Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-15775-2

A review on visual privacy preservation techniques for active and assisted living

Siddharth Ravi[1] · Pau Climent-Pérez[1] · Francisco Florez-Revuelta[1]

Received: 2 November 2022 / Revised: 27 March 2023 / Accepted: 2 May 2023

Abstract

This paper reviews the state of the art in visual privacy protection techniques, with particular attention paid to techniques applicable to the field of Active and Assisted Living (AAL). A novel taxonomy with which state-of-the-art visual privacy protection methods can be classified is introduced. Perceptual obfuscation methods, a category in this taxonomy, is highlighted. These are a category of visual privacy preservation techniques, particularly relevant when considering scenarios that come under video-based AAL monitoring. Obfuscation against machine learning models is also explored. A high-level classification scheme of privacy by design, as defined by experts in privacy and data protection law, is connected to the proposed taxonomy of visual privacy preservation techniques. Finally, we note open questions that exist in the field and introduce the reader to some exciting avenues for future research in the area of visual privacy.

Keywords Visual privacy preservation · Active and assisted living · Privacy by design · Perceptual obfuscation · Machine obfuscation · Facial de-identification

1 Introduction

Active and Assisted Living (AAL) systems aim to improve the quality of life for older adults and individuals with disabilities by leveraging information and communication technologies in a range of environments such as homes, workplaces, and public spaces. These systems integrate an array of sensors, which can be either worn by the user or installed in the

Pau Climent-Pérez and Francisco Florez-Revuelta contributed equally to this work.

✉ Siddharth Ravi
siddharth.ravi@ua.es

Pau Climent-Pérez
pcliment@dtic.ua.es

Francisco Florez-Revuelta
francisco.florez@ua.es

[1] Department of Computing Technology, University of Alicante,
San Vicente del Raspeig, Alicante 03690, Valencian Community, Spain

Published online: 04 July 2023

environment, to gather information about the individual's status and surroundings, enabling seamless interaction between the person and their environment. The data collected by these sensors is then processed by intelligent systems to offer tailored and advanced healthcare services.

The use of video-based devices in AAL is becoming increasingly common due to the application of computer vision techniques that enable the monitoring of environments and reporting of visual information. This is often the most direct and natural way of describing events, people, objects, actions, and interactions [119]. These advancements have transformed video cameras into 'smart cameras' and expanded their capabilities to tasks such as face recognition, object recognition and tracking, people identification, recognition of actions and activities of daily living, and even human behaviour analysis over an extended period [25, 35]. However, despite their potential benefits, their usage is currently limited, mainly due to ethical, legal, and privacy issues.

There are two major issues to address regarding visual privacy. One is identity protection, where the identity of the person in a visual is to be hidden from entities who might analyse the feed without the necessary access privileges. Following convention, these entities will be addressed as *adversaries* in this review. Adversaries can either be persons who view sensitive visuals without being provided with the necessary consent, or machine learning models that train on data collected without user consent. The second reason for visual privacy is the preservation of trust for persons who are monitored.

A typical AAL care home might be equipped with RGB cameras, the feeds of which are monitored and analysed to provide support to the residents in time of need. In these cases, the identity of the resident is of relatively less interest, as that is usually of a more public nature. A typical AAL care home resident, for example, could have given consent for them to be monitored by the home's personnel and their family for safety reasons. But a level of trust needs to be preserved for cameras to be deployed in privacy-sensitive settings. Borrowing the categorisation of privacy provided by Clarke [32, 33], what is crucial, however, is the need to preserve the resident's bodily privacy in various sensitive scenarios. Bodily privacy refers to the privacy regarding images of the body. More precisely, it considers the activities that are carried out, and the loss of privacy given the nature of some of these activities (e.g., nudity during showering, etc.). What is also of interest is to preserve the privacy of sensitive personal behaviour, such as a person's political activities, sexual habits, religious practices, and with the personal space required to facilitate such behaviour. To obtain and preserve this element of trust, visual privacy needs to be preserved at every stage of a system used for monitoring.

With this idea in focus, this document surveys the state of the art in visual privacy protection methods, with special attention paid to the concept of visual obfuscation. The dichotomy between identity protection and bodily privacy can also be observed in the classification scheme this paper proposes for visual privacy preservation techniques. *Perceptual obfuscation* methods (explained in Section 4.1) aim to preserve trust through the protection of bodily privacy. *Machine obfuscation* methods (explained in Section 4.2) are mainly aimed at the protection of identity from machine learning models.

This review introduces a framework with which visual privacy protection methods can be classified under, and introduces terminology that can be used to categorise methods developed to provide visual privacy. It attempts to capture the field in a broad sense, while also connecting the state-of-the-art in the field to the framework of privacy by design [24]. This is important, since privacy is a societal problem, rather than being a challenge that is purely technical in nature. Solutions that are deployed need to provide privacy from the ground up, while providing users with enough knowledge and options to control the flow of data which is obtained from their actions.

This work is meant to serve as more than merely a survey of the state-of-the-art. It seeks to provide the connection between high-level concepts defined in the area of privacy by design to the lower level taxonomy of methods proposed in this review. This is meant to introduce the reader to the idea of end-to-end privacy preserving systems to be used in environments like care homes, to highlight the practical relevance of privacy preserving technologies developed, and to push the field towards a place where more of the techniques developed through research are deployed in real-world scenarios. This is especially important when considering the ageing demographics in most developed nations, a trend that is expected to continue in the future. Considering this, there is the urgent need for more privacy to be imparted to the part of the population which will require monitoring to receive long-term care in private settings or in care homes.

1.1 Contributions

The central contributions of this review are as follows:

1. With emphasis on visual obfuscation methods, this paper reviews the state of the art in visual privacy protection methods.
2. It proposes a novel classification scheme to make sense of visual obfuscation methods.
3. This paper connects low-level concepts in the field of visual privacy to high-level concepts encountered when discussing privacy by design.

1.2 Review structure

The rest of the review is structured as follows. Section 2 looks at prior relevant reviews. Section 3 explores the state of the art in visual privacy protection methods. A novel classification scheme for the methods in this category is also introduced. Here, the review expands on those methods that are classified by the scheme under the categories of *intervention methods*, *blind vision*, *secure processing* and *data hiding* [104].

Section 4 explores in greater detail the state of the art in visual obfuscation methods, another subcategory of visual privacy protection methods that is essential to this review.

Section 5 explains the concept of privacy by design, a high-level concept in systems design essential to the creation of truly end-to-end private systems. In this section, the paper links together a categorisation scheme proposed for ensuring privacy by design to the scheme proposed in this review for categorising visual privacy protection methods.

Section 6 introduces the reader to performance evaluation setups used when measuring the efficacy of privacy preservation techniques. Important technical privacy metrics which are frequently employed are explored. It also introduces the reader to datasets that are commonly used to train models that work to impart visual privacy. Meta-studies are also explored which evaluate the real-life effectiveness of performance evaluation frameworks employed for privacy preservation techniques, through the use of user acceptance studies. Finally, Section 7 concludes the survey by introducing the reader to important future work to be conducted to advance the field.

2 Prior reviews

Prior work has attempted to systematise knowledge in the field of visual privacy preservation [88, 104, 121]. Padilla-López et al. (2015) [104] introduces the reader to a taxonomy of visual privacy preservation techniques seen in the literature. These are grouped under five major categories based on the manner in which they impart privacy, these being *Intervention Methods*, *Blind Vision*, *Secure Processing*, *Data Hiding*, and *Redaction methods*. Redaction methods are further subdivided into *image filtering, encryption, k-same family of algorithms, object / people removal*, and *visual abstraction*. The authors also provide a survey of privacy-aware intelligent monitoring systems as part of their review.

Another more recent work by Meden et al. (2021) [88] provides a taxonomy of methods for the area of biometric privacy enhancing technologies, paying particular attention to facial biometrics. The survey also introduces a taxonomy of biometric privacy enhancing techniques. The taxonomy of methods is grouped based on 6 criteria, namely - *the biometric attributes used*, *biometric utility*, referring to the usefulness of data for automatic extraction of various attributes like health indicators and identity information, *guarantees of reconstruction* from privacy-enhanced data, *target* from which the data is to be hidden, *type of mapping used* (reversible or irreversible mapping), and *type of data* the method is applied to. The classification scheme introduced along with the grouping criteria can be seen in Fig. 1.

The survey by Ribaric et al. (2016) [121] is a broader survey of the field of privacy preservation, touching on aspects of privacy for multimedia data, including both visual and non-visual (e.g. audio) data. The survey provides an overview of de-identification approaches for non-biometric identifiers (e.g. text, hairstyle, dressing style, licence plates), physiological identifiers (e.g. face, fingerprint, iris, ear), behavioural (e.g. voice, gait, gesture) and soft-biometric (e.g. body silhouette, gender, age, race, tattoo) identifiers in a multimedia context (Fig. 2). The authors then present examples of methods used to provide privacy to users based on these classifiers.

In contrast to the prior reviews in the field, this work seeks to present privacy preservation techniques that are meaningful in AAL applications. Therefore, the focus is on protecting bodily privacy, and is not concerned with whether the identity of the person is protected, as that is something commonly of a public nature. A broader exploration of the state of the art

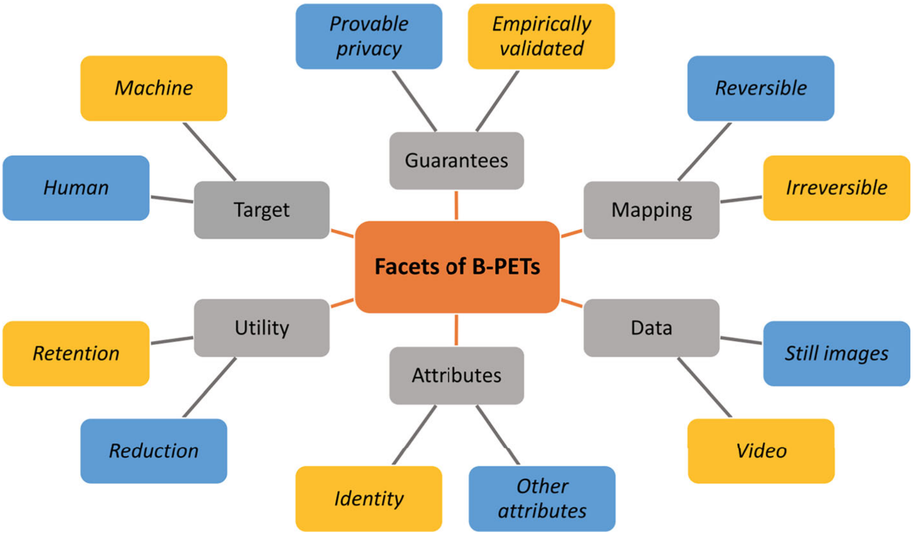


Fig. 1 Classification of Biometric Privacy Enhancing Technologies (Reprinted from [88])

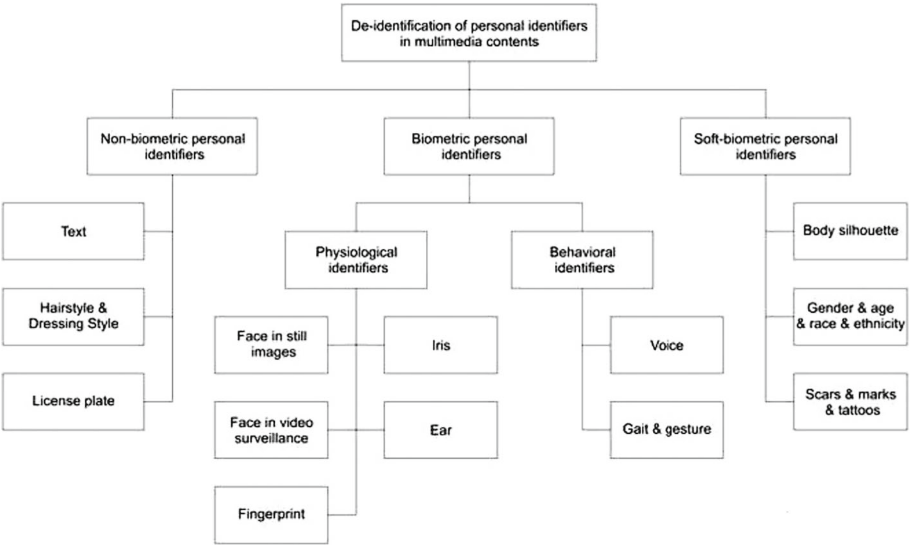

De-identification of personal identifiers in multimedia contents
Non-biometric personal identifiers
Biometric personal identifiers
Soft-biometric personal identifiers
Text
Hairstyle & Dressing Style
License plate
Physiological identifiers
Behavioral identifiers
Face in still images
Iris
Face in video surveillance
Ear
Fingerprint
Voice
Gait & gesture
Body silhouette
Gender & age & race & ethnicity
Scars & marks & tattoos

Fig. 2 Taxonomy of identifiers in multimedia content (Reprinted from [121])

is presented, tying together concepts from the privacy by design literature to ideas coming from computer vision.

As the focus of this review is on biometric identifiers that affect bodily privacy in the scenario of visuals from private settings or care home environments supporting AAL, some identifiers of direct importance here are behavioural identifiers (e.g., gait, gestures, actions, or activities), dressing styles, and body silhouettes. It might also be the case that wearable cameras are used to provide an AAL service. In this case, when the user moves out of the private environment, they might encounter other persons who might not have consented to being monitored. Hence, there is a need for stricter measures of privacy to be implemented through the obfuscation of other biometric identifiers. These are faces (in still and video images); gait, and gesture; scars, marks, and tattoos; and the hair, and dressing style. These have the potential to reveal the identity of passers-by to observers of the visual feed.

