Demographic Analysis from Biometric Data: Achievements, Challenges, and New Frontiers

Yunlian Sun, Man Zhang, Zhenan Sun, and Tieniu Tan, *Fellow, IEEE*

Abstract—Biometrics is the technique of automatically recognizing individuals based on their biological or behavioral characteristics. Various biometric traits have been introduced and widely investigated, including fingerprint, iris, face, voice, palmprint, gait and so forth. Apart from identity, biometric data may convey various other personal information, covering affect, age, gender, race, accent, handedness, height, weight, etc. Among these, analysis of demographics (age, gender, and race) has received tremendous attention owing to its wide real-world applications, with significant efforts devoted and great progress achieved. This survey first presents biometric demographic analysis from the standpoint of human perception, then provides a comprehensive overview of state-of-the-art advances in automated estimation from both academia and industry. Despite these advances, a number of challenging issues continue to inhibit its full potential. We second discuss these open problems, and finally provide an outlook into the future of this very active field of research by sharing some promising opportunities.

Index Terms—Demographic estimation, biometrics, human age estimation, gender classification, race recognition

1 Introduction

Biometrics is the technique of automatically recognizing individuals based on their biological or behavioral characteristics [1]. Since the pioneering work on automated biometric recognition using fingerprint proposed by Mitchell Trauring in 1963 [2], significant progress has been achieved in this field [3]. Diverse biometric traits have been introduced and widely investigated, including fingerprint, iris, face, voice, palmprint, gait and so on. Along with the technological evolution, more and more practical applications have benefited from biometrics, covering border and access control, unique identification for residents, video surveillance, forensics, etc.

Biometric traits are typically exploited for human identity recognition, through identity-related information extraction. However, there is much more than identity in biometric data. For example, humans can deduce from someone's face photo a wide range of social information, including approximate age, gender, race and affect. The voice of a person can reflect his/her age, gender, affect, and accent, while handwriting can be used to estimate human age, gender, and handedness.

- *Y. Sun is with the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Beijing 100190, China and also with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China. E-mail: yunlian.sun@nlpr.ia.ac.cn.*
- *M. Zhang, Z. Sun, and T. Tan are with the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Beijing 100190, China. E-mail: {zhangman, znsun, tnt}@nlpr.ia.ac.cn.*

Manuscript received 10 July 2016; revised 4 Feb. 2017; accepted 9 Feb. 2017. Date of publication 13 Feb. 2017; date of current version 5 Jan. 2018. Recommended for acceptance by M. Tistarelli.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPAMI.2017.2669035

This is also demonstrated by diverse research work in perception, cognition, psychology, neuroscience and psychophysics. In 1966, Ptacek et al. showed that listeners were able to differentiate voices of younger adults from aged speakers with impressive accuracy [4]. Bruce et al. explored the perceptual basis of human ability to categorize the sex of faces and found people were remarkably accurate at this task [5]. In this work, we are interested in the analysis of demographics from biometric data. Demography is the study of population dynamics. It encompasses the study of size, structure and distribution of populations, and how populations change over time due to births, deaths, migration, and aging. Demographic analysis can relate to whole societies or to smaller groups defined by criteria such as education, religion, or ethnicity. As illustrated in Fig. 1, we focus on human age, gender and race estimation from biometric data.

Biometric demographic analysis has very diverse practical applications. Targeted advertising is a typical one. If the knowledge of clients (e.g., age and gender) can be automatically estimated by either their faces or voice, customized products and services can be recommended (e.g., makeup to females and toys to children). Human-computer interaction (HCI) is another widely explored application, where automatic demographic analysis can make the interaction more socially competent. Owing to these emerging applications, there has been an increasing interest in this topic, resulting in a large number of attempts and initiatives in both scientific and industrial communities: collection of specific biometric databases [6], [7], [8], [9], [10]; numerous research work on demographic estimation from various biometric traits [11], [12], [13], [14]; commercial products (How-Old.net from Microsoft,[1] face attribute prediction from Face++[2]); demographic specific workshops and special issues in pattern recognition, signal processing and computer vision conferences

1. how-old.net
2. www.faceplusplus.com

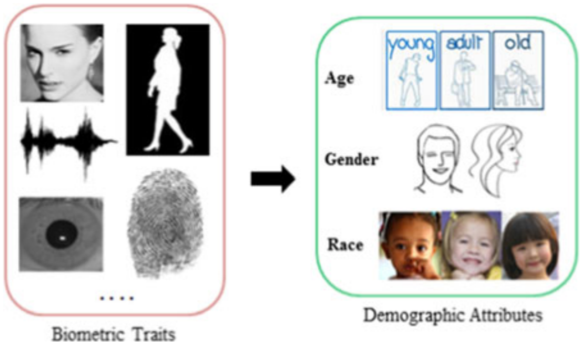

Fig. 1. Demographic attributes from biometric data (image source: Google Images).

and journals [15], [16]; organization of competitions [17], [18]. Thanks to these attempts, significant progress has been achieved. To gain a clear picture of the current panorama, we aim to, in this paper, keep track of these attempts and summarize the achievements.

Regarding biometric demographic estimation, there have been several survey papers in the literature. In [11], Fu et al. surveyed state-of-the-art techniques in face-based age synthesis and estimation prior to 2010. For gender recognition, Ng et al. presented a survey on computer vision-based methods, focusing on face and gait [14]. In 2014, Fu et al. provided a comprehensive review of face-based race estimation approaches [13]. Very recently, Han et al. offered a brief summary of existing demographic estimation methods from human face [12]. Notice that all these overviews center on either one demographic attribute or one biometric trait. The survey presented by Dantcheva et al. instead covers all soft biometrics (demographic, anthropometric, medical, material and behavioral attributes) from diverse biometric traits [19]. Another similar survey was organized by Nixon et al. [20], where some soft biometrics (demographic, clothing, and color attributes) are derived mainly from face and gait.

In spite of these efforts, there is no survey paper providing points of view about human perception for all the three tasks using various modalities. Second, no work presents a chain of chronological milestones related to the evolution of automated analysis. Third, for evaluating different systems, most surveys simply list estimation results (e.g., accuracy or error rate) in tabular form. Note that performance is a result of the algorithm as well as quality of biometric samples. Therefore, it is of significance to present details of both algorithm and data when talking about the state of the art. To address these issues, we organize this survey, which is more comprehensive, we believe, about covering the space of biometric modalities and demographic attributes than previous ones.

The rest of the paper is organized as follows: Section 2 presents some opinions about human demographic perception together with some popular applications. The historical development of automatic estimation is drawn in Section 3. Section 4 provides a comprehensive overview of existing techniques. Section 5 is devoted to the state-of-the-art performance. Several challenges faced by today's technology are listed in Section 6, followed by some future directions in Section 7. Finally, we conclude the survey in Section 8.

2 Background

In this section, we first provide some background work related to biometrics. Then, some interesting points about biometric demographic analysis will be presented from the view of human perception. A couple of typical applications will be further listed.

2.1 Biometrics

Biometrics is a technique developed for person recognition with reliability. Traditionally, people use token-based mechanisms (e.g., passport) and knowledge-based schemes (e.g., password) for verifying the identity. Compared with these traditional approaches, biometric systems exhibit more reliability owing to the distinctiveness, permanence, universality and invulnerability not a hundred percent satisfied by biometric modalities [1], which are often categorized as biological and behavioral traits. Biological traits include, but are not limited to fingerprint, iris, face, palmprint, Deoxribonucleic Acid (DNA), hand vein, palm vein, finger vein, periocular, ear, hand geometry, retina, sclera, Electrocardiograph (ECG), Electroencephalograph (EEG) and odour/scent. Behavioral traits are related to the pattern of human behavior, including voice, gait, handwriting, signature, typing rhythm, etc. In Fig. 2, we show diverse body traits that have been deployed in biometric systems or proposed in the literature.

Driven by explosively emerging real-world applications in access control, resident identification, surveillance and forensics, continuous efforts have been dedicated to tackle various problems involved in a biometric system, covering employing or designing application specific sensors to collect data, extracting representative features, developing robust matching algorithms, exploiting multimodal fusion strategies [21], investigating soft biometrics [22] and developing effective anti-spoofing approaches [23]. The evolution of biometrics is an on-going process. Nowadays biometric systems can achieve satisfactory performance in many applications, especially in controlled environments. In [3], Jain et al. give a clear depiction of its current state of the art.

2.2 Human Perception on Biometric Demographic Estimation

Estimating demographics from body traits is inherently a multidisciplinary task involving not merely fields of computer vision, pattern recognition and machine learning but also areas of perception, psychology, anthropometry, neuroscience and psychophysics. In this section, we provide some analytical understanding of human demographic perception.

2.2.1 *Age Perception*

Age estimation is the task of determining from various body traits the approximate age group (year range) or the scalar age value (year) for an individual. Human aging process is a slow, relentless, uncontrollable and irreversible process. Different people have different aging processes, which can be affected by both internal and external factors, including living style, working environment, health condition, etc.

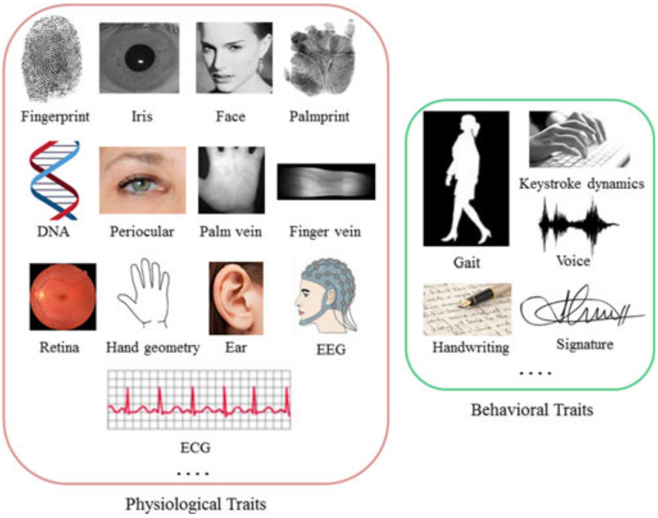

Fig. 2. Biological and behavioral body traits used for person recognition (image source: Google Images).

For example, in [25], Stone stated that facial aging can be accelerated by smoking, genetic predisposition, emotional stress, disease processes, dramatic changes in weight and exposure to extreme climates. Despite of the uncontrollability and personalization, human aging still has some general changes and shares some resemblances.

Human face provides us much information which is necessary and valuable to social intercourse. From the biological or anthropometric point of view, there are roughly two stages involved in facial aging which show large differences in facial growth and aging forms [26]. During the early development, from birth to adulthood, the greatest change is the craniofacial growth (i.e., shape change) [27]. Overall, the face size is getting larger gradually. The facial skin relatively does not change much. During adult aging, from adulthood to old age, the most perceptible change becomes skin aging (i.e., texture change). The skin becomes thinner, darker, less elastic, and more leathery. Wrinkles and blemishes gradually appear.

Voice is also a trait changing obviously throughout life. Advancing age produces physiologic changes that might alter the voice. These changes occur from birth to death and involve all parts of the vocal tract. During the childhood, the most significant change in voice results from the rapid growth of larynx, vocal folds and surrounding support structures. Throughout adult life, the mean fundamental frequency (F0) of females drops steadily from ≈ 225 Hz in $20 \sim 29$ years old to ≈ 195 Hz in $80 \sim 90$ years old [28]. For males, F0 drops until roughly the fifth decade, after which it rises gradually. Jitter is a measure of reflecting the periodicity of vocal fold vibration. In [29], Linville et al. reported higher mean jitter values in elderly than younger women. Significant jitter differences were also found between young and elderly men [30].

Another trait varying across ages is human gait. There have been studies revealing significant changes in gait patterns associated with the advancing age. In [31], [32], it was reported that gait speed decreased with increased age. Menz et al. found that elder subjects exhibited more conservative gait patterns [33], characterised by reduced velocity, shorter step length and increased step timing variability. These differences are particularly pronounced when walking on irregular surfaces.

2.2.2 Gender Perception

Gender prediction aims to determine if a person is a male or a female. Among various modalities, face and voice are probably the most widely investigated. Using photographs of adult faces with hair concealed, humans can achieve an accuracy as good as 96 percent in determining the gender [34]. Bruce et al. found the human ability to perform this categorization might be multiply determined by 2D, 3D, textural cues, and their interrelationships [5]. For gender from voice, Titze reported that adult males had pitches about an octave lower than adult females [35]. In [36], it was shown that listeners were able to correctly identify speakers' sex 88 percent of the time. Lass et al. indicated that F0 appeared to be a more important acoustic cue for sex identification than resonance characteristics [37].

Gait pattern exhibits also significant gender differences. Body sway, waist-hip ratio, and shoulder-hip ratio all are indicative of a walker's gender [38]. For example, males tend to swing their shoulders more than their hips, whereas females tend to walk in an opposite way. In [39], Kozlowski et al. showed that the sex of human walkers can be recognized without familiarity cues from displays of point-light sources mounted on major joints. Apart from using gait patterns, we human beings can discriminate genders when only static human body is available. In [40], Sheldon et al. qualitatively defined three components of body shapes, including endomorphy (soft and roundedness), mesomorphy (hardedness and muscularity), and ectomorphic (linearity and skinniness). In [41], Muñoz-Cachón et al. found that males displayed higher rates of mesomorphy, whereas endomorphy or relative body fat tended to be higher among females.

Handwriting is another gender informative trait. In [42], humans were asked to determine the writer's gender from a given handwriting document. An accuracy of about 68 percent was reported. Hand geometry shows also differences between males and females. In [43], Agnihotri et al. found that the average hand breadth and length were about 1 and 1.5 cm correspondingly greater in males than females.

2.2.3 Race Perception

Race is used to categorize humans into large and distinct populations or groups by heritable, phenotypic characteristics, geographic ancestry, physical appearance, and social status. Ethnicity is also commonly used to represent this categorization. Nevertheless, race and ethnicity are related to biological and sociological factors respectively. In this paper, we do not make a difference between them.

Race categorization is the task of predicting the racial group to which a person belongs. In [44], Hosoi et al. summarized some physical characteristics of three common racial groups, i.e., Asian (Mongoloid), European (Caucasoid), and African (Negroid). Asians generally have straight or slightly wavy hair and yellowish skin, whereas European people tend to have wavy or curly hair and light skin. For Africans, the skin is generally dark. Their hair usually has tight curls or heavy waves. Face is a popular trait in determining ethnicity with critical characteristics, including eyes, nose, lip, and facial skin [44]. For example, Asian people generally have narrow eyes with more single-edged eyelids, while Europeans generally have eyes with double eyelids.

