Demographic Bias in Biometrics: A Survey on an Emerging Challenge

Pawel Drozdowski, Christian Rathgeb, Antitza Dantcheva, Naser Damer,
Christoph Busch

Demographic Bias in Biometrics:
A Survey on an Emerging Challenge

Pawel Drozdowski[iD], Christian Rathgeb[iD], Antitza Dantcheva, Naser Damer[iD], and Christoph Busch[iD]

Abstract—**Systems incorporating biometric technologies have become ubiquitous in personal, commercial, and governmental identity management applications. Both cooperative (e.g., access control) and noncooperative (e.g., surveillance and forensics) systems have benefited from biometrics. Such systems rely on the uniqueness of certain biological or behavioral characteristics of human beings, which enable for individuals to be reliably recognized using automated algorithms. Recently, however, there has been a wave of public and academic concerns regarding the existence of systemic bias in automated decision systems (including biometrics). Most prominently, face recognition algorithms have often been labeled as "racist" or "biased" by the media, nongovernmental organizations, and researchers alike. The main contributions of this article are: 1) an overview of the topic of algorithmic bias in the context of biometrics; 2) a comprehensive survey of the existing literature on biometric bias estimation and mitigation; 3) a discussion of the pertinent technical and social matters; and 4) an outline of the remaining challenges and future work items, both from technological and social points of view.**

Index Terms—**Bias, bias estimation, bias mitigation, biometrics, demographics, fairness.**

I. INTRODUCTION

ARTIFICIAL intelligence (AI) systems increasingly support humans in complex decision-making tasks. Domains of interest include learning, problem solving, classifying, as well as making predictions and risk assessments. Automated algorithms have in many cases already outperformed humans and hence are used to support or replace human operators [1]. Those systems, referred to as "automated decision systems," can yield various benefits, e.g., increased efficiency and decreased monetary costs. At the same time, a number of ethical and legal concerns have been raised, specifically relating to *transparency, accountability, explainability, and*

fairness of such systems [2]. Automated algorithms can be utilized in diverse critical areas, such as criminal justice [3], healthcare [4], creditworthiness [5], and others [6], hence often sparking controversial discussions. This article focuses on *algorithmic bias and fairness in biometric systems w.r.t. demographic attributes*. In this context, an algorithm is considered to be biased if significant differences in its operation can be observed for different demographic groups of individuals (e.g., females or dark-skinned people), thereby privileging and disadvantaging certain groups of individuals.

A. Motivation

The interest and investment into biometric technologies are large and rapidly growing according to various market value studies [7]–[9]. Biometrics are utilized widely by governmental and commercial organizations around the world for purposes, such as border control, law enforcement and forensic investigations, voter registration for elections, as well as national identity management systems. Currently, the largest biometric system is operated by the Unique Identification Authority of India, whose national ID system (Aadhaar) accommodates almost the entire Indian population of 125 billion enrolled subjects at the time of this writing, see the online dashboard [10] for live data.

In recent years, reports of demographically unfair/biased biometric systems have emerged (see Section III), fueling a debate on the use, ethics, and limitations of related technologies between various stakeholders, such as the general population, consumer advocates, nongovernmental and governmental organizations, academic researchers, and commercial vendors. Such discussions are intense and have even raised demands and considerations that biometric applications should be discontinued in operation, until sufficient privacy protection and demographic bias mitigation can be achieved.[1,2,3,4] Algorithmic bias is considered to be one of the important open challenges in biometrics by Ross *et al.* [11].

B. Article Contribution and Organization

In this article, an overview of the emerging challenge of algorithmic bias and fairness in the context of biometric

Pawel Drozdowski, Christian Rathgeb, and Christoph Busch are with the da/sec—Biometrics and Internet Security Research Group, Hochschule Darmstadt, 64295 Darmstadt, Germany (e-mail: pawel.drozdowski@h-da.de).

Antitza Dantcheva is with Inria, 06904 Sophia Antipolis, France.

Naser Damer is with the Fraunhofer Institute for Computer Graphics Research IGD, 64283 Darmstadt, Germany, and also with the Mathematical and Applied Visual Computing, TU Darmstadt, 64289 Darmstadt, Germany.

[1]https://www.banfacialrecognition.com/

[2]https://www.cnet.com/news/facial-recognition-could-be-temporarily-banned-for-law-enforcement-use/

[3]https://www.theguardian.com/technology/2020/jan/17/eu-eyes-temporary-ban-on-facial-recognition-in-public-places

[4]https://www.biometricupdate.com/202001/eu-no-longer-considering-facial-recognition-ban-in-public-spaces

systems is presented. Accordingly, the biometric algorithms which might be susceptible to bias are summarized; furthermore, the existing approaches of bias estimation and bias mitigation are surveyed. This article additionally discusses other pertinent matters, including the potential social impact of bias in biometric systems, as well as the remaining challenges and open issues in this area.

The remainder of this article is organized as follows. The relevant background information is provided in Section II. Section III contains a comprehensive survey of the scientific literature on bias estimation and mitigation in biometric systems. Other relevant matters are discussed in Section IV, while concluding remarks and a summary are presented in Section V.

II. BACKGROUND

The following sections provide relevant background information with respect to (w.r.t.) the topic of bias in automated decision systems in general (Section II-A) and the basics of biometric systems (Section II-B). Furthermore, due to the sensitive nature of the matter at hand, Section II-C outlines the choices made w.r.t. the nomenclature used throughout this article.

A. Bias in Automated Decision Systems

In recent years, numerous concerns have been raised regarding the accuracy and fairness of automated decision-making systems. For instance, many studies regarding the risk assessment and welfare distribution tools found a number of issues concerning systemic bias and discrimination of the systems' predictions (e.g., against dark-skinned people). The impact of such automated decisions on the lives of the affected individuals can be tremendous, e.g., being jailed, denied a bail, parole, or welfare payments [2], [3], [12], [13]. Demographics-based bias and discrimination are especially concerning in this context, even if they occur unintentionally. One would intuitively expect that certain decisions be impacted exclusively by hard facts and evidence, and not factors often associated with discrimination—such as sex or race, or other context-specific discriminatory factors. Nonetheless, biases in decision making are a common occurrence; along with the notions of fairness, this topic has been extensively studied from the point of view of various disciplines, such as psychology, sociology, statistics, and information theory [14]–[16]. Recently, the field of bias and fairness in automated computer algorithms and machine learning has emerged [17], [18].

A good discussion of the topic of bias was provided by Danks and London [19], as well as Friedman and Nissenbaum [20], both of which explored various sources and types of bias in the context of computer systems. In many cases, the bias in the automated decision systems is directly related to the human designers or operators of a system. Semiautomatic decision systems are a good example of this. In such systems, a human decision maker can be aided by an algorithm (e.g., risk assessment). In such cases, errors in interpretation of the results of the system might occur; in other words, the human might misunderstand or misrepresent
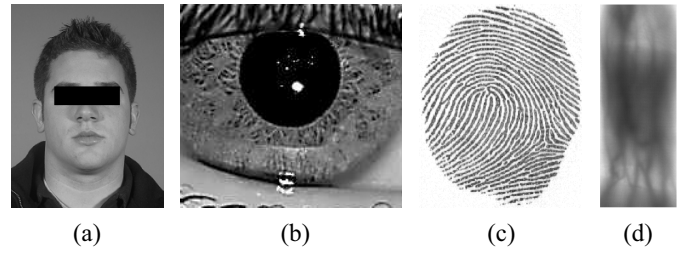

Fig. 1. Examples of biometric characteristics (images from publicly available research databases [27]–[30]). (a) Face. (b) Iris. (c) Fingerprint. (d) Veins.

the outputs or general functioning principles of an algorithm [21]–[23]. Furthermore, it has been shown that humans, in general, tend to over-rely on such automated systems, i.e., to overestimate the accuracy of their results [24]. While *human* cognitive biases are an important and actively researched topic, this article focuses exclusively on bias occurring in the context of *automated* algorithms themselves. Human cognitive biases have been analyzed, e.g., by Evans [14], whereas bias in human interactions with the automated system was explored, e.g., by Parasuraman and Manzey [25].

In the context of automated decision algorithms themselves, numerous potential bias causes exist. Most prominently, the *training data* could be skewed, incomplete, outdated, disproportionate, or have embedded historical biases, all of which are detrimental to algorithm training and propagate the biases present in the data. Likewise, the *implementation* of an algorithm itself could be statistically biased or otherwise, flawed in some way, for example, due to moral or legal norms, poor design, or data processing steps, such as parameter regularization or smoothing. For more details on the topic of algorithmic bias in general, the reader is referred to, e.g., [6], [19], and [20]. In the next sections, an introduction to biometric systems is provided, followed by a survey on algorithmic bias in such systems specifically.

B. Biometric Systems

Biometric systems aim at establishing or verifying the identity or demographic attributes of individuals. In the international standard ISO/IEC 2382-37 [26], "biometrics" is defined as: "automated recognition of individuals based on their biological and behavioral characteristics."

Humans possess, nearly universally, physiological characteristics which are highly distinctive and can, therefore, be used to distinguish between different individuals with a high degree of confidence. Example images of several prominent biometric characteristics are shown in Fig. 1.

Broadly speaking, an automated biometric system consists of: 1) a capture device (e.g., a camera), with which the biometric samples (e.g., images) are acquired; 2) a database which stores the biometric information and other personal data; 3) signal processing algorithms, which estimate the quality of the acquired sample, find the region of interest (e.g., a face) and extract the distinguishing features from it; and 4) comparison and decision algorithms, which enable ascertaining of similarity of two biometric samples by comparing

the extracted feature vectors and establishing whether or not the two biometric samples belong to the same source.

In the past, biometric systems typically utilized hand-crafted features and algorithms (i.e., texture descriptors, see Liu *et al.* [31]). Nowadays, the use of machine learning and deep learning has become increasingly popular and successful. Relevant-related works include [32]–[34], which achieved breakthrough biometric performances in facial recognition. Furthermore, promising results for deep learning-based fingerprint (see [35]) and iris (see [36]) recognition have also been achieved. For a review of deep learning techniques applied within biometrics, the reader is referred to as Sundararajan and Woodard [37]. For a highly comprehensive introduction to biometrics, the reader is referred to Li and Jain [38] and the handbook series [39]–[43].

C. Nomenclature

In this section, the nomenclature used throughout this article is explained. The authors note that demographic words, groups, and concepts, such as "gender," "sex," "race," and "ethnicity" can be extremely divisive and bear a heavy historical, cultural, social, political, or legislative load. The authors do not seek to define or redefine those terms; we merely report on the current state of the research. In the literature surveyed later on in this article, the following trends can be distinguished.

1) The terms gender and sex are often used in a binary and conflated manner. Readers interested in the possible consequences of this narrow approach are referred to [44].
2) Similarly, very often no real distinction between the terms race and ethnicity is made; moreover, the typical categorization is very coarse, only allowing for a small and finite (less than ten) possible racial/ethnic categories.
3) In general, and especially in the case of facial biometrics, demographic factors seem to be considered on the phenotypic basis, i.e., concerning the observable traits of the subjects (e.g., the color of the skin or masculine appearance).

Due to the demographic terms carrying a large amount of complexity and potential social divisiveness, the authors do not engage in those debates in this article, and merely reproduce and discuss the technical aspects of the current research. For the sake of consistency, certain decisions regarding the used nomenclature have to be made, especially since the surveyed literature does often seem to use the aforementioned demographic terms ambiguously or interchangeably.

Recently, in the context of biometrics, ISO/IEC has made the following separation [45][5]: while the term gender is defined as "the state of being male or female as it relates to social, cultural, or behavioral factors," the term sex is understood as "the state of being male or female as it relates to biological factors, such as DNA, anatomy, and physiology." The report also defines the term ethnicity as "the state of belonging to a group with a common origin, set of customs, or traditions," while the term race is not defined there. While

[5]Note that the document is currently in a draft stage.

the cultural and religious norms can certainly affect biometric operations, the surveyed literature mostly considers the appearance-based features and categorization—hence, the term race is used instead of ethnicity and the term sex is used instead of gender in accordance with ISO/IEC 22116 [45]. In the context of biometrics in general, the standardized biometric vocabulary is used (see ISO/IEC 2382-37 [26]). Finally, it is noted that a large part of the surveyed biometric literature follows the notions and metrics regarding the evaluation of biometric algorithms irrespective of the chosen biometric characteristic defined in ISO/IEC 19795-1 [46].

