Vis Comput (2018) 34:1617–1630
https://doi.org/10.1007/s00371-017-1428-z

SURVEY

Review on the effects of age, gender, and race demographics on automatic face recognition

Salem Hamed Abdurrahim[1] · Salina Abdul Samad[1] · Aqilah Baseri Huddin[1]

Published online: 23 August 2017

Abstract The performance of face recognition algorithms is affected by external factors and internal subject characteristics. Identifying these aspects and understanding their behaviors on performance can aid in predicting the performance of algorithms and in designing suitable acquisition settings at prospective locations to enhance performance. Factors that affect the performance of face recognition systems, such as pose, illumination, expression, and image resolution, are recognized as face recognition problems. These are substantially studied, and many algorithms have been developed to tackle these problems. However, the influence of population demographics (i.e., race, age, and gender) on face recognition performance has not received considerable attention. Early findings that deal with demographic influence give conflicting results. The studies conducted in the last decade resolve some of the contentions. Nonetheless, some findings have not reached consensus. Existing reviews on the influence of covariates are either outdated or do not cover the influence of demographic covariates on the performance of face recognition algorithms. This paper gives an intensive and focused review that covers recent research on demographic covariates. The effects of age, gender, and race covariates on face recognition are summarized based on these findings, and suggestions on the future direction of the field are given to have a significant understanding of these effects individually and their interactions with one another.

Keywords Face recognition · Demographic covariates · Race · Gender · Age

✉ Salem Hamed Abdurrahim
salemeez@siswa.ukm.edu.my

Salina Abdul Samad
salinasamad@ukm.edu.my

Aqilah Baseri Huddin
aqilah@ukm.edu.my

1 Department of Electrical, Electronic, and System Engineering, (UKM) Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

1 Introduction

In biometric systems, the human face has been a vital recognition modality and has been an active research area for the past several years because of the wide range of applications in crowd surveillance, security systems, border control, building accessibility, law enforcement, the identification of missing children, and the verification of duplicate enrollments. Moreover, the human face is widely used in many countries for identity verification via electronic passport gates and in visa screening by immigration departments [1–3].

The US Government has supported many research projects that investigate the performance of automatic face recognition algorithms [4]. Many papers have been published to measure the effects of covariates, such as illumination, pose, and expression on the performance of face recognition algorithms [5–9]. These findings provide useful information for a substantial understanding of the area of face recognition algorithms and their underlying face image formation as expressed through recognition performance. Understanding the factors that affect performance is essential in developing, evaluating, and operating face recognition algorithms. Studies corroborate that covariates commonly have effects on the performance of face recognition algorithms [10].

Viewing angle (pose) and lighting (illumination) are covariates that have the greatest effect on face recognition performance. Yale [11] and PIE [12] databases are developed to thoroughly study pose, illumination, and their interactions. The outcome is substantial literature on pose and illumination [10], which allow algorithms to outperform humans in matching face images across changes in illumination [4,13,14].

The look of a face does not only rely on pose, illumination, expression, or the resolution of the sensors, which can be controlled at the stage of face image enrollment, but it is also defined by age, gender, race, and distinctiveness of the face. These are attributes of a person being recognized and are person-specific covariates. Face recognition algorithms detect someone as a unique individual. The diversity of faces means that face recognition algorithms must perform through a backdrop of appearance variability; thus, face recognition algorithms targeted for real-world applications must function predictably over changes in the demographic composition of the anticipated application population [15,16].

Although the demographic covariates (i.e., age, race, and gender) affect face recognition performed by humans [17] and machines [18], the influence of population demographics on face recognition performance has not received considerable attention [6,8,18].

This paper reviews and summarizes what is knowing about the effects of demographic covariates on face recognition performance from literature. The conflicting results that have been reported are discussed and summarized. Moreover, suggestions are given for studying the demographic effects on face recognition.

2 Face recognition covariates

A covariate is a variable that has an effect of increasing the intra-class variation, decreasing the inter-class variation, or both. Examples of image covariates that affect the performance of face recognition are pose, illumination, expression, and image resolution. Covariates that can be controlled during image acquisition are the constraints of the face recognition system. Limitations introduced by these covariates are substantially studied in the literature, and many algorithms have been developed to address these problems.

The covariates of a person, which cannot be controlled, (i.e., race, gender, age groups, and aging), also affect the performance of face recognition algorithms. Aging has been researched extensively, and relevant databases have been collected to help researchers solve the aging problem.

2.1 Controlled covariates

Controlled covariates are the properties of an image, which can be controlled by the user in an environment or applications. For example, a traveler presents himself at an electronic passport gate at a border control; a photo is taken when the individual faces the camera and acts naturally without

Fig. 1 Examples of images showing pose, illumination, and expression variations from the FEI [19] and AR [20] databases

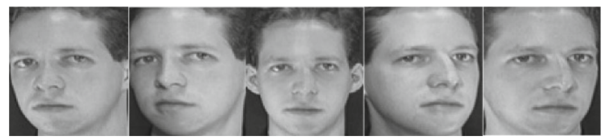

Fig. 2 Examples of face images with different poses from the AT&T database [23]

expression. Figure 1 is an example showing a natural face, a non-frontal posed, lighting variation, and an expressive face.

2.1.1 Pose

Pose in face recognition refers to face images whose poses are different from the gallery (known) images. Pose variation occurs during acquisition when a person is not facing or not looking straight at a camera. Different angles and locations appear in the captured image; thus, some of the face features are not visible for recognition. Accordingly, the accuracy of the face recognition algorithm will reduce significantly. The face image differences caused by poses are larger than the interpersonal differences used to differentiate identities [21]. Face recognition across pose variations has received considerable attention in the research community, and several promising approaches have been proposed for addressing the pose problem [21,22] (Fig. 2).

2.1.2 Illumination

Illumination variation in an uncontrolled environment is a challenging problem in face recognition. Depending on the position of the light source with respect to the camera and the captured 3D structure of the human face, facial image is viewed with several variations having shading and shadows. These variations in the visual aspects of a face can be larger than the variation caused by its other features [24], which affect the performance of face recognition algorithms [22,25]. The illumination variation has been extensively discussed in face detection and recognition research and has led researchers in the field to develop numerous methods to address the problem [26] (Fig. 3).

Fig. 3 Examples of varying illumination conditions from the Yale database [27]

2.1.3 Expression

The human face continually displays a series of facial expressions, unless she or he is quiet and motionless. Humans use these expressions to show their emotional status. These expression variations result in the deformation in local facial structure and the variations of the facial appearance and geometry [22] (Fig. 4).