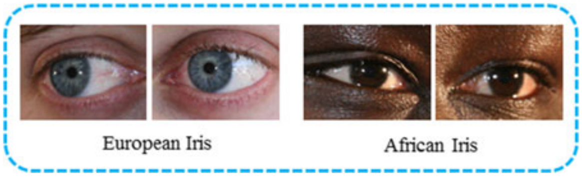

Fig. 3. Right and left iris images of one European subject and one African subject from the UBIRIS.v2 database [45].

Iris is also race informative. We illustrate in Fig. 3, the right and left iris images of one European and one African subjects from the UBIRIS.v2 database [45]. European iris usually exhibits bright and colorful appearance with clear texture. In contrast, African iris shows dark or brown appearance and it is impossible to obtain detailed texture with visible light.

2.3 Real-World Applications

Automated demographic analysis has many popular real-world applications. Here, we list several typical ones.

2.3.1 Human-Computer Interaction

More sophisticated HCI systems can be built if users' demographics can be collected automatically. Such systems are smarter and interact more naturally with humans. For instance, a communicative robotic can interact with users from different age/gender/race groups with particular preferences.

2.3.2 Security Control and Surveillance Monitoring

There are many security related situations where surveillance monitoring might play a vital role. A smart surveillance system with age estimation functionality can stop children from purchasing tobacco products from vending machines and entering bars or wine shops. By integrating gender estimation, such systems can restrict some areas to one gender only.

2.3.3 Multimedia Retrieval

For multimedia retrieval, demographic estimation can be used to locate specific individuals in video streams or crowd images. For example, when looking for photos of a specific person, estimating age as a preprocessing step will reduce the amount of search required.

2.3.4 Biometrics

Acting as soft biometrics, demographic attributes can be used for helping hard biometrics to further accelerate the matching process and improve the matching performance. For example, integrated with gender recognition, the time for searching the enrollment database can be reduced by half for an identification system.

2.3.5 Targeted Advertising

Targeted advertising is used to display advertisements which might interest consumers by simply analyzing which age group or gender they belong to. For example, the billboard may choose to show ads of cars when a male is detected, or dresses in the case of females.

3 Evolution of Automatic Demographic Estimation from Biometric Data

Since the work of Childers et al. in 1988 on automatically recognizing speakers' gender by voice [46], biometric demographic analysis has promoted a large amount of work on both developing automated algorithms and implementing practical systems. This has been specially the case for some of the most deployed modalities such as face, voice and gait. To get a clear picture of the progress, in the following, we list a chain of chronological milestones related to the evolution of each of the three demographic tasks. We are aware that other approaches merit being covered, however, due to space restrictions, we center only on those that, from our point of view, can better help the reader see the progress.

3.1 Historical Development of Automatic Age Estimation

The pioneering work on automated age estimation was started in the mid-1990s by Kwon and Lobo [48]. In their work, age categorization (babies, young adults and senior adults) was performed based on analysis of skin wrinkles and craniofacial changes. In Fig. 4, we list some major milestones.

In 2001, Davis investigated gait for visual discrimination of children from adults using stride-based properties [49]. Minematsu et al. instead derived age from voice using acoustic and prosodic features [50]. The Active Appearance Model (AAM) popularized by Lanitis et al. in 2002 considered both anthropometry/shape and texture/wrinkle of human face [51]. Other commonly used facial appearance models include Haar-like features [52], Local Binary Patterns (LBP) [53], Biologically Inspired Features (BIF) [57], etc. In [138], Metze et al. initialized an evaluation of different approaches on age grouping from telephony channel data. For describing human age, Yan et al. believe it makes more sense to express age as an interval than a fixed value [139]. Based on this, they learned an auto-structured regressor from uncertain nonnegative labels with face modality. For using voice, Bocklet et al. proposed to compute Gaussian Mixture Model (GMM) supervectors for each speaker [60]. In [58], [59], manifold embedding techniques were employed to learn a low-dimensional aging trend from face images.

In 2009, Gallagher et al. proposed a framework for facial age estimation by exploiting contextual features from group photos [239]. To address the low comparability of different paralinguistic algorithms developed for demographic estimation, Schuller et al. organized the first Challenge on Paralinguistics held at INTERSPEECH 2010 [17]. Towards using gait, Lu et al. computed Gabor features [63] from Gait Energy Image (GEI) [64] which is one of the most successful gait representations [155]. In 2011, an ordinal hyperplane ranking algorithm was presented for facial age estimation [65]. The first age estimation app for mobile marketing was launched by AppTech in 2012. It was developed for Android phones and can recognize the approximate age of a person from his/her face photo. In [212], Geng et al. estimated human age by regarding each face image as an instance associated with an age label distribution. In recent years, quite a few providers of biometric/face recognition technology have integrated age estimation module into

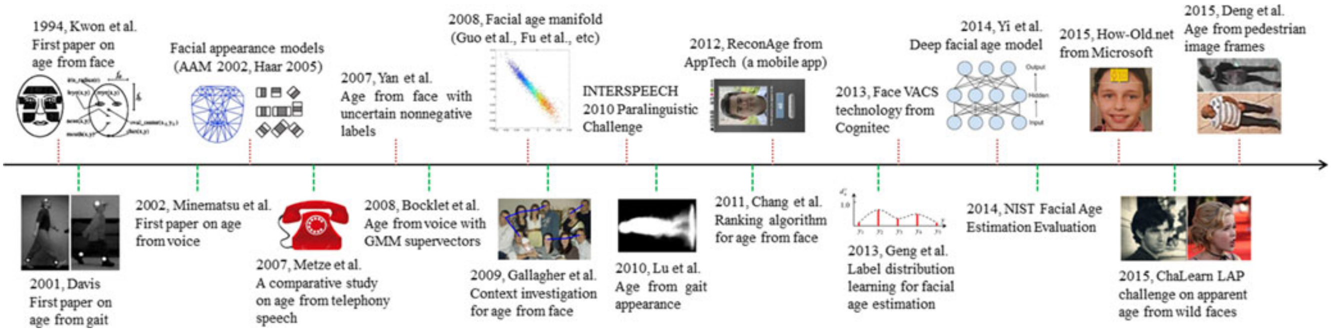

Fig. 4. Major milestones in the history of automatic age estimation from biometric data.

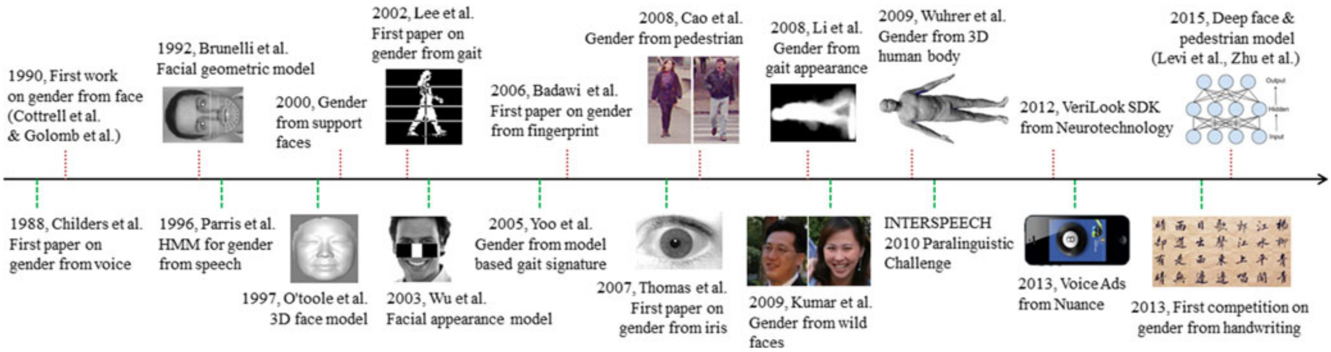

Fig. 5. Major milestones in the history of automatic gender estimation from biometric data.

their face recognition systems. Examples include Cognitec's FaceVACS-VideoScan[3] and Neurotechnology's VeriLook SDK.[4]

Motivated by the great success of deep neural networks in computer vision, language modelling and speech analysis, Yi et al. developed deep face models for age estimation [67]. In 2014, the National Institute of Standards and Technology (NIST) performed a large scale empirical evaluation of facial age estimation algorithms [66]. In 2015, Microsoft announced the How-Old.net website, on which users can upload their face photos and get estimated ages. To test current algorithms using "in the wild" faces (i.e., photos collected from web), ChaLearn organized the Looking At People (LAP) Challenge, focusing on apparent age estimation with age labels annotated by human assessors rather than real chronological ones [18]. Towards using human body, Deng et al. extracted color and texture features from pedestrian image [69].

3.2 Historical Development of Automatic Gender Estimation

Some major milestones in history of gender prediction are summarized in Fig. 5. In 1988, Childers et al. pioneered speech-based gender prediction using acoustic parameters of vowels and fricatives [46]. The face work started after two years with EMPATH and SEXNET [70], [71], where neural networks were applied directly to intensity features. Brunelli et al. instead extracted geometrical features from frontal faces [106]. In 1996, Parris et al. designed male and female speaker independent Hidden Markov Models (HMM) to model the spectral envelope [72]. Towards using face, O'Toole et al. investigated 3D head structure [74]. In 2000, Support Vector Machine (SVM) was employed to distinguish male from female faces [75]. The first work on gender detection from gait was presented by Lee et al, where segmented silhouette features were extracted across a gait sequence [76].

In 2003, Wu et al. extracted Haar-like features [77] to infer gender from face. In [79], Yoo et al. computed a model-based gait signature using joint angles and body points. The fingerprint work was pioneered by Badawi et al. in 2006 [81]. After this, Thomas et al. presented an iris approach [82]. Instead of using dynamic gait patterns, Cao et al. investigated pedestrian images to infer gender [83]. In [84], Average Gait Image (AGI) was computed from gait for gender recognition. In 2009, Kumar et al. designed a set of attribute classifiers using "in the wild" faces, with gender cue covered [85].

In [87], Wuhrer et al. investigated 3D human body shapes obtained by a range scanner. The Paralinguistic Challenge 2010 included also gender prediction [17]. In 2012, Neurotechnology implemented gender detection in their VeriLook SDK (a biometric SDK for face identification). Nuance,[5] a provider of speech recognition and language technology, announced a new mobile ad format, Voice Ads. This technology can distinguish speakers' gender and thus provide right ads. In 2013, Hassaïne et al. organized the first competition on gender prediction from handwriting documents [89]. Along with the popularity of deep learning, Levi et al. and Zhu et al. employed deep Convolutional Neural Networks (CNN) to derive gender from face and pedestrian images in [90], [91], respectively.

3. www.cognitec.com
4. www.neurotechnology.com
5. www.nuance.com/index.htm

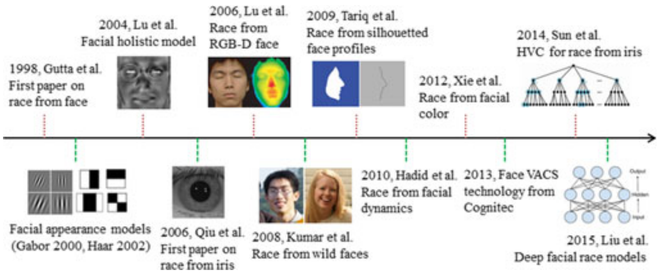

Fig. 6. Major milestones in the history of automatic race recognition from biometric data.

3.3 Historical Development of Automatic Race Estimation

Fig. 6 illustrates the evolution of race estimation. In 1998, Gutta et al. presented the pioneering work on recognizing race using face [92]. Similar to age and gender prediction, for facial race categorization, appearance models are popular representations, e.g., Gabor and Haar wavelets used in [93], [94]. In [95], Lu et al. applied Linear Discriminant Analysis (LDA) to face images of different scales for race detection. The first work on using iris was from Qiu et al. [96]. In 2006, range modality was investigated for race from face [80]. Kumar et al. instead downloaded face images from internet, labeled them with attributes such as gender, race, age, and hair color [97]. Race from these "in the wild" faces was then investigated.

In [98], Tariq et al. categorized racial groups by shape context matching of silhouetted face profiles. Hadid et al. instead extracted spatiotemporal features from face videos [99]. As presented in Section 2.2.3, skin tones are useful features for rough race classification. In [100], Xie et al. presented one such approach using facial color. In 2013, Cognitec released FaceVACS-VideoScan, the first technology to analyze the count, flow and demographics of people visible in video streams. Towards using iris, Sun et al. developed Hierarchical Visual Codebook (HVC) to encode texture primitives of iris images for race grouping [101]. In 2015, Liu et al. proposed a deep learning framework for predicting face attributes "in the wild" [102], where race was among the attributes.

4 Demographic Estimation from Biometric Data

In Fig. 7, an unified framework of typical biometric demographic systems is given. There are four main components: data acquisition, preprocessing, feature extraction, and demographic estimation. It should be noted that there are also systems which implement the last two components together.

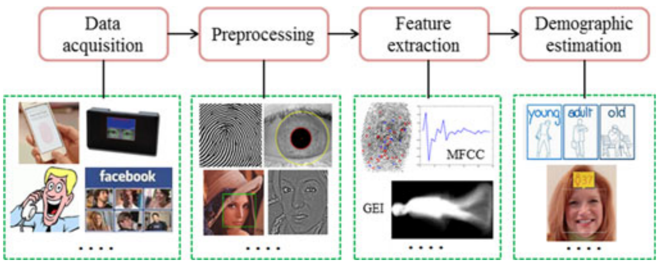

Fig. 7. Four main stages involved in typical biometric demographic estimation systems (image source: Google Images).

Fig. 8. Richness of biometric data in our society (image source: Google Images).

These systems like deep CNN architectures employed in [67], [90], [103], [216], [218], [219] are generally called end-to-end systems. In this section, we briefly present data acquisition and preprocessing components, while detail the existing work on feature extraction and estimation algorithms.

4.1 Biometric Data Acquisition

Biometric data can be captured directly from diverse sensors (2D camera, 3D scanner, video camera, thermal sensor, near infrared camera, microphone, etc.). To acquire high quality data, quite a few capturing devices have been developed. Early sensors were generally cumbersome and expensive. Moreover, they required a high degree of cooperativity from users. During the evolution, acquisition systems become more compact, affordable and friendly, making it possible to embed them in laptops, mobile phones and wearable devices (e.g., Google glass[6]). In [3], Jain et al. listed major turning points in the development of fingerprint, iris, and face capturing devices. Apart from directly acquiring data from biometric sensors, in recent years, researchers have started to collect data from the internet and social media. Social networks such as Facebook[7] and Twitter[8] have generated an unprecedented volume of photos and videos. For example, face images in the LFW database [24] and videos from Youtube[9] Faces database [104] were all collected from the internet. TV programs, movies and video games are also sources of biometric data. In Fig. 8, we illustrate the richness of biometric data in our society.