Obfuscation of some of these above-mentioned identifiers: scars, marks, and tattoos, and the hairstyle or dressing style have not been explored in the literature to the best of the authors' knowledge. Anonymisation techniques targeting other identifiers are explored in some depth in the next sections of this review, namely those concerning body silhouettes (using full-body de-identification), gait, and faces.

2.1 Methodology

Papers in the field of visual obfuscation reviewed in this work are listed in Table 1. Importance is given to research published in the field of *perceptual obfuscation*, as it is especially relevant for AAL. This work also puts more emphasis on work published after 2016, as it reviews the advances in the field which are not covered in the review by Padilla-López et al. [104]. Since the rise of deep learning, the field of computer vision has also undergone a revolutionary change. Arguably, most state-of-the-art methods proposed to impart visual privacy attempt

Table 1 Categorisation of visual obfuscation approaches reviewed

Category	Sub-category	Approach	Reference(s)
		Morphing	[71]
		Warping	[72]
	Image Filtering	Cartooning - mean shift / adaptive filter	[44], [43]
		Cartooning - using convolutional neural networks	[56]
		False Coloring	[31]
		PECAM	[153]
		Adaptive Blurring	[166]
Perceptual Obfuscation		Head Inpainting	[134]
	Facial De-Identification	Live Facial de-Identification	[47]
		AnonymousNet	[78]
		Gait anonymisation using deep learning	[140]
	Gait Anonymisation	STGAN	[141]
		Gait anonymisation from Low quality silhouettes	[139]
		Generative Full Body and Face De-Identification	[22]
		SMPLicit	[36]
		FrankMocap	[123]
	Total Body Abstraction	DensePose	[97]
		Dense correspondences using depth sensors	[137], [148] , [116]
		Dense correspondences using RGB images	[20], [171], [48]
		Object removal using PDE inspired algorithms	[110], [14], [125]
		Object Removal - Exemplar	[38]
		Object Removal - Hybrid	[15], [167], [29]
		Object Removal - Deep Learning	[157], [159], [66], [26], [77], [163], [102]

Table 1 continued

Category	Sub-category	Approach	Reference(s)
Machine Obfuscation	Evasion Attacks	Spectacles	[129]
		Adversarial stickers and patches	[68], [23], [152], [138]
	Poisoning Attacks	Clean label attacks	[127], [173], [131]
		Model Corruption	[132]

to do so through the use of deep learning. This is also reflected in the methods surveyed as part of this review.

Works surveyed were selected primarily through the use of Google Scholar. As the proposed taxonomy expands on the work proposed by Padilla-López et al. (2015) [104], filtering was done on works published in or after 2016. The keywords used for the searches include Visual privacy, survey, avatar, visual abstraction, SMPL, filter, privacy filter, facial privacy, face anonymization, full-body anonymization, body replacement, gait anonymization, and gait privacy.

This yielded search results that were then filtered based on the fit of the work, the publishing venue (filtering was done so that only Q1 and Q2 journals according to the Clarivate Journal Citation Reports[1] were selected, along with conferences that fall into the top quartile of conference rankings (A or A* from the Computing Research and & Education conference rankings[2] were selected)). Most exclusions were done based on assessing the relevance of the document at the title and abstract level, with fewer falling to the category of not fitting into the theme of the review.

Exceptions to these filtering rules were also applied, especially when there were only few publications in the area. For selecting works relating to gait anonymisation, for example, it was necessary to select papers from venues that fell outside the selection criterion as this is a research area that is arguably not widely explored in the literature.

3 Visual privacy preservation methods

Building on the taxonomy for visual privacy preservation methods introduced by Padilla-López et al. (2015) [104], this review categorises visual privacy preservation methods into 5 categories: *intervention methods*, *blind vision*, *secure processing*, *data hiding*, and *visual obfuscation* (Fig. 3).

3.1 Intervention methods

Intervention methods are those techniques that interfere during the data collection phase, preventing private visual data from being collected from the environment. Perez et al. [109] classify these methods under three categories - *sensor saturation*, *broadcasting commands*, and *context-based approaches*.

Sensor saturation methods impart privacy by feeding the input device's sensor a signal that is far more in amplitude than the maximum that the device can handle. Physical interventions that prevent the capture of private images under sensor saturation schemes are also present under this category. One of the most commonly used intervention methods of this type are commercial webcam covers, also known as privacy stickers for laptops and phone cameras. These are stickers that can be stuck onto the camera, and some can be closed and opened at will. The nature of the adhesive and the construction of the blocking mechanism differs between methods [11, 54, 64, 93, 94, 122].

The Blindspot system [106] consists of a camera lens tracking system that locates retroreflective CCD or CMOS cameras in the vicinity, along with directing a pulsing light at the camera's lens that distorts recorded visuals. Anti-paparazzi devices have also been devised

[1] https://jcr.clarivate.com/jcr/home

[2] http://portal.core.edu.au/conf-ranks/

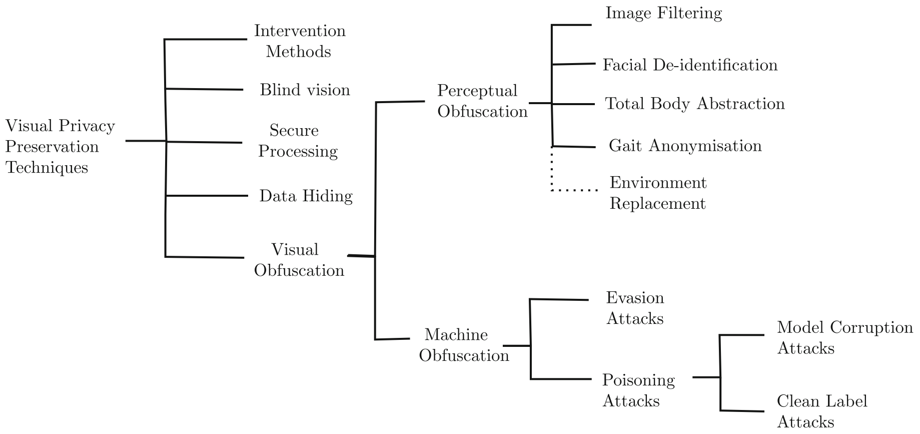


Fig. 3 A taxonomy of visual privacy preservation techniques for AAL. The topic of environmental privacy is connected with dotted lines to show that it is an under-researched but important topic

that qualify as intervention methods. Harvey and Knight [55] describe anti-paparazzi devices that are cloaked as fashionable clutch bags. These detect camera flashes with the use of light sensors along with IR sensors to detect autofocus lights. The intervention device then uses an array of LEDs to produce pulses of light bright enough to overexpose photos taken by the photographers.

Zhu et al. [174] created the concept of LiShield, which protects a physical scene against photographing. This is achieved through the use of smart LEDs, which emit specially constructed waveforms to illuminate a scene. The LEDs emit intensity modulated waveforms that are imperceptible to the human eye, but their waveforms are constructed in such a way as to interfere with the image sensors of mobile camera devices. Mobile phones have also started to be shipped with inbuilt mechanisms for sensor saturation-based intervention. Examples include the PinePhone [115] which comes with physical ‘kill switches’ for configuring its hardware. These can be individually configured to disable both its front and rear cameras, among other peripherals [114].

Broadcasting commands are another category of intervention methods, where devices broadcast commands using various communication protocols to disable input devices present around the subject. One example is Hewlett-Packard’s concept of a paparazzi-proof camera. This includes cameras with inbuilt facial recognition, which upon receiving a remote command, selectively blurs sensitive parts of images containing faces [113]. Broadcasting commands are considered less effective than their physical counterparts, because user consent[3] is required for these methods to work. Broadcasting commands are also arguably less popular as intervention methods than sensor saturation methods.

Context-based approaches are used by devices that use various methods of context recognition to understand the scene of data collection. Once recognised, the context is used to dictate whether data is to be collected or not by triggering software actions at the sensor level. One example of this is the *Virtual Walls* framework described by Kapadia et al. [65], where devices use contextual information such as GPS data to trigger software action like the disabling of sensors in the device. This allows users to control their digital footprint. To

[3] In this case the photographer’s as they are the users of the camera

the best of the authors' knowledge, this has not been implemented in commercial devices. Context-based approaches are also arguably less popular than other intervention methods.

3.2 Blind vision

Blind vision refers to the methods by which the processing of images and videos is done in an anonymous way [9, 10, 45, 126]. Blind vision methods allow commonly used computer vision tasks to be executed without compromising on the privacy of neither the algorithm used for computing, nor the data itself. Blind vision works through the use of secure multiparty computation (SMC) techniques, a subfield of cryptography that allows computations to be performed privately. This allows algorithms to be executed privately, but at the same time leads to the slowdown of computation due to the overhead involved.

3.3 Secure processing

Those privacy preservation methods that are not based on SMC, but which still can process visual information in a privacy respectful way, are classified in this review under secure processing. These refer to algorithms and queries where privacy is required in a unidirectional sense: the databases on which the queries are performed are usually public, but the query and its results are to be kept private. One relevant example is the image matching algorithm for private content-based image retrieval (PCBIR) [130]. Algorithms that reject visual information that is not necessary for processing are also considered by the authors to be under the framework of secure processing. As an example, consider the concept of using depth or thermal cameras as the sensor device in conducting privacy preserving machine learning. These devices allow the observer to glean some information from the visual feed (e.g., number of people in the room, the activity being performed etc) while hiding the most commonly utilised privacy-sensitive information (facial identity, location information, etc) [58]. The visual anonymisation strategy proposed in Al-Obaidi et al. [4] that still allows for human action recognition, is another example of an algorithm that comes under the umbrella of secure processing. The authors propose the use of an anonymization strategy resulting in the creation of highly anonymised silhouettes of the person being observed, thus allowing only the motion of body parts involving an action to be intelligible on the feed.

There are also secret sharing schemes that can be classified under secure processing, wherein inference is not done on the original data, but on privacy preserving derived data obtained from the original. One example is the scheme proposed by Upmanyu et al. [142], in which images are split into multiple privacy preserving parts, which can then be distributed across nodes. Algorithms can then be applied on these image parts privately. Homomorphic encryption schemes also figure into the space of secure processing. These allow data to be encrypted in such a way that algorithms can still be run with utility on the resulting encrypted data, thereby protecting privacy. Homomorphic encryption has been successfully applied in computer vision applications as well [16, 158].

3.4 Data hiding

Data hiding methods refer to privacy preservation methods that, in addition to modifying privacy-sensitive regions in images, aim to embed the original information inside the modified image so that the original can be retrieved if its need arises. Petitcolas et al. [111] provide a useful classification of data hiding methods. Under the process, embedded data (secret message) is hidden within another message (cover message) which in this case is a video frame. Thus, a marked message is obtained as a result of this hiding process. Data hiding techniques include steganography, digital watermarking, and fingerprinting. Steganography uses a key to allow the recovery of the secret message. Digital watermarking encodes the information about the ownership of an object by a visible pattern, such as a logo. Fingerprinting, conversely, hides serial numbers that uniquely identify an object inside an image, such that the owner of the copyright can detect violations of licence agreements. In the context of visual privacy protection, watermarking can be used to hide the sensitive attributes in an original video inside an obfuscated version. As an example, for facial privacy preservation, Yu and Babaguchi [160] hide real faces inside frames of a video where the real face has been replaced by a generated one. Quantisation index modulation [28] is used for the process of data hiding, and the original information can be retrieved using a secret key. This method, however, has limitations such as the artificial nature of the generated faces, and a lack of control for the generated expressions.

Depending on whether the method is fully reversible or not, data hiding techniques allow recovery of the original video to various extents. Fully reversible data hiding methods allow the original to be restored without information loss [100]. With non-reversible methods, the original image cannot be fully restored, but this usually means an increase in hiding capacity [155, 165].

PECAM [153] is a method that uses elements of data hiding for creating reversible privacy-preserving transformations of images. This is, however, a method which can be used in two different modalities where the system can either produce reversible image transformations or be irreversible. For this reason, in this review, PECAM has been categorised as a visual obfuscation method and is explained in more detail in Section 4.

4 Visual obfuscation

This work classifies methods that seek to hide sensitive visual information directly from adversaries under visual obfuscation methods. They are divided into two major categories, *perceptual obfuscation* and *machine obfuscation*, based on their intention and the type of adversary from whom the private data in an image is to be obfuscated. The landscape of visual obfuscation methods analysed in this review can be seen in Table 1.

The following sections deal with the state of the art in each of the major subcategories of perceptual obfuscation methods.

4.1 Perceptual obfuscation: Targetting human observers

In the case where obfuscation targets human observers, methods aim to impart visual privacy for users who wish to keep private from humans without the necessary access privileges, i.e. *perceptually* (therefore, ‘perceptual obfuscation’). The primary objective of this category of methods is to create images in which the privacy-sensitive elements are perceptually different from the original. Although the lines are blurred between some methods, these types of techniques can broadly be split into five subcategories of methods based on the result - Image filtering, facial de-identification, total body abstraction, gait anonymisation, and environment replacement. The latter, being an under-researched subject, is discussed in Section 7.1.1 of this review.

Perceptual obfuscation methods can also be either reversible in nature, where the original image can be retrieved after modification, or conversely be irreversible. A broad treatment of the classical literature in perceptual obfuscation is available in Padilla-López et al. [104].

4.1.1 Image filters

Image filtering is a class of perceptual obfuscation techniques that relies on the alteration/redaction of images in a way that imparts privacy to an image. Image filters can be applied globally to entire images, or to sensitive parts of images where privacy is required. The simplest forms of these filters are blurring and pixelation.

