

Article

# Demographic-Assisted Age-Invariant Face Recognition and Retrieval

Muhammad Sajid <sup>1,\*</sup>, Tamoor Shafique <sup>2</sup>, Sohaib Manzoor <sup>3</sup> , Faisal Iqbal <sup>1</sup>, Hassan Talal <sup>1</sup>, Usama Samad Qureshi <sup>1</sup> and Imran Riaz <sup>1</sup>

<sup>1</sup> Department of Electrical Engineering, Mirpur University of Science and Technology, Mirpur 10250 (AJK), Pakistan; faisal.iqbal84@gmail.com (F.I.); hassan.ee@must.edu.pk (H.T.); usama.ee@must.edu.pk (U.S.Q.); imran.ee@must.edu.pk (I.R.)

<sup>2</sup> Faculty of Computing, Engineering and Science, Staffordshire University, Stoke-on-Trent ST4 2DE, UK; t.shafique@ieee.org

<sup>3</sup> School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China; sohaibmanzoor@hust.edu.cn

\* Correspondence: sajid.ee@must.edu.pk; Tel.: +92-312-953-1096

Received: 17 April 2018; Accepted: 4 May 2018; Published: 8 May 2018



**Abstract:** Demographic estimation of human face images involves estimation of age group, gender, and race, which finds many applications, such as access control, forensics, and surveillance. Demographic estimation can help in designing such algorithms which lead to better understanding of the facial aging process and face recognition. Such a study has two parts—demographic estimation and subsequent face recognition and retrieval. In this paper, first we extract facial-asymmetry-based demographic informative features to estimate the age group, gender, and race of a given face image. The demographic features are then used to recognize and retrieve face images. Comparison of the demographic estimates from a state-of-the-art algorithm and the proposed approach is also presented. Experimental results on two longitudinal face datasets, the MORPH II and FERET, show that the proposed approach can compete the existing methods to recognize face images across aging variations.

**Keywords:** demographic estimation; facial asymmetry; face recognition; retrieval

## 1. Introduction

Recognition of face images is an important yet challenging problem. This challenge mainly includes pose, illumination, expression, and aging variations. Recognizing face images across aging variations is called age-invariant face recognition. Despite competitive performance of some existing methods, recognizing and retrieving face images across aging variations remains a challenging problem. Facial aging is a complex process owing to variations in the balance, proportions, and symmetry of the face with varying age, gender, and race [1,2]. More precisely, the age, gender, and race are correlated in characterizing the facial shapes [3]. In our routine life, we can estimate the age, gender, and race of our peers quite effectively and easily. However, a number of anthropometric studies such as [3] suggest significant facial morphological differences among race, gender, and age groups. The study suggests that the anthropometric measurements for female subjects are smaller than the corresponding measurements for male subjects. Another study [4] reported the broader faces and noses for Asians compared to North American Whites. Similarly, some studies on facial aging such as [5] suggest a change in facial size with change in age during craniofacial growth, with adults developing a triangular facial shape with some wrinkles. Some studies on sexual dimorphism reveal more prominent features for male faces compared to female faces [6]. Some studies suggest that different

populations exhibit different bilateral facial asymmetry. A relationship between facial asymmetry and sexual dimorphism has been revealed in [7]. In [8], it has been observed that facial masculinization covaries with bilateral asymmetry in males' faces. Facial asymmetry increases with increasing age, as observed in [9,10].

In this paper, we present demographic-assisted recognition and retrieval of face images across aging variations. More precisely, the proposed approach involves: (i) facial-asymmetry-based demographic estimation and (ii) demographic-assisted face recognition and retrieval. To this end, we first estimated the age group, gender, and race of a query face image using facial-asymmetry-based, demographic-aware features learned by convolutional neural networks (CNNs). Face images were then recognized and retrieved based on a demographic-assisted re-ranking approach.

The motivation of this study is to answer the following questions.

- (1) Does demographic-estimation-based re-ranking improve the face identification and retrieval performance across aging variations?
- (2) What is the individual role of gender, race, and age features in improving face recognition and retrieval performance?
- (3) Do deeply learned asymmetric features perform better or worse compared to existing handcrafted features?

We organize the rest of this paper as follows. Existing methods on demographic estimation and face recognition are presented in Section 2. The facial-asymmetry-based demographic estimation approach is illustrated in Section 3. Section 4 presents the demographic-assisted face recognition and retrieval approach with experiments and results. Discussion on the results is presented in Section 5, while the last section concludes this paper.

## 2. Related Works

Demographic estimation, face recognition, and retrieval have been extensively studied in the literature. In the following subsections, we provide an overview of some existing methods related to our work.

### 2.1. Demographic Estimation

The problem of demographic estimation has been studied extensively in the literature [11–23]. Existing demographic estimation approaches can be grouped into three main categories: landmarks-based approaches, texture-based approaches, and appearance-based approaches. Landmarks-based approaches, such as [11], use the distance ratios between facial landmarks to express the facial shapes of different age groups. However, manual annotation in facial landmark detection limits the usability of such approaches in automatic demographic estimation systems. Texture-based approaches, such as [16,17,20], utilize facial texture features, e.g., local binary patterns (LBP), Gabor, and biologically inspired features (BIF). Although used in many demographic estimation approaches, high feature dimensionality makes such approaches less efficient. The appearance-based approaches, such as [11,18,19], adopt facial appearance to discriminate among face images belonging to different demographic groups.

Recently, some deep-learning-based approaches to demographic estimation have been proposed [21,22]. In [21], deeply learned features have been used for accurate age estimation using age difference. In [22], weighted-LBP feature elements are used to estimate the age of a given face image in the presence of local latency. These methods suggest that deeply learned features have become the norm for demographic estimation and recognition tasks due to their ability to learn the best features for these tasks.

## 2.2. Recognition and Retrieval of Face Images across Aging Variations

There exists a sizable amount of literature on recognition of age-separated face images, such as [24–36]. The existing methods can be categorized as generative or discriminative. The generative methods, such as [24,26], rely on aging synthesis. Despite their success in discriminating face images across aging variations, these methods are parametric dependent, which limits their accuracy. In contrast, discriminative methods, such as [27–29], use multiple facial features for robust recognition across aging variations. Facial asymmetry, which corresponds to the shape, size, and location of the bilateral landmarks, is another important factor which increases with age [9,37]. During formative years, the human face is more symmetric, while in later years, facial asymmetry is more prominent. Facial-asymmetry-based features have been used in age-invariant face recognition in [10,38]. In [10], asymmetric facial features have been used along with local and global features for face recognition. In [38], age-group estimates have been used to enhance face recognition accuracy. Symmetrical characteristic of the face has been used in fear estimation in [39].

Recently, some deep learning methods, including [33,40–44], and data driven approaches, including [45,46], have been proposed for face recognition. In [33], coupled autoencoder networks have been used to recognize and retrieve face images with temporal variations. Similarly, [40–44] propose different neural networks to achieve age-invariant face recognition. Data-driven methods presented in [45,46] are based on cross-aging reference sets to perform face recognition across aging.

## 2.3. Recognition of Face Images Based on Demographic Estimates

To our knowledge, research on recognition of face images using demographic estimation is quite limited. Previously, age estimation and face recognition have been combined in [14,38]. However, these methods use only a single demographic attribute, i.e., age group to recognize face images.