4.2 Preprocessing

Subsequent to data acquisition, preprocessing follows. Due to variability in the acquisition environment (e.g., under different illumination conditions) and the interaction with acquisition devices (e.g., different head poses), the captured data usually appears very different. Meanwhile, the data might contain noisy information. Hence, preprocessing aims to detect/segment the valid data, normalize it, and further enhance the quality. Face detection and alignment are important procedures in face systems. For iris, localization, segmentation and normalization are generally performed. Normalization is used to unfold annular iris images to rectangle images of the same size. In voice systems, de-noising is usually performed for enhancement. If the speech is from multiple speakers, separation will then be essential to obtain

6. www.google.com/glass/start
7. www.facebook.com
8. twitter.com
9. www.youtube.com

each speaker's signal. Good quality fingerprint images are not always easy to be acquired, which might be corrupted or degraded due to skin variation. Hence, enhancement is necessary, generally involving three procedures: normalization (e.g., histogram equalization), binarization and thinning. Common gait preprocessing includes background substraction and silhouette normalization.

4.3 Feature Extraction

In this section, we provide detailed literature review on demographic informative representations from different biometric traits.

4.3.1 Representations from Face

Common facial representations for age, gender, and race recognition include geometric/anthropometric models, holistic models based on subspace/manifold learning, appearance models, 3D models/range modality, and deep models. Early demographic work usually adopted geometric/anthropometric models [48], [105], [106], [107], which are based on dimensions of the whole face/salient features (e.g., eyes, nose, mouth, etc.) and distance ratios measured from facial landmarks. Notice that only frontal faces can be used to compute such representations, which are sensitive to head pose variations. Another disadvantage is these models consider only facial geometry while ignore texture information. As a result, they are not appropriate for age estimation of adults.

Holistic models [58], [59], [108], [109], [110], [111] represent each face image as a single high-dimensional vector by concatenating gray values of all its pixels. Subspace/manifold learning algorithms are then employed to learn a low-dimensional representation, which can well capture the global information of the whole face. Holistic models are generally simple, efficient, and easy to be implemented. In contrast, appearance models are more complicated, and require more computational resource. Nevertheless, by considering both shape and texture, they are more robust to facial appearance variations. AAM [112], [113], [114], [115], Haar-like features [52], [77], LBP [53], [54], [55], [56], Gabor [44], [105], [116], BIF [12], [57], Discrete Cosine Transform (DCT) [117], [118], [195], and local directional pattern [119] all belong to appearance models.

Note that all the above representations are based on 2D RGB or intensity images. There are also approaches exploiting 3D faces. Examples include gender detection in [120], [121], [122], [123], [124] and ethnicity categorization in [125], [126], [127], [128]. By exploring the surface shape, 3D approaches show insensitivity to viewpoint and illumination variations. Along with the success of deep learning techniques in diverse areas and the availability of large number of training face images, researchers have turned their attention to learning facial representations by deep neural networks in recent years. These deep learned features [220], [223], [224] are more representative and discriminative for demographic tasks than traditional handcrafted and data-driven features. Training such models yet usually requires high computational cost.

Apart from the five common representations, in [129], [130], [131], [132], facial dynamics from video data were investigated for demographic estimation. In [88], [98], silhouetted face profiles were proved to be useful to gender and race classification. Facial color-based features are also race informative, which have been investigated in [100], [133]. Chen et al. instead explored face images obtained in near-infrared and thermal spectra for gender prediction [134].

4.3.2 Representations from Voice

Investigating paralinguistic information conveyed in speech signals offers a chance to determine a speaker's age and gender. In order to automatically recognize them, most researchers examine how the speaker said by extracting acoustic/prosodic features. Among these features, Linear Prediction Cepstral Coefficients (LPCC), Mel Frequency Cepstral Coefficients (MFCC) and Perceptual Linear Prediction (PLP) coefficients are the most popular ones [46], [50], [60], [72], [73], [78], [135], [136]. As stated in Section 2.2.2, females generally have higher pitch than males. In [72], [73], [135], authors verified the effectiveness of pitch on gender prediction. Shafran et al. instead employed pitch for age grouping [78]. Perturbation features like jitter and shimmer are also popular for demographic analysis [137], [138], [140]. Jitter is defined as micro-variations of F0, while shimmer represents micro-variations of amplitude. Other common acoustic/prosodic features include harmonicity (e.g., harmonics-to-noise ratio and noise-to-harmonics) [140], energy [135], [140], formants [46], [47], modulation cepstrum [141], and speech rate [50]. For demographic analysis using voice, generally, using single feature is hard to achieve satisfactory performance. Thus, most existing approaches employed multiple features.

4.3.3 Representations from Gait

Human gait representations for age and gender prediction can be roughly divided into three categories: point-light, model-based and appearance-based approaches. Early work generally adopted point-light representations from the aspect of biological motion, where a set of reflective markers are attached on the body. Locomotion features are then extracted from point-light trajectories. Examples include the work of Davis [49] and Begg et al. [142] for binary age group classification and the work of Davis et al. [143] and Troje [144] for gender prediction. One disadvantage of point-light is its obtrusiveness, which obliges users' cooperation. Model-based methods instead consider both shape and dynamics of human body. They represent body parts as different shapes, e.g., the stick model [79] and the ellipse model [76], [145], [146], [147], and then extract parameters of these shapes as gait features. Without fitting a model, appearance-based methods compute the representation directly from gait silhouette, i.e., using the whole motion pattern of human body. Widely used GEI [86], [148], [149], AGI [84], [150], and common signal transformation methods (e.g., Radon [151], [152], Fourier [61], [153] and wavelet transformations [154], [155]) all belong to this category. Appearance approaches are computationally efficient and work well with low-resolution gait sequences. In contrast, model-based methods usually require high-resolution gait data to accurately fit the physical model and thus demand for relatively high

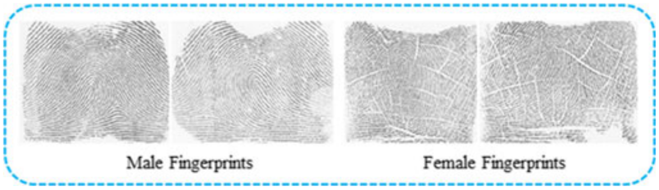

Fig. 9. Fingerprints of different genders (originally shown in [81]).

computational cost. However, they can well handle occlusion, noise, scale and rotation issues.

For gender from static human body, texture features are popular pedestrian representations, covering Histograms of Oriented Gradient (HOG) [83], [156], [157], [161], BIF [162], Gabor and Schmid [69], [163], LBP [161], etc. These features can well capture the local shape information with some invariance to translation and rotation variations. To infer the gender, humans use not merely body shape and hairstyle, but also additional cues such as clothes and accessories. Among these cues, clothing color plays an important role. In order to capture the color information, different color spaces have been exploited [69], [156], [161], [163], e.g., RGB, HSV, and YCbCr. Note that color alone is not discriminative enough for gender prediction from pedestrian. It is generally combined with texture features. Instead of exploiting 2D human body images, Wuhrer et al. and Tang et al. investigated 3D human body shapes [87], [164]. Compared with 2D models, 3D models show more robustness to variations of light and posture. Along with the success of deep learning, Zhu et al. proposed a multi-label deep CNN to predict gender and other attributes together in an unified framework from unconstrained pedestrian images [91].

For age from static human body, there is hardly any existing work. An interesting one is from Deng et al. [69], who performed age group classification from unconstrained pedestrian images using texture and color features.

4.3.4 Representations from Fingerprint

Fingerprint is a gender informative trait. Early work on this task generally extracted physical features [81], [165], including ridge count, pattern type concordance, ridge count asymmetry, Ridge Thickness to Valley Thickness Ratio (RTVTR), white line count, ridge density, and ridge width. As found in [81], female fingerprint is generally characterized by high RTVTR and high count of white lines (see Fig. 9), while male fingerprints generally have lower RTVTR and no or few white lines. Apart from physical features, texture features are also studied. In [170], discrete wavelet transform was employed to extract frequency features. Li et al. and Rattani et al. instead evaluated LBP and other texture features [172], [173].

4.3.5 Representations from Iris

For race categorization from iris, researchers generally compute texture features, e.g., Gabor energy [10], [96], statistics of spot and line filters [174], [175], [176] and Scale Invariant Feature Transform (SIFT) [101], [177]. Iris texture is also popular for gender discrimination. For example, Tapia et al. investigated the use of LBP in [178] and iris code in [235], and Bansal et al. combined wavelet features and statistics along angular and radial directions [179], [180]. In addition, geometric features which describe iris dimension have been also investigated in the literature [82].

4.3.6 Representations from Handwriting

Handwriting is a modality highly susceptible to biological factors, age, and social habits. This intuition has motivated its demographic study. By analyzing a set of micro features, covering gradient, structural, and concavity, Tomai et al. measured the performance of using individual characters for demographic tasks [181]. Bandi et al. instead investigated the whole handwritten document for gender and age group classification [182], where macro features such as slant, word gap and gray-scale threshold were examined. Micro features are good at describing fine details at character level, while macro features can capture more global characteristics of writers' individual writing habits and styles. In [183], on-line handwriting was explored for gender detection with the advantage of recording also the temporal information. Their on-line features include speed, writing direction, curvature, acceleration, etc. In the ICDAR gender competition [89], different features have been evaluated, covering curvatures, direction, tortuosities, edge-based directional features and chain codes.

4.3.7 Representations from Other Traits

Periocular refers to the region surrounding the eye which might or might not include the eyebrow. In the literature, there has been work on exploiting this region for demographic analysis using local appearance features. Examples include the work of Merkow et al. on gender classification using LBP [184] and the work of Lyle et al. on gender and ethnicity prediction using LBP, HOG and DCT [186]. Recognizing people by their ear has recently received significant efforts in diverse applications, especially when non-frontal faces are available. In [88], [187], appearance features of 2D ear were extracted for gender recognition. Lei et al. instead investigated 3D ear [188]. For gender from hand geometry, Amayeh et al. extracted region and boundary features from segmented hand silhouettes [189].

4.4 Demographic Estimation

With extracted representations, the next step is to estimate demographics. For age estimation, we can either determine a coarse age group (e.g., children, adult, and the elderly) to which the subject belongs or calculate his/her scalar age value. Coarse age group categorization is a classification problem. Various classifiers have been employed including Artificial Neural Network (ANN) [112], [137], [182], SVM [57], [60], [142], Adaboost [12], [53], [182], K-Nearest Neighbor (KNN) [61], [155], [181], GMM [50], [135], [136], etc. Gender and race prediction are also classification problems. Commonly used gender classifiers include SVM [69], [76], [86], [156], [157], [162], [163], [164], [165], [178], [179], [184], Adaboost [77], [158], [159], [160], [161], Decision Tree (DT) [82], Gradient Boosting Machine (GBM) [89], Bayesian classifier [166], LDA [167], [168], [169], KNN [164], [170], HMM [78], ANN [81], [164], [185], GMM [138], [141], [183], Random Forests (RF) [83], gaussian process classifier [171], etc. For race, SVM [10], [44], [80], [101], decision trees [214], Adaboost [96], and ANN [176], [186] are popular ones.

TABLE 1
Advanced Biometric Representations and Estimators for Demographic Analysis

		Face	Voice	Gait	Fingerprint	Iris	Handwriting
Age	F	Deep, Appearance F	Acoustic, Prosodic F	Gait: Appearance F Pedestrian: Texture, Color F	-	-	-
	E	Deep NN Hierarchical Framework Ranking Algorithm Label Distribution Learning	SVM, GMM	SVM	-	-	-
Gender	F	Deep, Appearance F	Acoustic, Prosodic F	Gait: Appearance F Pedestrian: Deep, Texture, Color F	Physical F Texture F	Texture F	Directional F Chain Codes Curvatures
	E	Deep NN, SVM, Adaboost	SVM, GMM	Gait: SVM Pedestrian: Deep NN, SVM	SVM ANN	SVM	GBM
Race	F	Deep, Appearance F	-	-	-	Texture F	-
	E	Deep NN, SVM, Adaboost	-	-	-	SVM	-

"F", "E", and "NN" respectively denote "Feature", "Estimator", and "Neural Network".
"-" represents there is no or little work on related tasks.

Among various classifiers, for demographic categorization, SVM is probably the most widely used one [190]. As a very effective method for general purpose supervised pattern classification, SVM finds an optimal separating hyperplane that provides superior generalization ability especially when working with high-dimensional data (e.g., biometric samples) and limited training data. Adaboost is another well-known classifier. It aggressively selects a small set of weak learners to form a stronger classifier [191], thus significantly boosting the performance.

For scalar age value estimation, most existing approaches consider a regression solution where a mapping function is learned explicitly between feature vectors and scalar age values. Support Vector Regression (SVR) is a popular regression algorithm [57], [116], [192]. Other regressors include linear/quadratic/cubic regression [51], ANN [193], RF [194], [220], kernel regression [195], [202], KNN regressor [203], locally adjusted robust regression [58], kernel partial least squares regression [204], Gaussian Process Regression (GPR) [153], etc. Instead of directly using regression, many systems employ a hierarchical framework [12], [112], [116], [220], [221], [222], which first roughly predicts an age group for the given sample then estimates its age in the specific group. Accordingly, both classification and regression algorithms are utilized in such systems. As human aging is perceived differently in different age groups, this hierarchical manner provides more accurate results than directly regressing the age.

Unlike regression, for scalar age value estimation, ranking algorithms exploit the relative order information among age labels for rank prediction [65], [205]. This ranking scheme works more effectively when training data is insufficient and imbalanced, since all training samples are exploited for building each age ranker. Another favorable property is its cost sensitivity, by investigating which, we can well capture the correlation among age labels so that samples with neighbouring age values share more than those further away. For example, if a person is 15 years old, the age label is more likely to be 14 or 16 years old than 10 or 20 years old. Thus, classifying his/her samples into different ages has different costs. For human facial age estimation, cost sensitivity was also studied for feature selection [206], [207] and cumulative attribute space learning [208]. Notice that the ranking scheme was explored not merely for direct age estimation as in [65], [205] but also for facial age subspace learning [209] and age difference investigation [210].

Besides ranking algorithms, label distribution learning is another effective technique for predicting scalar age value from limited training data [212], [213], [218]. It associates each sample with an age label distribution instead of a single age. The label distribution covers a certain number of age labels, representing the degree that each label describes the sample. Consequently, each sample contributes to not only the learning of its own age, but also the learning of its neighbouring ages. The uncertain nonnegative labels proposed in [139] and fuzzy age labels used in [211] share similar characteristics with label distribution.