Blurring filters slide a Gaussian kernel over an image, thereby using neighbourhood pixels to influence the values of a central pixel (Fig. 13f). Although widely used in applications as large as Google Maps, blurring has been shown to be ineffective for protecting identity against various deep learning-based attacks, even while appearing de-identified to human observers [87, 101]. For pixelation, a grid of a certain size is chosen for the sensitive pixels in an image. For each box in the grid, an average colour over all the pixels within the box is calculated and assigned to each pixel within the box (Fig. 13e).

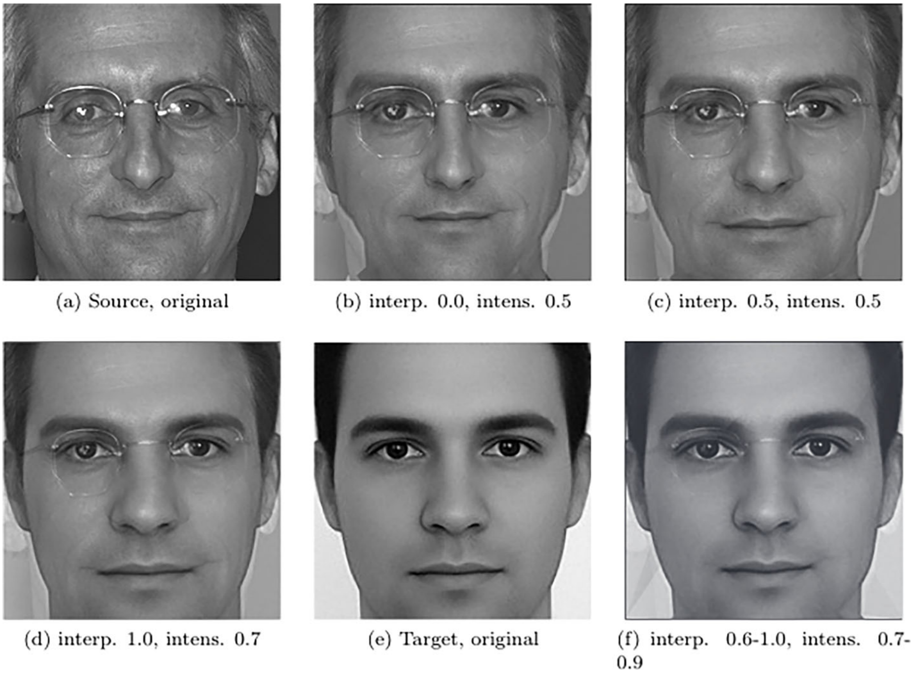
Image filtering has been widely used in the media, especially to obscure the identity of subjects who want to remain anonymous. These have, however, been primarily used offline due to difficulties caused due to target drift across frames, the possibility of over-filtering, and computing efficiency reasons. Real-time variants have, however, also been explored for use during live-streaming. Zhou and Pun [172], for example, created ‘Face Pixelation in Video Live Streaming´ (FPVLS) that allows for irrelevant[4] face tracking and pixelation in real time. The system utilises a multi-stage pipeline involving, in order, face detection and embedding networks [146, 168] to obtain facial embedding vectors, a clustering algorithm (Positioned Incremental Affinity Propagation) to associate the same person’s faces across frames, and a refinement stage involving a two-sample test based on the empirical likelihood ratio statistic to solve issues of drift in the proposed regions across frames.

These simpler image filtering techniques have, however, been shown in various studies to not be robust in providing privacy [70, 87, 90, 98]. Deblurring techniques have also been researched in literature [75, 124, 169]. It could be posited that these techniques can also be repurposed as attacks against images obfuscated using blurring filters. Commercial tools for deblurring have also been developed [67].

Morphing and warping are filtering techniques primarily used for facial anonymisation. In morphing [71], the input face is morphed into a target face (see Fig. 4). This is done using interpolation and intensity parameters, which are used to steer the positions of the keypoints in the input face towards the target. In warping [72], a set of keypoint parameters are determined using face detection techniques. These keypoints are then shifted according to a ’warping strength’ parameter. The new intensity values are determined using interpolation.

Çiftçi et al. [31] propose a false colour filter as a means of visual privacy for images, which involves converting RGB images to greyscale and mapping the pixel intensities to a set of RGB pixel values based on pre-defined colour palettes. The scheme is reversible, allowing the original image to be retrieved through storing a difference image and a sign image. The method is lightweight and can be applied to any RGB image, though it is vulnerable to attack through neural networks that learn the association between false colour pixels and the real

[4] The term ’irrelevant faces’ refers to the faces of people which are not of the primary subject being tracked in the video


(a) Source, original

(b) interp. 0.0, intens. 0.5

(c) interp. 0.5, intens. 0.5

(d) interp. 1.0, intens. 0.7

(e) Target, original

(f) interp. 0.6-1.0, intens. 0.7-0.9

Fig. 4 Morphing using various levels of interpolation and intensity parameters noted under each facial image (Reprinted from [71])

object's colours, compromising the privacy protection. Example results of the method are presented in Fig. 5.

Adaptive blurring [166] is an algorithm that blurs privacy-sensitive parts of videos using semantic segmentation masks. The algorithm uses DeepLab [27] to create segmentation masks and a scale-dependent Gaussian blur to blur the sensitive areas based on the mask. The algorithm also uses a custom symmetry-based strategy to guide the Gaussian blur application on object edges. The filter radius and standard deviation for the Gaussian blur kernel are set based on the estimated bounding box size. However, this approach does not account for camera distortion or depth uncertainty, potentially leading to under-blurring or over-blurring. Furthermore, commercial tools can deblur obfuscated images, reducing the security of the pipeline [67].

Cartooning has been proposed multiple times in literature as a method for filtering images for privacy reasons. Erdélyi et al. [44], for example, introduce a 'meanshift'-based method for cartooning. With this, they reduce the total number of colours and simplify the texture based on a neighbourhood pixel's property, and use edge recovery to preserve the sharpness of edges in the image. They also blur faces as part of the algorithm, and recolour parts of the image by shifting the hue as part of the final algorithm. Erdélyi et al. [43] also improve the previous work with the introduction of an adaptive filter, allowing users to determine the level of obfuscation. Hassan et al. [56] introduce a deep learning scheme for cartooning videos, by which privacy-sensitive objects in videos are replaced by abstract cartoon clip art. For this, a region convolutional neural network (R-CNN) [49] is used to get bounding boxes for the privacy-sensitive personal objects in the video. After selecting the right clip art and correcting for pose (the algorithm utilises the histogram of oriented gradients method [39]),

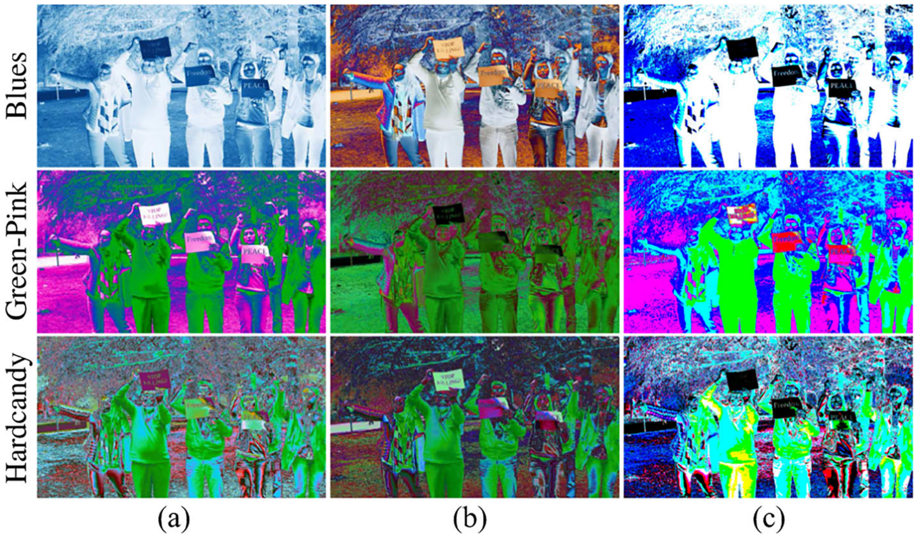


Fig. 5 False colouring done using the various palettes mentioned as row titles. The columns from left to right represent the final false image obtained after filtering, the difference image, and the sign image respectively (Reprinted from [31])

the clip art is inserted into the frame, creating privacy-preserving cartooning effects. Figure 6 shows the results before and after using the method.

Encryption methods for images can be viewed as image filtering that is reversible using a key. Naive encryption schemes treat images as textual data and encrypt the entire stream, leading to inefficiencies in real-time scenarios. To address this issue, selective encryption schemes have been proposed that only operate on specific parts of the image, reducing the total computation cost. Much of the classical literature in encryption is summarised in Padilla-López et al. [104]. One notable recent attempt at using encryption for visual privacy preservation is by Zhang et al. [164], who combine the concept of thumbnail preserving encryption (or TPEs [151]) which replaces the images with their approximate thumbnail as a

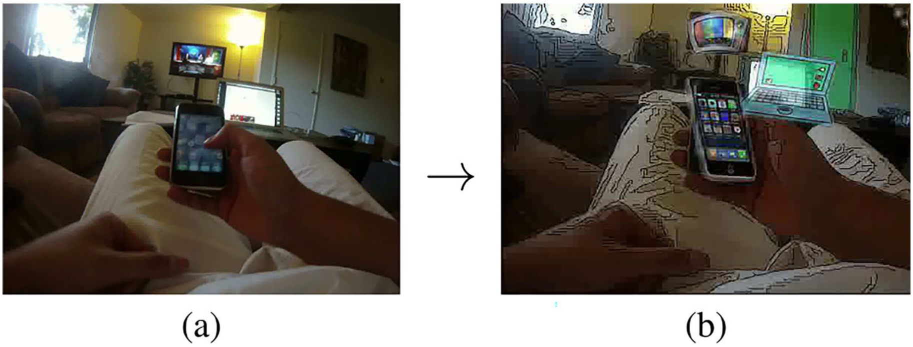


Fig. 6 Image filtering done using cartooning. (a) shows the original image, while (b) is the resulting image after the method has been applied (Reprinted from [56])

replacement that balanced privacy and utility, with chaotic systems that generate randomness for encrypting the frame. This reduced the time required for encryption and decryption[5].

PECAM [153] is a system that allows for reversible filtering transformations through the use of data hiding. The PECAM system is built for streaming, and allows for the creation of filtered images that can then be reconstructed if such a need arises. In this scheme, depending on whether the model is aiming to reconstruct the images after transformation, different directions in the pipeline are followed. A generator (referred to as a transformer in the paper) neural network and discriminator (termed reconstructor) network are trained using the cycle-consistent GAN approach. The transformer is used to generate filtered images, and the reconstructor is used to regenerate the originals if need be.

In the pipeline that requires reconstruction, a secret key is generated that is used by the transformer and the reconstructor to guide the transformations. This is embedded into the image using data hiding (steganography) as an alpha channel. This RGBA image is then fed to the generator network, which after compression produces a filtered image that preserves privacy. This filtered image can then be broadcast to viewers. This image can then be fed to the reconstructor to create a reconstruction of the original image. In the cases where reconstruction is not necessary, a lightweight network is used as the generator, which is created through model distillation of the original network. After compression, this student network outputs the filtered image that is broadcast to viewers.

One disadvantage of the PECAM network is that the network could cause privacy leakage, as it might not work well when the privacy-sensitive objects are close to the camera.

4.1.2 Facial de-identification

Facial de-identification involves generating artificial faces to protect facial features from identification. These artificial faces need to be blended into the original image. The traditional method for this task is to use the *k*-same family of algorithms [51, 52, 98].

State-of-the-art facial de-identification methods use Generative Adversarial Networks (GANs). One such method is by Sun et al. [134], which uses keypoint generation to condition an adversarial autoencoder (deep convolutional GANs). The scheme has two stages: the first uses either a feature-redacted blacked-out or blurred image or the original image as input. If the former, a landmark generator estimates facial landmarks as a heatmap; if the latter, a landmark detector extracts the heatmap. The second stage takes the concatenated heatmap and blacked-out original as input and generates realistic-looking faces through another adversarial DCGAN autoencoder. Figure 7 illustrates this method.

Gafni et al. [47] propose a live facial de-identification method for videos, where the system distances facial descriptors from a target image of the person provided to the system. Facial bounding boxes and keypoints are extracted from the video frame, and a similarity transformation matrix is obtained from these using an averaged face. The input face is transformed using this matrix and passed through an adversarial autoencoder network to obtain an output facial image and a mask. A linear per-pixel mixing of the input and output images is done, weighted by the transformed mask, and then merged into the original frame using the convex hull of facial keypoints to generate the final output. Figure 8 illustrates this method.