2.1.4 Occlusion

Occlusion can be caused by hair, eyeglasses, sunglasses, scarves, handkerchiefs, and hats. Face occlusion is one of the most challenging problems encountered in applications of automatic face recognition, because some parts of human faces, especially the facial features, are missing [29]. Face occlusion has received considerable attention in recent years [30] (Fig. 5).

2.1.5 Image resolution

A captured face image can be of different resolutions, which depend on how the image was captured. A video camera sometimes produces a low-resolution image, whereas a scanner can scan a document image of high resolution. An unstable camera or a lack of focus produces a blurry image, and insufficient exposure or aperture captures an image with low contrast [31]. Low-resolution images degrade the performance of face recognition, which makes accurate recognition challenging [32] (Fig. 6).

2.2 Uncontrolled covariates

Covariates, which are uncontrolled, inherit the properties of the face of a person, such as aging effects, age group, race, and gender.

Fig. 5 Example images showing occlusions from the AR database [20]

Fig. 6 Example images of different resolutions from the FRGC database [33]

2.2.1 Aging

Aging variation is a type of within-class appearance variation in human faces and refers to the face recognition problem where a large time difference exists between the acquired face images of the same person. Aging variation can be intense, which results in significant alterations in the overall facial appearance of individuals (Fig. 7). Even discriminatory facial characteristics can be affected significantly by changes in age [34]. Although all of the studies that dealt with the aging problem agree that performance deterioration occurs when the age gap increases between the query and the target image, the deterioration is only substantial when the gap is in years and not in months or days [10,35]. Most existing age-related studies for face image analysis focus on age estimation [36–39] and age simulation [37,40,41].

Researchers have started to focus on face recognition across ages only in recent years. However, face recognition remains a challenging problem [42]. This scenario is due to the existing databases used to study the aging effect on face recognition performance, databases such as MORPH and FG-net, which are small and contain other uncontrolled variations (e.g., pose and illumination) [43]. Figure 7 shows different images of the same individual at different time laps taken from the FG-NET database.

2.2.2 Age groups

Age groups refer to the difference between the ages and images of the two people involved. A person from any age group finds that they look similar to other people from the

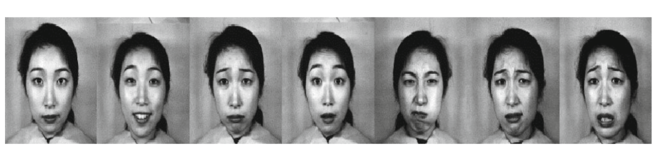

Fig. 4 Example images showing prototypic expressions from *left* to *right*: neutral, enjoyment, sadness, surprise, anger, disgust, and fear. Images are taken from the JAFFE Japanese Face Expression database [28]

Fig. 7 Images of the same individual at different ages from the FG-NET database [44]

same age group, as opposed to someone from another age group.

2.2.3 Gender

Men have faces that differ from women in local features and shapes. Men's chins are thicker than women's, whereas women's cheeks appear smoother than men's. Women's noses are commonly shaped smaller than men's. Men are also distinguished from women based on hairstyles and makeup [45]. Anthropological studies have also confirmed that the skeletal structure of males differs from that of females. However, boys and girls have similar skeletal structures, which makes gender classification difficult when intended for the young [46].

2.2.4 Race

The terms race and ethnicity are used interchangeably in the literature. However, race refers to the biological and physical formations (e.g., bone anatomy, hair, eye, and skin color). Different races can be represented by the difference among African-Americans, Caucasians, and East Asians. Ethnicity refers to the sociological differences: cultural heritage, nationality, language, regional traits, and religion. Different ethnicities can be represented by the difference between French people and Italians [47]. In the recognition field, we are substantially interested in race, which has an effect on performance due to the difference in face structure and color (Fig. 8).

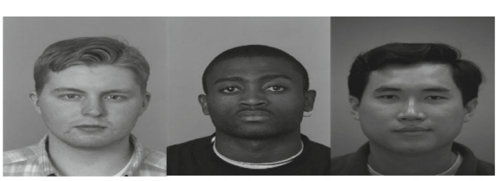

Fig. 8 Examples of three individuals from different races: the subject from the Caucasian race, African-American, and East Asian. Images are from the facial recognition technology (FERET) database [48]

3 Demographic covariate effects

3.1 Age groups

Since the emergence of the face recognition vendor test (FRVT) 2002, where test results concentrated on the effects of demographic covariates on the performance of face recognition, researchers have worked on age group cohorts and revealed some conflicting results. Almost all of the previous evaluations have demonstrated that older individuals are easier to recognize compared with younger ones [49].

Givens et al. [5] conducted some tests using the facial recognition technology (FERET) database and verified that young faces were more difficult to recognize compared with old ones; they assume that young faces have less character. Additional research was conducted using three different algorithms, and the results remained the same [6].

Givens et al. [6] studied how the covariates of the subjects may affect face recognition performance. They tested three well-known algorithms—principal component analysis, interpersonal image difference classifier, and elastic bunch graph matching algorithm—on the FERET database and evaluated their results with a statistical model. Their findings were consistent with those of other researchers; younger individuals are harder to recognize compared with older ones regardless of the algorithm used for recognition.

However, another research by Ho et al. [50] presented a different opinion. They suggested that the uneven distribution of ages in the databases used in the previous research led to different identifications for different ages. They used small samples in some of their results for some cases, but young people remained difficult to recognize.

Lui et al. [10] conducted an analysis on six covariates: age, elapsed time, gender, race, expression, and image resolution. These covariates were assumed important for facial recognition performance. Some of their results were consistent and proved all the previous results, but some were different. For example, older people were easier to recognize compared with younger ones. The result was not surprising because faces change over time. Considerable research is needed to explain the difference given that all of the previous results were consistent with one another.

Klare et al. [51] studied the demographic covariates using the Pinellas County Sheriff's Office (PCSO) database. This database is a collection of over one million mug shots used to analyze the effects of three demographic covariates of different cohorts, genders (male and female), races (Caucasians, African-Americans, and Hispanics), and ages (18–30 years old, 30–50 years old, and 50–70 years old). The present study tested the different demographic covariates using three commercial algorithms: Cognitec's FaceVACS v8.2, PittPatt v5.2.2, and Neurotechnology's MegaMatcher v3.1. They performed additional experiments with two non-trainable algorithms (local binary pattern-based and Gabor-based) and a trainable algorithm [spectrally sampled structural subspace feature (4SF)].

The database was partitioned equally between training and testing sets, and all sets had an equal number of males and females and African-Americans and Caucasians. The Hispanic race and old people were the least represented. They affirmed that all algorithm accuracies were affected by all three demographic covariates. The lowest accuracy is achieved specifically in the cohorts of African-American females, and young people 18–30 years of age when compared with the other cohorts.