4.5 Advanced Representations and Estimators

In order to find what algorithm to implement to best determine one specific attribute using one specific modality, based on the above analysis, we suggest here some biometric representations and estimators which either are the most widely used or achieve state-of-the-art performance. We believe this would be useful for the interested reader who intends to design their own demographic systems. In Table 1, we list advanced models for face, voice, gait, fingerprint, iris, and handwriting. For other modalities, since there are very few papers on related tasks, we do not give any recommendation. Age and race from handwriting are not considered likewise. Additionally, for age from face, we work on scalar age value estimation rather than age group classification, because most work on facial age estimation centers on the former. When deep neural network is used as an end-to-end architecture, it can be seen as an estimator. Otherwise, we can use them to learn representations [220], [223], [224], i.e., deep features.

5 State of the Art in Demographic Estimation

Demographic estimation is a rapidly evolving technology and demographic systems are experiencing continuous improvement in both performance and usability. In this section, we provide a clear assessment of the state of the art, with attention on public databases and estimation results.

Fig. 10. Sample images from the MORPH II face database [6].

TABLE 2
Machine and Human Performance on FG-NET and MORPH II Face Databases from [12]

	FG-NET		MORPH II	
	Machine	Human	Machine	Human
Age (MAE)	4.8 ± 6.2	4.7 ± 5.0	3.8 ± 3.3	6.3 ± 4.9
Gender (Accuracy)	-	-	97.6%	96.9%
Race (Accuracy)	-	-	99.1%	97.8%

race task is black versus white.

5.1 Evaluation of Demographic Estimation from Face

The state of the art of demographic estimation from face varies depending on the quality/variation of the face data. FG-NET [215] and MORPH [6] databases are two most well-used benchmarks for age estimation. FG-NET consists of about 1,002 high-resolution face images of 82 subjects from an age range of $0 \sim 69$. Face images in MORPH are organized into two albums. Album 2 (i.e., MORPH II) is available for academic purposes and contains over 55,000 images of 13,000 subjects. Besides age, subjects' gender and ethnicity are also provided in this album. Some example images from MORPH II are shown in Fig. 10. In [12], the machine (BIF + Adaboost + SVM/Hierarchical Age Estimator) performance was compared against human performance. We list the results in Table 2. For age estimation, the mean absolute error (MAE) in years is used as the measure, which is defined as the average of absolute errors between estimated age labels and ground truth labels. Using CNN fine-tuned on a large outside facial age dataset, Rothe et al. further achieved MAEs of **3.09** years on FG-NET and **2.68** years on MORPH II [216]. These encouraging results might be attributed to also good quality of face images from both databases, which are collected in controlled conditions.

The evaluation organized by NIST in 2014 provided an objective assessment of current automated facial age estimation technology [66]. It employed over 7 million face images from visas and law enforcement mugshots. On a subset of 6,172,395 visa images from an ethnically-homogeneous population spanning ages $0 \sim 99$, the best participant (a commercial system from Cognitec) obtains a MAE of **4.3** years. On an ethnically-heterogeneous population, most participants instead report higher MAEs, which suggests that ethnicity has an impact on age estimation. Specifically, South Americans tend to be overestimated in age, and Asians tend to be underestimated. For gender impact, the evaluation result indicates that age is more accurately estimated in males than females, with the tendency for adult females to be underestimated.

For age and gender from "in the wild" faces [196], [197], [198], [199], [200], [201], Eidinger et al. offered the Adience benchmark with face images acquired by mobile devices and uploaded to online repositories (Flickr albums[10]) [7]. These real-world face photos present large variation in appearance, lighting, head pose, resolution, quality, and more. In Fig. 11, we illustrate some example images after in-plane alignment. The database consists of 26,580 face images of 2,284 subjects. Age estimation here is a eight-group classification task with $0 \sim 2$, $4 \sim 6$, $8 \sim 12$, $15 \sim 20$, $25 \sim 32$, $38 \sim 43$, $48 \sim 53$, and $60+$. The evaluation protocol follows a 5-fold cross-validation (CV) scheme. Using deep CNNs, Rother et al. obtained an age accuracy of $\mathbf{64.0 \pm 4.2}$ percent [216]. The exact accuracy is though very low, the one-off accuracy gets around $\mathbf{96.6 \pm 0.9}$ percent, which means the predicted label is within the neighboring ± 1 groups. Similarly using deep CNNs, Levi et al. reported a gender accuracy of $\mathbf{86.8 \pm 1.4}$ percent on this database [90].

Fig. 11. Face images after in-plane alignment from the Adience database [7].

Fig. 12. Some example images from the ChaLearn LAP Challenge 2015 apparent age dataset [18].

TABLE 3
ChaLearn LAP 2015 Final Ranking with the Best Three Results

Rank	Algorithm	Error
1	CVL_ETHZ [216]	0.265
2	ICT-VIPL [218]	0.271
3	WVU_CVL [220]	0.295
4	Human Performance [18]	0.34

human performance is provide by organizers.

For apparent age from "in the wild" faces, in the ChaLearn LAP Challenge 2015, a dataset of 4,699 images was provided, exhibiting large variation of pose, illumination, expression, and quality (see Fig. 12). Each image was labeled by at least 10 users a real number from 0 to 100 years old. Their mean age is considered as the final apparent age. The test set contains 1,087 images. We list in Table 3 the best three results [217]. All the three participants trained deep age models using diverse outside face databases. The winning system reported an error rate of **0.265** [216]. These results clearly demonstrate the superiority of deep age models over human performance. In 2016, ChaLearn organized the second LAP Challenge [225]. The face dataset was extended to 7,591 images, from which 1,978 images were selected for test. The best performance with an error rate of **0.241** was achieved [226].

10. www.flickr.com

TABLE 4
Classification Accuracies (%) of Several Participant Algorithms in Paralinguistic Challenge 2010

	Participant	Age	Gender
Baseline	Schuller et al. [17]	48.9	81.2
Age Winner	Kockmann et al. [230]	52.4	83.1
Gender Winner	Meinedo et al. [135]	48.7	84.3

5.2 Evaluation of Demographic Estimation from Voice

For age and gender from voice, Schuller et al. organized the Paralinguistics Challenge 2010 [17], [62]. The "aGender" corpus with 47 hours of telephone speech in 65,364 single utterances of 945 speakers served to the evaluation [9], where 175 speakers were chosen for test. For age classification, four age groups are considered (child, young, adult and senior), while speakers' gender had to be determined from three groups (child, male and female). Table 4 shows the results of several participants. The baseline system was designed by Schuller et al. using acoustic features with SVM [17].

In order to boost the performance, most participants combined several sub-systems in which different features or classifiers were employed [135], [227], [228], [229], [230]. The age winner was from Kockmann et al. [230], who made use of utterance-based acoustic, prosodic and voice quality features provided by organizers together with their own frame-based acoustic features. The classification was based on GMM and SVM. For gender classification, the winning system was composed by six individual sub-systems trained with short and long term acoustic & prosodic features [135]. The authors further employed three outside corpora to incorporate more speaker variability and more diverse audio background conditions into their gender model, which was based on GMM, MLP, and SVM. The low accuracy of age task indicates age from telephone speech is a very difficult problem.

5.3 Evaluation of Demographic Estimation from Gait

For gender from gait, the most widely used benchmark is the CASIA B database with gait data captured from 124 subjects (93 males and 31 females) in indoor environments [8]. It contains large view variation from frontal to rear view. Some silhouette images of a male subject and a female subject are shown in Fig. 13. We list in Table 5 both human and machine performance on this database. All the algorithms used a subset of 31 females and 31 randomly selected males with only profile silhouette, following 31-fold CV. As observed, most machine algorithms achieve accuracies higher than human observers. It should be noted that this may be ascribed to the good quality of gait data, which is captured from profile view with normal clothes and without any bag.

Fig. 13. Multi-view silhouette images of one male subject (top row) and one female subject (bottom row) from the CASIA B gait database [8].

TABLE 5
Gender Classification Accuracies (%) on the CASIA B Gait Database

	Algorithm	Accuracy
Yu et al. [86]	Human Observers	95.47
Li et al. [84][a]	AGI + SVM	93.28
Yu et al. [86]	Segmented GEI + SVM	95.97
Hu et al. [154]	Gabor + HMM	96.77
Hu et al. [146]	Ellipse-fit Parameters + MCRF	98.39

[a] *is implemented in [86]; "MCRF" denotes "Mixed Conditional Random Field".*

For scalar age value estimation, the USF database is a public benchmark with biological age value available for each subject [231]. There are a total of 1,870 gait sequences from 122 subjects with age range $19 \sim 59$. Each subject presents variations in viewpoint, shoe type, walking surface, carrying condition, and elapsed time. To facilitate its use, the authors fixed one gallery subset as control, and further created 12 probe subsets to examine the effect of different variations. Table 6 presents MAEs of two algorithms on the Gallery and Probe A subsets with 79 subjects. Both algorithms follow the leave-one-person-out (LOPO) scheme. In [61], Makihara et al. examined multi-view gait silhouette for age group classification (children, adult male, adult female, and the elderly). On a self-collected database with 168 people ($4 \sim 75$ age range), using frequency domain features, they obtained an accuracy of **94** percent.

For gender from static human body, existing work used either pedestrian images or 3D human body shapes. The MIT pedestrian database is a popular benchmark for pedestrian detection [233]. In [83], Cao et al. labeled 600 images from this database as male and 288 images as female for gender classification. In Fig. 14, we illustrate several male and female samples from this database. Table 7 shows results of two algorithms, both following five-fold CV. To advance the state of the art in pedestrian attribute analysis, Deng et al. introduced the PEdesTrian Attribute (PETA) dataset [69], where both gender and age group labels are provided. Using SVM with texture and color features, on a test set of 7,600 images, they got a gender accuracy of about **86.5** percent. For age recognition, four age groups, i.e., $16 \sim 30$, $31 \sim 45$, $46 \sim 60$, and 60+, are

TABLE 6
MAEs on the USF Gait Database

	Algorithm	Gallery	Probe A
Lu et al. [155]	Gabor of GEI + MLG + ML-KNN	5.42	5.56
Lu et al. [232]	Manifold Analysis of GEI +Multiple Linear Regression	4.77	4.28

"MLG" denotes "Multilabel-Guided" and "ML-KNN" means "Multilabel KNN".

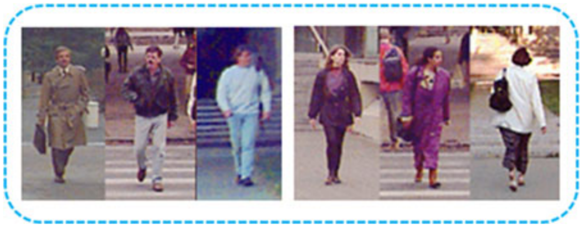

Fig. 14. Pedestrian images of male (left three) and female subjects (right three) from the MIT pedestrian database [233].

TABLE 7
Gender Recognition Results (%) on the MIT Pedestrian Dataset

	Algorithm	Accuracy
Cao et al. [83]	HOG + Ensemble Learning	75.0 ± 2.9
Guo et al. [162]	BIF + PCA + SVM	80.6 ± 1.2

considered. Their reported accuracies are respectively **86.8, 83.1, 80.1, and 93.8** percent.

For gender from 3D human body shapes, the CAESAR database[11] is a good benchmark [234], which includes full body shapes of about 4,500 civilians in North America and Europe captured by a range scanner. For each subject, 3D locations of 73 anthropometric landmarks are provided. Using SVM with all pairwise geodesic distances between landmarks, Wuhrer et al. reported an accuracy of at least **93** percent on a subset of 500 males and 500 females [87].

5.4 Evaluation of Demographic Estimation from Fingerprint

Most fingerprint-based gender approaches used self-collected databases. For example, in [81], fingerprints of 1,100 males and 1,100 females were scanned from their ink prints. On a test set of 52 males and 37 females, using RTVTR, white line count and ridge count features together leads to an accuracy of **87.64** percent with ANN. In [170], a database of 3,570 fingerprints was collected with 1,980 male fingerprints and 1,590 female ones. With KNN, a classification rate of **88.28** percent was reported using wavelet transform on a test of 1/3 of total fingerprints. To evaluate different texture descriptors, Rattani et al. organized a database of 166 males and 71 females [173]. On a test set of 116 males and 21 females, an accuracy of **80.4** percent was achieved by using LBP and SVM.

5.5 Evaluation of Demographic Estimation from Iris

For race from iris, Lagree et al. selected 1,200 iris images from the iris database collected by University of Notre Dame (UND) [174]. These images belong to 60 Asian and 60 Caucasian subjects, each with five left and five right iris images. A 10-fold CV scheme was followed. For gender classification, Tapia et al. presented the gender-from-iris (GFI) dataset, which contains 3,000 iris images of 750 males and 750 females [235]. Each subject has one left iris image and one right iris image. In Fig. 15, we show right and left iris images of a female subject and a male subject from this dataset. From the total 1,500 persons, 20 percent of males and 20 percent of females are randomly selected for test. The race and gender classification results on the two datasets are shown in Table 8.

11. store.sae.org/caesar

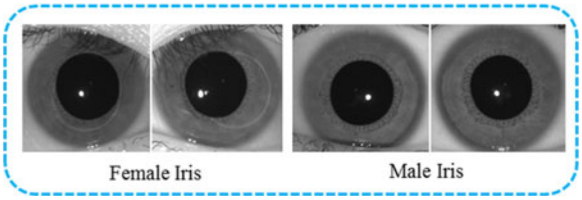

Fig. 15. Right and left iris of a female subject and a male subject from the GFI iris dataset [235].

TABLE 8
Race and Gender Classification Accuracies (%) on the UND and GFI Iris Datasets

	Algorithm	Dataset	Accuracy
Race	Lagree et al. [174] (Statistics of Filter Response + SVM)	UND	90.58
Gender	Tapia et al. [235] (Iris Code + SVM)	GFI	89.00

race classification is asian versus caucasian.

5.6 Evaluation of Demographic Estimation from Handwriting

For age and gender from handwritten documents, Bandi et al. employed the CEDAR letter database, which consists of more than 3,000 handwritten document images from more than 1,000 writers [182]. Each individual provided three samples of the same text. The test set for age classification (under the age of 24 versus above the age of 45) contains 350 documents, while there are 400 documents for gender task. Using Adaboost, they got accuracies of **86.6** and **77.5** percent, respectively.

The dataset used in the ICDAR gender competition contains a total of 475 writers' handwritten text [89], with both Arabic and English documents. Participants were asked to predict the gender of 193 writers. Among various classifiers used by participants, GBM is the most popular and achieves also the lowest logarithmic loss. The Area Under the receiver operator Curve (AUC) of the winning system is **87.1** percent, achieved by using GBM and a set of provided features. The organizers further investigated the importance of different features to gender prediction, and found edge-based directional features, chain codes, and curvatures obtained the best performance.