The approach by Li and Lin [78] is interesting for the way it straddles the worlds of both perceptual obfuscation and machine obfuscation (explored in Section 4.2). This method,

[5] TPEs allow for the creation of encrypted images that when are made into thumbnails exactly resemble the thumbnails of the original images. These can then be decrypted into their original versions using a decryption scheme

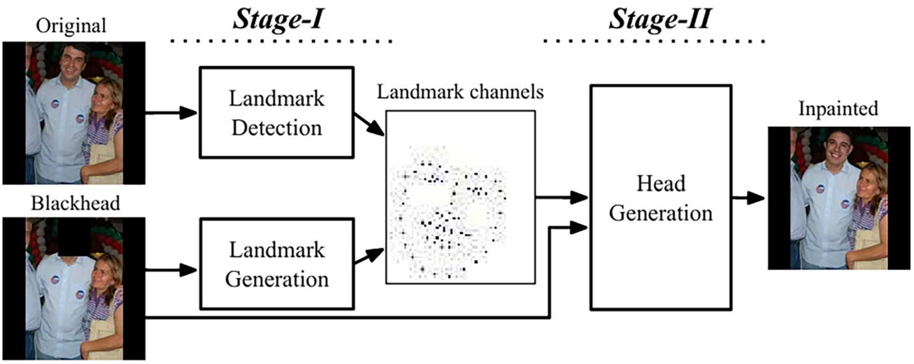


Fig. 7 Two-stage facial de-identification framework used by Sun et al. (2018) [134]. The first stage outputs a facial landmark heatmap, which is either generated or detected depending on input. This is then fed to a head generation network in the second stage along with the blackhead input image, and a generated head is inpainted into the image (Reprinted from [134])

named AnonymousNet, creates perceptually altered images based on knowledge of both the facial attributes of persons observed and the distribution of those attributes in the real world (approximated by the dataset in the image). The method aligns and crops faces using a neural net referred to by the authors as a *deep alignment network*, after which it does facial feature extraction using GoogleNet [136] and random forest models [19]. This is then used as input to a custom privacy preserving attribute selection algorithm, which obfuscates the features of the face and lets the outputs resemble the features of the real world in terms of their distribution. A de-identified face is then generated by a starGAN [30] model, conditioned by the features selected by the algorithm in the previous step. Finally, to obfuscate the outputs from machines, adversarial perturbation is done on the output image, using a universal perturbation vector defined by the DeepFool algorithm [96].

4.1.3 Total body abstraction

Total body abstraction methods aim to impart privacy by replacing the entire body of the subject in a visual with another generated one. Most methods under this category arguably

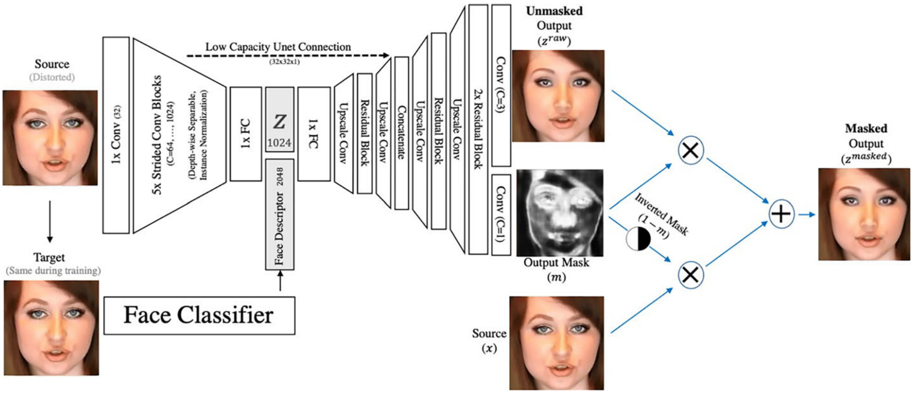


Fig. 8 Framework used by Gafni et al. (2019) [47]. The setup outputs a de-identified facial image with a similar pose, illumination, and expression to the original (Reprinted from [47])

use semantic segmentation methods to segment out humans from frames, and then subsequently replace these with abstractions such as avatars. Other visual abstractions include silhouettes, where a binary mask of the person is obtained (and sometimes modified for various purposes); invisibility, where inpainting techniques are used to replace the person with the environment/background [34]; and background subtraction, where a background image is generated and subtracted from the current frame to obtain a mask of the foreground object (here a person) of interest [95, 120].

One particularly interesting total body abstraction method relied on the use of generative adversarial models to generate full-body replacements. The approach by Brkic et al. [22] uses conditional GANs (DCGANs) to synthesise entire bodies of subjects, while the faces are generated using deep convolutional GAN models. The conditional GAN was trained on pairs of segmentation masks and images, and is trained to operate on segmentations with different levels of detail, from simple silhouette blobs to full-body segmentations with detailed tags for individual garments. The results from applying the method can be seen in Fig. 9

State-of-the-art human body pose estimation methods relying on the fitting of 3D avatars to humans in frames can also serve to impart visual privacy. These mostly build on the Skinned Multi-Person Linear (SMPL) model [84]. SMPL is created to be fast and to operate with standard rendering engines, producing realistic looking avatars that do not produce the unnatural joint deformation effects commonly seen in other avatar fitting schemes. Blend shapes are represented in the scheme as a vector of concatenated vertex offsets. An artist created mesh of 6890 vertices and 23 joints is obtained. The mesh used for the rendering uses the same topology for men and women. The model also comes with other options such as a spatially variant resolution and a skeletal rig. SMPL is, however, a function solely of joint angles and face parameters. It does not consider some bodily actions such as breathing, facial motions or actions, muscle tension, or changes independent of skeletal joint angles and overall shape. SMPL also does not generalise well to account for all the variations found in people's body shapes, and produces unnatural deformations of blend shapes.

A recent example of a method devised using SMPL is Frankmocap [123], capable of both hand and body capture and replacement in real time. Since the pose of hands is harder to estimate than most parts of the body as they are small, the authors also built a custom 3D monocular hand capture method that uses the hand part of the SMPL model to achieve this task. One drawback of this scheme is that garments are not modelled for the avatar.

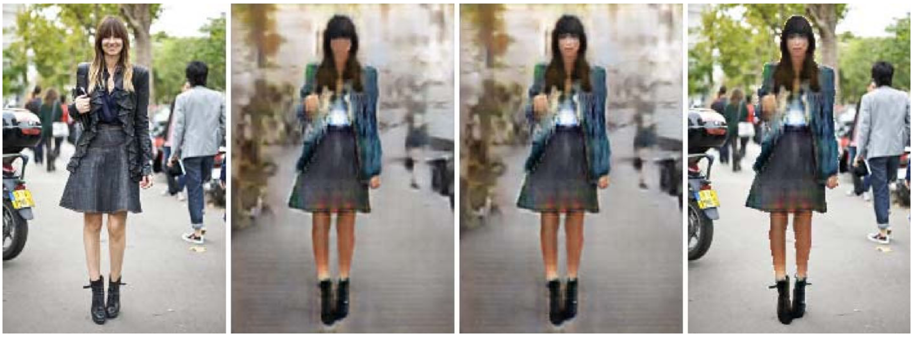


Fig. 9 Results from using the full-body de-identification method (reprinted from [22]). From left to right are the outputs of various stages of the pipeline: The original image, a de-identified full-body image, the result after addition of a synthetic face, and after blending into the original background

Most advancements in avatar fitting have focussed solely on returning the SMPL parameters which stand in for the 3D body meshes, ignoring the garments worn. Some advances over the standard SMPL model have focused on modelling garments worn by the person. One such recent model is the SMPLicit [36]. This approach specifically models garment topologies on top of the SMPL model. Garments are predicted through the use of a semantically interpretable latent vector. The objective is to then be able to influence the looks of garments by manipulating this interpretable vector. SMPL-X [107] is another extension of the SMPL model, which generates avatars with fully articulated hands and facial expressions. The Sparse Trained Articulated Human Body Regressor (STAR [103]) improves the SMPL by producing more realistic deformations, and with only 20% of the model parameters required for the SMPL. The model also generalises better to account for the variations in the body shapes of the human population.

The creation of a dense correspondence between images and surface-based representations is another active area of research. Some works have utilised depth images [116, 137, 148], and others have employed RGB images to correspond to objects [20, 48, 171].

One noteworthy example using RGB images is the DensePose [97] framework. The authors set about annotating persons appearing in the COCO dataset [80] through the use of human annotators utilising a novel annotation pipeline, thereby creating a 'DensePose-COCO' dataset. They then set about training deep neural networks to learn the associations between RGB image pixels and the surface points of human bodies. The authors use a Mask-RCNN segmentation model [57] and couple it with a Dense regression system (DenseReg) [5] for the task. DensePose has also been successfully employed in protecting visual privacy in AAL settings. Climent-Pérez and Florez-Revuelta [34] create various privacy preserving visualisations using a union of masks obtained from DensePose and a Mask-RCNN model, along with the original RGB image used as input for the models (See Figs. 12 and 13).

Object/People Removal Various algorithms are available to remove privacy-sensitive objects and individuals from frames, which are referred to as total body substitution methods. After removal, a gap is left, which is then filled with a generated background using inpainting methods to create a coherent image. Image inpainting methods usually rely on information from surrounding areas to fill in the gaps. In video inpainting, information from previous frames can be used to inpaint subsequent frames, but temporal consistency between frames must be maintained, which is referred to as background modelling in the literature.

There are various techniques that have been created for image inpainting. Paunwala [61] classifies these into *partial differential equation-based methods*, *exemplar-based methods*, and *hybrid methods*. The authors introduce a category of *deep learning based inpainting schemes*, which have been increasingly used since the creation of generative adversarial networks.

PDE-inspired algorithms - Algorithms in this category utilise geometric information to do inpainting of the gaps, by looking at the image inpainting process as one of heat diffusion. Several types of PDE-inspired algorithms exist, notably anisotropic diffusion [110], diffusion-based image inpainting [14], and total variational inpainting [125].

Exemplar-based methods - Initially created by Criminisi et al. [38], these algorithms gather information from nearby regions or a database of images to fill in missing areas. Texture synthesis is a subset of this category, where synthetic textures from one part of an image are used to fill missing regions in another part of the image. Texture synthesis is slower than other patch-based methods, as it performs inpainting on a pixel-by-pixel basis.

Hybrid Approaches - Hybrid approaches combine the advantages of both PDE-based methods and exemplar-based methods to create better inpainting results. Examples include the approach by Bertalmio et al. [15], and the wavelet decomposition-based methods by Zhang and Dai [167] and Cho and Bui [29]

'Deep Learning'-based methods - Although their use in the scenario of object removal is scarce, deep learning models have increasingly been used for image inpainting tasks. These typically make use of generative adversarial networks, to create realistic looking inpainted results [157, 159]. Similar approaches which have also utilised deep learning to do video inpainting include [26, 66, 77, 102, 163].

4.1.4 Gait anonymisation

Gait is a unique biomarker used to identify individuals [13, 17, 82, 86, 147, 170], and gait anonymisation is a newer area of research. deeper treatment of the subject of gait recognition can be seen in the work by Wan et al. [144]. Video surveillance anonymisation tools often use filters like pixelation and blurring, , and then assume the gait to be anonymised in the process [3]. However, these approaches result in an artificial-looking video and are vulnerable to targeted attacks.

The approach proposed by Tieu et al. [140] suggests using deep neural networks to generate an anonymised gait. The algorithm inputs the original gait from the visual feed along with a specially created 'noise gait' to a convolutional neural network, which outputs an anonymising contour vector. The contour vector is processed to produce the anonymised gait, which is then placed back into the original scene.

With the rise of generative adversarial models capable of state-of-the-art generative capabilities, newer literature has focussed on leveraging their power to produce anonymised gaits. Tieu et al. [141] create spatio-temporal generative models that can obfuscate gaits present in videos, creating natural-looking sequences. This architecture makes use of one generator and two discriminators. The generator accepts the original gait and random noise to generate anonymised gaits. The first discriminator is a spatial discriminator which accepts a contour vector extracted from frames of the gait, and tries to distinguish the shape of real gaits from generated gaits at each frame. The results improve the naturalness of the shape of the generated gait. The second discriminator is a temporal discriminator, which distinguishes between the temporal continuity of the real gait and a generated gait. This determines whether the generated gait moves smoothly. A contour sequence is fed through a long short-term memory network [59], the outputs of nodes of which are concatenated to form one input vector for the network. A binary anonymised gait is obtained through the generation process, which is then colourised to merge into the original background.

This process is known to work only on high-quality silhouette inputs, and fails notably with low-quality silhouettes. Tieu et al. (2019) [139] expand on this work by creating a colourisation network, in addition to a different STGAN-based generator-discriminator architecture defined in [141]. Through this approach, the authors were able to provide gait anonymisation for low-quality silhouettes as well (Fig. 10).

4.2 Machine obfuscation: Targetting algorithms

This review classifies algorithms that aim to protect user privacy from machine learning algorithms as machine obfuscation techniques. These techniques employ generative models, specifically GANs, and are commonly referred to as attacks since they aim to attack the validity of deep learning models used for automated analysis.

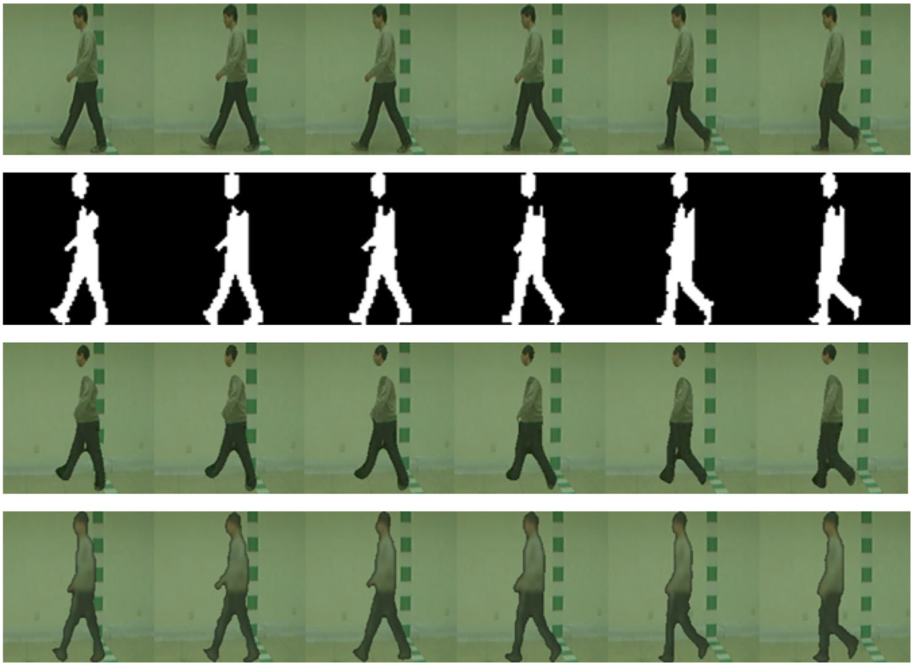


Fig. 10 From top to bottom, the original gait, low-quality silhouette of the gait, results from applying STGAN [141] and the results from applying the improved method proposed in [139] (Reprinted from [139])

Machine obfuscation attacks can be split into two different types - *Poisoning attacks* and *Evasion attacks* [131]. Their objective is to create imperceptible changes in images that cause misclassification in machine recognition models. These changes should also be perceptually pleasing to evade humans from detecting their presence, and to be useful for sharing on popular photo sharing applications.