The above presented methods suggest that there is vast scope to develop demographic-assisted methods to age-invariant face recognition and retrieval. To this end, we propose to use facial-asymmetry-based deeply learned features to estimate the gender, race, and age group of a given face image. The same features are then used to re-rank face images to achieve superior recognition and retrieval results. Such an approach is called demographic-assisted face recognition and retrieval. Experiments were conducted on two standard datasets, the MORPH II [47] and FERET [48]. MORPH II is a longitudinal face dataset, with 55,134 face images of more than 13,000 individual subjects. FERET is another aging dataset, which contains 3540 frontal face images of 1196 individual subjects. The dataset contains a gallery called *fa* set and two distinct aging subsets called *dup-I* and *dup-II*, containing face images with small and large aging variations, respectively.

The key statistics of MORPH II and FERET datasets are given in Table 1, while some sample face images from these datasets are shown in Figure 1.

**Table 1.** The key statistics of MORPH II and FERET datasets.

Attributes	MORPH II	FERET		
		Fa Set	Dup I Set	Dup II Set
# Subjects	13,000+	1196	243	75
# Images	55,134	1196	722	234
# Average images/subject	4	1	2.97	3.12
Temporal variations	Minimum: 1 day Maximum: 1681 days	N/A	Images were acquired later in a time with maximum temporal variation of 1031 days as compared to corresponding gallery images	Images were acquired at least 1.5 years later than corresponding gallery images
Age range	16–77		10–60+	
Gender	Male, Female		Male, Female	
Race	White, others		White, others	

			
Age	40      42	50      52	30      32
Gender	Female	Male	Male
Race	White	White	White

			
Age	54      56	47      48	50      54
Gender	Male	Female	Male
Race	Black	White	Black

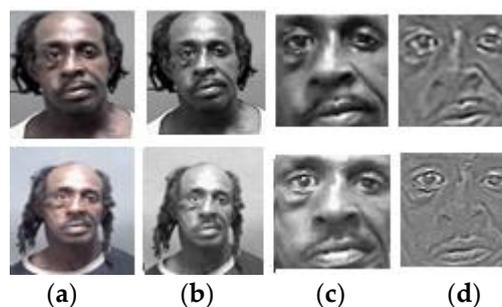
**Figure 1.** Example face image pairs with age, gender, and race information from FERET (top row), and MORPH II (bottom row).

### 3. Demographic Estimation

The proposed approach for demographic estimation consists of two main components: (i) preprocessing and (ii) learning demographic informative features using CNNs as described in the following subsections.

#### 3.1. Preprocessing

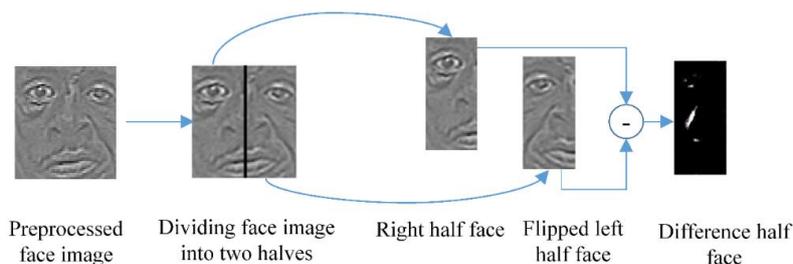
Face images normally contain various appearance variations including scale, translation, rotation, illumination, and color cast. To compensate for such variations, we adopted following preprocessing steps: (i) to correct unwanted color cast, we used a luminance model adopted by NTSC and JPEG [20]; (ii) to mitigate the effects of in-plane rotation and translation, face images were aligned based on fixed eye locations detected by a publicly available tool, the Face++ [49]; (iii) the aligned face images were cropped to a common size of  $128 \times 128$  pixels; and (iv) to correct illumination variations due to shadows and underexposure, we used difference of Gaussian filtering [50]. Figure 2 shows face preprocessing results for two sample face images of a subject from MORPH II dataset.



**Figure 2.** Face preprocessing (a) input images (b) gray-scale images (c) images aligned based on fixed eye locations (d) illumination correction.

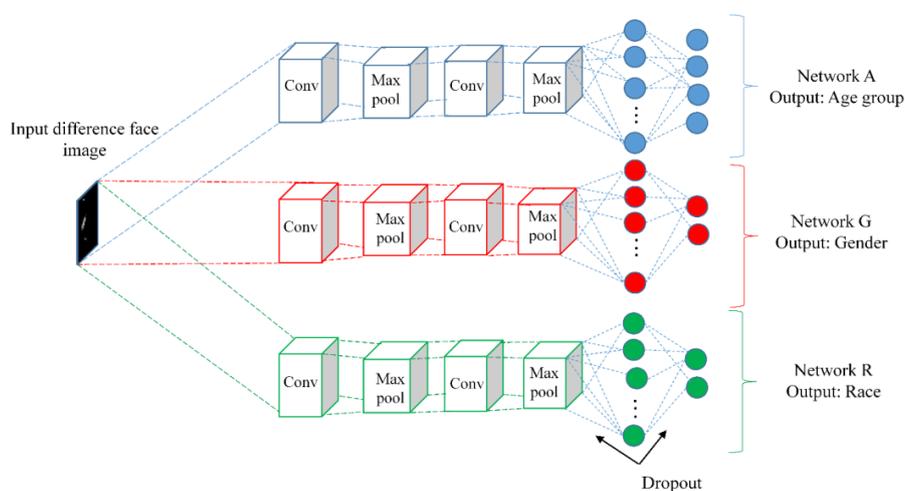
### 3.2. Demographic Features Extraction

Since we aimed to extract asymmetric facial features for demographic estimation (age group, gender, and race), it was necessary to have a face representation to extract such features. For this purpose, each preprocessed face image of size  $128 \times 128$  pixels was divided into  $128 \times 64$  pixels, left and right half face, such that difference between two halves was minimal. The flipped image of the left half face was subtracted from the right half face, resulting in  $128 \times 64$  pixels left–right difference half face. The difference half-face image contained the asymmetric facial variations and was used to extract the demographic informative features. Figure 3 illustrates the extraction of a difference half face from a given preprocessed face image.



**Figure 3.** Extraction of difference half-face image.

To compute demographic informative features from the difference half-face image, we trained three convolutional neural networks, *A*, *G*, and *R*, each for age group, gender, and race classification tasks, respectively. More precisely, networks *A*, *G*, and *R* represented CNNs that took difference half-face images as input and output class labels with corresponding softmax layers. Network *A* was trained for the age classification task such that input face images can be classified into one of the four age groups. Similarly, networks *G* and *R* were trained for binary classification tasks of gender and race, respectively. Each network consisted of two convolutional layers (conv) followed by batch normalization and max pooling steps. Two fully connected layers were placed at the end of each network. The first fully connected layer consisted of 1024 units and the second layer acted as the output layer with the softmax function. The last layer of network *A* consisted of four units, each for one age group. Similarly, the networks *G* and *R* contained the last fully connected layer with two units each for the binary classification tasks of gender and race of input image. Figure 4 illustrates the CNNs used for demographic estimation, including age group, gender, and race for an arbitrary input difference face image.



