Age, Gender and Race Estimation from Unconstrained Face Images

Hu Han, *Member, IEEE* and Anil K. Jain, *Fellow, IEEE*

***Abstract*— Automatic estimation of demographic attributes (*e.g.,* age, gender, and race) from a face image is a topic of growing interest with many potential applications. Most prior work on this topic has used face images acquired under constrained and cooperative scenarios. This paper addresses the more challenging problem of automatic age, gender, and race estimation from real-life face images (face images in the wild) acquired in unconstrained conditions. Given an input face image, we first normalize it by performing pose and photometric corrections. Biologically inspired features (BIF) are then extracted from the normalized face image, including both the central face region and the surrounding context region. Given this representation, three different Support Vector Machines (SVM) are used to predict the age group (or exact age), gender, and race of a subject. Experimental results on two large public-domain unconstrained face databases (Images of Groups and LFW) show that the proposed approach significantly outperforms the state-of-the-art methods. Our results also highlight that extraction of demographic attributes from face images in the wild is a difficult problem.**

***Index Terms*— Unconstrained demographic estimation, age group, gender and race classification, face alignment, illumination correction, 3D face modeling, Difference of Gaussisans, MTurk crowdsourcing**

I. Introduction

A person's face contains important clues for social interaction, providing a wide variety of useful information, including the person's identity, age, gender, race, expression, etc. Over the past 50 years, significant advances have been made on extracting discriminative features in a face image to determine the subject's identity [1]. In recent years, several applications have emerged that make use of demographic or soft biometric traits (*e.g.,* age, gender, and race). These applications include access control, re-identification in surveillance videos, law enforcement, integrity of face images in social media, intelligent advertising, and human-computer interaction. As a result, studies on the exploration of various attributes (other than the identity) embodied in a face image, such as age, gender, and race, have drawn increasing attention [2]–[5]. While state-of-the-art demographic estimation methods are able to attain a mean absolute error (MAE) of about 4 years for age estimation[1], and more than 95% accuracy for gender and race classifications, most of these studies have utilized face images captured in rather controlled sensing and cooperative subject scenarios [6], such as the MORPH [7] face database. However,

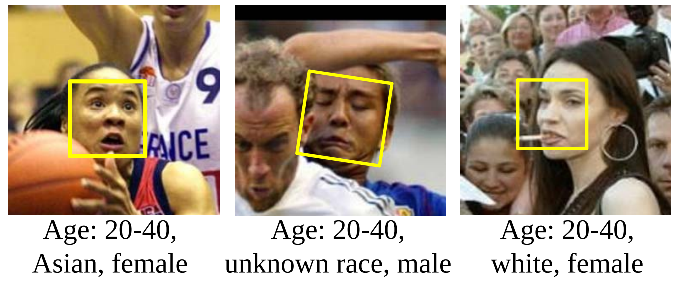


Fig. 1. Examples of unconstrained face images from the LFW database[2] with the human estimates of the age range, gender, and race collected using the Amazon Mechanical Turk (MTurk)[3] crowdsourcing service. Note that even humans are often not sure about specific demographic attributes (*e.g.,* the race of the subject in the second image) of a subject's face image that are captured in unconstrained and uncooperative scenarios.

in many of the applications mentioned above, especially video surveillance, the available face images of a person of interest are most likely to be captured under unconstrained and uncooperative scenarios (see Fig. 1). To close the gap between the applicability of published methods and the requirements of real world applications, researchers have attempted to develop new approaches for demographic estimation that are robust to unconstrained face images [8]–[13]. But, even these methods have some limitations. For example, most of them estimate a single demographic attribute with less than satisfactory performance. Additionally, no demographic estimation results have been reported on the Labeled Faces in the Wild (LFW) database [14], the most popular public-domain database for unconstrained face images.

In this paper, we present an integrated framework for age, gender, and race estimation from unconstrained face images. Our approach involves (i) face normalization consisting of pose and photometric corrections of an input face image, (ii) feature extraction from the normalized face image, including both the central face region and the surrounding contextual information (*e.g.,* facial shape, ears, and hair style), and (iii) age group (or exact age), gender, and race estimation using Support Vector Machines (SVM).

Unlike previous studies where low-resolution face images (*e.g.,* interpupillary distance (IPD) smaller than 24 pixels) were excluded from their evaluations [11], [13], we evaluate the proposed approach using the entire *Images of Groups* database [8] containing $28,231$ face images. Additionally, we use an extended version (LFW+) of the public-domain LFW database[4] containing $15,699$ face images to perform age,

H. Han and A. K. Jain are with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, USA. E-mail: {hhan, jain}@msu.edu.

[1]The mean absolute difference between the estimated ages and the ground-truth (real) ages is about four years.

[2]`http://vis-www.cs.umass.edu/lfw/`

[3]`www.mturk.com`

Age group	0–2	3–7	8–12	13–19	20–36	37–65	66+	Total
Female	439	771	378	956	7,767	3,604	644	14,559
Male	515	824	494	736	7,281	3,213	609	13,672
Total	954	1,595	872	1,692	15,048	6,817	1,253	28,231

(a) *Images of Groups* (IPD: 26.5 ± 17.5, FDR: 99.0%, FQS: 4.3 ± 1.7)

Age group	0–20	21–40	41–60	61+	Total
Female	114	1,685	1,011	165	2,975
Male	95	2,501	5,021	2,641	10,258
Black	17	532	354	219	1,122
White	169	3,368	5,140	2,368	1,1045
Asian	23	284	537	219	1,063
Unknown	0	2	1	0	3
Total	209	4,186	6,032	2,806	13,233

(b) LFW (IPD: 60.0 ± 0.2, FDR: 99.7%, FQS: 4.6 ± 1.7)

Age group	0–20	21–40	41–60	61+	Total
Female	1,248	1,685	1,011	165	4,109
Male	1,427	2,501	5,021	2,641	11,590
Black	40	532	354	219	1,145
White	1,497	3,368	5,140	2,368	12,373
Asian	1,126	284	537	219	2,166
Unknown	12	2	1	0	15
Total	2,675	4,186	6,032	2,806	15,699