Those limitations and imprecisions of the nomenclature notwithstanding, due to the potential of real and disparate impacts [47] of automated decision systems including biometrics, it is imperative to study the bias and fairness of such algorithms w.r.t. the demographic attributes of the population, regardless of their precise definitions.

III. BIAS IN BIOMETRIC SYSTEMS

To facilitate discussions on algorithmic fairness in biometric systems, Howard *et al.* [48] introduced the following two terms.

1) *Differential performance* concerns the differences in (genuine and/or impostor) score distributions between the demographic groups. Those effects are closely related to the so-called "biometric menagerie" [49]–[51]. While the menagerie describes the score distributions being statistically different for specific individual subjects, the introduced term describes the analogous effect for different demographic groups of subjects.
2) *Differential outcomes* relate to the decision results of the biometric system, i.e., the differences in the false-match and false-nonmatch rates at a specific decision threshold.

Given that these terms have been introduced relatively recently, the vast majority of surveyed literature have not (directly) used them, instead *ad hoc* methodologies based on existing metrics were used. However, Grother *et al.* [52] presented a highly comprehensive study of the demographic effects in biometric recognition, conducting their benchmark utilizing the terms and notions above. A standardization effort in this area under the auspices of ISO/IEC is ongoing [45].

Before surveying the literature on bias estimation and mitigation (Sections III-C and III-D, respectively), this section begins with an outline of biometric algorithms which might be affected by bias (Section III-A), as well as of covariates which might affect them (Section III-B).

A. Algorithms

Similar to other automated decision systems, human biases have been shown to exist in the context of biometrics. The so-called "other-race effect" has long been known to affect the human ability to recognize faces [53]. As previously stated, the cognitive biases of humans are out of scope for this article, as it focuses on the biases in the algorithms themselves. The processing pipeline of a biometric system can consist of various algorithms depending on the application scenario and the chosen biometric characteristic. Said algorithms might be
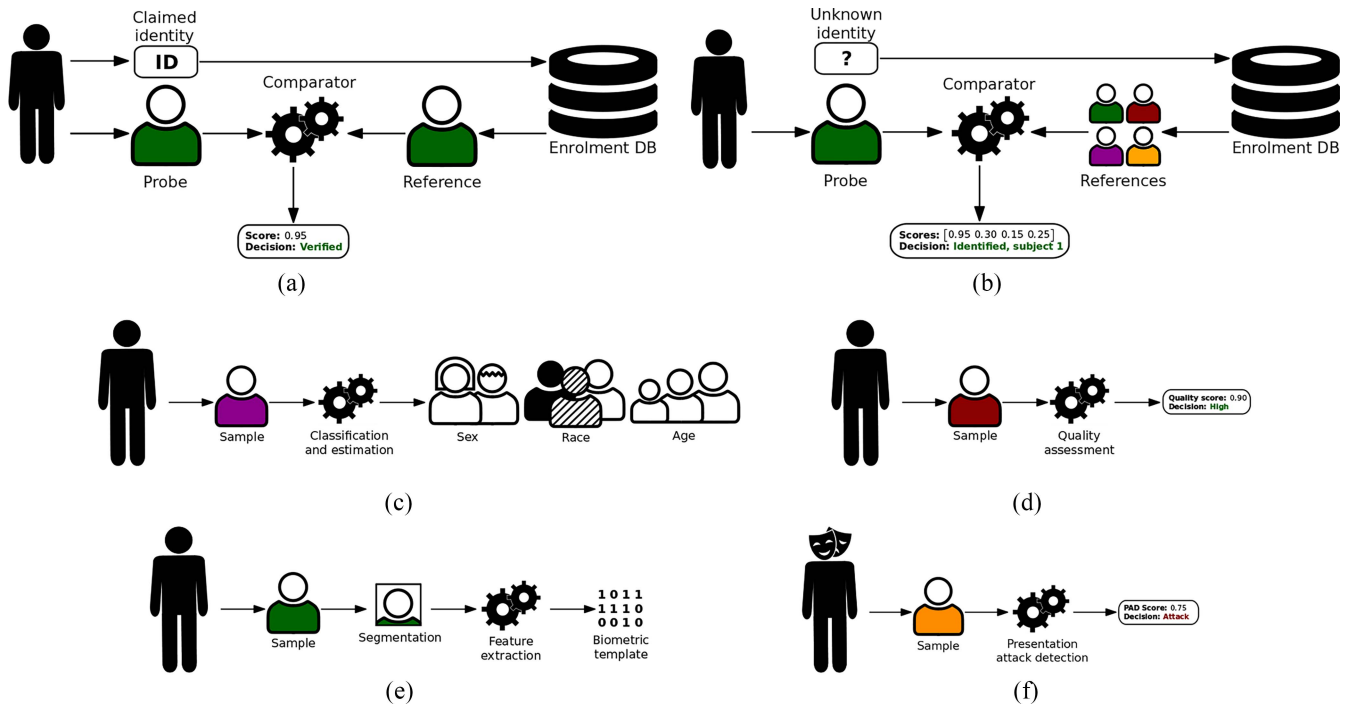
Fig. 2.   Conceptual overview of algorithms used in biometric systems. (a) Verification. (b) Identification. (c) Classification and estimation. (d) Quality assessment. (e) Segmentation and feature extraction. (f) PAD.

subject to algorithmic bias w.r.t. certain covariates, which are described in Section III-B. Below, the most important algorithms used in the context of biometrics are described and visualized conceptually in Fig. 2.

One of the most prevalent uses of biometrics is recognition. Here, distinguishing features of biometric samples are compared to ascertain their similarity. Such systems typically seek to: 1) determine if an individual is who they claim to be (i.e., one-to-one comparison) or 2) to determine the identity of an individual by searching a database (i.e., one-to-many search). Accordingly, the following two scenarios might be used in biometric recognition.

1) *Verification:* Referring to the "process of confirming a biometric claim through biometric comparison" [26], [46].
2) *Identification:* Referring to the "process of searching against a biometric enrolment database to find and return the biometric reference identifier(s) attributable to a single individual" [26], [46].

The biometric samples are a rich source of information beyond the mere identity of the data subject. Another use case of biometrics is the extraction of auxiliary information from a biometric sample, primarily using the following algorithms.

1) *Classification and Estimation:* Referring to the process of assigning demographic or other labels to biometric samples [54], [55]. Prior to recognition or classification tasks, the system must acquire and preprocess the biometric sample(s). Here, most prominently, the following algorithms might be used.
2) *Segmentation and Feature Extraction:* Referring to the process of locating the region of interest and extracting a set of biometric features from a biometric sample [38].

3) *Quality Assessment:* Referring to the process of quantifying the quality of an acquired biometric sample [56], [57].
4) *Presentation Attack Detection (PAD):* Referring to the "automated determination of a presentation attack," i.e., detecting a "presentation to the biometric data capture subsystem with the goal of interfering with the operation of the biometric system" [58], [59].

B. Covariates

Broadly, three categories of covariates relevant to the effectiveness of the biometric algorithms can be distinguished.

1) *Demographic:* Referring to, e.g., the sex, age, and race of the data subject.
2) *Subject Specific:* Referring to the behavior of the subject (e.g., pose or expression, and use of accessories, such as eyewear or makeup), as well as their interaction with the capture device (e.g., distance from a camera or pressure applied to a touch-based sensor).
3) *Environmental:* Referring to the effects of the surroundings on the data acquisition process (e.g., illumination, occlusion, and resolution of the images captured by the sensor).

Fig. 3 shows the example images of the aforementioned covariates using the facial biometric characteristic. While there do exist studies that investigate environmental and subject-specific covariates (e.g., [60]), this article concentrates on the demographic covariates.

C. Estimation

Table I summarizes the existing research in the area of bias estimation in biometrics. The table is organized conceptually
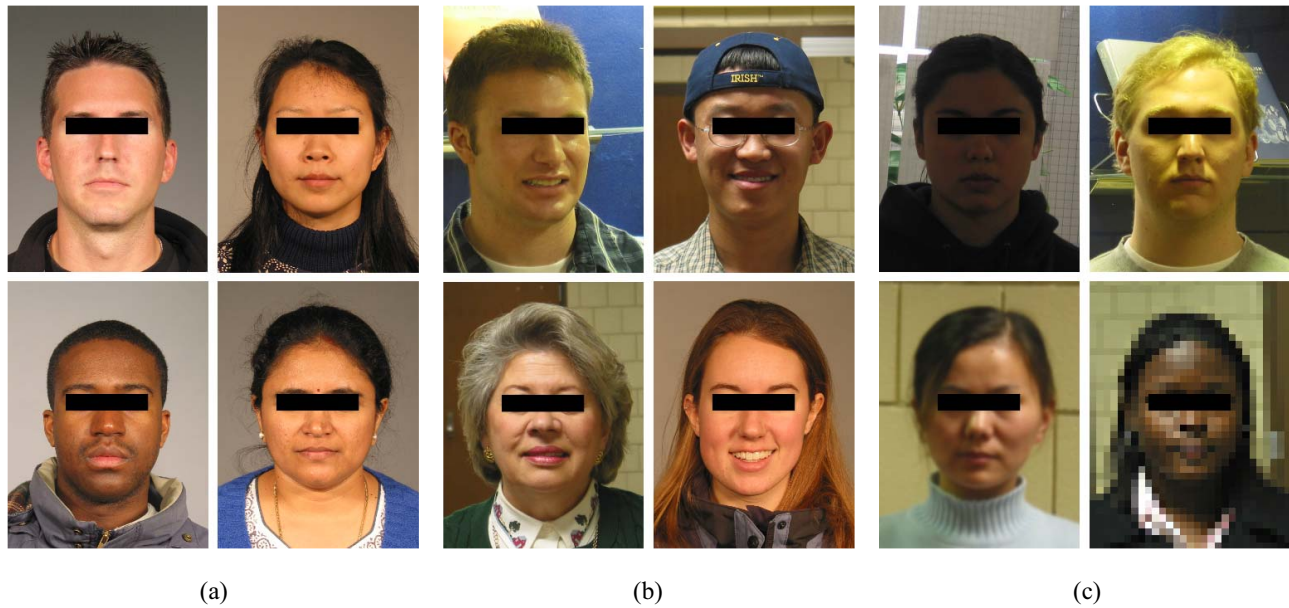
(a) (b) (c)

Fig. 3. Example images of covariates which might influence a biometric system utilizing facial information (images from a publicly available research database [27]). Black rectangles were added in an effort to respect individual anonymity and privacy. (a) Demographic (different sex, age, and race). (b) Subject specific (different pose and expression, and use of makeup and accessories). (c) Environmental (different lighting conditions, sharpness, and resolution).

as follows: the studies are divided by biometric characteristic and listed chronologically. The third column lists the algorithms (recall Section III-A) evaluated by the studies, while the covariates (recall Section III-B) considered in the studies are listed in the next column. Finally, the last column outlines the key finding(s) of the studies. Wherever possible, those were extracted directly from the abstract or summary sections of the respective studies.

By surveying the existing literature, the following trends can be distinguished.

1) Most existing studies conducted the experiments using *face*-based biometrics. There are significantly fewer studies on other modalities (primarily fingerprint).

2) The majority of studies concentrated on biometric *recognition* algorithms (primarily verification), followed by quality assessment and classification algorithms.

3) Some scenarios have barely been investigated, e.g., PAD.

4) The existing studies predominantly considered the *sex covariate*; the race covariate is also often addressed (possibly due to the recent press coverage [134], [135]). The age covariate is least often considered in the context of bias in the surveyed literature. The impact of aging on biometric recognition is an active field of research, but out of scope for this article. The interested reader is referred to, e.g., [73], [106], and [136]–[139].

5) Many studies focused on general *accuracy* rather than distinguishing between false-positive and false-negative errors. Recent works [48], [52] introduced and used the useful concepts of "false-positive differentials" and "false-negative differentials" in demographic bias benchmarks.

6) A significant number of studies (e.g., [48], [52], and [82]) conducted evaluations on *sequestered databases and/or commercial systems*. Especially the

results of Grother *et al.* [52] in the context of an evaluation conducted by the National Institute of Standards and Technology (NIST) were valuable due to the realistic/operational nature of the data, the large scale of used databases, as well as the testing of state-of-the-art commercial and academic algorithms. However, reproducing or analyzing their results may be impossible due to the unattainability of data and/or tested systems.