**Figure 4.** Illustration of networks *A*, *G*, and *R* for demographic classification tasks.

### 3.3. Experimental Results of Demographic Estimation

We performed age group, gender, and race classification on difference half-face images using simple CNN models. To this end, each difference face image of size  $h \times w$  was passed through pretrained networks  $A$ ,  $G$ , and  $R$ , each containing two convolutional layers. The choice of the simple CNN models was motivated by the abstract nature of the asymmetric facial information present in the difference half-face images. The output  $L^p$  of a layer  $p$  was a  $h^p \times w^p \times f^p$  feature map, where  $f$  is number of filters in a convolutional layer. Finally, a set  $S = \{s_1^p, s_2^p, \dots, s_{h^p, w^p}^p\}$  was obtained at each location of the feature map for the networks  $A$ ,  $G$ , and  $R$ . More precisely, we obtained three sets of feature vectors,  $S_A = \{s_{1A}^p, s_{2A}^p, \dots, s_{h^p, w^p}^p\}$ ,  $S_G = \{s_{1G}^p, s_{2G}^p, \dots, s_{h^p, w^p}^p\}$ , and  $S_R = \{s_{1R}^p, s_{2R}^p, \dots, s_{h^p, w^p}^p\}$  for age, gender, and race features, respectively. To avoid overfitting, we used a cross-validation strategy to obtain error-aware outputs from each network. Due to significantly imbalanced race distribution in MORPH II and FERET datasets, we performed binary race estimation between White and other races (Black, Hispanic, Asian, and African-American). For each classification task, the results are reported in the form of a confusion matrix, as illustrated in following subsections.

#### 3.3.1. Results on MORPH II Dataset

We used a subset consisting of 20,000 face images from MORPH II dataset. We used 10,000 face images in training, while 10,000 face images were used in test sets with age ranges of 16–20, 21–30, 31–45, and 46–60+. The performance of the proposed and the Face++ methods [27] for age group, gender and race estimation is reported in Table 2a–c in the form of confusion matrices. It is worthwhile to mention that the results were calculated using a 5-fold cross-validation methodology. An overall age group estimation accuracy of 94.50% with a standard deviation of 1.1 was achieved for our method compared to the state-of-the-art Face++, which achieves an overall accuracy of 89.15% with a standard deviation of 1.5. In case of Face++, the subjects (particularly in the age range of 46–60+) are found to be more baffled with subjects in younger age groups, mainly because of use of facial make-up, resulting in a younger appearance of the subjects. In contrast, asymmetric features used in the proposed method are difficult to manipulate by using facial make-up and hence the subjects are less confused with the corresponding younger age groups. The comparative results of the gender estimation task of the proposed approach and the Face++ are shown in Table 2b. It can be seen that our method achieved an overall accuracy of 82.35% with a standard deviation of 0.6 compared to the Face++, which achieved an overall accuracy of 77.80% with a standard deviation of 0.4. An interesting observation is higher misclassification rates for females than males achieved by the proposed approach. One possible explanation for this higher misclassification is frequent extrinsic variations of female faces, such as facial make-up and varying eyebrow styles.

**Table 2.** Confusion matrices showing classification accuracies of the proposed and Face++ methods for (a) age group, (b) gender, and (c) race estimation tasks on MORPH II dataset.

(a)								
Actual Age Groups	Predicted Age Groups							
	Proposed				Face++			
	16–20	21–30	31–45	46–60+	16–20	21–30	31–45	46–60+
16–20	93.45	5.58	0.49	0.48	88.80	7.54	2.43	1.21
21–30	1.68	95.02	2.50	0.80	3.59	92.04	2.32	2.04
31–45	0.20	1.42	96.00	2.38	0.75	3.93	93.53	1.78
46–60+	2.53	1.00	2.91	93.56	3.20	5.28	9.28	82.24
Overall accuracy	94.50 ± 1.1				89.15 ± 1.5			

Table 2. Cont.

(b)				
Actual Gender	Predicted Gender			
	Proposed		Face++	
	Male	Female	Male	Female
Male	86.00	14.00	80.70	19.30
Female	21.23	78.77	25.09	74.91
Overall accuracy	$82.35 \pm 0.6$		$77.80 \pm 0.4$	

(c)				
Actual Race	Predicted Race			
	Proposed		Face++	
	White	Others	White	Others
White	89.61	10.39	84.00	16.00
Others	29.19	70.81	31.00	69.00
Overall accuracy	$80.21 \pm 1.18$		$76.50 \pm 1.2$	

The race estimation experiments conducted on MORPH II dataset classify subjects between White and other races. The race estimation results are shown in Table 2c, depicting the superior performance of our method compared to the Face++. The proposed method achieved an overall accuracy of 80.21% with a standard deviation of 1.18 compared to the Face++, which achieved an overall accuracy of 76.50% with a standard deviation of 1.2. It can be seen that proposed approach is better at estimating the White race compared to the other races.

### 3.3.2. Results on FERET Dataset

We use 1196 frontal face images from *fa* set of FERET dataset. Five hundred and ninety-eight images were used in training, while 598 images were used in testing for demographic estimation tasks. The demographic estimation performance of the proposed approach and the Face++ is reported in Table 3a–c in the form of confusion matrices. The proposed method gave an overall age group estimation accuracy of 83.88% with a standard deviation of 1.00 across a 5-fold cross validation. In contrast, the Face++ achieved an overall age group accuracy of 81.52% for the same experimental protocol.

For gender estimation task, the proposed method achieved an overall estimation accuracy of 82.72% with a standard deviation of 0.9 compared to 76.28% achieved by the Face++. It can be observed that the misclassification rate of female subjects is greater than male subjects in the case of the FERET dataset, similar to the trend observed for subjects in MORPH II dataset.

Our method achieved an overall race estimation accuracy of 79.80% with a standard deviation of 0.9 across 5-folds, compared to the race estimation accuracy of 74.56% achieved by the Face++ for the same experimental protocol. The proposed approach gave a higher misclassification rate for the White race compared to other races (26.00% vs. 14.39%), which can be attributed to large within group variety of other races.

**Table 3.** Confusion matrices showing the classification accuracies of the proposed and the Face++ methods for (a) age group, (b) gender, and (c) race estimation accuracies on FERET dataset.

(a)								
Actual Age Groups	Predicted Age Groups							
	Proposed				Face++			
	10–20	21–30	31–45	46–60+	10–20	21–30	31–45	46–60+
10–20	86.00	6.01	4.93	3.06	83.33	8.54	6.40	1.73
21–30	3.50	85.23	6.00	5.27	3.84	83.84	12.32	0
31–45	0.29	6.00	81.70	12.01	3.57	3.57	80.94	11.90
46–60+	0.20	5.20	12.00	82.60	0	8.67	13.33	78.00
Overall accuracy	83.88 ± 1.00				81.52 ± 1.1			

(b)				
Actual Gender	Predicted Gender			
	Proposed		Face++	
	Male	Female	Male	Female
Male	83.00	17.00	78.94	21.06
Female	17.56	82.44	26.39	73.61
Overall accuracy	82.72 ± 0.9		76.28 ± 0.5	