5.7 Evaluation of Demographic Estimation from Other Traits

For gender and ethnicity from periocular, in [236], Lyle et al. collected a set of periocular images from high-resolution frontal face images of the FRGC database [237]. There are a total of 2,116 samples from 404 subjects. With a five-fold CV scheme, using LBP and SVM obtained accuracies of **96.67** percent for gender (57 percent males versus 43 percent females) and **93.8** percent for ethnicity classification (22 percent Asian versus 78 percent non-Asian). Merkow et al.

instead examined low-resolution periocular images collected from web for gender task [184]. The database contains 468 male images and 468 female images. Following five-fold CV, they got an accuracy of **84.9** percent by using intensity features and SVM. For gender from ear, Zhang et al. selected 942 face profile images of 302 subjects from the UND biometrics data sets Collection F [88], [238], with 562 male and 380 female images. Following 10-fold CV, they achieved an accuracy of **91.78**$\pm$**4.66** percent by using texture features and SVM. For gender from hand geometry, Amayeh et al. [189] built a small dataset with 20 males and 20 females. With a leave-one-out CV scheme, an accuracy of **98** percent was reported by using Fourier descriptor and LDA classifier.

6 Open Problems

As reviewed above, a great amount of work has been carried out concerning various aspects of demographic analysis. Comprehensive efforts have been devoted to machine estimation from both academia and industry. Moreover, different public evaluations have shown that some techniques are able to achieve very promising performance. Nevertheless, many systems, even commercial products, fail to accurately estimate demographics in a number of applications. Here, we discuss some still open issues involved in this topic.

6.1 Unconstrained Biometric Demographic Estimation

In many biometric applications, it is not easy to impose constraints on how to acquire data, resulting in a large amount of unconstrained data. Estimating demographics from such data is far from being solved. One classical example of unconstrained sensing environment is video surveillance, where biometric samples acquired are generally of poor quality and low resolution, with strong illumination variation, large pose/view changes, and occlusions. Another well-known source is internet and social media, e.g., photos in Facebook, videos in Youtube, diverse TV programs and movies. These "in the wild" samples present a great variety of variations, making it extremely hard to detect demographics. Latent fingerprints in forensic applications and faces after makeup [240] are also common challenging samples.

To well handle this problem, it is in great demand to collect unconstrained data. Several such face databases have been made publicly available in order to facilitate the study of analyzing demographics from unconstrained face, e.g., Gallagher group photos [239], Adience [7], and faces in ChaLearn 2015 [18]. Pedestrian images are also unconstrained samples for gender and age prediction, e.g., MIT [233] and PETA datasets [69]. For other modalities, there is rather little work yet on investigating unconstrained data.

6.2 Correlation Among Attributes from Single Trait

Some modalities contain information of more than one attribute. For instance, face is informative for all the three attributes, while voice, gait, and handwriting convey age and gender cues. It is unclear about how these attributes affect each other, there is only limited work on studying their interrelationships, though. Here we give several examples. In [241], Guo et al. observed degraded facial age estimation performance when conducted across gender and ethnicity. The NIST age evaluation studied also the ethnicity and gender impact [66]. Guo et al. [242] and Farinella et al. [244] instead explored how ethnicity and gender interact with each other in human face. In order to build a robust facial gender classifier, Bekios-Calfa et al. studied dependence between gender and age [245]. Towards gait-based age estimation, Lu et al. included gender cue in age label encoding and achieved better results [155].

Another problem is how to estimate multiple attributes in an unified framework. For example, in [53], Yang et al. predicted the three attributes independently from face, following an assumption that there is no relation among them. Guo et al. [246] and Yi et al. [67] instead presented frameworks that can deal with the mutual influence implicitly and thus estimate the three cues jointly. In spite of these efforts, whether and how to incorporate the correlation among different attributes require further investigation, especially for other traits.

6.3 Correlation Among Age Labels

Human age is strongly correlated and biometric samples with neighboring age values share more than those further apart. For instance, a human face of 60 years old looks more similar to that of 55 than that of 15. To exploit this observation, most existing approaches seek a regression solution to scalar age value estimation. These approaches, though, fail to well capture complicated correlation among age labels when there is no adequate training data available. We discussed in Section 4.4 several techniques specially developed for this issue, including ranking algorithms, label distribution learning, and cost sensitive strategies. Such approaches are yet limited and investigated only for face modality. More efforts are therefore needed to well study this issue.

6.4 Limited Labeled Data for Age Model Developing

For learning an age model robust to the uncontrollable and personalized age progression, one big challenge is the acquisition of labeled data with either biological or apparent age values available. Existing databases either include very few ages for each individual (shallow) or provide data of very few subjects (narrow). For example, the deep but not broad FG-NET and the shallow but broad MORPH II databases. A deep and broad database allows for both the study of individual age progression and the extraction of common aging trend among different subjects. Collecting such databases is thus crucial but very difficult in practice. Privacy issue is one concern. The long time it may take is another problem. Although for some traits like face, nowadays we can easily collect a huge amount of unlabeled data from internet, manually labelling is tedious and error prone. This makes the study of biometric age estimation extremely hard.

6.5 Vague Concept of Age & Race Groups

For age group/race classification, one important issue is the definition of age/race groups. As presented in Section 5, different age groups have been utilized for evaluation of different algorithms, e.g., four-group classification in Paralinguistic Challenge 2010, seven-group in Gallagher group photos, and eight-group in Adience benchmark. However, it is not investigated whether using different age group concepts affects the performance. It is thus indefinite which concept offers a sounder evaluation platform. For this issue, Dibeklioğlu et al. proposed an adaptive age grouping approach which defines age groups automatically [131]. Likewise, race groups are loosely defined due to intermingling of distinct races, e.g., two-category classification in [12], [174], three-category in [44], five-category in [242], and even deciphering Chinese, Japanese and Koreans in [243]. Again, it is unexplored in biometric community which concept we should follow for race tasks. Another challenging issue regarding age group/race classification lies in the ambiguity. For example, in Adience database, should we classify a person of 35 years old into Group $25 \sim 32$ or Group $38 \sim 43$? And for race, there are quite a few subjects in the world whose parents come from different races, making it difficult to precisely determine their race groups. An interesting attempt was from Zhong et al. [247], which viewed ethnicity categorization as a fuzzy problem and then assigned each face a reasonable membership degree.

7 Opportunities

Bearing in mind all lessons learned from the existing work, in this section, we list a couple of promising future directions, by pointing out several key factors that will play a dominant role in shaping the future.

7.1 Biologically Inspired Model Investigation

Given the remarkable ability of humans for demographic analysis, it is favorable to look to biologically inspired models for improving machine performance. Among existing techniques, Gabor and BIF are two such models. Gabor wavelets, whose kernels are similar to 2D receptive field profiles of mammalian cortical simple cells, have been widely used for computing face, iris and gait representations. BIF is similarly designed following models of primate visual system. Another two biologically inspired models are artificial neural networks and human visual attention [248], [249]. Using deep neural networks, encouraging results have been reported for face and pedestrian-based demographic systems. With the availability of large amount of biometric data and powerful computational hardware, it will be of significant interest to investigate these models using different modalities. This is particularly significant when working with unconstrained data.

7.2 Unlabeled Data & Context Investigation

As stated above, one major difficulty of current age estimation comes from limited labeled data. This is also the case of gender and race grouping. A straightforward solution is to exploit unlabeled data. There have been some such attempts on using human face, owing to the availability of large amount of data from web sources. Some researchers manually/automatically annotated unlabeled samples and used them as new training data for semi-supervised learning, e.g., the work of Cherniavsky et al. for gender & age classification [250] and the work of Liu et al. for scalar age value estimation [211]. Some tried to approximate age ranks/difference of unlabeled data, which offered some weak supervision to age estimation [209], [210]. Ni et al. instead crawled face images from web by a set of age related text queries which then serve as age labels of crawled data [202]. Many recent deep models pretrained for object/face recognition belong to also such attempts [102], [213], [218], [219], [220]. Besides unlabeled data, context is another informative resource. For instance, Gallagher et al. estimated age and gender by exploiting first name priors [251] and social context [239]. In [103], object correspondence in successive video frames was examined as weak supervision for demographic analysis. Song et al. investigated also video context to enforce age consistency of multi-view face images [252].

7.3 Apparent Demographic Analysis Investigation

In 2009, Gao et al. estimated subjective ages from consumer face images [68]. In 2015, the ChaLearn LAP Challenge boosted the research on face-based apparent age estimation. Besides age, Davis et al. estimated perceptual gender from point-light trajectories of human gait [143]. These attempts all belong to apparent demographic work. Despite being strongly correlated with each other, an apparent attribute of a person may be very different from the corresponding biological one. For instance, in experiments of Davis et al. [143], seven among 50 persons are falsely perceived as their opposite genders. This is particularly the case for age detection, e.g., some people might look/sound younger than their actual ages. As a result, existing methods developed for biological demographics may be not optimal for apparent tasks. When developing new methodology, two issues should be taken in account. First, there is often no adequate training data with perceptual/apparent demographic labels available. Second, annotated/subjective labels might not be consistent among different assessors.

7.4 Feature Modeling & Metric Learning Investigation

In Section 4.3, we described various features extracted from different traits. Many existing approaches employ directly these “primitive” features. They are demographic informative, though, designing useful algorithms to model or organize them can result in more advanced descriptors which generally possess some attractive characteristics, e.g., discriminative, pose/view invariant, cost sensitive, robust to limited training data, etc. We give several examples here. In [118], Zhuang et al. applied HMM supervectors to DCT features, achieving a representation better capturing spatial structure of human face. Yan et al. instead modeled DCT by GMM and obtained a facial representation robust to misalignment [117]. In [206], Li et al. learned an ordinal discriminative representation from Gabor facial features to preserve both local manifold structure of data and ordinal information among age labels. The aging pattern subspace built on AAM facial features can work well with incomplete training data [113]. Therefore, it will be interesting to investigate such feature modeling techniques using diverse modalities. This is also the case of distance metric learning [203], [207], [253], by investigating which we can obtain a metric more suitable to the task at hand rather than directly using euclidean distance or cosine similarity.

7.5 Fusion of Multiple Modalities

Most existing work on demographic analysis employs a single modality. One issue involved in such unimodal systems is how to contend with noises in sensed data, e.g., a voice sample altered by cold or a low-resolution face image. A simple but useful solution is the employment of multiple modalities, e.g., gender determination using fusion of face & ear in [88], face & gait in [148], and face & voice in [254]. As demographic systems are expected to operate in unconstrained conditions, fusing different modalities certainly offers an alternative to provide supplementary information, and thus to render them more robust.

7.6 Integration with Mobile & Wearable Devices

As technology progresses, new hardware continues to emerge. Rapid improvements in both computing and storage allow deploying more powerful algorithms to process data. Nowadays, more and more mobile & wearable devices support biometric technology for quick user verification. It is intriguing to include also demographic functionality in these devices. Nevertheless, there is far less such work in both academia and industry, than biometric authentication. An appealing example is the Google Glass software from Fraunhofer Institute[12] which can read the emotion of everyone you talk to and tell you their age and gender.

8 Conclusion

Biometric data conveys a wealth of personal information. Demographic attributes are among these signals. Over the decades, we have witnessed tremendous efforts devoted to biometric demographic analysis, with a good deal of progress achieved. This survey provided a comprehensive overview of these efforts and achievements. We began by listing some interesting points on human demographic perception and then traced the history of automatic demographic estimation. A systematic review of the state-of-the-art was then provided. Finally, we highlighted some critical challenges and offered some insights into the future.

Acknowledgments

We would like to thank the associate editor and anonymous reviewers for their constructive comments and significant efforts spent to help us for further improving the survey. This work was supported in part by the National Key Research and Development Program of China (Grant No. 2016YFB1001000) and the National Natural Science Foundation of China (Grant No. 61603391, 61603385). Man Zhang is the corresponding author.

References

[1] A. K. Jain, A. Ross, and K. Nandakumar, *Introduction to Biometrics*. Berlin, Germany: Springer, 2011.
[2] M. Trauring, "Automatic comparison of finger-ridge patterns," *Nature*, vol. 197, pp. 938–940, 1963.
[3] A. K. Jain, K. Nandakumar, and A. Ross, "50 years of biometric research: Accomplishments, challenges, and opportunities," *Pattern Recognit. Lett.*, vol. 79, pp. 80–105, 2015.
[4] P. H. Ptacek and E. K. Sander, "Age recognition from voice," *J. Speech Language Hearing Res.*, vol. 9, no. 2, pp. 273–277, 1966.

12. www.iis.fraunhofer.de/en.html

[5] V. Bruce, A. M. Burton, E. Hanna, P. Healey, O. Mason, and A. Coombes, "Sex discrimination: How do we tell the difference between male and female faces?" *Perception*, vol. 22, no. 2, pp. 131–152, 1993.
[6] K. Ricanek Jr and T. Tesafaye, "MORPH: A longitudinal image database of normal adult age-progression," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recognit.*, 2006, pp. 341–345.
[7] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Trans. Inf. Forensics Secur.*, vol. 9, no. 12, pp. 2170–2179, Dec. 2014.
[8] S. Yu, D. Tan, and T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2006, pp. 441–444.
[9] F. Burkhardt, M. Eckert, W. Johannsen, and J. Stegmann, "A database of age and gender annotated telephone speech," in *Proc. Int. Conf. Language Resources Eval.*, 2010, pp. 1562–1565.
[10] X. Qiu, Z. Sun, and T. Tan, "Learning appearance primitives of iris images for ethnic classification," in *Proc. IEEE Int. Conf. Image Process.*, 2007, pp. II-405–II-408.
[11] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 1955–1976, Nov. 2010.
[12] H. Han, C. Otto, X. Liu, and A. Jain, "Demographic estimation from face images: Human versus machine performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 6, pp. 1148–1161, Jun. 2015.
[13] S. Fu, H. He, and Z. Hou, "Learning race from face: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 12, pp. 2483–2509, Dec. 2014.
[14] C. B. NG, Y. H. Tay, and B. M. Goi, "Vision-based human gender recognition: A survey," *CoRR*, vol. abs/1204.1611, 2012.
[15] What's in a Face, in *Proc. Eur. Conf. Comput. Vis.*, 2012. [Online]. Available: https://sites.google.com/site/wiaf2012/call-for-papers
[16] International workshop on soft biometrics, in *Proc. Eur. Conf. Comput. Vis.*, 2014. [Online]. Available: https://sites.google.com/site/softbiometrics2014/
[17] B. Schuller, et al., "The INTERSPEECH 2010 paralinguistic challenge," in *Proc. INTERSPEECH*, 2010, pp. 2795–2798.
[18] S. Escalera, et al., "ChaLearn looking at people 2015 new competitions: Age estimation and cultural event recognition," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2015, pp. 1–8.
[19] A. Dantcheva, P. Elia, and A. Ross, "What else does your biometric data reveal? A survey on soft biometrics," *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 3, pp. 441–467, Mar. 2016.
[20] M. S. Nixon, P. L. Correia, K. Nasrollahi, T. B. Moeslund, A. Hadid, and M. Tistarelli, "On soft biometrics," *Pattern Recognit. Lett.*, vol. 68, pp. 218–230, 2015.
[21] A. Ross and A. K. Jain, "Multimodal biometrics: An overview," in *Proc. IEEE Eur. Conf. Signal Process.*, 2004, pp. 1221–1224.
[22] A. K. Jain, S. C. Dass, and K. Nandakumar, "Soft biometric traits for personal recognition systems," in *Proc. Int. Conf. Biometric Authentication*, 2004, pp. 731–738.
[23] S. Marcel, M. S. Nixon, and S. Z. Li, *Handbook of Biometric Anti-Spoofing*. Berlin, Germany: Springer, 2014.
[24] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts, Amherst, MA, USA, Tech. Rep. 07–49, vol. 1, no. 2, 2007.
[25] A. Stone, "The aging process of the face & techniques of rejuvenation," 2010. [Online]. Available: http://www.aaronstonemd.com/Facial_Aging_Rejuvenation.shtm
[26] L. S. Mark, J. B. Pittenger, H. Hines, C. Carello, R. E. Shaw, and J. T. Todd, "Wrinkling and head shape as coordinated sources of age-level information," *Perception Psychophysics*, vol. 27, no. 2, pp. 117–124, 1980.
[27] N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2006, pp. 387–394.
[28] W. S. Brown, R. J. Morris, H. Hollien, and E. Howell, "Speaking fundamental frequency characteristics as a function of age and professional singing," *J. Voice*, vol. 5, no. 4, pp. 310–315, 1991.
[29] S. E. Linville and H. B. Fisher, "Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females," *J. Acoustical Soc. America*, vol. 78, no. 1, pp. 40–48, 1985.
[30] K. A. Wilcox and Y. Horii, "Age and changes in vocal jitter," *J. Gerontology*, vol. 35, no. 2, pp. 194–198, 1980.