The following common findings for the evaluated biometric algorithms can be discerned.

1) *Recognition:* One result which appears to be mostly consistent across surveyed studies is that of worse biometric performance (both in terms of false positives and false negatives) for female subjects (see [52], [67]). Furthermore, several studies associated race as a major factor influencing biometric performance. However, the results were not attributed to a specific race being inherently more challenging. Rather, the country of software development (and presumably the training data) appears to play a major role; in this context, the evidence of the other-race effect in facial recognition has been found [65], e.g., algorithms developed in Asia were more easily recognizing Asian individuals and conversely algorithms developed in Europe were found to be more easily recognizing Caucasians. Finally, the age has been determined to be an important factor as well—especially the very young subjects were a challenge (with effects of aging also playing a major role). Grother *et al.* [52] presented hitherto the largest and most comprehensive study of demographic bias in biometric recognition. Their benchmark showed that false-negative differentials usually vary by a factor of less than 3 across the benchmarked algorithms. On the other hand, the

TABLE I
SUMMARY OF STUDIES CONCERNING BIAS ESTIMATION IN BIOMETRIC SYSTEMS

Reference	Characteristic	Algorithm(s)	Covariate(s)	Key Findings
Beveridge *et al.* [61]	Face	Verification	Sex, age, race	Better biometric performance for older subjects, males, and East Asians.
Lui *et al.* [62]	Face	Verification	Sex, age, race	Meta-analysis of previous studies.
Guo *et al.* [63]	Face	Age estimation	Sex, race	Large impact of the training data composition on the system accuracy.
Grother *et al.* [64]	Face	Verification	Sex	More false-non-matches at fixed FMR for females than for males.
Phillips *et al.* [65]	Face	Verification	Race	Varying results depending on the demographic origin of the algorithm and demographic structure of the data subjects.
O'Toole *et al.* [66]	Face	Verification	Sex, race	The concept of "yoking" in experimental evaluation to demonstrate the variability of algorithm performance estimates.
Klare *et al.* [67]	Face	Verification	Sex, age, race	Lower biometric performance for females, young, and black cohorts.
Givens *et al.* [68]	Face	Verification	Sex, age, race	Better biometric performance for Asian and older subjects.
Beveridge *et al.* [69]	Face	Verification	Sex, race	Better biometric performance for males and Asian subjects.
Ricanek *et al.* [70]	Face	Verification	Age	Poor biometric performance for children.
El Khiyari *et al.* [71]	Face	Verification	Sex, age, race	Lower biometric performance for female, 18-30 age group, and dark-skinned subjects.
Deb *et al.* [72]	Face	Verification	Sex, race	Algorithm dependent effects of the covariates.
Best-Rowden *et al.* [73]	Face	Verification	Sex, age, race	Lower comparison scores for females.
Buolamwini *et al.* [74]	Face	Sex and race classification	Race	Highest accuracy for males and light-skinned individuals; worst accuracy for dark-skinned females.
Deb *et al.* [75]	Face	Verification, identification	Age	Child females easier to recognise than child males.
Michalski *et al.* [76]	Face	Verification	Age	Large variation of biometric performance across age and ageing factors in children. Poor biometric performance for very young subjects.
Abdurrahim *et al.* [77]	Face	Verification	Sex, age, race	Lower biometric performance for females, inconsistent results w.r.t. age and race.
Rhue *et al.* [78]	Face	Emotion classification	Race	Negative emotions more likely to be assigned to dark-skinned males.
Lu *et al.* [79]	Face	Verification	Sex, age, race	Lower biometric performance for females; better biometric performance for middle-aged.
Raji *et al.* [80]	Face	Sex and race classification	Sex, race	Lower accuracy for dark-skinned females.
Srinivas *et al.* [81]	Face	Verification, identification	Sex, age	Lower biometric performance for females and children.
Cook *et al.* [82]	Face	Verification	Sex, age, race	Genuine scores tend to be worse for females than males.
Hupont *et al.* [83]	Face	Verification	Sex, race	Highest biometric performance for white males, lowest for Asian females.
Denton *et al.* [84]	Face	Classification	CelebA attributes	Generative adversarial model which can reveal biases in a face attribute classifier.
Garcia *et al.* [85]	Face	Verification, presentation attack detection	Sex, race	Higher inter-subject distance for Caucasians than other groups; morphing attacks more successful for Asian females.
Nagpal *et al.* [86]	Face	Verification	Age, race	Training data dependent own-age and own-race effect in deep learning-based systems.
Krishnapriya *et al.* [87]	Face	Quality, verification	Race	Lower rate of ICAO compliance [88] for the dark-skinned cohort, fixed decision thresholds not suitable for cross-cohort biometric performance benchmark.
Muthukumar [89]	Face	Sex classification	Race	Lower accuracy for dark females; importance of not only skin type, but also luminance in the images on the results.
Srinivas *et al.* [90]	Face	Verification, identification	Age	Lower biometric performance for children.
Vera-Rodriguez *et al.* [91]	Face	Verification	Sex	Lower biometric performance for females.
Howard *et al.* [48]	Face	Verification	Sex, age, race	Evaluates effects of population homogeneity on biometric performance.
Wang *et al.* [92]	Face	Verification	Race	Higher biometric performance for Caucasians.
Serna *et al.* [93]	Face	Verification	Sex, race	Better biometric performance for male Caucasians; large impact of the training data composition on the system accuracy.
Cavazos *et al.* [94]	Face	Verification	Sex, race	Higher false match rate for Asians compared to Caucasians at operationally relevant fixed decision thresholds; data-driven anomalies might contribute to system bias.
Grother *et al.* [52]	Face	Verification, identification	Sex, age, race	Large-scale benchmark of commercial algorithms. Algorithm dependent false positive differentials w.r.t. race. Consistently elevated false positives for female, elderly and very young subjects. Algorithm specific false negative differentials, also correlated with image quality.
Robinson *et al.* [95]	Face	Verification	Sex, race	Highest biometric performance for males and Caucasians.
Albiero *et al.* [96]	Face	Verification	Sex	Lower biometric performance for females. Negative impact of facial cosmetics on (female) genuine scores distribution. Minor impact of expression, pose, hair occlusion, and imbalanced datasets on bias.
Krishnapriya *et al.* [97]	Face	Verification, identification	Race	Lower biometric performance for females, higher false match rate for African-Americans, and higher false non-match rate for Caucasians at fixed, operationally relevant decision threshold.
Terhörst *et al.* [98]	Face	Quality	Age, race	Bias in quality scores for demographic and non-demographic characteristics is significant. Bias is transferred from face recognition to face image quality.
Hicklin *et al.* [99]	Fingerprint	Quality	Sex	Lower sample quality for females.
Sickler *et al.* [100]	Fingerprint	Quality	Age	Lower sample quality for the elderly.
Modi *et al.* [101]	Fingerprint	Quality, verification	Age	Lower sample quality and biometric performance for the elderly.
Modi *et al.* [102]	Fingerprint	Quality, verification	Age	Lower sample quality and biometric performance for the elderly.
Frick *et al.* [103]	Fingerprint	Quality, verification	Sex	Higher sample quality and biometric performance for males.
O'Connor *et al.* [104]	Fingerprint	Quality, verification	Sex	Higher sample quality for males, higher biometric performance for females.
Schumacher *et al.* [105]	Fingerprint	Quality, verification	Age	Lower sample quality and biometric performance for children.
Yoon *et al.* [106]	Fingerprint	Quality, verification	Sex, age, race	Negligible correlations between sample quality and subject age; sex and race have a marginal impact on comparison scores, whereas subject's age has a non-trivial impact for genuine scores.
Galbally *et al.* [107], [108]	Fingerprint	Quality, verification	Age	On average, low quality for children under 4 years and elderly (70+ years), medium quality for children between 4 and 12 years. Lowest biometric performance in youngest children, then elderly.
Preciozzi *et al.* [109]	Fingerprint	Quality, verification	Age	Lower sample quality and biometric performance for young children.
Drozdowski *et al.* [110]	Fingervein	Verification	Sex, age	No statistically significant biases detected.
Fang *et al.* [111]	Iris	Presentation attack detection	Sex	Better PAD rates for males. Maps differential performance/outcome concepts to PAD.
Xie *et al.* [112]	Palmprint	Sex classification	Sex	Higher accuracy for females.
Uhl *et al.* [113]	Palmprint	Verification	Age	Lower biometric performance for very young subjects.
Brandão *et al.* [114]	Unconstrained	Pedestrian detection	Sex, age	Higher miss rate for children.

false-positive differentials were much more prevalent (albeit not universal) and often larger, i.e., varying by two to three orders of magnitude across the benchmarked algorithms.[6] Most existing studies considered biometric verification, with only a few addressing biometric identification. Estimating bias in biometric identification is nontrivial, due to the contents of the screening database being an additional variable factor susceptible to bias.

[6]Note that this is a very high-level summary to illustrate the general size of the demographic differentials. The experimental results are much more nuanced and complex, as well as dependent on a number of factors in the used data, experimental setup, and the algorithms themselves.

Specifically, in addition to potential biases in the biometric algorithms themselves, certain biases stemming from data acquisition might occur and be propagated (e.g., historical and societal biases having impact on the demographic composition of a criminal database). Consequently, demographic bias estimation in biometric identification is an interesting and important item for future research.

2) *Classification and Estimation:* Scientific literature predominantly studied face as the biometric characteristic, since the facial region contains rich information from which demographic attributes can be estimated. Several studies showed a substantial impact of sex and race on

TABLE II
SUMMARY OF STUDIES CONCERNING BIAS MITIGATION IN BIOMETRIC SYSTEMS

Reference	Characteristic	Algorithm(s)	Method(s)
Guo *et al.* [63]	Face	Age classification	Dynamic classifier selection based on the demographic attributes.
Klare *et al.* [67]	Face	Verification, identification	Balanced training dataset or dynamic matcher selection based on the demographic attributes.
Guo *et al.* [115]	Face	Verification, identification	Imbalanced learning.
Ryu *et al.* [116]	Face	Sex and race classification	Twofold transfer learning, balanced training dataset.
Hasnat *et al.* [117]	Face	Verification	Imbalanced learning.
Deb *et al.* [75]	Face	Verification, identification	Training fine-tuning.
Michalski *et al.* [76]	Face	Verification	Dynamic decision threshold selection.
Alvi *et al.* [118]	Face	Sex, age, and race classification	Bias removal from neural network embeddings.
Das *et al.* [119]	Face	Sex, age, and race classification	Multi-task neural network with dynamic joint loss.
Acien *et al.* [120]	Face	Verification, identification	Suppression of deep learning features related to sex and race.
Amini *et al.* [121]	Face	Detection	Unsupervised learning, sampling probabilities adjustment.
Lu *et al.* [79]	Face	Verification	Curating training data (noisy label removal) using automatic sex estimation and clustering.
Terhörst *et al.* [122], [123]	Face	Sex and age classification	Suppression of demographic attributes.
Gong *et al.* [124]	Face	Verification; sex, age, and race classification	Disentangled representation for identity, sex, age, and race reduces bias for all estimations.
Kortylewski *et al.* [125]	Face	Verification	Synthetic data use in algorithm training.
Krishnapriya *et al.* [87]	Face	Verification	Cohort-dependent decision thresholds.
Srinivas *et al.* [90]	Face	Verification	Score-level fusion of algorithms.
Vera-Rodriguez *et al.* [91]	Face	Verification	Covariate-specific or covariate-balanced training.
Wang *et al.* [126]	Face	Verification	Reinforcement learning, balanced training datasets.
Robinson *et al.* [95]	Face	Verification, identification	Learning subgroup-specific thresholds mitigate the bias and boost overall performance.
Bruveris *et al.* [127]	Face	Verification	Weighted sampling and fine-grained labels.
Smith *et al.* [128]	Face	Sex and age classification	Data augmentation for model training.
Terhörst *et al.* [129]	Face	Verification	Individual fairness through fair score normalisation.
Terhörst *et al.* [130]	Face	Verification, identification	Comparison-level bias-mitigation by learning a fairness-driven similarity function.
Gottschlich *et al.* [131]	Fingerprint	Verification, identification	Modelling fingerprint growth and rescaling.
Preciozzi *et al.* [109]	Fingerprint	Quality, verification	Rescaling and bi-cubic interpolation as preprocessing.
Bekele *et al.* [132]	Unconstrained	Soft-biometric classification	Weighing to compensate for biases from imbalanced training dataset.
Wang *et al.* [133]	Unconstrained	Classification	Introduces concepts of dataset and model leakage; adversarial debiasing network.

the accuracy of demographic attribute classification. In particular, numerous commercial algorithms exhibited significantly lower accuracy w.r.t. dark-skinned female subjects (see [74], [80]). Research on the classification of sex from iris and periocular images exists, but biases in those algorithms have not yet been studied. Additionally, it is not clear if such classifiers rely on actual anatomical properties or merely the application of mascara [140].