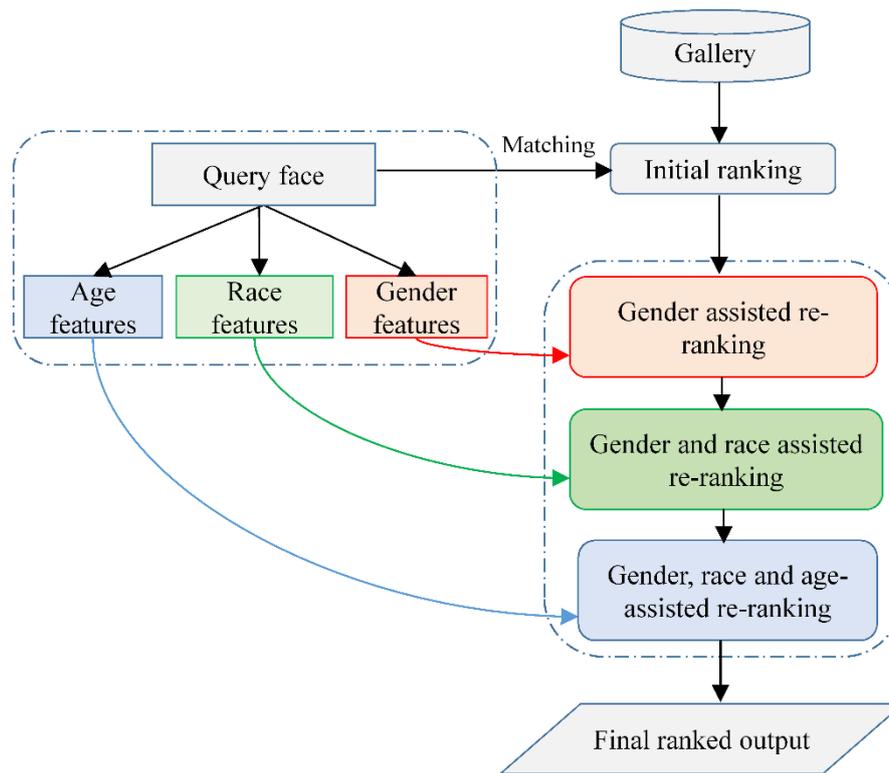
  

(c)				
Actual Race	Predicted Race			
	Proposed		Face++	
	White	Others	White	Others
White	85.61	14.39	82.00	18.00
Others	26.00	74.00	32.88	67.12
Overall accuracy	79.80 ± 0.9		74.56 ± 0.9	

#### 4. Recognition and Retrieval of Face Images across Aging Variations

The second part of this study deals with the demographic-assisted face recognition and retrieval approach. Our proposed method for face recognition and retrieval includes the following steps.

- (i) A query face image is first matched against a gallery using deep CNN (dCNN) features. To extract these features, we used VGGNet [51]. Particularly, we used a variant of VGGNet called VGG-16, which contains 16 layers with learnable weights, including 13 convolution and 3 fully connected layers. The VGGNet used a filter of size  $3 \times 3$ . The combination of two  $3 \times 3$  filters resulted in a receptive field of  $5 \times 5$ , simulating a larger filter but retaining the benefits of smaller filters. We selected VGGNet due to its better performance for the desired task with relatively simpler architecture. The matching stage returned top  $k$  matches from gallery against the query face image.
- (ii) Extract demographic features (age, gender, and race) using three CNNs  $A$ ,  $G$  and  $R$ .
- (iii) The top  $k$  matched face images are then re-ranked by using gender, race and age features as shown in Figure 5.

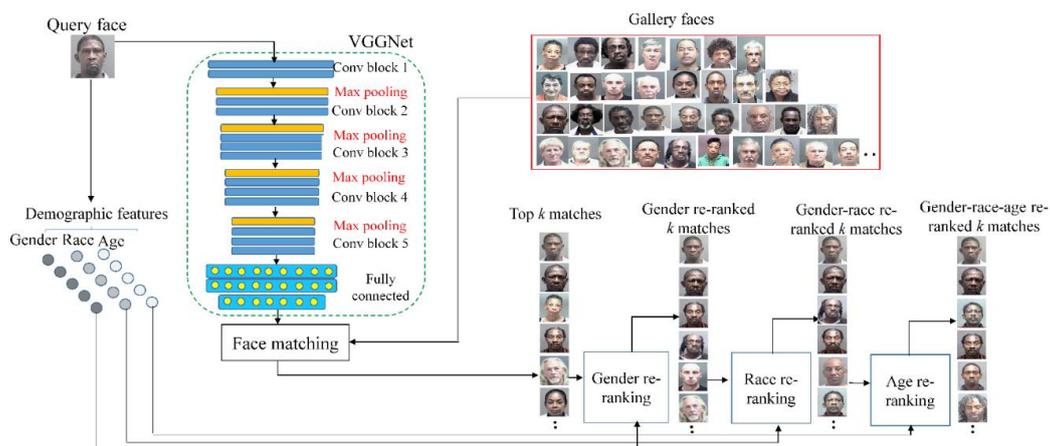


**Figure 5.** The proposed demographic-assisted re-ranking pipeline.

We explain demographic-assisted re-ranking in the following steps:

- (i) Re-ranking by gender features: In this step, the top  $k$  matched face images are re-ranked based on the gender-specific features,  $S_G = \{s_{1G}^p, s_{2G}^p, \dots, s_{h^p, w^p}^p\}$ , of the query face image, returning gender re-ranked top  $k$  matches. The re-ranking applied in this step helps to refine the initial top  $k$  matches based on the gender features of the query face image. The resulting re-ranked images are called gender re-ranked  $k$  matches.
- (ii) Re-ranking by race features: The gender re-ranked face images are again re-ranked using race-specific features,  $S_R = \{s_{1R}^p, s_{2R}^p, \dots, s_{h^p, w^p}^p\}$ , resulting in the gender–race re-ranked top  $k$  matches. The re-ranking applied in this step helps to refine the gender re-ranked  $k$  matches obtained in the first step based on the race of the query face image. The resulting re-ranked images are called gender–race re-ranked  $k$  matches.
- (iii) Re-ranking by age features: Finally, the gender–race re-ranked face images are re-ranked using age-specific features,  $S_A = \{s_{1A}^p, s_{2A}^p, \dots, s_{h^p, w^p}^p\}$ , returning gender–race–age re-ranked top  $k$  matches. The re-ranking applied in final step helps to refine the gender–race re-ranked face images based on the aging features of the query face image. The resulting re-ranked images are called gender–race–age re-ranked top  $k$  matches. This step produces the final ranked output as shown in Figure 5.

The complete block diagram of the proposed demographic-assisted face recognition and retrieval approach is shown in Figure 6.



**Figure 6.** The block diagram of the proposed face recognition and retrieval approach.

#### 4.1. Evaluation

The performance of our method is evaluated in terms of identification accuracy and mean average precision (mAP), as illustrated in the following subsections.