3.2 Gender

Information about the influence of gender on face recognition algorithms and whether men's faces affect algorithms differently than women's faces is limited before the work done by Gross et al. [52]. Few reports on the subject exist, and the effects of the training samples were rarely reported. Moreover, the proportion of men to women in the database is often not mentioned [49,52].

Gross et al. [52] first used the AR database, which was small and has only 126 individuals but has a nearly balanced ratio of men and women (70 males against 56 females). Their findings were that women were easier to recognize compared with men by an average recognition rate of 5% difference. The recognition rate for women was 91.66%. They were surprised by the results and concluded that a significantly large database was needed to confirm their findings.

Lui et al. [10] conducted a meta-analysis on demographic covariates research between 2001 and 2008. The findings corroborated a small difference in the ratio of recognition between men and women. Nevertheless, men had a minimally better performance. Men were easily recognized in six studies, whereas women were easily recognized in five studies. No significant effect was reported in the other six studies. The recognition ratio gap between men and women decreased with age. Lui et al. concluded that the gender effects are not universal but are rather influenced by the algorithms and the dataset, the database was either of small size or was not divided equally between races, gender and age, and all studies used databases without controlling other covariates.

The effects of gender did not speed up until additional databases were available with considerable variations of gender and race. Guo et al. [53] studied the influence of gender on age estimation and confirmed that the performance was significantly affected by female subjects. Guo and Mu [54] conducted an experiment on the influence of the crossing of race and gender on age estimation. They observed that the crossing of gender and race significantly affected the accuracy of the algorithms' performance on age estimation. Both experiments were conducted on significantly large databases with an adequate number of subjects in terms of gender and race.

Grother et al. [55] also studied the gender effects on face recognition performance using commercial algorithms (i.e., L1 identity solution, Neurotechnology, Toshiba, Cognitec, NEC, Pittsburgh Pattern Recognition, and Sagem), which were submitted to Multi-Biometric Evaluation 2010 organized by the National Institute of Standard and Technology (NIST). The tests were evaluated using the law enforcement agency (LEO) database, which is a collection of mug shots collected by law enforcement agencies. Five of the seven algorithms tested were more accurate on males than on females, which was confirmed by Klare et al. [51], who found that females, African-Americans, and young people were difficult to recognize for all the algorithms used. The present study was the first attempt to examine the effects of gender or age used for training face recognition algorithms.

Ngan and Grother [46] recently used algorithms from Cognitec, Neurotechnology, NEC, Tsinghua University, MITRE, and Zhuhai-Yisheng; the algorithms were submitted previously to FRVT to be evaluated for gender classification. Tests were carried out on five different datasets (i.e., mug shot images collected over the years by different LEOs, visa images from visa applications, public database consisting of unconstrained facial images collected from the web, public database containing images collected from Flicker GROUPS labeled face in the wild (LWF), and sketch image datasets from the FERET database). They confirmed that gender classification for men is highly accurate for all algorithms when having balanced datasets of both genders.

3.3 Race

Race is another demographic covariate that can have a significant effect on the performance of the face recognition system. The role race plays in these systems must be considered in the field of computer vision and by physiologists [2,52,56] because of the differences in the shapes of faces across races [57].

The representative of a population in an intended application venue may vary from 100%, representing the majority of one race to different degrees of percentages that represent the presence of other races. This race variation in populations may also differ at different places (e.g., border crossing and airport gates) in different seasons (e.g., tourist season) or even at the same day (e.g., flights that carry a different racial group arriving in the morning and other flights that carry a different racial group arriving from a second location in the afternoon). Therefore, the demographic similarity between the faces of the same race decreases the matched face pairs, which affects the performance of an algorithm negatively [58]. Table 1 summarizes the work that has dealt with demographic covariates.

Table 1 Major work that dealt with demographic covariates

Authors	Year	Covariate	Cohort	Database	Algorithm
Akhtar et al. [59]	2013	Age, race, gender, glasses, and facial hair under the influence of aging	171 whites versus 460 nonwhites, 515 males versus 116 females, and 130 old versus 550 young	MORPH	Local binary pattern (LBP), multiscale local binary pattern, local phase quantization, local ternary pattern, elastic bunch graph matching (EBGM), scale invariant feature transform, and speed up robust feature
Mahalingam and Kambhamettu [57]	2012	Face verification of age-separated images with various internal (age, gender, and race) and external (pose, expression, facial hair, and glasses) factors on performance	Caucasians, African-Americans, Hispanics, Indians, and Japanese people	FG-net, MORPH, Indian face DB, and JAFFE	Haar local binary pattern (HLBP) + AdaBoost, HLBP + random forest for the effects of race, linear (PCA), and nonlinear HLBP + AdaBost are used.
O'tool et al. [58]	2012	The influence of aging on different facial components and demographic categories	Caucasians versus East Asians and Males versus Females	FRVT 2006 dataset (Notre Dame-sadia)	Fusion of the top three commercial algorithms submitted to FRVT 2006
Klare et al. [51]	2012	Age, gender, race, and their impact on the training of FR algorithms	Equal # of individuals: 8000 black and white individuals, 8000 males and females, and three different age categories	PCSO	Three commercials, two non-trainable (LBP and Gabor), and one trainable (4SF)
Beveridge et al. [16]	2010	Gender influence on race classification	Males and females, mainly Caucasians with some East Asians	FRVT 2006 dataset (Notre Dame-sadia)	Top three commercial algorithms submitted to FRVT 2006
Philips et al. [4]	2010	Tested the other-race-effect phenomena on humans and machines	Caucasians versus East Asians	FRVT 2006 dataset (Notre Dame-sadia)	Fusion of the eight algorithms submitted to FRVT 2006
Lui et al. [10]	2009	Meta-analysis on age, race, and gender covariates	N/A	N/A	N/A
Beveridge et al. [9]	2009	Age, gender, race, distance between the eyes and the image focus	Males and females, mainly Caucasians with some East Asians	FRVT 2006 dataset (Notre Dame-sadia)	Fusion of the top three commercial algorithms submitted to FRVT 2006
Ho et al. [50]	2007	Age using an age-balanced gallery	Mainly Caucasians and some East Asians and Black Americans	FERET	PCA
Givens et al. [7]	2005	Age, gender, bangs, facial hair, eyes, and their influence on verifications	Mainly Caucasians and some East Asians and Black Americans	FERET	PCA using cosine distance measure
Givens et al. [5,6]	2003, 2004	Age, gender, race, skin, glasses, facial hair, makeup, bangs, expression, mouth, and eyes	Mainly Caucasians and some Asians and Africans	FERET	PCA, interpersonal image difference classifier, and EBGM

3.3.1 Other-race effect

Physiological researchers [17,60,61] affirmed that people can identify a face of their own race much easier than a face of other races.