3) *Quality Assessment:* Most existing studies conducted experiments using fingerprint-based biometrics. This could be partially caused by the standardization of reliable fingerprint quality assessment metrics [141], whereas this remains an open challenge for the face characteristic [142]. The existing fingerprint quality assessment studies consistently indicated that the extreme ranges of the age distribution (infants and elderly) can pose a challenge for current systems [108]. Correlations between the quality metrics of facial images (obtained using state-of-the-art estimators) and demographic covariates were recently pointed out in a preliminary study [98]. Additional nonobvious, hidden biases can also occur. For example, the presence of eyeglasses [143], [144] or contact lenses [145] lowers the sample quality and biometric performance under objective metrics in iris recognition systems. The demographics disproportionately afflicted with myopia (i.e., most likely to wear corrective eyewear) are those from "developed" countries and East Asia [146]. Admittedly, the inability of the algorithms to compensate for the presence of corrective eyewear might be argued not to be a bias *per se*. This argument notwithstanding, specific demographic groups could clearly be disadvantaged in this case—either by increased error rates or the requirement for a more elaborate (especially for contact lenses) interaction with the acquisition device. Issues, such as

this one push the boundaries of what might be considered biased or fair in the context of biometric systems and constitute an interesting area of future technical and philosophical research.

In addition, it is necessary to point out potential issues in surveyed studies, such as the following.

1) *Differences in experimental setups*, used toolchains and datasets, training–testing data partitioning, imbalanced datasets, etc.
2) Statistical significance of the results due to relatively *small size of the used datasets* in most cases (except, e.g., [52] and [107]).
3) *Lack* of a single *definition of bias/fairness* (see also Section IV-A), as well as a standardized methodology and metrics for conducting evaluations.
4) Difficulty of sufficiently *isolating the influence* of demographic factors from other important covariates (e.g., pose and illumination).
5) Potential for *bias propagation* from the previous steps of the pipeline (e.g., data acquisition).

Nevertheless, some results appear to be intuitive, e.g., worse accuracies for women. These could be due to numerous reasons, such as: larger intraclass variations due to makeup [147], occlusion by hairstyle and accessories, or pose differences due to women being shorter than men and cameras being calibrated with the height of men. Likewise, lower sample quality of infant fingerprints makes sense due to anatomical constraints and the fact that the size of the fingerprint area is considered as a relevant factor for fingerprint sample quality. In order to acquire high-quality fingerprint samples from very young data subjects, specialized hardware may be necessary (see [148]).

D. Mitigation

Table II summarizes the existing research in the area of bias mitigation in biometrics. Similar to the above, related

TABLE III
SUMMARY OF THE EXISTING DATASETS FOR BIAS-RELATED
RESEARCH IN BIOMETRICS

Reference	Characteristic	Size (images)	Details
Ricanek *et al.* [154]	Face	55.134	Ageing research database with demographic labels.
Azzopardi *et al.* [155]	Face	946	Subset of FERET dataset balanced w.r.t. sex.
Buolamwini *et al.* [74]	Face	1.270	Images of parliamentarians balanced w.r.t. sex and race. One image per subject, *i.e.* not suitable for biometric recognition.
Alvi *et al.* [118]	Face	14.000	Scraped images balanced w.r.t. race.
Alvi *et al.* [118]	Face	60.000	Subset of IMDB dataset balanced w.r.t. sex and age.
Morales *et al.* [153]	Face	139.677	Subset of MegaFace dataset balanced w.r.t. sex and race.
Merler *et al.* [156]	Face	964.873	Demographic and geometric annotations for selected images from YFCC-100M dataset.
Hupont *et al.* [83]	Face	10.800	Subset of CWF and VGG datasets balanced w.r.t. sex and race.
Kärkkäinen *et al.* [157]	Face	108.501	Subset of YFCC-100M dataset balanced w.r.t. sex, race, and age.
Wang *et al.* [92]	Face	40.607	Subset of MS-Celeb-1M dataset balanced w.r.t. race.
Robinson *et al.* [95]	Face	20.000	Subset of LFW dataset balanced w.r.t. sex and race.
Albiero *et al.* [96]	Face	42.134	Subset of AFD dataset balanced w.r.t. sex.

work here focuses predominantly on the face as biometric characteristic. In this context, mainly recognition and classification algorithms have been analyzed. Generally speaking, the existing approaches can be assigned to the following categories.

1) *Training:* Learning-based methods have experienced a tremendous growth in accuracy and popularity in recent years. As such, the training step is of critical importance for the used systems and mitigation of demographic bias. The existing techniques mainly rely on demographically balanced training datasets (e.g., [92]) and synthetic data to enhance the training datasets (e.g., [125]), as well as learning specialized loss or similarity functions (e.g., [130]). A number of balanced training datasets have been released to the research community, as shown in Table III.

2) *Dynamic Selection:* Deviating from preventing demographic bias, some methods attempted to employ a bias-aware approach. Examples in this category include the dynamic selection of the recognition algorithms (e.g., [63]) or decision thresholds (e.g., [87]) based on demographic attributes of the individual subjects.

In addition to the categories above, other approaches may be considered in the context of bias mitigation. For example, modeling of factors, such as fingerprint growth, can be used to improve the biometric recognition performance for children (see [109]) and to mitigate the effects of aging (see [149]). Other examples include de-identification and anonymization methods (see [150], [151]), whose primary use case is privacy protection in biometrics. Such methods aim to remove, change, or obfuscate certain information (e.g., demographics) either from the image (e.g., [152]) or feature (e.g., [120] and [153]) domain, often through a form of adversarial learning. One could hypothesize that a system trained on such data might not exhibit biases w.r.t. the de-identified demographic covariates. However, the validity of such hypotheses has not yet been ascertained experimentally.

IV. DISCUSSION

In this section, several issues relevant to the topic of this article are discussed. Concretely, Section IV-A addresses the topic of algorithmic fairness in general, while Section IV-B does so in the context of biometrics specifically. Section IV-C illustrates the importance of further research on algorithmic bias and fairness in biometrics by describing the social impact of demographically biased systems.

A. Algorithmic Fairness in General

The challenge of fairness is common in machine learning and computer vision, i.e., it is by no means limited to biometrics. A survey focusing on issues and challenges associated with algorithmic fairness was conducted among industry practitioners by Holstein *et al.* [158]. For a comprehensive overview of bias in automated algorithms in general, the reader is referred to, e.g., [18] and [159]. In addition to algorithmic fairness, algorithmic transparency, explainability, interpretability, and accountability (see [160]–[163]) have also been heavily researched in recent years both from the technical and social perspective. The current research in the area of algorithmic fairness concentrates on the following topics.

1) Theoretical and formal definitions of bias and fairness (see [18], [164], [165]).
2) Fairness metrics, software, and benchmarks (see [166]–[168]).
3) Societal, ethical, and legal aspects of algorithmic decision making and fairness therein (see [1], [169]–[172]).
4) Estimation and mitigation of bias in algorithms and datasets (see [173]–[178]).

Despite decades of research, there exists no single agreed coherent definition of algorithmic fairness. In fact, dozens of formal definitions (see [164], [165]) have been proposed to address different situations and possible criteria of fairness.[7] Certain definitions, which are commonly used and advocated for, are even provably mutually exclusive [179]. Therefore, depending on the definition of fairness one chooses to adopt, a system can effectively always be shown to exhibit some form of bias. As such, the "correct" approach is essentially application dependent. This in turn necessitates a keen domain knowledge and awareness of those issues from the system operators and stakeholders, as they need to select the definitions and metrics of fairness relevant to their particular use case. Research in this area strongly suggests that the notion of fairness in machine learning is context sensitive [180], [181]; this presumably also applies to the field of biometrics, especially for machine learning-based systems. In the next section, the notions of fairness and bias are discussed in the context of biometrics specifically based on the literature surveyed in Section III.

---

[7]See also https://towardsdatascience.com/a-tutorial-on-fairness-in-machine-learning-3ff8ba1040cb and https://fairmlbook.org/ for visual tutorials on bias and fairness in machine learning.

B. Algorithmic Fairness in Biometrics

Although the topic of demographic bias and fairness in biometrics has emerged relatively recently, it has quickly established itself as an important and popular research area. Several high-ranking conferences featured special sessions,[8],[9],[10] NIST conducted large-scale evaluations [52], while ISO/IEC is currently preparing a technical report on this subject [45]. Likewise, a significant number of scientific publications has appeared on this topic (surveyed in Section III). The existing studies concentrated on face-based biometrics—more research is urgently needed for other biometric characteristics, e.g., fingerprints [182]. The existing studies primarily address the following aspects.

1) Evaluations with the aim of quantitatively ascertaining the degree of demographic bias in various biometric algorithms.
2) Methods that seek to mitigate the effects of demographic bias in various biometric algorithms.

The existing bias estimation studies have uncovered new trends w.r.t. algorithmic bias and fairness in biometric algorithms (recall Section III-C). However, it should be noted as described as follows.

1) In many cases, the biases were algorithm specific, i.e., while given the same benchmark dataset, some algorithms exhibited a bias (e.g., lower biometric performance for a certain demographic group), others did not. In aggregate, however, the existing studies did seem to agree on certain points, as described in Section III-C.
2) While a high relative increase in error rates for a certain demographic group may appear quite substantial, its importance in absolute terms could be negligible, especially for very accurate algorithms which hardly make any errors whatsoever [52].

Those caveats notwithstanding, the commitment of the academic researchers and commercial vendors to researching algorithmic fairness is especially important for the public perception of biometric technologies. The field of algorithmic fairness in the context of biometrics is in its infancy and a large number of issues are yet to be comprehensively addressed (see Section IV-A).

1) Limited theoretical work has been conducted in this field specifically focusing on biometrics. Indeed, the majority of publications surveyed in Section III do not approach the notions of bias and fairness rigorously; rather, they tend to concentrate on an equivalent of some of the simpler statistical definitions, such as group fairness and error rate parity. Extending the existing estimation and mitigation works, for example, to consider other and more complex notions of fairness (see [129]) could be seen as important future work in the field. Likewise, investigating tradeoffs between biometric performance, fairness, user experience, social perceptions, monetary costs, and other aspects of the biometric systems might be of interest.
2) In addition to empiric studies (especially in the case of bias mitigation, see Section III-D), stricter theoretical approaches need to be pursued in order to provably demonstrate the bias-mitigating properties of the proposed methods.
3) Isolating the effects of the demographic factors from other confounding factors (i.e., the environmental and subject-specific covariates, such as illumination and use of accessories) is a challenging task, which is not sufficiently addressed in many existing studies. An example of a study which partially addressed those issues in a systematic manner is the work of Grother *et al.* [52].
4) More complex analyses based on demographic attributes and combinations thereof (intersectionality) could be conducted for a more detailed and nuanced view of demographic biases in biometric systems.
5) Comprehensive independent benchmarks utilizing various algorithmic fairness measurement methodologies and metrics are, as of yet, lacking. Only recently, in [52], first independent benchmarks of biometric recognition algorithms have been conducted. Similar and more extensive benchmarks for other biometric algorithms (recall Section III-A) are needed.
6) Large-scale datasets designed specifically for bias-related research need to be collected. The existing datasets only pertain to face-based biometrics (see Table III).
7) Humans are known to exhibit a broad range of biases [14], [25]. The influence of those factors on the biometric algorithm design, interactions with and use of biometric systems, as well as perceptions of biometric systems could be investigated.
8) Most of the surveyed studies did not explicitly provide information about ethics approval. Future works could improve on those practices, especially considering the sensitive nature of the research topic at hand.

In the next section, the possible consequences of failing to appropriately address the issues of algorithmic fairness in biometrics are discussed.