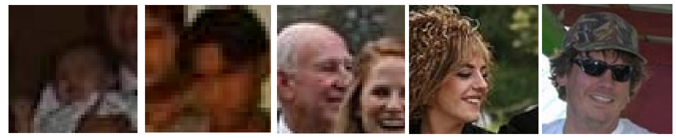
(c) LFW+ (IPD: 43.7 ± 8.0, FDR: 99.6%, FQS: 4.6 ± 1.6)

Age group	0–20	21–40	41–60	61+	Total
#Images	730	212	53	7	1,002

(d) FG-NET (IPD: 122.1 ± 16.0, FDR: 99.8%, FQS: 4.8 ± 1.5)

TABLE I

THE AGE GROUP, GENDER, AND RACE DISTRIBUTIONS OF SUBJECTS IN THE (*a*) *Images of Groups*, (*b*) LFW , (*c*) LFW+, AND (D) FG-NET DATABASES, AND VALUES OF THE INTERPUPILLARY DISTANCE (IPD, IN PIXELS), AND FACE DETECTION RATE (FDR) AND FACE QUALITY SCORE (FQS) BY PITTPATT BASED ON 1,000 RANDOMLY SELECTED FACE IMAGES FROM EACH DATABASE. AS A COMPARISON, THE IPD, FDR AND FQS OF THE MORPH DATABASE (1,000 RANDOMLY SELECTED FACE IMAGES) ARE 95.5 ± 24.0, 100%, AND 5.7 ± 1.3, RESPECTIVELY.



(a) Face images from the Images of Groups database

(b) Face images from the LFW database

(c) Face images from the LFW+ database (the extended part)

(d) Face images from the FG-NET database

Fig. 2. Examples of unconstrained face images from the (a) *Images of Groups* database (total of 28,231 face images), (b) LFW database (total of 13,233 face images), (c) the extended part in LFW+ databases (2,466 images of subjects in age group 0–20), and (d) FG-NET database (total of 1,002 face images).

gender, and race estimation.

The main contribution of this paper is to estimate age, gender and race from unconstrained face images. Most of the previous work focused only on MORPH and FG-NET databases. This is the first paper that gives complete and detailed results on age (EXACT age and age group), gender and race estimation for the LFW database. Novelty of this work consists of (i) cascade face normalization, and (ii) utilizing face context information. Face normalization handles pose and illumination variations in unconstrained faces. Additional contributions: (i) LFW+ database with EXACT age, gender, and race for each subject, and (ii) demographic estimation under cross-database scenarios.

II. Related Work

A. Unconstrained Demographic Face Database

To our knowledge, the most commonly used public-domain unconstrained face database for demographic estimation is the *Images of Groups* database [8], which is a collection of face images of groups of people (group images) from Flickr. In all, there are 5,080 group images containing 28,231 faces. All the face positions in the group images are provided; 86% of the faces are detected automatically, and the remaining localized manually. Age group and gender of each face image were manually labeled. Seven age categories were used: 0–2, 3–7, 8–12, 13–19, 20–36, 37–65, and 66+. This dataset is very challenging for demographic estimation tasks as many of the face images have low image resolution (the median IPD is only 19 pixels, and 25% of the faces have an IPD smaller than 13 pixels), and many face images have large facial appearance variations due to pose, illumination, etc.

In addition to the *Images of Groups* database, we also use the well-known Labeled Faces in the Wild (LFW) database [14] for unconstrained age, gender, and race estimation. Since the number of subjects in age group 0–20 in the LFW database is only 209 (see Table I (b)), we extended the LFW database by collecting 2,466 unconstrained face images in the age range 0–20 using Google Images search service[5]. Age group $[0, 20]$ is useful (e.g., vending machines that dispense alcohol/cigarette) for binary age classification ($<$ 21 vs. 21+). The extended LFW database (LFW+) contains 15,699 unconstrained face images of about 8,000 subjects.

Age, gender and race information is not provided for the face images in LFW+ database. Therefore, we utilized the

[4]The subjects in the LFW database are mainly celebrities, public figures, etc. which explains why most of the subjects are in the age range 30–70. We populate the LFW database by collecting 2,466 unconstrained face images from the Internet in the age range 0–20.

[5]We first use the keywords such as "baby", "kid", and "teenagers" to find $\sim$5,000 images of interest from Google Images. The Viola-Jones [15] face detector is then applied to generate a set of candidate faces. Finally, we manually remove false face detections as well as subjects that appear to be older than 20.

Publication	Approach (feature extraction and classification)	Face database #images (training; testing)	Classification accuracy
Gallagher and Chen [8]	Raw intensity, face position, image structure, and social relationship; Gaussian Maximum Likelihood	*Images of Groups* Age group: $(3,500;\ 1,050)$ Gender: $(23,218;\ 1,881)$	Age group[1]: 42.9% Gender: 74.1%
Kumar *et al.* [9]	Color, raw intensity, edge magnitude, and gradient direction; Support Vector Machines	Columbia Face Database[2] (n/a, n/a)	Binary age group[3]: 80.3%–93.0% Gender: 85.8% Binary race group[3]: 91.5%–94.6%
Demirkus *et al.* [10]	SIFT features and temporal relationship; Bayesian temporal model	McGill Videos Database[4] $(4,450;\ 9,000$ frames)	Gender: 90%
Shan [11]	Local Binary Patterns and Gabor features; Support Vector Machines	*Images of Groups* Age group: $(2,080;\ 664)$ Gender: $(9,336;\ 2,744)$	Age group[1]: 55.9% Gender: 77.4%
Ylioinas *et al.* [12]	Variants of Local binary patterns; Support Vector Machines	*Images of Groups* $(3,500;\ 1,050)$	Age group[1]: 51.7%
Alnajar *et al.* [13]	Orientation histogram of local gradients; Learning-based encoding	*Images of Groups* $(2,080;\ 664)$	Age group[1]: 56.5%
Proposed method	Pose and illumination correction, and biologically inspired features (BIF); Support Vector Machines	*Images of Groups* $(22,585;\ 5,646)$[5] LFW+ $(12,559;\ 3,140)$[5] FG-NET (990; 12)[5]	*Images of Groups* Age group[1]: 68.1% Gender: 87.1% LFW+ Age group[1]: 66.7%, Gender: 95.4%, Race: 90.8% FG-NET Exact age: 4.5 years MAE

TABLE II

A SUMMARY OF PUBLISHED METHODS ON DEMOGRAPHIC (AGE, GENDER AND RACE) ESTIMATION FROM UNCONSTRAINED FACE IMAGES.