4.2.1 Poisoning attacks

Poisoning attacks are a type of machine obfuscation attack that aims to disrupt machine learning models by introducing specific ’poisoned’ images during the training process. These attacks can be categorised into ’clean label’ attacks and ’model corruption’ attacks.

Clean Label Attacks Clean label attacks involve the creation of adversarial noise to make machine learning models misclassify a specific image or set of images containing the person4 [127, 173]. The adversarial noise is created in a specific way to alter the feature space used by the models for recognition, causing them to classify unaltered images incorrectly during testing.

Most clean label attacks work on the possible misclassification of a single preselected image that is introduced, although exceptions do exist. Shan et al. [131] developed *Fawkes*, which is one such approach through which users can produce ‘cloaked’ images of themselves through the addition of imperceptible adversarial noise. These then cause machine learning models trained on the cloaked images to misclassify normal images of the user.

Model Corruption Attacks A model corruption attack aims to distort the feature space of images in such a way that upon using the altered images, it reduces the overall accuracy of the trained model [132]. The objective of model corruption attacks are to prevent unauthorised data collection and model training. One disadvantage of these types of attacks is that they are more easily detectable because the presence of such an attack would be readily reflected in the drop in overall model accuracy seen.

4.2.2 Evasion attacks

Evasion attacks create images that are difficult for image recognition systems to identify. These commonly rely on the creation of adversarial examples through the use of physical artefacts, which upon being shown to cameras during capture increases the chances of the subject being misidentified. Prominent examples of this sort include wearables like a specially crafted pair of spectacles [129], adversarial stickers [68] (See Fig. 11), or adversarial patches [23, 138, 152] that increase the chances of misidentification.

The downside of these types of attacks is that these are obvious to a human observer of the footage. Techniques that use adversarial models to alter faces to avoid detection can also be classified under evasion attacks, while in this survey, these are moved to perceptual obfuscation techniques as they alter the appearance of the person in obvious ways, and are usually primarily aimed at human adversaries. The lines are blurred, however, as they can be created to fool machine recognition systems as well.

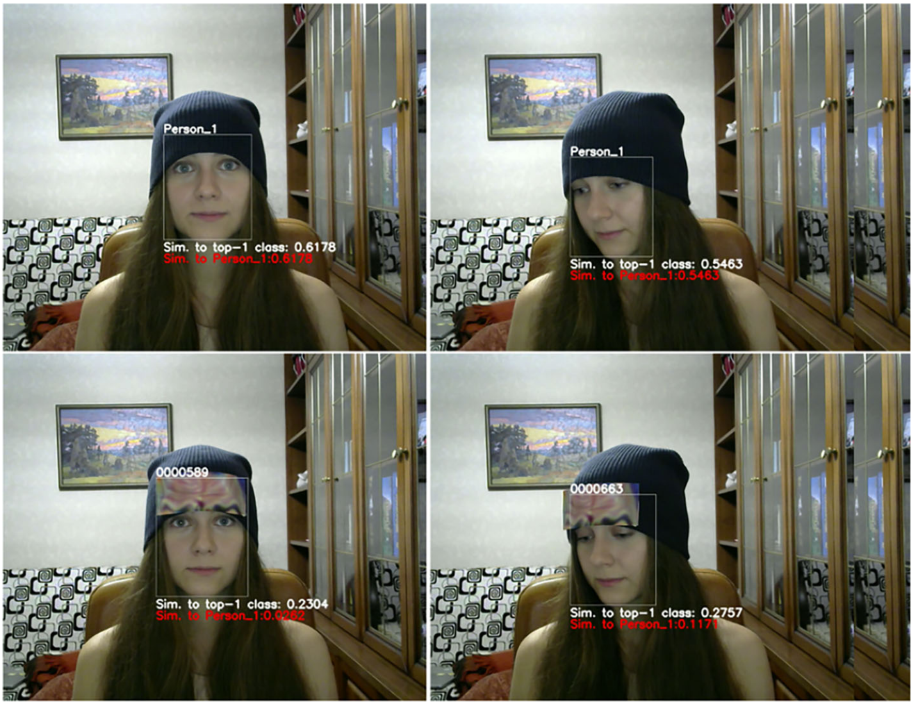


Fig. 11 Results from using the AdvHat method described in [68]. The top row shows the images without the use of the adversarial sticker, and the second row shows the results after the sticker (printed on the hat) is used. As the results printed on the images show, use of the sticker causes misclassification (Reprinted from [68])

Evasion attacks are not to be confused with intervention methods. While evasion attacks prevent machine learning algorithms from recognition through the use of hardware, these do not prevent the collection of the data itself. Intervention methods, on the other hand, use specialised hardware to interfere during the data collection stage, preventing private data from ever being sent to the subsequent stages of the pipeline.

4.3 Privacy protecting pipelines

Research has also been conducted to create end-to-end pipelines that aim to preserve visual privacy through the combination of various techniques in visual privacy preservation. One notable example is by Climent-Pérez and Florez-Revuelta [34] (see Fig. 12). Here, the authors accept an RGB image as input, creating with it a Densepose [97] and Mask R-CNN [57] masks. Using these representations along with a background model created after using a union of the two masks as input, the authors produce five privacy preserving representations, namely the avatar, blurring, invisibility, embossing, and pixelation. These preserve privacy to differing extents, and the footage can be broadcast to users depending on access privileges. The results from the application of the pipeline on a frame from the Toyota Smarthomes dataset [40] can be seen in Fig. 13.

5 Privacy by design

Privacy by Design is a systems design concept defined by Cavoukian et al. [24], which advances the view that privacy cannot be ensured through compliance with regulatory frameworks, and must instead stem from an organisation's default mode of operation. The concept is accomplished through adhering to the following 7 principles:

1. Proactive not Reactive; Preventative not Remedial - Systems ought to be created that prevent privacy invasive events before they occur.

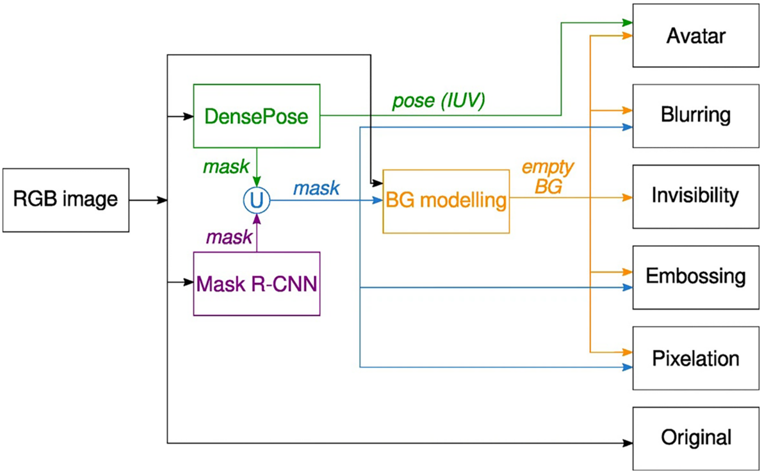


Fig. 12 An illustration of a pipeline that accepts RGB images and applies various privacy preserving filters according to access privileges (reprinted from [34])

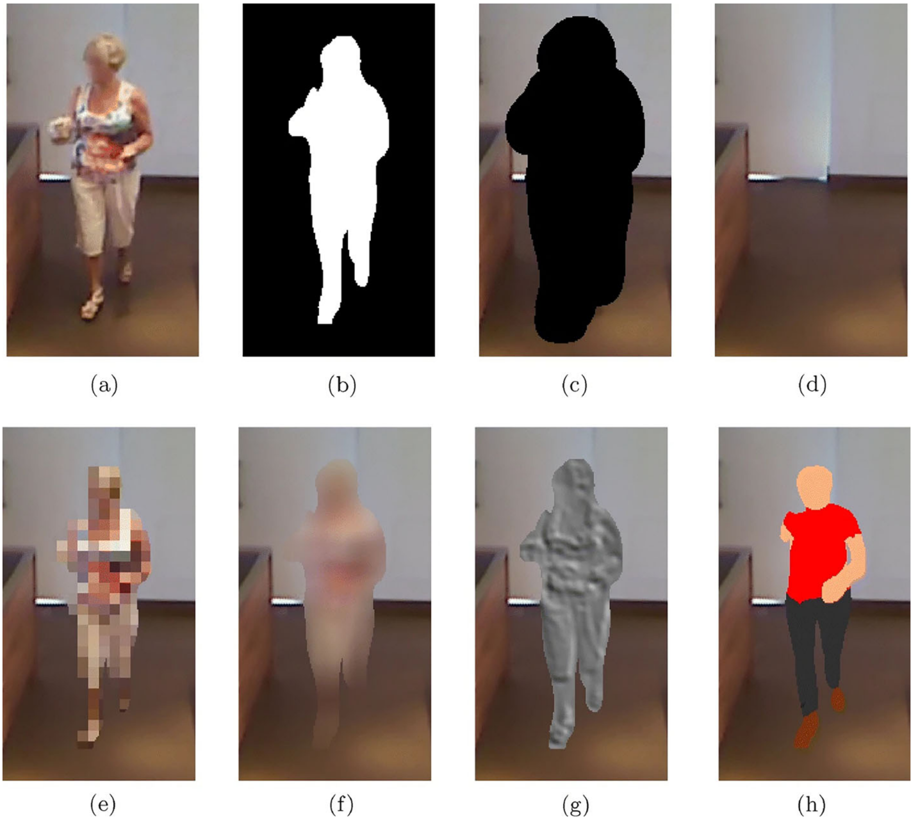


Fig. 13 Example frame from the Toyota Smarthome dataset, within the workflow of the method proposed in [34]. (a) shows the original frame; (b) shows the union mask obtained for this frame; (c) shows the background image fed to the background updating scheme; (d) through (h) show results after applying the different filters (in order - invisibility, pixelation, blurring, embossing, and avatar). (Reprinted from [34])

2. Privacy as the Default Setting - In any business practice or IT system, an individual's privacy is automatically protected even if they perform no actions.
3. Privacy Embedded into Design - Privacy is embedded into the core design and architecture of IT systems, and into the surrounding business practices.
4. Full Functionality (Positive-Sum, not Zero-Sum) - False dichotomies, such as that of privacy vs security, is avoided. It is the goal of the system to accommodate the legitimate interests of both the user and the service provider.
5. End-to-End Security (Full Lifecycle Protection) - The system architecture ensures that strong security measures which are essential to ensuring privacy are established, extending through the entire lifecycle of the data.
6. Visibility and Transparency (Keep it Open) - Components of the system are created in a way as to be visible and transparent to users and data providers. This ensures verification of the objective that the business is operating according to its stated promises.
7. Respect for User Privacy (Keep it User-Centric) - The system is architected in such a way that the interests of the individual is upheld. This is done through providing strong privacy defaults, appropriate notice, and user-friendly options.

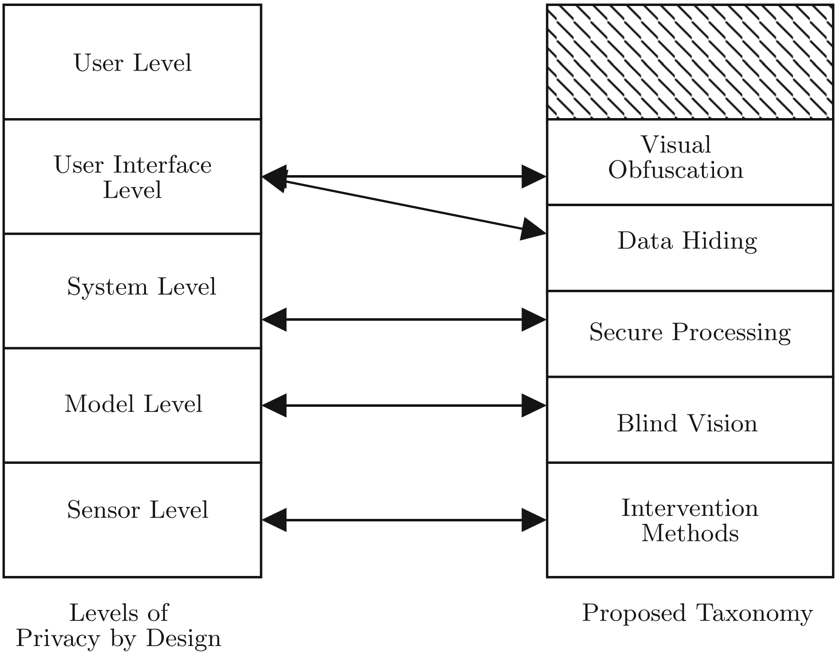


Fig. 14 Connection between the levels of Privacy by Design [92] and visual privacy protection methods

Based on different design elements present in lifelogging technologies, Mihailidis & Colonna [92] created a classification schema that separates privacy by design into levels. According to the schema, components in a pipeline acting at each level must be compliant with existing data protection rules for the system to adhere to the notion of privacy by design. The most basic of these is at the *sensor level*. Moving upwards in scope, they can be specified as *model level*, *system level*, *user interface level*, and at the most abstract, privacy at the *user level*. For clarity, this is connected to the taxonomy of visual privacy preservation methods presented in Section 3. The correspondence between both taxonomies can be seen in Fig. 14, and is further explained in subsequent subsections.