##### 4.1.1. Face Recognition Experiments

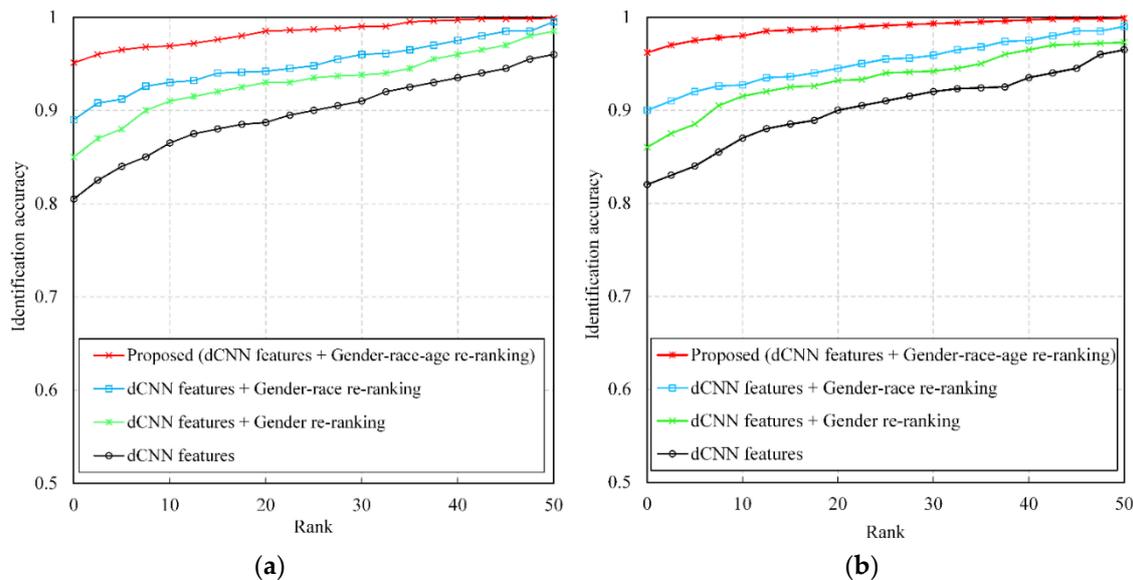
In the first evaluation, we used rank-1 identification accuracy as an evaluation metric for face recognition performance on MORPH II and FERET datasets, as described in the following experiments.

**Face Recognition Experiments on MORPH II Dataset:** In case of MORPH II dataset, we used 20,000 face images of 20,000 subjects, as a gallery, while 20,000 older face images are used as the probe set. To extract gender, race, and aging features, the CNN models were trained on gallery face images. The rank-1 recognition accuracies for face identification experiments are reported in Table 4. We achieved a rank-1 recognition accuracy of 80.50% for the probe set when dCNN features were used to match the face images. Re-ranking face images with gender features resulted in a rank-1 identification accuracy of 84.81%. Similarly, gender–race re-ranking yielded the rank-1 identification accuracy of 89.00%. Finally, the proposed approach with gender–race–age re-ranking gave a rank-1 identification accuracy of 95.10%. We also show the cumulative match characteristic (CMC) curves for this series of experiments in Figure 7a.

**Table 4.** Rank-1 recognition accuracies of the proposed and existing methods.

Approach	Rank-1 Recognition Accuracy (%)	
	MORPH II	FERET
(i) CARC [45]	84.11	85.98
(ii) GSM1 [52]	83.33	85.00
(iii) GSM2 [52]	93.73	94.23
(iv) Score-space-based fusion [10]	72.40	66.66
(v) Age-assisted face recognition [38]	85.00	78.60
(vi) dCNN features	80.50	82.00
(vii) dCNN features + gender re-ranking	84.81	85.91
(viii) dCNN features + gender–race re-ranking	89.00	90.00
(ix) Proposed approach with demographic re-ranking (actual age groups, gender, and race)	96.53	97.01
(x) Proposed approach with demographic re-ranking (estimated age groups, gender, and race)	95.10	96.21

**Face Recognition Experiments on FERET Dataset:** In case of FERET dataset, we used standard gallery and probe sets to evaluate the performance of our method. We used 1196 face images from *fa* set as the gallery, while 956 face images from dup I and dup II sets were used as probe set. Gender, race, and age features were extracted by training the CNN models on the gallery set. The rank-1 recognition accuracies are reported in Table 4. The dCNN features to match the probe and gallery face images yielded a rank-1 recognition accuracy of 82.00%. Gender re-ranking, and gender–race re-ranking, yielded rank-1 identification accuracies of 85.91% and 90.00%, respectively. The proposed approach with gender–race–age re-ranking yielded the highest rank-1 identification accuracy of 96.21%. The CMC curves for this series of experiments are shown in Figure 7b.



**Figure 7.** Cumulative match characteristic (CMC) curves showing the identification performance of the proposed methods on (a) MORPH II (b) FERET dataset.

#### 4.1.2. Comparison of Face Recognition Results with State-Of-The-Art

We compared the results of the proposed face-recognition approach with existing methods, including the score-space-based fusion approach presented in [10], age-assisted face recognition [38], CARC [45], GSM1 [52], and GSM2 [52]. The score-space-based approach presented in [10] uses facial-asymmetry-based features to recognize face images across aging variations without demographic estimation. Age-assisted face recognition [38] leverages the age group estimates to enhance the recognition accuracy across aging variations. CARC [45] is a data-driven coding framework, called cross-age reference coding, that is suggested to retrieve and recognize face images across aging variations. The generalized similarity models (GSM1 and GSM2) aim to retrieve face images by implementing a cross-domain visual matching strategy followed by incorporation of a similarity measure matrix into a deep architecture.

For the same experimental setup, the score-space-based fusion [10] achieved recognition accuracies of 72.40% on MORPH II and 66.66% on FERET dataset. The age-assisted face-recognition-based approach [38] achieved 85.00% on MORPH II and 78.60% on FERET dataset. CARC achieved rank-1 identification accuracies of 84.11% and 85.98% on MORPH II and FERET datasets, respectively. GSM1 yielded the rank-1 identification accuracies of 83.33% and 85.00% for MORPH II and FERET datasets, respectively. Similarly, GSM2 achieved rank-1 identification accuracies of 93.73% and 94.23% on MORPH II and FERET datasets, respectively. We also present an analysis for the error introduced by the age group, gender, and race estimation of probe images compared to

the actual age-groups, gender, and race in recognizing face images both for MORPH II and FERET datasets (see row viii, Table 4).

#### 4.1.3. Face Retrieval Experiments

In the second evaluation, mAP was used as an evaluation metric for the face retrieval performance of the proposed method. Following existing face-retrieval methods [33,53,54] for a given a set of  $p$  query face images,  $Q = \{y_q^1, y_q^2, \dots, y_q^p\}$ , and a gallery set with  $N$  face images, the average precision for  $y_q^i$  is defined as:

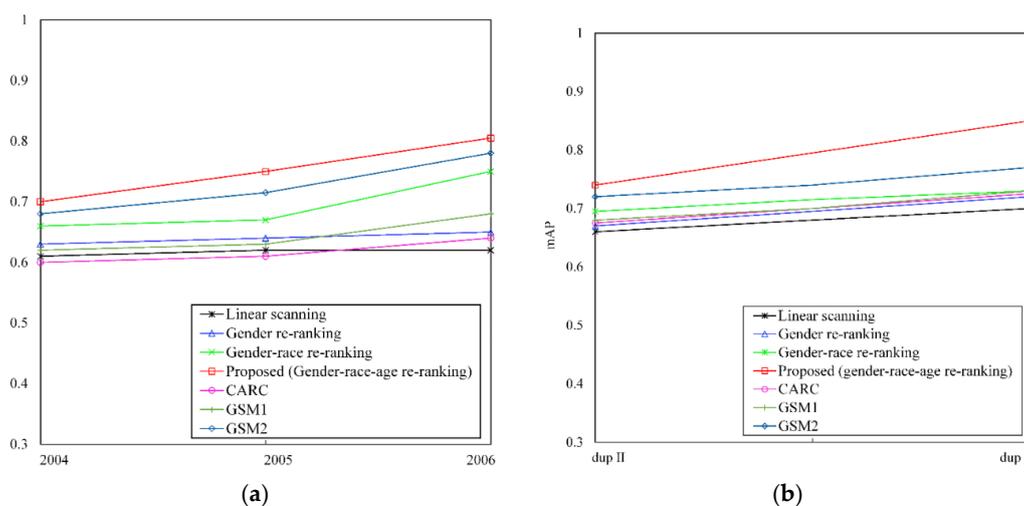
$$avgP(y_q^i) = \sum_{j=1}^N P(y_q^i, j) [C(y_q^i, j) - C(y_q^i, j-1)] \quad (1)$$

