of misclassified images in the ethnicity layer on the gender layer. Table 17 shows the results of gender classification when we use the ethnicity layer of the hierarchical estimator and the ground-truth ethnicity. From these results, we can observe that the possible misclassification in the ethnicity layer can slightly deteriorate the performance of gender and age estimation.

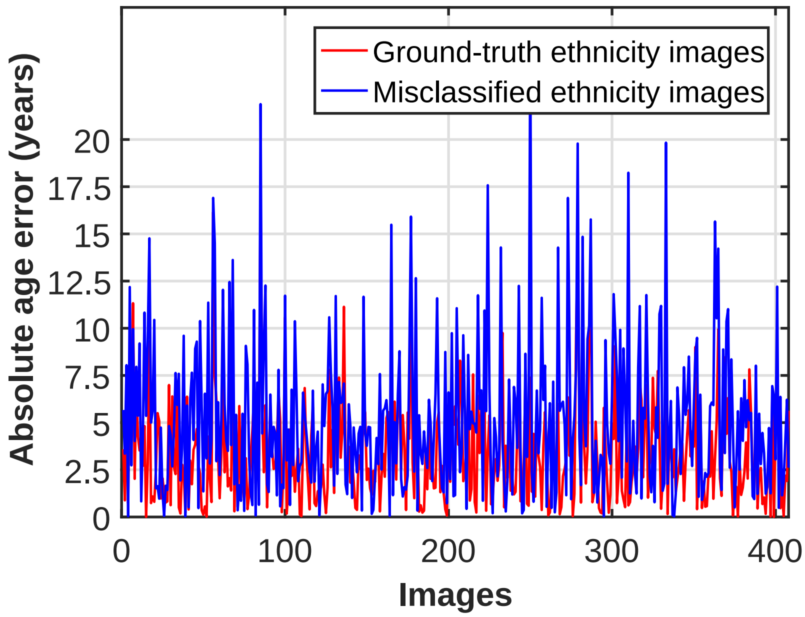


Figure 16: Absolute age error of misclassified ethnicity images and ground-truth ethnicity images.

Table 17: Gender classification using predicted and ground-truth ethnicities.

	Ground-truth ethnicity	Predicted ethnicity
Gender rate	99.2%	98.9%

4.4 Real-time application

We implement our approach as a C++ application using OpenCV[8] and DLIB libraries[9]. We used PML BSIF features. The structure of the system is composed of: (i) face preprocessing which includes detection and normalization, (ii) feature extraction and selection, and (iii) the demographic estimator which contains one SVM for ethnicity classification, 3 SVMs for gender classification and 6 SVRs for age estimation. We emphasize that, at testing phase, the system will apply 2 SVMs and 1 SVR each time. The CPU time associated with each part is given in Table 18. The application works at 30 fps, if there is more than one face it is best to ig-

[8]http://opencv.org/

[9]http://dlib.net/

nored some frames. The test is carried out on a laptop DELL 7510 Precision (Xeon Processor E3-1535M v5 ,8M Cache, 2.90 GHz, 64GB RAM, Ubuntu 16.04)..

Among the approaches which worked on the joint demographic attributes, only (Guo & Mu, 2014) and (Han et al., 2015) reported the CPU time associated with their developed methods. Guo (Guo & Mu, 2014) has reported a 623 ms for the testing phase for one image excluding the time of feature extraction. Their test was on Lenovo Thinkpad X61 (Intel Core 2, Duo CPU T8100, @2.1GHz, 4G RAM, Windows Vista 64). In the other hand Han (Han et al., 2015) has got 90 ms on execution time for one image without excluding any part of the prototype. They using a laptop with 2.9 GHz dual-core Intel Core i7 processor and 8G RAM for the testing the system.

We emphasize the fact that the CPU time of proposed state-of-the-art demographic estimation methods is not always reported in the literature. Indeed, most researchers in this field were focusing on improving the accuracy of their methods and systems.

Table 18: CPU time (ms) of the proposed approach.

Face preprocessing	Feature extraction and selection	Demographic estimation	Total
9	12	7	28

This application is developed to support the facial recognition system by using this information as facial soft biometric traits. Moreover, the application can be used in security for the governments or to get statistics in social media and commercial centers.

