Privacy–Enhancing Face Biometrics: A Comprehensive Survey

Blaž Meden, Peter Rot, Philipp Terhörst, Naser Damer, Arjan Kuijper, Walter J. Scheirer, Arun Ross, Peter Peer and Vitomir Štruc

Abstract—Biometric recognition technology has made significant advances over the last decade and is now used across a number of services and applications. However, this widespread deployment has also resulted in privacy concerns and evolving societal expectations about the appropriate use of the technology. For example, the ability to automatically extract age, gender, race, and health cues from biometric data has heightened concerns about privacy leakage. Face recognition technology, in particular, has been in the spotlight, and is now seen by many as posing a considerable risk to personal privacy. In response to these and similar concerns, researchers have intensified efforts towards developing techniques and computational models capable of ensuring privacy to individuals, while still facilitating the utility of face recognition technology in several application scenarios. These efforts have resulted in a multitude of privacy–enhancing techniques that aim at addressing privacy risks originating from biometric systems and providing technological solutions for legislative requirements set forth in privacy laws and regulations, such as GDPR. The goal of this overview paper is to provide a comprehensive introduction into privacy–related research in the area of biometrics and review existing work on *Biometric Privacy–Enhancing Techniques* (B–PETs) applied to face biometrics. To make this work useful for as wide of an audience as possible, several key topics are covered as well, including evaluation strategies used with B–PETs, existing datasets, relevant standards, and regulations and critical open issues that will have to be addressed in the future.

Index Terms—Biometrics, face recognition, privacy, privacy–enhancing techniques.

1 Introduction

Face recognition, now in widespread use, provides an excellent measure of security and convenience to users — *but at what cost to privacy?* By definition, biometric data is linked to distinct individuals and, consequently, contains a considerable amount of personal information that can be extracted and automatically inferred from the data itself. While biometric data was traditionally collected, processed, and stored by dedicated biometric systems, the widespread availability of mobile devices, consumer cameras, and other forms of imaging technology has made it easier than ever to capture and share biometric data online. When it comes to faces, billions of images (and videos) are being uploaded to various social media platforms, such as Facebook, Instagram, Twitter, or YouTube, on a monthly basis, creating massive image collections that can be processed and analyzed using sophisticated computer vision and biometric recognition techniques. As these collections can easily be used to infer sensitive information about individuals, to profile selected users or even identify people and match them against other existing databases (without their knowledge and consent), such collections represent a considerable privacy risk. [1].

- *B. Meden and P. Peer are with the Faculty of Computer and Information Science, University of Ljubljana. E-mail: blaz.meden@fri.uni-lj.si*
- *P. Rot and V. Štruc are with the Faculty of Electrical Engineering, University of Ljubljana.*
- *P. Terhörst, N. Damer and A. Kuijper are with the Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany, and also with the Department of Computer Science, Technical University of Darmstadt, Darmstadt, Germany.*
- *W. J. Scheirer is with the University of Notre Dame*
- *A. Ross is with Michigan State University*

Manuscript received February, 2021; revised: right away.

To illustrate the potential privacy–related pitfalls associated with image collections gathered by existing platforms, consider the 2020 news coverage of *Clearview AI*, a private company that scraped facial images from social media sites and created large–scale face recognition technology capable of matching images of subjects to their social media accounts and linking individuals across various databases and services [2]. The fact that Clearview AI monetized personal (biometric) data without people's consent made headlines across news media and raised social awareness about the capabilities of existing face recognition technology and the privacy intrusions the technology can cause. Examples like this clearly demonstrate the need for (privacy–oriented) mechanisms that can efficiently protect biometric data and limit the amount of information that can be inferred from it using (semi–)automatic recognition techniques.

Privacy concerns over biometric recognition systems are, however, not limited only to social media. Similar concerns can also been voiced for other areas, such as smart phones, mobile applications, smart camera networks, IoT, e–commerce sites, health services, and related application domains. In fact, face recognition technology and its use by law enforcement have recently become so controversial (due to fairness and privacy concerns) that major software companies, such as Microsoft, Amazon, and IBM, are currently considering putting a moratorium on their face recognition programs [3]. These developments and the associated societal expectations about the appropriate use of biometric recognition technology have resulted in a growing interest of the research community in *technological* solutions capable of balancing the benefits of biometric systems on the one hand, and the privacy of individuals on the other. Such solutions could, for instance, facilitate the use of face

(a) Original (b) Blurred (c) Pixelated (d) Deidentified

Fig. 1: Examples of the visual effect of different biometric privacy enhancing techniques (B–PETs) applied to face images. B–PETs aim to remove sensitive information from facial images to protect privacy. In the presented examples, both (b) blurring and (c) pixelation try to conceal identity information, but do so unselectively – all information about the subject is concealed. The more recent deidentification procedure (based on face swapping) in (d), on the other hand, conceals identity information, but preserves other relevant information contained in the data, such as gender, facial expression, pose, and other similar visual attributes. The goal of this paper is to present a comprehensive review of the field of privacy enhancement in the context of biometric systems – with a focus on face biometrics.

recognition in criminal cases that involve large collections of face images from the Internet, e.g., human trafficking, sex crimes, and war crimes, but do so in a manner that does not compromise the privacy of innocent subjects in recovered photos.

Issues related to privacy in biometric recognition systems are also being addressed by governments around the globe, which are starting to regulate the collection, use, processing, and storage of biometric data. The European Union (EU), for example, in 2016 passed the General Data Protection Regulation (GDPR)[1], which defines rules for data protection (including biometric data) and security across Europe and foresees heavy fines for noncompliance. GDPR considers biometric data within a special category of personal data that requires the highest level of protection and allows processing of such data only in a limited number of scenarios and only with appropriate privacy–related safeguards in place. Similar legislative developments are also underway in the United States. In California, for example, the California Consumer Privacy Act (CCPA) came into effect in 2018 and is considered by many the potential model for a US–wide data privacy law [4]. Similarly to GDPR, CCPA includes provisions that affect access to, storage, and handling of biometric data. Illinois, Washington, and Texas all have similar legislation already in place, with other US states expected to follow. While the goal of these laws and privacy acts is to provide a legal framework for the use of biometric systems, they also capitalize on the importance (and expectations of legislators) of addressing biometric privacy from a technological perspective.

A considerable amount of research has been done over recent years to address the privacy–related issues, needs, and legislative requirements (associated with biometric data) discussed above, including work on imaging sensors with built–in privacy protection [5], [6], deidentification techniques for biometric data [7], [8], [9], [10], adversarial approaches capable of confounding (automatic) recognition techniques [11], [12], [13], schemes that allow for privacy–preserving data sharing [14], [15], template protection techniques [16], cancelable biometrics [17], [18], and others. A significant portion of this work share a common characteristic in that they try to mitigate privacy concerns by reducing the biometric utility of the captured data. In other words, they attempt to remove (or conceal) certain (sensitive) information from the biometric data, while leaving other useful information unchanged. Techniques that follow this idea are related to the *data minimization* principle set forth in GDPR, which suggests that the amount of personal data that needs to be processed for a specific purpose must be kept minimal. Since different terminology is used in the literature to describe such techniques, we refer to them collectively as *Biometric Privacy–Enhancing Techniques* (B–PETs) in this survey [12], [19], [20], [21], [22].

A few examples of B–PETs applied to facial images are presented in Fig. 1. Here, the input face image in Fig. 1 (a) is processed with a blurring and pixelation operation in Fig. 1 (b) and (c), respectively. While such an approach contributes towards privacy protection, it also renders all other information contained in the image more or less useless. The more recent B–PET in Fig. 1 (d) conceals identity information, but also preserves other visual attributes, such as gender, facial expression, pose, and the like. Thus, it offers a compromise between privacy and the biometric utility of the data. Research in biometric privacy enhancement is today focused on such selective techniques, able to remove (or conceal) specific biometric attributes, while not affecting others, on solutions that offer provable privacy guarantees, on techniques that confound automatic recognition approaches and related privacy challenges applied not only to visual data, but also other components and data representations within biometric recognition systems. While the interest in privacy enhancing techniques is growing and substantial research

1. Accessible from: https://gdpr-info.eu/

effort has been directed towards solutions in this area, a broad and well–structured overview on this important and timely research topic is still missing from the literature.

In this paper we aim to fill this gap and present a comprehensive overview on privacy enhancement in the context of facial biometrics. We focus on facial imagery because of the popularity and widespread deployment of face recognition technology, wide scope of applications facilitated by facial images and broad spectrum of privacy concerns associated with facial biometrics. However, the discussion across a significant portion of the paper is intentionally kept general, as many of the concepts addressed apply not only to facial data, but also to other types of biometric data. The survey is meant to be more than just a summary of existing privacy–enhancing techniques and aims to *i)* introduce the reader to the main concepts, characteristics, and challenges associated with privacy enhancement (and protection) of biometric data, *ii)* define a taxonomy for existing biometric privacy–enhancing techniques, *iii)* outline issues associated with quantifying performance, and *iv)* point the reader to important regulations and standards. As such, the goal of the survey is to serve as a reference document for researchers from the biometrics and computer vision communities, but also for privacy scholars, legislators, and practitioners interested in the topic of biometric privacy.

Existing surveys related to biometric privacy address important topics, such as template protection [16][2], deidentification [7], [8], [23], visual privacy in the context of social media [24], and related areas [25], privacy in video surveillance systems [26], [27], [28] and video redaction [29], This work complements the listed surveys and makes the following distinct contributions:

- It presents a comprehensive introduction into the field of privacy enhancement as applied to facial imagery and introduces a taxonomy of existing biometric privacy enhancing techniques (B–PETs).
- It present a review of over 200 references related to B–PETs, and provides an in–depth analysis of the surveyed methods.
- It identifies open issues and challenges that need to be addressed in the future.

2 The Facets of Biometric Privacy

2.1 Privacy and Biometric Systems

The topic of privacy has historically been studied across multiple disciplines, including social, legal, and political sciences, philosophy, ethics, and more recently in the context of information technology and biometric systems. Since the concept of privacy is relevant to several fields, no universally applicable definition exists and the exact meaning of the term "privacy" very much depends on the context in which it is studied. Clarke [48], for example, suggests to avoid the common notion of privacy as some sort of (moral or legal) *right*, but to rather think of it as an *"interest that individuals have in sustaining personal space, free from interference by other people and organizations"*. Alongside this definition, Clarke also provides a categorization of the different types of privacy [48], [49]:

- *Privacy of the person (or bodily privacy)*, which is related to the integrity of a person's body. Threats to this type of privacy include physical intrusions, such as torture or compulsory medical treatment, immunisation, and, as noted by Finn et al. [50], also forced biometric measurement.
- *Privacy of personal behaviour (or media privacy)*, which is concerned with sensitive behavioral information, such as political activities, sexual habits, or religious practices, but also with the personal space (private or public) needed to facilitate such behaviour [49].
- *Privacy of personal communications (or interception privacy)*, which is associated with the ability to communicate freely using various means (e.g., verbal, written, gestured, electronic) without being monitored by third parties [48].
- *Privacy of personal data (or data privacy)*, which is related to the general availability of personal data and the ability to execute control over one's personal data and its use by third parties. This type of privacy is (often jointly with privacy of personal communications) also referred to as *information privacy* [48], [50].

Modern biometric systems are, in general, able to extract a considerable amount of information about multiple attributes (e.g., identity, soft biometric attributes, health–related attributes, behavioural attributes) from biometric data and can, therefore, impact privacy across all categories listed above. However, they are most often associated with risks related to data or information privacy. Data privacy is usually the main type of privacy regulated within existing data–protection legislation and various privacy laws (e.g., GDPR).

Clarke's definition of privacy as an *interest of individuals* implies that it has to be balanced against other interests (e.g., personal interests, third party interests, societal interests), which may be of economical, political, societal (e.g, national security, crime investigations), or technological nature [48]. The process of balancing privacy against such competing interests is called *privacy protection*.

The literature on privacy protection in the context of biometrics is split between two main research directions:

- *Data security:* Solutions from this group are usually concerned with data encryption schemes and aim to mitigate risks associated with unauthorized data access or interception of data during transmission. Examples of solutions from this group include template protection schemes and cancelable biometrics [16], [17], [18], [51], [52], [53]. Such solutions typically aim to secure the data, but not necessarily the information that can be extracted from the data. Thus, if the biometric matching (or classification) is not conducted in the encrypted domain, sensitive information may still get exposed during decryption.
- *Privacy enhancement:* Solutions from this group (which are also at the heart of this paper) try to modify the biometric data or elements of biometric systems to limit the amount of information that

2. Note that this topic is not covered by our definition of B–PETs, as it follows different assumptions.

TABLE 1: Overview of the main terminology used in the literature to describe techniques that try to contribute towards privacy protection by reducing the biometric utility of biometric data. In this survey we use *Biometric Privacy Enhancing Techniques (B–PETs)* as an umbrella term to describe such approaches.

Process (technique)	Explanation and examples	Observation, specifics
Deidentification	Ribaric et al. [8] define *deidentification* as *"the process of concealing or removing personal identifiers (information that can be used for identification), or replacing them with surrogate personal identifiers in biometric data, in order to prevent the disclosure and use of data for purposes unrelated to the purpose for which the data was originally collected."* Examples of deidentification technique applied to biometric data include [15], [30], [31], [32], [33].	Most often used to describe techniques that try to remove (or conceal) identity from biometric data, but sometimes also target other attributes, such as soft–biometrics.
Anonymization	In the context of biometrics the term *anonymization* is used in multiple contexts. The term is often considered to relate to anynomization of identity and describe an irreversible form of deidentification [8], [23]. However, it is also used to indicate the reduction of biometric utility in terms of soft–biometrics – e.g., [13]. In [34] anonymization is defined as *"a process which enhances privacy by reducing the uniqueness of an identity."*	The term is not used consistently in the biometric literature, but is often considered to be close to (or a synonym for) deidentification.
Redaction	In the computer vision literature the term *redaction* refers to technology that can remove certain privacy–sensitive aspects of visual data [29], [35]. It describes techniques that go beyond solely (human–centric) biometrics and may apply to processing of other visual categories, such as text, license plates, images of identity documents, etc.	When applied to biometric data (such as faces) redaction techniques use simple processes such as blurring or masking to reduce utility.
Obfuscation	The term *obfuscation* is often used to describe techniques that alter biometric data in a way that causes miss detections, or in other words, helps individuals to avoid biometric recognition, but also in the context of privacy protection [36]. Obfuscation techniques have been studied across different biometric modalities, including speech [37], fingerprints [38], irises [39], and facial images [40].	Similarly to deidentification reduces biometric utility, but commonly with the goal of evading biometric recognition.
Obscuration	*Obscuration* tries do reduce biometric utility by obscuring certain aspects of biomateric data. The term appears in the literature as a synonym for techniques that conduct deidentification and is often used interchangeably with deidentification. It is popular mostly with face–related research, e.g., [41], [42], [43].	An alternative term for deidentification used mainly to describe techniques that try to obscure identity information in facial images.
Soft–biometric privacy enhancement	*Soft–biometric privacy enhancement* refers to a group of methods that tries to remove or suppress information about soft–biometric attributes in biometric data – see e.g., [20], [44], [45], [46]. Techniques from this group are sometimes also called deidentification techniques for soft–biometric identifiers [8].	Describes techniques that reduce the utility of biometric data by targeting soft–biometric attributes, such as age, gender, and ethnicity.
Controllable privacy	The concept of *controllable privacy*, introduced in [47], characterizes targeted privacy enhancing techniques that allow to control which specific aspects (and to what extend) of the biometric data should be removed or concealed. Techniques in line with this concept allow the user to explicitly define, which attributes of the biometric data to conceal and which to preserve.	The concept is related to the characteristics of privacy enhancement and applies to different attributes of the biometric data, e.g., identity, age, gender, ethnicity.

can be extracted automatically. Thus, they try to balance privacy against the biometric utility of the data. Examples include data masking approaches, various obfuscation schemes, deidentification techniques, solutions ensuring so–called soft–biometric privacy and other related approaches [7], [8], [11], [45], [54].

Note that solutions from the two groups are by no means mutually exclusive and can in practice be combined to improve both security as well as privacy in biometric systems.

Research on the topics discussed above is critical for addressing potential privacy risks originating from biometric recognition systems. As noted by Solove [55], privacy threats can in general arise during *i)* data collection, *ii)* storage, and processing as well as during *iii)* data dissemination. In the context of biometric systems these (direct and indirect) threats most often relate to [55], [56]:

- *Linkage (direct):* Biometric data can be matched across different datasets (e.g., social media, police records, financial services, health databases), linking various sources of information and generating aggregated data that is often greater than the sum of its parts [57]. Such linked data may disclose personal information not available in any of the individual datasets, posing considerable privacy risks to individuals.
- *Secondary use or function creep (direct):* When biometric data is collected, consent is typically given for the data to be used for a specific purpose (e.g., authentication, time–and–attendance monitoring). However, the data may also be used for alternative purposes, such as for the extraction of sensitive information (e.g., health status, ethnic origin, age estimation) that enables secondary use cases, such as user profiling, targeted advertising, and other unsolicited applications that intrude on individuals' privacy [58]. Secondary use in the context of biometrics is also related to unsolicited *information disclosure* – the disclosure of true facts about individuals. For example, health information or other sensitive information extracted from biometric data can raise serious privacy concerns when shared with third parties.
- *Insecurity (indirect):* Biometric data needs to be kept secure to avoid *data breaches* that may eventually lead to privacy threats either through data linkage or function creep effects. Data insecurity in biometric systems may result in unauthorized access to financial services, social media, mobile phones, and personal computers, and impact privacy across different privacy categories discussed above.
- *Exclusion (indirect):* Third parties collecting, processing, or storing biometric data sometimes fail to inform individuals about the fact that they have access to individuals' biometric data. This exclusion poses privacy threats with respect to unauthorized use of the data, disclosure of sensitive information, or linkage and is therefore also often addressed by privacy laws.

For a more comprehensive coverage of existing privacy threats, the reader is referred to the work of Solove [55].

2.2 Biometric Privacy Enhancement

Privacy protection in datasets containing structured data (e.g., medical, school, or government records) is usually addressed by data anonymization or redaction techniques that selectively remove (replace, conceal) sensitive information and only retain information needed for a particular purpose [23]. While such techniques have been found useful tools for privacy protection, they are not easily applicable to biometric data.

Biometric data is by default non–deterministic (and unstructured) and commonly processed using automated machine learning techniques. It is, therefore, often not clear what part (and particular aspect) of the data is used to extract or infer potentially sensitive information. Consider facial images, for example. A classifier aiming to extract gender or ethnicity information may use spatially local information, such as the shape of the eyes or mouth, but also global characteristics, such as facial appearance, face geometry or skin–tone. Removing (or concealing) sensitive information is therefore not a straight forward task. This process becomes even more challenging when derived data representations are considered, i.e., templates. As shown by recent work [59], [60], contemporary biometric templates encode a variety of information that is entangled in the features and, therefore, difficult to remove or hide in a selective manner.

One possibility to address these issues is to completely remove all information that could be used by biometric recognition techniques. At the image–level, for example, early privacy protection methods tried to hide sensitive information by placing black patches over individuals or replacing image regions corresponding to individuals by uninformative surrogate images [10], [15], [61]. While such approaches ensure perfect privacy protection, they also completely destroy the utility of the data. No useful information remains in such "privacy protected" data, not even information that bears no privacy risks at all.

To overcome this limitation, more recent techniques try to strike a balance between privacy protection and the biometric utility of the data. Here, the term biometric utility is used in the broadest possible sense, to describe the usefulness of the data for automatic extraction of various attributes, including identity information, soft–biometrics, health indicators, behavioural cues, and other similar characteristics. The main idea behind these techniques is to selectively remove (or conceal, suppress, replace) information about specific (potentially sensitive) attributes[3], while preserving information on attributes not critical from a privacy perspective. While considerable research has been directed towards such techniques in recent years, the terminology used in the literature differs from paper to paper. Table 1 provides a summary of the main terminology used in this field and a brief explanation of the most important terms.

For the sake of consistency, we refer to the techniques listed in Table 1 collectively as *Biometric Privacy Enhancing Techniques (B–PETs)*. Thus, we use the term B–PETs to describe techniques that weigh privacy against biometric utility and illustrate this trade–off in Fig. 2. On one end of the trade–off we have raw (Fig. 2 right) unprocessed biometric data. Such data has complete biometric utility, but also poses the highest privacy risk. On the other end of the trade–off (Fig. 2 left) is biometric data that has no biometric utility (e.g., masked facial images) but, therefore, offers the highest level of privacy protection. In between the two ends is data, from which some biometric attributes can been inferred, while others cannot. Such data offers a compromise between biometric utility and privacy and is typically generated using biometric privacy–enhancing techniques.

3. Which attributes are targeted for removal (concealment, suppression) is defined by the application domain and purpose, for which the data were initially collected.

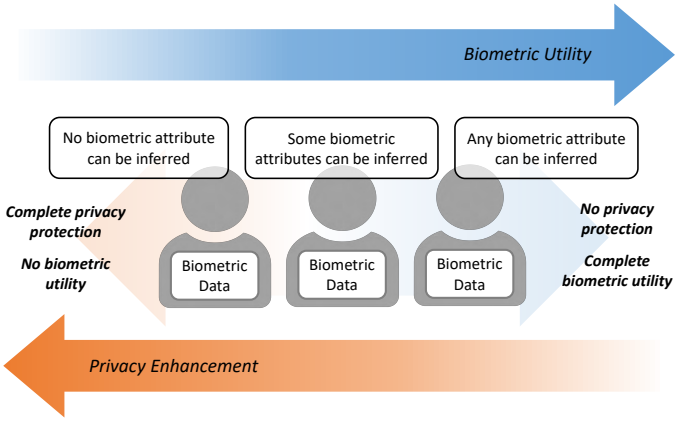


Fig. 2: Biometric data (e.g., facial images) contains a considerable amount of potentially sensitive personal information that can be inferred from the data automatically using modern machine learning techniques. Biometric privacy enhancing techniques (B–PETs) try to address privacy concerns associated with such data by reducing its biometric utility, i.e., by removing sensitive information from the data, while preserving other useful information.

2.3 Characteristics of B–PETs

Existing B–PETs can be grouped according to a number of different criteria, including *i)* the data they apply to, *ii)* the type of mapping they use, *iii)* the biometric attributes they target, *iv)* the way biometric utility is addressed, *v)* the guarantees they provide regarding the possibility of reconstructing information from the privacy enhanced data, and *vi)* whether biometric attributes are concealed from humans and/or machines. An overview of this categorization is presented in Fig. 3.

2.3.1 Input data

The first categorization of B–PETs relates to the type input data that is used as the basis for privacy enhancement. Here, B–PETs can be applied to either *videos* or *still images*. The type of input data affects a number of aspects that B–PETs need to consider.

With video sequences, for example, it is crucial that (all) biometric data in every frame is subjected to privacy enhancement. If only a single instance of the data (e.g., a face) is not processed, the video still poses a privacy risk and may still allow to extract sensitive information about individuals. Due to this characteristic, simple techniques

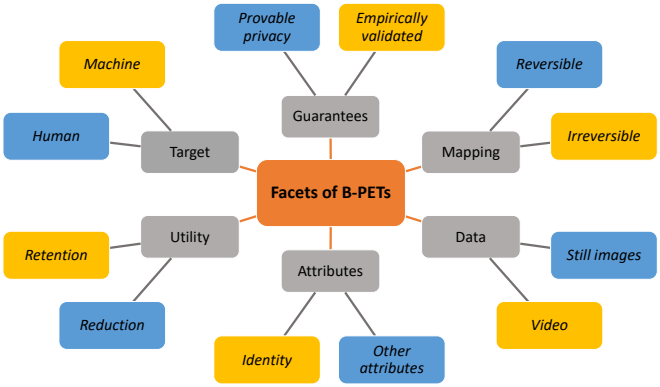


Fig. 3: The facets of biometric privacy enhancing techniques (B–PETs). Existing B–PETs exhibit different characteristics and can be categorized according to various criteria – criteria are shown in grey (best viewed in color).

such as blurring and masking (e.g., [62], [63]) are most often used in conjunction with robust object detectors for privacy enhancement in videos, as opposed to still images, where typically higher quality data is available that facilitates more elaborate B–PETs relying, for example, on different statistical models [10], [64], [65], generative deep–learning models [32], [33], or adversarial examples [11], [13].

Furthermore, if privacy enhancement is applied on derived representations (e.g., on templates) the type of input data may affect how these representations are defined. For static images, for example, a single feature representation (template) may be available, whereas videos may be represented using multiple feature representations, which warrants a different approach to privacy enhancement [20], [66].

2.3.2 Mapping

Privacy enhancement may at the coarsest level be defined as a mapping f that takes a biometric datum x as input and returns a privacy enhanced version x_E with possibly reduced biometric utility, i.e., $x_E = f(x)$. Depending on the type of mapping, B–PETs can be categorized as:

- *Irreversible:* B–PETs from this group usually apply a mapping that is difficult to invert. Once privacy enhancement is applied to biometric data, the original data is hard (or ideally impossible) to reconstruct. Note, however, that the irreversiblity is often not guaranteed, as for example, with template protection methods, because of the requirement to preserve some of the biometric utility of the original data.
- *Reversible:* B–PETs from this group incorporate mechanisms that allow them to invert the mapping used for privacy enhancement. A common approach here is to encode information about the privacy enhancement process or the original data using cryptography and then reverse the mapping if needed using data decryption, see e.g., [36], [67]. Because it is possible to reverse the privacy enhancement, reversible B–PETs depend heavily on the security of the adopted cryptographic solutions.

2.3.3 Biometric Attributes

The primary goal of biometric recognition systems is to link biometric traits to individuals. Consequently, the focus of the majority of existing B–PETs is on removing (or concealing) identity information from biometric data with the goal of privacy protection. This is also evidenced by the vast body of work on deidentification techniques, which mostly targets so called *personal identifiers* for deidentification, e.g., [10], [64], [65], [68]. Here, the term *personal identifier* is used in the privacy literature to characterize (any and all) information that can be used to link data to individuals.

More recently, researchers also started looking at privacy enhancement approaches that target and try to obfuscate soft–biometric attributes, such as gender, age, or ethnicity[4] – e.g., [44], [45], [70]. B–PETs from this group are, for example, relevant in the context of social media. People are in general willing to share their images and selfies (revealing their identity) online, but are less keen on the consequental privacy intrusions, such as targeted advertising that is often facilitated by an automatic analysis of the demographic attributes of the shared images. Attributes, such as soft–biometrics, are often referred to as *quasi–identifiers* by privacy scholars, because they provide partial information on individuals, but cannot be linked to a specific person in an unambiguous manner.

While other types of attributes, such as behavioural cues or health indicators could also be targeted by B–PETs, work in this direction is limited in the literature – see [71] for one of the few exceptions.