[1]Age group classification accuracies reported on the *Images of Groups* and LFW+ databases are based on seven (0–2, 3–7, 8–12, 13–19, 20–36, 37–65, and 66+) and four (0–20, 21–40, 41–60, and 61+) age groups, respectively. [2]Two subsets (FaceTracer[8] and PubFig[9]) of the Columbia Face database were made publicly available; however many URLs for downloading the face images of the FaceTracer dataset are no longer active, and human labels for age, gender and race attributes of the Pubfig dataset are not provided with the face images. [3]Performance is reported for binary group classifications, such as baby vs. non-baby, and Asian vs. non-Asian, but the overall accuracies for age group and race classifications are not reported. [4]While the website indicates that this database will be made publicly available, at the time of writing this paper, it is not yet available. [5]We perform evaluations using the entire database with 5-fold cross-validation; the numbers of training and testing images reported here are the average in one-fold test.

MTurk crowdsourcing service to collect the human labels for exact age, gender, and race of each face image in the LFW+ database. To determine how many human workers are enough for each Human Intelligence Task (HIT), we randomly selected 100 images from the FG-NET database [16] (with known ground-truth age), and asked 10 MTurk workers to provide age estimates. We then randomly selected 3, 6, and 9 human estimates from these 10 MTurk estimates and calculated their MAEs of age estimation. Our analysis shows that the age estimates based on 3, 6, and 9 MTurk workers are very close to those by all the 10 MTurk workers.[6] Therefore, we utilized only 3 MTurk workers per HIT. For age group classification on the LFW+ database, we consider four age groups: 0–20, 21–40, 41–60, and 60+, which generally correspond to four life stages of a person, *i.e.,* young, young adult, middle adult, and old adult.[7] Another reason why different age groups are considered is that they are application dependent in practice.

We also perform age estimation on the public-domain FG-NET database [16] to evaluate the generalization ability of the proposed approach to semi-constrained sensing scenarios. Age, gender and race distributions of the subjects in the *Images of Groups*, LFW, LFW+ and FG-NET databases are listed in Table I, and example face images from these databases are shown in Fig. 2. Values (in Table I) of the interpupillary distance (IPD), and face detection rate (FDR) and face quality score (FQS) by a Commercial Off-The-Shelf (COTS) SDK (PittPatt) based on $1,000$ randomly selected face images from each database show that the *Images of Groups* and LFW+ databases are more challenging than the FG-NET and MORPH face databases for automatic demographic estimation.

B. Literature Review

We summarize published methods for demographic estimation from unconstrained face images in Table II. To the best of our knowledge, Gallagher and Chen [8] were the first to study demographic (age group and gender) estimation from unconstrained face images. Both facial features (*e.g.,* raw pixel values) and contextual features (*e.g.,* absolute and relative face positions, image structure, juxtaposition of neighboring faces) were used in [8], and a Gaussian Maximum Likelihood (GML) classifier was learnt to perform age group and gender classification. However, race classification results are not available because the race information is not provided with the subjects in the *Images of Groups* database.

[6]The MAEs of age estimation by 3, 6, 9, and 10 MTurk workers are 4.7, 4.4, 4.3, and 4.3 years, respectively.

[7]`http://en.wikipedia.org/wiki/Young_adult_(psychology)`

[8]`www.cs.columbia.edu/CAVE/databases/facetracer`

[9]`www.cs.columbia.edu/CAVE/databases/pubfig`

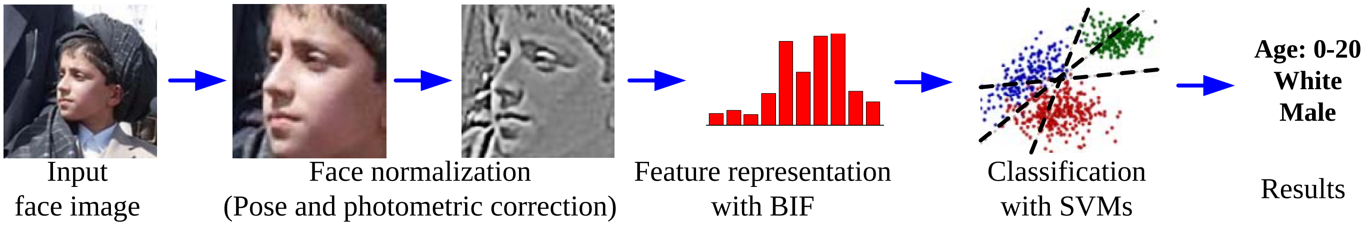


Fig. 3. Proposed approach for demographic (age, gender, and race) estimation from unconstrained face images consists of three main stages: face normalization for pose and illumination correction, feature representation with biologically inspired features (BIF), and demographic classification using SVM classifiers.

Kumar *et al.* [9] introduced the use of describable visual attributes for face verification and image search, where age group, gender, and race were defined as three of the 73 attributes. Image color, intensity, edge magnitude and edge orientation were extracted from a face image, followed by feature normalization and aggregation. Per attribute SVM classifiers were trained to determine whether a face image of a subject possesses a particular attribute, *e.g.,* baby or not a baby. However, only binary classification accuracies were reported for age group classification, and only a small subset of $5,000$ face images (the FaceTracer database[8]) from their Columbia Face database was released with age, gender and race information labeled by MTurk workers.

The *Images of Groups* database released in [8] has facilitated the research on demographic estimation from unconstrained face images [11]–[13]. Shan [11] proposed to use LBP and Gabor features with SVM classifiers to perform age group and gender classification. Ylioinas *et al.* [12] used a combination of local binary pattern (LBP) variants with a SVM classifier to perform age group classification. Alnajar *et al.* [13] used code learning with low-level features such as the orientation histogram of local gradients to perform age group classification. Demirkus *et al.* [10] studied the problem of gender classification on their own proprietary database consisting of unconstrained video sequences of 30 subjects. However, all these methods except for [11] only studied the estimation of a single demographic attribute using relatively small face datasets (see Table II).