5.1 Sensor level

Sensor level privacy preservation techniques prevent the capture of sensitive data in visual feeds using various software and hardware implements. These mechanisms can prevent the capture of sensitive content in the first place by the camera. This can also be implemented at the software level, as a filter to clear the captured images of protected content before the images are stored to disk. Intervention methods (Section 3.1) can be grouped under the umbrella of intervention methods, as these intervene during the data collection phase to protect the privacy of users and environments.

5.2 Model level

To observe model level privacy, methods are created that preserve privacy for users while at the same time enabling models to infer information from data. Also termed as privacy-preserving data mining (PPDM), these techniques aim to create privacy in such a way that unintended third parties cannot make sense out of protected attributes in data, while also removing sensitive knowledge that has been mined from the data.

Since blind vision methods (see Section 3.2) help in processing the data securely, these schemes can be considered under model level methods, as they contribute to the model level privacy of the pipeline. Since blind vision techniques also allow inferring from data while preserving privacy, it could also be noted as contributing to the system level privacy of a pipeline. Another example of a technique that contributes to the model level privacy of a pipeline is *federated learning* [69], a technique used for the private training of machine learning models.

5.3 System level

For system level privacy preservation, techniques need to be developed so that the data used in the pipeline becomes secure, and that user consent for the use of the data in the pipeline is traceable. Traceability requires two components [92]. The first is that personal data can be traced to when user consent for its usage was recorded. Secondly, the flow of the data to various sources should also be traceable. This is essential because withdrawal of consent is an important facet of privacy laws like the GDPR [37]; upon withdrawal of consent, actions have to be taken by the authorised administrator to comply with the request. For this reason, system level privacy is not only an essential concept, but also an arguably overlooked one that is critical to managing the legal requirements surrounding the use of data in machine learning projects.

Additionally, an important facet to system level privacy is the creation of secure databases that protect against information breaches. State-of-the-art techniques like homomorphic encryption allow for machine learning models to infer from the data privately. Boulemtafes et al. [18] provide a more in-depth treatment on the subject of privacy preserving deep learning. Techniques under secure processing (see Section 3.3) can be considered as contributing to the system level privacy in a system that enforces privacy by design, as for system level privacy, it is required that the data remains secure inside the pipeline. Secure processing techniques assist the pipeline in this regard. It is, however, unclear whether techniques categorised as secure processing also fall under model level privacy preservation schemes as they do allow models to infer information from the data, while also preserving user privacy.

5.4 User interface level

Privacy provided at the user interface level prevents the exposure of privacy-sensitive images or parts of images in various scenarios. Under the classification of privacy preservation methods proposed in this review, techniques under the category of visual obfuscation (Section 4) can be mentioned as adding to user interface level privacy of pipelines. Data hiding methods also contribute to the user interface level privacy of a pipeline because, according to definition, these act to restrict the exposure of private visual information within the image, differing from the former category by the strategy with which the hiding of sensitive information is performed.

5.5 User level

User level privacy measures empower users by helping them manage their data. These also help users understand the privacy risks involved with the sharing of their data, and also give them mechanisms through which they can control the disclosure of their data. User level privacy is ensured through various educative measures, such as through the use of clear and easy to understand privacy disclosures and agreements. The creation of transparent dashboards through which users can control their data usage is another measure. The regular collection, analysis, and incorporation of user feedback into the pipeline is also a measure to incorporate user level privacy into the pipeline.

6 Performance evaluation

For the case of visual obfuscation techniques, the type of performance evaluation used depends on the adversary. In systems to perform machine obfuscation, image quality metrics [108] are popularly used. Since the objective of machine obfuscation techniques is to create images that are perceptually similar to the original, image quality metrics are employed to ascertain the (dis)-similarity of the two images. As for perceptual obfuscation, where the adversary is a human observer, a more empirical evaluation is often used. Human feedback is commonly sought for this purpose through the deployment of targeted surveys. Machine recognition systems are also often employed in the case of facial de-identification tasks.

The following subsections deal with the most commonly used metrics in the literature. Popular datasets used during evaluation are also explained.

6.1 Technical privacy metrics

There are different types of privacy metrics that have been employed for measuring the performance of privacy preservation methods. Wagner and Eckhoff [143] refer to eight categories of metrics used to measure privacy in various contexts. We classify technical privacy metrics into two strains: those which measure an adversary's estimates to gauge how private a dataset is, and those metrics which gauge privacy according to a variable independent of adversarial estimates.

6.1.1 Indistinguishability metrics

Indistinguishability metrics measure whether an adversary can distinguish between two outcomes of a *privacy mechanism*, and gather information about the dataset's composition from the differences between the outcomes. One commonly used indistinguishability metric is differential privacy [42], which is nowadays extensively used in the securing of databases.

Dwork et al. [42] define differential privacy as a promise made by a data holder/curator to a data subject. The promise is defined as follows:

> You will not be affected, adversely or otherwise, by allowing your data to be used in any study or analysis, regardless of what other studies, datasets or information sources are available.

When differential privacy is implemented for a specific database, it ensures protection against differencing attacks that can reveal information about a specific user in the database.

By assuring differential privacy, the designer is ensuring that upon removal of a record containing a specific user's information, queries executed against the database do not produce a different output from when the same query was executed on the version of the database with the user's record present.

In obfuscation tasks, a commonly used metric is the accuracy of machine recognition systems, which [143] classify into error-based metrics. This looks at how often a machine recognition system like Amazon Rekognition engine [7] can identify subjects in images that have been visually obfuscated. It is usually the case that a simple tally is used as the metric, counting the number of times the subject of interest is detected.

Of particular interest to the concept of perceptual obfuscation are metrics that are independent of adversary. These are solely dependent on observable or measurable differences between two data points or sets of data.

6.1.2 Data similarity metrics

One such category proposed by Wagner and Eckhoff [143] is data similarity. These include metrics that measure the similarity within a dataset through the formation of equivalence classes, or between two sets of data. Some common types include k-anonymity [135] and its variants, namely l-diversity [85] and t-closeness [79].

k-Anonymity - k-anonymity is one of the most widely used metrics to evaluate privacy and defines itself regarding quasi-identifiers inside a database. *Quasi-identifiers* are attributes that can be taken together to identify an individual. Examples of this include the postcode or the birthdate in a personal database. In the case of a facial features database, this can refer to features like glasses, shapes of facial features like noses and the face itself. The metric is defined as follows -

> A database is private if each record, k, in the database is indistinguishable from at least $k - 1$ records in the database with quasi-identifiers.

Upon satisfaction of k-anonymity, a person's record can only be chosen from a database with a probability of $1/k$.

l-Diversity - Proposed to address the limitations of k-anonymity, l-diversity is defined as follows -

> For the equivalence class representing a set of records with the same values for quasi-identifiers, it should have at least l 'well-represented' values for the sensitive attribute.

'Well represented' values are commonly defined as whether an equivalence class has l distinct values for the sensitive attribute, without considering the frequency of values.

t-Closeness - To prevent attacks on privacy by adversaries with knowledge of global distribution of sensitive attributes inside a database, Li et al. [79] devised the measure of t-closeness. This measure updates k-anonymity as follows.

> The distribution of sensitive values, S_E, in an equivalence class E shall be close to its distribution, S inside the entire database.

6.1.3 Machine recognition scores

Particularly in the context of facial de-identification, machine recognition is commonly employed as a metric to gauge the effectiveness of obfuscation methods. Machine recognition algorithms work by scoring how often a trained recognition algorithm can identify a de-identified subject. In the context of facial recognition, the most commonly used API services are the Google Vision API [50], Microsoft Azure Face API [91], Amazon Rekognition [7] and Face++ [89]. Simple scoring systems are mostly used for these metrics, often a simple tally of the recognised attribute in the case of attribute recognition, or the recognised activity category in the case of an activity recognition task on obfuscated frames.

For gait obfuscation, custom metrics are usually employed. Tieu et al. [139] craft custom automatic evaluation strategies that seek to measure the difference between a standard gait and a generated one. They employ a *frame score* and a *video score* to measure the differences. The *frame score* measures the degree to which the shape of the object in the frame looks human. For this, they employed a pretrained YOLO model [118] that detects and classifies objects in an image. The authors compute the probability that a person in a frame belongs to the 'person' class. The *video score* measures the degree to which the gait in the video looks like a humanoid walking. A pretrained ResNeXT-101 [154] was used for this purpose, which classifies actions in the video. The probability that the action in the video corresponds to the 'video' class is measured and reported for this score.

6.1.4 Human recognition scores

To evaluate the effectiveness of privacy preservation methods, researchers often employ human feedback alongside machine recognition algorithms. Questionnaires are commonly used to gather targeted feedback, consisting of a set of questions with pre-defined response options or free-form filling sections. Online services like Mechanical Turk [6] and Prolific [117] are often used to gather responses from targeted audiences.

Çiftçi et al. [31] and Padilla-López et al. [105] both used targeted questionnaires to gather feedback on the efficacy of privacy preservation methods. Çiftçi et al. focused on face recognition and activity recognition tasks after image filtering using the 'false colors' method, while Padilla-López et al. used various perceptual privacy preservation methods, including blurring, pixelation, embossing, silhouette, skeleton, and an avatar, and asked participants to identify visual attributes of obfuscated subjects such as hair and skin colour and facial expressions.

6.2 User acceptance studies

The acceptance of privacy preservation technology is an important concept that is often examined in studies. Wilkowska et al. [149] conducted a study that compared the perspectives of German and Turkish participants on lifelogging technologies and the visual obfuscation techniques used on their feeds. The study included representative images obfuscated in five different ways, ranging from low to high levels of privacy protection. Participants were asked to provide feedback on the images and answer questions about their preferences for different visualization modes. The study aimed to determine whether cultural influences affect perceptions of privacy preservation technologies and which visualisation mode is the most preferred among participants.

6.3 Datasets

The research community has employed several datasets for the task of measuring visual privacy. The most commonly used datasets consist of RGB images or video streams. It is also popular to curate subsets of these datasets for various targetted experiments. In this section, various datasets that are used for validating the efficacy of privacy preservation methods are listed, along with details of their composition and the papers that use these sets for experimentation[6].

For the case of facial anonymisation, some popular datasets used are the following:

Facial Recognition Technology (FERET) dataset [112] - Containing 14,126 facial stills of 1,199 people, FERET is a publicly available dataset from the US Army. For every facial image, the coordinates for the centres of the eyes and tip of the nose are provided. Examples of privacy preservation methods using FERET for validation include [31].

People in photo albums (PIPA) dataset [162] - is a dataset consisting of over 6,000 images of around 2,000 persons, with only half of the images being of persons from a frontal frame of reference. This creates a challenging task, as recognition systems are mostly trained on frontal imagery. The dataset contains people in a good variety of poses, activities, and scenery. One example of a method validated using PIPA is the method proposed by Sun et al. [134].

AT&T Database of Faces [8] -The AT&T database of faces contains 400 grayscale images of 40 individuals of resolution 92×112. The dataset contains 10 images of each individual, taken under a variety of conditions including varied lighting, different expressions, and different facial details. One example of a privacy protection scheme that uses this dataset for testing is that by Fan [46].

Facescrub [99] is a large dataset consisting of slightly more than 65,000 facial images of 530 celebrities collected from online publications. Only URLs are distributed for copyright reasons[7]. Shan et al. [131] proposes a scheme that makes use of this dataset while testing.

PubFig images dataset [74] - This is a dataset of images of public figures (celebrities and politicians) obtained from the internet. The dataset consists of around 60,000 images, with around 300 images per individual. Shan et al. [131] and Sharif et al. [129] are notable examples of papers using the PubFig images dataset.

CelebFaces Attributes (CelebA) dataset [81] - Used for facial attribute estimation in the process of training facial de-identification methods, this dataset contains 202,599 images and 10,177 identities of celebrities. Each image has around 40 boolean attribute labels. Li and Lin [78] is notable for making use of the CelebA dataset for testing.

Labeled Faces in the Wild (LFW) dataset [60] is another dataset containing ≈13,000 images of faces collected from the web. 1,680 individuals in the set have two or more distinct images of themselves represented in the dataset. Several alternative datasets of faces in the wild have also been proposed, some notable ones being *Fine-grained LFW* [41], *LFWGender* [63], and *LFW3D*. Zhang et al. [166] proposes a method that is notable for using the LFW dataset during testing.

Generic image recognition and object detection datasets are often used in validating the efficacy of privacy preservation schemes, mostly in the case of machine obfuscation schemes. Some commonly used ones are the following.

[6] It is to be noted that a number of these datasets presented are aimed at measuring the efficacy of machine obfuscation methods

[7] The original dataset contains URLs to 100000 images, with a number of URLs broken due to missing media.

Modified NIST (MNIST) [76] - MNIST is an extremely popular dataset consisting of images of handwritten digits collected from census bureau employees and high school students in the USA. The entire dataset consists of 70,000 images in total. Abadi et al. [1] proposes a scheme that is benchmarked using the MNIST dataset.