[31] J. E. Himann, D. A. Cunningham, P. A. Rechnitzer, and D. H. Paterson, "Age-related changes in speed of walking," *Med. Sci. Sports Exercise*, vol. 20, no. 2, pp. 161–166, 1988.

[32] R. W. Bohannon, "Comfortable and maximum walking speed of adults aged 20–79 years: Reference values and determinants," *Age Ageing*, vol. 26, no. 1, pp. 15–19, 1997.

[33] H. B. Menz, S. R. Lord, and R. C. Fitzpatrick, "Age-related differences in walking stability," *Age Ageing*, vol. 32, no. 2, pp. 137–142, 2003.

[34] A. M. Burton, V. Bruce, and N. Dench, "What's the difference between men and women? evidence from facial measurement," *Perception*, vol. 22, no. 2, pp. 153–176, 1993.

[35] I. R. Titze, "Physiologic and acoustic differences between male and female voices," *J. Acoustical Soc. America*, vol. 85, no. 4, pp. 1699–1707, 1989.

[36] R. O. Coleman, "Male and female voice quality and its relationship to vowel formant frequencies," *J. Speech Language Hearing Res.*, vol. 14, no. 3, pp. 565–577, 1971.

[37] N. J. Lass, K. R. Hughes, M. D. Bowyer, L. T. Waters, and V. T. Bourne, "Speaker sex identification from voiced, whispered, and filtered isolated vowels," *J. Acoustical Soc. America*, vol. 59, no. 3, pp. 675–678, 1976.

[38] G. Mather and L. Murdoch, "Gender discrimination in biological motion displays based on dynamic cues," *Proc. Roy. Soc. London B: Biol. Sci.*, vol. 258, no. 1353, pp. 273–279, 1994.

[39] L. T. Kozlowski and J. E. Cutting, "Recognizing the sex of a walker from a dynamic point-light display," *Perception Psychophysics*, vol. 21, no. 6, pp. 575–580, 1977.

[40] W. H. Sheldon, S. S. Stevens, and W. B. Tucker, *The Varieties of Human Physique: An Introduction to Constitutional Psychology*, vol. 1. New York, NY, USA: Harper, 1940.

[41] M. J. Muñoz-Cachón, I. Salces, M. Arroyo, L. Ansotegui, A. M. Rocandio, and E. Rebato, "Body shape in relation to socio-economic status in young adults from the Basque Country," *Collegium Antropologicum*, vol. 31, no. 4, pp. 963–968, 2007.

[42] S. Hamid and K. M. Loewenthal, "Inferring gender from handwriting in Urdu and English," *J. Soc. Psychology*, vol. 136, no. 6, pp. 778–782, 1996.

[43] A. K. Agnihotri, B. Purwar, N. Jeebun, and S. Agnihotri, "Determination of sex by hand dimensions," *Internet J. Forensic Sci.*, vol. 1, no. 2, pp. 12–24, 2006.

[44] S. Hosoi, E. Takikawa, and M. Kawade, "Ethnicity estimation with facial images," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recognit.*, 2004, pp. 195–200.

[45] H. Proença, S. Filipe, R. Santos, J. Oliveira, and L. A. Alexandre, "The UBIRIS.v2: A database of visible wavelength iris images captured on-the-move and at-a-distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 8, pp. 1529–1535, Aug. 2010.

[46] D. G. Childers, K. Wu, K. S. Bae, and D. M. Hicks, "Automatic recognition of gender by voice," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 1988, pp. 603–606.

[47] Y. L. Shue and M. Iseli, "The role of voice source measures on automatic gender classification," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2008, pp. 4493–4496.

[48] Y. H. Kwon and N. D. V. Lobo, "Age classification from facial images," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 1994, pp. 762–767.

[49] J. W. Davis, "Visual categorization of children and adult walking styles," in *Proc. 3rd Int. Conf. Audio Video-Based Biometric Person Authentication*, 2001, pp. 295–300.

[50] N. Minematsu, M. Sekiguchi, and K. Hirose, "Automatic estimation of one's age with his/her speech based upon acoustic modeling techniques of speakers," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2002, pp. I-137–I-140.

[51] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 442–455, Apr. 2002.

[52] S. K. Zhou, B. Georgescu, X. S. Zhou, and D. Comaniciu, "Image based regression using boosting method," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2005, pp. 541–548.

[53] Z. Yang and H. Ai, "Demographic classification with local binary patterns," in *Proc. IAPR Int. Conf. Biometrics*, 2007, pp. 464–473.

[54] H. C. Lian and B. L. Lu, "Multi-view gender classification using local binary patterns and support vector machines," in *Proc. Int. Symp. Neural Netw.*, 2006, pp. 202–209.

[55] E. Mäkinen and R. Raisamo, "An experimental comparison of gender classification methods," *Pattern Recognit. Lett.*, vol. 29, no. 10, pp. 1544–1556, 2008.

[56] L. A. Alexandre, "Gender recognition: A multiscale decision fusion approach," *Pattern Recognit. Lett.*, vol. 31, no. 11, pp. 1422–1427, 2010.

[57] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 112–119.

[58] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1178–1188, Jul. 2008.

[59] Y. Fu and T. S. Huang, "Human age estimation with regression on discriminative aging manifold," *IEEE Trans. Multimedia*, vol. 10, no. 4, pp. 578–584, Jun. 2008.

[60] T. Bocklet, A. Maier, J. G. Bauer, F. Burkhardt, and E. Nöth, "Age and gender recognition for telephone applications based on GMM supervectors and support vector machines," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2008, pp. 1605–1608.

[61] Y. Makihara, H. Mannami, and Y. Yagi, "Gait analysis of gender and age using a large-scale multi-view gait database," in *Proc. Asian Conf. Comput. Vis.*, 2010, pp. 440–451.

[62] B. Schuller, et al., "Paralinguistics in speech and language–state-of-the-art and the challenge," *Comput. Speech Language*, vol. 27, no. 1, pp. 4–39, 2013.

[63] T. S. Lee, "Image representation using 2D Gabor wavelets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 959–971, Oct. 1996.

[64] J. Man and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 2, pp. 316–322, Feb. 2006.

[65] K. Y. Chang, C. S. Chen, and Y. P. Huang, "Ordinal hyperplanes ranker with cost sensitivities for age estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 585–592.

[66] M. Ngan and P. Grother, "Face recognition vendor test (FRVT) performance of automated age estimation algorithms," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, NIST Interagency Rep. 7995, 2014.

[67] D. Yi, Z. Lei, and S. Z. Li, "Age estimation by multi-scale convolutional network," in *Proc. Asian Conf. Comput. Vis.*, 2014, pp. 144–158.

[68] F. Gao and H. Ai, "Face age classification on consumer images with Gabor feature and fuzzy LDA method," in *Proc. IAPR Int. Conf. Biometrics*, 2009, pp. 132–141.

[69] Y. Deng, P. Luo, C. C. Loy, and X. Tang, "Learning to recognize pedestrian attribute," *CoRR*, vol. 1501.00901, 2015.

[70] G. W. Cottrell and J. Metcalfe, "EMPATH: Face, emotion, and gender recognition using Holons," in *Proc. Advances Neural Inf. Process. Syst.*, 1990, pp. 564–571.

[71] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski, "SEXNET: A neural network identifies sex from human faces," in *Proc. Conf. Advances Neural Inf. Process. Syst.*, 1990, pp. 572–577.

[72] E. S. Parris and M. J. Carey, "Language independent gender identification," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 1996, pp. 685–688.

[73] T. Huang, Y. Yang, and Z. Wu, "Combining MFCC and pitch to enhance the performance of the gender recognition," in *Proc. IEEE Int. Conf. Signal Process.*, 2006, vol. 1.

[74] A. J. O'toole, T. Vetter, N. F. Troje, and H. H. Bülthoff, "Sex classification is better with three-dimensional head structure than with image intensity information," *Perception*, vol. 26, no. 1, pp. 75–84, 1997.

[75] B. Moghaddam and M. H. Yang, "Gender classification with support vector machines," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recognit.*, 2000, pp. 306–311.

[76] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recognit.*, 2002, pp. 148–155.

[77] B. Wu, H. Ai, and C. Huang, "LUT-based adaboost for gender classification," in *Proc. 4th Int. Conf. Audio Video-Based Biometric Person Authentication*, 2003, pp. 104–110.

[78] I. Shafran, M. Riley, and M. Mohri, "Voice signatures," in *Proc. IEEE Workshop Automat. Speech Recognit. Understanding*, 2003, pp. 31–36.

[79] J. H. Yoo, D. Hwang, and M. S. Nixon, "Gender classification in human gait using support vector machine," in *Proc. Adv. Concepts Intell. Vis. Syst.*, 2005, pp. 138–145.

[80] X. Lu, H. Chen, and A. K. Jain, "Multimodal facial gender and ethnicity identification," in *Proc. IAPR Int. Conf. Biometrics*, 2006, pp. 554–561.

[81] A. M. Badawi, M. Mahfouz, R. Taqdross, and R. Jantz, "Fingerprint-based gender classification," in *Proc. Int. Conf. Image Process. Comput. Vis. Pattern Recognit.*, 2006, pp. 41–46.

[82] V. Thomas, N. V. Chawla, K. W. Bowyer, and P. J. Flynn, "Learning to predict gender from iris images," in *Proc. IEEE Int. Conf. Biometrics: Theory Appl. Syst.*, 2007, pp. 1–5.

[83] L. Cao, M. Dikmen, Y. Fu, and T. S. Huang, "Gender recognition from body," in *Proc. ACM Int. Conf. Multimedia*, 2008, pp. 725–728.

[84] X. Li, S. J. Maybank, S. Yan, D. Tao, and D. Xu, "Gait components and their application to gender recognition," *IEEE Trans. Syst. Man Cybern. C Appl. Rev.*, vol. 38, no. 2, pp. 145–155, Mar. 2008.

[85] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2009, pp. 365–372.

[86] S. Yu, T. Tan, K. Huang, K, Jia, and X. Wu, "A study on gait-based gender classification," *IEEE Trans. Image Process.*, vol. 18, no. 8, pp. 1905–1910, Aug. 2009.

[87] S. Wuhrer, C. Shu, and M. Rioux, "Posture invariant gender classification for 3D human models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2009, pp. 33–38.

[88] G. Zhang and Y. Wang, "Hierarchical and discriminative bag of features for face profile and ear based gender classification," in *Proc. IEEE Int. Joint Conf. Biometrics*, 2011, pp. 1–8.

[89] A. Hassaïne, S. Al Maadeed, J. Aljaam, and A. Jaoua, "ICDAR 2013 competition on gender prediction from handwriting," in *Proc. Int. Conf. Document Anal. Recognit.*, 2013, pp. 1417–1421.

[90] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2015, pp. 34–42.

[91] J. Zhu, S. Liao, D. Yi, Z. Lei, and S. Z. Li, "Multi-label CNN based pedestrian attribute learning for soft biometrics," in *Proc. IAPR Int. Conf. Biometrics*, 2015, pp. 535–540.

[92] S. Gutta, H. Wechsler, and P. J. Phillips, "Gender and ethnic classification of face images," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recognit.*, 1998, pp. 194–199.

[93] M. J. Lyons, J. Budynek, A. Plante, and S. Akamatsu, "Classifying facial attributes using a 2-D Gabor wavelet representation and discriminant analysis," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recognit.*, 2000, pp. 202–207.

[94] G. Shakhnarovich, P. A. Viola, and B. Moghaddam, "A unified learning framework for real time face detection and classification," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recognit.*, 2002, pp. 14–21.

[95] X. Lu and A. K. Jain, "Ethnicity identification from face images," in *Proc. Int. Soc. Opt. Photon.*, 2002, pp. 114–123.

[96] X. Qiu, Z. Sun, and T. Tan, "Global texture analysis of iris images for ethnic classification," in *Proc. IAPR Int. Conf. Biometrics*, 2006, pp. 411–418.

[97] N. Kumar, P. Belhumeur, and S. Nayar, "FaceTracer: A search engine for large collections of images with faces," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 340–353.

[98] U. Tariq, Y. Hu, and T. S. Huang, "Gender and ethnicity identification from silhouetted face profiles," in *Proc. IEEE Int. Conf. Image Process.*, 2009, pp. 2441–2444.

[99] A. Hadid and M. Pietikäinen, "Demographic classification from face videos using manifold learning," *Neurocomputing*, vol. 100, pp. 197–205, 2013.

[100] Y. Xie, K. Luu, and M. Savvides, "A robust approach to facial ethnicity classification on large scale face databases," in *Proc. IEEE Int. Conf. Biometrics: Theory Appl. Syst.*, 2012, pp. 143–149.

[101] Z. Sun, H. Zhang, T. Tan, and J. Wang, "Iris image classification based on hierarchical visual codebook," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1120–1133, Jun. 2014.

[102] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 3730–3738.