2.3.4 Biometric Utility

There are two different strategies used with existing B–PETs with respect to data utility:

- *Utility reduction:* The first strategy aims to modify the biometric data in such a way that sensitive information is removed, while other aspects of the data are preserved. Consider adversarial face deidentification techniques, such as [72], as an example. These techniques introduce minute changes to face images, so that automatic identity recognition is impaired, whereas other tasks (e.g., gender or age recognition) are still feasible. Earlier techniques that follow this strategy often reduce utility unselectively through blurring or masking, which typically makes it challenging to extract any meaningful information from the data. Regardless of whether this strategy is used in an attribute–selective manner or not, it produces privacy enhanced data with reduced biometric utility.
- *Utility retention:* The second strategy aims to generate surrogate data that corresponds to the original biometric data in a number of preselected attributes. This strategy is usually used with generative models, such as Generative Adversarial Networks (GANs),

4. The terms *gender* and *sex* have been used interchangeably in the biometric literature. Note that gender is a social or cultural construct, while sex is based on biological characteristics. Similarly, the terms *race* and *ethnicity* have also been used interchangeably in the literature. An exact definition of either of these two terms appears to be a subject of debate [69].

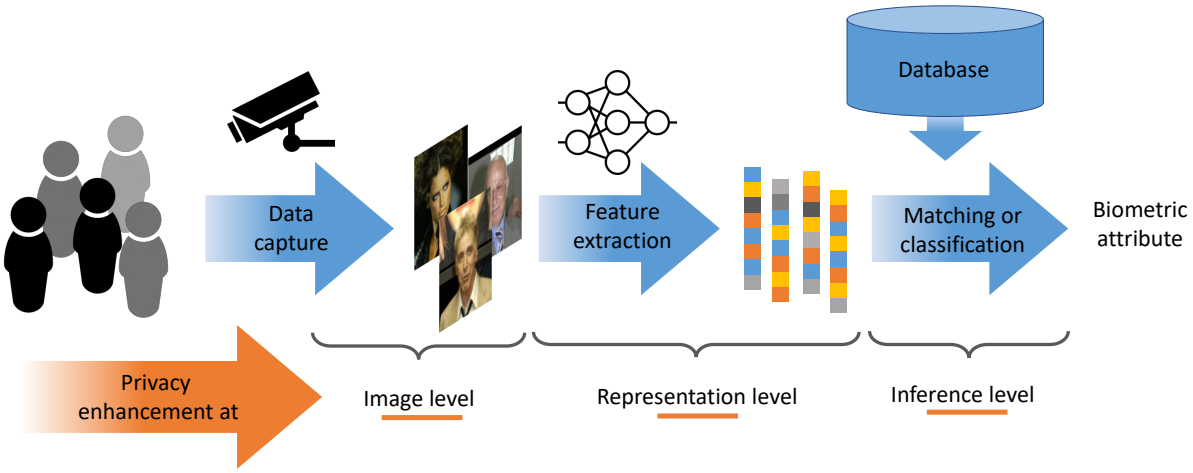


Fig. 4: Possible application points of biometric privacy enhancing techniques (B–PETs) in the context of biometric recognition systems. B–PETs can be applied at either the *image* level, *representation* level, or at the *inference* level. The main functional steps of a biometric system are shown in blue and the application points of B–PETs are depicted in orange. The figure is best viewed in color.

that are capable of generating synthetic data with predefined characteristics, e.g., [33]. With this strategy, specific attributes of the data are retained, while others are typically artificially generated.

2.3.5 Guarantees

One of the most important issues associated with B–PETs is how to quantify privacy–protection performance. This issue is related to the risk of inferring information about removed (or concealed) attributes[5], which is notoriously difficult to estimate in an objectively manner. One possibility used regularly in the literature is to empirically validate that information on certain attributes cannot be inferred from the privacy–enhanced data. This approach typically involves biometric recognition experiments with preselected matchers (or classifiers) using standard biometric datasets – see Section 4 for more information on performance evaluation.

However, since different types of biases may be involved in empirical evaluations, researchers are increasingly looking into privacy enhancing techniques that offer formal (quantifiable) privacy guarantees. B–PETs that offer such guarantees are usually referred to as techniques with *provable privacy*. While initial attempts using the concept of ϵ–*differential privacy* were recently presented, e.g., [73], [74], to ensure provable privacy with biometric data, most of the existing work on this topic centers around k–*anonymity*.

Here, k–anonymity [14], [15] is commonly used with B–PETs that aim to conceal identity information and has, to the best of our knowledge, not yet been extended to other visual attributes. With the k–anonymity model, privacy enhancement is defined over a closed set of N data samples, i.e., $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$, from which a new, privacy enhanced set of data $\mathcal{X}_\mathcal{P}$ is generated, i.e., $f : \mathcal{X} \mapsto \mathcal{X}_\mathcal{P} \in \mathbb{R}^N$. To provide anonymity guarantees, the mapping is typically implemented in such a way that groups of k samples from $\mathcal{X}$ are replaced by one and the same sample in $\mathcal{X}_\mathcal{P}$. As a result, only N/k distinct data samples are present in $\mathcal{X}_\mathcal{P}$ and the probability of linking any privacy enhanced sample from $\mathcal{X}_\mathcal{P}$ to the originals in $\mathcal{X}$ equals $1/k$. Note that this probability bound is only valid if the samples in $\mathcal{X}$ belong to exactly N distinct identities.

Existing incarnations of the k–anonymity model differ mostly in the definition of the surrogate samples that populate $\mathcal{X}_\mathcal{P}$. The seminal k–Same algorithm [15], for example, used centroids of clusters of k facial images as surrogate faces for $\mathcal{X}_\mathcal{P}$. Since the cluster centroids preserve visual elements of all k images of any given cluster, the surrogate faces retain some of the utility of the original raw facial images. Later techniques extended this idea beyond pixel averaging and defined cluster centroids on the parameters of generative models, moved beyond clusters and incorporated various strategies to preserve the biometric utility of the data. Examples of such extensions are, for instance, presented in [31], [64], [65], [75], [76].

We also note that for the task of deidentifying entries in relational databases the concept of k–anonymity was already extended to other provable privacy schemes, such as L–diversity [77], t–closeness [78], p–sensitive k–anonymity [79], and others. However, these privacy models have seen limited application on visual data so far.

2.3.6 Target

B–PETs can also be categorized based on the target of the privacy enhancement, i.e., *humans* or *machines*. While some algorithms are designed to ensure privacy from human observers, others are targeting automated machine learning models only. Deidentification techniques, such as [15], [31], [33], for example, typically change the visual appearance of facial images and, primarily target human observers. Conversely, privacy–enhancing techniques that use adversarial noise as a privacy mechanism, e.g. [13], introduce minor (often imperceivable) visual changes into the images and, hence, aim to make only automatic attribute inference infeasible, while not affecting human perception. Both types of B–PETs are designed to meet specific requirements and address distinct application scenarios. However, it needs to

5. We refer to this risk as *attribute recovery risk* in this work.

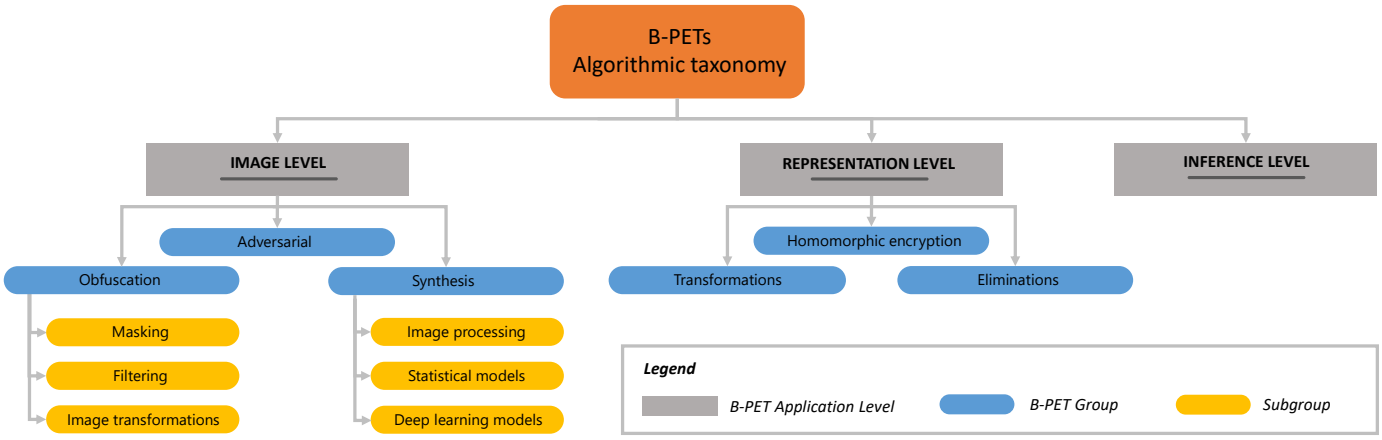


Fig. 5: Algorithmic taxonomy of existing biometric privacy enhancing techniques (B–PETs). The proposed taxonomy is tied to the functional structure of biometric recognition systems and partitions B–PETs into image–level, representation–level, and inference–level techniques.

be noted that B–PETs targeting human observers also ensure privacy with respect to automatic recognition techniques (in most cases), while this is not necessarily the case the other way around.

Human–targeted B–PETs commonly operate directly on image (or video) data and try to alter facial appearances for privacy protection. Similarly, machine–targeted B–PETs also often operate at the image level. However, machine–targeted techniques can also be applied at the later stages of a biometric recognition system, e.g., on the representation or inference levels, as discussed in the following section.

2.4 Algorithmic Taxonomy

For our algorithmic taxonomy, we consider possible application points of B–PETs within biometric recognition systems, as illustrated in Fig. 4. Thus, we classify existing B–PETs into techniques that operate at: *i)* the image level, *ii)* the representation level, and *iii)* the inference level. The main characteristics of these three groups are summarized below:

- *Image–level techniques:* Most of the existing work on privacy enhancement in biometrics focuses on the problem of *visual privacy* and is, therefore, concerned with B–PETs that operate at the image–level. We use the term *image–level* to account for techniques that process either still images or videos, but manipulate image-level data for privacy enhancement. Image–level techniques aim to enhance privacy by altering visual data using either obfuscation, adversarial, or synthesis techniques, as illustrated in Fig. 5. Solutions from this group include a wide variety of techniques, ranging from simple low-level B–PETs typically deployed on the onboard processing logic of contemporary imaging sensors, such as smart cameras [5], [6], [54], [80], to more elaborate techniques capable of ensuring better trade–offs between data utility and privacy protection, but at the cost of higher computational complexity. A few application examples of image-level B–PETs are shown in Figs. 6, 7, and 8.
- *Representation–level techniques:* With deployed biometric systems access to the raw (image–level) data is usually not possible. Once users enroll in a system their biometric data is stored in the form of compact templates – descriminative representations derived from the enrollment data. Representation–level techniques try to ensure that no sensitive information can be extracted from these templates and, consequently, that the data stored is only used for the intended purpose. As shown in Fig. 5, techniques from this group can in general be partitioned into transformation and elimination based methods, where the former aim to suppress some targeted aspect of the data by transforming the original templates into another form, while the latter try to remove elements (i.e., features) of the templates that are most informative with respect to the targeted attribute. Representation–level B–PETs also include solutions based on homomorphic encryption, which operate at the intersection between data security and privacy enhancement and use cryptographic solutions to ensure that the stored representations (i.e., biometric templates) are used only for a predefined purpose.
- *Inference–level techniques:* Privacy enhancement may also be applied during the matching or classification stages in a biometric system, a.k.a, during *inference*. Here, some properties of the matching or classification procedure are commonly exploited to ensure that the data is only used for the intended purpose, see e.g., [22]. Thus, inference–level B–PETs typically modify the biometric template as well as the comparison/classification procedure used to derive a similarity/comparison score in the biometric system with the goal of privacy enhancement. Unlike image–level B–PETs, techniques from this group exclusively target automatic machine learning models and not humans. The value of inference–level B–PETs is to two–fold: *i) privacy-by-design:* such techniques may be used during the design of biometric recognition systems to add another layer of privacy and ensure

that biometric data is used only for the intended purpose, e.g., identity recognition/authentication, and *ii*) *retrofitting:* for retrofitting existing biometric systems that may not have been designed with privacy assurances in mind. With such legacy systems, inference–level B–PETs can be utilized to improve the level of privacy with minimal intervention into the deployed components. In this regard, inference–level techniques share characteristics with representation–level B–PETs that can also be applied on top of existing installations by training/building/introducing an additional privacy–oriented layer in the overall system design [81].
Inference–level techniques were only recently presented in the literature and are, therefore, significantly less represented than B–PETs operating at other levels.

A high–level comparison of some of the characteristics of the three categories of B–PETs discussed above is presented in Table 2. In the table we provide our view on the B–PET categories in terms of *i)* computational complexity, *ii)* suitability of the different groups for removal (or suppression) of different attributes, *iii)* the possibility to incorporate formal privacy schemes with privacy guarantees, *iv)* the level of privacy enhancement that can be achieved, and *v)* the amount of biometric utility that can be preserved when applying privacy enhancement. Note that image-level techniques include a broad range of techniques, so a range of values is provided for this group.

The taxonomy presented above considers functional aspects of biometric systems and the corresponding data representations, where privacy enhancement can be applied. Another important aspect of privacy protection in the context of biometric systems is database privacy, where the number and type of queries that can be issued to a database is limited. The purpose is to mitigate the possibility of deducing sensitive attributes of an individual while still permitting the extraction and computation of aggregate statistics. A number of differential privacy schemes have been developed in other fields for this purpose [82], [83], [84], [85]. In this survey, we do not discuss this topic since it has been sparingly used in the biometrics literature compared to other schemes.

3 Survey of B–PETs for Face Biometrics

Biometric privacy enhancing techniques (B–PETs) are in general applicable to different biometric characteristics, as evidenced by the vast body of research on this topic e.g., [32], [86], [87], [88], [89], [90], [91], [92], [93], [94], [95]. However, a considerable portion of the existing work is focusing on facial biometrics, mainly because of the multitude of (sensitive) information that can be inferred automatically by analyzing facial appearances and the fact that the analysis can be done without the cooperation and consent of individuals. In this section, we review different approaches to privacy enhancement with facial data. We discuss in detail each of the categories introduced in the previous section, i.e., image–level, representation–level, and inference–level techniques, and provide examples of B–PETs for each of these categories.

3.1 Image–level techniques

Image–level techniques represent the broadest category of B–PETs in our taxonomy and at the coarsest level can be grouped into: *i) obfuscation*, *ii) adversarial*, and *iii) synthesis* approaches, as also illustrated in Fig. 5. Details on the three groups are provided in the following sections.

3.1.1 Obfuscation techniques

The first group of image–level B–PETs, i.e., obfuscation techniques, is based on computationally simple privacy mechanism that are commonly applicable to both still images as well video data. Technique from this group typically degrade the quality of the original images (and videos) to such a degree that face recognition or attribute classification becomes unfeasible or (at least) impaired. As a result, the obfuscated data offers higher levels of privacy protection, but at the expense of reduced biometric utility. Many of the existing obfuscation techniques also utilize cryptography to secure the original data and incorporate mechanisms to invert the privacy enhancements. A few illustrative examples of the application of obfuscation techniques are presented in Fig. 6. In our taxonomy we further partition obfuscation techniques into three sub-groups that use either: *i) masking*, *ii) filtering*, or *iii) image transformations* for biometric privacy enhancement. The three subgroups are discussed in the sections below.

Masking techniques

The first subgroup of obfuscation-based B–PETs tries to reduce the biometric utility of facial data by masking informative regions in the facial imagery. The main idea behind such *masking techniques* is to conceal either facial parts or the entire face region using masks or other abstract shapes with the goal of hiding identity information – illustrated in Fig. 6 (a). However, as a result of the masking operation, information on other biometric attributes is also concealed and the generated privacy enhanced data is usually considered to have only limited (or no) biometric utility. Because masking is a simple operation from a computational point of view, techniques from this subgroups are highly suitable for implementation on embedded devices, smart cameras, and other low-resource platforms.

An example of a masking–based B–PET was presented by Chinomi et al. in [96], [97]. Here, the authors describe *PriSurv*, a surveillance system that contributes towards privacy protection in video footage by replacing image regions corresponding to people (or their faces) with various types of masks, shapes, or even background pixels. Background subtraction is used to identity subjects in video and an adaptive masking procedure is employed to mask out faces in accordance with a predefined *privacy policy*. The policy defines the amount of utility to preserve and determines the type of *visual abstraction* (e.g., semi–transparent or opaque mask) to use to conceal people in the surveillance data.

Another solution that relies on masking for privacy enhancement was proposed by Zhang et al. [54] in the form of the *Anonymous Camera*. The presented camera features built–in CCD (RGB) and IR imaging sensors, a cold mirror, and a liquid crystal on silicon (LCoS) device. It performs accurate real–time masking of human faces based on information

TABLE 2: High–level comparison of existing categories of biometric privacy enhancing techniques (B–PETs). The table compares the computational complexity of B–PETs, the suitability for removing (suppressing) information on either identity or soft–biometrics, the possibility to devise provable privacy schemes, the level of privacy enhancement achieved and the level of biometric utility the techniques are able to ensure when suppressing other attributes. Image– and representation–level techniques cover a broad spectrum of conceptually different techniques with diverse characteristics, a range of values is, therefore, provided for these categories.

B–PET category	Complexity	Suitability for removal of information on		Guarantees	Privacy enhancement	Utility preservation
		Identity	Soft biometrics			
Image–level	L–H	M–H	M–H	L–M	M	L–M
Representation–level	L–M	M	H	L–M	M	M–H
Inference–level	L	L	M	M	H	H

*Symbol explanation: L – low, M – medium, H – high.

from the thermal image – captured by the IR part of the camera. Because faces usually appear as warmer objects in the data, heat signatures are used to identify image regions for masking. A similar idea was also described by Shiff et al. in [98] with their *Respectful camera* design. Here however, visual color–markers are used to find sensitive image regions with faces for masking instead of heat signatures.

Chen et al. [99] described an automatic face masking technique for obscuring human faces in video. In this work, a privacy enhancing technique is presented that detects faces and the video and then tracks them across the entire video sequence. To enhance privacy, obfuscation masks are created and applied on the detected face locations at the detected scales with the goal of obfuscating identity information. Robustness of the privacy enhancing procedure is ensured through the use of an efficient face detection approach that relies on simple background subtraction and a head–shoulder detector.

In [100], Wang et al. demonstrated how face masking (and blurring) techniques can be used for privacy enhancement in real-world applications. The authors describe the so-called *RTFace* system, which consists of two components: a Face Trainer and a Privacy Mediator, and supports face detection, tracking, path–based anonymization, and whitelisting of enrolled users. The first component of RTFace, the Face Trainer (OpenFace [101] based), is used to recognize faces and enroll new users, whereas the second component, the Privacy Mediator, is utilized to determine whether detected faces should be visible (whitelisted) or not (masked with patches). In [102], Das et al. applied the idea of RTFace to IoT–based infrastructures and provided users with the explicit option to choose whether or not they want their faces to be masked (a.k.a. denatured) in the video streams.

An interesting (reversible) masking–based B–PET was presented by Yuan and Ebrahimi [36], [103]. The technique is based on *JPEG transmorphing* and supports arbitrary visual manipulations of selected image regions (i.e., typically faces) – not only masking. In the first step, selected image regions are masked or otherwise abstracted to conceal sensitive information. A binary mask, corresponding to the concealed image regions, is created next and used to extract the original (unmasked) regions from the input image. Finally, the original image regions are secured using a symmetric encryption scheme with a secret key. The process of transmorphing is completed when the mask and encrypted sub–image are inserted into the obfuscated JPEG image. To reconstruct the original content from the transmorphed image, data decryption is used.

Masking techniques are in general easy to implement and are, hence, often used as baselines in various privacy–related performance evaluations, e.g., [104], [105], [106].

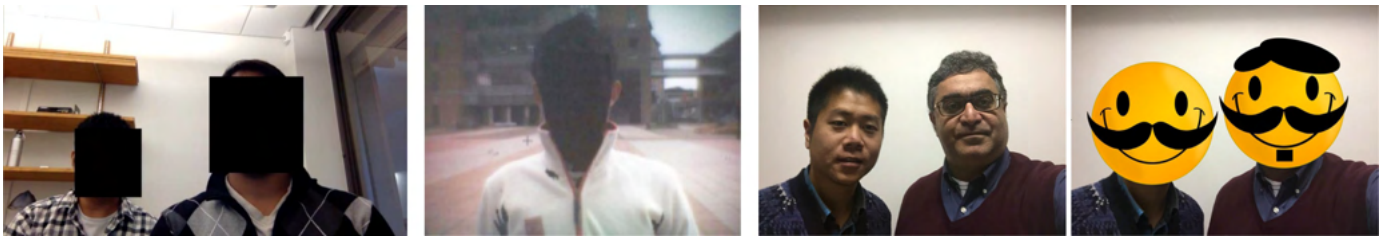
Filtering techniques

The second subgroup of obfuscation-based B–PETs processes regions of interest (ROIs) in video frames or still images by applying selected (linear or non–linear) filtering operations on the input data. The filtering operation typically degrades the quality of the imagery, which, in turn, affects its biometric utility. Typical filters used include blurring and averaging filters as well as gradient operators. B–PETs from this subgroup also often incorporate additional processing steps (such as binarization) to further degrade the quality of the source images. The privacy enhancement is typically targeting identity information, but in general impacts other biometric attributes as well. While some implementations include face detection techniques that drive the privacy enhancement towards specific image regions, others simply filter the entire image and do not rely on the identification of specific privacy–sensitive ROIs. A few illustrative examples of filtering techniques are presented in Fig. 6 (b). We note that in the privacy–related image processing literature, the term *privacy filter*[6] is often used to describe any privacy protection approach even if the underlying mechanism includes data transformations, scrambling, masking, and other similar approaches. In our taxonomy, on the other hand, we only consider filtering techniques that utilize standard image filters, defined with some sort of convolutional kernel.

A typical example of a B–PET relying on such a filtering technique is deployed in the *Deidentification camera*, described in [80]. The camera aims to ensure real–time privacy protection in videos by implementing a deidentification pipeline that consists of five steps: background subtraction, person detection, person tracking, segmentation, and deidentification. The near real–time deidentification process (10–11 frames per second) is based on Gaussian blurring and binarization, and is implemented entirely on a low-resource (OMAP4–based) embedded platform.

Wang et al. [41] introduced an obfuscation technique that again exploits blurring for privacy protection. The procedure uses a tracking approach based on mean–shift and

6. The term is most often used in relation to privacy enhancement in the area of video surveillance.



(a) Masking (left to right): Das et al. [102], Zhang et al. [54], and Yuan and Ebrahimi [36], [103].

(b) Filtering (left to right): Erdélyi et al. [107], Winkler and Rinner [6], and Fradi et al. [108].

(c) Image transformations (left to right): Ruchaud and Dugelay [109], Kobayashi et al. [110], and Korshunov and Ebrahimi [111].

Fig. 6: Visual examples of the application of obfuscation-based B–PETs relying on: (a) masking, (b) filtering, and (c) image transformations. As can be seen, obfuscation techniques commonly use (computationally) simple privacy enhancing techniques that usually aim at hiding identity information. Only limited biometric utility (if any) is preserved with such B–PETs.

active contours that makes it applicable to video sequences. An AdaBoost–based face detector is utilized to initiate the tracking and a background subtraction procedure is devised to constrain the face search region. Due to the design of the procedure, the algorithm exhibits robustness to changes in scale, pose, and partial occlusion of the facial regions. Several conceptually similar techniques, exploiting blurring to achieve privacy enhancement have also been presented in the literature, e.g., [112], [113].

To address privacy concerns in crowd monitoring applications, while still preserving utility, Fradi et al. [108] introduced the concept of context–enhanced filters. The idea here is to use crowd–density information to define the level of privacy enhancement that should be applied to different parts of the input data. To this end, a crowd–density estimation procedure is designed around (local) FAST features and optical flow. A HOG–based (head) detector is utilized to identify sensitive regions in the input images. The computed information (ROI positions and crowd–density maps) is then used with blurring and pixelation for adaptive privacy enhancement, enabling higher–levels of privacy protection in less crowded areas and lower–levels in areas with many people. A similar adaptive scheme was also proposed by Letournel et al. [114] in the form of a face deidentification technique with expression preservation capabilities. The technique is based on variational adaptive filtering, where the filtering preserves key facial features (i.e., the eyes, the lips, and their corners) but conceals facial identity.

Another notable technique from this subgroup, based on *adaptive cartooning*, was presented by Erdélyi et al. in [107]. The technique converts original input images (or selected regions) into abstract representations with cartoonish appearance and, consequently, decreased biometric utility. The cartooning process includes blurring, edge detection, mean–shift filtering, and in the final step a (edge–intensity based) weighted superposition of the generated cartoon image and the corresponding original. Because of the computationally simple steps involved with this technique, it is suited for deployment in embedded smart cameras. The authors also suggest that privacy enhancement with the cartooning approach retains some utility because it allows to infer behavioural information, i.e., actions remain perceptible.

A filtering approach, designed around gradient operators, was implemented as the privacy mechanism of choice in the *TrustCam* – a privacy–preserving smart camera, introduced by Winkler and Rinner in [6]. TrustCam uses edge detection with standard gradient operators to generate surrogate ROIs for privacy protection in visual data. To ensure confidentiality (i.e., data security), the ROIs are encrypted using cryptographic keys. This process limits access to the encrypted data only to camera operators with suitable credentials. Representing faces with edge information ensures

a certain level of privacy protection, but also preserves the biometric utility of the data to some extent, as evidenced by existing work building on gradient information for recognition, e.g. [115], [116], [117], [118], [119].

Due to privacy concerns over videos captured by Micro Aerial Vehicles (MAVs or mini drones), researchers also started looking at obfuscation-based filtering techniques depoyable on onboard cameras of MAVs. Bonetto et al. [63], for example, studied the efficiency of different privacy filters for privacy protection in video footage captured by mini–drone based video surveillance systems. Sarwar et al. [120], on the other hand, presented an approach that filters facial regions in video footage captured by MAVs adaptively as a function of image resolution. The proposed privacy filter is applied only if the input resolution is high enough to be used with a face matcher. The approach is designed to be robust to *reconstruction attacks* aimed at reversing the effect of the privacy filter and presents an extension of the authors' earlier work, relying on *adaptive Gaussian blurring* [121].

While a vast majority of work related to filtering-based B-PETs focuses on video data and an unselective reduction of biometric utility [122], [123], more recent research also started investigating filtering techniques capable of preserving information on specific biometric attributes. To this end, Ye et al. [124], for example, studied various privacy enhancement techniques (including blurring) that protect privacy, on the one hand, but also allow for reliable age estimation, on the other.