where  $(y_q^i, j)$  is precision at the  $j$ -th position for  $y_q^i$  and  $C(y_q^i, j)$  is the recall for the same position, such that  $y_q^i(C(0) = 0$ . Finally, the mAP for entire query dataset  $Q$  is defined as:

$$mAP(Q) = \frac{1}{|Q|} (avgP(y_q^i)), \quad i = 1, 2, \dots, p \quad (2)$$

**Retrieval Experiments on MORPH II Dataset:** We selected face images of 780 distinct subjects from the MORPH II dataset to constitute a test set with images acquired in the years 2007 as the query set, while images acquired in the years 2004, 2005, and 2006 formed three distinct training subsets. The cosine similarity metric was used to calculate the matching scores between two given face images. The mAPs for face matching using linear scan, gender re-ranking, gender-race re-ranking, and the proposed approach are shown in Figure 8a for three test sets containing face images acquired in the years 2004, 2005, and 2006. The results suggest that the proposed approach achieved the highest mAP on all three test sets compared to gender re-ranking and gender-race re-ranking-based retrieval methods.

**Retrieval Experiments on FERET Dataset:** In case of FERET dataset, we used  $fa$  as training, while dup I and dup II were used as test sets. The mAPs for face matching using the linear scan, gender re-ranking, gender-race re-ranking, and the proposed approach are shown in Figure 8b for the dup I and dup II test sets. The results show that the proposed approach gave the highest mAP on all three test sets compared to gender re-ranking and gender-race re-ranking-based retrieval methods.



**Figure 8.** Face retrieval performance of the proposed and existing methods on (a) MORPH II (b) FERET dataset.

#### 4.1.4. Comparison of Face Retrieval Results with State-Of-The-Art

The performance of our approach is compared with close competitors, including CARC [45], and generalized similarity models [52] (GSM1 and GSM2) in Figure 8a,b in terms of mAP for MORPH II and FERET test sets, respectively. It is evident that the proposed approach outperformed the existing methods. The superior performance can be attributed to the demographic-estimation-based re-ranking compared to the linear scan employed in CARC [45], GSM1 [52], and GSM2 [52].

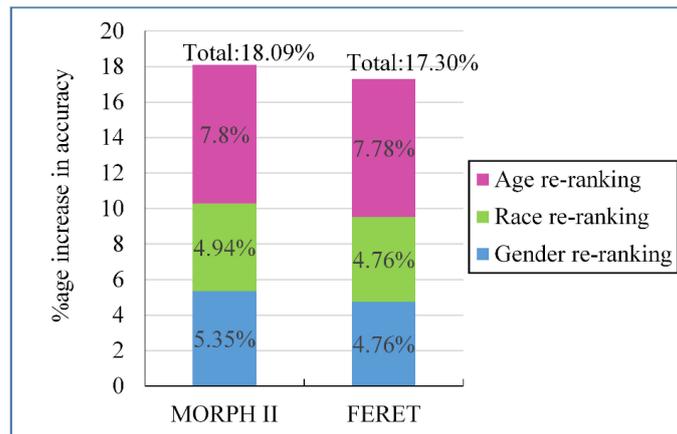
### 5. Results Related Discussion

Based on the above presented comparison, we make the following key observations.

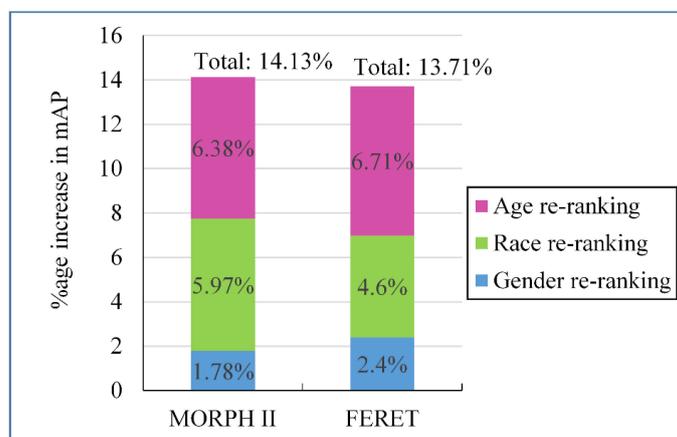
- (i) We have encountered more complicated tasks of demographic-assisted face recognition and retrieval compared to existing methods which consider either demographic estimation or face recognition.
- (ii) The existing methods to age-invariant face recognition and retrieval lack the analysis of impact of demographic features on face recognition and retrieval performance. In contrast, the current study is the first of its kind analyzing the impact of demographic information on face recognition and retrieval performance. In Figures 9 and 10, we illustrate the impact of gender-, race-, and aging-features-based re-ranking on recognition and retrieval accuracies, respectively. One can observe that aging features have the most significant impact on both the identification accuracy and mAP compared to the race and gender features for MORPH II and FERET datasets. This is because aging features are heterogeneous in nature, i.e., more person-specific [36] thus helpful in recognizing face images across aging. In contrast, race and gender features are more population-specific and less person specific, as suggested in [3,4].
- (iii) The comparative results for face recognition and retrievals tasks suggest that demographic re-ranking can improve the recognition accuracies effectively. In all presented experiments on MORPH II and FERET datasets, gender–race–age re-ranking yields superior accuracies compared to gender and gender–race re-ranking. This is because face images are first re-ranked-based on gender features. In this re-ranking stage, query face image is matched with face images of a specific gender only. The gender-re-ranked face images are then re-ranked-based on race features resulting in gender–race re-ranked results. In gender–race re-ranking, the query face image is matched with face images of a specific gender and race. Finally, the gender–race re-ranked face images are re-ranked-based on gender, race and age features. In this final re-ranking stage, query face image is matched with face images of a specific gender, race, and age group, yielding the highest recognition accuracies.
- (iv) The retrieval results suggest that demographic-assisted re-ranking yields superior mAPs compared to the linear scan approach, where a query face is matched against the entire gallery, resulting in lower mAP accuracies. In contrast, the proposed gender–race–age re-ranking makes it possible to match a query against face images of specific gender, race, and age group, resulting in higher mAPs. For example, the linear scan approach gives mAP of 61.00% for the MORPH II query set acquired in the year 2004 (refer to Figure 8a), compared to mAP of 70.00% obtained using the proposed approach for the same query set.
- (v) Motivated by the fact that facial asymmetry is a strong indicator of age group, gender, and race [7–9], we used demographic estimates to enhance recognition and retrieval performance. The demographic assisted face recognition and retrieval accuracies are benchmarked with accuracies obtained from a dCNN model. It is observed that demographic assisted accuracies are better than the benchmark results, suggesting the effectiveness of the proposed approach. To the best of our knowledge, the current work is the first of its kind that uses demographic features with face recognition and a retrieval algorithm to achieve better accuracies.
- (vi) Through extensive experiments, we have demonstrated the effectiveness of our approach against the existing methods including CARC [45], GSM1 [52], and GSM2 [52] on MORPH II and FERET

datasets. The superior performance of our approach can be attributed to the deeply learned asymmetric facial features and demographic re-ranking strategy to recognize and retrieve face images across aging variations. In contrast, the most related competitors, including CARC, GSM1, and GSM2, consider only aging information to recognize and retrieve face images across aging variations.