Conclusion

This paper presents a novel approach for automatic demographic estimation that combines Pyramid Multi-Level local features obtained by two different texture descriptors LPQ and BSIF. Furthermore, in order to improve the estimation accuracy of the demographic attributes, the proposed approach adopts a hierarchical estimator. Several performance comparisons are presented.

These comparisons show that the results of the proposed automatic demographic estimation are better than those obtained by most of the state-of-the-art approaches in different well-known databases (MORPH II, PAL, IoG, LFW, FERET). Also, the proposed approach performed well in the challenging databases. Moreover, the developed real-time application demonstrates the ability and efficiency of our approach for extracting the demographic information using a regular computer.

As a future work, we envision the use of deep features obtained from pre-trained neural networks. We also envision the improvement of the hierarchical demographic estimator by improving the architecture of the hierarchy and by introducing more sophisticated feature selection methods. Concerning the application, we suggest using multiple threads with a synchronization between the different threads. We may include a face tracker in order to predict the demographic information from video sequences.

References

Antipov, G., Baccouche, M., Berrani, S.-A., & Dugelay, J.-L. (2016, June). Apparent age estimation from face images combining general and children-specialized deep learning models. In *The ieee conference on computer vision and pattern recognition (cvpr) workshops.*

Antipov, G., Berrani, S.-A., & Dugelay, J.-L. (2016). Minimalistic cnn-based ensemble model for gender prediction from face images. *Pattern Recognition Letters*, *70*, 59 - 65. doi: 10.1016/j.patrec.2015.11.011

Bekhouche, S., Ouafi, A., Benlamoudi, A., Taleb-Ahmed, A., & Hadid, A. (2015). Automatic age estimation and gender classification in the wild. In *Proceeding of the international conference on automatic control, telecommunications and signals icats'15.* doi: 10.13140/RG.2.1.1893.8323

Bekhouche, S., Ouafi, A., Taleb-Ahmed, A., Hadid, A., & Benlamoudi, A. (2014). Facial age estimation using bsif and lbp. In *Proceeding of the first international conference on electrical engineering iceeb14.* doi: 10.13140/RG.2.1.1933.6483/1

Bekhouche, S. E., Ouafi, A., Benlamoudi, A., Taleb-Ahmed, A., & Hadid, A. (2015, May). Facial age estimation and gender classification using multi level local phase quantization. In *Control, engineering information technology (ceit), 2015 3rd international conference on* (p. 1-4). doi: 10.1109/CEIT.2015.7233141

Chang, K. Y., Chen, C. S., & Hung, Y. P. (2011, June). Ordinal hyperplanes ranker with cost sensitivities for age estimation. In *Computer vision and pattern recognition (cvpr), 2011 ieee conference on* (p. 585-592). doi: 10.1109/CVPR.2011.5995437

Correa, T., Hinsley, A. W., & de Ziga, H. G. (2010). Who interacts on the web?: The intersection of users personality and social media use. *Computers in Human Behavior*, *26*(2), 247 - 253. doi: http://dx.doi.org/10.1016/j.chb.2009.09.003

Dago-Casas, P., Gonzlez-Jimnez, D., Yu, L. L., & Alba-Castro, J. L. (2011, Nov). Single- and cross- database benchmarks for gender classification under unconstrained settings. In *Computer vision workshops (iccv workshops), 2011 ieee international conference on* (p. 2152-2159). doi: 10.1109/ICCVW.2011.6130514

Eidinger, E., Enbar, R., & Hassner, T. (2014, Dec). Age and gender estimation of unfiltered faces. *Information Forensics and Security, IEEE Transactions on*, *9*(12), 2170-2179. doi: 10.1109/TIFS.2014.2359646

Escalera, S., Torres, M., Martinez, B., Baró, X., Escalante, H. J., et al. (2016). Chalearn looking at people and faces of the world: Face analysis workshop and challenge 2016. In *Proceedings of ieee conference on computer vision and pattern recognition workshops.*

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., & Lin, C.-J. (2008). Liblinear: A library for large linear classification. *The Journal of Machine Learning Research*, *9*, 1871–1874.