Image transformations

The last subgroup of obfuscation-based B–PETs relies on various image transformations to conceal (remove or obscure) sensitive regions in facial images or video. These transformations include image subsampling [134], scrambling [136], [139], [152], [153], [161], mosaicing [110], [160], warping [142], morphing [111], foveation [151], half–toning [159], image puzzling [147], steganography, and others. Techniques from this group are often tied to various compression standards [145] and exploit selected characteristics of the standards for privacy enhancement. Transformation–based techniques, like filtering approaches, try to reduce the biometric utility of facial data by altering its visual characteristics. This process is typically not selective and affects all biometric attributes to a similar extent. Existing techniques often focus on data security aspects (encryption and data hiding) and mechanisms for reversing the effect of privacy enhancement. A few application examples of transformation–based B–PETs are presented in Fig. 6 (c).

One of the early solutions utilizing image transformations for privacy enhancement was presented by Cucchiara et al. in [134]. In this work, the authors present an active, privacy–aware surveillance system built around Pan–Tilt–Zoom cameras. The system included a privacy enhancing mechanism that pixelated (subsampled) detected faces, while still allowing to monitor large areas and observe information on the number of surveilled people, their interactions, and behaviour as well as other information less critical from a privacy perspective.

Another transformation–based solution, called *PrivacyCam*, was presented by Chattopadhyay and Boult in [5]. The privacy enhancement used by PrivacyCam is implemented on a digital signal processor (DSP) and relies on simple computer vision techniques and encryption of the quantized DCT (Discrete Cosine Transform) coefficients used in compression standards, such as JPEG [135]. The mechanism implemented in PrivacyCam reduces the biometric utility of facial data, but also incorporates cryptographic means to ensure that the enhancement procedure is reversible.

Building on the idea of manipulating DCT coefficients, Kobayashi et al. [110] proposed the *Privacy Protection Surveillance Camera System* (PPSCS) that allows users to request to be hidden from surveillance. At the core of PPSCS is a privacy enhancing technique that combines reversible mosaicing with reversible watermarking. Detected faces are first divided into 8×8 pixel blocks (similar to a JPEG processing) and quantized DCT coefficients are computed for each of the blocks. The values of low–frequency pixels are encrypted by AES and embedded in the high–frequency pixels. Finally, a reversible watermarking procedure is used to embed the encrypted data in the DCT domain of the mosaic image. The camera system was later extended by Hoshino et al. [160] into the *improved PPSCS* (IPPCS) design addressing some of the shortcomings of PPSCS. Several other notable works on privacy enhancement relying on the manipulation of JPEG–like DCT–coefficients, were also presented in the literature, e.g., [132], [133], [151]. Many of these operate with specific compression standards, such as H.264/AVC, e.g., [109], [136], [139], [145], [146].

Another representative approach from the subgroup of transformation–based B-PETs was presented by Korshunov and Ebrahimi in [142]. Here, privacy protections is achieved through a warping procedure that distorts facial appearances and in turn reduces the biometric utility of facial images. The authors demonstrate that face detection is still feasible with the warped images, while face recognition performance (with a selected classifier) is degraded significantly. In [111], Korshunov and Ebrahimi report similar findings, when morphing is used for privacy protection instead of warping.

Ruchaud and Dugelay [150] proposed a privacy protection technique that obfuscates regions–of–interest (ROIs) in images (e.g., faces) using surrogate ROIs that only encode shape information. The technique relies on a combination of steganography and scrambling. Once the scrambling is applied on the pixels of the ROI, their most significant bits are hidden in the least significant bits of the target image. The privacy enhancement procedure is designed to be reversible, while ensuring suitable levels of privacy protection through the surrogate images.

Chriskos et al. [156] described an transformation–based B–PET designed for hindering automatic face detection. The proposed *face–detection hindering* technique (as called by the authors) introduces artifacts into face images (e.g., noise and projections) that impair face detection, but still preserves enough information, so that faces remain intelligible for human observers.

Dadkhah et al. [159] studied possibilities for applying different *half–toning algorithms* with the goal of avoiding unwanted (automated) face detection and recognition. Half–toning transforms the standard grey–level pixel intensities of the input images into black and white dots in a way that preserves the intelligibility of images for human observes,

TABLE 3: High–level comparison of the surveyed obfuscation privacy–enhancing techniques. The table summarizes the techniques in terms of the type of input data they were applied to, the targeted attribute to be concealed and the targeted attribute to be preserved (if any), reversibility (Rev.) and privacy guarantees (Gua.), the privacy mechanism applied, the strategy used with respect to biometric utility, the target of the privacy enhancement, and the test data selected for experimentation. Techniques are listed in chronological order for each sub–group.

Group	Techniques	Applied to		Attribute		Rev.	Gua.	Mechanism	Utility	Target	Test dataset
		Video	Stills	Concealed	Preserved						
Masking	Schiff et al., 2007 [98]	✓	✓	ID	NSA	✗	✗	Opaque masks	RD	H	IHD
	Chinomi et al. (*PriSurv*), 2008 [96]	✓	✓	ID	NSA	✗	✗	Adaptive masks	RD	H	IHD
	Chen et al. (EMHI), 2009 [99]	✓	✗	ID	NSA	✗	✗	Opaque masks	RD	H	IHD
	Zhang et al., 2014 [54]	✓	✓	ID	NSA	✗	✗	Opaque masks	RD	H	IHD
	Yuan and Ebrahimi, 2017 [36], [103]	✗	✓	ID	NSA	✓	✗	Arbitrary masks	RD	H	PIPA [125]
	Wang et al. (*RTFace*), 2017 [100]	✓	✓	ID	NSA	✗	✗	Denaturing masks	RD	H	LFW [126]
	Das et al., 2017 [102]	✓	✓	ID	NSA	✗	✗	Opaque masks	RD	H	IHD
Filtering	Zhao and Stasko, 1998 [112]	✓	✓	ID	NSA	✗	✗	Various filters	RD	H	IHD
	Boyle et al., 2000 [113]	✓	✓	ID	NSA	✗	✗	Blurring/pixelation	RD	H	IHD
	Wang et al., 2008 [41]	✓	✓	ID	NSA	✗	✗	Blurring	RD	H	IHD
	Mrityunjay and Narayanan, 2011 [80]	✓	✓	ID	NSA	✗	✗	Blurring, binarization	RD	H	IHD
	Agrawal et al., 2010 [122]	✓	✗	ID	NSA	✗	✗	Various filters	RD	H	CAVIAR; BEHAVE [127]
	Winkler and Rinner, 2010 [6]	✓	✓	ID	NSA	✓	✗	Gradient (edge) filtering	RD	H	IHD
	Korshunov and Ebrahimi, 2012 [123]	✓	✓	ID	NSA	✗	✗	Various filters	RD	H	IHD
	Fradi et al., 2013 [108]	✓	✗	ID	NSA	✗	✗	Blurring, pixelation	RD	H	PETS2009 [128] and others
	Erdélyi et al., 2014 [107]	✓	✓	ID	NSA	✗	✗	Cartooning	RD	H	PEViD [129]
	Letournel et al., 2015 [114]	✓	✓	ID	NSA	✗	✗	Adaptive filtering	RD	H	LFW [126]
	Bonetto et al., 2015 [63]	✓	✓	ID	NSA	✗	✗	Various filters	RD	H	IHD (Mini–drone)
	Sarwar et al., 2016 [121]	✓	✓	ID	NSA	✗	✗	Adaptive blurring	RD	H	Dataset from [130]
	Ye et al., 2018 [124]	✗	✓	ID	Age	✗	✗	Various filters	RD	H	Adience [131]
	Sarwar et al., 2019 [120]	✓	✓	ID	NSA	✗	✗	Adaptive blurring	RD	H	LFW [126], IHD
Transforms	Dufaux and Ebrahimi, 2004 [132]	✓	✓	ID	NSA	✓	✗	Scrambling	RD	H	IHD
	Martínez–Ponte et al., 2005 [133]	✓	✓	ID	NSA	✗	✗	Masking	RD	H	IHD
	Cucchiara et al., 2006 [134]	✓	✗	ID	NSA	✗	✗	Pixelation	RD	H	IHD
	Chattopadhyay and Boult, 2007 [5], [135]	✓	✓	ID	NSA	✓	✗	Scrambling	RD	H	IHD
	Dufaux and Ebrahimi, 2008 [136]	✓	✗	ID	NSA	✓	✗	Scrambling	RD	H	VTM [137]
	Xuan and Jiang, 2009 [138]	✓	✓	ID	NSA	✓	✗	Color watermark	RD	H	VTM [137]
	Tong et al., 2010 [139]	✓	✗	ID	NSA	✓	✗	Scrambling	RD	H	VTM [137]
	Cichowski and Czyzewski, 2011 [140]	✓	✓	ID	NSA	✓	✗	Pixel relocation	RD	H	IHD
	Rahman et al., 2012 [141]	✓	✓	ID	NSA	✓	✗	Scrambling	RD	H	IHD
	Korshunov and Ebrahimi, 2013 [142]	✓	✓	ID	NSA	✗	✗	Warping	RD	M	Yale Faces [143]
	Korshunov and Ebrahimi, 2013 [111]	✓	✓	ID	NSA	✗	✗	Morphing	RD	H	FERET [144]
	Wang et al., 2013 [145]	✓	✗	ID	NSA	✓	✗	Scrambling	RD	H	VTM [137]
	Su et al., 2013 [146]	✓	✗	ID	NSA	✓	✗	Partial scrambling	RD	H	VTM [137]
	Kobayashi et al., 2014 [110]	✓	✓	ID	NSA	✓	✗	Reversible mosaicking	RD	H	IHD
	Bhattarai et al., 2014 [147]	✗	✓	ID	NSA	✗	✗	Noise overlay	RD	M	LFW [126]
	Melle and Dugelay, 2014 [148]	✓	✓	ID	NSA	✓	✗	Scrambling	RD	H	ORL [149]
	Ruchaud and Dugelay, 2015 [150]	✓	✓	ID	NSA	✓	✗	Edge image, binarization	RD	H	Mini–drone [63]
	Alonso–Pérez et al., 2016 [151]	✓	✓	ID	NSA	✗	✗	Foveation	RD	H	FERET [144]
	Jiang et al., 2016 [152], [153]	✓	✓	ID	NSA	✓	✗	Scrambling	RD	H	ORL [149], PIE [154], PUBFIG [155]
	Ruchaud and Dugelay, 2017 [109]	✓	✓	ID	NSA	✓	✗	Adaptive scrambling	RD	H	FERET [144], VTM [137]
	Chriskos et al., 2017 [156]	✓	✓	ID	NSA	✗	✗	Noise, SVD transform	RD	M	XM2VTS [157], dataset from [158]
	Dadkhah et al., 2018 [159]	✓	✓	ID	NSA	✗	✗	Half–toning	RD	M	IHD
	Hoshino et al., 2018 [160]	✓	✓	ID	NSA	✓	✗	Reversible mosaicking	RD	H	IHD
	Liu et al., 2018 [161]	✓	✓	ID	NSA	✓	✗	Scrambling	RD	H	IHD
	Fan, 2018 [73]	✓	✓	ID	NSA	✗	✓	Pixelation	RD	H	ORL [149], MOT [162], PETS [128]
	Fan, 2019 [163]	✓	✓	ID	NSA	✗	✓	SVD transform	RD	H	PIPA [125], ORL [149]

Symbol explanation: ID – identity, NSA – no specific attribute (information removed on all attributes), RD – reduction, IHD – In–house dataset, VTM – Video Test Media, H – Human, M – Machine.

but not necessarily for machine learning techniques. The work investigates the impact of multiple half–toning techniques for privacy enhancement, including Floyd–Steinberg dithering for RGB images, and Stucki dithering, Bayern half–toning and Jarvis half–toning for graycale images.

An interesting approach to privacy enhancement based on watermarking was proposed by Xuan and Jiang in [138]. The approach treats detected facial regions as characteristic watermark information and embeds them into a target image using a (large capacity) watermarking algorithm. After face detection, compression, and recoding are used to generate a bipolar watermark sequence that can be embedded into the target image for privacy protection. The approach is reversible and, hence, allows to extract the original facial "watermark" from the bipolar watermark sequence.

A solution for video–based privacy enhancement with image transformations was proposed by Cichowski and Czyzewski in [140]. Here, the authors introduce a reversible and perceptually lossless algorithm for *anonymizing video streams*. The algorithm uses a dedicated (reversible) *pixel relocation* technique for ROI hashing and a semi–blind watermarking procedure for data embedding. Pixel relocation changes pixels locations within a ROI and ensures that visual information within the ROI is obscured. The semi–blind watermarking, on the other hand, is employed to hide the coordinates of the anonymized ROI in the target image.

Liu et al. [161] introduced an approach to privacy enhancement in video that relies on skin color information to detect face–region candidates and an SVM–based classifier to refine candidate regions and determine final face locations. (Reversible) scrambling procedures are utilized to modify both spatial positions of pixels within a face region as well as their values. This process effectively conceals identity information but, as suggested by the authors, still allows to recognize actions in the processed videos. This characteristic comes as a side effect of the obfuscation step, which only targets facial regions and not full body ROIs.

Rahman et al. [141] developed a privacy enhancing obfuscation procedure targeting surveillance videos. The procedure uses chaos cryptography for data scrambling of

sensitive ROIs in an image and is fully reversible. According to the authors, the proposed procedure supports multiple levels of abstraction in data hiding and therefore allows for the implementation of various privacy policies (i.e., obfuscation of different ROIs in the image) depending on the credentials of the person examining the surveillance video. The procedure is computationally efficient and, therefore, highly suitable for video–based privacy enhancement.

Melle and Dugelay [148] devised a reversible scrambling technique applicable to both image and video data. The technique processes privacy sensitive ROIs using an adaptive codebook that consists of a set of background patches (image regions without sensitive information) processed with affine transformations. The main idea of this work is to exploit image self–similarities to encode the image and combine the encoding scheme with a scrambling procedure for privacy enhancement. The authors demonstrate that different trade–offs between privacy–protection and intelligibility can be achieved by varying the scrambling strength.

While the techniques discussed above rely on privacy mechanisms without privacy guarantees, a formal obfuscation technique based on the concept of differential privacy was recently presented in the literature, i.e., [73]. This work showed that it is possible to develop efficient privacy enhancing algorithms (sanitizers) based on the basic differential privacy framework [164] that are able to conceal sensitive information from input images by adding controlled randomness to the data. Based on this observation, the authors of [73] developed a differentially private pixelization scheme for CCTV–based image data with the goal of protecting individuals, objects, as well as their features. The proposed method was shown to reduce the success rate of re–identification attacks and was later extended to other obfuscation schemes in [163].

3.1.2 Comparison of obfuscation techniques

A high–level summary of the surveyed obfuscation techniques is presented in Table 3. The table compares the reviewed techniques in terms of *i)* input data the methods were applied to, *ii)* the biometric attribute targeted for removal and preservation, *iii)* whether the techniques are reversible or not, *iv)* whether they provide privacy guarantees, *v)* the mechanism used for privacy enhancement, *vi)* the utility preservation strategy utilized, *vii)* the target of the privacy enhancement, and *viii)* the datasets the techniques were tested on. As can be seen, obfuscation techniques mostly target identity information during privacy enhancement and typically do not try explicitly to preserve any specific biometric attribute. Any utility preserved after privacy enhancement (if at all) is commonly a side effect of the selected mechanism utilized and not a targeted characteristic of the B–PETs from this group. However, because many of the reviewed techniques are designed specifically for concealing facial regions in video data, they are commonly advertised in the literature as being able to preserve behavioral information. Various mechanisms have been used in the literature with obfuscation–based B–PETs, but they share a common characteristic in that they are computationally simple and (for the most part) do not come with provable privacy guarantees. Many solutions in this group are also reversible and allow to reconstruct the original visual content from the obfuscated data. In terms of experimental evaluations, these are typically conducted on in–house datasets (IHD), but also on standard face datasets used regularly in the face recognition literature.

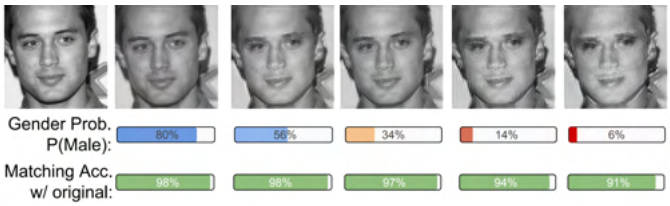


Fig. 7: Visual example of an adversarial approach to biometric privacy enhancement, i.e., FlowSAN from [12]. As can be seen, adversarial approaches introduce small perturbations into facial images, so that certain biometric attributes are suppressed (in this case gender), while others are preserved (e.g., identity in the example above).

3.1.3 Adversarial approaches

The second group of image–level privacy enhancing techniques uses adversarial approaches to reduce the biometric utility of facial images and impair recognition of biometric attributes with selected classifiers. Techniques from this group rely on various strategies from the field of adversarial machine learning (i.e., adversarial perturbations, adversarial examples, adversarial noise, etc.) to alter facial data, so that the privacy enhanced imagery retains some utility, but has an adverse effect on the performance of automatic classification models. An illustrative example of an adversarial privacy enhancing technique is presented in Fig. 7.

Adversarial approaches represent a recent group of B–PETs that was popularized mostly due to advances in deep learning, even though similar ideas have been explored before the deep learning era. Techniques from this group are often designed to be highly selective, targeting a specific attribute (or more) for suppression, while preserving others. They typically require a pretrained classification model during training, which raises questions with respect to the generalization abilities of these techniques. Depending on the strategy adopted, existing techniques from this group either try to force the selected classifier to produce incorrect predictions (e.g., to flip labels for binary recognition tasks [13]) or to generate low–probability classification results, e.g., [12]. A shortcoming of adversarial techniques is their computational complexity, which makes them difficult to apply to large scale data in a time–efficient manner.

Several techniques have been proposed in the literature recently following the idea of adversarial learning. Wu et al. [166], for example, proposed an *adversarial training framework* for privacy–preserving visual recognition. The main idea of this work is to learn active degradations that can be used to transform video inputs and generate privacy enhanced data. The adversarial learning procedure is based on two competing objectives, where one strives to preserve data utility (actions in this work), while the second aims to ensure privacy protection by removing information on sensitive biometric attributes, such as face identity. A similar solution was also presented in [71].

TABLE 4: High–level comparison of existing image–level B–PETs relying on adversarial learning. The table compares the surveyed techniques with respect to the data they were applied to, the attribute they are aiming to conceal and to preserve, reversibility (Rev.) and privacy guarantees (Gua.), the privacy mechanism and utility preservation strategy utilized, the target of the enhancement, and the test dataset(s) employed for evaluation. Techniques are listed in chronological order.

Adversarial techniques	Applied to		Attribute		Rev.	Gua.	Mechanism	Utility	Target	Test dataset
	Video	Stills	Concealed	Preserved						
Mirjalili and Ross, 2017 [44]	✗	✓	GD	ID	✗	✗	SAP	RD	M	MUCT [165], LFW [126]
Wu et al. (Adversarial Privacy), 2018 [166]	✓	✓	ID	Behaviour	✗	✗	LD	RD	H	SBU [167], UCF101 [168], VISPR [169]
Huang et al. (GAP), 2018 [170]	✗	✓	GD	ID	✗	✗	LD	RD	M	GENKI [171]
Mirjalili et al. (SAN), 2018 [172]	✗	✓	GD	ID	✗	✗	SAP	RD	M	CelebA [173], MUCT [165], LFW [126], AR–face [174]
Mirjalili et al., 2018 [11]	✗	✓	GD, A, ET	ID	✗	✗	SAP	RD	M	CelebA [173], MORPH [175], LFW [126], MUCT [165], RaFD [176]
Chhabra et al., 2018 [13]	✗	✓	Arbitrary	ID	✗	✗	AP	RD	M	CelebA [173], MUCT [165], LFW [126]
Mirjalili et al. (FlowSAN), 2019 [12]	✗	✓	GD	ID	✗	✗	SAP	RD	M	CelebA [173], MORPH [175], MUCT [165], RaFD [176]
Chatzikyriakidis et al., 2019 [72]	✗	✓	ID	NSA	✗	✗	AP	RD	M	CelebA [173]
Mirjalili et al. (PrivacyNet), 2020 [69]	✗	✓	GD, A, ET	ID	✗	✗	SAP	RD	M	CelebA [173], MORPH [175], UTKFace [177], MUCT [165], RaFD [176]

Symbol explanation: GD – Gender, A – Age, ET – Ethnicity, ID – identity, RD – reduction, NSA – no specific attribute (information removed on all attributes), LD – Learned degradation
AP – Adversarial perturbation, SAP – Semi–adversarial perturbation, IDR – image degradation, H – Human, M – Machine.

In line with the idea discussed above, Huang et al. in [170] introduced a privacy–enhancing techniques, named *Generative Adversarial Privacy* (GAP). To ensure privacy, GAP relies on two deep learning models, where the first, the Privatizer, is a generator–style network that aims to alter input images and the second, the Adversary, is a classification model that tries to predict privacy–sensitive (in this case gender) information from the images produced by the Privatizer. Similarly to [166], the GAP model is learnt by optimizing a training objective that takes both data utility and privacy protection into account. Soft–biometric privacy experiments on the GENKI dataset are presented to demonstrate the feasibility of the approach.

Many existing adversarial techniques try to introduce minute changes (adversarial perturbations) into facial images to confound selected classification models. Following this idea, Mirjalili et al. [44] applied Delaunay triangulation on facial landmarks and optimized pixel intensities (i.e., color information) within the triangulated facial mesh in such a way that a pretrained gender classifier generates incorrect predictions. Because the color changes are minute, human observers can still correctly determine gender from the images, but the considered gender classifier cannot. The algorithm was not explicitly designed to preserve identity information, but the recognition performance (identity) was not significantly impacted as shown in the paper.

In their follow up work [172], the same authors proposed so–called *Semi–Adversarial Networks* (SANs), deep learning models that are again able to confound gender classifiers, but also allow for the privacy enhanced data to be used for identity verification – without significant loss in verification performance. One shortcoming of SANs is that they do not generalize well to unseen gender classifiers, i.e., classifiers that were not incorporated in the training phase may still accurately classify gender. To ensure generalizability to such unseen classifiers, the authors proposed a SAN extension, called *FlowSAN*. FlowSAN [12] applies multiple SAN transformations sequentially to improve the ability to generalize. An ensemble of SANs was also investigated in [11].

To fool soft–biometric classifiers, Chhabra et al. [13] proposed a framework aiming to conceal a *predefined set of facial attributes*. The privacy enhancement with this approach is based on the Carlini–Wagner L2 attack that adds adversarial noise to the input images, making it difficult for the selected classifiers to automatically infer the attributes targeted with the privacy enhancement. Similar to the SAN models discussed above, this approach yields promising results with the classifiers considered when generating perturbations, however, it does not generalize well to arbitrary classifiers.

A privacy enhancing technique built around Generative Adversarial Networks (GANs), called *PrivacyNet* [69], was presented in 2020. PrivacyNet is a GAN–based semi–adversarial network that targets multiple attribute classifiers simultaneously and in turn suppresses multiple attributes in facial images, while preserving identitiy information and, hence, facilitates verification. The generator of PrivacyNet is optimized to generate suitable adversarial noise, while the discriminator aims to detect if an input image is real or modified. The main contribution of PrivacyNet is the extension of SANs to multiple attributes and the integration into a GAN framework that ensures that the generated privacy enhanced images appear as natural as possible.

Solutions from this group have also been applied for deidentification techniques and not only soft–biometric privacy problems. Chatzikyriakidis et al. [72], for instance, proposed the *Penalized Fast Gradient Value Method* (P–FGVM), and applied it conceal identity information in still images. This idea of this work is similar to the idea used with SAN models, as adversarial examples are created that are able to retain most of the original facial appearance, while producing images that are typically missclassified by selected (pretrained) face recognition models. The deidentified face images using this adversarial approach yield high misclassification rates, but can still be recognized by human observers as they are visually very similar to the originals.

A high–level summary of the surveyed adversarial B–PETs in presented in Table 4. Note that techniques from this group have predominantly been applied to still images and less so to video data. Unlike other types of B–PETs, adversarial techniques have been applied for soft–biometric privacy problems as well as for deidentification. However, they are not designed to be reversible, even though defenses to adversarial examples, e.g., [178], [179], may be exploited to counteract privacy enhancement. Existing work in this area has, to the best of our knowledge, so far not looked into provable privacy schemes, and relied mostly of empirical evaluation to establish the privacy protection level ensured by the techniques.

3.1.4 Synthesis approaches

The last group of image–level privacy enhancing techniques relies on image synthesis and typically generate synthetic

facial data with predefined characteristics to be used as surrogates for the original face images (or videos). A few examples of the visual effect of synthesis–based B–PETs are presented in Fig. 8. Synthesis techniques are among the most widely studied B–PETs in the field of face biometrics and are commonly implemented based on: *i) standard image processing techniques*, *ii) statistical models*, and *iii) deep learning* solutions. Below we provide an overview of existing work from each of these three subgroups.

Image processing approaches

Techniques in this subgroup use standard image processing approaches, such as blending, morphing, mosaicing, or swapping, to generate synthetic images with enhanced privacy and typically operate on raw image pixels (and not parameterized models). The seminal k–Same–Pixel approach from Newton et al. [15], for instance, relies on the formal k–anonymity model and generates surrogate faces for deidentification of a subject–specific closed set of face images by averaging groups of k faces that are found to be similar (i.e., that appear within the same cluster of k faces after applying a clustering procedure). Each face in the input set is then replaced with the average face of the cluster it belongs to. This procedure is illustrated in Fig. 8 (b). Because faces in every given cluster are replaced with the same surrogate image, it is impossible to link the deidentified images to the original identities with a probability higher than $1/k$, where k is a hyper parameter of this approach. Utility preservation with the k–Same–Pixel approach is ensured by the surrogate–generation procedure, which combines visual characteristics of all averaged faces. We note that this approach is typically applied only to tightly cropped facial areas (without hair, shoulders, and other body parts that may appear in the image), so the privacy guarantees only apply to the cropped part of the image and not for others. Several extensions to the k–Same–Pixel approach were presented in the literature and we discuss these in later sections.

A similar approach, but without formal privacy guarantees, was also proposed by Bitouk et al. [181]. Here, the authors create a large face library and then deidentify images by swapping faces in the original input images with the closest face candidate from the created library. The replacement is based on similarities of different image attributes, such as pose, resolution, blur, and the like. Facial alignment is done by using six annotated fiducial points (eye and mouth corners). Finally, recoloring and relightning is applied to match the original illumination conditions on the newly swapped face as much as possible. The procedure is able to produce visually convincing deidentified faces and can also preserve a certain level of data utility.

Following the same idea, Mosaddegh et al. [182] developed a part–based deidentification scheme by aggregating facial components from different donors. Different from the approach in [181], this work partitioned the facial image into multiple components and swapped each component separately. Such a parts–based approach provides a certain level of flexibility in the construction of the donor library, but also makes sure that the generated (surrogate) face images does not bear similarities with any real person. To ensure visually convincing results, the approach utilizes Poisson image blending, which mitigates artifacts and produces photo–realistic deidentification results. A similar solution was concurrently also proposed by Xu et al. in [183]. However, this approach was also applied to video data.