[103] M. Yang, S. Zhu, F. Lv, and K. Yu, "Correspondence driven adaptation for human profile recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 505–512.

[104] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 529–534.

[105] L. Wiskott, J. M. Felous, N. Krüger, and C. von der Malsburg, "Face recognition and gender determination," in *Proc. Int. Workshop Automat. Face Gesture Recognit.*, 1995, pp. 92–97.

[106] R. Brunelli and T. Poggio, "HyberBF networks for gender classification," in *Proc. DARPA Image Understanding Workshop*, 1992, pp. 311–314.

[107] A. Samal, V. Subramani, and D. Marx, "Analysis of sexual dimorphism in human face," *J. Vis. Commun. Image Representation*, vol. 18, no. 6, pp. 453–463, 2007.

[108] H. Abdi, D. Valentin, B. Edelman, and A. J. O'Toole, "More about the difference between men and women: Evidence from linear neural networks and the principal-component approach," *Perception*, vol. 24, no. 5, pp. 539–562, 1995.

[109] J. Bekios-Calfa, J. M. Buenaposada, and T. Baumela, "Revisiting linear discriminant techniques in gender recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 4, pp. 858–864, Apr. 2011.

[110] K. Balci and V. Atalay, "PCA for gender estimation: Which eigenvectors contribute?" in *Proc. IEEE Int. Conf. Pattern Recognit.*, 2002, pp. 363–366.

[111] A. B. Graf and F. A. Wichmann, "Gender classification of human faces," in *Proc. Int. Workshop Biologically Motivated Comput. Vis.*, 2002, pp. 491–500.

[112] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Trans. Syst. Man Cybern. B Cybern.*, vol. 34, no. 1, pp. 621–628, Feb. 2004.

[113] X. Geng, Z. H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2234–2240, Dec. 2007.

[114] Y. Saatci and C. Twon, "Cascaded classification of gender and facial expression using active appearance models," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recognit.*, 2006, pp. 393–398.

[115] N. P. Costen, M. Brown, and S. Akamatsu, "Sparse models for gender classification," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recognit.*, 2004, pp. 201–206.

[116] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim, "Age estimation using a hierarchical classifier based on global and local facial features," *Pattern Recognit.*, vol. 44, no. 6, pp. 1262–1281, 2011.

[117] S. Yan, M. Liu, and T. S. Huang, "Extracting age information from local spatially flexible patches," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2008, pp. 737–740.

[118] X. Zhuang, X. Zhou, M. Hasegawa-Johnson, and T. Huang, "Face age estimation using patch-based hidden Markov model supervectors," in *Proc. IEEE Int. Conf. Pattern Recognit.*, 2008, pp. 1–4.

[119] T. Jabid, M. H. Kabir, and O. Chae, "Gender classification using local directional pattern (LDP)," in *Proc. IEEE Int. Conf. Pattern Recognit.*, 2010, pp. 2162–2165.

[120] L. Ballihi, B. Ben Amor, M. Daoudi, A. Srivastava, and D. Aboutajdine, "Boosting 3-D-geometric features for efficient face recognition and gender classification," *IEEE Trans. Inf. Forensics Secur.*, vol. 7, no. 6, pp. 1766–1779, Dec. 2012.

[121] X. Han, H. Ugail, and I. Palmer, "Gender classification based on 3D face geometry features using SVM," in *Proc. IEEE Int. Conf. CyberWorlds*, 2009, pp. 114–118.

[122] J. Wu, W. A. Smith, and E. R. Hancock, "Facial gender classification using shape-from-shading," *Image Vis. Comput.*, vol. 28, no. 6, pp. 1039–1048, 2010.

[123] T. Huynh, R. Min, and J. L. Dugelay, "An efficient LBP-based descriptor for facial depth images applied to gender recognition using RGB-D face data," in *Proc. Asian Conf. Comput. Vis. Workshops*, 2012, pp. 133–145.

[124] B. Xia, B. B. Amor, and M. Daoudi, "Exploring the magnitude of human sexual dimorphism in 3D face gender classification," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2014, pp. 697–710.

[125] G. Toderici, S. M. Omalley, G. Passalis, T. Theoharis, and I. A. Kakadiaris, "Ethnicity-and gender-based subject retrieval using 3-D face-recognition techniques," *Int. J. Comput. Vis.*, vol. 89, no. 2/3, pp. 382–391, 2010.

[126] H. Ding, D. Huang, Y. Wang, and L. Chen, "Facial ethnicity classification based on boosted local texture and shape descriptions," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recognit.*, 2013, pp. 1–6.

[127] O. Ocegueda, T. Fang, S. K. Shah, and I. A. Kakadiaris, "3D face discriminant analysis using Gauss-Markov posterior marginals," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 728–739, Mar. 2013.

[128] D. Huang, H. Ding, C. Wang, Y. Wang, G. Zhang, and L. Chen, "Local circular patterns for multi-modal facial gender and ethnicity classification," *Image Vis. Comput.*, vol. 32, no. 12, pp. 1181–1193, 2014.

[129] A. Hadid and M. Pietikäinen, "Combining appearance and motion for face and gender recognition from videos," *Pattern Recognit.*, vol. 42, no. 11, pp. 2818–2827, 2009.

[130] A. Hadid, "Analyzing facial behavioral features from videos," in *Proc. Int. Workshop Human Behavior Understanding*, 2011, pp. 52–61.
[131] H. Dibeklioğlu, F. Alnajar, A. A. Salah, and T. Gevers, "Combining facial dynamics with appearance for age estimation," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1928–1943, Jun. 2015.
[132] A. Dantcheva and F. Brémond, "Gender estimation based on smile-dynamics," *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 3, pp. 719–729, Mar. 2017.
[133] S. Roomi, S. L. Virasundarii, S. Selvamegala, S. Jeevanandham, and D. Hariharasudhan, "Race classification based on facial features," in *Proc. IEEE National Conf. Comput. Vis. Pattern Recognit. Image Process. Graph.*, 2011, pp. 54–57.
[134] C. Chen and A. Ross, "Evaluation of gender classification methods on thermal and near-infrared face images," in *Proc. IEEE Int. Joint Conf. Biometrics*, 2011, pp. 1–8.
[135] H. Meinedo and I. Trancoso, "Age and gender classification using fusion of acoustic and prosodic features," in *Proc. INTERSPEECH*, 2010, pp. 2818–2821.
[136] C. A. Müller and F. Burkhardt, "Combining short-term cepstral and long-term pitch features for automatic recognition of speaker age," in *Proc. INTERSPEECH*, 2007, pp. 2277–2280.
[137] C. Müller, F. Witting, and J. Baus, "Exploiting speech for recognizing elderly users to respond to their special needs," in *Proc. Eur. Conf. Speech Commun. Technol.*, 2003, pp. 1305–1308.
[138] F. Metze, et al., "Comparison of four approaches to age and gender recognition for telephone applications," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2007, pp. IV-1089–IV-1092.
[139] S. Yan, H. Wang, X. Tang, and T. S. Huang, "Learning auto-structured regressor from uncertain nonnegative labels," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2007, pp. 1–8.
[140] M. Wolters, R. Vipperla, and S. Renals, "Age recognition for spoken dialogue systems: Do we need it?" in *Proc. INTERSPEECH*, 2009, pp. 1435–1438.
[141] J. Ajmera and F. Burkhardt, "Age and gender classification using modulation cepstrum," in *Proc. Odyssey: Speaker Language Recognit. Workshop*, 2008, p. 25.
[142] R. K. Begg, M. Palaniswami, and B. Owen, "Support vector machines for automated gait classification," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 5, pp. 828–838, May 2005.
[143] J. W. Davis and H. Gao, "Gender recognition from walking movements using adaptive three-mode PCA," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshop*, 2004, pp. 9–9.
[144] N. F. Troje, "Decomposing biological motion: A framework for analysis and synthesis of human gait patterns," *J. Vis.*, vol. 2, no. 5, pp. 371–378, 2002.
[145] G. Huang and Y. Wang, "Gender classification based on fusion of multi-view gait sequences," in *Proc. Asian Conf. Comput. Vis.*, 2007, pp. 462–471.
[146] M. Hu, Y. Wang, Z. Zhang, and D. Zhang, "Gait-based gender classification using mixed conditional random field," *IEEE Trans. Syst. Man Cybern. B Cybern.*, vol. 41, no. 5, pp. 1429–1439, Oct. 2011.
[147] R. Martín-Félez, R. A., Mollineda, and J. S., Sánchez, "Towards a more realistic appearance-based gait representation for gender recognition," in *Proc. Int. Conf. Pattern Recognit.*, 2010, pp. 3810–3813.
[148] C. Shan, S. Gong, and P. W. McOwan, "Fusing gait and face cues for human gender recognition," *Neurocomputing*, vol. 71, no. 10, pp. 1931–1938, 2008.
[149] J. Lu and Y. P. Tan, "Uncorrelated discriminant simplex analysis for view-invariant gait signal computing," *Pattern Recognit. Lett.*, vol. 31, no. 5, pp. 382–393, 2010.
[150] J. Lu, G. Wang, and P. Moulin, "Human identity and gender recognition from gait sequences with arbitrary walking directions," *IEEE Trans. Inf. Forensics Secur.*, vol. 9, no. 1, pp. 51–61, Jan. 2014.
[151] L. Chen, Y. Wang, Y. Wang, and D. Zhang, "Gender recognition from gait using radon transform and relevant component analysis," in *Proc. 5th Int. Conf. Emerging Intell. Comput. Technol. Appl.*, 2009, pp. 92–101.
[152] F. B. Oskuie and K. Faez, "Gender classification using a novel gait template: Radon transform of mean gait energy image," in *Proc. 8th Int. Conf. Image Anal. Recognit.*, 2011, pp. 161–169.
[153] Y. Makihara, M. Okumura, H. Iwama, and Y. Yagi, "Gait-based age estimation using a whole-generation gait database," in *Proc. IEEE Int. Joint Conf. Biometrics*, 2011, pp. 1–6.
[154] M. Hu, Y. Wang, Z. Zhang, and Y. Wang, "Combining spatial and temporal information for gait based gender classification," in *Proc. Int. Conf. Pattern Recognit.*, 2010, pp. 3679–3682.
[155] J. Lu and Y. P. Yan, "Gait-based human age estimation," *IEEE Trans. Inf. Forensics Secur.*, vol. 5, no. 4, pp. 761–770, Dec. 2010.
[156] M. Collins, J. Zhang, P. Miller, and H. Wang, "Full body image feature representations for gender profiling," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2009, pp. 1235–1242.
[157] L. Bourdev, S. Maji, and J. Malik, "Describing people: A poselet-based approach to attribute classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 1543–1550.
[158] A. Lapedriza, M. J. Marín-Jiménez, and J. Vitrià, "Gender recognition in non controlled environments," in *Proc. Int. Conf. Pattern Recognit.*, 2006, pp. 834–837.
[159] S. Baluja and H. A. Rowley, "Boosting sex identification performance," *Int. J. Comput. Vis.*, vol. 71, no. 1, pp. 111–119, 2007.
[160] E. Mäkinen and R. Raisamo, "Evaluation of gender classification methods with automatically detected and aligned faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 541–547, Mar. 2008.
[161] J. Zhu, S. Liao, Z. Lei, D. Yi, and S. Li, "Pedestrian attribute classification in surveillance: Database and evaluation," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2013, pp. 331–338.
[162] G. Guo, G. Mu, and Y. Fu, "Gender from body: A biologically-inspired approach with manifold learning," in *Proc. Asian Conf. Comput. Vis.*, 2009, pp. 236–245.
[163] R. Layne, T. M. Hospedales, and S. Gong, "Towards person identification and re-identification with attributes," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 402–412.
[164] J. Tang, X. Liu, H. Cheng, and K. M. Robinette, "Gender recognition using 3-D human body shapes," *IEEE Trans. Syst., Man, Cybern. C Appl. Rev.*, vol. 41, no. 6, pp. 898–908, Nov. 2011.
[165] M. Verma and S. Agarwal, "Fingerprint based male-female classification," in *Proc. Int. Workshop Comput. Intell. Secur. Inf. Syst.*, 2009, pp. 251–257.
[166] M. Toews and T. Arbel, "Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1567–1581, Sep. 2009.
[167] Z. Sun, G. Bebis, X. Yuan, and S. J. Louis, "Genetic feature subset selection for gender classification: A comparison study," in *Proc. Int. Workshop Appl. Comput. Vis.*, 2002, pp. 165–170.
[168] A. Jain and J. Huang, "Integrating independent components and linear discriminant analysis for gender classification," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recognit.*, 2004, pp. 159–163.
[169] Z. Yang, M. Li, and H. Ai, "An experimental study on automatic face gender classification," in *Proc. Int. Conf. Pattern Recognit.*, 2006, pp. 1099–1102.
[170] P. Gnanasivam and D. S. Muttan, "Fingerprint gender classification using wavelet transform and singular value decomposition," *CoRR*, vol. 1205.6745, 2012.
[171] H. C. Kim, D. Kim, Z. Ghahramani, and S. Y. Bang, "Appearance-based gender classification with gaussian processes," *Pattern Recognit. Lett.*, vol. 27, no. 6, pp. 618–626, 2006.
[172] X. Li, X. Zhao, Y. Fu, and Y. Liu, "Bimodal gender recognition from face and fingerprint," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 2590–2597.
[173] A. Rattani, C. Chen, and A. Ross, "Evaluation of texture descriptors for automated gender estimation from fingerprints," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2014, pp. 764–777.
[174] S. Lagree and K. W. Bowyer, "Ethnicity prediction based on iris texture features," in *Proc. Midwest Artif. Intell. Cogn. Sci. Conf.*, 2011, pp. 225–230.
[175] S. Lagree and K. W. Bowyer, "Predicting ethnicity and gender from iris texture," in *Proc. IEEE Int. Conf. Technol. Homeland Secur.*, 2011.
[176] A. Zarei and D. Mou, "Artificial neural network for prediction of ethnicity based on iris texture," in *Proc. IEEE Int. Conf. Mach. Learn. Appl.*, 2012, pp. 514–519.
[177] H. Zhang, Z. Sun, T. Tan, and J. Wang, "Ethnic classification based on iris images," in *Proc. Biometric Recognit.*, 2011, pp. 82–90.
[178] J. E. Tapia, C. A. Perez, and K. W. Bowyer, "Gender classification from iris images using fusion of uniform local binary patterns," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2014, pp. 751–763.
[179] A. Bansal, R. Agarwal, and R. K. Sharma, "SVM based gender classification using iris images," in *Proc. IEEE Int. Conf. Comput. Intell. Commun. Netw.*, 2012, pp. 425–429.

[180] A. Bansal, R. Agarwal, and R. K. Sharma, "Predicting gender using iris images," *Res. J. Recent Sci.*, vol. 3, no. 4, pp. 20–26, 2014.