Fazl-Ersi, E., Mousa-Pasandi, M. E., Laganire, R., & Awad, M. (2014, Oct). Age and gender recognition using informative features of various types. In *Image processing (icip), 2014 ieee international conference on* (p. 5891-5895). doi: 10.1109/ICIP.2014.7026190

Fu, Y., & Huang, T. S. (2008, June). Human age estimation with regression on discriminative aging manifold. *IEEE Transactions on Multimedia*, *10*(4), 578-584. doi: 10.1109/TMM.2008.921847

Gallagher, A., & Chen, T. (2009, June). Understanding images of groups of people. In *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference* *on* (p. 256-263). doi: 10.1109/CVPR.2009.5206828

Gallagher, A. C., & Chen, T. (2008, June). Clothing cosegmentation for recognizing people. In *Computer vision and pattern recognition, 2008. cvpr 2008. ieee conference on* (p. 1-8). doi: 10.1109/CVPR.2008.4587481

Geng, X., Yin, C., & Zhou, Z. H. (2013, Oct). Facial age estimation by learning from label distributions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(10), 2401-2412. doi: 10.1109/TPAMI.2013.51

Geng, X., Zhou, Z.-H., & Smith-Miles, K. (2007, Dec). Automatic age estimation based on facial aging patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *29*(12), 2234-2240. doi: 10.1109/TPAMI.2007.70733

Geng, X., Zhou, Z.-H., Zhang, Y., Li, G., & Dai, H. (2006). Learning from facial aging patterns for automatic age estimation. In *Proceedings of the 14th acm international conference on multimedia* (pp. 307–316). New York, NY, USA: ACM. doi: 10.1145/1180639.1180711

Gilani, S. Z., Shafait, F., & Mian, A. (2013, Nov). Biologically significant facial landmarks: How significant are they for gender classification? In *Digital image computing: Techniques and applications (dicta), 2013 international conference on* (p. 1-8). doi: 10.1109/DICTA.2013.6691488

Gunay, A., & Nabiyev, V. (2008, Oct). Automatic age classification with lbp. In *Computer and information sciences, 2008. iscis '08. 23rd international symposium on* (p. 1-4). doi: 10.1109/ISCIS.2008.4717926

Gunay, A., & Nabiyev, V. V. (2007, June). Automatic detection of anthropometric features from facial images. In *2007 ieee 15th signal processing and communications applications* (p. 1-4). doi: 10.1109/SIU.2007.4298656

Günay, A., & Nabiyev, V. V. (2016). Information sciences and systems 2015: 30th international symposium on computer and information sciences (iscis 2015). In

H. O. Abdelrahman, E. Gelenbe, G. Gorbil, & R. Lent (Eds.), (pp. 295–304). Cham: Springer International Publishing. doi: 10.1007/978-3-319-22635-4_27

Guo, G., & Mu, G. (2011, June). Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression. In *Computer vision and pattern recognition (cvpr), 2011 ieee conference on* (p. 657-664). doi: 10.1109/CVPR.2011.5995404

Guo, G., & Mu, G. (2013, April). Joint estimation of age, gender and ethnicity: Cca vs. pls. In *Automatic face and gesture recognition (fg), 2013 10th ieee international conference and workshops on* (p. 1-6). doi: 10.1109/FG.2013.6553737

Guo, G., & Mu, G. (2014). A framework for joint estimation of age, gender and ethnicity on a large database. *Image and Vision Computing*, *32*(10), 761 - 770. (Best of Automatic Face and Gesture Recognition 2013) doi: 10.1016/j.imavis.2014.04.011

Hadid, A., & Pietikinen, M. (2013). Demographic classification from face videos using manifold learning. *Neurocomputing*, *100*, 197 - 205. (Special issue: Behaviours in video) doi: 10.1016/j.neucom.2011.10.040

Han, H., & Jain, A. K. (2014, July). *Age, gender and race estimation from unconstrained face images* (Tech. Rep. No. MSU-CSE-14-5). East Lansing, Michigan: Department of Computer Science, Michigan State University.

Han, H., Otto, C., Liu, X., & Jain, A. K. (2015, June). Demographic estimation from face images: Human vs. machine performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(6), 1148-1161. doi: 10.1109/TPAMI.2014.2362759

Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007, October). *Labeled faces in the wild: A database for studying face recognition in unconstrained* *environments* (Tech. Rep. No. 07-49). University of Massachusetts, Amherst.