A conceptually different approach to image processing based privacy enhancement was introduced by Othman and Ross in [45]. The work aimed at ensuring soft–biometric privacy (hiding gender information), while still enabling identity verification, through a face morphing procedure. The authors used facial landmarks to generate a triangular facial mesh for a pair of faces (e.g., one male, one female) and then implemented a morphing procedure that combined the two images, so that gender information is combined by the morphing step. Since the contribution of each face can easily be controlled, different trade–offs between gender obfuscation and identity preservation can be achieved.

Statistical models

Techniques in this subgroup perform privacy enhancement with the help of generative statistical models, such as Principal Component Analysis (PCA) [184] or Active Appearance Models (AAMs) [185]. Statistical models were introduced in this field as an alternative to the image processing techniques discussed in the previous section mostly due to their flexibility and the potential for improving the quality of the generated surrogate images and preserving more of the utility of the original input images. An application example of an AAM–based privacy enhancing technique is shown in Fig. 8 (c).

Many of the techniques in this subgroup are based on the k–anonymity model. In [15], Newton et al. presented the *k–Same–Eigen* technique (in addition to k–Same–Pixel discussed earlier), which extends the idea of k–anonymity from the pixel space to the PCA subspace. Clusters of k subjects are hence defined in the eigenspace of the training data and the surrogate images for deidentification are created from the averages of the PCA coefficients corresponding to all k faces in each given cluster. Compared to k–Same–Pixel, this solution produces less ghosting effects in the surrogate images, but still provides provable privacy guarantees. A detailed evaluation of k–Same algorithms was presented in [186].

An extension of Newton's work, called *k–Same–Select*, was later presented by Gross et al. in [10] and aimed at improving the utility of the deidentified faces. With k–Same–Select a face dataset is first partitioned into mutually exclusive groups of faces in accordance with a so–called utility function (e.g., measuring face similarity in terms of facial expressions). A k–Same algorithm is then utilized on each face group separately, resulting in deidentified faces conforming to the k–anonymity model that still preserve information on the biometric attributes targeted by the utility function.

To improve the quality of the generated surrogate faces, Gross et al. [64], proposed the *k–Same–M* approach. This privacy enhancing technique extends the concept of the k–Same family of algorithm to AAMs. Here, surrogate faces for deidentification are created by averaging over AAM model parameters, instead of pixels or PCA coefficients, such as with the k–Same–Pixel or k–Same–Eigen techniques.

(a) Reference images for (b), (c), (d), (e), and (f).

(b) Newton et al. (k–Same–Pixel) [15]

(c) Gross et al. (k–Same–M) [64]

(d) Meden et al. (k–Same–Net) [33]

(e) Hukelås et al. (DeepPrivacy) [180]

(f) Ren et al. [71]

Fig. 8: Application examples of different synthesis–based approaches to privacy enhancement. The images in (a) represent the input face images, the examples from (b)–(f) show corresponding privacy enhanced faces using different B–PETs from the literature. Note that the generated surrogate faces typically retain some utility, e.g., part of visual appearance in (b), (c), and (f), and also facial expressions in (d). In all examples, artificial face images are generated to replace the original data, so with the goal of concealing identity information.

Because AAMs combine a shape and an appearance (texture) model, synthesized faces from the averaged AAM parameters appear visually convincing and exhibit minimal artifacts – see Fig. 8 (c). A more complex multi–factor model capitalizing on data utility was later presented by the same authors in [30]. Implementational issues related to the k–Same–M approach were studied by Prinosil et al. [75], who proposed several heuristics to improve the visual quality of the deidentified images.

Meng et al. [31] introduced another AAM–based approach to deidentification aimed at utility preservation. The *k–Same–furthest* technique is again based on k–anonymity, but instead of generating surrogate images based on the closest set of faces (i.e., faces in a cluster) it identifies most dissimilar faces (i.e., faces furthest away in a feature space) and utilizes those for deidentification. To ensure that utility is preserved (facial expressions in this case), k–Same–furthest then clones the facial expression from the original face and maps it to the surrogate using a dedicated expression transfer procedure. Multiple extension of this approach focusing on different aspects of the deidentification procedure were presented in the literature, e.g., [76], [65].

Concurrently with [31], Du et al. [187] proposed *GARP–Face*, an AAM–based model that also preserves utility in accordance with the utility retention strategy. Thus, GARP–Face first extracts attributes (e.g., age, gender, ethnicity) of the input face and then generates a k–anonymity compliant surrogate image for deidentification in accordance with the identified attributes. In this way, the technique is able to conceal identity information, but retain a predefined set of soft biometric attributes.

Again targeting data utility, Chi and Hu [188] proposed an AAM–based approach to face deidentification that aimed to split the AAM parameter space into an identity subspace and a residual subspace. The presented approach, called *identity subspace decomposition* (ISD), is based on k–anonymity and allows to separate sensitive identity information from information that relates to data utility, in turn allowing to efficiently deidentify facial images, while still preserving information on facial expressions. Research following a similar idea was also presented in [189].

A notable work from the family of k–anonymity models was described by Sim and Zhang in [47]. The authors employ a subspace decomposition technique, called *Multimodal Discriminant Analysis* (MMDA) to decouple AAM parameters that control different facial attributes in order to selectively alter some facial attributes while retaining others. MMDA offers significantly higher flexibility compared to competing AAM–based deidentification schemes and other related decomposition approaches. The importance of this work is in the introduction of the notion of *controllable privacy*, where a user can specify, which attributes to conceal and which to preserve, something that is important, for example, when sharing images online and across social media. Similar work using MMDA for deidentification was also presented in [190].

A number of B–PETs that do not offer formal privacy guarantees, but utilize statistical models in the process of privacy enhancement, have also been studied in the literature. Samaržija et al., for example, described a technique based on the q–far deidentification concept in [191]. The technique uses AAMs for deidentification and swaps faces by mapping the texture triangles from the surrogate to the target faces. The authors argue that a one–to–one type of deidentification scheme are often preferable over standard many–to–one schemes (e.g., k–anonymity), as it does not need to consider other identities during deidentification and is able to generate surrogate faces that look more natural [191]. In conclusion, the standard many–to–one scheme ensures better privacy protection (with guarantees), whereas the proposed one–to–one scheme is more flexible and faster when new faces need to be deidentified, because no joint

processing of the entire dataset is needed.

To address privacy protection in videos, Meng et al. [192] designed a face deidentification pipeline built around statistical models capable of preserving facial expressions. The work focused mostly on ensuring identity consistency across frames, which is achieved by applying the same identity shift (the difference between a person's original neutral face and its de–identified version) to all face instances of the same person. The approach employs AAMs for face modelling, a constrained local neural field (CLNF) technique [193] for landmark detection and the k–Diff–furthest algorithm for face deidentification.

Deep learning approaches

Deep learning techniques represent the latest synthesis–based techniques used for biometric privacy enhancement. Techniques and models from this subgroup are typically considerably more complex than standard statistical models discussed in the section above and, as a results, are able to synthesise visually convincing and photo–realistic surrogate faces both for still images as well as for video data. Moreover, because deep learning models can be trained to exhibit a multitude of different characteristics, they are able to ensure competitive trade–offs between privacy protection and utility preservation, as illustrated by the examples in Fig. 8 (d)–(f). Deep models with provable privacy guarantees have also been explored in the literature recently. We note that some of the synthesis–based solutions in this subgroup also incorporate ideas from adversarial machine learning. These, however, are usually limited to the use of Generative Adversarial Networks (GANs) [231] typically utilized to ensure photo–realism of the synthesized faces.

One of the earliest examples from this subgroup was presented by Chi and Hu in [205]. This work extends the *identity subspace decomposition* technique (discussed in the previous section) to deep learning models. At the heart of the approach is a deep learning model that extracts so–called *facial identity–preserving* (FIP) features from input images and reconstructs faces from the computed FIP features. A k–Same based deidentification scheme is utilized to generate average FIP features from clusters of k subjects, leading to a deep learning based k–Same deidentification technique with utility preservation capabilities.

Another deep learning based B–PET was proposed by Rafique et al. in [206]. Here, the authors explored Gaussian–Bernoulli Restricted Boltzmann Machines (GBRBMs) [232], [233] as means for face deidentification and exploited the generative nature of GBRBMs to design two separate deidentification solutions. The first is capable of pixelating faces with the goal of hiding identity, whereas the second (auto–encoder based) allows to produce smoother, higher quality images that also preserve information on the expressed (facial) emotions.

Brkić et al. [32], [207] described an approach that goes beyond deidentification of facial regions only, but instead performs full–body deidentification. To this end a deep convolutional generative adversarial network (DCGANs) and the Viola–Jones face detector are utilized and combined with a GraphCut segmentation procedur to detect, segment, synthesize and finally deidentify faces, hairstyles as well as human silhouettes – via clothing color deidentification.

Several deep learning solutions inspired by k–anonymity have also been proposed recently. Meden et al. [9], for example, designed a face deidentification pipeline around Generative Neural Networks (GNN) consisting of the Viola–Jones face detector, a VGG16–based feature extractor and a GNN based face renderer – for surrogate face generation. This ad–hoc pipeline was then extended to the *k–Same–Net* model, which incorporated elements from the k–Same family of algorithms and was empirically shown to offer a solid compromise between privacy protection and data utility – preserving facial expressions without visual artifacts [33], [61]. While k–Same–Net was inspired by the k–anonymity model, a formal proof of the privacy guarantees associated with the model was not provided.

Building on the idea from [33], Guo et al. [211] introduced *k–Dive–Net*, a deep learning based deidentification models, that improved on the k–Same–Net model by integrating diversity into the deidentification procedure. As a result of this modification, the deidentified faces appear distinct from each other despite building on the k–Same family of techniques. Another interesting extension of the k–Same–Net model, called *k–Same–Siamese–GAN*, was also presented recently in [218].

A synthesis–based approach to privacy enhancement relying on *facial attribute transfer* was designed by Li and Lyu [219]. With this approach, facial attributes (mildly correlated with identity) are transferred from a source image to a donor face, preserving attribute information while hiding identity in the process. The model is implemented with an encoder–decoder network topology. The same encoder but separate decoders are used with each identity. This design choice is made to ensure that the encoder of the attribute–transfer model captures identity independent attributes (shared among all face identities) and each of the individual decoders preserves identity–dependent attributes, which can be mapped onto the target faces.

More recently, privacy-enhancing solutions based on GANs started to appear in the literature. Chen et al. [212], for instance, presented the *Privacy–Preserving Representation–Learning Variational Generative Adversarial Network* or PPRL–VGAN for short. The model relies on a variational auto–encoder (VAE) that serves as the generator within the adversarial learning framework as well as on multiple discriminators that ensure that the generated faces appear realistic and exhibit the desired properties in terms of facial attributes (e.g., identity, facial expression). To facilitate facial synthesis a disentangled representation is learned in the latent space of the auto–encoder. Once trained, the model is able to perform face image synthesis, face image completion, and even facial expression morphing, in addition to deidentification.

Ren et al. [71] studied privacy enhancement in the context of action recognition systems. The authors proposed a GAN–based multi–task framework that learns to generate surrogate facial regions through a combination of learning objectives related to action recognition, identity classification, and face modification. Because action recognition also heavily relies on consistent visual information obtained from facial areas, standard face deidentification methods have a negative impact on action recognition performance. The presented solution therefore aims to keep the synthesized faces as close as possible to the originals, while

TABLE 5: High–level summary and comparison of the surveyed synthesis–based image–level B–PETs. Information is provided on the type of data the techniques were applied to, the attributes to be concealed and preserved, reversibility (Rev.) and privacy guarantees (Gua.), privacy mechanism used with the B–PETs, the utility preservation strategy, privacy target, and the dataset the techniques were tested on. Techniques are ordered chronologically for each sub–group.

Summary of Synthesis techniques											
Subgroup	Techniques	Applied to		Attribute		Rev.	Gua.	Mechanism	Utility	Target	Test dataset
		Video	Stills	Concealed	Preserved						
Image processing	Newton et al. (*k*–Same–Pixel), 2005 [15]	✗	✓	ID	NSA	✗	✓	Image averaging	RD	H	FERET [144]
	Bitouk et al., 2008 [181]	✗	✓	ID	Appearance	✗	✗	Replacement	RT	H	IHD
	Othman and Ross, 2014 [45]	✗	✓	GD	ID	✗	✓	Morphing	RT	M	MUCT [165]
	Mosaddegh et al., 2015 [182]	✓	✓	ID	NSA	✗	✗	Composing parts	RT	H	MultiPIE [194]
	Xu et al., 2015 [183]	✓	✓	ID	EX	✗	✗	Replacement	RT	H	FEI [195], Caltech Faces [196]
Statistical models	Gross et al. (*k*–Same–Select), 2005 [10]	✗	✓	ID	EX, GD	✗	✓	Selective averaging	RD	H	FERET [144]
	Mercier et al., 2005 [189]	✓	✓	ID	EX	✗	✗	AAM	RT	H	MMI [197]
	Gross et al. (*k*–Same–M), 2008 [30]	✗	✓	ID	EX, GD	✗	✓	AAM	RT	H	MultiPIE [194]
	Driessen and Durmuth, 2013 [186]	✗	✓	ID	NSA	✗	✓	PCA	RD	M	FERET [144]
	Chi and Hu, 2014 [188]	✗	✓	ID	EX	✗	✓	AAM	RT	H	JAFFE [198]
	Samaržija et al., 2014 [191]	✓	✓	ID	NSA	✗	✗	AAM	RT	H	IMM [199]
	Du et al. (GARP–Face), 2014 [187]	✗	✓	ID	EX, ET, GD	✗	✓	AAM	RT	H	MORPH [175]
	Meng et al. (*k*–Same–furthest–FET), 2014 [31]	✗	✓	ID	EX	✗	✓	AAM	RT	H	IMM [199]
	Sun et al. (*k*–Diff–furthest), 2015 [76]	✗	✓	ID	NSA	✗	✓	AAM	RT	H	IMM [199], LFPW [200]
	Sim and Zhang, 2015 [47]	✗	✓	Arbitrary	Arbitrary	✗	✓	AAM	RT	H	IHD
	Ohana et al., 2016 [201]	✗	✓	ID	NSA	✗	✗	AAM	RT	M	IHD
	Meng et al. (*k*–SameClass–Eigen), 2017 [65]	✗	✓	ID	EX	✗	✓	AAM	RT	H	BU–3DFE [202]
	Meng et al., 2017 [192]	✓	✗	ID	EX	✗	✓	AAM	RT	H	UNBC–McMaster [203]
	Prinosil et al., 2017 [75]	✗	✓	ID	NSA	✓	✓	AAM	RT	H	IHD
	Wang et al., 2018 [190]	✗	✓	Arbitrary	Arbitrary	✗	✓	AAM	RT	H	CK+ [204]
Deep learning	Chi and Hu, 2015 [205]	✗	✓	ID	NSA	✗	✓	Feature averaging	RT	H	MultiPIE [194]
	Rafique et al., 2016 [206]	✗	✓	ID	NSA	✗	✗	RBM	RT	H	Yale Faces [143]
	Brkić et al., 2017 [32], [207]	✓	✓	ID	NSA	✗	✗	GAN	RT	H	ChokePoint [208]
	Meden et al., 2017 [9]	✓	✓	ID	EX	✗	✗	GNN	RT	H	ChokePoint [208]
	Meden et al. (*k*–Same–Net), 2018 [33], [61]	✗	✓	ID	EX	✗	✗	GNN	RT	H	XM2VTS [157], CK+ [204]
	Ren et al., 2018 [71]	✓	✓	ID	Behaviour	✗	✗	GAN	RT	H	LFW [126], DALY [209], JHMDB [210]
	Guo et al. (*k*–Dive–Net), 2018 [211]	✗	✓	ID	EX	✗	✗	GNN	RT	H	BU–3DFE [202]
	Chen et al. (PPRL–VGAN), 2018 [212]	✓	✓	ID	EX	✗	✗	GAN	RT	H	FERG [213], MUG [214]
	Sun et al., 2018 [215], [216]	✗	✓	ID	EX	✗	✗	AE, GAN	RT	H	PIPA [125]
	Wu et al. (PP–GAN), 2018 [217]	✗	✓	ID	A, ET	✗	✗	AE, GAN	RT	H	MORPH [175]
	Pan et al. (*k*–SS–GAN), 2019 [218]	✗	✓	ID	EX	✗	✗	GAN	RT	H	RaFD [176], CelebA [173]
	Li and Lyu, 2019 [219]	✓	✓	ID	EX	✗	✗	Attribute transfer	RT	H	LFW [126], PIPA [125]
	Li and Lin (AnonymousNet), 2019 [220]	✗	✓	ID	Arbitrary	✗	✗	GAN	RT	H	CelebA [173]
	Hao et al. (UP–GAN), 2019 [221]	✗	✓	ID	A, EX, GD	✗	✗	GAN	RT	H	FaceScrub [222]
	Hukelås et al. (DeepPrivacy), 2019 [180]	✓	✓	ID	NSA	✗	✗	GAN	RT	H	WIDER Face [223], IHD (FDF)
	Croft et al., 2019 [74]	✗	✓	ID	GD	✗	✗	GNN	RT	H	RaFD [176], KDEF [224]
	Maximov et al. (CIAGAN), 2020 [225]	✓	✓	ID	EX	✗	✗	AE, GAN	RT	H	CelebA [173], MOTS [226], LFW [126]
	Cho et al. (CLEANIR), 2020 [227]	✓	✓	ID	EX	✗	✗	AE	RT	H	LFW [126], JAFFE [198], MUG [214]
	Proença (UU–Net), 2020 [228]	✓	✓	ID	ET, EX, GD	✓	✗	AE	RT	H	IHD (BIODI), MARS [229], P–DESTRE [230]

Symbol explanation: ID – identity, GD – gender, A – age, ET – ethnicity, EX – expression, RD – reduction, RT – retention, IHD – In house dataset, NSA – no specific attribute (information removed on all attributes), AAM – Active Appearance Model, PCA – Principal Component Analysis, GNN – Generative Neural Network, GAN – Generative Adversarial Network, AE – Autoencoder, RBM – Restricted Boltzmann Machine, H – Human, M – Machine.

still efficiently concealing identity information. Unlike many other works in this area, this research is focused on video data and not only still images.

Another solution exploiting GANs for image synthesis was proposed by Wu et al. [217] in the form of the *Privacy–Protective GAN* (PP–GAN). PP–GAN is a dedicated face deidentification model capable of generating high–fidelity, visually convincing deidentification results that retained a high structural similarity with the input images. Unlike *k*–Same–like algorithms, the model is able to generate deidentified faces from a single input image (instead of a set of images). PP–GAN uses a U–Net–based encoder–decoder network as the generator, a (relatively shallow) conv net as the discriminator. The model is learned in an adversarial setting using multiple learning objectives that ensure the quality of the generated deidentified output images.

A GAN–based face replacement framework for face obfuscation was introduced by Sun et al. in [216]. This work combines a data–driven method to image rendering with a parametric face model (a 3D Morphable Model – 3DMM [234]) in a two–stage procedure. In the first stage a surrogate face is rendered using the 3DMM. This rendering procedure conceals identity information, but preserves utility (e.g., facial expressions). In the second stage, the complete head is inpainted with a GAN model using the rendered face and the rest of the original input image.

A notable approach to face deidentification, called *AnonymousNet*, was designed by Li and Lin [220]. AnonymousNet combines several ideas from the literature into a powerful solution that can generate convincing and photo realistic face deidentification results. In the first step, the framework uses a large set of attribute classifiers to identify aspects of the faces that need to be concealed. Next, it uses a GAN model to generate surrogate faces with natural appearance and then maps the generated surrogates back into the originals. Finally, it adds adversarial noise (using DeepFool [235]) to add another layer of privacy enhancement to the overall framework. Despite building on different formal privacy schemes, the authors do not discuss whether AnonymousNet also comes with privacy guarantees.

A recent B–PET designed for utility preservation was introduced by Hao et al. in [221]. The presented *Utility–Preserving GAN* (UP–GAN) again generates surrogate face for deidentification, but is designed to preserve multiple soft–biometric attributes that do not reveal identity, such as age, gender, skin tone, pose, and expression. The generated faces appear natural and are used in a face swapping procedure to achieve deidentification. UP–GAN is benchmarked against *k*–Same, *k*–Same–Net, blurring, and pixelation at different resolutions and is shown not to yield highly competitive performance.

In [180], Hukelås et al. studied a GAN–based *face anonymization approach* that removes identity information from images through an inpainting procedure. The model used in this work is closely related to inpainting approaches based on conditional GANs (c–GAN) and is able to generate

surrogate faces based on conditional information in the form of a sparse set of facial landmarks (defining pose) and image background. Thus, the surogate faces are seamlessly incorporated into the source image during deidentification.

Conditional GANs were also used by Maximov et al. for their Conditional Identity Anonymization GAN model (CIAGAN) [225]. CIAGAN, in essence, implements a face swapping procedure using an encoder-decoder type generator network trained in an adversarial manner. For anonymization, the model first locates and masks the facial region in the input image using a landmarking technique and then inpaints the masked out region with a face of an artificially generated surrogate identity. During inpaitning, the output identity is determined by a one-hot encoding vector fed to the latent space of the CIAGAN model.

Next to GANs, other generative deep models have also been considered when developing B–PETs for facial images. In [227], for instance, Cho et al. described *CLEANIR*, a Variational Auto–Encoder (VAE) [252] based face deidentification technique. With CLEANIR, a disentangled representation (separating identity from other attributes) is learned in the VAE latent space. This disentangled latent space then allows to modify (and in turn conceal) identity information independently from other attributes and generate images that are identical to the originals in all aspects except identity. The resulting deidentified images are convincing, although not as sharp as those generated by GAN–based approaches.

Another approach based on auto-encoders, called *UU–Net*, was recently proposed by Proença [228]. In this work, a reversible deidentification is described, where two consecutive U–Nets are trained, one for deidentification and the second one for reversing the deidentification process. Thus, the first U–Net generates privacy-enhanced video data that facilitate video analytics and data sharing, whereas the second is intended to be private and available to authorities in case the original data is needed.

Attempts have also been made to devise privacy–enhancing techniques based on differential privacy. The work of Croft et al. [74] presents an initial attempt toward this goal using deep learning synthesis–based B–PETs. Here, the authors introduce a general framework for face obfuscation using differential privacy and apply it to a generative neural network (GNN). The proposed approach exhibits several desirable characteristics over techniques relying on by formal privacy models, such as k–anonymity, in that it is resilient to composition attacks (special type of linkage attack[7]) and does not need subject–specific set of images to perform obfuscation.

3.1.5 Comparison of synthesis techniques

Table 5 presents a summary and high–level overview of the techniques surveyed in this section. As can be seen, existing synthesis techniques use various privacy mechanism to remove or conceal information on biometric attributes from facial data. The latest approaches proposed in the literature are influenced heavily by advancements in deep learning and typically rely on adversarial techniques or generative deep models. B–PETs in this group are often able to target any selected specific biometric attribute, though the majority of research is still focused on deidentification and obfuscation of identity information. Existing solutions are mostly not reversible, but may come with privacy guarantees. Interestingly, deep learning solutions with provable privacy are so far still limited in the literature, but some pioneering work is currently being done here as well.

7. See Section 4.1.3 for details on attack models.

3.2 Representation–level techniques

Different from image–level techniques, which mostly focus on removing identity information from facial images, representation–level techniques predominantly focus on the removal of other (typically soft) biometric attributes, such as gender, age, or ethnicity, from biometric templates. The reason for this conceptual difference is in the fact that templates are typically constructed explicitly for the goal of identity inference (with consent), and should, therefore, ideally be free of other potentially sensitive information. We partition representation–level B–PETs into three distinct groups, i.e., *i) transformation–based* techniques, *ii) elimination–based* techniques, and techniques based on *iii) homomorphic encryption*. We describe the three groups in–depth in the following sections.

3.2.1 Transformation techniques

Transformation techniques try to remove sensitive information from biometric templates by transforming the feature space in such a way that information on predefined attributes is suppressed as much as possible or, ideally, is removed from the feature space altogether. This group also includes techniques that try to learn (from scratch) image representations without sensitive information. In essence, techniques following this strategy try to built models capable of transforming input face images into image representations containing only information that is not critical from a privacy perspective. Transformation–based techniques are in general highly selective and usually able to suppress only the targeted attributes, while not affecting others.

Several techniques have been proposed in the literature that fall into this group and exploit different mechanisms to ensure privacy. Feutry et al. [236], for example, proposed an *adversarial anonymization method* that allows to learn facial representations useful for expression, but not identity recognition. The core contribution of this work is a novel training objective that allows to simultaneously learn a predictor/classifier for a selected attribute (facial expressions in this work), while preventing the representation to be predictive of sensitive data attributes, e.g., identity.

Terhörst et al. [70] introduced a *Cosine–Sensitive Noise* (CSN) transformation to ensure soft–biometric privacy and suppress gender and age information in biometric templates. With CSN, a specific type of noise is added to the face representations such that soft–biometric information is masked, while identity information is not. CSN works in an unsupervised manner and unlike competing techniques from the literature does not need large amounts of annotated data to learn the privacy enhancement.

Morales et al. [46] proposed a supervised approach to soft–biometric privacy by learning dedicated deep models, called *SensitiveNets*. SensitiveNets use a modified version of a triplet loss objective function to learn feature space

TABLE 6: High–level comparison of the surveyed representation–level B–PETs. The table provides information on the attributes the techniques are trying to conceal, type of supervision, mechanism used for privacy enhancement, target of the privacy enhancement, and dataset(s) used to evaluate the technique.