III. Proposed Approach

The proposed approach for age group, gender, and race estimation from unconstrained face images consists of three main components (Fig. 3): (i) face normalization consisting of pose and photometric corrections, (ii) feature representation with biologically inspired features (BIF), and (iii) demographic estimation using SVM.

A. Pose and Illumination Correction

We apply 2D affine transformation based on the two eye locations to correct pose variations in unconstrained face images. Given an input 2D face images with eye locations (Fig. 4), *e.g.,* $p_1(x_1, y_1)$ and $p_2(x_2, y_2)$, and a target pose-corrected face image with fixed eye locations, *e.g.,* $p'_1(x'_1, y'_1)$ and $p'_2(x'_2, y'_2)$, the 2D affine transformation can be formulated

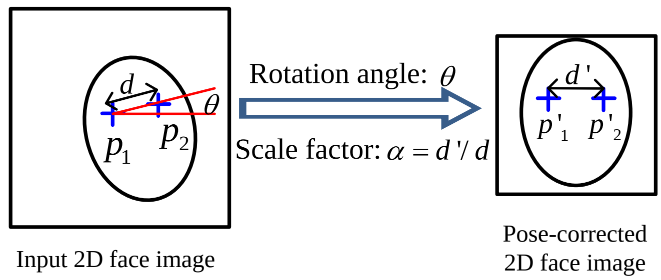


Fig. 4. Pose correction using 2D affine transformation based on the two eye locations.

as

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \alpha \ R \begin{bmatrix} x \\ y \end{bmatrix} = \alpha \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix},$$

where the 2D rotation matrix R is determined by the rotation angle θ, and scaling factor $\alpha = d'/d$. 2D affine transformation based pose correction is able to handle the in-plane rotations, and the pose-corrected face images are supposed to be upright with the eyes in fixed positions. One possible reason is that 3D models constructed from unconstrained face images introduce facial artifacts during the texture mapping process.

Illumination variations due to shadow and underexposure can also lead to significant changes in facial appearance. As a result, the intra-person differences due to large illumination variations may even be larger than inter-person differences. Following the success of Difference of Gaussians (DoG) filtering based illumination normalization methods in face recognition applications [17], [18], we also use DoG filtering ($\sigma_1 = 1.0, \sigma_2 = 0.2$) to perform photometric correction on the pose-corrected face images. Additionally, DoG filtering also increases the visibility of facial details present in a blurred face image.

Figure 5 shows the pose and photometric correction results for two unconstrained face images of one subject from the LFW+ database. We find that the normalized face images are well aligned, and illumination variations due to shadow have been effectively suppressed.[10] We also tried to align individual face images using a COTS 3D face modeling system. Figure 6 shows the alignment of two unconstrained face images where the 2D affine transformation does not work very well, but the 3D modeling method provides really good alignment. For example, although eyes in the two aligned face images by 2D affine transformation are in the same locations, the remaining

Fig. 5. Face normalization for two face images of one subject from the LFW+ database with pose and photometric corrections.

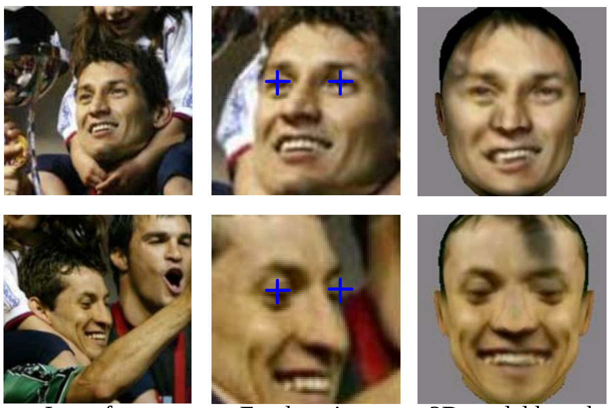


Fig. 6. Alignment results of two unconstrained face images of one subject from the LFW+ database by (i) 2D affine transformation (based on eye locations) and (ii) 3D face model (A COTS 3D modeling system, CyberExtruder Aureus 3D SDK[11], is used in our approach).

parts of the face are misaligned with each other. By contrast, the two aligned face images by 3D model have much better correspondence with each other in most of the facial area.

B. Feature Representation

In constrained face acquisition conditions, published methods often utilize the central facial region (Fig. 7 (b)) to perform demographic estimation [4], [5], [19]. However, in unconstrained conditions, the contextual region around a face may complement the central facial region by providing additional information such as face contour, ears, hair style, etc. While relative and absolute face positions, image structure, and social relationship in a group photograph were defined as contextual features in [8], they are not useful for demographic estimation from a single face image. We do not explicitly partition a face image into facial and contextual regions, and then extract features from them. Instead, we utilize the same

Fig. 7. An input 2D face images (a, 250×250) and its facial region determined by tight cropping (b, 60×60), and facial region with context determined by loose cropping (c, 60×60). Contextual information such as facial contour, ears, and part of the hair style is included in (c).

face descriptor, biologically inspired feature (BIF) descriptor [19], to extract features from "facial region with context" (Fig. 6 (c)).[11] In the first layer of BIF, we apply Gabor filtering to a normalized face image with 12 scales and 8 directions. We then apply a max pooling operator between every two successive scales to aggregate the features, resulting in the second layer features with 6 scales and 8 directions. Finally, all the features in the second layer are concatenated into a single feature vector.

C. Classification Method

We use three different SVM classifiers with RBF kernel to perform age group, gender, and race classifications. For exact age estimation, we follow a coarse-to-fine approach, which has been shown to achieve better accuracy than direct age regression [5], [20]. For SVM implementation, we use a publicly available LIBSVM library [21], where parameters γ and c are optimized using a grid search on the training set.

The same features are used in all three SVMs. Labels for training age group, gender, and race classifiers are $0, 1, \cdots, N$, $0, 1$, and $0, 1$, respectively; N is the number of age groups minus 1.