Most of the research done by physiologists on the other-race-effect phenomena agrees that the result is consistent across cultural and racial diversities. However, no clear contention is shown about the social or cognitive drives that implicate this effect [62].

Givens et al. [6] used the FERET database on three well-known algorithms (i.e., principle component analysis, an interpersonal image difference classifier, and elastic bunch graph matching algorithm); non-Caucasians are significantly recognized with all algorithms even when the system trained on the majority race, which is Caucasian.

Beveridge et al. [8] studied the difference on the verification rate of different races on 345 matched pairs, wherein most were Caucasians and East Asians. The rate of successful verification was good for East Asians, which has been verified by Phillips et al. [63], whose study was motivated by a surprising result in the 2006 NIST FRVT, where numerous tested algorithms for face recognition from commercial and academic contenders showed similar characteristics. Algorithms handed in for testing by East Asians performed better on East Asians compared with algorithms developed by western countries, which is also true for Caucasians. Algorithms developed by western countries obtained a significant result.

3.3.2 Majority race training effect

Most researchers expected that the training set for a face recognition algorithm affects its performance with respect to different demographic groups.

All face recognition algorithms submitted for evaluation in the FERT 2002 evaluation had prior training on sets by vendors. Therefore, the algorithms were generalized and not tuned to specific face gallery. That is, datasets used for training were not restricted. All algorithms, thus, had prior training on different datasets. Nevertheless, all systems were affected by the covariates, which suggests that training does not play a role in the effect of covariates on the algorithm performance [49].

Mahalingam and Kambhamettu [57] examined the influence of race on face verification performance. The performance deteriorated when the system was trained with multiracial groups compared with the performance when the system was trained with one race and tested with the same or different racial groups.

Grother et al. [55] studied the race effect using commercial algorithms (i.e., L1 identity solution, Neurotechnology, Toshiba, Cognitec, NEC, Pittsburgh Pattern Recognition, and Sagem), which were submitted to Multi-Biometric Evaluation 2010 organized by the NIST. The tests were evaluated using the LEO database. They corroborated that the race affects the performance of each algorithm. However, some algorithms found that African-Americans are easier to recognize compared with Caucasians, whereas other algorithms found that East Asians and Indians were the most difficult to recognize. These conflicting results are probably due to the different training sets used to train each algorithm prior to testing.

Lui et al. [10] presented a meta-analysis for six covariates (i.e., age, gender, and race). However, the studies could not confirm whether one race has more or less influence on recognition compared with the other. The reason for the race effects reported in the literature is that the researchers used imbalanced datasets between the racial groups (i.e., races are unevenly presented). Therefore, the race that has the least representatives is the easiest race to be recognized. Another explanation for the phenomenon is that intrinsic differences exist between races, which led to the performance differences. One of the suggestions to solve this contention is a study to test the racial effect on face recognition performance, where all races are evenly distributed in the database. However, such database does not exist, and such research is yet to be reported in the literature.

The three consecutive studies by Beveridge et al. on the effects of covariates on the algorithm performance using a dataset in which Caucasians were the majority race, followed by East Asians. Verified that all races were easy to recognize, except for African-Americans, although their sample size was relatively small compared with East Asians or Caucasians. However, Beveridge et al. found that East Asians were always the easiest to recognize [8,9,16].

This was also verified by Klare et al. [51], who went further and were the first to examine the impact of demographic training on face recognition algorithms. All three demographic covariates were tested: (gender [males and females], with three race cohorts [Caucasians, African-Americans, and Hispanic], and age; individuals were partitioned into three age groups [18–30, 30–50, and 50–70 years old]). The data subset was from MORPH-II with an equal number of all covariates involved in the test. The performance of all algorithms was consistent in that they yielded low recognition rates on cohorts of females, African-Americans, and young individuals 18–30 years old.

Also Phillips et al [63] in the 2006 NIST face recognition algorithms tests, where the tests were on a range of false accept rates, found an advantage for the East Asian faces at a low false accept rate, which is typical in any security application. The second experiment used a small equal number of faces of both races. Both algorithms were good on Caucasians although Western algorithms showed better performance. This proves an advantage for East Asian faces. The results are consistent with the results from the FRVT 2006 data, where only matched face pairs were considered. The authors also suggested that the advantage for the East Asian faces is due to the low false acceptance rate. The authors also thought that all algorithms have been tested on a majority of Caucasian faces given that the algorithms have been submitted to FRGC prior to FRVT 2006 and that the dataset used at FRVT was collected at the same site as FRGC, which had 70% of Caucasian faces and only 22% of East Asian faces. Consequently, all algorithms had prior experience with the majority of Caucasians, which means that training has no effect on other-race effect.

Akhtar et al. [59] evaluated the effects of demographic covariates evaluated on six baseline facial representations based on local features (local binary patterns, multiscale local binary patterns, local phase quantization, local ternary patterns, elastic bunch graph matching, scale invariant feature transform, and speeded up robust feature). These algorithms do not require training for learning and efficient performance. The dataset used for testing consisted of 631 subjects of covariates—race (171 subjects of Caucasians origin and 460 others), age (550 young subjects and 130 old subjects), and gender (515 males and 116 females) from the MORPH database. Their findings validate that older people were easier to recognize compared with younger people and that males were easier to recognize compared with females, which were consistent with all other studies. However, the authors raised their concern given that their result on race covariate was contrary to other results, where Caucasians were easier to recognize compared with non-Caucasians. The majority of non-Caucasians in the MORPH database are African-Americans with few of the other races, which is consistent with the study by Beveridge et al. [9], which revealed that Caucasians were easier to recognize compared with African-Americans.

4 Demographic covariate interactions

In a study by Phillips et al. [49], FRVT 2002 used the HCInt database, which mainly had Mexicans with some East Asians and had an equal number of men and women. They tested the effects of gender on the performance of algorithms, but the database contained more young people than old people. They found that males were easier to recognize compared with females. They performed two additional experiments based on their findings. The first experiment evaluated the gender effect, whereas the other evaluated the interaction between gender and age. In the first experiment, they divided the dataset into two, with an equal number of men and women. The rate of identification for the men-only data set was greater than that for the women-only dataset.

In the second experiment, they created a single dataset, which had an equal number of men and women. The dataset was portioned to an equal number of age categories. The results were consistent with the first experiment. Age, gender, and the interaction between the two affect the performance of face recognition algorithms. They also noticed that the performance gap between males and females shrunk by age. Lui et al. [10] confirmed that age interacts with gender (i.e., the effects of gender decrease as people age).