Group	Techniques	Attribute		Supervision	Mechanism	Utility	Target	Dataset
		Concealed	Preserved					
Transformation	Feutry et al., 2018 [236]	ID	EX	Supervised	Adversarial objective	RD	M	JAFFE [198]
	Terhörst et al. (CSN), 2019 [70]	GD, A, ET	ID	Unsupervised	Noise addition	RD	M	ColorFERET [144]
	Morales et al. (SensitiveNets), 2019 [46]	GD; ET	ID	Supervised	Triple loss learning	RD	M	DiveFace [46], LFW [126], CelebA [173]
Elimination	Terhörst et al. (IVE), 2019 [66]	GD; A	ID	Supervised	Feature elimination	RD	M	ColorFERET [144]
	Bortolato et al. (PFRNet), 2020 [20]	GD	ID	Supervised	Disentanglement, elimination	RD	M	LFW [126], CelebA [173], Adience [131]
Homomorphic	Erkin et al., 2009 [237]	Arbitrary	ID	n/a	Paillier and DGK encryption	RD	M	AT&T (ORL) [149]
	Sadeghi et al., 2009 [238]	Arbitrary	ID	n/a	Garbled circuit scheme	RD	M	IHD
	Rahulamathavan et al., 2012 [239]	Arbitrary	EX	n/a	Paillier encryption	RD	M	JAFFE [198], MUG [214]
	Troncoso–Pastoriza et al., 2013 [240]	Arbitrary	ID	n/a	Gentry scheme	RD	M	XM2VTS [157], FERET [144], LFW [126]
	Xiang et al., 2016 [241]	Arbitrary	ID	n/a	FHE [242]	RD	M	IHD
	Ma et al., 2017 [243]	Arbitrary	ID	n/a	Paillier encryption	RD	M	LFW [126], Faces94 [244]
	Boddeti, 2018 [245]	Arbitrary	ID	n/a	BFV	RD	M	LFW [126], IJB-A [246], IJB-B [247], CASIA [248]
	Drozdowski et al., 2019 [249]	Arbitrary	ID	n/a	BFV, CKKS	RD	M	FERET [144]
	Kolberg et al., 2020 [250]	Arbitrary	ID	n/a	BFV, CKKS, NTRU	RD	M	FERET [144]
	Yang et al., 2021 [251]	Arbitrary	ID	n/a	CKKS	RD	M	LFW [126]

Symbol explanation: GD – gender, A – age, ET – ethnicity, ID – identity, EX – expression, RD – reduction, M – machine, DGK – Damgård, Geisler and Krøigaard cryptosystem, HSC – Hilbert Space-filling Curves, FHE – Fully Homomorphic Encryption, BFV – Brakerski/Fan-Vercauteren encryption, CKKS – Cheon-Kim-Kim-Song encryption, NTRU – N–th degree truncated polynomial ring.
n/a – not applicable, since the techniques are not learning based.

transformations that ensure, that sensitive information is removed from biometric templates. The approach is evaluated for suppression of gender and ethnicity information and is shown to ensure competitive performance on three public benchmarks.

3.2.2 Elimination techniques

Unlike transformation–based techniques, elimination techniques do not try to construct a feature space, in which information on certain attributes is suppressed, but instead aim to remove (or eliminate) elements of the image representation that carry most of the information on the targeted soft–biometric attribute(s). A strong (implicit) assumption made by techniques from this group is that the features in the computed face representations are mutually independent or at least the information on different biometric attributes is not distributed equally across the image representations.

A notable elimination technique, called *Incremental Variable Elimination* (IVE) was proposed by Terhörst et al. in [66]. IVE gradually eliminates components of biometric templates that contribute most to the prediction of a chosen attribute (i.e., age or gender). The algorithm is based on a decision tree ensemble that scores each variable in the face representation with respect to its importance for a specific recognition task. Variables most affecting attribute classification are then excluded from the representation. The authors show that it is possible to discard a considerable amount of information on sensitive attributes, while still maintaining high recognition accuracy. A shortcoming of this approach is the fact that after each elimination step a new impact–estimation model needs to be trained, which slows down the the elimination process.

To have better control over the information contained in the face representations prior to elimination, Bortolato et al. [20] proposed to learn a disentangled feature representation using a deep learning model, called *PFRNet*. PFRNet is an autoencoder that accepts original face templates as an input and then generates a disentangled representations in its latent space, in which identity information is separated from the attribute–related information. To ensure that sensitive information is removed, the attribute–related part of the latent space is discarded, while the identity–related part is used for verification purposes.

3.2.3 Homomorphic encryption techniques

Techniques from this group rely on homomorphic encryption to ensure that face representations, e.g., templates, are used only for the intended purpose. The main idea behind these techniques is to encrypt data in such a way that specific calculations are still possible in the encrypted domain. This allows for the design of template–comparison functions (applicable only to a predefined task) without the need for data decryption. Techniques from this group combine characteristics from both *data security* and *privacy–enhancing techniques* and are included in this survey due to their relevance for privacy protection. Homomorphic encryption schemes can in general be partitioned into three main sub–groups, i.e. [245], [254]:

- *Partially Homomorphic Encryption* (PHE) schemes that allow for *a single* mathematical operation (addition or multiplication) to be performed in the encrypted domain an *unlimited number of times*.
- *Somewhat Homomorphic Encryption* (SHE) schemes that allow to perform *different* mathematical operations (additions and mutliplications) in the encrypted domain a *limited number of times*.
- *Fully Homomorphic Encryption* (FHE) schemes that allow to conduct *different* mathematical operations (additions and mutliplications) directly in the encrypted domain an *unlimited number of times*. FHE schemes represent the most general type of homomorphic encryption approaches, but due to their generality are also computationally intensive.

One of the earliest B–PETs based on homomorphic encryption was presented by Erkin et al. in [237]. Here, the authors describe a PHE scheme that allows for executing Eigenface–based [184] matching directly in the encrypted domain. The presented scheme combines Eigenfaces with the Paillier [255] and Damgård, Geisler and Krøigaard (DGK) [256] cryptosystems and enables projecting facial images into an Eigen–space, comparing queries to templates in the database, and finding matching identities from the database. Because only a matching function is defined in the encrypted domain and the database is assumed to be private, this scheme allows for identity inference but

TABLE 7: High–level comparison of the surveyed inference–level B–PETs.The table provides information on the attributes the techniques are trying to conceal, type of supervision, mechanism used for privacy enhancement, privacy target, and dataset(s) used to evaluate the technique.

Techniques	Attribute		Supervision	Mechanism	Utility	Target	Test dataset
	Concealed	Preserved					
Terhörst et al. (NFR), 2020 [253]	GD, A, ET	ID	Supervised†	Negative templates	RT	M	ColorFERET [144], Adience [131]
Terhörst et al. (PE–MIU), 2020 [22]	GD, ET	ID	Unsupervised	Minimum information units	RD	M	LFW [126], Adience [131], ColorFERET [144]

Symbol explanation: GD – gender, A – age, ET – ethnicity, ID – identity, RD – reduction, RT – retention, M – machine.
† Supervision is applied for an intermediate step, not the privacy enhancement.

not (necessarily) extraction of other biometric attributes. A similar (but computationally simpler) SHE–based approach based on Garbled Circuits [257] and again focused on Eigenface was later proposed by Sadeghi et al. [238] and improved through the use of a fully homomorphic encryption (FHE) [242] scheme by Xiang et al. in [241].

Rahulamathavan et al. [239] introduced a PHE–based scheme for recognizing facial expressions in the encrypted domain. The schemes uses Local Fisher Discriminant Analysis (LFDA) to extract facial features and the Paillier cryptosystem [255] to perform homomorphic encryption. The authors show that comparable expression–recognition performance can be achieved in the original (plain) and encrypted domains under a suitable selection of hyper–parameters. Because the subspace (i.e., feature extractor) used is trained to be discriminative in terms of facial expressions, the scheme is expected to perform poorly for other biometric recognition tasks, such as, identity recognition for example.

Ma et al. [243] described a PHE encrypted face verification system, where the main idea is to extract facial features using deep neural networks and then encrypt the computed features with the Paillier cryptosystem [255]. To calculate distances during verification (between two encrypted vectors of facial features), a Hamming distance is used. The reported results show that verification performance in the original (plain) and encrypted domains is comparable. Because the complete solution is designed as a multi party system that only allows for distance calculations, other use cases of the data beyond similarity score computations are not possible.

Boddeti [245] presented a face recognition framework based on fully homomorphic encryption (FHE) for protecting user privacy in and preventing information leakage from biometric templates. The framework enables template matching in the encrypted domain and is based on the Fan–Vercauteren encryption scheme [258]. To trade–off matching accuracy and computational complexity, the author also explores batch–processing and dimensionality reduction of deep facial features. A similar FHE scheme [259] defined over Gabor features was presented earlier by Troncoso–Pastoriza et al. [240].

More recently, Drozdowski et al. [249] explored the use of homomorphic encryption in biometric systems designed for identification rather than verification as most prior work. The presented system combines template protection with existing homomorphic encryption schemes (Cheon-Kim-Kim-Song – CKKS [260] and Brakerski/Fan-Vercauteren – BFV [258]) and is demonstrated to result in comparable identification performance to the original and quantised feature vectors. In a follow–up to this work, Kolberg et al. [250], also investigated another homomorphic encryption scheme, i.e., N–th degree truncated polynomial ring (NTRU) [261]. A conceptually similar study based on CKKS homomorphix encryption was discussed by Yuan et al. in [251].

Homomorphic encryption is relevant also beyond the field of biometrics. The reader is referred to some of the existing surveys, e.g., [254], [262], [263] for more detailed information on this topic.

3.2.4 Comparison of representation–level techniques

Table 6 provides a high–level comparison of different charateristics of representation–level B–PETs. The characteristics compared include *i)* the targeted soft–biometric attribute(s) the techniques are trying to remove, *ii)* the attribute they are trying to preserve, *iii)* whether supervision is used during training (when training is needed), *v)* the mechanism used for privacy enhancement, *vi)* the strategy used to reduce the utility of the biometric data, *vii)* the target of the privacy enhancement, and *viii)* the test dataset the techniques were tested on. As can be seen, all surveyed techniques aim to reduce the biometric utility of the facial representation and in most cases preserve identity information. While different mechanisms are typically used for privacy enhancement for the transformation and elimination–based methods, these techniques share common characteristics in that they are computationally simple and, thus, applicable with limited computational resources. Homomorphic encryption based methods, on the other hand, typically do not require a learning stage, but are computationally more intensive. Most often these techniques are used to ensure template comparisons for (identity) recognition purposes, though work is out there on enabling recognition of other attributes (e.g., facial expressions) as well.

3.3 Inference–level techniques

Inference–level privacy enhancing techniques represent the most recent category of B–PETs. Consequently, only a few solutions from this category can currently be found in the literature. However, given their computational simplicity, on the one hand, and their efficiency, on the other, it is expected that research on this type of B–PETs will intensify in the future. Different from other categories of B–PETs, inference–level solutions don't alter only the biometric representation, i.e., the templates, but also modify the inference procedure (matching) to ensure privacy. Techniques from this group have so far only been applied for soft–biometric privacy, as also shown in Table 7, while approaches for removing identity information have not yet been proposed. Existing inference–level B–PETs are not reversible and do not come with formal privacy guarantees.

A notable solution from this group was presented by Terhörst et al. in [253]. The main idea of this work is to use the concept of *negative face recognition* (NFR) to ensure privacy. Using NFR two templates are generated for a given input face image. The first is a standard face template generated, for example, by recent deep face recognition models, whereas the second is a so–called negative template, which is stored in the systems database. The negative template is created by populating it with features that are intentionally different (but in a valid range) from the ones in the original (positive, standard) template. In this way, the negative template encodes (random) attribute information that is not present in the input image and the template itself exhibits only small similarities with the original (positive) template. Due to the design of the negative template, it is hard to infer sensitive soft–biometric information from the stored template. Identity verification, on the other hand, is still feasible, as the similarity computed for a mated comparison is by design smaller (on average) than the similarity for a non-mated template comparison. Thus, a special type of matching procedure is used that is based on dissimilarities – hence the name *negative face recognition*. The authors demonstrate the feasibility of their approach in soft–biometric privacy experiments (with age, gender, and ethnicity) on multiple face datasets.

Another inference–level technique was introduced recently by the same authors [22]. The proposed *Privacy–Enhancing* face recognition approach is based on *Minimum Information Units* (PE–MIU) and is completely unsupervised. As a consequence, PE–MIU is not limited to a predefined set of attributes that are targeted for suppression, but ensures privacy for any biometric attribute beyond identity. Similarly to the approach from [253], a special type of template is created by first partitioning the computed (original) template into smaller parts (called minimum information units – MIUs), and then randomly shuffling the location of the MUIs in the generated template. Because each template is encoded differently (using another random shuffling operation), a malicious user trying to infer soft–biometric information cannot learn a classification model from the stored templates even in the worst case white–box scenario. The approach still allows to conduct identity verification by designing a matching procedure that incorporates an (optimal) alignment step between a given "live" template and the MIUs of the modified template stored in the biometric system. The efficiency of the technique was presented in experiments demonstrating gender and ethnicity suppression with respect to verification performance using challenging datasets, such as LFW, Adience, and ColorFERET.

4 Evaluating Performance

A key issues with biometric privacy enhancing techniques (B–PETs) is how to quantify performance. As illustrated by the literature survey in the previous sections, many of the existing B–PETs rely on image processing and machine learning techniques and do not come with formal provable privacy guarantees. The performance of such techniques is, therefore, usually evaluated empirically using suitable datatsets and performance measures. Typical evaluations in this area include experiments that test for *i)* the efficiency of the privacy enhancement, *ii)* the biometric utility preserved after privacy enhancement, and *iii)* the robustness to attempts to reverse the privacy enhancement. However, we note that the evaluation methodology is not yet standardized and often varies from paper to paper. In the following sections, we provide an overview of the evaluation methodologies used most frequently in the literature, we describe existing evaluation frameworks and briefly review the datasets utilized in the literature for evaluation of biometric privacy enhancing techniques.

4.1 Evaluation methodology

4.1.1 Evaluating privacy enhancement efficiency

B–PETs use various mechanisms to remove (conceal) sensitive information from biometric data (at the image, representation, or inference levels). However, due to the characteristics of biometric data, its inherent variability and the (often) data–driven approach to privacy enhancement, this process is never perfect. Even with elaborate B–PETs, the possibility remains that sensitive information can (to some extent) still be recovered from the privacy–enhanced data.

To assess the efficiency of privacy enhancement, most of the existing work relies on automatic recognition techniques trained for extracting various attributes, such as identity, age, gender, ethnicity, or others, from the biometric data, e.g., [10], [20], [22], [33], [44], [45], [46], [69], [105], [106]. Experiments are typically conducted on both, the original and privacy enhanced data, and performance differences between the two groups of data are reported. In verification scenarios (with still images), for example, this involves matching the original and privacy enhanced data against the galleries stored in the database of a biometric system. For soft–biometric privacy problems the evaluation comprises recognition experiments aimed at predicting soft–biometric attributes from the original and privacy–enhanced data, and so on. Such an evaluation process generates two sets of results that are then compared to assess the efficiency of privacy enhancement. The performance assessment typically includes comparisons of performance curves, such as receiver operating charateristics (ROC) curves for binary recognition problems (e.g., identity verification, gender recognition) and cumulative match score curves (CMCs) for multi–class problems (e.g., identification, ethnicity recognition), or scalar accuracy measures, such as equal error or rank–1 recognition rates.

While the evaluation methodology described above is most widely used in the literature, attempts have also been made to combine results generated on the original and privacy–enhanced data into scalar measures that quantify privacy enhancement in the form of a single number. Korshunov et al. [105], for example, proposed to measure the *privacy gain* (PG) produced by B–PETs (aiming at identity suppression) as:

$$PG = (1 - R_p) - (1 - R_o), \tag{1}$$

where, R denotes the recognition performance, o stands for the results on the original data, and p for the results obtained on the privacy enhanced data. PG is positive if the privacy enhancement degrades the recognition performance compared to the original data and equals 1 in the ideal case

when the performance drops from $R_o = 1$ to $R_p = 0$ due to application of the evaluated B–PET.

A similar performance measure was also proposed by Terhörst et al. [253] in the context of soft–biometric privacy. Here, the authors defined an attribute *suppression rate* (SR) that measures the difference in attribute–prediction accuracy with and without privacy enhancement, i.e., A_p and A_o, respectively:

$$SR = \frac{A_o - A_p}{A_o}. \tag{2}$$

The presented evaluation methodology and performance scores are used predominantly with automatic recognition techniques. However, several studies also employ *human–centered experiments* to compute performance metrics and evaluate how the privacy enhancement impacts human perception of sensitive information in biometric data – see [264] for a notable example for video data. An aspect that is also often of interest is the *computational complexity and processing speed*, which are frequently assessed when evaluating B–PET efficiency, see [100] for an example.

4.1.2 Evaluating biometric utility

B–PETs are designed to provide a trade–off between privacy protection and biometric data utility. When evaluating performance, it is, therefore, paramount to consider the efficiency of privacy enhancement with respect to the preserved utility of the biometric data.

Because the meaning of the term *biometric utility* is very much application dependent, standard evaluation strategies typically define a *secondary task* that needs to be evaluated after privacy enhancement. With video surveillance systems, for example, utility is often linked to the ability to extract behavioural information [120], with deidentification techniques, utility is frequently measured with the ability to infer facial expressions after deidentification [9], [10], while for soft–biometric privacy problems, utility is routinely defined with the recognition performance after the removal of selected soft–biometric attributes [20], [66]. To quantify the amount of preserved utility, existing work is commonly looking at the performance differences achieved with automatic recognition techniques on the selected secondary task with the original and privacy–enhanced data. The smaller this difference, the higher the level of utility preservation.

While separate experiments are usually conducted to assess the efficiency of privacy enhancement and utility preservation aspects of B–PETs, it is crucial that the results are interpreted jointly. To facilitate such comparisons, the so–called *privacy–gain identity–loss coefficient* (PIC) was recently introduced for measuring soft–biometric privacy performance [70]:

$$PIC = \frac{AE_p - AE_o}{AE_o} - \frac{RE_p - RE_o}{RE_o}. \tag{3}$$

The privacy–utility trade–off measured by PIC pins attribute prediction errors AE against verification (or recognition) errors RE on both, the original (o) and privacy enhanced (p) data. Positive PIC values suggest that the privacy gain is higher than the loss in recognition performance, with higher values indicating better privacy enhancement. While originally proposed for evaluating soft–biometric privacy techniques, the idea behind the coefficient is in general applicable to other privacy enhancing problems as well.

Alternative strategies to quantifying utility preservation also rely on measuring the structural similarity (SSIM) or peak–signal–to–noise ratio (PSNR) between the original in privacy enhanced data. Here, the similarity with the original data is used as a proxy for biometric utility. This strategy is most often seen with image–level privacy enhancing techniques applied to surveillance footage, where computationally simple techniques, such as filtering, are used for privacy enhancement [106]. It is less suitable for more complex techniques that rely, for example, on image synthesis or other generative approaches.

4.1.3 Evaluating robustness

Another important aspect of B–PETs is their robustness to attempts at reconstructing (or recovering) the concealed information. The evaluation methodologies described in the previous two sections often assume a *"vanilla"* scenario, where recognition experiments are conducted on the privacy–enhanced data with machine learning models trained on the original data. In deidentification settings, for example, a pretrained recognition model is typically applied on deidentified images to assess the efficiency of privacy enhancement. While such an approach is reasonable for application domains, where it can be assumed that biometric data will only be processed by automatic recognition techniques, (e.g., for profiling and tagging in social media, for targeted advertising, etc.), it is critical that the robustness of privacy enhancing techniques is evaluated in more comprehensive attack scenarios as well.

Several such attack scenarios are considered in the literature in addition to the vanilla scenario discussed above, including [23], [120]:

- *Parrot or imitation attacks:* This scenario assumes a white box attack, where an adversary trying to recover sensitive information has access to the privacy enhancement mechanism. To extract information about biometric attributes, the adversary applies the privacy mechanism on some training data and learns a classifier in the privacy enhanced domain. In recognition settings, the adversary applies the privacy enhancement on the gallery images and then simply matches data in the privacy enhanced domain [9], [33], [120].
- *Reconstruction attacks:* This scenario again assumes a white box attack where the adversary has access to the privacy enhancement procedure. However, instead of matching the privacy enhancement on the reference data, the adversary now tries to invert it. Evaluating robustness to reconstruction attacks, hence, involves solving an inverse (privacy enhancement) problem and evaluating the amount of sensitive information contained in the recovered data [120]. Alternatively, a reconstruction attack may be implemented within a grey or black box attack scenario, where the reconstruction attempt is made based on limited or no knowledge of the mechanism used for privacy enhancement, see, e.g., [43].
- *Linkage attacks:* The last scenario assumes an attacker is able to link the privacy enhanced data with other

sources of information that, in combination, reveal sensitive details bout the original data. These sources of information can represent parts of the biometric data not considered critical from a privacy perspective, prior knowledge, multiple instances of the same privacy–enhanced data, or external sources of information, for which a link with the privacy–enhanced data can be established [23], [74], [265], [266].

The first two types of attacks are relatively straight forward to implement and are used regularly to evaluate the robustness of privacy enhancing techniques, e.g., [9], [43]. The robustness to linkage attacks, on the other hand, is more difficult to evaluate and may require assessment of data not subject to privacy enhancement. For example, clothing, background, body parts, or hairstyle in facial images are often excluded from privacy enhancement, but may reveal sensitive information, as shown in [265], [267].

We note that both, the vanilla–type evaluations as well as the attack experiments discussed above, aim at estimating the risk associated with recovering suppressed biometric attributes. This is risk of often referred to as *re–identification risk* in the deidentification literature, but we use the term *attribute recovery risk* in this work to also account for attributes other than identity when discussing performance evaluations.

4.2 Evaluation frameworks and performance studies

Several evaluation frameworks have been presented over the years with the goal of providing standardized experimental methodologies for the assessment of B–PETs. Below, we provide a short summary of the most relevant ones.

Dufaux and Ebrahimi [268] were among the first to propose using a standardized face recognition framework to assess the efficiency of different privacy enhancing techniques. In their study, the authors utilized the CSU Face Identification Evaluation System (FIES) and the FERET dataset to evaluate the efficiency of five different deidentification methods (pixelization, Gaussian blurring, scrambling by random sign inversion, and scrambling by random permutation) and found that simple blurring and pixelation offers only limited privacy protection, while scrambling techniques were deemed most efficient. The value of this work is in the attempt to quantify privacy protection using automated face matchers.

Korshunov et al. [105] extended the evaluation framework of Dufaux and Ebrahimi [268] and capitalized on the need to evaluate B–PETs both from a privacy protection as well as from an utility preservation perspective. Face detection performance was used as a metric for data utility (or intelligibility) and (the inverse of) face recognition performance as a metric for privacy protection. Three privacy enhancing techniques were considered, i.e., blurring, pixelation, and masking, and evaluated on three face datasets, i.e., FERET, SCFace, and ChokePoint.

Badii et al. [104], [269] presented a holistic privacy impact assessment framework for B–PETs applied to video surveillance data. The authors proposed an evaluation framework based on five different criteria, i.e., efficacy, consistency, disambiguity, intelligibility, and aesthetics, that measured different aspects of the privacy enhancement with specifics of surveillance technology in mind. To demonstrate the feasibility of their framework, the authors evaluated several simple privacy enhancement methods (masking, blurring, pixelation, resampling, and scrambling), and found scrambling to be the most efficient.

Erdélyi et al. [106] described an objective evaluation framework for assessing visual privacy enhancing methods with video surveillance data. The framework defined performance scores and evaluation protocols that measure the privacy protection achieved as well as the preserved levels of data utility. In addition to the traditional frame–by–frame evaluation approach used regularly in the literature with video data, the framework also introduced two new evaluation approaches based on aggregated and fused frames. The application of the framework was demonstrated with eight selected privacy enhancing techniques.

More recently, Terhörst et al. [21] introduced a comprehensive framework for the evaluation of soft–biometric privacy enhancing techniques. Here, the authors defined separate evaluation protocols for training–free and learning–based techniques based on Kerckhoffs's principle of cryptography. They also proposed performance measures to quantify performance as well as recognition–attribute plots to visualize results in a comprehensive manner.

4.3 Datasets

Despite the considerable research effort directed toward B–PETs, there are currently only few datasets available that are dedicated specifically to the problem of (visual) privacy in face biometrics. An overwhelming majority of the research on B–PETs for face biometrics is, therefore, conducted on standard face datasets typically used for face recognition, face analysis, and related tasks. In this section we present an overview of the datasets most frequently utilized in experiments with B–PETs from the literature. The reader is referred to the summary tables in Section 3 for information on the datasets used in specific works.

4.3.1 Standard datasets

Research on image–level B–PETs applied to video footage is often concerned with privacy in surveillance scenarios and, therefore, uses standard video surveillance datasets, such as ChokePoint [208] and VTM [137] for performance evaluations. In house datasets that are not publicly available are also utilized frequently [54], [63], [98], [110], [134], [270].

For B–PETs (image–, representation–, and inference–level) operating on still images evaluations are routinely done on standard face datasets. This includes both earlier datasets, captured in controlled, laboratory–like conditions with well illuminated faces captured mostly in frontal pose, without varying facial expressions and without occlusions, as well as more recent datasets, typically collected from the web (a.k.a., in the wild), which exhibit a significantly higher degree of variability. Examples from the first group of controlled datasets include AT&T (ORL) [149], Yale Faces [143], AR–face [174], JAFFE [198], KDEF [224], Caltech Faces [196], FERET and ColorFERET [144], CMU PIE [154], XM2VTS [157], IMM [199], MMI [197], BU–3DFE [202], MORPH [175], CK+ [204], FEI [195], RaFD [176], MUCT [165], MUG [214],

MultiPIE [194], UNBC–McMaster [203] and FERG [213]. Representative examples from the second group of unconstrained dataset are LFW [126], PUBFIG [155], LFPW [200], Adience [131], FaceScrub [222], CelebA [173], PIPA [125], WIDER Face [223] and UTKFace [177].

4.3.2 Dedicated datasets

Existing datasets collected for studying problems in the area of visual privacy, such as PicAlert [271], YourAlert [272], Campus Face Set [273] VISPR [169], Visual Redactions [35], or VizWiz–Priv [274], (in a significant part) also include facial images, but typically aim to address privacy concerns associated with visual data that go beyond biometric privacy enhancement. As a result, only limited work related to B–PETs has been done with these datasets so far.

A notable dataset dedicated to research on B–PETs is the Privacy Evaluation Video Dataset (PEViD) from [129]. The dataset is commonly used with research on B–PETs for video surveillance applications. PEViD ships with XML–based annotations of various privacy regions, including faces, accessories, skin regions, hair, body silhouettes, and other personal information, including their descriptions. The dataset consists of 65 video sequences (16 seconds each) of full HD resolution covering different video surveillance scenarios: walking, fighting, stealing, and dropping bag, in outdoor and indoor environments, as well as during day and night conditions. A subset containing 20 video sequences of the dataset was annotated using the ViPER–GT annotation tool. An extension of the dataset with ultra high definition (UHD) video was presented in [275].

In [46], Morales et al. presented another dataset for B–PET evaluton, called DiveFace. The dataset is designed for studying soft–biometric privacy problems and contains still image data on 24,000 identities. DiveFace was created from the publicly available MegaFace dataset MF2 [276]. The data is grouped and balanced according to gender (male, female) and ethnicity (three groups) labels and is, therefore, well suited for research on soft–biometric privacy.

Another relevant dataset, called PA–HMDB51, was recently introduced in [277]. The dataset is designed for development of biometric privacy enhancing techniques aimed at privacy protection (in videos) across five labeled privacy attributes (i.e., skin color, gender, faces, nudity, and personal relationship). The utility preservation aspect is covered by the ability to conduct action recognition. Thus, videos with various actions are included in the dataset. The entire dataset contains 580 videos with frame–level annotations.