- (vii) The superior performance of our method compared to the discriminative approach suggested in [10] can be attributed to two key factors. First, the proposed approach employs deeply learned face features in contrast to handcrafted features used in [10]. Second, the proposed approach uses the gender and race features in addition to the aging features. The demographic-assisted face recognition also surpasses the face identification results achieved by the age-assisted face recognition presented in [38], suggesting the discriminative power of the gender and race features in addition to aging features in recognizing face images with temporal variations.
- (viii) Face identification and retrieval results suggest that the proposed approach is well suited for age-invariant face recognition, owing to the discriminative power of population-specific gender and race features.
- (ix) To analyze the errors introduced by demographic estimation in the proposed face recognition and retrieval approach, we reported the rank-1 identification accuracies and mAPs for the scenario if actual age group, gender, and race information (ground truth) is used (refer to row viii of Table 4). The meager difference between accuracies obtained from actual and estimated age group, gender, and race features suggests the efficacy of the proposed demographic-estimation-based approach.



**Figure 9.** Impact of the demographic re-ranking on rank-1 identification accuracies for MORPH II and FERET datasets.



**Figure 10.** Impact of the demographic re-ranking on mAP for MORPH II and FERET datasets.

## 6. Conclusions

The human face contains a number of age-specific features. Facial asymmetry is one of such intrinsic facial feature, which is a strong indicator of age group, gender, and race. In this work, we have proposed a demographic-assisted face recognition and retrieval approach. First, we estimated facial-asymmetry-based age group, gender, and race of a query face image using CNNs. The demographic features are then used to re-rank face images. The experimental results suggest that, firstly, facial asymmetry is a strong indicator of age group, gender, and race. Secondly, the demographic features can be used to re-rank face images to achieve superior recognition accuracies. The proposed approach yields superior mAP accuracies by matching a query face image against gallery face images of a specific gender, race, and age group. Thirdly, the deeply learned face features can be used to achieve superior face recognition and retrieval performance compared to the handcrafted features. Finally, the study suggests that among aging, gender, and race, the aging features play a significant role in recognizing and retrieving age-separated face images, owing to their person-specific nature. The experimental results on two longitudinal datasets suggest that the proposed approach can compete with the existing methods to recognize and retrieve face images across aging variations. Future work may include expanding the role of other facial attributes in demographic estimation and face recognition.

**Author Contributions:** M.S. and T.S. conceived the idea and performed the analysis. M.S., and F.I. contributed in the write up of the manuscript, performed experiments and prepared the relevant Figures and Tables. S.M. gave useful insights during experimental analysis. H.T., U.S.Q., and I.R. provided their expertise in revising the manuscript. All authors were involved in preparing the final manuscript.

**Acknowledgments:** The authors are thankful to the sponsors of publicly available FERET and MORPH II datasets used in this study.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Mayes, E.; Murray, P.G.; Gunn, D.A.; Tomlin, C.C.; Catt, S.D.; Wen, Y.B.; Zhou, L.P.; Wang, H.Q.; Catt, M.; Granger, S.P. Environmental and lifestyle factors associated with perceived age in Chinese women. *PLoS ONE* **2010**, *5*, e15273. [CrossRef] [PubMed]
2. Balle, D.S. Anatomy of Facial Aging. Available online: <https://drballe.com/conditions-treatment/anatomy-of-facial-aging-2> (accessed on 28 January 2017).
3. Zhuang, Z.; Landsittel, D.; Benson, S.; Roberge, R. Facial anthropometric differences among gender, ethnicity, and age groups. *Ann. Occup. Hyg.* **2010**, *54*, 391–402. [PubMed]
4. Farkas, L.G.; Katic, M.J.; Forrest, C.R. International anthropometric study of facial morphology in various ethnic groups/races. *J. Craniofac. Surg.* **2005**, *16*, 615–646. [CrossRef] [PubMed]
5. Ramanathan, N.; Chellappa, R.; Biswas, S. Computational methods for modeling facial aging: A survey. *J. Vis. Lang. Comput.* **2009**, *20*, 131–144. [CrossRef]
6. Bruce, V.; Burton, A.M.; Hanna, E.; Healey, P.; Mason, O.; Coombs, A. Sex discrimination: How do we tell the difference between male and female faces? *Perception* **1993**, *22*, 131–152. [CrossRef] [PubMed]
7. Little, A.C.; Jones, B.C.; Waitt, C.; Tiddem, B.P.; Feinberg, D.R.; Perrett, D.I.; Apicella, C.L.; Marlowe, F.W. Symmetry is related to sexual dimorphism in faces: Data across culture and species. *PLoS ONE* **2008**, *3*. [CrossRef] [PubMed]
8. Steven, W.; Randy, T. Facial masculinity and fluctuating symmetry. *Evol. Hum. Behav.* **2003**, *24*, 231–241.
9. Morrison, C.S.; Phillips, B.Z.; Chang, J.T.; Sullivan, S.R. The Relationship between Age and Facial Asymmetry. 2011. Available online: <http://meeting.nesps.org/2011/80.cgi> (accessed on 28 April 2017).
10. Sajid, M.; Taj, I.A.; Bajwa, U.I.; Ratyal, N.I. The role of facial asymmetry in recognizing age-separated face images. *Comput. Electr. Eng.* **2016**, *54*, 255–270. [CrossRef]
11. Fu, Y.; Huang, T.S. Human age estimation with regression on discriminative aging manifold. *IEEE Trans. Multimed.* **2008**, *10*, 578–584. [CrossRef]