C. Social Impact

Numerous studies described the potential of real harms as a consequence of biased algorithmic decision-making systems [169], [183] in general. Regarding biometric systems, in particular, facial recognition technologies have been the main focus of such discussions (see [184]). Concerning the notions of bias and fairness, in addition to being context sensitive (recall Section IV-A), one might argue the impact assessments to also be purpose sensitive. Specifically, depending on the application scenario, the impact and importance of systemic biases might differ significantly. As an example, consider an application of biometrics in cooperative access control systems or personal devices. A demographic bias in such a system might cause a certain demographic group to be inconvenienced through additional authentication attempt(s) being

---

[8]https://sites.google.com/view/wacv2020demographics
[9]https://sites.google.com/site/eccvbefa2018
[10]https://dasec.h-da.de/wp-content/uploads/2020/01/EUSIPCO2020-ss_bias_in_biometrics.pdf

necessary due to false-negative errors. On the other hand, the stakes are much higher in, e.g., a state surveillance scenario. There, demographic biases could directly cause substantial personal harms, e.g., higher (unjustified) arrest rates [12], due to false-positive errors. At the same time, it is also clear that the biometric recognition technology can be highly accurate. Taking the recently contested facial recognition as an example, given prerequisites, such as a high-resolution camera, proper lighting and image quality controls, as well as high-quality comparison algorithms, the absolute error rates can become vanishingly small [52], thereby potentially rendering the relative imbalance of error rates across demographic groups insignificant.

It should be noted that there are no indications of the algorithmic biases in biometrics being deliberately put into the algorithms by design; rather, they are typically a result of the used training data and other factors. In any case, one should also be mindful, that as any technology, biometrics could be used in malicious or dystopian ways (e.g., privacy violations through mass surveillance [185] or "crime prediction" [186]). Consequently, a framework for human impact assessments [187] should be developed for biometrics as soon as possible. A proactive and cognizant approach could foster awareness among the citizens and policymakers, as well as contribute to minimizing the potential negative perception of the biometric technology and innovation by individuals and society as a whole.

In a broader context, algorithmic bias and fairness is one of the topics in the larger discourse on the ethical design in AI systems [188], most prominently encompassing:

1) transparency;
2) accountability;
3) explainability;
4) fairness.

Currently, the legal and societal scrutiny of the technologies utilizing automated decision systems seem to be insufficient. However, recent legislation in the European Union [189], [190] constitutes a step in that direction. Below, several social and technological provisions, which might be considered in this context, are listed.

1) Carefully *selecting the data used to train* the algorithms is the first and perhaps the most important step: inherent biases in training data should be avoided wherever possible. Furthermore, the size of the dataset matters—some systems have been reported to be trained on very small datasets (in the order of thousands of items), which are usually wholly insufficient to show that an approach generalizes well.
2) Higher degree of *transparency* and/or independent insight into data and algorithms, as well as validation of the results could be established to foster the public trust and acceptance of the systems.
3) Thresholds for *acceptable accuracy* (i.e., how much the systems can error) could be established legally (potentially in a system purpose-sensitive manner), as well as reviewed and validated periodically.
4) Special *training of the systems' personnel* could be established to make them aware of the potential issues

and to establish proper protocols for dealing with them.
5) *Due diligence* could be legally expected from vendors of such systems, i.e., in reasonably ensuring some or all of the aforementioned matters and rectifying problems as they come up. Additionally, certain accountability provisions could be incorporated to further facilitate this.

The issues of fairness (including algorithmic fairness) are complicated from the point of view of the legislation—a somewhat deep understanding of statistics, formal fairness definitions, and other concepts is essential for an informed discourse. Furthermore, the ethical and moral perceptions and decisions are not uniform across different population demographics and by geographical location (see [191]). This reinforces an important dilemma regarding the regulation of automated decision systems—since many situations are morally and ethically ambiguous to humans, how should they be able to encode ethical decision making into laws? Once that issue is somehow surmounted, there also remains the issue of feasibility of technical solutions, as described in the previous two sections.

Currently, many laws and rules exist (international treaties, constitutions of many countries, and employment law) which aim to protect against generic discrimination on the basis of demographics [192]. However, historically, the enforcement of those has been fraught with difficulties and controversies. In this context, the algorithmic decision systems are merely one of the most recent and technologically advanced cases. The policymakers and other stakeholders will have to tackle it in the upcoming years in order to develop a legal framework similar to those already governing other areas and aspects of the society [193].

V. SUMMARY

This article has investigated the challenge of demographic bias in biometric systems. Following an overview of the topic and challenges associated therewith, a comprehensive survey of the literature on bias estimation and mitigation in biometric algorithms has been conducted. It has been found that demographic factors can have a large influence on various biometric algorithms and that current algorithms tend to exhibit some degree of bias w.r.t. certain demographic groups. Most effects are algorithm dependent, but some consistent trends do also appear (as discussed in Section III-C). Specifically, many studies point to a lower biometric performance for females and youngest subjects in biometric recognition systems, as well as lower classification accuracy for dark-skinned females in the classification of demographic attributes from facial images. It should be noted that many of the studies conducted their experiments using relatively small datasets, which emphasize the need for large-scale studies. In general, a broad spectrum of open technical (and other) challenges exists in this field (see Section IV).

Biased automated decision systems can be detrimental to their users, with issues ranging from simple inconveniences, through disadvantages, to lasting serious harms. This relevance notwithstanding, the topic of algorithmic fairness is still

relatively new, with many unexplored areas and few legal and practical provisions in existence. Recently, a growing academic and media coverage has emerged, where the overwhelming consensus appears to be that such systems need to be properly assessed (e.g., through independent benchmarks), compelled to some degree of transparency, accountability, and explainability in addition to guaranteeing some fairness definitions. Furthermore, it appears that, in certain cases, legal provisions might need to be introduced to regulate these technologies.

Automatic decision systems (including biometrics) are experiencing a rapid technological progress, thus simultaneously holding a potential of beneficial and harmful applications, as well as unintentional discrimination. Zweig *et al.* [17] even argued that the issues (including, but not limited to bias and fairness) concerning algorithmic decision systems are directly related to the so-called "quality of democracy" measure of countries. As such, developing proper frameworks and rules for such technologies is a large challenge which the policymakers and the society as a whole must face in the upcoming future [194], [195].

REFERENCES

[1] F. Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA, USA: Harvard Univ. Press, 2015.

[2] O. A. Osoba and W. Welser, *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*. Santa Monica, CA, USA: Rand Corporat., 2017.

[3] A. L. Washington, "How to argue with an algorithm: Lessons from the COMPAS-ProPublica debate," *Colorado Technol. Law J.*, vol. 17, pp. 131–160, Mar. 2018.

[4] K.-H. Yu and I. S. Kohane, "Framing the challenges of artificial intelligence in medicine," *BMJ Qual. Safety*, vol. 28, no. 3, pp. 238–241, Mar. 2019.

[5] M. Hurley and J. Adebayo, "Credit scoring in the era of big data," *Yale J. Law Technol.*, vol. 18, no. 1, p. 5, Apr. 2017.

[6] C. Castelluccia and D. Le Métayer, "Understanding algorithmic decision-making: Opportunities and challenges," Institut Nat. de Recherche en Informatique et en automatique, Rocquencourt, France, Rep. PE 624.261, Mar. 2019.

[7] A. Bhutani and P. Bhardwaj, "Biometrics market size by application," Global Market Insights, Pune, India, Rep. GMI493, Aug. 2017.

[8] "Biometric system market by authentication type—Global forecast to 2023," Marketsandmarkets, Pune, India, Rep. SE 3449, Jul. 2018.

[9] D. Thakkar. (Aug. 2018). *Global Biometric Market Analysis: Trends and Future Prospects*. Accessed: May 12, 2020. [Online]. Available: https://www.bayometric.com/global-biometric-market-analysis/

[10] Unique Identification Authority of India. *Aadhaar Dashboard*. Accessed: May 12, 2020. [Online]. Available: https://www.uidai.gov.in/aadhaar_dashboard/

[11] A. Ross *et al.*, "Some research problems in biometrics: The future beckons," in *Proc. Int. Conf. Biometr. (ICB)*, Jun. 2019, pp. 1–8.

[12] C. Garvie, *The Perpetual Line-Up: Unregulated Police Face Recognition in America*, Center Privacy Technol., Georgetown Law, Washington, DC, USA, Oct. 2016.

[13] C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, NY, USA: Broadway Books, 2016.

[14] J. S. B. T. Evans, *Bias in Human Reasoning: Causes and Consequences*. London, U.K.: Lawrence Erlbaum Assoc., 1989.

[15] R. R. Banks, J. L. Eberhardt, and L. Ross, "Discrimination and implicit bias in a racially unequal society," *California Law Rev.*, vol. 94, no. 4, pp. 1169–1190, Jul. 2006.

[16] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*. vol. 1. New York, NY, USA: Springer, Feb. 2009.

[17] K. A. Zweig, G. Wenzelburger, and T. D. Krafft, "On chances and risks of security related algorithmic decision making systems," *Eur. J. Security Res.*, vol. 3, pp. 1–23, Apr. 2018.

[18] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," Sep. 2019. [Online]. Available: arXiv:1908.09635.

[19] D. Danks and A. J. London, "Algorithmic bias in autonomous systems," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, Aug. 2017, pp. 4691–4697.

[20] B. Friedman and H. Nissenbaum, "Bias in computer systems," *IEEE Trans. Inf. Syst.*, vol. 14, no. 3, pp. 330–347, Jul. 1996.

[21] S. Lansing, "New York state COMPAS-probation risk and need assessment study: Examining the recidivism scale's effectiveness and predictive accuracy," New York State Division Criminal Justice Services, Albany, NY, USA, Rep., Sep. 2012.

[22] S. Desmarais and J. Singh, *Risk Assessment Instruments Validated and Implemented in Correctional Settings in the United States*. New York, NY, USA: Council State Govt. Justice Center, Mar. 2013.

[23] A. Chouldechova, "Fair prediction with disparate impact: A study of bias in recidivism prediction instruments," *Big Data*, vol. 5, no. 2, pp. 153–163, Jun. 2017.

[24] K. L. Mosier, L. J. Skitka, S. Heers, and M. Burdick, "Automation bias: Decision making and performance in high-tech cockpits," *Int. J. Aviation Psychol.*, vol. 8, no. 1, pp. 47–63, Jan. 1998.

[25] R. Parasuraman and D. H. Manzey, "Complacency and bias in human use of automation: An attentional integration," *Human Factors*, vol. 52, no. 3, pp. 381–410, Jun. 2010.

[26] *ISO/IEC 2382-37:2017. Information Technology—Vocabulary—Part 37: Biometrics*, 2nd ed. Geneva, Switzerland: Int. Org. Stand., Feb. 2017.

[27] P. J. Phillips *et al.*, "Overview of the face recognition grand challenge," in *Proc. Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2005, pp. 947–954.

[28] A. Kumar and A. Passi, "Comparison and combination of iris matchers for reliable personal authentication," *Pattern Recognit.*, vol. 43, no. 3, pp. 1016–1026, Mar. 2010.

[29] J. Ortega-Garcia *et al.*, "MCYT baseline corpus: A bimodal biometric database," *IEE Proc. Vis. Image Signal Process.*, vol. 150, no. 6, pp. 395–401, Dec. 2003.

[30] B. T. Ton and R. N. J. Veldhuis, "A high quality finger vascular pattern dataset collected using a custom designed capturing device," in *Proc. IEEE Int. Conf. Biometr. (ICB)*, Jun. 2013, pp. 1–5.

[31] L. Liu, J. Chen, P. Fieguth, G. Zhao, R. Chellappa, and M. Pietikäinen, "From BoW to CNN: Two decades of texture representation for texture classification," *Int. J. Comput. Vis.*, vol. 127, no. 1, pp. 74–109, Jan. 2019.

[32] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 1701–1708.

[33] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 815–823.

[34] O. M. Parkhi *et al.*, "Deep face recognition," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, Sep. 2015, pp. 1–6.

[35] Y. Tang, F. Gao, J. Feng, and Y. Liu, "FingerNet: An unified deep network for fingerprint minutiae extraction," in *Proc. Int. Joint Conf. Biometr. (IJCB)*, Oct. 2017, pp. 108–116.

[36] K. Nguyen, C. Fookes, A. Ross, and S. Sridharan, "Iris recognition with off-the-shelf CNN features: A deep learning perspective," *IEEE Access*, vol. 6, pp. 18848–18855, 2018.