5 Relevant Standards and Regulations

As privacy is often considered as a fundamental human right [278], [279], [280], several regulations have been adapted to ensure privacy in different applications. Moreover, to motivate consistent understanding of privacy-related concepts and technologies, a number of standards have been developed, or are currently under development. In the following, Section 5.1 describes privacy regulations and standards that are explicitly designed for biometric data. Section 5.2 elaborates on regulations and standards related to privacy in biometrics but are not explicitly designed for biometric data. Table 8 provides a summary of the documents we base our discussion upon.

5.1 Privacy in Biometrics

In 2019, the Biometric Institute, which include a number of biometric end-users and vendors, proposed a comprehensive privacy guideline for biometrics [281] taking into account the new General Data Protection Regulation (GDPR) [284]. The guideline addresses complaints of users who suffered discrimination or damage as a result of biometric systems. It demands stronger privacy protections for automated data collection and defines the role of audits and privacy impact assessments. Moreover, it provides advice on the management of data breaches and capitalizes on the users right to delete their biometric records. In total, 16 principles are described in the guideline to maintain a strong privacy environment.

Several ISO standards have been proposed to ensure the privacy in the context of biometric systems, such as ISO/IEC 30137-1:2019 [282], ISO/IEC TR 24741:2018 [283], and ISO/IEC 24745:2011 [287]. In ISO/IEC 30137-1:2019 [282], the use of biometrics in video surveillance systems is described. It focuses on the design and specifications for these systems, among others, to ensure privacy. ISO/IEC TR 24741:2018 [283] focuses on the application aspects of biometrics. In terms of privacy, it contains a set of principles including that individuals be informed about the purpose of the collected data, who is requesting it, how their data is protected from unauthorized access and modification, and how long the data will be stored. Concerning privacy–enhancing mechanisms, it mainly focuses on template protection properties, such as irreversibility, unlinkability, and renewability. Similarly, ISO/IEC 24745:2011 [287] provides requirements and guidelines for secure and privacy-compliant management and processing of biometric data.

In 2013, the Organisation for Economic Cooperation and Development (OECD) renewed their 30-year-old privacy recommendations in a new privacy framework [286]. This took place due to significant changes in the volume of personal data used, the range of analytics involving this data, their social and economic value, the extended threats to privacy, the global availability of personal data, and the number of actors capable of putting privacy at risk. Two themes can be found in the updated Guidelines. The first is the focus on practical implementations of privacy protection and the second is the need for greater efforts to address the global dimension of privacy through enhanced interoperability. Consequently, many new concepts were introduced such as national privacy strategies, privacy management programs, and security breach notifications. The guidelines especially focus "on the human body as information" in the context of identity and biometrics, including face biometrics.

In 2017, the Japanese Act on the Protection of Personal Information was enforced [285]. The act states that individual identification codes, such as facial features, have to be treated carefully. The capture of such data requires a public announcement of the purpose or to notify the subject directly. Additionally, in case of data breaches involving privacy-sensitive data, relevant authorities as well as the Personal Information Protection Commission have to be provided with details of the breach.

In the United States, Illinois was one of the first states that regulates the collection of biometric information. In

TABLE 8: An overview of regulations and standards related to privacy in biometric systems. The top rows refer specifically to biometrics. The bottom rows refer to general privacy-protection regulations and standards that effects biometric systems processes.

Year	Name	Type
2019	Privacy Guidelines of the Biometric Institute [281]	Guideline
2019	ISO/IEC 30137-1:2019 Information technology — Use of biometrics in video surveillance systems [282]	ISO document
2018	ISO/IEC TR 24741:2018 Information technology — Biometrics — Overview and application [283]	ISO document
2018	General Data Protection Regulation (GDPR) of the European Union [284]	Regulation
2017	Japanese Act on the Protection of Personal Information [285]	Regulation
2013	Organisation for Economic Cooperation and Development (OECD) Privacy Framework [286]	Regulation
2011	ISO/IEC 24745:2011 Information technology — Security techniques — Biometric information protection [287]	ISO document
2008	Biometric Information Privacy Act (BIPA) [288]	Regulation
2019	ISO/IEC 27701:2019 Security techniques — Privacy information management — Requirements and guidelines [289]	ISO document
2018	ISO/IEC 20889:2018 Privacy enhancing data de-identification terminology and classification of techniques [290]	ISO document
2018	California Consumer Privacy Act (CCPA) [291]	Regulation
2017	ISO/IEC 29134:2017 Information technology — Security techniques — Guidelines for privacy impact assessment [292]	ISO document
2015	ISO/IEC 29190:2015 Information technology — Security techniques — Privacy capability assessment model [293]	ISO document
2011	ISO/IEC 29100:2011 Information technology — Security techniques — Privacy framework [294]	ISO document
1950	Article 8 of the European Convention of Human Rights [295]	Regulation
1948	Article 7 of the United Nations Universal Declaration of Human Rights [280]	Regulation

2008, the Biometric Information Privacy Act (BIPA) [288] was passed by the Illinois General Assembly. This act requires companies doing business in Illinois to obtain consent from individuals before the collection of their biometric data.

One of the most important regulations to ensure individual's privacy is the General Data Protection Regulation (GDPR) of the European Union (EU) [284]. EU's GDPR establishes a harmonized framework within the EU, including the right to be forgotten, unambiguous, and affirmative consent. Moreover, it imposes penalties for non-compliance with these rules. This applies for "GDPR special category data" including biometric data used to identify individuals. GDPR protects EU citizens, as well as long-term residents, from the distribution of their information without their consent [296]. Because of its sensitiveness, the collection, storage, transmission, and processing of biometric data needs a legitimate and lawful reason [297]. However, the regulation contains specific exceptions, e.g. if the consent has been explicitly given or if it is critical for legal claims, public health, or social security. GDPR aims to give back control over personal data to European citizens, while simplifying the regulatory framework for companies [296]. It explicitly states that consent must be given before the collection of personal data and individuals can withdraw their consent at any time. If commercial entities do not make significant efforts to secure the "GDPR special category data", they can be hit with massive penalties. It is important to note that the GDPR applies to Non-EU organizations if they process personal data of EU subjects. This gives the GDPR a global reach.

5.2 General privacy standards and regulations

There are many regulations and standards that addresses privacy without explicitly discussing biometric systems. In ISO/IEC 27701:2019 [289], requirements and guidelines for information management of privacy-sensitive information is specified for establishing, implementing, maintaining, and continually improving a privacy-specific information security management system. ISO/IEC 20889:2018 [290], defines the terminology of de-identification methodologies and classifies these techniques according to their characteristics and applicability for reducing the risk of re-identification. In ISO/IEC 29134:2017 [292], guidelines for privacy impact assessments (PIA) in the context of security applications are given as well as the structure and contents of a PIA report. ISO/IEC 29190:2015 [293] provides organizations with guidance on how to assess their capability to manage privacy-related processes with a privacy capability assessment model.

While in the United States, no general and comprehensive federal law regulates the handling of all privacy-sensitive information, such as biometric data, some states have introduced their own laws. Representative for these, we emphasize the California Consumer Privacy Act (CCPA). The CCPA enhances privacy rights and consumer protection for residents of California. Since California is the fifth-largest economy in the world and home of several large technology firms, it is considered as a trend-setting state for data protection and privacy in the US [296]. The CCPA provides residents of California with several rights, such as the right of disclosure or access, or the right to be forgotten. In March 2020, a New York State law Stop Hacks and Improve Electronic Data Security (SHIELD) became effective. The SHIELD act requires the implementation of a cyber-security programs and protective measures for residents of the New York state [298].

The European Convention of Human Rights guarantees the "right to respect for private and family life" [278]. Aspects of the right to private life include the physical and psychological integrity of a person, personal data, reputation, names, and face images [279]. Also in the United Nations Universal Declaration of Human Rights [280] privacy is defined as a fundamental right for humans. More precisely, it states that "no one shall be subjected to arbitrary interference with his privacy, family, home or correspondence" [299]. This demonstrates that privacy is an inherently important property that should be especially be included in

security and biometrics, and thus face recognition systems.

6 Open Issues and Future Challenges

While the interest in privacy enhancing solutions for biometric systems is growing and significant progress has been made in this field over recent years, there are still several open issues that need to be addressed as part of future research activities. In this section, we provide a summary of the main challenges currently associated with biometric privacy enhancing techniques in our view.

6.1 Quantifiable and provable privacy

The ability to objectively quantify performance of privacy enhancing techniques is of paramount importance for a number of reasons: *i)* it allows to assess the quality of the privacy enhancement in its suitability for deployment in biometric systems, *ii)* it enables comparisons between competing solutions, and *iii)* it facilitates technological advances by providing clear efficiency and success criteria. Currently, different strategies and evaluation methodologies are used in the literature to quantify the performance of biometric privacy enhancing techniques, which makes it difficult to objectively compare results and identify prospective future research directions. While a cross–section of the (currently) most frequently used evaluation strategies in this field was presented in Section 4, we note that there is still not generally accepted consensus on how to quantify performance. Research into evaluation methodologies and performance measures is, therefore, needed to move the field forward and provide common performance criteria that apply to B–PETs at the image, representation, and inference levels and a wide range of biometric attributes. Standardization efforts aimed the performance quantization may also be needed.

Another important aspect of B–PETs related to performance is privacy guarantees. Existing provable privacy models, such as k–anonymity, are based on strong assumptions that make it difficult to apply them to real–world problems. It is therefore not clear how to extend such models: *i)* beyond closed–sets of still images, e.g., to video data or open–set problems, *ii)* to attributes other than identity, or *iii)* to B–PETs operating at the representation or inference levels. Novel ideas are needed to provide privacy guarantees while ensuring biometric utility for more realistic deployment scenarios. Privacy enhancing models built around ϵ–*differential privacy* [82] may be able to relax some of the assumption associated with k–anonymity, but the research in this direction is so far still limited [73], [74].

6.2 Generalization and robustness

Contemporary B–PETs are often built (or learned) in a supervised, data–driven manner. As such, they suffer from similar shortcomings as related machine learning models in that they *i)* need to generalize to different classification models (e.g., used in an attack scenario to recover sensitive information), *ii)* be applicable to input data of different characteristics, and *iii)* provide a consistent trade–off between privacy and biometric utility across a wide variety of conditions. Devising privacy–enhancement techniques that generalize well across all outlined settings is a challenging tasks that is not solved yet to a satisfactory extent. Techniques built in an unsupervised manner e.g., [22], [253] or ensemble solutions, e.g., [12], may offer ideas that can help with the generalization ability of future B–PETs.

The robustness of B–PETs is another challenge that will have to be addressed in the future. Some of the existing work in this area considers privacy attacks when evaluating the performance of B–PETs. However, the majority of evaluation in the literature assumes *vanilla* or *zero–effort* experimental scenarios, where it is assumed that no attempt is made to recover the concealed biometric attributes. With experimental evaluations focusing more on the attack scenarios, discussed in Section 4.1.3, we expect the robustness of B–PETs to become a central research topic going forward.

6.3 Controllable privacy

Existing B–PETs are usually designed to either remove specific attributes from the data (for utility reduction strategies) or alternatively, to retain specific attribute and remove all others (for utility retention strategies). Thus, a decision about which attributes to keep and which to remove is typically made up front. While such privacy enhancing mechanisms are most common in the literature, they assume that the attributes most critical from a privacy perspective: *i)* are known in advance, *ii)* do not change over time, and *iii)* are equally important to all individuals and application scenarios. B–PETs, built around the recently introduced concept of *controllable privacy*, e.g., [47], relax these assumptions and incorporate mechanism that allow individuals to explicitly specify which attributes of their data to conceal and which to preserve. Such an approach ensures a higher level of flexibility and is, for example, relevant in the context of social media, where different people may have different preferences about which information to make publicly available in which not when sharing images online. While B–PETs in line with this concept may come at the expense of more complex privacy enhancing mechanism, we expect to see increased interest in this direction from the biometric community in the future.

6.4 Public benchmarks

To date, most of the research on B–PETs is conducted on datasets collected originally for either face recognition or video surveillance. Only a limited number of dedicated public datasets exists that aim at evaluating B–PETs, and even those don't necessarily come with predefined experimental protocols and performance measures. To move the field forward, large, representative public datasets with a well defined experimental methodology are needed. Such dataset allow to compare different privacy–enhancing techniques under a common framework and in consistent settings. Moreover, to the best of knowledge, there are currently no datasets available that would aim at the evaluation of privacy–enhancing techniques at the representation or inference–level. We note again that performance evaluation protocols have been proposed for soft–biometric privacy techniques recently [21], but these are not tied to any public dataset yet.

6.5 Multi–level privacy enhancement

Research on B–PETs is usually limited to a single application level within biometric systems. Thus, techniques have been proposed in the literature that operate at either the image, representation, or inference level. However, it may be beneficial to apply privacy enhancement across multiple levels. For example, representation level techniques may be applied jointly with inference level approaches, resulting in solutions with better characteristics. Such multi–level techniques may also help to address challenges related to generalization and robustness.

6.6 Fairness and bias

The issue of fairness in data-driven computer vision systems has gained prominence, especially due to the presence of potential biases in the training data [300]. Since biometric systems rely heavily on training data, it is imperative that evaluation methodologies explicitly address the issue of fairness and bias [301], [302], [303], [304].

The erosion of user privacy isn't the only critique being leveled against face recognition at the present time. Researchers have shown that some of the datasets used for face recognition algorithm development have a bias towards lighter skinned faces, causing face recognition algorithms to exhibit significant differences in matching accuracy between darker and lighter skin colors [305]. Other work in this vein has gone on to demonstrate this specific problem in Amazon's Rekognition face analysis platform [306]. Thus, this is not merely a hypothetical problem — real face recognition applications have been demonstrated to exihibit biases in their performance. It is expected that this problem will also manifest itself in B-PETs, but as of this writing, that particular aspect has not been studied to any great extent in B-PETs technology. Work must be done to avoid bias and promote fairness in B-PETs before they are deployed in real-world applications.

6.7 Visual privacy beyond faces

The interest in privacy–enhancing mechanisms is not limited only to facial images. With the development of automatic recognition techniques and improvements in their capabilities, it is today possible to extract potentially sensitive information from a wide variety of biometric modalities, e.g., [307], [308], [309], [310]. B–PETs are, therefore, becoming increasingly relevant across the board and research efforts, focusing on privacy protection with modalities other than faces are expected to intensify in the future. While work on this topic is already underway, e.g., [32], [86], [89], [92], [93], [95], we anticipate to see synergies between these research activities and exchange of ideas among solutions targeting specific modalities going forward.

Furthermore, while research on B–PETs has traditionally focused on a single modality at the time, multiple modalities will have to be considered jointly during privacy enhancement, as multi–modal biometric system become more prevalent. Future research is, therefore, expected to focus increasingly on multi–modal B–PETs and consider correlations and interdependencies between the information contained in different biometric traits.

7 Conclusion

With further improvements in biometric recognition technology and its deployment in an ever increasing number of applications domains, privacy concerns associated with biometrics are only expected to increase in the years to come [311]. A considerable amount of research has already been done to address such concerns and make the benefits of automated recognition techniques available to individuals, while also taking measures to minimize the impact of the technology on individuals' privacy. As discussed in this survey paper, a significant portion of this research is focused on biometric privacy enhancing techniques (B–PETs) that aim to remove sensitive information from biometric data, while preserving only essential information, needed for a specific purpose. Such techniques try to strike a balance between the utility of the biometric data and the level of privacy protection for individuals.

This survey aimed at presenting a comprehensive introduction into privacy–related research in biometrics and reviewing existing work on B–PETs for face biometrics. As seen from the literature review, techniques have been proposed over the years that operate at different levels of a typical biometric processing pipeline and exploit a wide range of ideas and approaches to privacy enhancement. While some of these techniques target raw biometric data, i.e., facial images, others aim at modifying derived representations, i.e., templates, or even components of the biometric system, i.e., the matching and/or classification stages. These techniques typically target different application scenarios, but share a common characteristic in that they strive to mitigate privacy concern originating from the use of biometric systems.

To place the ongoing research on B–PETs in a broader context, the survey also discussed existing regulations and standards related to privacy and biometrics. These include recent privacy laws and regulations that have a direct impact on future requirements of biometric systems, but also standards (both for biometric as well as more general information technology) that impact privacy enhancing solutions and their deployment. These initiatives demonstrate the interest of legislative and standardization bodies to provide (legal and technological) frameworks for the development future privacy–sensitive biometric solutions, but also point to the importance of the topic of privacy in modern society.

The progress in B-PETs has been evident over the last few years, but there are still multiple key issues that have to be addressed in the future to make them applicable widely in real–life (deployed) biometric systems. As discussed above, a critical component that can drive the field forward is a well defined evaluation methodology with clear goals and performance criteria as well (challenging) large–scale benchmarks publicly available to the research community. Current research often relies on different evaluation strategies and non–standard datasets that make it difficult to compare results across publications and identify prospective research directions. B–PETs that offer a high degree of flexibility, generalize well over classification models and are robust with respect to a wide variety of characteristics of the input data are also needed. These and related challenges will undoubtedly represent research priorities related to

biometric privacy enhancement in the years to come.

ACKNOWLEDGMENTS

This research was supported in parts by the ARRS Project J2–1734 "Face deidentification with generative deep models", and ARRS Research Programs P2–0250 (B) "Metrology and Biometric Systems" and P2–0214 (A) "Computer Vision". This research work has also been partially funded by the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE. Ross was supported by the National Science Foundation under Grant 1618518.

REFERENCES

[1] A. Acquisti, R. Gross, and F. D. Stutzman, "Face recognition and privacy in the age of augmented reality," *Journal of Privacy and Confidentiality*, vol. 6, no. 2, 2014.

[2] D. Alba. "A.C.L.U. Accuses Clearview AI of Privacy 'Nightmare Scenario' ". [Online]. Available: https://www.nytimes.com/2020/05/28/technology/clearview-ai-privacy-lawsuit.html

[3] R. Heilweil, "Big tech companies back away from selling facial recognition to police. That's progress." 2020, accessed: 2021-02-04. [Online]. Available: https://www.vox.com/recode/2020/6/10/21287194/amazon-microsoft-ibm-facial-recognition-moratorium-police

[4] Thales Group Website. Biometric Data and Data Protection Regulations (GDPR and CCPA). [Online]. Available: https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/biometrics/biometric-data

[5] A. Chattopadhyay and T. E. Boult, "PrivacyCam: a Privacy Preserving Camera Using uCLinux on the Blackfin DSP," in *Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–8.

[6] T. Winkler and B. Rinner, "Trustcam: Security and privacy-protection for an embedded smart camera based on trusted computing," in *International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2010, pp. 593–600.

[7] S. Ribaric and N. Pavesic, "An overview of face de-identification in still images and videos," in *Automatic Face and Gesture Recognition (FG)*, vol. 04, 2015, pp. 1–6.

[8] S. Ribaric, A. Ariyaeeinia, and N. Pavesic, "De-identification for privacy protection in multimedia content: A survey," *Signal Processing: Image Communication*, vol. 47, pp. 131 – 151, 2016.

[9] B. Meden, R. C. Mallı, S. Fabijan, H. K. Ekenel, V. Štruc, and P. Peer, "Face deidentification with generative deep neural networks," *IET Signal Processing*, vol. 11, no. 9, pp. 1046–1054, 2017.

[10] R. Gross, E. Airoldi, B. Malin, and L. Sweeney, "Integrating utility into face de-identification," in *International Conference on Privacy Enhancing Technologies (PETS)*, Berlin, Heidelberg, 2006, pp. 227–242.

[11] V. Mirjalili, S. Raschka, and A. Ross, "Gender privacy: An ensemble of semi adversarial networks for confounding arbitrary gender classifiers," *CoRR*, vol. abs/1807.11936, 2018.

[12] V. Mirjalili, S. Raschka, and A. Ross, "FlowSAN: Privacy-Enhancing Semi-Adversarial Networks to Confound Arbitrary Face-Based Gender Classifiers," *IEEE Access*, vol. 7, 2019.

[13] S. Chhabra, R. Singh, M. Vatsa, and G. Gupta, "Anonymizing k-facial attributes via adversarial perturbations," in *International Joint Conferences on Artificial Intelligence (IJCAI)*, 2018.

[14] L. Sweeney, "K-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 5, pp. 557–570, 2002.

[15] E. M. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying face images," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 2, pp. 232–243, 2005.

[16] K. Nandakumar and A. K. Jain, "Biometric template protection: Bridging the performance gap between theory and practice," *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 88–100, 2015.

[17] V. M. Patel, N. K. Ratha, and R. Chellappa, "Cancelable biometrics: A review," *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 54–65, 2015.

[18] C. Rathgeb and A. Uhl, "A survey on biometric cryptosystems and cancelable biometrics," *EURASIP Journal on Information Security*, vol. 2011, no. 1, 2011.

[19] P. Rot, P. Peer, and V. Štruc, "PrivacyProber: Assessment and Detection of Soft–Biometric Privacy–Enhancing Techniques," *under review*, 2021.

[20] B. Bortolato, M. Ivanovska, P. Rot, J. Križaj, P. Terhörst, N. Damer, P. Peer, and V. Štruc, "Learning privacy-enhancing face representations through feature disentanglement," in *Automatic Face and Gesture Recognition (FG)*, 2020.

[21] P. Terhörst, M. Huber, N. Damer, P. Rot, F. Kirchbuchner, V. Struc, and A. Kuijper, "Privacy evaluation protocols for the evaluation of soft-biometric privacy-enhancing technologies," in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2020.

[22] P. Terhörst, K. Riehl, N. Damer, P. Rot, B. Bortolato, F. Kirchbuchner, V. Štruc, and A. Kuijper, "Pe-miu: A training-free privacy-enhancing face recognition approach based on minimum information units," *IEEE Access*, vol. 8, pp. 93 635–93 647, 2020.

[23] S. L. Garfinkel, "De-identification of personal information," *NIST-Technology Internal Report*, vol. 8053, pp. 1–46, 2015.

[24] N. N. G. d. Andrade, A. Martin, and S. Monteleone, "All the better to see you with, my dear": Facial recognition and privacy in online social networks," *IEEE Security Privacy*, vol. 11, no. 3, pp. 21–28, 2013.

[25] J. R. Padilla-López, A. A. Chaaraoui, and F. Flórez-Revuelta, "Visual privacy protection methods: A survey," *Expert Systems with Applications*, vol. 42, no. 9, pp. 4177 – 4195, 2015.

[26] T. Winkler and B. Rinner, "Privacy and security in video surveillance," in *Intelligent Multimedia Surveillance*, 2013, pp. 37–66.

[27] ——, "Security and Privacy Protection in Visual Sensor Networks: A Survey," *ACM Computing Surveys*, vol. 47, no. 1, 2014.

[28] T. D. Räty, "Survey on contemporary remote surveillance systems for public safety," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 5, pp. 493–515, 2010.

[29] S. Sah, A. Shringi, R. Ptucha, A. M. Burry, and R. P. Loce, "Video redaction: a survey and comparison of enabling technologies," *Journal of Electronic Imaging*, vol. 26, no. 5, 2017.

[30] R. Gross, L. Sweeney, F. de la Torre, and S. Baker, "Semi-supervised learning of multi-factor models for face de-identification," in *Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8.

[31] L. Meng, Z. Sun, A. Ariyaeeinia, and K. L. Bennett, "Retaining expressions on de-identified faces," in *International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2014, pp. 1252–1257.

[32] K. Brkič, I. Sikirič, T. Hrkač, and Z. Kalafatič, "I know that person: Generative full body and face de-identification of people in images," in *Computer Vision and Pattern Recognition (CVPR) Workshops*, 2017, pp. 1319–1328.

[33] B. Meden, Z. Emersic, V. Struc, and P. Peer, "κ-same-net: Neural-network-based face deidentification," in *International Work Conference on Bioinspired Intelligence (IWOBI)*, 2017, pp. 1–7.

[34] B. Yang, L. Rajbhandari, C. Busch, and X. Zhou, "Privacy implications of identity references in biometrics databases," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, 2012, pp. 25–30.

[35] T. Orekondy, M. Fritz, and B. Schiele, "Connecting pixels to privacy and utility: Automatic redaction of private information in images," in *Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 8466–8475.

[36] L. Yuan and T. Ebrahimi, "Image privacy protection with secure jpeg transmorphing," *IET Signal Processing*, vol. 11, pp. 1031–1038, 2017.

[37] F. Alegre, G. Soldi, and N. Evans, "Evasion and obfuscation in automatic speaker verification," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 749–753.

[38] S. Yoon, J. Feng, and A. K. Jain, "Altered fingerprints: Analysis and detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 3, pp. 451–464, 2012.

[39] S. Thavalengal, R. Vranceanu, R. G. Condorovici, and P. Corcoran, "Iris pattern obfuscation in digital images," in *International Joint Conference on Biometrics (IJCB)*, 2014, pp. 1–8.

[40] T. I. Dhamecha, A. Nigam, R. Singh, and M. Vatsa, "Disguise detection and face recognition in visible and thermal spectrums," in *International Conference on Biometrics (ICB)*, 2013, pp. 1–8.

[41] J.-G. Wang, A. Suwandy, and W.-Y. Yau, "Face obscuration in a video sequence by integrating kernel-based mean-shift and active contour," in *International Conference on Control, Automation, Robotics and Vision (ICARCV)*, 2008, pp. 2314–2318.

[42] R. Cucchiara, A. Prati, and R. Vezzani, "A system for automatic face obscuration for privacy purposes," *Pattern Recognition Letters*, vol. 27, no. 15, pp. 1809–1815, 2006.

[43] H. Hao, D. Güera, J. Horváth, A. R. Reibman, and E. J. Delp, "Robustness analysis of face obscuration," in *Automatic Face and Gesture Recognition (FG)*, 2020, pp. 176–183.

[44] V. Mirjalili and A. Ross, "Soft biometric privacy: Retaining biometric utility of face images while perturbing gender," in *International joint conference on biometrics (IJCB)*, 2017, pp. 564–573.

[45] A. Othman and A. Ross, "Privacy of facial soft biometrics: Suppressing gender but retaining identity," in *European Conference on Computer Vision Workshops (ECCVW)*, 2014, pp. 682–696.

[46] A. Morales, J. Fierrez, R. Vera-Rodriguez, and R. Tolosana, "Sensitivenets: Learning agnostic representations with application to face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 6, pp. 2158–2164, 2020.

[47] T. Sim and L. Zhang, "Controllable face privacy," in *Automatic Face and Gesture Recognition (FG)*, 2015, pp. 1–8.

[48] R. Clarke, "Internet privacy concerns confirm the case for intervention," *Communications of the ACM*, vol. 42, no. 2, pp. 60–67, 1999.

[49] ——, "What's' privacy'," in *Australian law reform commission workshop (ALRCW)*, vol. 28, 2006.

[50] R. L. Finn, D. Wright, and M. Friedewald, "Seven types of privacy," in *European data protection: coming of age*, 2013, pp. 3–32.

[51] A. Ross and A. Othman, "Visual cryptography for biometric privacy," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 1, pp. 70–81, 2011.

[52] A. Ross and A. A. Othman, "Visual cryptography for face privacy," in *Biometric Technology for Human Identification VII*, vol. 7667, 2010, p. 76670B.