Huerta, I., Fernandez, C., Segura, C., Hernando, J., & Prati, A. (2015). A deep analysis on age estimation. *Pattern Recognition Letters*, *68, Part 2*, 239 - 249. (Special Issue on Soft Biometrics) doi: 10.1016/j.patrec.2015.06.006

Huo, Z., Yang, X., Xing, C., Zhou, Y., Hou, P., Lv, J., & Geng, X. (2016, June). Deep age distribution learning for apparent age estimation. In *The ieee conference on computer vision and pattern recognition (cvpr) workshops.*

Jain, A. K., Dass, S. C., & Nandakumar, K. (2004). *Can soft biometric traits assist user recognition?* (Vol. 5404). doi: 10.1117/12.542890

Kannala, J., & Rahtu, E. (2012, Nov). Bsif: Binarized statistical image features. In *Pattern recognition (icpr), 2012 21st international conference on* (p. 1363-1366).

Kazemi, V., & Sullivan, J. (2014). One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 1867–1874).

Kumar, N., Belhumeur, P., & Nayar, S. (2008). Facetracer: A search engine for large collections of images with faces. In D. Forsyth, P. Torr, & A. Zisserman (Eds.), *Computer vision – eccv 2008: 10th european conference on computer vision, marseille, france, october 12-18, 2008, proceedings, part iv* (pp. 340–353). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-540-88693-8_25

Kwon, Y. H., & da Vitoria Lobo, N. (1994, Jun). Age classification from facial images. In *Computer vision and pattern recognition, 1994. proceedings cvpr '94., 1994 ieee computer society conference on* (p. 762-767). doi: 10.1109/CVPR.1994.323894

Lanitis, A., Taylor, C. J., & Cootes, T. F. (2002, Apr). Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(4), 442-455. doi: 10.1109/34.993553

Levi, G., & Hassncer, T. (2015, June). Age and gender classification using convolutional neural networks. In *2015 ieee conference on computer vision and pattern recognition workshops (cvprw)* (p. 34-42). doi: 10.1109/CVPRW.2015.7301352

Li, C., Liu, Q., Liu, J., & Lu, H. (2012, June). Learning ordinal discriminative features for age estimation. In *Computer vision and pattern recognition (cvpr), 2012 ieee conference on* (p. 2570-2577). doi: 10.1109/CVPR.2012.6247975

Luu, K., Seshadri, K., Savvides, M., Bui, T. D., & Suen, C. Y. (2011, Oct). Contourlet appearance model for facial age estimation. In *Biometrics (ijcb), 2011 international joint conference on* (p. 1-8). doi: 10.1109/IJCB.2011.6117601

Mansanet, J., Albiol, A., & Paredes, R. (2016). Local deep neural networks for gender recognition. *Pattern Recognition Letters*, *70*, 80 - 86. doi: 10.1016/j.patrec.2015.11.015

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 630–633. doi: 10.3758/BF03206543

Mkinen, E., & Raisamo, R. (2008). An experimental comparison of gender classification methods. *Pattern Recognition Letters*, *29*(10), 1544 - 1556. doi: 10.1016/j.patrec.2008.03.016

Mu, G., Guo, G., Fu, Y., & Huang, T. S. (2009, June). Human age estimation using bio-inspired features. In *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on* (p. 112-119). doi: 10.1109/CVPR.2009.5206681

Nguyen, D. T., Cho, S. R., & Park, K. R. (2014). Future information technology: Futuretech 2014. In H. J. J. J. Park, Y. Pan, C.-S. Kim, & Y. Yang (Eds.), (pp. 433–438). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-55038-6_67

Nguyen, D. T., Cho, S. R., Shin, K. Y., Bang, J. W., & Park, K. R. (2014). Comparative study of human age estimation with or without preclassification of gender and facial expression [Journal Article]. *The Scientific World Journal*, *2014*, 15. doi: 10.1155/2014/905269

Ojansivu, V., & Heikkilä, J. (2008). Image and signal processing: 3rd international conference, icisp 2008. cherbourg-octeville, france, july 1 - 3, 2008. proceedings. In (pp. 236–243). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-540-69905-7_27

Park, U., & Jain, A. K. (2010, Sept). Face matching and retrieval using soft biometrics. *IEEE Transactions on Information Forensics and Security*, *5*(3), 406-415. doi: 10.1109/TIFS.2010.2049842