[37] K. Sundararajan and D. L. Woodard, "Deep learning for biometrics: A survey," *ACM Comput. Surveys*, vol. 51, no. 3, pp. 1–34, Jul. 2018.

[38] S. Z. Li and A. K. Jain, *Encyclopedia of Biometrics*. New York, NY, USA: Springer, 2015.

[39] A. K. Jain, P. Flynn, and A. Ross, *Handbook of Biometrics*. New York, NY, USA: Springer, 2007.

[40] S. Z. Li and A. K. Jain, *Handbook of Face Recognition*. London, U.K.: Springer, 2004.

[41] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*. London, U.K.: Springer, 2009.

[42] K. Bowyer and M. J. Burge, *Handbook of Iris Recognition*. London, U.K.: Springer, 2016.

[43] A. Uhl, S. Marcel, C. Busch, and R. N. J. Veldhuis, *Handbook of Vascular Biometrics*. London, U.K.: Springer, 2020.

[44] O. Keyes, "The misgendering machines: Trans/HCI implications of automatic gender recognition," *ACM Human–Comput. Interact.*, vol. 2, p. 88, Nov. 2018.

[45] *Information Technology—Biometrics—Identifying and Mitigating the Differential Impact of Demographic Factors in Biometric Systems*, ISO/IEC Standard WD TR 22116.

[46] ISO/IEC JTC1 SC37 Biometrics, *Information Technology—Biometric Performance Testing and Reporting—Part 1: Principles and Framework, International Organization for Standardization and International Electrotechnical Committee*, ISO/IEC Standard 19795-1:2006, Apr. 2006.

[47] M. Feldman, S. A. Friedler, J. Moeller, C. Scheidegger, and S. Venkatasubramanian, "Certifying and removing disparate impact," in *Proc. ACM Int. Conf. Knowl. Disc. Data Min.*, Aug. 2015, pp. 259–268.

[48] J. J. Howard, Y. B. Sirotin, and A. R. Vemury, "The effect of broad and specific demographic homogeneity on the imposter distributions and false match rates in face recognition algorithm performance," in *Proc. IEEE Int. Conf. Biometr. Theory Appl. Syst. (BTAS)*, Sep. 2019, pp. 1–8.

[49] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, "Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation," in *Proc. Int. Conf. Spoken Lang. Process.*, Dec. 1998, pp. 1351–1355.

[50] N. Yager and T. Dunstone, "Worms, chameleons, phantoms and doves: New additions to the biometric menagerie," in *Proc. IEEE Workshop Autom. Identification Adv. Technol. (AutoID)*, Jun. 2007, pp. 1–6.

[51] J. Daugman and C. Downing, "Searching for Doppelgängers: Assessing the universality of the IrisCode impostors distribution," *IET Biometr.*, vol. 5, no. 2, pp. 65–75, Jun. 2016.

[52] P. Grother, M. Ngan, and K. Hanaoka, "Ongoing face recognition vendor test (FRVT) part 3: Demographic effects," Nat. Inst. Stand. Technol., Gaithersburg, MA, USA, Rep. NISTIR 8280, Dec. 2019.

[53] N. Furl, P. J. Phillips, and A. J. O'Toole, "Face recognition algorithms and the other-race effect: Computational mechanisms for a developmental contact hypothesis," *Cogn. Sci.*, vol. 26, no. 6, pp. 797–815, Nov. 2002.

[54] A. Dantcheva, P. Elia, and A. Ross, "What else does your biometric data reveal? A survey on soft biometrics," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 3, pp. 441–467, Mar. 2016.

[55] Y. Sun, M. Zhang, Z. Sun, and T. Tan, "Demographic analysis from biometric data: Achievements, challenges, and new frontiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 2, pp. 332–351, Feb. 2018.

[56] ISO/IEC JTC1 SC37 Biometrics, *Information Technology—Biometric Sample Quality—Part 1: Framework, International Organization for Standardization and International Electrotechnical Committee*, ISO/IEC Standard 29794-1:2016, Sep. 2016.

[57] S. Bharadwaj, M. Vatsa, and R. Singh, "Biometric quality: A review of fingerprint, iris, and face," *EURASIP J. Image Video Process.*, vol. 2014, no. 1, p. 34, Jul. 2014.

[58] ISO/IEC JTC1 SC37 Biometrics, *Information Technology—Biometric Presentation Attack Detection—Part 1: Framework, International Organization for Standardization and International Electrotechnical Committee*, ISO/IEC Standard 30107-1:2016, Jan. 2016.

[59] S. Marcel, M. S. Nixon, J. Fierrez, and N. Evans, *Handbook of Biometric Anti-spoofing: Presentation Attack Detection*. Cham, Switzerland: Springer, 2019.

[60] A. Kortylewski, B. Egger, A. Schneider, T. Gerig, A. Morel-Forster, and T. Vetter, "Empirically analyzing the effect of dataset biases on deep face recognition systems," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 2093–2102.

[61] J. R. Beveridge, G. H. Givens, P. J. Phillips, and B. A. Draper, "Factors that influence algorithm performance in the face recognition grand challenge," *Proc. Comput. Vis. Image Understand.*, vol. 113, no. 6, pp. 750–762, Jun. 2009.

[62] Y. M. Lui, D. Bolme, B. A. Draper, J. R. Beveridge, G. Givens, and P. J. Phillips, "A meta-analysis of face recognition covariates," in *Proc. IEEE Int. Conf. Biometr. Theory Appl. Syst. (BTAS)*, Sep. 2009, pp. 1–8.

[63] G. Guo and G. Mu, "Human age estimation: What is the influence across race and gender?" in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2010, pp. 71–78.

[64] P. Grother, G. W. Quinn, and P. J. Phillips, "Report on the evaluation of 2D still-image face recognition algorithms," Nat. Inst. Stand. Technol., Gaithersburg, MA, USA, Rep. NISTIR 7709, Aug. 2010.

[65] P. J. Phillips, F. Jiang, A. Narvekar, J. Ayyad, and A. J. O'Toole, "An other-race effect for face recognition algorithms," *Trans. Appl. Perception*, vol. 8, no. 2, pp. 1–11, Jan. 2011.

[66] A. J. O'Toole, P. J. Phillips, X. An, and J. Dunlop, "Demographic effects on estimates of automatic face recognition performance," *IEEE Image Vis. Comput.*, vol. 30, no. 3, pp. 169–176, Mar. 2012.

[67] B. F. Klare, M. J. Burge, J. C. Klontz, R. W. V. Bruegge, and A. K. Jain, "Face recognition performance: Role of demographic information," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1789–1801, Oct. 2012.

[68] G. H. Givens, J. R. Beveridge, P. J. Phillips, B. Draper, Y. M. Lui, and D. Bolme, "Introduction to face recognition and evaluation of algorithm performance," *Comput. Stat. Data Anal.*, vol. 67, pp. 236–247, Nov. 2013.

[69] J. R. Beveridge *et al.*, "Report on the FG 2015 video person recognition evaluation," in *Proc. IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, vol. 1, May 2015, pp. 1–8.

[70] K. Ricanek, S. Bhardwaj, and M. Sodomsky, "A review of face recognition against longitudinal child faces," in *Proc. Int. Conf. Biometr. Special Interest Group (BIOSIG)*, Sep. 2015, pp. 15–26.

[71] H. El Khiyari and H. Wechsler, "Face verification subject to varying (age, ethnicity, and gender) demographics using deep learning," *J. Biometr. Biostat.*, vol. 7, no. 323, pp. 11–16, Nov. 2016.

[72] D. Deb, L. Best-Rowden, and A. K. Jain, "Face recognition performance under aging," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 46–54.

[73] L. Best-Rowden and A. K. Jain, "Longitudinal study of automatic face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 1, pp. 148–162, Jan. 2018.

[74] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Proc. ACM Conf. Fairness Accountability Transparency*, Jan. 2018, pp. 77–91.

[75] D. Deb, N. Nain, and A. K. Jain, "Longitudinal study of child face recognition," in *Proc. IEEE Int. Conf. Biometr. (ICB)*, Feb. 2018, pp. 225–232.

[76] D. Michalski, S. Y. Yiu, and C. Malec, "The impact of age and threshold variation on facial recognition algorithm performance using images of children," in *Proc. IEEE Int. Conf. Biometr. (ICB)*, Feb. 2018, pp. 217–224.

[77] S. H. Abdurrahim, S. A. Samad, and A. B. Huddin, "Review on the effects of age, gender, and race demographics on automatic face recognition," *Vis. Comput.*, vol. 34, no. 11, pp. 1617–1630, Aug. 2018.

[78] L. Rhue, "Racial influence on automated perceptions of emotions," in *Proc. Soc. Sci. Res. Netw.*, Nov. 2018, p. 11.

[79] B. Lu, J. Chen, C. D. Castillo, and R. Chellappa, "An experimental evaluation of covariates effects on unconstrained face verification," *IEEE Trans. Biometr. Behav. Identity Sci.*, vol. 1, no. 1, pp. 42–55, Jan. 2019.

[80] I. D. Raji and J. Buolamwini, "Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products," in *Proc. ACM Conf. AI Ethics Soc. (AIES)*, Jan. 2019, pp. 429–435.

[81] N. Srinivas, M. Hivner, K. Gay, H. Atwal, M. King, and K. Ricanek, "Exploring automatic face recognition on match performance and gender bias for children," in *Proc. IEEE Winter Appl. Comput. Vis. Workshops (WACVW)*, Jan. 2019, pp. 107–115.

[82] C. M. Cook, J. J. Howard, Y. B. Sirotin, J. L. Tipton, and A. R. Vemury, "Demographic effects in facial recognition and their dependence on image acquisition: An evaluation of eleven commercial systems," *IEEE Trans. Biometr. Behav. Identity Sci.*, vol. 1, no. 1, pp. 32–41, Feb. 2019.

[83] I. Hupont and C. Fernández, "DemogPairs: Quantifying the impact of demographic imbalance in deep face recognition," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2019, pp. 1–7.

[84] E. Denton, B. Hutchinson, M. Mitchell, and T. Gebru, "Detecting bias with generative counterfactual face attribute augmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1–10.

[85] R. V. Garcia, L. Wandzik, L. Grabner, and J. Krueger, "The harms of demographic bias in deep face recognition research," in *Proc. Int. Conf. Biometr. (ICB)*, Jun. 2019, pp. 1–6.

[86] S. Nagpal, M. Singh, R. Singh, M. Vatsa, and N. Ratha, "Deep learning for face recognition: Pride or prejudiced?" Jun. 2019. [Online]. Available: arXiv:1904.01219.

[87] K. S. Krishnapriya, K. Vangara, M. C. King, V. Albiero, and K. Bowyer, "Characterizing the variability in face recognition accuracy relative to race," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 2278–2285.

[88] ISO/IEC JTC1 SC37 Biometrics, *Information Technology—Biometric Data Interchange Formats—Part 5: Face Image Data, International Organization for Standardization and International Electrotechnical Committee*, ISO/IEC Standard 19794-5:2005, Jun. 2011.

[89] V. Muthukumar, "Color-theoretic experiments to understand unequal gender classification accuracy from face images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 2286–2295.

[90] N. Srinivas, K. Ricanek, D. Michalski, D. S. Bolme, and M. King, "Face recognition algorithm bias: Performance differences on images of children and adults," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 2269–2277.

[91] R. Vera-Rodriguez, M. Blazquez, A. Morales, E. Gonzalez-Sosa, J. C. Neves, and H. Proença, "FaceGenderID: Exploiting gender information in DCNNs face recognition systems," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 2254–2260.

[92] M. Wang, W. Deng, J. Hu, X. Tao, and Y. Huang, "Racial faces in-the-wild: Reducing racial bias by information maximization adaptation network," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, Nov. 2019, pp. 692–702.

[93] I. Serna, A. Morales, J. Fierrez, M. Cebrian, N. Obradovich, and I. Rahwan, "Algorithmic discrimination: Formulation and exploration in deep learning-based face biometrics," Dec. 2019. [Online]. Available: arXiv:1912.01842.

[94] J. G. Cavazos, P. J. Phillips, C. D. Castillo, and A. J. O'Toole, "Accuracy comparison across face recognition algorithms: Where are we on measuring race bias?" Dec. 2019. [Online]. Available: arXiv:1912.07398.