A series of tests [53,54] on the influence of demographics on the estimate of age and race affirmed that the performance changed tremendously between crossing age and gender and non-crossing when estimating age, which means that covariates interact with one another.

Guo and Mu [64] conducted an experiment to find out whether an interaction with age and gender variations exists when performing race classification. They used the MORPH-II database, which had a majority of African-Americans, but constructed a dataset of an equal size of race, gender, and age groups. They corroborated that, when classifying people by race, performance would decrease by 6–8% if the training had been done using females and the testing using males. However, the performance shows no significant change if testing was done in reverse (i.e., training using males and testing using females). This scenario suggests a difference in the recognition influence between males and females and that demographic covariates interact with one another.

Ngan and Grother [46] recently conducted some gender classification tests on 240 thousand visa images taking under controlled illumination, pose, and facial expression. The classification of males was more accurate than females, and the algorithms' performance decreased in classifying females as age increased. Gender classification was more accurate for adult males (aged 21–60) than for young males (aged 0–10), which suggested that gender and age were interacting. The results also corroborated that East Asians were the most difficult to classify among other races and that East Asian males were commonly classified as females. These results suggested a strong interaction among the three demographics: age, gender, and, race. Their results relied on empirical observations without knowing the cause for the results.

Farinella and Dugelay [65] conducted some tests using a dataset of 200 males and females and 200 Caucasians and non-Caucasians from two databases: FERET [48] and TRECVID [66]. However, the number of races and gender was uneven because the FERET database is small and mainly contains Caucasians. The algorithms used three different feature extractors (i.e., pixel-based local binary pattern LBP, and HOG histogram-oriented gradients). They used a support vector machine as the classifier. They confirmed that gender and race have no effect on each other during classification.

5 Demographic databases

Face database is a collection of face images that can be used to test and evaluate the performance of face recognition systems [31]. The results of algorithms tested are a reflection of the database used to develop them [48]. The following are the most used databases for evaluating the algorithm performance under the influence of demographic covariates.

5.1 Face and gesture recognition research network

The face and gesture recognition research network (FG-NET) provides an image database containing face images that show a number of subjects at different ages. The FG-NET is one of the first publicly available face databases with real ages provided for each subject. Moreover, the FG-NET has played an important role in assisting researchers in investigating the effects of aging on facial appearance. This database contains 1,002 face images of only 82 subjects, with approximately 12 images per subject at different ages, the minimum age of 0 (<12 months), and the maximum age of 69. The age distribution of this database is strongly biased to younger ages (<18 years). In addition, this database contains no data elements on key parameters that affect the appearance of faces across adulthood [67].

5.2 AR

AR [20] is a publicly available database, which has a balanced ratio of males and females. This database consists of over 3000 images for 126 persons—70 are males, and 56 are females. The images are colored and are of good quality resolution. Images were photographed under different controlled lighting conditions with various expressions and occlusions, and the subjects were photographed twice over the interval of two weeks.

5.3 MORPH

MORPH is a large database of mug shots collected in real-world conditions (not a controlled collection). Each shot has associated metadata that contains age, gender, race, height, and weight, which are important covariates for understanding the diversities of the human face information [68]. This database is divided into two sections: MORPH Albums 1 and 2. MORPH Album 1 is small and contains 1690 images from 625 different subjects (∼2.7 images/subject). MORPH Album 2 has approximately 55,000 face images of more than 13,000 subjects, in which approximately 77% of the images are Black faces; 19% are Caucasians; and the remaining 4% includes Hispanics, Asians, and Indian. The distribution of gender and race is uneven although the database is large. The male to female ratio is approximately 5.5:1, and the Black to Caucasian ratio is approximately 5:1 [68].

5.4 Facial recognition technology

The FERET [48] database is of frontal facial images collected between the years 1993 and 1996 under the FERET program, sponsored by the US Department of Defense [69]. This database contains images of over 1000 individuals with some metadata recorded with the images such as facial points, ethnicity, date of birth, gender, and illumination [65] and is distributed by the NIST. FERET has been used for studying gender classification but is not well suited to study age or race influence on algorithms given that the race distribution in this database is significantly biased to Caucasians and that the age distribution of individuals is highly concentrated toward young subjects (e.g., 20, 30, and 40) [70].

5.5 PCSO

PCSO is a large database collected from the State of Florida by the Pinellas County Sheriff's Office. This database contains mug shot images with metadata, which include the image capture date, date of birth, gender, and race of mainly Caucasian to which considerable attention is paid, followed by African-Americans. In addition, PCSO has been used extensively in demographic research [15].

5.6 Asian face image

The Asian Face Image database contains face images of 103 individuals with equal numbers of males and females of Asians, which are mainly Korean. The images were taken with different lighting conditions, poses, and expressions [71].

5.7 Cross-age celebrity dataset

Cross-age celebrity dataset [42] is a large collection of the images of celebrities whose age range is from 16 to 62 with the age gap of up to 10 years for each celebrity name collected from the Internet. This dataset contains more than 160000 images for 2000 celebrities.

5.8 Ethnicity, gender, and age database

Ethnicity, gender, and age (EGA) face database [72] is an integration of six different datasets, namely FERET, FRGC, CASIA-Face V5, FEI, JAFFE, and Indian Face DB, into a single-face database. EGA contains over 2000 images of 469 subjects. This database is organized according to the age, gender, and race of subjects. Most images are frontal with limited extrinsic effects, such as pose, illumination, and expression. Moreover, EGA is divided into five races, namely Caucasians, East Asians, Black Africans, Indians, and Latinos. Each race was further divided into two genders: male and female. These two groups were further divided into three age categories: young, adult, and middle-aged. Table 2 shows the list of available databases commonly used to test face recognition algorithms, which include the effects of gender and race.

6 Observations and suggestions

6.1 Age

A general agreement from the reviewed literature states that young individuals are more difficult to recognize by face recognition systems compared with old people. However, most databases used had fewer old people than young people, which means that the number of individuals in each age category is inconsistent. This led some researchers to argue about the difference in the performance of the algorithms. The only way to clear this dispute is to use a database that has an adequate and equal number of all ages spread over the database.

6.2 Gender

The research done has a general agreement on the effects of gender on the performance of face recognition algorithms. Tests have been conducted using trainable and non-trainable algorithms. Most results confirmed that females are harder to recognize compared with males. Some researchers assume that the lower recognition of females compared with males is due to other covariates, such as makeup that was not controlled during the studies [6]. Some studies corroborated that gender interacts with other covariates to affect the performance of recognition algorithms. For example, the effects of females on the algorithm performance decreased as age increases [10,49].