[53] M. Kumar and N. Kumar, "Cancelable biometrics: a comprehensive survey," *Artificial Intelligence Review*, pp. 1–44, 2019.

[54] Y. Zhang, Y. Lu, H. Nagahara, and R. Taniguchi, "Anonymous camera for privacy protection," in *International Conference on Pattern Recognition (ICPR)*, 2014, pp. 4170–4175.

[55] D. J. Solove, "A taxonomy of privacy," *University of Pennsylvania Law Review*, vol. 154, p. 477, 2006.

[56] J. Breebaart, B. Yang, I. Buhan-Dulman, and C. Busch, "Biometric template protection," *Datenschutz und Datensicherheit-DuD*, vol. 33, no. 5, pp. 299–304, 2009.

[57] J. E. Cohen, "Examined lives: Informational privacy and the subject as object," *Stan. L. Rev.*, vol. 52, p. 1373, 1999.

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

[59] P. Dhar, A. Bansal, C. D. Castillo, J. Gleason, P. J. Phillips, and R. Chellappa, "How are attributes expressed in face dcnns?" in *Automatic Face and Gesture Recognition (FG)*, 2020, pp. 85–92.

[60] P. Terhörst, D. Fährmann, N. Damer, F. Kirchbuchner, and A. Kuijper, "Beyond identity: What information is stored in biometric face templates?" in *International Joint Conference on Biometrics (IJCB)*, 2020, pp. 1–10.

[61] B. Meden, Ž. Emeršič, V. Štruc, and P. Peer, "k-same-net: k-anonymity with generative deep neural networks for face deidentification," *Entropy*, vol. 20, no. 1, p. 60, 2018.

[62] O. Sarwar, A. Cavallaro, and B. Rinner, "Temporally smooth privacy-protected airborne videos," in *International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 6728–6733.

[63] M. Bonetto, P. Korshunov, G. Ramponi, and T. Ebrahimi, "Privacy in mini-drone based video surveillance," in *Automatic Face and Gesture Recognition (FG)*, 2015, pp. 1–6.

[64] R. Gross, L. Sweeney, F. de la Torre, and S. Baker, "Model-based face de-identification," in *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2006, pp. 161–161.

[65] L. Meng and A. Shenoy, "Retaining expression on de-identified faces," in *International Conference on Speech and Computer (SPECOM)*, 2017, pp. 651–661.

[66] P. Terhörst, N. Damer, F. Kirchbuchner, and A. Kuijper, "Suppressing gender and age in face templates using incremental variable elimination," in *International Conference on Biometrics (ICB)*, 2019, pp. 4–7.

[67] S. Çiftçi, A. O. Akyüz, and T. Ebrahimi, "A reliable and reversible image privacy protection based on false colors," *IEEE Transactions on Multimedia*, vol. 20, no. 1, pp. 68–81, 2018.

[68] O. Gafni, L. Wolf, and Y. Taigman, "Live face de-identification in video," in *International Conference on Computer Vision (ICCV)*, 2019, pp. 9378–9387.

[69] V. Mirjalili, S. Raschka, and A. Ross, "PrivacyNet: Semi-adversarial Networks for Multi-attribute Face Privacy," *IEEE Transactions on Image Processing*, vol. 29, pp. 9400–9412, 2020.

[70] P. Terhörst, N. Damer, F. Kirchbuchner, and A. Kuijper, "Unsupervised privacy-enhancement of face representations using similarity-sensitive noise transformations," *Applied Intelligence*, pp. 1–18, 2019.

[71] Z. Ren, Y. J. Lee, and M. S. Ryoo, "Learning to anonymize faces for privacy preserving action detection," in *European Conference on Computer Vision (ECCV)*, 2018, pp. 620–636.

[72] E. Chatzikyriakidis, C. Papaioannidis, and I. Pitas, "Adversarial face de-identification," in *International Conference on Image Processing (ICIP)*, 2019, pp. 684–688.

[73] L. Fan, "Image pixelization with differential privacy," in *IFIP Annual Conference on Data and Applications Security and Privacy*, 2018, pp. 148–162.

[74] W. L. Croft, J.-R. Sack, and W. Shi, "Differentially private obfuscation of facial images," in *Machine Learning and Knowledge Extraction*, 2019, pp. 229–249.

[75] J. Prinosil, P. Kriz, K. Riha, M. K. Dutta, and A. Issac, "Facial image de-identification using active appearance model," in *International Conference on Emerging Trends in Computing and Communication Technologies (ICETCCT)*, 2017, pp. 1–5.

[76] Z. Sun, L. Meng, and A. Ariyaeeinia, "Distinguishable de-identified faces," in *Automatic Face and Gesture Recognition (FG)*, 2015, pp. 1–6.

[77] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "L-diversity: Privacy Beyond K-Anonymity," *ACM Transactions on Knowledge Discovery from Data*, vol. 1, no. 1, 2007.

[78] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," in *International Conference on Data Engineering (ICDE)*, 2007, pp. 106–115.

[79] A. Campan, T. M. Truta, and N. Cooper, "P-Sensitive K-Anonymity with Generalization Constraints," *Transactions on Data Privacy*, vol. 3, no. 2, pp. 65–89, 2010.

[80] Mrityunjay and P. J. Narayanan, "The de-identification camera," in *National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, 2011, pp. 192–195.

[81] P. Terhörst, "Mitigating soft-biometric driven bias and privacy concerns in face recognition systems," Ph.D. dissertation, Technische Universität Darmtadt, 2021.

[82] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, pp. 211–407, 2014.

[83] F. D. McSherry, "Privacy integrated queries: an extensible platform for privacy-preserving data analysis," in *ACM SIGMOD International Conference on Management of data*, 2009, pp. 19–30.

[84] C. Dwork, "Differential privacy: A survey of results," in *Theory and Applications of Models of Computation*, 2008, pp. 1–19.

[85] I. Dinur and K. Nissim, "Revealing information while preserving privacy," in *ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (PODS)*, 2003, pp. 202–210.

[86] H. Zhang, H. Zhou, W. Jiao, J. Shi, Q. Zang, J. Sun, and J. Zhang, "Biological features de-identification in iris images," in *International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN)*, 2018, pp. 67–71.

[87] S. Li and A. C. Kot, "Fingerprint combination for privacy protection," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 2, pp. 350–360, 2012.

[88] L. Lugini, E. Marasco, B. Cukic, and J. Dawson, "Removing gender signature from fingerprints," in *International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2014, pp. 1283–1287.

[89] D. Lee and K. N. Plataniotis, "A novel eye region based privacy protection scheme," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 1845–1848.

[90] J. Chaudhari, S. C. Sen-ching, and M. V. Venkatesh, "Privacy protection for life-log system," in *Workshop on Signal Processing Applications for Public Security and Forensics (SAFE)*, 2007, pp. 1–5.

[91] Q. Jin, A. R. Toth, T. Schultz, and A. W. Black, "Speaker de-identification via voice transformation," in *Workshop on Automatic Speech Recognition & Understanding (ASRU)*, 2009, pp. 529–533.

[92] T. Justin, V. Štruc, S. Dobrišek, B. Vesnicer, I. Ipšić, and F. Mihelič, "Speaker de-identification using diphone recognition and speech synthesis," in *Automatic Face and Gesture Recognition (FG)*, vol. 4, 2015, pp. 1–7.

[93] D. Marčetić, S. Ribarić, V. Štruc, and N. Pavešić, "An experimental tattoo de-identification system for privacy protection in still images," in *International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2014, pp. 1288–1293.

[94] P. Agrawal and P. Narayanan, "Person de-identification in videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 3, pp. 299–310, 2011.

[95] M. Faundez-Zanuy and J. Mekyska, "Privacy of online handwriting biometrics related to biomedical analysis," *IET in User-Centric Privacy and Security in Biometrics*, 2017.

[96] K. Chinomi, N. Nitta, Y. Ito, and N. Babaguchi, "Prisurv: Privacy protected video surveillance system using adaptive visual abstraction," in *Advances in Multimedia Modeling (MMM)*, Berlin, Heidelberg, 2008, pp. 144–154.

[97] N. Babaguchi, T. Koshimizu, I. Umata, and T. Toriyama, "Psychological study for designing privacy protected video surveillance system: Prisurv," in *Protecting Privacy in Video Surveillance*, 2009, pp. 147–164.

[98] J. Schiff, M. Meingast, D. K. Mulligan, S. Sastry, and K. Goldberg, "Respectful cameras: detecting visual markers in real-time to address privacy concerns," in *International Conference on Intelligent Robots and Systems (IROS)*, 2007, pp. 971–978.

[99] D. Chen, Y. Chang, R. Yan, and J. Yang, "Protecting personal identification in video," in *Protecting Privacy in Video Surveillance*, 2009, pp. 115–128.

[100] J. Wang, B. Amos, A. Das, P. Pillai, N. Sadeh, and M. Satyanarayanan, "A Scalable and Privacy-Aware IoT Service for Live Video Analytics," in *ACM on Multimedia Systems Conference (MMSys)*, ser. MMSys'17, New York, NY, USA, 2017, pp. 38–49.

[101] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "Openface: an open source facial behavior analysis toolkit," in *Winter Conference on Applications of Computer Vision (WACV)*, 2016, pp. 1–10.

[102] A. Das, M. Degeling, X. Wang, J. Wang, N. Sadeh, and M. Satyanarayanan, "Assisting users in a world full of cameras: A privacy-aware infrastructure for computer vision applications," in *Computer Vision and Pattern Recognition (CVPR) Workshops*, 2017, pp. 1387–1396.

[103] L. Yuan and T. Ebrahimi, "Image transmorphing with jpeg," in *IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 3956–3960.

[104] A. Badii, A. Al-Obaidi, M. Einig, and A. Ducournau, "Holistic privacy impact assessment framework for video privacy filtering technologies," *Signal & Image Processing*, vol. 4, no. 6, p. 13, 2013.

[105] P. Korshunov, A. Melle, J.-L. Dugelay, and T. Ebrahimi, "A framework for objective evaluation of privacy filters in video surveillance," *Proceedings of SPIE Volume 8856*, 2013.

[106] Á. Erdélyi, T. Winkler, and B. Rinner, "Privacy protection vs. utility in visual data," *Multimedia Tools and Applications*, vol. 77, no. 2, pp. 2285–2312, 2018.

[107] A. Erdélyi, T. Barát, P. Valet, T. Winkler, and B. Rinner, "Adaptive cartooning for privacy protection in camera networks," in *International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2014, pp. 44–49.

[108] H. Fradi, V. Eiselein, I. Keller, J. Dugelay, and T. Sikora, "Crowd context-dependent privacy protection filters," in *International Conference on Digital Signal Processing (DSP)*, 2013, pp. 1–6.

[109] N. Ruchaud and J. Dugelay, "Aseppi: Robust privacy protection against de-anonymization attacks," in *Computer Vision and Pattern Recognition (CVPR) Workshops*, 2017, pp. 1352–1359.

[110] K. Kobayashi, K. Iwamura, K. Kaneda, and I. Echizen, "Surveillance camera system to achieve privacy protection and crime prevention," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, 2014, pp. 463–466.

[111] P. Korshunov and T. Ebrahimi, "Using face morphing to protect privacy," in *International Conference on Advanced Video and Signal Based Surveillance (ICAVSBS)*, 2013, pp. 208–213.

[112] Q. A. Zhao and J. T. Stasko, "Evaluating image filtering based techniques in media space applications," in *ACM conference on Computer Supported Cooperative Work CSCW*, Seattle, Washington, United States, 1998, pp. 11–18.

[113] M. Boyle, C. Edwards, and S. Greenberg, "The effects of filtered video on awareness and privacy," in *ACM conference on Computer Supported Cooperative Work (CSCW)*, 2000, pp. 1–10.

[114] G. Letournel, A. Bugeau, V. T. Ta, and J. P. Domenger, "Face de-identification with expressions preservation," in *International Conference on Image Processing (ICIP)*, 2015, pp. 4366–4370.

[115] A. Albiol, D. Monzo, A. Martin, J. Sastre, and A. Albiol, "Face recognition using hog–ebgm," *Pattern Recognition Letters*, vol. 29, no. 10, pp. 1537–1543, 2008.

[116] M. Dahmane and J. Meunier, "Emotion recognition using dynamic grid-based hog features," in *Automatic Face and Gesture Recognition (FG)*, 2011, pp. 884–888.

[117] Y. Pang, Y. Yuan, X. Li, and J. Pan, "Efficient hog human detection," *Signal Processing*, vol. 91, no. 4, pp. 773–781, 2011.

[118] O. Déniz, G. Bueno, J. Salido, and F. De la Torre, "Face recognition using histograms of oriented gradients," *Pattern Recognition Letters*, vol. 32, no. 12, pp. 1598–1603, 2011.

[119] N.-S. Vu, "Exploring patterns of gradient orientations and magnitudes for face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 2, pp. 295–304, 2012.

[120] O. Sarwar, B. Rinner, and A. Cavallaro, "A privacy-preserving filter for oblique face images based on adaptive hopping gaussian mixtures," *IEEE Access*, vol. 7, pp. 142 623–142 639, 2019.

[121] ——, "Design space exploration for adaptive privacy protection in airborne images," in *International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2016, pp. 159–165.

[122] P. Agrawal and P. J. Narayanan, "Person de-identification in videos," in *Asian Conference on Computer Vision (ACCV)*, 2010.

[123] P. Korshunov, C. Araimo, F. D. Simone, C. Velardo, J. L. Dugelay, and T. Ebrahimi, "Subjective study of privacy filters in video surveillance," in *International Workshop on Multimedia Signal Processing (MMSP)*, 2012, pp. 378–382.

[124] L. Ye, B. Li, N. Mohammed, Y. Wang, and J. Liang, "Privacy-preserving age estimation for content rating," in *International Workshop on Multimedia Signal Processing (MMSP)*, 2018, pp. 1–6.

[125] N. Zhang, M. Paluri, Y. Taigman, R. Fergus, and L. Bourdev, "Beyond frontal faces: Improving person recognition using multiple cues," in *Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 4804–4813.

[126] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," University of Massachusetts, Amherst, Tech. Rep. 07-49, 2007.

[127] S. Blunsden and R. Fisher, "The behave video dataset: ground truthed video for multi-person," *Annals of the BMVA*, vol. 2010, no. 4, pp. 1–11, 2010.

[128] J. Ferryman and A. Shahrokni, "Pets2009: Dataset and challenge," in *International Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, 2009, pp. 1–6.

[129] P. Korshunov and T. Ebrahimi, "Pevid: privacy evaluation video dataset," *Proceedings of SPIE Volume 8856*, vol. 8856, p. 9, 2013.

[130] H.-J. Hsu and K.-T. Chen, "Face recognition on drones: Issues and limitations," in *ACM Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use*, 2015, pp. 39–44.

[131] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2170–2179, 2014.

[132] F. Dufaux and T. Ebrahimi, "Video surveillance using jpeg 2000," in *Applications of Digital Image Processing XXVII*, vol. 5558. International Society for Optics and Photonics, 2004, pp. 268–275.

[133] I. Martínez-Ponte, X. Desurmont, J. Meessen, and J. françois Delaigle, "Robust human face hiding ensuring privacy," in *International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS*, 2005.

[134] R. Cucchiara, A. Prati, and R. Vezzani, "Advanced video surveillance with pan tilt zoom cameras," in *International Workshop on Visual Surveillance (IWVS)*, 2006, pp. 334–352.

[135] T. E. Boult, "Pico: Privacy through invertible cryptographic obscuration," in *Computer Vision for Interactive and Intelligent Environment (CVIIE)*, 2005, pp. 27–38.

[136] F. Dufaux and T. Ebrahimi, "H.264/avc video scrambling for privacy protection," in *International Conference on Image Processing (ICIP)*, 2008, pp. 1688–1691.

[137] C. Montgomery *et al.*, "Xiph. org video test media (derf's collection), the xiph open source community, 1994," *Online, https://media. xiph. org/video/derf*.

[138] M. Xuan and J. Jiang, "Video security algorithm aiming at the need of privacy protection," in *International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, vol. 5, 2009, pp. 473–477.

[139] L. Tong, F. Dai, Y. Zhang, and J. Li, "Prediction restricted h.264/avc video scrambling for privacy protection," *Electronics Letters*, vol. 46, no. 1, pp. 47–49, 2010.

[140] J. Cichowski and A. Czyzewski, "Reversible video stream anonymization for video surveillance systems based on pixels relocation and watermarking," in *International Conference on Computer Vision Workshops (ICCVW)*, 2011, pp. 1971–1977.

[141] S. M. M. Rahman, M. A. Hossain, H. Mouftah, A. El Saddik, and E. Okamoto, "Chaos-cryptography based privacy preservation technique for video surveillance," *Multimedia Systems*, vol. 18, no. 2, pp. 145–155, 2012.

[142] P. Korshunov and T. Ebrahimi, "Using warping for privacy protection in video surveillance," in *International Conference on Digital Signal Processing (DSP)*, 2013, pp. 1–6.

[143] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[144] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi, "The FERET evaluation methodology for face recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, 2000.

[145] Y. Wang, M. O'Neill, and F. Kurugollu, "Privacy region protection for h.264/avc by encrypting the intra prediction modes without drift error in i frames," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 2964–2968.

[146] P. Su, W. Chen, S. Shiau, C. Wu, and A. Y. S. Su, "A privacy protection scheme in h.264/avc by data hiding," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2013, pp. 1–7.

[147] B. Bhattarai, A. Mignon, F. Jurie, and T. Furon, "Puzzling face verification algorithms for privacy protection," in *International Workshop on Information Forensics and Security (WIFS)*, 2014, pp. 66–71.

[148] A. Melle and J. Dugelay, "Scrambling faces for privacy protection using background self-similarities," in *International Conference on Image Processing (ICIP)*, 2014, pp. 6046–6050.

[149] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Workshop on Applications of Computer Vision (WACV)*, 1994, pp. 138–142.

[150] N. Ruchaud and J.-L. Dugelay, "Privacy protection filter using stegoscrambling in video surveillance," in *MediaEval*, 2015.

[151] V. E. Alonso, R. A. Enríquez-Caldera, and L. E. Sucar, "Foveation: an alternative method to simultaneously preserve privacy and information in face images," *Journal of Electronic Imaging*, vol. 26, no. 2, 2017.

[152] R. Jiang, A. Bouridane, D. Crookes, M. E. Celebi, and H. Wei, "Privacy-protected facial biometric verification using fuzzy forest learning," *IEEE Transactions on Fuzzy Systems*, vol. 24, no. 4, pp. 779–790, 2016.

[153] R. Jiang, S. Al-Maadeed, A. Bouridane, D. Crookes, and M. E. Celebi, "Face recognition in the scrambled domain via salience-aware ensembles of many kernels," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1807–1817, 2016.

[154] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression (pie) database," in *Automatic Face and Gesture Recognition (FG)*, 2002, pp. 53–58.

[155] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *International Conference on Computer Vision (ICCV)*, 2009, pp. 365–372.

[156] P. Chriskos, J. Munro, V. Mygdalis, and I. Pitas, "Face detection hindering," in *Global Conference on Signal and Information Processing (GlobalSIP)*, 2017, pp. 403–407.

[157] K. Messer, J. Kittler, M. Sadeghi, S. Marcel, C. Marcel, S. Bengio, F. Cardinaux, C. Sanderson, J. Czyz, L. Vandendorpe, S. Srisuk, M. Petrou, W. Kurutach, A. Kadyrov, R. Paredes, B. Kepenekci, F. B. Tek, G. B. Akar, F. Deravi, and N. Mavity, "Face Verification Competition on the XM2VTS Database," in *Audio- and Video-based Biometric Person Authentication (AVBPA)*, 2003, pp. 964–974.

[158] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference (BMVC)*, 2015.

[159] S. Dadkhah, M. Koeppen, S. Sadeghi, and K. Yoshida, "Bad ai: Investigating the effect of half-toning techniques on unwanted face detection systems," in *International Conference on New Technologies, Mobility and Security (NTMS)*, 2018, pp. 1–5.

[160] K. Hoshino, K. Iwamura, and K. Kaneda, "Improvement of privacy protection surveillance camera system and its applications," in *International Conference on Mobile and Secure Services (MobiSecServ)*, 2018, pp. 1–5.

[161] S. Liu, L. Kong, and H. Wang, "Face detection and encryption for privacy preserving in surveillance video," in *Pattern Recognition and Computer Vision (PRCV)*, 2018, pp. 162–172.

[162] L. Lealtaixé, A. Milan, I. Reid, S. Roth, and K. Schindler, "Mot challenge 2015: Towards a benchmark for multi-target tracking," *arXiv:1504.01942*, 2015.

[163] L. Fan, "Practical image obfuscation with provable privacy," in *IEEE International Conference on Multimedia and Expo (ICME)*, 2019, pp. 784–789.

[164] C. Dwork, "Differential privacy," in *Automata, Languages and Programming*, 2006, pp. 1–12.

[165] S. Milborrow, J. Morkel, and F. Nicolls, "The MUCT Landmarked Face Database," *Pattern Recognition Association of South Africa*, 2010, http://www.milbo.org/muct.

[166] Z. Wu, Z. Wang, Z. Wang, and H. Jin, "Towards privacy-preserving visual recognition via adversarial training: A pilot study," in *European Conference on Computer Vision (ECCV)*, 2018.

[167] K. Yun, J. Honorio, D. Chattopadhyay, T. L. Berg, and D. Samaras, "Two-person interaction detection using body-pose features and multiple instance learning," in *Computer Vision and Pattern Recognition (CVPR) Workshops*, 2012, pp. 28–35.

[168] K. Soomro, A. R. Zamir, and M. Shah, "Ucf101: A dataset of 101 human actions classes from videos in the wild," *arXiv:1212.0402*, 2012.

[169] T. Orekondy, B. Schiele, and M. Fritz, "Towards a visual privacy advisor: Understanding and predicting privacy risks in images," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 3706–3715, 2017.

[170] C. Huang, P. Kairouz, X. Chen, L. Sankar, and R. Rajagopal, "Generative Adversarial Privacy," *arXiv:1807.05306*, 2018.

[171] MPLab. The MPLab GENKI Database. [Online]. Available: http://mplab.ucsd.edu

[172] V. Mirjalili, S. Raschka, A. M. Namboodiri, and A. Ross, "Semi-adversarial networks: Convolutional autoencoders for imparting privacy to face images," in *International Conference on Biometrics (ICB)*, Gold Coast, Australia, 2018, pp. 82–89.

[173] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *International Conference on Computer Vision (ICCV)*, 2015.

[174] A. M. Martinez, "The AR face database," CVC Technical Report #24, 1998.

[175] K. Ricanek and T. Tesafaye, "Morph: a longitudinal image database of normal adult age-progression," in *Automatic Face and Gesture Recognition (FG)*, 2006, pp. 341–345.

[176] O. Langner, R. Dotsch, G. Bijlstra, D. Wigboldus, S. Hawk, and A. van Knippenberg, "Presentation and validation of the radboud faces database," *Cognition&Emotion*, vol. 24, no. 8, pp. 1377–1388, 2010.

[177] S. Y. Zhang, Zhifei and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *Computer Vision and Pattern Recognition (CVPR)*, 2017.

[178] M. Lecuyer, V. Atlidakis, R. Geambasu, D. Hsu, and S. Jana, "Certified robustness to adversarial examples with differential privacy," in *IEEE Symposium on Security and Privacy (SP)*, 2019, pp. 656–672.

[179] C. Guo, M. Rana, M. Cisse, and L. Van Der Maaten, "Countering adversarial images using input transformations," *arXiv:1711.00117*, 2017.

[180] H. Hukkelås, R. Mester, and F. Lindseth, "Deepprivacy: A generative adversarial network for face anonymization," *arXiv:1909.04538*, 2019.

[181] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar, "Face Swapping: Automatically Replacing Faces in Photographs," *ACM Transactions on Graphics*, vol. 27, no. 3, pp. 39:1–39:8, 2008.

[182] S. Mosaddegh, L. Simon, and F. Jurie, *Photorealistic Face De-Identification by Aggregating Donors' Face Components*, 2015, pp. 159–174.

[183] W. Xu, S. S. Cheung, and N. Soares, "Affect-preserving privacy protection of video," in *International Conference on Image Processing (ICIP)*, 2015, pp. 158–162.

[184] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[185] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

[186] B. Driessen and M. Dürmuth, *Achieving Anonymity against Major Face Recognition Algorithms*, 2013, pp. 18–33.

[187] L. Du, M. Yi, E. Blasch, and H. Ling, "GARP-Face: Balancing privacy protection and utility preservation in face de-identification," in *International Joint Conference on Biometrics (JCB)*, 2014, pp. 1–8.

[188] H. Chi and Y. H. Hu, "Facial image de-identification using identiy subspace decomposition," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 524–528.

[189] H. Mercier, P. Dalle, and I.-U. P. Sabatier, "Face analysis: identity vs. expressions," in *International Society for Gesture Studies Symposium*, 2005.

[190] X. Wang, C. Xiong, Q. Pei, and Y. Qu, "Expression preserved face privacy protection based on multi-mode discriminant analysis," *CMC: computers, materials & continua*, vol. 57, no. 1, pp. 107–121, 2018.

[191] B. Samaržija and S. Ribaric, "An approach to the de-identification of faces in different poses," in *International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2014, pp. 1246–1251.

[192] L. Meng, Z. Sun, and O. T. Collado, "Efficient approach to de-identifying faces in videos," *IET Signal Processing*, vol. 11, no. 9, pp. 1039–1045, 2017.

[193] T. Baltrusaitis, P. Robinson, and L.-P. Morency, "Constrained local neural fields for robust facial landmark detection in the wild," in *International Conference on Computer Vision Workshops (ICCVW)*, 2013, pp. 354–361.

[194] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image Vision Comput.*, vol. 28, no. 5, p. 807–813, 2010.

[195] C. E. Thomaz and G. A. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image and Vision Computing*, vol. 28, no. 6, pp. 902 – 913, 2010.

[196] Markus Weber. (1999) Frontal Faces Database. [Online]. Available: http://www.vision.caltech.edu/html-files/archive.html

[197] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *International Conference on Multimedia and Expo (ICME)*, 2005, pp. 5 pp.–.

[198] M. Lyons, M. Kamachi, and J. Gyoba. The Japanese Female Facial Expression (JAFFE) Database. [Online]. Available: https://doi.org/10.5281/zenodo.3451524

[199] M. M. Nordstrøm, M. Larsen, J. Sierakowski, and M. B. Stegmann, "The IMM Face Database - An Annotated Dataset of 240 Face Images," Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Tech. Rep., 2004. [Online]. Available: http://www2.compute.dtu.dk/pubdb/pubs/3160-full.html

[200] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2930–2940, 2013.

[201] M. Ohana, O. Dunkelman, S. Gibson, and M. Osadchy, "Honeyfaces: Increasing the security and privacy of authentication using synthetic facial images," *arXiv:1611.03811*, 2016.