12. Lu, K.; Seshadri, K.; Savvides, M.; Bu, T.; Suen, C. Contourlet Appearance Model for Facial Age Estimation. 2011. Available online: <https://pdfs.semanticscholar.org/bc82/a5bfc6e5e8fd77e77e0ffaadedb1c48d6ae4.pdf> (accessed on 28 April 2017).
13. Bekios-Calfa, J.; Buenaposada, J.M.; Baumela, L. Revisiting linear discriminant techniques in gender recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 858–864. [[CrossRef](#)] [[PubMed](#)]
14. Wu, T.; Turaga, P.; Chellappa, R. Age estimation and face verification across aging using landmarks. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 1780–1788. [[CrossRef](#)]
15. Hadid, A.; Pietikainen, M. Demographic classification from face videos using manifold learning. *Neurocomputing* **2013**, *100*, 197–205. [[CrossRef](#)]
16. Guo, G.; Mu, G. Joint estimation of age, gender and ethnicity: CCA vs. PLS. In Proceedings of the 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, 22–26 April 2013.
17. Tapia, J.E.; Perez, C.A. Gender classification based on fusion of different spatial scale features selected by mutual information from histogram of LBP, intensity, and shape. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 488–499. [[CrossRef](#)]
18. Choi, S.E.; Lee, Y.J.; Lee, S.J.; Park, K.R.; Kim, J.K. Age estimation using a hierarchical classifier based on global and local facial features. *Pattern Recognit.* **2011**, *44*, 1262–1281. [[CrossRef](#)]
19. Geng, X.; Yin, C.; Zhou, Z.-H. Facial age estimation by learning from label distributions. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *35*, 2401–2412. [[CrossRef](#)] [[PubMed](#)]
20. Han, H.; Otto, C.; Liu, X.; Jain, A.K. Demographic estimation from face images: Human vs. machine performance. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 1148–1161. [[CrossRef](#)] [[PubMed](#)]
21. Hu, Z.; Wen, Y.; Wang, J.; Wang, M.; Hong, R.; Yan, S. Facial age estimation with age difference. *IEEE Trans. Image Process.* **2017**, *26*, 3087–3097. [[CrossRef](#)] [[PubMed](#)]
22. Jadid, M.A.; Sheij, O.S. Facial age estimation under the terms of local latency using weighted local binary pattern and multi-layer perceptron. In Proceedings of the 4th International Conference on Control, Instrumentation, and Automation (ICCIA), Qazvin, Iran, 27–28 January 2016.
23. Liu, K.-H.; Yan, S.; Kuo, C.-C.J. Age estimation via grouping and decision fusion. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 2408–2423. [[CrossRef](#)]
24. Geng, X.; Zhou, Z.; Smith-Miles, K. Automatic age estimation based on facial aging patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 2234–2240. [[CrossRef](#)] [[PubMed](#)]
25. Ling, H.; Soatto, S.; Ramanathan, N.; Jacobs, D. Face verification across age progression using discriminative methods. *IEEE Trans. Inf. Forensics Secur.* **2010**, *5*, 82–91. [[CrossRef](#)]
26. Park, U.; Tong, Y.; Jain, A.K. Age-invariant face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 947–954. [[CrossRef](#)] [[PubMed](#)]
27. Li, Z.; Park, U.; Jain, A.K. A discriminative model for age-invariant face recognition. *IEEE Trans. Inf. Forensics Secur.* **2011**, *6*, 1028–1037. [[CrossRef](#)]
28. Yadav, D.; Vatsa, M.; Singh, R.; Tistarelli, M. Bacteria foraging fusion for face recognition across age progression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 23–28 June 2013.
29. Sungatullina, D.; Lu, J.; Wang, G.; Moulin, P. Discriminative Learning for Age-Invariant Face Recognition. In Proceedings of the IEEE International Conference and Workshops on Face and Gesture Recognition, Shanghai, China, 22–26 April 2013.
30. Ramanathan, N.; Chellappa, R. Modeling age progression in young faces. In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006; pp. 387–394.
31. Deb, D.; Best-Rowden, L.; Jain, A.K. Face recognition performance under aging. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017.
32. Machado, C.E.P.; Flores, M.R.P.; Lima, L.N.C.; Tinoco, R.L.R.; Franco, A.; Bezerra, A.C.B.; Evison, M.P.; Aure, M. A new approach for the analysis of facial growth and age estimation: Iris ratio. *PLoS ONE* **2017**, *12*. [[CrossRef](#)] [[PubMed](#)]
33. Xu, C.; Liu, Q.; Ye, M. Age invariant face recognition and retrieval by coupled auto-encoder networks. *Neurocomputing* **2017**, *222*, 62–71. [[CrossRef](#)]

34. Park, U.; Tong, Y.; Jain, A.K. Face recognition with temporal invariance: A 3D aging model. In Proceedings of the 8th IEEE International Conference on Automatic Face & Gesture Recognition, Amsterdam, The Netherlands, 17–19 September 2008.
35. Yadav, D.; Singh, R.; Vatsa, M.; Noore, A. Recognizing age-separated face images: Humans and machines. *PLoS ONE* **2014**, *9*. [[CrossRef](#)] [[PubMed](#)]
36. Best-Rowden, L.L.; Jain, A.K. Longitudinal study of automatic face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 148–162. [[CrossRef](#)] [[PubMed](#)]
37. Cheong, Y.W.; Lo, L.J. Facial asymmetry: Etiology, evaluation and management. *Chang Gung Med. J.* **2011**, *34*, 341–351. [[PubMed](#)]
38. Sajid, M.; Taj, I.A.; Bajwa, U.I.; Ratyal, N.I. Facial asymmetry-based age group estimation: Role in recognizing age-separated face images. *J. Forensic Sci.* **2018**. [[CrossRef](#)] [[PubMed](#)]
39. Lee, K.W.; Hong, H.G.; Park, K.R. Fuzzy system-based fear estimation based on symmetrical characteristics of face and facial feature points. *Symmetry* **2017**, *9*, 102. [[CrossRef](#)]
40. Zhai, H.; Liu, C.; Dong, H.; Ji, Y.; Guo, Y.; Gong, S. Face verification across aging based on deep convolutional networks and local binary patterns. In *International Conference on Intelligent Science and Big Data Engineering*; Springer: Cham, Switzerland, 2015.
41. Wen, Y.; Li, Z.; Qiao, Y. Age invariant deep face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
42. El Khiyari, H.; Wechsler, H. Face recognition across time lapse using convolutional neural networks. *J. Inf. Secur.* **2016**, *7*, 141–151. [[CrossRef](#)]
43. Liu, L.; Xiong, C.; Zhang, H.; Niu, Z.; Wang, M.; Yan, S. Deep aging face verification with large gaps. *IEEE Trans. Multimed.* **2016**, *18*, 64–75. [[CrossRef](#)]
44. Lu, J.; Liong, V.E.; Wang, G.; Moulin, P. Joint feature learning for face recognition. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 1371–1383. [[CrossRef](#)]
45. Chen, B.-C.; Chen, C.-S.; Hsu, W.H. Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. *IEEE Trans. Multimed.* **2015**, *17*, 804–815. [[CrossRef](#)]
46. Chen, B.-C.; Chen, C.-S.; Hsu, W.H. Cross-Age Reference Coding for Age-Invariant Face Recognition and Retrieval. 2014. Available online: <http://bcsiriuschen.github.io/CARC/> (accessed on 28 January 2017).
47. Ricanek, K.; Tesafaye, T. MORPH: A longitudinal image database of normal adult age-progression. In Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06), Southampton, UK, 10–12 April 2006.
48. FERET Database. Available online: <http://www.itl.nist.gov/iad/humanid/feret> (accessed on 15 September 2014).
49. Face++ API. Available online: <http://www.faceplusplus.com> (accessed on 30 January 2017).
50. Ha, H.; Shan, S.; Chen, X.; Gao, W. A comparative study on illumination preprocessing in face recognition. *Pattern Recognit.* **2013**, *46*, 1691–1699.
51. Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep face recognition. In Proceedings of the British Machine Vision Conference, Swansea, UK, 7–10 September 2015.
52. Lin, L.; Wang, G.; Zuo, W.; Xiangchu, F.; Zhang, L. Cross-domain visual matching via generalized similarity measure and feature learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1089–1102. [[CrossRef](#)] [[PubMed](#)]
53. Wu, Z.; Ke, Q.; Sun, J.; Shum, H.-Y. Scalable face image retrieval with identity-based quantization and multireference reranking. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 1991–2001. [[PubMed](#)]
54. Jegou, H.; Douze, M.; Schmid, C. Product quantization for nearest neighbor search. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 117–128. [[CrossRef](#)] [[PubMed](#)]