6.3 Race

Research relating to race effects on face recognition algorithms presented many conflicting results.

- The race that is heavily presented in the database produced the best result.
- The least represented race in the database produced the best result.
- Caucasians are easier to recognize compared with non-Caucasians.
- Non-Caucasians are easier to recognize compared with Caucasians.

The only way to prove which of the statements is correct is to test the influence of races on the same algorithms using the same number of individuals for all races.

6.4 Database

The biggest hurdle researchers encounter when dealing with covariate effects is the absence of relevant databases for the study of demographic covariates that comprise all of the requirements, including an adequate number of faces with a wide range of ages for each individual, a wide range of ages spread over all databases, and an equal distribution of gender and race. The databases presently used by researchers when studying the intrinsic user-related feature (e.g., race and gender) lack composition or diversity of the significant requirement to extensively study the influence of demographic covariates on face recognition algorithms.

Many factors affect the recognition of a face, such as identity, illumination, facial expression, pose, age, occlusion, gender, race, and facial hair. A database of an adequate size that possesses carefully controlled variation of these factors is needed to develop an algorithm that is robust to the influence of these factors [73]. Moreover, the availability of public databases is important for the advancement of the research field given that they allow researchers to quickly become engaged in research work. The availability of public databases may have a great impact on the field in cases where data collection demands an intensive task [74]. However, many of the databases are made specifically for the algorithms under development [1]. None of the available databases are specialized for demographic covariates, even other existing specialized databases, such as the aging database, contain other covariates, such as illumination and pose [59]. Therefore, the sole impact of any covariate on the degradation of performance is difficult to assess. A database should have high-quality images of several races and should be large enough for the results to be substantially general. The ratio between males and females should be balanced and represent all ages.

Collecting databases relies on volunteers who are willing to have their faces photographed or to submit their personal collection of photographs to the database collectors. Therefore, databases are small because few people are willing to jeopardize their privacy for something that has no return value to them. Databases are heavily influenced by young individuals because they rely on volunteers, which are commonly young. Databases are gathered in a certain geographical location; thus, all presently used databases are heavily biased toward a single race.

One encouraging exception is an attempt to collect a relevant demographic database, EGA database, which is a pre-annotated face datasets by Riccio et al. [72]. These datasets consist of subjects collected from six different

Table 2 Available databases commonly used to test face recognition algorithms, which include the effects of gender and race

Database	No. of people, no. of faces	Race	Age	Gender	Time lapse	Availability	Controlled variation
MORPH-II	>13000, 55000	77% African-Americans, 19% whites, 4% others	16–99	46645 males, 8487 females	164 days	Free	Age
AR	126, >4000	Caucasians	N/A	70 males, 56 females	2 weeks	Free	Pose, expression, illumination
FERET	>1000, >14000	Mainly Caucasians	Young (20, 30, and 40)	Males, females	2 years	Free	Pose, expression, illumination
FG-NET	82, 1002 images	Caucasians	0–69, biased toward young <40	48 males, 34 females	N/A	Free	Age
PCSO	>45000, >1 million mug shots	15996 Caucasians, 26457 African-Americans, 4439 others	17–68 uniform age distribution	Males, females	N/A	Free	Pose, expression, illumination
EGA	469, >2000	Caucasian, East Asian, African, Indian, Hispanic	32.6% young, 48.5% adult and 18.9% middle-aged	Males, females	N/A	Free (collect each dataset individually)	Pose, expression, illumination
Notre Dame dataset from FERET	>300, >15000	Caucasians, East Asians	18–40	Males, females	11 weeks	Free	N/A
ASIAN DATABASE	103, 1751	103 people mainly Korean	N/A	53 males, 50 females	N/A	Private	Pose, expression, illumination
CAS-PEAL	1040, 99594	Chinese faces	N/A	595 males, 445 females	N/A	Free	Pose, expression, illumination
CADO	2000, 16000	Caucasians	16–62	Males, females	10 years	Free	Age

databases: CASIA-Face, FEI, FERET, FRGC, JAFFE, and Indian Face Database. Its diversity includes many images of multi-race, gender, and age. However, the dataset collected remains unbalanced in terms of race and age, which contain several young people and Caucasians and few African-Americans and old individuals. Another drawback is that the authors are unable to submit the EGA database for release because of an agreement with the original owners of the databases not to redistribute them. Accordingly, users must follow the link of the EGA database collector, download each database individually, and obtain usage permission from the original owners. Moreover, the acquisition settings are different for each database, which underscores the overall difficulty of finding the right database for testing the effects of covariates on face recognition algorithms.

7 Conclusions

One conclusion, which is consistent in most research conducted on the influence of demographic covariates, is that recognition accuracy of males is greater than that of females. The recognition ratio gap between men and women degraded with age. Second, older people are easier to recognize compared with younger people. Inconsistencies in terms of race have been reported to draw a general conclusion although other race and the majority of the training race effects have been observed. Moreover, studies suggest a strong interaction among the three demographics: race, gender, and age. The main obstacle to a comprehensive study of covariate effects is the lack of a dedicated database that comprises an adequate number of faces with a wide range of ages for each individual, a wide range of ages spread over all the databases, and an equal distribution of gender and race.