IV. Experimental Results

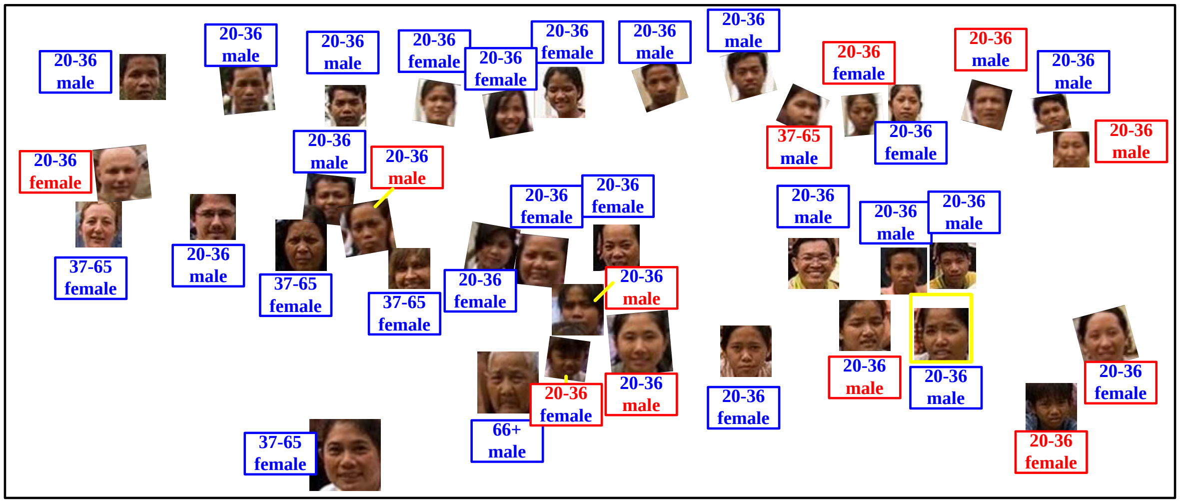
We perform age group and gender classifications on the *Images of Groups* [8] database, and age group, gender and race classifications as well as exact age estimation on the LFW+ database. On each face database, the proposed demographic estimation approach is evaluated using a five-fold cross-validation protocol.

Generalization ability of the proposed approach is evaluated through cross-database age group and gender classification between the *Images of Groups* and LFW+ databases, and the exact age estimation on the FG-NET database.

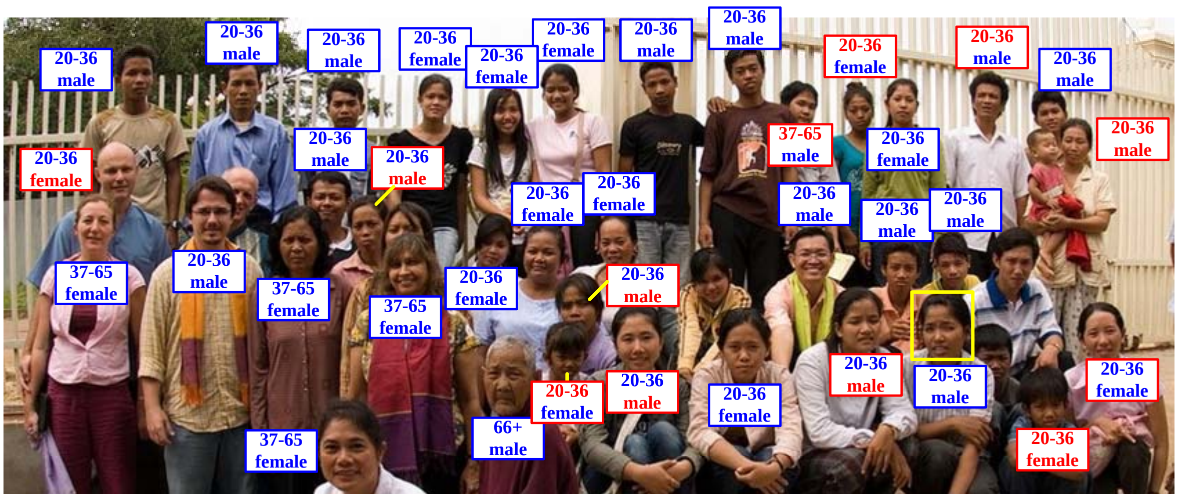
For age group, gender, and race classification tasks, we report the confusion matrix and the overall classification accuracy. For exact age estimation, we report the cumulative score (CS) [22] curve and the mean absolute error (MAE) [23].

[10] Image color was used for attribute classification in [9]; however, we notice image (skin) color is not always reliable due to improper "white" balance. Thus, we first convert color face images into gray-scale and then perform photometric correction on the gray-scale face images.

[11] Age group classification accuracy on LFW+ database using facial region with context is 2.1% higher than that using only the facial region.



(a) Individual face images used as input to the proposed approach



(b) The complete image containing individual subjects

Fig. 8. Examples of correct (shown in blue) and incorrect (shown in red) age group and gender classifications for one image from the *Images of Groups* database. The ground-truth gender of a female subject (yellow rectangle) in this image has been mislabeled as male.

A. *Results on Images of Groups Database*

The performance of the proposed methods for age group and gender classifications on the *Images of Groups* database is shown in Table III. For age group classification, subjects in age groups 8–12, 13–19, and 37–65 are found to be easily confused with subjects in their neighboring age groups, particularly the 20–36 age group (see Table III (a)). We attribute this to the following reasons: (i) the number of subjects in different age groups in the *Images of Groups* database is significantly imbalanced (*e.g.*, 53% of all the subjects are in the age group 20–36), and (ii) age groups of the subjects labeled by humans can be incorrect (see the face image highlighted with yellow rectangle in Fig. 8) and inconsistent.[12] Low image resolution in the *Images of Groups* database is another challenge that is difficult to compensate for by our face normalization approach. However, the proposed approach still achieves significantly better overall age group classification accuracy (68.1%) than the state-of-the-art methods (56.5% reported in [13]). Again, while only a subset (2,744 face images) of the *Images of Groups* database was used in [13], our results are based on the entire database (28,231 face images).