[95] J. P. Robinson, G. Livitz, Y. Henon, C. Qin, Y. Fu, and S. Timoner, "Face recognition: Too bias, or not too bias?" Feb. 2020. [Online]. Available: arXiv:2002.06483.

[96] V. Albiero, K. S. Krishnapriya, K. Vangara, K. Zhang, M. C. King, and K. W. Bowyer, "Analysis of gender inequality in face recognition accuracy," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 81–89.

[97] K. S. Krishnapriya, V. Albiero, K. Vangara, M. C. King, and K. W. Bowyer, "Issues related to face recognition accuracy varying based on race and skin tone," *IEEE Trans. Technol. Soc.*, vol. 1, no. 1, pp. 8–20, Mar. 2020.

[98] P. Terhörst, J. N. Kolf, N. Damer, F. Kirchbuchner, and A. Kuijper, "Face quality estimation and its correlation to demographic and non-demographic bias in face recognition," Apr. 2020. [Online]. Available: arXiv:2004.01019.

[99] R. A. Hicklin and C. L. Reedy, "Implications of the IDENT/IAFIS image quality study for visa fingerprint processing," Mitretek Syst., Falls Church, VA, USA, Rep., Oct. 2002.

[100] N. C. Sickler and S. J. Elliott, "An evaluation of fingerprint image quality across an elderly population vis-a-vis an 18–25 year old population," in *Proc. IEEE Int. Carnahan Conf. Security Technol.*, Oct. 2005, pp. 68–73.

[101] S. K. Modi and S. J. Elliott, "Impact of image quality on performance: Comparison of young and elderly fingerprints," in *Proc. Int. Conf. Recent Adv. Soft Comput. (RASC)*, Jul. 2006, pp. 449–454.

[102] S. K. Modi, S. J. Elliott, J. Whetsone, and H. Kim, "Impact of age groups on fingerprint recognition performance," in *Proc. IEEE Workshop Autom. Identification Adv. Technol. (AutoID)*, Jun. 2007, pp. 19–23.

[103] M. Frick, S. K. Modi, S. Elliott, and E. P. Kukula, "Impact of gender on fingerprint recognition systems," in *Proc. Int. Conf. Inf. Technol. Appl. (ICITA)*, 2008, pp. 717–721.

[104] K. O'Connor and S. J. Elliott, "The impact of gender on image quality, Henry classification and performance on a fingerprint recognition system," in *Proc. Int. Conf. Inf. Technol. Appl. (ICITA)*, 2011, pp. 304–307.

[105] G. Schumacher, "Fingerprint recognition for children," Joint Res. Centre, Ispra, Italy, Rep. EUR 26193 EN, Sep. 2013.

[106] S. Yoon and A. K. Jain, "Longitudinal study of fingerprint recognition," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 28, pp. 8555–8560, Jul. 2015.

[107] J. Galbally, R. Haraksim, and L. Beslay, "Fingerprint quality: A life-time story," in *Proc. IEEE Int. Conf. Biometr. Special Interest Group (BIOSIG)*, Sep. 2018, pp. 1–5.

[108] J. Galbally, R. Haraksim, and L. Beslay, "A study of age and ageing in fingerprint biometrics," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 5, pp. 1351–1365, May 2019.

[109] J. Preciozzi *et al.*, "Fingerprint biometrics from newborn to adult: A study from a national identity database system," *IEEE Trans. Biometr. Behav. Identity Sci.*, vol. 2, no. 1, pp. 68–79, Jan. 2020.

[110] P. Drozdowski *et al.*, "Demographic bias: A challenge for fingervein recognition systems?" Apr. 2020. [Online]. Available: arXiv:2004.01418.

[111] M. Fang, N. Damer, F. Kirchbuchner, and A. Kuijper, "Demographic bias in presentation attack detection of iris recognition systems," Mar. 2020. [Online]. Available: arXiv:2003.03151.

[112] Z. Xie, Z. Guo, and C. Qian, "Palmprint gender classification by convolutional neural network," *IET Comput. Vis.*, vol. 12, no. 4, pp. 476–483, May 2018.

[113] A. Uhl and P. Wild, "Comparing verification performance of kids and adults for fingerprint, palmprint, hand-geometry and digitprint biometrics," in *Proc. Int. Conf. Biometr. Theory Appl. Syst. (BTAS)*, Sep. 2009, pp. 1–6.

[114] M. Brandão, "Age and gender bias in pedestrian detection algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1–4.

[115] Y. Guo and L. Zhang, "One-shot face recognition by promoting underrepresented classes," Microsoft, Redmond, WA, USA, Rep. MSR-TR-2017-27, Jul. 2017.

[116] H. J. Ryu, H. Adam, and M. Mitchell, "InclusiveFaceNet: Improving face attribute detection with race and gender diversity," in *Proc. ACM Workshop Fairness Accountability Transparency Mach. Learn.*, Jul. 2018, pp. 1–6.

[117] M. Hasnat *et al.*, "von Mises–Fisher mixture model-based deep learning: Application to face verification," Dec. 2017. [Online]. Available: arXiv:1706.04264.

[118] M. Alvi, A. Zisserman, and C. Nellåker, "Turning a blind eye: Explicit removal of biases and variation from deep neural network embeddings," in *Proc. Eur. Conf. Comput. Vis. Workshops (ECCVW)*, Sep. 2018, pp. 556–572.

[119] A. Das, A. Dantcheva, and F. Bremond, "Mitigating bias in gender, age and ethnicity classification: A multi-task convolution neural network approach," in *Proc. Eur. Conf. Comput. Vis. Workshops (ECCVW)*, Sep. 2018, pp. 1–13.

[120] A. Acien, A. Morales, R. Vera-Rodriguez, I. Bartolome, and J. Fierrez, "Measuring the gender and ethnicity bias in deep models for face recognition," in *Proc. Ibero Amer. Congr. Pattern Recognit. (CIARP)*, Nov. 2018, pp. 584–593.

[121] A. Amini, A. P. Soleimany, W. Schwarting, S. N. Bhatia, and D. Rus, "Uncovering and mitigating algorithmic bias through learned latent structure," in *Proc. ACM Conf. AI Ethics Soc. (AIES)*, Jan. 2019, pp. 289–295.

[122] P. Terhörst, N. Damer, F. Kirchbuchner, and A. Kuijper, "Suppressing gender and age in face templates using incremental variable elimination," in *Proc. Int. Conf. Biometr. (ICB)*, Jun. 2019, pp. 1–8.

[123] P. Terhörst, N. Damer, F. Kirchbuchner, and A. Kuijper, "Unsupervised privacy-enhancement of face representations using similarity-sensitive noise transformations," *Appl. Intell.*, vol. 49, no. 8, pp. 3043–3060, Aug. 2019.

[124] S. Gong, X. Liu, and A. K. Jain, "DebFace: De-biasing face recognition," Mar. 2019. [Online]. Available: arXiv:1911.08080.

[125] A. Kortylewski, B. Egger, A. Schneider, T. Gerig, A. Morel-Forster, and T. Vetter, "Analyzing and reducing the damage of dataset bias to face recognition with synthetic data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 2261–2268.

[126] M. Wang and W. Deng, "Mitigate bias in face recognition using skewness-aware reinforcement learning," Nov. 2019. [Online]. Available: arXiv:1911.10692.

[127] M. Bruveris, J. Gietema, P. Mortazavian, and M. Mahadevan, "Reducing geographic performance differentials for face recognition," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 98–106.

[128] P. Smith and K. Ricanek, "Mitigating algorithmic bias: Evolving an augmentation policy that is non-biasing," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 90–97.

[129] P. Terhörst, J. N. Kolf, N. Damer, F. Kirchbuchner, and A. Kuijper, "Post-comparison mitigation of demographic bias in face recognition using fair score normalization," Feb. 2020. [Online]. Available: arXiv:2002.03592.

[130] P. Terhörst, M. L. Tran, N. Damer, F. Kirchbuchner, and A. Kuijper, "Comparison-level mitigation of ethnic bias in face recognition," in *Proc. IEEE Int. Workshop Biometr. Forensics (IWBF)*, Apr. 2020, pp. 1–6.

[131] C. Gottschlich, T. Hotz, R. Lorenz, S. Bernhardt, M. Hantschel, and A. Munk, "Modeling the growth of fingerprints improves matching for adolescents," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 1165–1169, Sep. 2011.

[132] E. Bekele, C. Narber, and W. Lawson, "Multi-attribute residual network (MAResNet) for soft-biometrics recognition in surveillance scenarios," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2017, pp. 386–393.

[133] T. Wang, J. Zhao, M. Yatskar, K.-W. Chang, and V. Ordonez, "Balanced datasets are not enough: Estimating and mitigating gender bias in deep image representations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 5309–5318.

[134] M. Orcutt. (Jul. 2016). *Are Face Recognition Systems Accurate? Depends on Your Race.* [Online]. Available: https://www.technologyreview.com/s/601786/are-face-recognition-systems-accurate-depends-on-your-race/

[135] J. Snow. (Jul. 2018). *Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots.* [Online]. Available: https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28

[136] M. Fairhurst, *Age Factors in Biometric Processing.* Inst. Eng. Technol., London, U.K., 2013.

[137] K. W. Bowyer and E. Ortiz, "Critical examination of the IREX VI results," *IET Biometr.*, vol. 4, no. 4, pp. 192–199, Dec. 2015.

[138] P. Grother, J. R. Matey, and G. W. Quinn, "IREX VI: Mixed-effects longitudinal models for iris ageing: Response to bowyer and ORTIZ," *IET Biometr.*, vol. 4, no. 4, pp. 200–205, Dec. 2015.

[139] M. Arnold, C. Busch, and H. Ihmor, "Investigating performance and impacts on fingerprint recognition systems," in *Proc. IEEE SMC Inf. Assurance Workshop*, Jun. 2005, pp. 1–7.

[140] A. Kuehlkamp, B. Becker, and K. Bowyer, "Gender-from-Iris or gender-from-mascara?" in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2017, pp. 1151–1159.

[141] ISO/IEC JTC1 SC37 Biometrics, *Information Technology—Biometric Sample Quality—Part 4: Finger Image Data, International Organization for Standardization and International Electrotechnical Committee*, ISO/IEC Standard 29794-4:2017, Sep. 2015.

[142] J. Galbally, P. Ferrara, R. Haraksim, A. Psyllos, and L. Beslay, "Study on face identification technology for its implementation in the Schengen information system," Joint Res. Centre, Ispra, Italy, Rep. JRC-34751, Jul. 2019.

[143] P. Drozdowski, F. Struck, C. Rathgeb, and C. Busch, "Detection of glasses in near-infrared ocular images," in *Proc. IEEE Int. Conf. Biometr. (ICB)*, Feb. 2018, pp. 202–208.

[144] D. Osorio-Roig, P. Drozdowski, C. Rathgeb, A. Morales-González, E. Garea-Llano, and C. Busch, "Iris recognition in visible wavelength: Impact and automated detection of glasses," in *Proc. IEEE Int. Conf. Signal Image Technol. Internet Syst. (SITIS)*, Nov. 2018, pp. 542–546.

[145] S. E. Baker, A. Hentz, K. W. Bowyer, and P. J. Flynn, "Degradation of iris recognition performance due to non-cosmetic prescription contact lenses," *Comput. Vis. Image Understand.*, vol. 114, no. 9, pp. 1030–1044, Sep. 2010.

[146] E. Dolgin, "The myopia boom," *Nature*, vol. 519, no. 7543, pp. 276–278, Mar. 2015.

[147] C. Rathgeb, A. Dantcheva, and C. Busch, "Impact and detection of facial beautification in face recognition: An overview," *IEEE Access*, vol. 7, pp. 152667–152678, 2019.

[148] Y. Koda, A. Takahashi, K. Ito, T. Aoki, S. Kaneko, and S. M. Nzou, "Development of 2,400 ppi fingerprint sensor for capturing neonate fingerprint within 24 hours after birth," in *Proc. IEEE Int. Conf. Biometr. Special Interest Group (BIOSIG)*, Sep. 2019, pp. 1–7.

[149] R. Haraksim, J. Galbally, and L. Beslay, "Fingerprint growth model for mitigating the ageing effect on children's fingerprints matching," *Pattern Recognit.*, vol. 88, pp. 614–628, Apr. 2019.

[150] S. Ribarić and N. Pavešić, "De-identification for privacy protection in biometrics," in *User-Centric Privacy and Security in Biometrics*. London, U.K.: Inst. Eng. Technol., Nov. 2017, pp. 293–324.