References

1. Li, S.Z., Jain, A.K.: Handbook of Face Recognition. Springer, New York (2005)
2. O'Toole, A.J., Natu, V.: Computational perspectives on the other-race effect. Vis. Cognit. **21**, 1121–1137 (2013)
3. Fu, S., He, H., Hou, Z.-G.: Learning race from face: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **36**, 2483–2509 (2014)
4. Phillips, P.J., Scruggs, W.T., O'Toole, A.J., Flynn, P.J., Bowyer, K.W., Schott, C.L., Sharpe, M.: FRVT 2006 and ICE 2006 large-scale experimental results. IEEE Trans. Pattern Anal. Mach. Intell. **32**, 831–846 (2010)
5. Givens, G., Beveridge, J.R., Draper, B.A., Bolme, D.: A statistical assessment of subject factors in the PCA recognition of human faces. In: Conference on 2003 Computer Vision and Pattern Recognition Workshop, CVPRW'03, pp. 96–96 (2003)
6. Givens, G., Beveridge, J.R., Draper, B.A., Grother, P., Phillips, P.J.: How features of the human face affect recognition: a statistical comparison of three face recognition algorithms. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2004, vol. 2, p. 381 (2004)
7. Givens, G.H., Beveridge, J.R., Draper, B.A., Phillips, P.J.: Repeated measures glmm estimation of subject-related and false positive threshold effects on human face verification performance. In: IEEE Computer Society Conference on 2005 Computer Vision and Pattern Recognition-Workshops. CVPR Workshops, pp. 40–40 (2005)
8. Beveridge, J.R., Givens, G.H., Phillips, P.J., Draper, B.A., Lui, Y.M.: Focus on quality, predicting FRVT 2006 performance. In: 8th IEEE International Conference on FG'08 Automatic Face and Gesture Recognition 2008, pp. 1–8 (2008)
9. Beveridge, J.R., Givens, G.H., Phillips, P.J., Draper, B.A.: Factors that influence algorithm performance in the face recognition grand challenge. Comput. Vis. Image Underst. **113**, 750–762 (2009)
10. Lui, Y.M., Bolme, D., Draper, B.A., Beveridge, J.R., Givens, G., Phillips, P.J.: A meta-analysis of face recognition covariates. Biometrics: Theory, Applications, and Systems. In: IEEE 3rd International Conference on 2009 BTAS'09, pp. 1–8 (2009)
11. Lee, K.-C., Ho, J., Kriegman, D.J.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans. Pattern Anal. Mach. Intell. **27**, 684–698 (2005)
12. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-pie. Image Vis. Comput. **28**, 807–813 (2010)
13. O'Toole, A.J., Phillips, P.J., Narvekar, A.: Humans versus algorithms: Comparisons from the face recognition vendor test 2006. In: 8th IEEE International Conference on 2008 Automatic Face and Gesture Recognition, FG'08, pp. 1–6 (2008)
14. O'Toole, A.J., Phillips, P.J., Jiang, F., Ayyad, J., Pénard, N., Abdi, H.: Face recognition algorithms surpass humans matching faces over changes in illumination. IEEE Trans. Pattern Anal. Mach. Intell. **29**, 1642–1646 (2007)
15. Otto, C., Han, H., Jain, A.: How does aging affect facial components?. In: Computer Vision-ECCV 2012. Workshops and Demonstrations, pp. 189–198 (2012)
16. Beveridge, J.R., Givens, G.H., Phillips, P.J., Draper, B.A., Bolme, D.S., Lui, Y.M.: FRVT 2006: Quo vadis face quality. Image Vis. Comput. **28**, 732–743 (2010)
17. Malpass, R.S., Kravitz, J.: Recognition for faces of own and other race. J. Personal. Soc. Psychol. **13**, 330 (1969)
18. Furl, N., Phillips, P.J., O'Toole, A.J.: Face recognition algorithms and the other-race effect: computational mechanisms for a developmental contact hypothesis. Cognit. Sci. **26**, 797–815 (2002)
19. Thomaz, C.E., Giraldi, G.A.: A new ranking method for principal components analysis and its application to face image analysis. Image Vis. Comput. **28**, 902–913 (2010)
20. Martinez, A.M.: The AR face database. CVC Technical Report 24 (1998)
21. Zhang, X., Gao, Y.: Face recognition across pose: a review. Pattern Recognit. **42**, 2876–2896 (2009)
22. Bhatt, H.S., Singh, R., Vatsa, M.: Covariates of face recognition, technical report submitted to (IIITD) Delhi Institutional Repository, 2015. https://repository.iiitd.edu.in/jspui/bitstream/handle/123456789/214/IIITD-TR-2015-002.pdf;jsessionid=81603766B00FAC5FA143C5AC6E24A979?sequence=1 (2015)
23. Samaria, F.S., Harter, A.C.: Parameterisation of a stochastic model for human face identification. In: Proceedings of the Second IEEE Workshop on 1994 Applications of Computer Vision, pp. 138–142 (1994)
24. Zou, X., Kittler, J., Messer, K.: Illumination invariant face recognition: a survey. Biometrics: theory, applications, and systems. In: First IEEE International Conference on 2007 BTAS , pp. 1–8 (2007)
25. Cheng, Y., Jiao, L., Cao, X., Li, Z.: Illumination-insensitive features for face recognition. Vis. Comput. 1–11 (2017). doi:10.1007/s00371-017-1357-x
26. Makwana, R.M.: Illumination invariant face recognition: a survey of passive methods. Proc. Comput. Sci. **2**, 101–110 (2010)

27. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intell. **23**, 643–660 (2001)
28. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with gabor wavelets. In: Proceedings of Third IEEE International Conference on 1998 Automatic Face and Gesture Recognition, pp. 200–205 (1998)
29. Ekenel, H.K.: A robust face recognition algorithm for real-world applications. Karlsruhe University, PhD thesis, Samsun Turkey (2009)
30. Jia, H., Martinez, A.M.: Face recognition with occlusions in the training and testing sets. In: 8th IEEE International Conference on 2008 Automatic Face and Gesture Recognition, FG'08, pp. 1–6 (2008)
31. Gorodnichy, D.O.: Face databases and evaluation. In: Li, Stan Z., Jain, Anil K. (eds.) Encyclopedia of Biometrics, pp. 295–303. Springer, Berlin (2009)
32. Sezer, O.G., Altunbasak, Y., Ercil, A.: Face recognition with independent component-based super-resolution. Electron. Imag. **2006**, 607705–607705 (2006)
33. Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. In: IEEE Computer Society Conference on 2005 Computer Vision and Pattern Recognition, CVPR, pp. 947–954 (2005)
34. Lanitis, A., Taylor, C.J.: Towards automatic face identification robust to ageing variation. In: Proceedings of Fourth IEEE International Conference on 2000 Automatic Face and Gesture Recognition, pp. 391–396 (2000)
35. Guo, G., Mu, G., Ricanek, K.: Cross-age face recognition on a very large database: the performance versus age intervals and improvement using soft biometric traits. In: 20th International Conference on 2010 Pattern Recognition (ICPR), pp. 3392–3395 (2010)
36. Han, H., Otto, C., Jain, A.K: Age estimation from face images: human versus machine performance. In: International Conference on 2013 Biometrics (ICB), pp. 1–8 (2013)
37. Geng, X., Zhou, Z.-H., Smith-Miles, K.: Automatic age estimation based on facial aging patterns. IEEE Trans. Pattern Anal. Mach. Intell. **29**, 2234–2240 (2007)
38. Choi, S.E., Lee, Y.J., Lee, S.J., Park, K.R., Kim, J.: Age estimation using a hierarchical classifier based on global and local facial features. Pattern Recognit. **44**, 1262–1281 (2011)
39. Li, Y., Peng, Z., Liang, D., Chang, H., Cai, Z.: Facial age estimation by using stacked feature composition and selection. Vis. Comput. **32**, 1525–1536 (2016)
40. Lanitis, A., Taylor, C.J., Cootes, T.F.: Toward automatic simulation of aging effects on face images. IEEE Trans. Pattern Anal. Mach. Intell. **24**, 442–455 (2002)
41. Wang, J., Shang, Y., Su, G., Lin, X.: Age simulation for face recognition. In: Proceedings of International Conference on Pattern Recognition, pp 913–916 (2006)
42. Chen, B.-C., Chen, C.-S., Hsu, W.H.: Cross-age reference coding for age-invariant face recognition and retrieval. In: Computer Vision-ECCV 2014, pp. 768–783. Springer, Berlin (2014)
43. Park, U., Tong, Y., Jain, A.K.: Age-invariant face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **32**, 947–954 (2010)
44. Facial image processing and analysis. https://fipa.cs.kit.edu/433.php#Downloads (2015)
45. Luckman, A.J., Allinson, N.M., Ellis, A.W., Flude, B.M.: Familiar face recognition: a comparative study of a connectionist model and human performance. Neurocomputing **7**, 3–27 (1995)
46. Ngan, M., Grother, P.: Face recognition vendor test (FRVT) performance of automated gender classification algorithms. National Institute of Standards and Technology, 2015. http://dx.doi.org/10.6028/NIST.IR.8052 (2015)
47. Ethnicity versus race- difference and comparison. http://www.diffen.com/difference/Ethnicity_vs_Race (2016)
48. Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.J.: The FERET database and evaluation procedure for face-recognition algorithms. Image Vis. Comput. **16**, 295–306 (1998)
49. Phillips, P.J., Grother, P., Micheals, R., Blackburn, D.M., Tabassi, E., Bone, M.: Report, Face recognition vendor test 2002. In: IEEE International Workshop on 2003 Analysis and Modeling of Faces and Gestures, AMFG , p. 44 (2003)
50. Ho, W.H., Watters, P., Verity, D.: Are younger people more difficult to identify or just a peer-to-peer effect. In: Computer Analysis of Images and Patterns, pp. 351–359 (2007)
51. Saavedra, M.B.F.: Evaluation methodologies for security testing of biometric systems beyond technological evaluation (2013)
52. Gross, R., Shi, J., Cohn, J.F.: Quo vadis face recognition? Technical report submitted to, Carnegie Mellon University, The Robotics Institute, Pittsburgh (2001)
53. Guo, G., Mu, G., Fu, Y., Dyer, C., Huang, T.: A study on automatic age estimation using a large database. In: IEEE 12th International Conference on 2009 Computer Vision, pp. 1986–1991 (2009)
54. Guo, G., Mu, G.: Human age estimation: What is the influence across race and gender?. In: IEEE Computer Society Conference on 2010 Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 71–78 (2010)
55. Grother, P.J., Quinn, G.W., Phillips, P.J.: Report on the evaluation of 2D still-image face recognition algorithms. NIST Interag. Rep. **7709**, 106 (2010)
56. Chellappa, R., Wilson, C.L., Sirohey, S.: Human and machine recognition of faces: a survey. Proc. IEEE **83**, 705–741 (1995)
57. Mahalingam, G., Kambhamettu, C.: Face verification of age separated images under the influence of internal and external factors. Image Vis. Comput. **30**, 1052–1061 (2012)
58. O'Toole, A.J., Phillips, P.J., An, X., Dunlop, J.: Demographic effects on estimates of automatic face recognition performance. Image Vis. Comput. **30**, 169–176 (2012)
59. Akhtar, Z., Rattani, A., Hadid, A., Tistarelli, M.: Face recognition under ageing effect: a comparative analysis. In: Processing of Image Analysis ICIAP 2013, pp. 309–318. Springer (2013)
60. Chiroro, P., Valentine, T.: An investigation of the contact hypothesis of the own-race bias in face recognition. Q. J. Exp. Psychol. **48**, 879–894 (1995)
61. Kelly, D.J., Quinn, P.C., Slater, A.M., Lee, K., Ge, L., Pascalis, O.: The other-race effect develops during infancy evidence of perceptual narrowing. Psychol. Sci. **18**, 1084–1089 (2007)
62. Meissner, C.A., Brigham, J.C.: Thirty years of investigating the own-race bias in memory for faces: a meta-analytic review. Psychol. Public Policy Law **7**, 3 (2001)
63. Phillips, P.J., Jiang, F., Narvekar, A., Ayyad, J., O'Toole, A.J.: An other-race effect for face recognition algorithms. ACM Transact. Appl. Percept. (TAP) **8**, 14 (2011)
64. Guo, G., Mu, G.: A study of large-scale ethnicity estimation with gender and age variations. In: IEEE Computer Society Conference on 2010 Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 79–86 (2010)
65. Farinella, G., Dugelay, J.-L.: Demographic classification: Do gender and ethnicity affect each other?. In: International Conference on 2012 Informatics, Electronics and Vision (ICIEV), pp. 383–390 (2012)
66. Smeaton, A.F., Over, P., Kraaij, W.: Evaluation campaigns and TRECVid. In Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval, pp. 321–330 (2006)
67. Panis, G., Lanitis, A., Tsapatsoulis, N., Cootes, T.: Overview of research on facial ageing using the FG-NET ageing database. IET Biom. **5**(2), 37–46 (2015). doi:10.1049/iet-bmt.2014.0053
68. Ricanek, Jr, K., Tesafaye, T.: Morph: A longitudinal image database of normal adult age-progression. In: 7th International Conference

on 2006 Automatic Face and Gesture Recognition, FGR, pp. 341–345 (2006)
69. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face-recognition algorithms. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1090–1104 (2000)
70. Han, H., Otto, C., Liu, X., Jain, A.K.: Demographic estimation from face images: human versus machine performance. IEEE Transact. Pattern Anal. Mach. Intell. **37**(6), 1148–1161 (2015). doi:10.1109/TPAMI.2014.2362759
71. Dong, H., Gu, N.: Asian face image database PF01,Technical Report submitted to Pohang University of Science and Technology, Korea 2001. http://citeseerx.ist.psu.edu/viewdoc/download?rep=rep1&type=pdf&doi=10.1.1.115.9985. Technical Report (2001)
72. Riccio, D., Tortora, G., De, Marsico, M., Wechsler, H.: EGA—ethnicity, gender and age, a pre-annotated face database. In: IEEE Workshop on 2012 Biometric Measurements and Systems for Security and Medical Applications (BIOMS), pp. 1–8 (2012)
73. Gross, R.: Face databases. In: Li, S., Jain, A. (eds.) Handbook of Face Recognition, pp. 301–327. Springer, Berlin (2005)
74. Panis, G., Lanitis, A.: An overview of research activities in facial age estimation using the FG-NET aging database. In: Workshops on Computer Vision-ECCV, pp. 737–750 (2014)