The best known gender classification accuracy reported on the *Images of Groups* database is 77.4% [11] (see Table II). The proposed approach achieves a much higher accuracy of 87.1%. Additionally, while only a subset (12,080 face images) of the *Images of Groups* database was used in [11], our results are based on the entire database (28,231 face images). Regarding the per gender accuracy, the misclassification error of females (14.1%) is higher than that of males (11.6%).

Figure 8 shows the age group and gender classification results on one group photograph from the *Images of Groups* database. There are 37 subjects in this image whose face locations, age group and gender are provided in the *Images of Groups* database. Among these 37 subjects, the proposed approach gives correct age group and gender classifications for 31 and 30 subjects, respectively.

[12] By inconsistent, we mean that (i) different subjects in the same age group can be categorized into different age groups, and (ii) the same face image can be categorized into different age groups by different people.

Overall Acc. **68.1±0.2**	0–2	3–7	8–12	13–19	20–36	37–65	66+
0–2	**72.6**	22.2	0.2	0.2	4.5	0.3	0.0
3–7	10.0	**66.5**	6.5	2.0	13.6	1.3	0.1
8–12	1.1	35.4	**17.4**	6.5	36.0	3.1	0.5
13–19	0.1	4.0	4.1	**7.6**	79.9	4.1	0.2
20–36	0.0	0.5	0.2	0.4	**90.2**	8.5	0.2
37–65	0.1	0.5	0.2	0.1	49.8	**46.4**	2.9
66+	0.0	0.4	0.2	0.2	11.8	49.2	**38.2**

(a) Confusion matrix for age group classification

Overall Acc. **87.1±0.4**	Female	Male
Female	**85.9**	14.1
Male	11.6	**88.4**

(b) Confusion matrix for gender classification

TABLE III

CONFUSION MATRICES OF THE PROPOSED APPROACH FOR (a) AGE GROUP, AND (b) GENDER CLASSIFICATION ACCURACIES (IN %) FOR THE *Images of Groups* DATABASE USING 5-FOLD CROSS-VALIDATION.

Overall Acc. **66.7±1.1**	0–20	21–40	41–60	61+
0–20	**82.8**	13.7	3.0	0.5
21–40	4.5	**71.6**	23.0	0.9
41–60	1.1	16.8	**70.9**	11.2
61+	0.3	2.4	55.2	**42.1**

(a) Confusion matrix for age group classification

Overall Acc. **95.4±1.1**	Female	Male
Female	**87.3**	12.7
Male	2.3	**97.7**

(b) Confusion matrix for gender classification

Overall Acc. **90.8±2.8**	White	Other
White	**95.5**	4.5
Other	33.2	**66.8**

(c) Confusion matrix for race classification

TABLE IV

CONFUSION MATRICES OF THE PROPOSED APPROACH FOR (a) AGE GROUP, (b) GENDER, AND (c) RACE CLASSIFICATION ACCURACIES (IN %) ON THE LFW+ DATABASE USING 5-FOLD CROSS-VALIDATION.

B. Results on LFW+ Database

The performance of the proposed methods for age group, gender and race classification on the LFW+ database is shown in Table IV. For age group classification, the subjects in the age group 61+ tend to be confused with subjects in age group 41–60 (see Table IV (a)). One possible explanation for this misclassification is that most of the subjects (72%) in LFW+ are public figures that are in the age range 30–70, who tend to use facial makeup, resulting in their younger appearance than their real age. The proposed approach nevertheless achieves an overall age group classification accuracy of 66.7% with a standard deviation of 1.1 across five folds. To the best of our knowledge, no results have been reported for age group classification with more than two groups for the LFW database. Considering that face images in both the *Images of Groups* and LFW+ databases are unconstrained, age group classification accuracies of 66.7% and 68.1%, in our opinion, are quite good.[13]

For gender classification on the LFW+ database, the proposed approach achieves an impressive accuracy of 95.4% with a standard deviation of 1.1 across five folds. Similar to the gender classification results on the *Images of Groups* database, misclassification rate of females (12.7%) is higher than that of males (2.3%). While no gender classification accuracy has been reported in the literature on the LFW database, gender classification results reported on other unconstrained face databases by the state-of-the-art methods can be used as a reference. For example, the gender classification accuracy reported on the Columbia Face database (with many of the same types of variations as the LFW database) is 85.8% [9], which is 10% lower than that achieved by the proposed approach on LFW+ database.

Due to the significantly biased race distribution in the LFW+ database (78.8% of the subjects are white (see Table I)), we perform only binary race classification: White vs. Other (Black, Asian, and Unknown). However, training for multi-class race classifier is straightforward. The proposed approach achieves a 2-class race classification accuracy of 90.8% with a standard deviation of 2.8 across five folds. Misclassifications of subjects in the "other" race (33.2%) are higher than that for white subjects (4.5%). We attribute this to (i) the significantly imbalanced race distribution ($12,373$ whites vs. $3,326$ "others"), and (ii) large within-group diversity of the "other" race group. To our knowledge, there are no race classification results reported on the LFW database. Again, race classification accuracy on the Columbia Face database (91.5% for White v. rest) can be used as a reference. This is similar to the accuracy of the proposed approach for White vs. Other classification on the LFW+ database (90.8%).

Figure 11 (a) shows the cumulative score (CS) of age estimation by the proposed approach. Given a tolerance of 7.5-year absolute error, ages of 64% of the face images in the LFW+ database can be correctly estimated. The MAE of age estimation on LFW+ database by the proposed approach is 7.3 years, which is significantly higher than the MAE on constrained face databases (*e.g.,* 4.0 years MAE on the MORPH database reported in [4]). This experiment shows that exact age estimation from unconstrained face images is a very difficult problem.

Examples of correct and incorrect age, gender, and race estimates by the proposed approach on the LFW+ database are shown in Fig. 9. We notice inaccurate labels provided by MTurk workers. As an example, for the second face image in Fig. 9 (b), the gender of the female baby was labeled as male.

We also perform age group, gender, and race classification on the LFW database following the standard 10-fold protocol [14]. Age group, gender, and race classification accuracies

[13] We also consider different age categories, such as 0–14, 15–24, 25–44, 45–64, and 65+ suggested by the United Nations (`http://unstats.un.org/unsd/publication/SeriesM/SeriesM_74e.pdf`). The proposed approach achieves an overall classification accuracy of 64.1% for these five age groups on the LFW+ database.