[181] C. I. Tomai, D. M. Kshirsagar, and S. N. Srihari, "Group discriminatory power of handwritten characters," in *Proc. Int. Soc. Opt. Photon.*, 2003, pp. 116–123.

[182] K. Bandi and S. N. Srihari, "Writer demographic classification using bagging and boosting," in *Proc. Int. Graphonomics Soc. Conf.*, 2005, pp. 133–137.

[183] M. Liwicki, A. Schlapbach, P. Loretan, and H. Bunke, "Automatic detection of gender and handedness from on-line handwriting," in *Proc. IEEE Int. Conf. Graphonomics Soc.*, 2007, pp. 179–183.

[184] J. Merkow, B. Jow, and M. Savvides, "An exploration of gender identification using only the periocular region," in *Proc. IEEE Int. Conf. Biometrics Theory Appl. Syst.*, 2010, pp. 1–5.

[185] S. Tamura, H. Kawai, and H. Mitsumoto, "Male/female identification from 8×6 very low resolution face images by neural network," *Pattern Recognit.*, vol. 29, no. 2, pp. 331–335, 1996.

[186] J. R. Lyle, P. E. Miller, S. J. Pundlik, and D. L. Woodard, "Soft biometric classification using local appearance periocular region features," *Pattern Recognit.*, vol. 45, no. 11, pp. 3877–3885, 2012.

[187] R. Khorsandi and M. Abdel-Mottaleb, "Gender classification using 2-D ear images and sparse representation," in *Proc. IEEE Workshop Appl. Comput. Vis.*, 2013, pp. 461–466.

[188] J. Lei, J. Zhou, and M. Abdel-Mottaleb, "Gender classification using automatically detected and aligned 3D ear range data," in *Proc. IAPR Int. Conf. Biometrics*, 2013, pp. 1–7.

[189] G. Amayeh, G. Bebis, and M. Nicolescu, "Gender classification from hand shape," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2008, pp. 1–7.

[190] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 1995.

[191] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Proc. Eur. Conf. Comput. Learn. Theory*, 1995, pp. 23–37.

[192] K. Luu, K. Seshadri, M. Savvides, T. D. Bui, and C. Y. Suen, "Contourlet appearance model for facial age estimation," in *Proc. IEEE Int. Joint Conf. Biometrics*, 2011, pp. 1–8.

[193] H. Fang, P. Grant, and M. Chen, "Discriminant feature manifold for facial aging estimation," in *Proc. Int. Conf. Pattern Recognit.*, 2010, pp. 593–596.

[194] A. Montillo and H. Ling, "Age regression from faces using random forests," in *Proc. IEEE Int. Conf. Image Process.*, 2009, pp. 2465–2468.

[195] S. Yan, Z. Zhou, M. Liu, M. Hasegawa-Johnson, and T. S. Huang, "Regression from patch-kernel," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.

[196] F. Alnajar, C. Shan, T. Gevers, and J. M. Geusebroek, "Learning-based encoding with soft assignment for age estimation under unconstrained imaging conditions," *Image Vis. Comput.*, vol. 30, no. 12, pp. 946–953, 2012.

[197] W. Gao and H. Ai, "Face gender classification on consumer images in a multiethnic environment," in *Proc. IAPR Int. Conf. Biometrics*, 2009, pp. 169–178.

[198] C. Shan, "Learning local binary patterns for gender classification on real-world face images," *Pattern Recognit. Lett.*, vol. 33, no. 4, pp. 431–437, 2012.

[199] S. Jia and N. Cristianini, "Learning to classify gender from four million images," *Pattern Recognit. Lett.*, vol. 58, pp. 35–41, 2015.

[200] R. Ranjan, V. M. Patel, and R. Chellappa, "HyperFace: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," *CoRR*, vol. 1603.01249, 2016.

[201] G. Antipov, S. A. Berrani, and J. L. Dugelay, "Minimalistic CNN-based ensemble model for gender prediction from face images," *Pattern Recognit. Lett.*, vol. 70, pp. 59–65, 2016.

[202] B. Ni, Z. Song, and S. Yan, "Web image mining towards universal age estimator," in *Proc. ACM Int. Conf. Multimedia*, 2009, pp. 58–94.

[203] B. Xiao, X. Yang, Y. Xu, and H. Zha, "Learning distance metric for regression by semidefinite programming with application to human age estimation," in *Proc. ACM Int. Conf. Multimedia*, 2009, pp. 451–460.

[204] G. Guo and G. Mu, "Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 657–664.

[205] K. Y. Chang and C. S. Chen, "A learning framework for age rank estimation based on face images with scattering transform," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 785–798, Mar. 2015.

[206] C. Li, Q. Liu, J. Liu, and H. Lu, "Learning ordinal discriminative features for age estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2570–2577.

[207] W. L. Chao, J. Z. Liu, and J. J. Ding, "Facial age estimation based on label-sensitive learning and age-oriented regression," *Pattern Recognit.*, vol. 46, no. 3, pp. 628–641, 2013.

[208] K. Chen, S. Gong, T. Xiang, and C. C. Loy, "Cumulative attribute space for age and crowd density estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 2467–2474.

[209] Y. L. Chen and C. T. Hsu, "Subspace learning for facial age estimation via pairwise age ranking," *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 12, pp. 2164–2176, Dec. 2013.

[210] Z. Hu, Y. Wen, J. Wang, M. Wang, R. Hong, and S. Yan, "Facial age estimation with age difference," *IEEE Trans. Image Process.*, vol. PP, no. 99, p. 1, 2016, doi: 10.1109/TIP.2016.2633868.

[211] J. Liu, Y. Ma, L. Duan, F. Wang, and Y. Liu, "Hybrid constraint SVR for facial age estimation," *Signal Process.*, vol. 94, pp. 576–582, 2014.

[212] X. Geng, C. Yin, and Z. H. Zhou, "Facial age estimation by learning from label distributions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 10, pp. 2401–2412, Oct. 2013.

[213] X. Yang, et al., "Deep label distribution learning for apparent age estimation," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2015, pp. 102–108.

[214] S. Gutta, J. R. Huang, P. Jonathon, and H. Wechsler, "Mixture of experts for classification of gender, ethnic origin, and pose of human faces," *IEEE Trans. Neural Netw.*, vol. 11, no. 4, pp. 948–960, Jul. 2000.

[215] The FG-NET aging database," 2002. [Online]. Available: http://www-prima.inrialpes.fr/FGnet/html/benchmarks.html

[216] R. Rother, R. Timofte, and L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," *Int. J. Comput. Vis.*, pp. 1–14, 2016.

[217] S. Escalera, et al., "ChaLearn looking at people 2015: Apparent age and cultural event recognition datasets and results," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2015, pp. 243–251.

[218] X. Liu, et al., "AgeNet: Deeply learned regressor and classifier for robust apparent age estimation," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2015, pp. 258–266.

[219] K. Zhang, L. Tan, Z. Li, and Y. Qiao, "Gender and smile classification using deep convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2016, pp. 34–38.

[220] Y. Zhu, Y. Li, G. Mu, and G. Guo, "A study on apparent age estimation," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2015, pp. 267–273.

[221] K. Luu, K. Ricanek, T. D. Bui, and C. Y. Suen, "Age estimation using active appearance models and support vector machine regression," in *Proc. IEEE Int. Conf. Biometrics Theory Appl. Syst.*, 2009, pp. 1–5.

[222] P. Thukral, K. Mitra, and R. Chellappa, "A hierarchical approach for human age estimation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2012, pp. 1529–1532.

[223] X. Wang, R. Guo, and C. Kambhamettu, "Deeply-learned feature for age estimation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2015, pp. 534–541.

[224] X. Wang and C. Kambhamettu, "Age estimation via unsupervised neural networks," in *Proc. IEEE Int. Conf. Automat. Face Gesture Recognit.*, 2015, vol. 1, pp. 1–6.

[225] S. Escalera, et al., "ChaLearn looking at people and faces of the world: Face analysis workshop and challenge 2016, " in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2016, pp. 706–713.

[226] G. Antipov, M. Baccouche, S.-A. Berrani, and J.-L. Dugelay, "Apparent age estimation from face images combining general and children-specialized deep learning models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2016, pp. 801–809.

[227] F. Lingenfelser, J. Wagner, T. Vogt, J. Kim, and E. André, "Age and gender classification from speech using decision level fusion and ensemble based techniques," in *Proc. INTERSPEECH*, 2010, pp. 2798–2801.

[228] T. Bocklet, G. Stemmer, V. Zeissler, and E. Nöth, "Age and gender recognition based on multiple systems-early versus late fusion," in *Proc. INTERSPEECH*, 2010, pp. 2830–2833.

[229] M. Li, C. S. Jung, and K. J. Han, "Combining five acoustic level modeling methods for automatic speaker age and gender recognition," in *Proc. INTERSPEECH*, 2010, pp. 2826–2829.
[230] M. Kockmann, L. Burget, and J. Černocký, "Brno University of technology system for interspeech 2010 paralinguistic challenge," in *Proc. INTERSPEECH*, 2010, pp. 2822–2825.
[231] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The HumanID gait challenge problem: Data sets, performance, and analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 162–177, Feb. 2005.
[232] J. Lu and Y. P. Tan, "Ordinary preserving manifold analysis for human age and head pose estimation," *IEEE Trans. Human-Mach. Syst.*, vol. 43, no. 2, pp. 249–258, Mar. 2013.
[233] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio, "Pedestrian detection using wavelet templates," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1997, pp. 193–199.
[234] D. Burnsides, M. Boehmer, and K. Robinette, "3-D landmark detection and identification in the CAESAR project," in *Proc. IEEE Int. Conf. 3-D Digit. Imag. Model.*, 2001, pp. 393–398.
[235] J. Tapia, C. Perez, and K. Bowyer, "Gender classification from the same iris code used for recognition," *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 8, pp. 1760–1770, Aug. 2016.
[236] J. R. Lyle, P. E. Miller, S. J. Pundlik, and D. L. Woodard, "Soft biometric classification using periocular region features," in *Proc. IEEE Int. Conf. Biometrics Theory, Appl. Syst.*, 2010, pp. 1–7.
[237] P. J. Phillips, et al., "Overview of the face recognition grand challenge," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 947–954.
[238] P. Yan and K. Bowyer, "Empirical evaluation of advanced ear biometrics," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2005, pp. 41–41.
[239] A. Gallagher and T. Chen, "Understanding images of groups of people," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 256–263.
[240] C. Chen, A. Dantcheva, and A. Ross, "Impact of facial cosmetics on automatic gender and age estimation algorithms," in *Proc. IEEE Int. Conf. Comput. Vis. Theory Appl.*, 2014, vol. 2, pp. 182–190.
[241] G. Guo and G. Mu, "Human age estimation: What is the influence across race and gender?" in *Proc. IEEE Int. Workshop Anal. Model. Faces Gestures*, 2010, pp. 71–78.
[242] G. Guo and G. Mu, "A study of large-scale ethnicity estimation with gender and age variations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2010, pp. 79–86.
[243] Y. Wang, H. Liao, Y. Feng, X. Xu, and J. Luo, "Do they all look the same? deciphering Chines, Japanese, and Koreans by fine-grained deep learning," *CoRR*, vol. 1610.01854, 2016.
[244] G. Farinella and J. L. Dugelay, "Demographic classification: Do gender and ethnicity affect each other?" in *Proc. IEEE Int. Conf. Inf. Electron. Vis.*, 2012, pp. 383–390.
[245] J. Bekios-Calfa, J. M. Buenaposada, and L. Baumela, "Robust gender recognition by exploiting facial attributes dependencies," *Pattern Recognit. Lett.*, vol. 36, pp. 228–234, 2014.
[246] G. Guo and G. Mu, "A framework for joint estimation of age, gender and ethnicity on a large database," *Image Vis. Comput.*, vol. 32, no. 10, pp. 761–770, 2014.
[247] C. Zhong, Z. Sun, and T. Tan, "Fuzzy 3D face ethnicity categorization," in *Proc. IAPR Int. Conf. Biometrics*, 2009, pp. 386–393.
[248] K. Xu, et al., "Show, attend and tell: Neural image caption generation with visual attention," in *Proc. Int. Conf. Mach. Learn.*, vol. 14, pp. 77–81, 2015.
[249] F. Juefei-Xu, E. Verma, P. Goel, A. Cherodian, and M. Savvides, "DeepGender: Occlusion and low resolution robust facial gender classification via progressively trained convolutional neural networks with attention," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2016, pp. 68–77.
[250] N. Cherniavsky, I. Laptev, J. Sivic, and A. Zisserman, "Semi-supervised learning of facial attributes in video," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 43–56.
[251] A. C. Gallagher and T. Chen, "Estimating age, gender, and identity using first name priors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.
[252] Z. Song, B. Ni, D. Guo, T. Sim, and S. Yan, "Learning universal multi-view age estimator using video context," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2011, pp. 241–248.
[253] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell, "Distance metric learning with application to clustering with side-information," in *Proc. Advances Neural Inf. Process. Syst.*, 2003, vol. 15, pp. 505–512.
[254] L. Walavalkar, M. Yeasin, A. Narasimhamurthy, and R. Sharma, "Support vector learning for gender classification using audio and visual cues," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 17, no. 3, pp. 417–439, 2003.

Yunlian Sun received the ME degree in computer science and technology from the Harbin Institute of Technology, China, in 2010 and the PhD degree in ingegneria elettronica, informatica e delle telecomunicazioni from the University of Bologna, Italy, in 2014. She is currently a postdoctoral researcher in the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China. Her research interests include biometrics, pattern recognition, and computer vision.

Man Zhang received the BE degree in information engineering from Beijing University of Posts and Telecommunications, China and the PhD degree in computer application technology from CASIA, in 2008, and 2013, respectively. She is currently an assistant professor in the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China. Her research interests include biometrics, pattern recognition, and image processing.

Zhenan Sun received the BE degree in industrial automation from Dalian University of Technology, China, the MS degree in system engineering from Huazhong University of Science and Technology, China, and the PhD degree in pattern recognition and intelligent systems from CASIA, in 1999, 2002, and 2006, respectively. He is currently a professor in the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China. His research interests include biometrics, pattern recognition, and computer vision. He is a fellow of the IAPR.

Tieniu Tan received the BSc degree in electronic engineering from Xian Jiaotong University, China, in 1984, and the MSc and PhD degrees in electronic engineering from Imperial College London, United Kingdom, in 1986 and 1989, respectively. He is currently a professor in the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China. His research interests include biometrics, image and video understanding, information hiding, and information forensics. He is a fellow of the CAS, the TWAS, the BAS, the IEEE, the IAPR, the UK Royal Academy of Engineering, and the Past President of IEEE Biometrics Council.