[202] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3d facial expression database for facial behavior research," in *Automatic Face and Gesture Recognition (FG)*, 2006, p. 211–216.

[203] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, and I. Matthews, "Painful data: The unbc-mcmaster shoulder pain expression archive database," in *Automatic Face and Gesture Recognition (FG)*, 2011, pp. 57–64.

[204] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Computer Vision and Pattern Recognition (CVPR) Workshops*, 2010, pp. 94–101.

[205] H. Chi and Y. H. Hu, "Face de-identification using facial identity preserving features," in *Global Conference on Signal and Information Processing (GlobalSIP)*, 2015, pp. 586–590.

[206] M. A. Rafique, M. S. Azam, M. Jeon, and S. Lee, "Face-deidentification in images using restricted boltzmann machines," in *International Conference for Internet Technology and Secured Transactions (ICITST)*, 2016, pp. 69–73.

[207] K. Brkić, T. Hrkać, Z. Kalafatić, and I. Sikirić, "Face, hairstyle and clothing colour de-identification in video sequences," *IET Signal Processing*, 2017.

[208] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell, "Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition," in *Biometrics Workshop, Computer Vision and Pattern Recognition (CVPR) Workshops*, 2011, pp. 81–88.

[209] P. Weinzaepfel, X. Martin, and C. Schmid, "Human action localization with sparse spatial supervision," *arXiv:1605.05197*, 2016.

[210] H. Jhuang, J. Gall, S. Zuffi, C. Schmid, and M. J. Black, "Towards understanding action recognition," in *International Conference on Computer Vision (ICCV)*, 2013, pp. 3192–3199.

[211] S. Guo, S. Feng, Y. Li, S. An, and H. Dong, "Integrating diversity into neural-network-based face deidentification," in *Chinese Control Conference (CCC)*, 2018, pp. 9356–9361.

[212] J. Chen, J. Konrad, and P. Ishwar, "Vgan-based image representation learning for privacy-preserving facial expression recognition," in *Computer Vision and Pattern Recognition (CVPR) Workshops*, 2018, pp. 1570–1579.

[213] D. Aneja, A. Colburn, G. Faigin, L. Shapiro, and B. Mones, "Modeling stylized character expressions via deep learning," in *Asian Conference on Computer Vision (ACCV)*, 2016, pp. 136–153.

[214] N. Aifanti, C. Papachristou, and A. Delopoulos, "The mug facial expression database," in *International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, 2010, pp. 1–4.

[215] Q. Sun, L. Ma, S. Joon Oh, L. Van Gool, B. Schiele, and M. Fritz, "Natural and effective obfuscation by head inpainting," in *Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 5050–5059.

[216] Q. Sun, A. Tewari, W. Xu, M. Fritz, C. Theobalt, and B. Schiele, "A Hybrid Model for Identity Obfuscation by Face Replacement," in *European Conference on Computer Vision (ECCV)*, 2018.

[217] Y. Wu, F. Yang, and H. Ling, "Privacy-Protective-GAN for Face De-identification," *arXiv:1806.08906*, 2018.

[218] Y.-L. Pan, M.-J. Haung, K.-T. Ding, J.-L. Wu, and J.-S. Jang, "k-Same-Siamese-GAN: k-Same Algorithm with Generative Adversarial Network for Facial Image De-identification with Hyperparameter Tuning and Mixed Precision Training," *arXiv:1904.00816*, 2019.

[219] Y. Li and S. Lyu, "De-identification without losing faces," *arXiv:1902.04202*, 2019.

[220] T. Li and L. Lin, "Anonymousnet: Natural face de-identification with measurable privacy," *arXiv:1904.12620*, 2019.

[221] H. Hao, D. Güera, A. R. Reibman, and E. J. Delp, "A utility-preserving gan for face obscuration," *arXiv:1906.11979*, 2019.

[222] H. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets," in *International Conference on Image Processing (ICIP)*, 2014, pp. 343–347.

[223] S. Yang, P. Luo, C. C. Loy, and X. Tang, "Wider face: A face detection benchmark," in *Computer Vision and Pattern Recognition (CVPR)*, 2016.

[224] D. Lundqvist, A. Flykt, and A. Öhman, "The karolinska directed emotional faces (KDEF)," *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet*, vol. 91, no. 630, pp. 2–2, 1998.

[225] M. Maximov, I. Elezi, and L. Leal-Taixé, "CiaGAN: Conditional identity anonymization generative adversarial networks," in *Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 5447–5456.

[226] P. Voigtlaender, M. Krause, A. Osep, J. Luiten, B. B. G. Sekar, A. Geiger, and B. Leibe, "Mots: Multi-object tracking and segmentation," in *Computer Vision and Pattern Recognition (CVPR)*, 2019.

[227] D. Cho, J. H. Lee, and I. H. Suh, "CLEANIR: Controllable Attribute-Preserving Natural Identity Remover," *Applied Sciences*, vol. 10, no. 3, p. 1120, 2020.

[228] H. Proença, "The uu-net: Reversible face de-identification for visual surveillance video footage," *arXiv:2007.04316*, 2020.

[229] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian, "Mars: A video benchmark for large-scale person re-

identification," in *European Conference on Computer Vision (ECCV)*, 2016, pp. 868–884.

[230] S. Kumar, E. Yaghoubi, A. Das, B. Harish, and H. Proença, "The p-destre: A fully annotated dataset for pedestrian detection, tracking, re-identification and search from aerial devices," *arXiv:2004.02782*, 2020.

[231] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems (NIPS)*, 2014, pp. 2672–2680.

[232] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[233] S. Keronen, K. Cho, T. Raiko, A. Ilin, and K. Palomäki, "Gaussian-bernoulli restricted boltzmann machines and automatic feature extraction for noise robust missing data mask estimation," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 6729–6733.

[234] V. Blanz and T. Vetter, "Face Recognition Based on Fitting a 3D Morphable Model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, 2003.

[235] S. M. Moosavi-Dezfooli, A. Fawzi, and P. Frossard, "DeepFool: A Simple and Accurate Method to Fool Deep Neural Networks," in *Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2574–2582.

[236] C. Feutry, P. Piantanida, Y. Bengio, and P. Duhamel, "Learning anonymized representations with adversarial neural networks," *arXiv:1802.09386*, 2018.

[237] Z. Erkin, M. Franz, J. Guajardo, S. Katzenbeisser, I. Lagendijk, and T. Toft, "Privacy-preserving face recognition," in *Privacy Enhancing Technologies Symposium (PETS)*, 2009, pp. 235–253.

[238] A.-R. Sadeghi, T. Schneider, and I. Wehrenberg, "Efficient privacy-preserving face recognition," in *International Conference on Information Security and Cryptology (ICISC)*, 2009, pp. 229–244.

[239] Y. Rahulamathavan, R. C.-W. Phan, J. A. Chambers, and D. J. Parish, "Facial expression recognition in the encrypted domain based on local fisher discriminant analysis," *IEEE Transactions on Affective Computing (TAC)*, vol. 4, no. 1, pp. 83–92, 2013.

[240] J. R. Troncoso-Pastoriza, D. González-Jiménez, and F. Pérez-González, "Fully private noninteractive face verification," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 7, pp. 1101–1114, 2013.

[241] C. Xiang, C. Tang, Y. Cai, and Q. Xu, "Privacy-preserving face recognition with outsourced computation," *Soft Computing*, vol. 20, no. 9, pp. 3735–3744, 2016.

[242] Z. Li and T. H. Lai, "Deterministic fully homomorphic encryptions for privacy preserving cloud computing," 2012.

[243] Y. Ma, L. Wu, X. Gu, J. He, and Z. Yang, "A secure face-verification scheme based on homomorphic encryption and deep neural networks," *IEEE Access*, vol. 5, pp. 16 532–16 538, 2017.

[244] K. Rujirakul, C. So-In, and B. Arnonkijpanich, "PEM-PCA: A parallel expectation-maximization PCA face recognition architecture," *The Scientific World Journal*, vol. 2014, 2014.

[245] V. N. Boddeti, "Secure face matching using fully homomorphic encryption," in *International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2018, pp. 1–10.

[246] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, and A. K. Jain, "Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a," in *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1931–1939.

[247] C. Whitelam, E. Taborsky, A. Blanton, B. Maze, J. Adams, T. Miller, N. Kalka, A. K. Jain, J. A. Duncan, K. Allen *et al.*, "Iarpa janus benchmark-b face dataset," in *IEEE conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 90–98.

[248] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *arXiv:1411.7923*, 2014.

[249] P. Drozdowski, N. Buchmann, C. Rathgeb, M. Margraf, and C. Busch, "On the application of homomorphic encryption to face identification," in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2019, pp. 1–5.

[250] J. Kolberg, P. Drozdowski, M. Gomez-Barrero, C. Rathgeb, and C. Busch, "Efficiency analysis of post-quantum-secure face template protection schemes based on homomorphic encryption," in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2020, pp. 1–4.

[251] T. Yang, Y. Zhang, J. Sun, and X. Wang, "Privacy enhanced cloud-based facial recognition," *Neural Processing Letters*, pp. 1–9, 2021.

[252] C. Doersch, "Tutorial on variational autoencoders," *arXiv:1606.05908*, 2016.

[253] P. Terhörst, M. Huber, N. Damer, F. Kirchbuchner, and A. Kuijper, "Unsupervised enhancement of soft-biometric privacy with negative face recognition," *arXiv:2002.09181*, 2020.

[254] B. Pulido-Gaytan, A. Tchernykh, J. M. Cortés-Mendoza, M. Babenko, G. Radchenko, A. Avetisyan, and A. Y. Drozdov, "Privacy-preserving neural networks with homomorphic encryption: Challenges and opportunities," *Peer-to-Peer Networking and Applications*, vol. 14, no. 3, pp. 1666–1691, 2021.

[255] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*, 1999, pp. 223–238.

[256] I. Damgard, M. Geisler, and M. Kroigard, "A correction to'efficient and secure comparison for on-line auctions'," *International Journal of Applied Cryptography (IJACT)*, vol. 1, no. 4, pp. 323–324, 2009.

[257] A. C.-C. Yao, "How to generate and exchange secrets," in *Annual Symposium on Foundations of Computer Science*, 1986, pp. 162–167.

[258] J. Fan and F. Vercauteren, "Somewhat practical fully homomorphic encryption," *International Association for Cryptologic Research (IACR) Cryptology ePrint Archive*, vol. 2012, p. 144, 2012.

[259] C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Stanford University, 2009.

[260] J. H. Cheon, A. Kim, M. Kim, and Y. Song, "Homomorphic encryption for arithmetic of approximate numbers," in *International Conference on the Theory and Application of Cryptology and Information Security (ASIACRYPT)*, 2017, pp. 409–437.

[261] J. Hoffstein, J. Pipher, and J. H. Silverman, "Ntru: A ring-based public key cryptosystem," in *International Algorithmic Number Theory Symposium (ANTS)*, 1998, pp. 267–288.

[262] A. Acar, H. Aksu, A. S. Uluagac, and M. Conti, "A survey on homomorphic encryption schemes: Theory and implementation," *ACM Computing Surveys*, vol. 51, no. 4, pp. 1–35, 2018.

[263] P. Martins, L. Sousa, and A. Mariano, "A survey on fully homomorphic encryption: An engineering perspective," *ACM Computing Surveys*, vol. 50, no. 6, pp. 1–33, 2017.

[264] V. Fetzer, J. Müller-Quade, and T. Nilges, "A Formal Treatment of Privacy in Video Data," in *European Symposium on Research in Computer Security (ESORICS)*, 2016, pp. 406–424.

[265] S. J. Oh, R. Benenson, M. Fritz, and B. Schiele, "Faceless person recognition: Privacy implications in social media," in *European Conference on Computer Vision (ECCV)*, 2016, pp. 19–35.

[266] J. Henriksen-Bulmer and S. Jeary, "Re-identification attacks—a systematic literature review," *International Journal of Information Management*, vol. 36, no. 6, pp. 1184–1192, 2016.

[267] Z. Sun, L. Meng, A. Ariyaeeinia, X. Duan, and Z. Tan, "Privacy protection performance of de-identified face images with and without background," in *International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2016, pp. 1354–1359.

[268] F. Dufaux and T. Ebrahimi, "A framework for the validation of privacy protection solutions in video surveillance," in *International Conference on Multimedia and Expo (ICME)*, 2010, pp. 66–71.

[269] A. Badii, T. Ebrahimi, C. Fedorczak, P. Korshunov, T. Piatrik, V. Eiselein, and A. Al-Obaidi, "Overview of the mediaeval 2014 visual privacy task," *CEUR Workshop Proceedings*, vol. 1263, 2014.

[270] O. Sarwar, "Facial privacy protection in airborne recreational videography," Ph.D. dissertation, Queen Mary University of London, 2019.

[271] S. Zerr, S. Siersdorfer, J. Hare, and E. Demidova, "Privacy-aware image classification and search," in *Conference on Research and Development in Information Retrieval (SIGIR)*, 2012, pp. 35–44.

[272] E. Spyromitros-Xioufis, S. Papadopoulos, A. Popescu, and Y. Kompatsiaris, "Personalized privacy-aware image classification," in *International Conference on Multimedia Retrieval (ICMR)*, 2016, pp. 71–78.

[273] A. Frome, G. Cheung, A. Abdulkader, M. Zennaro, B. Wu, A. Bissacco, H. Adam, H. Neven, and L. Vincent, "Large-scale privacy protection in google street view," in *International Conference on Computer Vision (ICCV)*, 2009, pp. 2373–2380.

[274] D. Gurari, Q. Li, C. Lin, Y. Zhao, A. Guo, A. Stangl, and J. P. Bigham, "Vizwiz-priv: a dataset for recognizing the presence and purpose of private visual information in images taken by blind

people," in *Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 939–948.
[275] P. Korshunov and T. Ebrahimi, "UHD video dataset for evaluation of privacy," in *International Workshop on Quality of Multimedia Experience (QoMEX)*, 2014, pp. 232–237.
[276] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, "The megaface benchmark: 1 million faces for recognition at scale," in *Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 4873–4882.
[277] H. Wang, Z. Wu, Z. Wang, Z. Wang, and H. Jin, "Privacy-preserving deep visual recognition: An adversarial learning framework and a new dataset," *arXiv:1906.05675*, 2019.
[278] Council of Europe, "Protocol to the convention for the protection of human rights and fundamental freedoms (european convention on human rights) as amended by protocol no. 11," Strasbourg, 1988.
[279] European Court of Human Rights, "Guide on article 8 of the european convention on human rights," Technical Report, 2018.
[280] The United Nations, *Universal Declaration of Human Rights*, 1948.
[281] Biometrics-Institute, "First universal privacy guidelines for biometrics," https://www.biometricsinstitute.org/privacyguidelines/, 2019, accessed: 2020-05-26.
[282] ISO/IEC 30137-1, "Information technology — Use of biometrics in video surveillance systems — Part 1: System design and specification," International Organization for Standardization, Standard, 2019.
[283] ISO/IEC TR 24741, "Information technology — Biometrics — Overview and application," International Organization for Standardization, Standard, 2018.
[284] P. Voigt and A. v. d. Bussche, *The EU General Data Protection Regulation (GDPR): A Practical Guide*, 1st ed., 2017.
[285] "Privacy measures of biometrics businesses," NEC Technical Journal, Technical Report, 2018.
[286] Organisation for economic cooperation and development, "The OECD privacy framework," https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf, 2013, accessed: 2020-05-27.
[287] ISO/IEC 24745, "Information technology — Security techniques — Biometric information protection," International Organization for Standardization, Standard, 2011.
[288] 740 ILCS/14, "Biometric Information Privacy Act (BIPA)," Illinois General Assembly, Public Act 095-994, 2008.
[289] ISO/IEC 27701, "Security techniques — Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management — Requirements and guidelines," International Organization for Standardization, Standard, 2019.
[290] ISO/IEC 20889, "Privacy enhancing data de-identification terminology and classification of techniques," International Organization for Standardization, Standard, 2018.
[291] P. Bukaty, *The California Consumer Privacy Act (CCPA): An implementation guide*, 2019. [Online]. Available: http://www.jstor.org/stable/j.ctvjghvnn
[292] ISO/IEC 29134, "Information technology — Security techniques — Guidelines for privacy impact assessment," International Organization for Standardization, Standard, 2017.
[293] ISO/IEC 29190, "Information technology — Security techniques — Privacy capability assessment model," International Organization for Standardization, Standard, 2015.
[294] ISO/IEC 29100, "Information technology — Security techniques — Privacy framework," International Organization for Standardization, Standard, 2011.
[295] Council of Europe, "European convention for the protection of human rights and fundamental freedoms, as amended by protocols nos. 11 and 14," https://www.refworld.org/docid/3ae6b3b04.html, 1950, accessed: 2020-05-27.
[296] Thales Group, "Biometric data and data protection regulations (GDPR and CCPA)," https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/biometrics/biometric-data, 2020, accessed: 2020-05-27.
[297] HIPAA Journal, "What is GDPR special category data?" https://www.hipaajournal.com/what-is-gdpr-special-category-data/, 2018, accessed: 2020-05-27.
[298] L. Jehl and A. Friel, "CCPA and GDPR comparison chart," Thomson Reuters Practical Law, Technical Report, 2019.
[299] I. Wagner and D. Eckhoff, "Technical privacy metrics: A systematic survey," *ACM Computing Surveys*, vol. 51, no. 3, pp. 57:1–57:38, 2018. [Online]. Available: https://doi.org/10.1145/3168389
[300] A. Ross, S. Banerjee, C. Chen, A. Chowdhury, V. Mirjalili, R. Sharma, T. Swearingen, and S. Yadav, "Some research problems in biometrics: The future beckons," in *International Conference on Biometrics (ICB)*, 2019, pp. 1–8.
[301] K. K. S, K. Vangara, M. C. King, V. Albiero, and K. Bowyer, "Characterizing the variability in face recognition accuracy relative to race," in *Computer Vision and Pattern Recognition (CVPR) Workshops*, 2019.
[302] 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 Transactions on Technology and Society*, vol. 1, no. 1, pp. 8–20, 2020.
[303] V. Albiero, K. K.S., K. Vangara, K. Zhang, M. C. King, and K. W. Bowyer, "Analysis of gender inequality in face recognition accuracy," in *IEEE Winter Conference on Applications of Computer Vision (WACV) Workshops*, 2020.
[304] S. Gong, X. Liu, and A. K. Jain, "Jointly de-biasing face recognition and demographic attribute estimation," in *European Conference on Computer Vision (ECCV)*. Springer, 2020, pp. 330–347.
[305] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Conference on fairness, accountability and transparency*, 2018, pp. 77–91.
[306] J. Buolamwini, "Response: Racial and Gender bias in Amazon Rekognition — Commercial AI System for Analyzing Faces." https://medium.com/@Joy.Buolamwini/response-racial-and-gender-bias-in-amazon-rekognition-commercial-ai-system-for-analyzing-faces-a289222eeced, 2019, accessed: 2020-11-06.
[307] S. Lagree and K. W. Bowyer, "Predicting ethnicity and gender from iris texture," in *Technologies for Homeland Security (HST)*, 2011, pp. 440–445.
[308] A. Kuehlkamp and K. Bowyer, "Predicting gender from iris texture may be harder than it seems," in *Winter Conference on Applications of Computer Vision (WACV)*, 2019, pp. 904–912.
[309] C. Xu, Y. Makihara, R. Liao, H. Niitsuma, X. Li, Y. Yagi, and J. Lu, "Real-time gait-based age estimation and gender classification from a single image," in *Winter Conference on Applications of Computer Vision (WACV)*, 2021, pp. 3460–3470.
[310] E. Marasco, L. Lugini, and B. Cukic, "Exploiting quality and texture features to estimate age and gender from fingerprints," in *Biometric and Surveillance Technology for Human and Activity Identification XI*, vol. 9075, 2014.
[311] R. Van Noorden, "The ethical questions that haunt facial-recognition research," *Nature*, vol. 587, no. 7834, pp. 354–358, 2020.

Blaž Meden is a PhD candidate and a researcher at the Computer Vision Laboratory, Faculty of Computer and Information Science, University of Ljubljana, Slovenia. He received his Bachelor's (2013) and Master's (2016) degrees from the Faculty of Computer and Information Science and is currently working on his PhD in computer science at the University of Ljubljana. His PhD research is focused on facial privacy protection and generative deep learning approaches. More broadly, he is also interested in image based biometrics, image processing, and computer vision. Blaž is also a reviewer for a number of conferences and journals, such as IET Biometrics, Entropy, IEEE Access, IMAVIS, CVWW, ISPA, CSAE, and IWOBI.

Peter Rot is a researcher at the Laboratory for Machine Intelligence, Faculty of Electrical Engineering, and the Laboratory for Computer Vision, Faculty of Computer and Information Science, University of Ljubljana, Slovenia. He received his Bachelor's (2015) and Master's (2018) degrees from the Faculty of Computer and Information Science and he is currently pursuing a PhD in computer science at the University of Ljubljana. During his undergraduate studies he completed several internships, including internships at the Jozef Stefan Institute (Ljubljana, Slovenia) and Philips (Belfast, UK). His research interests include privacy aspects of face biometrics, sclera recognition, and deep learning. Peter is a student member of the IEEE and a reviewer for top-tier conferences and journals, e.g., TIFS, FG and CVWW.

Philipp Terhörst completed his studies in physics at the Technical University of Darmstadt in 2017. Since then, he has been working in the Smart Living & Biometric Technologies department at the Fraunhofer Institute for Computer Graphics Research (IGD) as a researcher. In the context of his doctorate, his field of work includes research at machine learning and biometric solutions, specialising on reliable, privacy-preserving, and bias-mitigating face recognition. He is author of several publications in conferences and journals such as CVPR and IEEE Access and is regularly serving as a reviewer (e.g. for TPAMI, TIP, BTAS, and ICB). For his scientific work, he received several awards such as the EAB Biometrics Industry Award 2020 from the European Association for Biometrics for his dissertation or the IJCB 2020 Qualcomm PC Chairs Choice Best Student Paper Award. Moreover, he is involved in the Software Campus program, a management program of the Federal Ministry of Education and Research (BMBF).

Naser Damer is a senior researcher at the competence center Smart Living & Biometric Technologies, Fraunhofer IGD. He received his master of science degree in electrical engineering from the Technical University of Kaiserslautern (2010) and his PhD in computer science from the Technical University of Darmstadt (2018). He is a researcher at Fraunhofer IGD since 2011 performing applied research, scientific consulting, and system evaluation. His main research interests lie in the fields of biometrics, machine learning and information fusion. He published more than 80 scientific papers in these fields. Dr. Damer is a Principal Investigator at the National Research Center for Applied Cybersecurity ATHENE in Darmstadt, Germany. He serves as a reviewer for a number of journals and conferences and as an associate editor for the Visual Computer journal. He represents the German Institute for Standardization (DIN) in ISO/IEC SC37 biometrics standardization committee.

Arjan Kuijper received the M.Sc. degree in applied mathematics from Twente University, The Netherlands, the Ph.D. degree from UtrechtUniversity, The Netherlands, and the Habitation degree from TU Graz, Austria. He was an Assistant Research Professor with the IT University of Copenhagen, Denmark, and a Senior Researcher with RICAM, Linz, Austria. His research interests include all aspects of mathematics-based methods for computer vision, graphics, imaging, pattern recognition, interaction, and visualization. He is a member of the management of Fraunhofer IGD, where he is responsible for scientific dissemination. He holds the Chair in mathematical and applied visual computing with TU Darmstadt. He is the author of over 300 peer-reviewed publications,the Associate Editor of CVIU, PR, and TVCJ, the Secretary of the International Association for Pattern Recognition (IAPR), and serves both as a Reviewer for many journals and conferences, and as a program committee member and organizer of conferences.

Walter Scheirer received the M.S. degree in computer science from Lehigh University, in 2006, and the Ph.D. degree in engineering from the University of Colorado, Colorado Springs, CO, USA, in 2009. He is an Associate Professor with the Department of Computer Science and Engineering, University of Notre Dame. Prior to joining the University of Notre Dame, he was a Postdoctoral Fellow with Harvard University from 2012 to 2015, and the Director of Research and Development with Securics, Inc., from 2007 to 2012. His research interests include computer vision, machine learning, biometrics, and digital humanities.

Arun Ross received the B.E. (Hons.) degree in computer science from BITS Pilani, India, and the M.S. and Ph.D. degrees in computer science and engineering from Michigan State University. He is the John and Eva Cillag Endowed Chair with the College of Engineering and a Professor in the Department of Computer Science and Engineering, Michigan State University. He was the faculty of West Virginia University between 2003 and 2012 where he received the Benedum Distinguished Scholar Award for excellence in creative research and the WVU Foundation Outstanding Teaching Award. Dr. Ross is a recipient of the NSF CAREER Award and was designated a Kavli Fellow by the U.S. National Academy of Sciences in 2006. He received the JK Aggarwal Prize in 2014 and the Young Biometrics Investigator Award in 2013 from the International Association of Pattern Recognition. He is the coauthor of the textbook Introduction to Biometrics and the monograph Handbook of Multibiometrics.

Peter Peer is a Professor of computer science at the University of Ljubljana, Slovenia, where he heads the Computer Vision Laboratory, coordinates the double degree study program with the Kyungpook National University, South Korea, and serves as a vice-dean for economic affairs. He received his doctoral degree in computer science from the University of Ljubljana in 2003. Within his post-doctorate he was an invited researcher at CEIT, Donostia – San Sebastian, Spain. His research interests include biometrics, color constancy, image segmentation, detection, recognition and real-time computer vision applications. He participated in several national and EU funded R&D projects and published more than 100 research papers in leading international peer reviewed journals and conferences. He is co-organizer of the Unconstrained Ear Recognition Challenge and Sclera Segmentation Benchmarking Competition. He serves as an Associated Editor of IEEE Access and IET Biometrics. He is a member of the EAB, IAPR, and IEEE, where he also served as a chairman of the Slovenian IEEE Computer chapter for four years.

Vitomir Štruc is an Associate Professor at the University of Ljubljana, Slovenia. He received his doctoral degree from the Faculty of Electrical Engineering in Ljubljana in 2010. Vitomir's research interests include problems related to biometrics, computer vision, image processing, pattern recognition, and machine learning. He (co-)authored more than 100 research papers for leading international peer reviewed journals and conferences in these and related areas. He served in different capacities on the organizing committees of several top-tier vision conferences, including IEEE Face and Gesture, ICB, WACV, and IJCB. Vitomir is a Senior Area Editor for the IEEE Transactions on Information Forensics and Security, and an Associate Editor for Pattern Recognition, Signal Processing, and IET Biometrics. He served as an Area Chair for WACV 2018, 2019, 2020, ICPR 2018, Eusipco 2019, and FG 2020. Dr. Štruc is a member of the IEEE, IAPR, EURASIP, Slovenia's national contact point for the EAB and the current president of the Slovenian Pattern Recognition Society (Slovenian branch of IAPR).