[151] S. Chhabra, R. Singh, M. Vatsa, and G. Gupta, "Anonymizing *k*-facial attributes via adversarial perturbations," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, Jul. 2018, pp. 656–662.

[152] V. Mirjalili, S. Raschka, A. Namboodiri, and A. Ross, "Semi-adversarial networks: Convolutional autoencoders for imparting privacy to face images," in *Proc. Int. Conf. Biometr. (ICB)*, Feb. 2018, pp. 82–89.

[153] A. Morales, J. Fierrez, and R. Vera-Rodriguez, "SensitiveNets: Learning agnostic representations with application to face recognition," Feb. 2019. [Online]. Available: arXiv:1902.00334.

[154] K. Ricanek and T. Tesafaye, "MORPH: A longitudinal image database of normal adult age-progression," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit. (FGR)*, Apr. 2006, pp. 341–345.

[155] G. Azzopardi, A. Greco, and M. Vento, "Gender recognition from face images with trainable COSFIRE filters," in *Proc. IEEE Int. Conf. Adv. Video Signal Surveillance (AVSS)*, Aug. 2016, pp. 235–241.

[156] M. Merler, N. Ratha, R. S. Feris, and J. R. Smith, "Diversity in faces," Apr. 2019. [Online]. Available: arXiv:1901.10436.

[157] K. Kärkkäinen and J. Joo, "FairFace: Face attribute dataset for balanced race, gender, and age," Aug. 2019. [Online]. Available: arXiv:1908.04913.

[158] K. Holstein, J. W. Vaughan, H. Daumé, M. Dudik, and H. Wallach, "Improving fairness in machine learning systems: What do industry practitioners need?" in *Proc. ACM Conf. Human Factors Comput. Syst. (CHI)*, May 2019, pp. 1–16.

[159] M. Du, F. Yang, N. Zou, and X. Hu, "Fairness in deep learning: A computational perspective," Aug. 2019. [Online]. Available: arXiv:1908.08843.

[160] N. Diakopoulos, "Accountability in algorithmic decision making," *Commun. ACM*, vol. 59, no. 2, pp. 56–62, Jan. 2016.

[161] J. A. Kroll, S. Barocas, E. W. Felten, J. R. Reidenberg, D. G. Robinson, and H. Yu, "Accountable algorithms," *Univ. Pennsylvania Law Rev.*, vol. 165, p. 66, Nov. 2016.

[162] M. T. Ribeiro, S. Singh, and C. Guestrin, "'why should I trust you?': Explaining the predictions of any classifier," in *Proc. ACM Int. Conf. Knowl. Disc. Data Min.*, 2016, pp. 1135–1144.

[163] J. New and D. Castro, "How policymakers can foster algorithmic accountability," Center Data Innov., Washington, DC, USA, Rep., May 2018.

[164] S. Verma and J. Rubin, "Fairness definitions explained," in *Proc. ACM Int. Workshop Softw. Fairness (FairWare)*, May 2018, pp. 1–7.

[165] B. Hutchinson and M. Mitchell, "50 years of test (un)fairness: Lessons for machine learning," in *Proc. ACM Conf. Fairness Accountability Transparency (FAT)*, Jan. 2019, pp. 49–58.

[166] M. Hardt, E. Price, and N. Srebro, "Equality of opportunity in supervised learning," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Dec. 2016, pp. 3315–3323.

[167] P. Saleiro *et al.*, "Aequitas: A bias and fairness audit toolkit," Nov. 2018. [Online]. Available: arXiv:1811.05577.

[168] R. K. E. Bellamy *et al.*, "AI fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias," Oct. 2018. [Online]. Available: arXiv:1810.01943.

[169] Z. Tufekci, "Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency," *J. Telecommun. High Technol. Law*, vol. 13, p. 203, 2015.

[170] S. Corbett-Davies, E. Pierson, A. Feller, S. Goel, and A. Huq, "Algorithmic decision making and the cost of fairness," in *Proc. ACM Int. Conf. Knowl. Disc. Data Min.*, Aug. 2017, pp. 797–806.

[171] FRA Focus, "BigData: Discrimination in data-supported decision making," Eur. Union Agency Fundam. Rights, Vienna, Austria, Rep. TK-02-18-634-EN-N, May 2018.

[172] FRA Focus, "Data quality and artificial intelligence—Mitigating bias and error to protect fundamental rights," Eur. Union Agency Fundam. Rights, Vienna, Austria, Rep. TK-01-19-330-EN-N, Jun. 2019.

[173] A. Torralba and A. A. Efros, "Unbiased look at dataset bias," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 1521–1528.

[174] R. Zemel, Y. Wu, K. Swersky, T. Pitassi, and C. Dwork, "Learning fair representations," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Feb. 2013, pp. 325–333.

[175] S. Shaikh, H. Vishwakarma, S. Mehta, K. R. Varshney, K. N. Ramamurthy, and D. Wei, "An end-to-end machine learning pipeline that ensures fairness policies," Oct. 2017. [Online]. Available: arXiv:1710.06876.

[176] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, *Learning From Imbalanced Data Sets*. Cham, Switzerland: Springer, Oct. 2018.

[177] B. H. Zhang, B. Lemoine, and M. Mitchell, "Mitigating unwanted biases with adversarial learning," in *Proc. ACM Conf. AI Ethics Soc. (AIES)*, Feb. 2018, pp. 335–340.

[178] P. C. Roy and V. N. Boddeti, "Mitigating information leakage in image representations: A maximum entropy approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2586–2594.

[179] S. A. Friedler, C. Scheidegger, and S. Venkatasubramanian, "On the (im)possibility of fairness," Sep. 2016. [Online]. Available: arXiv:1609.07236.

[180] B. Green and L. Hu, "The myth in the methodology: Towards a recontextualization of fairness in machine learning," in *Proc. Int. Conf. Mach. Learn. Debates Workshop*, Jul. 2018, pp. 1–5.

[181] L. T. Liu, S. Dean, E. Rolf, M. Simchowitz, and M. Hardt, "Delayed impact of fair machine learning," in *Proc. Int. Joint Conf. Artif. Intell.*, Aug. 2018, pp. 6196–6200.

[182] E. Marasco, "Biases in fingerprint recognition systems: Where are we at?" in *Proc. IEEE Int. Conf. Biometr. Theory Appl. Syst. (BTAS)*, Sep. 2019, pp. 1–5.

[183] K. Kirkpatrick, "Battling algorithmic bias: How do we ensure algorithms treat us fairly?" *Commun. ACM*, vol. 59, no. 10, pp. 16–17, Oct. 2016.

[184] N. Hallowell, L. Amoore, S. Caney, and P. Waggett, "Ethical issues arising from the police use of live facial recognition technology," Interim Report of the Biometrics and Forensics Ethics Group Facial Recognition Working Group, Rep., Feb. 2019.

[185] E. J. Kindt, *Privacy and Data Protection Issues of Biometric Applications*. Dordrecht, The Netherlands: Springer, 2016.

[186] X. Wu and X. Zhang, "Automated inference on criminality using face images," Nov. 2016. [Online]. Available: arXiv:1611.04135.

[187] R. A. Calvo, D. Peters, and S. Cave, "Advancing impact assessment for intelligent systems," *Nat. Mach. Intell.*, vol. 2, no. 2, pp. 89–91, Feb. 2020.

[188] J. Bryson and A. Winfield, "Standardizing ethical design for artificial intelligence and autonomous systems," *Computer*, vol. 50, no. 5, pp. 116–119, May 2017.

[189] European Parliament, "Regulation (EU) 2016/679," *Official J. Eur. Union*, vol. L119, pp. 1–88, Apr. 2016.

[190] B. Goodman and S. Flaxman, "European Union regulations on algorithmic decision-making and a right to explanation," Aug. 2016. [Online]. Available: arXiv:1606.08813.

[191] E. Awad *et al.*, "The moral machine experiment," *Nature*, vol. 563, no. 7729, pp. 59–64, Nov. 2018.

[192] European Union Agency for Fundamental Rights and Council of Europe, *Handbook on European Non-Discrimination Law*. Eur. Union, Brussels, Belgium, Feb. 2018.

[193] A. P. Nutman, V. C. Bennett, C. R. Friend, M. J. Van Kranendonk, and A. R. Chivas, "Algorithm and blues," *Nature*, vol. 537, no. 7621, pp. 535–538, Sep. 2016.

[194] D. K. Citron, "Technological due process," *Washington Univ. Law Rev.*, vol. 85, p. 66, Sep. 2007.

[195] D. K. Citron and F. A. Pasquale, "The scored society: Due process for automated predictions," *Washington Law Rev.*, vol. 89, pp. 1–34, Apr. 2014.

Antitza Dantcheva received the Ph.D. degree in signal and image processing from Eurecom/Telecom ParisTech, Paris, France, in 2011.

He is a Research Scientist (CRCN) with the STARS Team, Inria, Sophia Antipolis, France. She has received the French National Research Agency JCJC Young Researcher Grant in 2017. Her research interests include computer vision and face analysis, where she has been working in appearance and dynamic analysis for healthcare and security.

Dr. Dantcheva was a recipient of the Best Presentation Award (ICME'11), the Best Poster Award (ICB'13), the Tabula Rasa Spoofing Award in 2013, the Best Paper Award (Runner Up) (ISBA'17), and the Best Poster Award (FG'19). She was in the winning team of the ECCV 2018 Challenge on Bias Estimation in Face Analysis.

Naser Damer received the Ph.D. degree in computer science from the Technical University of Darmstadt, Darmstadt, Germany, in 2018.

He is a Senior Researcher with Fraunhofer IGD, Darmstadt. He lectures with the Technical University of Darmstadt. He is a Principal Investigator with the National Research Center for Applied Cybersecurity ATHENE, Darmstadt. His main research interests lie in the fields of biometrics, machine learning, and information fusion. He published more than 60 scientific papers in the above fields.

Dr. Damer serves as an Associate Editor for *Visual Computer* and represents the German Institute for Standardization (DIN) in ISO/IEC SC37 Biometrics Standardization Committee.

Pawel Drozdowski is currently pursuing the Ph.D. degree with the Norwegian University of Science and Technology, Gjøvik, Norway.

He is a Researcher with the Faculty of Computer Science, Hochschule Darmstadt, Darmstadt, Germany. He has coauthored over 15 technical publications in the field of biometrics.

Mr. Drozdowski won the Best Student Paper Runner-Up Award (WIFS'18) and the Best Poster Award (BIOSIG'19). He represents the German Institute for Standardization (DIN) in ISO/IEC SC37 JTC1 SC37 on biometrics.

Christian Rathgeb received the Ph.D. degree in computer science from the University of Salzburg, Salzburg, Austria, in 2012.

He is a Senior Researcher with the Faculty of Computer Science, Hochschule Darmstadt, Darmstadt, Germany. He is a Principal Investigator with the National Research Center for Applied Cybersecurity ATHENE, Darmstadt. He has coauthored over 100 technical publications in the field of biometrics. His research includes pattern recognition, iris and face recognition, security aspects of biometric systems, secure process design, and privacy enhancing technologies for biometric systems.

Dr. Rathgeb is a winner of the European Biometrics Research Award in 2012, the Austrian Award of Excellence in 2012, the Best Poster Paper Awards (IJCB'11, IJCB'14, and ICB'15), and the Best Paper Award Bronze (ICB'18).

Christoph Busch received the Ph.D. degree in computer graphics from the Technical University of Darmstadt, Germany, in 1997.

He is a Member of the Norwegian University of Science and Technology, Gjøvik, Norway. He holds a joint appointment with Hochschule Darmstadt, Darmstadt. Furthermore, he has been lecturing Biometric Systems with the Technical University of Denmark, Lyngby, Denmark, since 2007. He has coauthored more than 500 technical papers and has been a speaker with international conferences.

Dr. Busch served for several conferences and journals as a reviewer (e.g., ACM-SIGGRAPH, ACM-TISSEC, IEEE CG&A, and IEEE PAMI). He is also an appointed member of the editorial board of *IET Biometrics* and the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY. Furthermore, he chairs the German Biometrics Working Group and is a Board Member of the European Association for Biometrics. He chairs the German standardization body on Biometrics (DIN—NIA37) and is a convenor of WG3 in ISO/IEC JTC1 SC37 on Biometrics.