Patel, B., Maheshwari, R., & Balasubramanian, R. (2015). Multi-quantized local binary patterns for facial gender classification. *Computers & Electrical Engineering*, -. doi: 10.1016/j.compeleceng.2015.11.004

Ranjan, R., Zhou, S., Chen, J. C., Kumar, A., Alavi, A., Patel, V. M., & Chellappa, R. (2015, Dec). Unconstrained age estimation with deep convolutional neural networks. In *2015 ieee international conference on computer vision workshop (iccvw)* (p. 351-359). doi: 10.1109/ICCVW.2015.54

Ren, H., & Li, Z. N. (2014, Aug). Gender recognition using complexity-aware local features. In *Pattern recognition (icpr), 2014 22nd international conference on* (p. 2389-2394). doi: 10.1109/ICPR.2014.414

Ricanek, K., & Tesafaye, T. (2006, April). Morph: a longitudinal image database of normal adult age-progression. In *Automatic face and gesture recognition, 2006. fgr 2006. 7th international conference on* (p. 341-345). doi: 10.1109/

FGR.2006.78

Shan, C. (2010). Learning local features for age estimation on real-life faces. In *Proceedings of the 1st acm international workshop on multimodal pervasive video analysis* (pp. 23–28). New York, NY, USA: ACM. doi: 10.1145/1878039.1878045

Shan, C. (2012). Learning local binary patterns for gender classification on real-world face images. *Pattern Recognition Letters*, *33*(4), 431 - 437. (Intelligent Multimedia Interactivity) doi: 10.1016/j.patrec.2011.05.016

Sung Eun, C., Youn Joo, L., Sung Joo, L., Kang Ryoung, P., & Jaihie, K. (2010). A comparative study of local feature extraction for age estimation [Conference Proceedings]. In *Control automation robotics vision (icarcv), 2010 11th international conference on* (p. 1280-1284). doi: 10.1109/ICARCV.2010.5707432

Tapia, J. E., & Perez, C. A. (2013, March). Gender classification based on fusion of different spatial scale features selected by mutual information from histogram of lbp, intensity, and shape. *IEEE Transactions on Information Forensics and Security*, *8*(3), 488-499. doi: 10.1109/TIFS.2013.2242063

Uricar, M., Timofte, R., Rothe, R., Matas, J., & Van Gool, L. (2016, June). Structured output svm prediction of apparent age, gender and smile from deep features. In *The ieee conference on computer vision and pattern recognition (cvpr) workshops.*

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer vision and pattern recognition, 2001. cvpr 2001. proceedings of the 2001 ieee computer society conference on* (Vol. 1, p. I-511-I-518 vol.1). doi: 10.1109/CVPR.2001.990517

Wang, W., Chen, W., & Xu, D. (2011, May). Pyramid-based multi-scale lbp features for face recognition. In *Multimedia and signal processing (cmsp), 2011* *international conference on* (Vol. 1, p. 151-155). doi: 10.1109/CMSP.2011.37

Yang, Z., & Ai, H. (2007a). Advances in biometrics: International conference, icb 2007, seoul, korea, august 27-29, 2007. proceedings. In S.-W. Lee & S. Z. Li (Eds.), (pp. 464–473). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-540-74549-5_49

Yang, Z., & Ai, H. (2007b). Advances in biometrics: International conference, icb 2007, seoul, korea, august 27-29, 2007. proceedings. In S.-W. Lee & S. Z. Li (Eds.), (pp. 464–473). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-540-74549-5_49

Ylioinas, J., Hadid, A., & Pietikainen, M. (2012, Nov). Age classification in unconstrained conditions using lbp variants. In *Pattern recognition (icpr), 2012 21st international conference on* (p. 1257-1260).

Zhang, K., Tan, L., Li, Z., & Qiao, Y. (2016, June). Gender and smile classification using deep convolutional neural networks. In *The ieee conference on computer vision and pattern recognition (cvpr) workshops.*

Zhang, W., Smith, M. L., Smith, L. N., & Farooq, A. (2016). Gender and gaze gesture recognition for human-computer interaction. *Computer Vision and Image Understanding*, *149*, 32 - 50. doi: 10.1016/j.cviu.2016.03.014