CIFAR-10 [73] - Another popular dataset is CIFAR-10, consisting of of a total of 60,000 images of size 32×32. Labels of the dataset consists of either animals (e.g., cats, dogs etc.), or vehicles (e.g., planes, cars, etc.). Abadi et al. [1] proposes a scheme that uses the CIFAR-10 dataset for validation.

YouTube 8M video dataset [2] - The YouTube 8M dataset is a video dataset composed of around 8million videos, approximately 500,000 hours of content, annotated in a multi-label format with 4,800 distinct labels. These labels are machine generated and human curated, with 1.9 billion video frame-level annotations. The entities in videos are also categorised, with some categories represented in the dataset being 'Arts & Entertainment', 'Games', 'People & Society', and 'Books & Literature'. Wong et al. [150] proposed a privacy preservation scheme that notably uses the YouTube 8M video dataset for testing.

In the setting of gait anonymisation, the **CASIA-B gait dataset** [161] is one that is arguably the most popular. This dataset contains 124 individuals in total, with 110 sequences (10 sequences each for each of 11 viewing angles from 0° to 180°). Tieu et al. [140] create a gait anonymisation scheme that uses the CASIA-B dataset for validation.

In the context of full-body de-identification, the following datasets are commonly used:

Clothing Co-Parsing dataset [156] - This dataset consists of 2,098 high resolution, street fashion images. Pixel-level segmentations of individual garments and skin are available for ≈1000 of the images. 59 segmentation tags defining various garment types, e.g., blazer, cardigan, sweatshirt etc., are used in this dataset. Brkić et al. [22] makes use of the clothing co-parsing dataset to test their full-body privacy preservation scheme.

Human3.6M dataset [62] - This dataset consists of 3.6 million video frames of actors performing actions in a controlled setting. 3D joint positions, the laser scans of the actors, and their corresponding 3D poses are available as annotations. The dataset utilises a static camera angle for the recordings. Brkić et al. [21] proposed a privacy protection scheme that utilised this dataset for testing purposes.

Toyota Smarthomes dataset [40] - This is a dataset of slightly more than 16,000 video clips, of 31 activity classes performed by 18 seniors in a smart home setting. The dataset is labelled with both coarse and fine-grained labels and contains heavy class imbalances, high intra-class variation, simple as well as composite activities, and activities with similar motion and of variable duration. Climent-Pérez and Florez-Revuelta [34] use the Toyota Smarthomes dataset to validate their privacy preservation scheme.

NTU RGB+D dataset [128] - Containing 60 different action classes including daily, interaction-based, and health-related actions, this is a large-scale dataset for RGB+D human action recognition, containing greater than 56,000 samples and 4,000,000 frames, collected from 40 distinct subjects. Wang et al. [145] use this dataset to test the efficacy of their privacy preserving action recognition method. An extended version of this dataset was published by J. Liu et al. [83].

7 Conclusion and future directions

This work reviews the state of the art in visual privacy preservation methods. A low-level taxonomy of visual privacy preservation methods is introduced, and the categories under the taxonomies were subsequently explored. Special attention was given to visual obfuscation methods, these being of most relevance to AAL applications. The taxonomy is then connected to a high-level classification scheme of the levels of privacy by design.

Visual obfuscation methods are categorised into two categories in this review based on the targets from whom the algorithms are seeking to hide private information: *perceptual obfuscation* and *machine obfuscation* methods. Perceptual obfuscation seeks to perceptually alter images in ways that unauthorised human observers who view the visual feed are thwarted. By contrast, machine obfuscation methods try to hide privacy-sensitive elements from machine learning algorithms. These seek to alter the feature space of images in ways that machine recognition systems are thwarted, while also perceptually changing the visuals to the least possible extent.

As these are two different directions of research, algorithms can also be built such that they perform both machine and perceptual obfuscation. The capability of performing reversible transformations through secure pipelines is another promising direction for research. This is useful in the case when reversibility is required, such as for an arbiter (a judge, a doctor, etc.) to view unedited footage to obtain full information about a specific scenario.

7.1 Technical questions

In the context of visual privacy preservation, numerous technical challenges remain to be addressed. One major challenge is to create real-time pipelines that impart privacy. Most of the existing state-of-the-art methods rely on computationally intensive pipelines. To create real-time privacy protection, methods have to be made more lightweight.

There are also some widely used cameras that are arguable not sufficiently researched in literature from the perspective of privacy preservation. Egocentric/wearable cameras have been touted as a method to protect identity, but this poses problems if the environment contains objects (e.g. mirrors) that reveals one's personal attributes. This also introduces issues when bystanders come into the visual field; bystanders would typically not have given permission for them to be captured on camera. This poses ethical and legal challenges, in addition to technical ones, especially when egocentric cameras are utilised [53].

Omnidirectional cameras have fisheye lenses that provide the user with a mostly non-occluded view of an entire room based on its placement (usually on the ceiling). However, object detection algorithms have not typically been trained to detect on images from distorted lenses. Privacy preservation algorithms that rely on detection as part of the pipeline are therefore summarily excluded from use on these streams. Other non-standard cameras (thermal, infrared) also face similar problems. Therefore, the authors call for more research to create privacy preserving algorithms that work on non-standard cameras.

Some identifiers have also been arguably addressed less in the literature. Gait is one such example, and to the authors' knowledge, only a few papers have attempted to create gait anonymisation algorithms. Environmental identifiers are also another.

7.1.1 Privacy of the environment

Although included in this review as a sub-category of perceptual obfuscation, literature searches show that environmental privacy is an under-researched area, but arguably one that is critical to the ensuring of visual privacy. Most of the existing methods that impart privacy target people and their visible attributes. However, objects in the environment are also required to be obfuscated if the identity of the person is to be protected. Objects like credit cards and address labels create privacy risks if not obfuscated. Cartooning is one type of method that can provide environmental privacy, as it can replace objects in the environment with privacy protected elements.

Some methods do provide environmental privacy as a side effect of their use. As an example, consider a blurring filter. When a blurring filter is applied to an image as a whole, textural information is lost, which might lead to smaller privacy-sensitive objects such as credit cards (and specifically the numbers printed on them) being obfuscated. Depending on the parameters used for the blurring, larger objects in the environment might still contribute to privacy leakages.

Commercial products which aim to detect and obfuscate personally identifiable text that occurs in images do exist [133]. These include phone numbers, email addresses, links and URLs, and social media accounts that occur as visible text inside images.

7.2 Social and legal aspects of privacy

There is also the urgent need to understand the methods from social and legal perspectives. There needs to be studies to ascertain the level of acceptance of different perceptual obfuscation methods among the monitored subjects. It is also unclear as to the extent of the acceptability of reversible transformations for the subjects being monitored. Although there are several methods that reconstruct obfuscated images, the acceptability of reconstructed images through a reverse transformation pipeline that contains embedded stochasticity is an especially interesting one to study. In a setting such as that of a court or in forensics, as reconstruction is an imperfect process, there is always the possibility of information loss. It is unclear if such images are viable for presentation in such circumstances. There also needs to be more studies that detail the relationship between human perception and the metrics that are used to measure perceptual obfuscation. Although there are some studies that do this, there is a distinct need for more wide-ranging targeted studies to be performed.

The concept of a ‘privacy paradox’ also needs to be investigated. It is a known phenomenon that people act in contrast to what they believe their privacy preferences are, especially when it comes to their online behaviour [12]. Users claim to be concerned about their online privacy, but they do little to protect their personal data. If this is also the case for visual data like that used in AAL applications, then the gathering of subjective data about user preferences through a medium such as questionnaires should be called into question. It could mean that better ways of gauging preferences should be created and deployed. It could also mean that existing studies that gauge privacy preferences ought to be re-evaluated.

Acknowledgements This work is part of the visuAAL project on Privacy-Aware and Acceptable Video-Based Technologies and Services for Active and Assisted Living (https://www.visuaal-itn.eu/). This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 861091. The authors would also like to acknowledge the contribution of COST Action CA19121 - GoodBrother, Network on Privacy-Aware Audio- and Video-Based Applications for Active and Assisted Living (https://goodbrother.eu/), supported by COST (European Cooperation in Science and Technology) (https://www.cost.eu/).

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Data Availability Not applicable

Declarations

Competing interests The authors declare they have no competing interests.

References

1. Remagnino P, Foresti GL, Ellis T (2004) Ambient intelligence: A novel paradigm. Springer, New York USA
2. Chaaraoui AA, Climent-Pérez P, Flórez-Revuelta F (2012) A review on vision techniques applied to human behaviour analysis for ambientassisted living. Expert systems with applications 39(12):10873–10888 . https://doi.org/10.1016/j.eswa.2012.03.005
3. Climent-Pérez P, Spinsante S, Mihailidis A, Florez-Revuelta F (2020) A review on video-based active and assisted living technologies for automated lifelogging, vol 139, p 112847 . https://doi.org/10.1016/j.eswa.2019.112847. https://www.sciencedirect.com/science/article/pii/S0957417419305494
4. Clarke R (1999) Internet privacy concerns confirm the case for intervention, vol 42. Association for Computing Machinery, New York, NY USA, pp 60–67. https://doi.org/10.1145/293411.293475
5. Alp Guler R, Trigeorgis G, Antonakos E, Snape P, Zafeiriou S, Kokkinos I (2017) DenseReg: Fully convolutional dense shape regression In-The-Wild. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2017.280
6. Cavoukian A, et al. (2009) Privacy by design: The 7 foundational principles, vol 5. p 12
7. Padilla-Lóópez JR, Chaaraoui AA, Flórez-Revuelta F (2015) Visual privacy protection methods: A Survey, vol 42. pp 4177–4195 . https://doi.org/10.1016/j.eswa.2015.01.041. https://www.sciencedirect.com/science/article/pii/S0957417415000561
8. Ribaric S, Ariyaeeinia A, Pavesic N (2016) De-identification for privacy protection in multimedia content: A Survey, vol 47. pp 131–151 . https://doi.org/10.1016/j.image.2016.05.020. https://www.sciencedirect.com/science/article/pii/S0923596516300856
9. Meden B, Rot P, Terhörst P, Damer N, Kuijper A, Scheirer WJ, Ross A, Peer P, Štruc V (2021) Privacy-enhancing face biometrics: A comprehensive survey, vol 16. pp 4147–4183. https://doi.org/10.1109/TIFS.2021.3096024
10. Perez AJ, Zeadally S, Griffith S (2017) Bystanders' privacy, vol 19. pp 61–65 . https://doi.org/10.1109/MITP.2017.42
11. Jonsson KS, Bergthorsdottir SH (2016) Webcam privacy shield. Google patents. US Patent 9,465,276
12. Barth S, de Jong MDT (2017) The privacy paradox -Investigating discrepancies between expressed privacy concerns and actual online behavior. A Systematic literature review 34:1038–1058. https://doi.org/10.1016/j.tele.2017.04.013 www.sciencedirect.com/science/article/pii/S0736585317302022
13. Haddad WS (2017) Detachable lens shuttering apparatus for use with a portable communication device. Google patents. US Patent 9,571,708
14. Miller K (2020) Electronic device privacy cover. Google patents. US Patent 10,816,878
15. Mitskog TF, Ralston RA (2012) Camera blocker for a device with an integrated camera that uses a thin film organic polymer. Google Patents. US Patent App. 13/477,485
16. Bian S, Wang T, Hiromoto M, Shi Y, Sato T (2020) ENSEI: Efficient secure inference via frequency-domain homomorphic convolution for privacy-preserving visual recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr42600.2020.00942
17. Patel SN, Summet JW, Truong KN (2009) BlindSpot: Creating capture-resistant spaces. Springer London, pp 185–201 . https://doi.org/10.1007/978-1-84882-301-311
18. Harvey A, Knight H (2009) Anti-paparazzi fashion. www.marilynmonrobot.com Accessed: 30 June 2021