[14] `http://www.cs.columbia.edu/CAVE/databases/pubfig/download/lfw_attributes.txt`

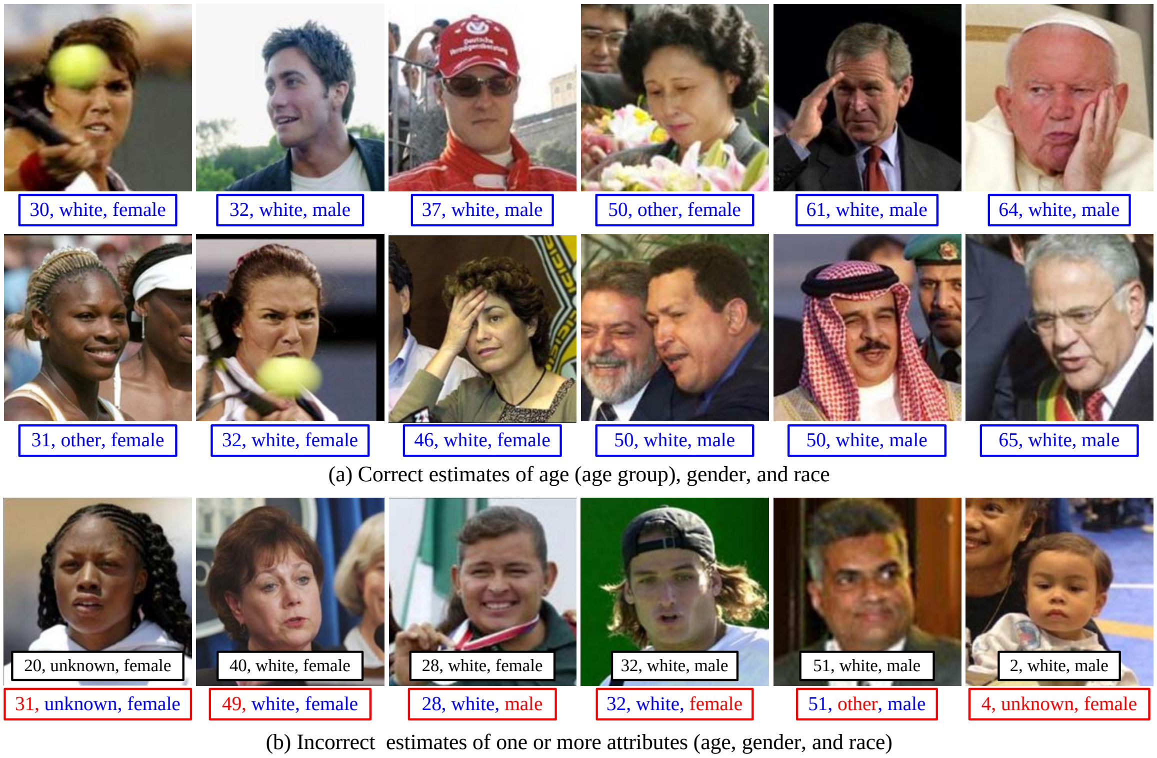


Fig. 9. Examples of correct (a) and incorrect (b) age, gender and race estimations by the proposed approach on the LFW+ database. For face images with one or more incorrect attribute estimation (shown in red) in (b), their true (crowdsourced) attributes are also shown within individual face images.

for just the LFW database using the LFW 10-fold protocol are 64.8%, 93.8%, and 88.4%, with std. of 1.9, 1.2, and 1.4, respectively. While age group, gender and race classification results on the LFW database were given in [9][14], only the estimates of age group, gender and race were provided, but NOT the ground-truth. So, age group, gender and race classification accuracies of [9] on LFW are not known.

By using 3D face modeling method to do the face alignment, we further achieve about 1% improvement in age group classification on the LFW+ database. Examples where 3D model based face alignment leads to correct and incorrect age group classifications of subjects in the LFW+ database are shown in Fig. 10. While 3D model based face alignment improves the overall age group classification accuracy, 3D face models reconstructed from kid's face image or low-resolution face images may introduce artificial facial textures, resulting incorrect age group classifications.

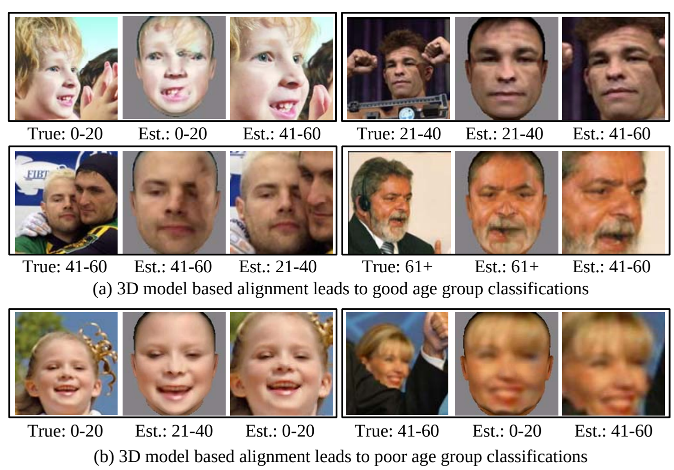


Fig. 10. Examples where 3D model based face alignment leads to correct (a) and incorrect (b) age group classifications of subjects in the LFW+ database. The three face images of each subject are the original 2D face image, aligned 2D face image using 3D model, and aligned 2D face image using 2D affine transformation, respectively.

C. Generalization Ability

We evaluate the generalization ability of the proposed approach from the following aspects: (i) cross-database testing between *Images of Groups* and LFW+ databases;[15] and (ii) evaluation of the proposed approach on the public-domain FG-NET database.

On the *Images of Groups* database, the proposed approach with cross-database testing achieves accuracies of 51.8% and 78.2% for age group and gender classification tasks, respectively. On the LFW+ database, the proposed approach with cross-database testing achieves accuracies of 48.0% and 88.1% for age group and gender classification tasks, respectively. As expected, the performance with cross-database testing is significantly lower than that with intra-database testing. However, we believe these cross-database testing results (not reported in other published methods) are reasonable, particularly when

[15]The demographic estimation method is trained on one face database, and tested on a different face database.

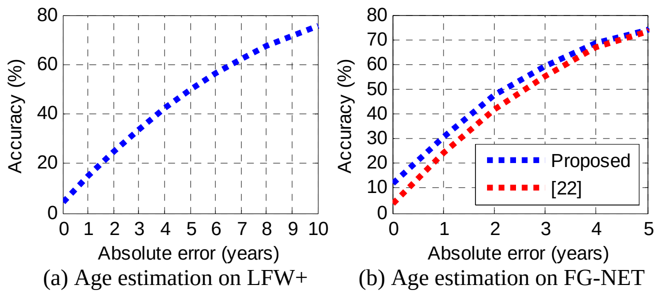


Fig. 11. Cumulative scores of age estimation by the proposed approach on (a) the unconstrained LFW+ database, and (b) the semi-constrained FG-NET database. The MAEs of the age estimate on LFW+ and FG-NET databases are 7.5, and 4.5 years, respectively.

compared to the intra-database testing results in Table II.

It is useful to determine how the proposed approach generalizes to other face databases with less-constrained sensing conditions, such as the public-domain FG-NET database.[16] We perform exact age estimation on the FG-NET database following the widely used leave-one-person-out (LOPO) protocol [2], [3], [24]. While the state-of-the-art methods reported a lower MAE (4.1 years [24] with LOPO protocol) than the proposed approach (4.5 years MAE), the propose approach performs consistently better than [24] in an operational absolute error range of 0–5 years (see Fig. 11 (b)).

V. Conclusions

This paper addresses the problem of automatic age, gender, and race estimation from real-life face images acquired in unconstrained conditions. We perform pose and photometric corrections to normalize an input unconstrained face image. Biologically inspired features are then extracted from the normalized face image, including both the central face region and the surrounding context region. Support Vector Machine classifiers are used to predict the age group (or exact age), gender, and race of a subject.

We evaluate the proposed approach on the public-domain *Images of Groups* and FG-NET databases and an extended version (LFW+) of the LFW database, where about $2,500$ unconstrained face images of subjects in the age group 0–20 are added to the LFW database. Age, gender, and race information of each subject in the LFW+ database has been collected using Amazon Mechanical Turk crowdsourcing service. Experimental results show that demographic estimation from unconstrained face images remains a very difficult problem; the proposed approach, nevertheless, significantly outperforms the state-of-the-art methods in unconstrained scenarios, and generalizes well to the scenarios of cross-database testing and semi-constrained sensing conditions.

References

[1] S. Z. Li and A. K. Jain, editors. *Handbook of Face Recognition*. New York: Springer, 2nd edition, 2011.
[2] Y. Fu, G. Guo, and T.S. Huang. Age synthesis and estimation via faces: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(11):1955–1976, Nov. 2010.
[3] X. Geng, C. Yin, and Z.-H. Zhou. Facial age estimation by learning from label distributions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(10):2401–2412, Oct. 2013.
[4] G. Guo and G. Mu. Joint estimation of age, gender and ethnicity: CCA vs. PLS. In *Proc. FG*, 2013.
[5] H. Han, C. Otto, and A. K. Jain. Age estimation from face images: Human vs. machine performance. In *Proc. ICB*, 2013.
[6] M. Ngan and P. Grother. Performance of automated age estimation algorithms. Technical Report NIST Interagency Report 7995, 2014.
[7] K. Ricanek and T. Tesafaye. MORPH: A longitudinal image database of normal adult age-progression. In *Proc. FG*, 2006.
[8] A.C. Gallagher and T. Chen. Understanding images of groups of people. In *Proc. IEEE CVPR*, 2009.
[9] N. Kumar, A.C. Berg, P.N. Belhumeur, and S.K. Nayar. Describable visual attributes for face verification and image search. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(10):1962–1977, Oct. 2011.
[10] M. Demirkus, M. Toews, J.J. Clark, and T. Arbel. Gender classification from unconstrained video sequences. In *Proc. CVPRW*, 2010.
[11] C. Shan. Learning local features for age estimation on real-life faces. In *Proc. ACM Workshop on MPVA*, 2010.
[12] J. Ylioinas, A. Hadid, and M. Pietikäinen. Age classification in unconstrained conditions using LBP variants. In *Proc. ICPR*, 2012.
[13] F. Alnajar, C. Shan, T. Gevers, and T. Geusebroek. Learning-based encoding with soft assignment for age estimation under unconstrained imaging conditions. *Image Vision Comput.*, 30(12):946–953, Dec. 2012.
[14] G.B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, 2007.
[15] P. Viola and M.J. Jones. Robust real-time face detection. *Int. J. Comput. Vision*, 57(2):137–154, May 2004.
[16] Face and Gesture Recognition Working group. FG-NET aging database, 2002. `http://www-prima.inrialpes.fr/FGnet/html/benchmarks.html`.
[17] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. Image Process.*, 19(6):1635–1650, Jun. 2010.
[18] H. Han, S. Shan, X. Chen, and W. Gao. A comparative study on illumination preprocessing in face recognition. *Pattern Recognit.*, 46(6):1691–1699, 2013.
[19] G. Guo, G. Mu, Y. Fu, and T.S. Huang. Human age estimation using bio-inspired features. In *Proc. IEEE CVPR*, 2009.
[20] S.E. Choi, Y.J. Lee, S.J. Lee, K.R. Park, and J. Kim. Age estimation using a hierarchical classifier based on global and local facial features. *Pattern Recogn.*, 44(6):1262–1281, Jun. 2011.
[21] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2:1–27, 2011.
[22] X. Geng, Z.-H. Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(12):2234–2240, Dec. 2007.
[23] A. Lanitis, C.J. Taylor, and T.F. Cootes. Toward automatic simulation of aging effects on face images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):442–455, Apr. 2002.
[24] K. Luu, K. Seshadri, M. Savvides, T.D. Bui, and C.Y. Suen. Contourlet appearance model for facial age estimation. In *Proc. IJCB*, 2011.

[16] The FG-NET database can be considered as a semi-constrained database because the pose, expression and illumination variations are somewhat constrained.