19. Zhu S, Zhang C, Zhang X (2017) Automating visual privacy protection using a smart LED. In Proceedings of the 23rd annual international conference on mobile computing and networking. MobiCom '17, Association for computing machinery. New York, NY USA, pp 329-342. https://doi.org/10.1145/3117811.3117820
20. Bristow H, Valmadre J, Lucey S (2015) Dense semantic correspondence where every pixel is a classifier. In Proceedings of the IEEE international conference on computer vision (ICCV). https://doi.org/10.1109/iccv.2015.458
21. Brkić K, Hrkać T (2017) Kalafatić Z. Protecting the privacy of humans in video sequences using a computer vision-based De-identification pipeline 87:41–55. https://doi.org/10.1016/j.eswa.2017.05.067 www.sciencedirect.com/science/article/pii/S0957417417303986
22. Pilu M (2007) Detector for use with data encoding pattern. Google patents. US patent App. 11/491,174
23. Kapadia A, Henderson T, Fielding JJ, Kotz D (2007) Virtual walls: Protecting digital privacy in pervasive environments. In LaMarca A, Langheinrich M, Truong KN (eds) Pervasive computing. Springer Berlin Heidelberg, pp 162–179 . https://doi.org/10.1007/978-3-540-72037-910
24. Avidan S, Butman M (2006) Blind vision. In Leonardis A, Bischof H, Pinz A (eds) Computer vision - ECCV 2006. Springer Berlin Heidelberg, pp 1–13. https://doi.org/10.1007/11744078
25. Chaaraoui AA, Climent-Pérez P, Flórez-Revuelta F (2012) A review on vision techniques applied to human behaviour analysis for ambientassisted living. Expert systems with applications 39(12):10873–10888. https://doi.org/10.1016/j.eswa.2012.03.005
26. Chang Y-L, Liu ZY, Lee K-Y, Hsu W (2019) Free-form video inpainting with 3D gated convolution and temporal PatchGAN. In Proceedings of the IEEE/CVF international conference on computer vision (ICCV). https://doi.org/10.1109/iccv.2019.00916
27. Sadeghi A-R, Schneider T,Wehrenberg I (2010) Efficient privacy-preserving face recognition. In Lee D, Hong S (eds) Information, security and cryptology - ICISC 2009. Springer, Berlin Heidelberg, pp 229–244. https://doi.org/10.1007/978-3-642-14423-316
28. Shashank J, Kowshik P, Srinathan K, Jawahar CV (2008) Private content based image retrieval. In 2008 IEEE conference on computer vision and pattern recognition, pp 1–8 . https://doi.org/10.1109/CVPR.2008.4587388
29. Heitzinger T, Kampel M (2021) IPT: A dataset for identity preserved tracking in closed domains. In 2020 25th international conference on pattern recognition (ICPR), pp 8228–8234. https://doi.org/10.1109/ICPR48806.2021.9412979
30. Al-Obaidi S, Al-Khafaji H, Abhayaratne C (2020) Modeling temporal visual salience for human action recognition enabled visual anonymity preservation, vol 8. pp 213806–213824 . https://doi.org/10.1109/ACCESS.2020.3039740
31. Upmanyu M, Namboodiri AM, Srinathan K, Jawahar CV (2009) Efficient privacy preserving video surveillance. In 2009 IEEE 12th international conference on computer vision. pp 1639–1646 . https://doi.org/10.1109/ICCV.2009.5459370
32. Yonetani R, Naresh Boddeti V, Kitani KM, Sato Y (2017) Privacy-preserving visual learning using doubly permuted homomorphic encryption. In Proceedings of the IEEE international conference on computer vision (ICCV) . https://doi.org/10.1109/iccv.2017.225
33. Bian S, Wang T, Hiromoto M, Shi Y, Sato T (2020) ENSEI: Efficient secure inference via frequency-domain homomorphic convolution for privacy-preserving visual recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR) . https://doi.org/10.1109/cvpr42600.2020.00942
34. Petitcolas FAP, Anderson RJ, Kuhn MG (1999) Information hiding-a survey, vol 87. pp 1062–1078. https://doi.org/10.1109/5.771065
35. Yu X, Babaguchi N (2007) Privacy preserving: Hiding a face in a face. In Asian conference on computer vision. Springer, pp 651–661. https://doi.org/10.1007/978-3-540-76390-164
36. Chen B, Wornell GW (2001) Quantization index modulation methods for digital watermarking and information embedding of multimedia, vol 27. Springer, pp 7–33
37. Ni Z, Shi Y-Q, Ansari N, Su W (2006) Reversible data hiding. IEEE Trans Circ Syst Video Tech 16(3):354–362 . https://doi.org/10.1109/tcsvt.2006.869964
38. Yabuta K, Kitazawa H, Tanaka T (2005) A new concept of security camera monitoring with privacy protection by masking moving objects. In pacific-Rim conference on multimedia. Springer, pp 831–842 . https://doi.org/10.1007/1158177273
39. Zhang W, Cheung S-CS, Chen M (2005) Hiding privacy information in video surveillance system. In IEEE international conference on image processing 2005, vol 3. p 868. https://doi.org/10.1109/icip.2005.1530530. IEEE

40. Das S, Dai R, Koperski M, Minciullo L, Garattoni L, Bremond F, Francesca G (2019) Toyota smarthome: Real-world activities of daily living. In Proceedings of the IEEE/CVF international conference on computer vision (ICCV). https://doi.org/10.1109/iccv.2019.00092
41. Korshunov P, Ebrahimi T (2013) Using face morphing to protect privacy. In 2013 10th IEEE international conference on advanced video and signal based surveillance, pp 208–213 https://doi.org/10.1109/AVSS.2013.6636641
42. Korshunov P, Ebrahimi T (2013) Using warping for privacy protection in video surveillance. In 2013 18th international conference on digital signal processing (DSP), pp 1–6 https://doi.org/10.1109/ICDSP.2013.6622791
43. Erdélyi, Á, Winkler T, Rinner B (2013) Serious fun: Cartooning for privacy protection. In MediaEval
44. Erdélyi Á, Barát T, Valet P, Winkler T, Rinner B (2014) Adaptive cartooning for privacy protection in camera networks. In 2014 11th IEEE international conference on advanced video and signal based surveillance (AVSS), pp 44–49. https://doi.org/10.1109/AVSS.2014.6918642
45. Hassan ET, Hasan R, Shaffer P, Crandall D, Kapadia A (2017) Cartooning for enhanced privacy in lifelogging and streaming videos. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) workshops. https://doi.org/10.1109/cvprw.2017.175
46. Çiftçi S, Akyüz AO, Ebrahimi T (2018) A reliable and reversible image privacy protection based on false colors, vol 20. pp 68–81. https://doi.org/10.1109/TMM.2017.2728479
47. Zhang Z, Cilloni T, Walter C, Fleming C (2021) Multi-scale, class-generic, privacy-preserving video. vol 10. https://doi.org/10.3390/electronics10101172. https://www.mdpi.com/2079-9292/10/10/1172
48. Gaur U, Manjunath BS (2017) Weakly supervised manifold learning for dense semantic object correspondence. In Proceedings of the IEEE international conference on computer vision (ICCV). https://doi.org/10.1109/iccv.2017.192
49. Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2014.81
50. Li T, Lin L (2019) AnonymousNet: Natural face De-identification with measurable privacy. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR) workshops . https://doi.org/10.1109/cvprw.2019.00013
51. Tieu N-DT, Nguyen HH, Nguyen-Son H-Q, Yamagishi J, Echizen I (2017) An approach for gait anonymization using deep learning. In 2017 IEEE workshop on information forensics and security (WIFS), pp 1–6. https://doi.org/10.1109/WIFS.2017.8267657
52. Tieu N-DT, Nguyen HH, Nguyen-Son H-Q, Yamagishi J, Echizen I (2019) Spatio-temporal generative adversarial network for gait anonymization, vol 46. pp 307–319. https://doi.org/10.1016/j.jisa.2019.03.002. https://www.sciencedirect.com/science/article/pii/S2214212618304629
53. Tieu N-DT, Nguyen HH, Fang F, Yamagishi J, Echizen I (2019) An RGB gait anonymization model for low-quality silhouettes. In: 2019 asia-pacific signal and information processing association annual summit and conference (APSIPA ASC), pp 1686–1693 . https://doi.org/10.1109/APSIPAASC47483.2019.9023188
54. Brkic K, Sikiric I, Hrkac T, Kalafatic Z (2017) I know that person: Generative full body and face De-identification of people in images. In 2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW), pp 1319–1328. https://doi.org/10.1109/CVPRW.2017.173
55. Corona E, Pumarola A, Alenya G, Pons-Moll G, Moreno-Noguer F(2021) SMPLicit: Topology-aware generative model for clothed people. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 11875–11885. https://doi.org/10.1109/cvpr46437.2021.01170
56. Rong Y, Shiratori T, Joo H (2021) FrankMocap: A monocular 3D whole-body pose estimation system via regression and integration. In IEEE international conference on computer vision workshops . https://doi.org/10.1109/iccvw54120.2021.00201
57. He K, Gkioxari G, Dollar P, Girshick R (2017) Mask r-CNN. In, (2017) IEEE International conference on computer vision (ICCV). IEEE. https://doi.org/10.1109/iccv.2017.322
58. Taylor J, Shotton J, Sharp T, Fitzgibbon A (2012) The vitruvian manifold: Inferring dense correspondences for one-shot human pose estimation. In 2012 IEEE conference on computer vision and pattern recognition, pp 103–110. https://doi.org/10.1109/CVPR.2012.6247664
59. Wei L, Huang Q, Ceylan D, Vouga E, Li H (2016) Dense human body correspondences using convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2016.171
60. Pons-Moll G, Taylor J, Shotton J, Hertzmann A, Fitzgibbon A (2015) Metric regression forests for correspondence estimation, vol 113. Springer, pp 163–175. https://doi.org/10.1007/s11263-015-0818-9

61. Bristow H, Valmadre J, Lucey S (2015) Dense semantic correspondence where every pixel is a classifier. In Proceedings of the IEEE international conference on computer vision (ICCV). https://doi.org/10.1109/iccv.2015.458
62. Zhou T, Krahenbuhl P, Aubry M, Huang Q, Efros AA (2016) Learning dense correspondence via 3D-guided cycle consistency. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2016.20
63. Gaur U, Manjunath BS (2017) Weakly supervised manifold learning for dense semantic object correspondence. In Proceedings of the IEEE international conference on computer vision (ICCV) . https://doi.org/10.1109/iccv.2017.192
64. Perona P, Malik J (1990) Scale -space and edge detection using anisotropic diffusion, vol 12. pp 629–639. https://doi.org/10.1109/34.56205
65. Bertalmio M, Sapiro G, Caselles V, Ballester C (2000) Image inpainting. In Proceedings of the 27th annual conference on computer graphics and interactive techniques. SIGGRAPH' 00, ACM press/addison-wesley publishing Co. USA, pp 417–424. https://doi.org/10.1145/344779.344972
66. Kim D, Woo S, Lee J-Y, Kweon IS (2019) Deep video inpainting. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2019.00594
67. Criminisi A, Perez P, Toyama K (2004) Region filling and object removal by exemplar-based image inpainting, vol 13. pp 1200–1212 .https://doi.org/10.1109/TIP.2004.833105
68. Bertalmio M, Vese L, Sapiro G, Osher S (2003) Simultaneous structure and texture image inpainting, vol 12. pp 882–889 . https://doi.org/10.1109/TIP.2003.815261
69. Zhang H, Dai S (2012) Image inpainting based on wavelet decomposition, vol 29. 2012 international workshop on information and electronics engineering. pp 3674–3678. https://doi.org/10.1016/j.proeng.2012.01.551. https://www.sciencedirect.com/science/article/pii/S1877705812005619
70. Korshunov P, Ooi WT (2011) Video quality for face detection, recognition, and tracking, vol 7. Association for computing machinery, New York, NY USA. https://doi.org/10.1145/2000486.2000488
71. Yeh RA, Chen C, Yian Lim T, Schwing AG, Hasegawa-Johnson M, Do MN (2017) Semantic image inpainting with deep generative models. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2017.728
72. Yu J, Lin Z, Yang J, Shen X, Lu X, Huang TS (2018) Generative image inpainting with contextual attention. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2018.00577
73. Kim D, Woo S, Lee J-Y, Kweon IS (2019) Deep video inpainting. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2019.00594
74. Chang Y-L, Liu ZY, Lee K-Y, Hsu W (2019) Free-form video inpainting with 3D gated convolution and temporal PatchGAN. In Proceedings of the IEEE/CVF international conference on computer vision (ICCV). https://doi.org/10.1109/iccv.2019.00916
75. Kupyn O, Martyniuk T, Wu J, Wang Z (2019) Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In Proceedings of the IEEE/CVF international conference on computer vision (ICCV). https://doi.org/10.1109/iccv.2019.00897
76. Zhang H, Mai L, Xu N, Wang Z, Collomosse J, Jin H (2019) An internal learning approach to video inpainting. In Proceedings of the IEEE/CVF international conference on computer vision (ICCV) . https://doi.org/10.1109/iccv.2019.00281
77. Lee S, Oh SW, Won D, Kim SJ (2019) Copy-and-paste networks for deep video inpainting. In Proceedings of the IEEE/CVF international conference on computer vision (ICCV). https://doi.org/10.1109/iccv.2019.00451
78. Sharif M, Bhagavatula S, Bauer L, Reiter MK (2016) Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition. In Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. CCS' 16, Association for computing machinery. New York NY USA, pp 1528–1540. https://doi.org/10.1145/2976749.2978392
79. Komkov S, Petiushko A (2021) AdvHat: Real-world adversarial attack on ArcFace face ID system. In 2020 25th international conference on pattern recognition (ICPR), pp 819–826. https://doi.org/10.1109/ICPR48806.2021.9412236
80. Brown TB, Mané D, Roy A, Abadi M, Gilmer J (2017) Adversarial patch. https://arxiv.org/pdf/1712.09665.pdf
81. Liu Z, Luo P, Wang X, Tang X (2015) Deep learning face attributes in the wild. In Proceedings of the IEEE International conference on computer vision (ICCV). https://doi.org/10.1109/iccv.2015.425
82. Thys S, Van Ranst W, Goedeme T (2019) Fooling automated surveillance cameras: Adversarial patches to attack person detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR) workshops. https://doi.org/10.1109/cvprw.2019.00012

83. Shafahi A, Huang WR, Najibi M, Suciu O, Studer C, Dumitras T, Goldstein T (2018) Poison frogs! targeted clean-label poisoning attacks on neural networks. In Proceedings of the 32nd international conference on neural information processing systems. NIPS'18, Curran Associates Inc. Red Hook, NY USA, pp 6106–6116
84. Loper M, Mahmood N, Romero J, Pons-Moll G, Black MJ (2015) SMPL: A skinned multi-person linear model, vol 34. Association for computing machinery, New York, NY USA. https://doi.org/10.1145/2816795.2818013
85. Shen J, Zhu X, Ma D (2019) TensorClog: An imperceptible poisoning attack on deep neural network applications, vol 7. pp 41498–41506 . https://doi.org/10.1109/ACCESS.2019.2905915
86. Makihara Y, Sagawa R, Mukaigawa Y, Echigo T, Yagi Y (2006) Gait recognition using a view transformation model in the frequency domain. In: Leonardis A, Bischof H, Pinz A (eds) Computer vision - ECCV 2006. Springer, Berlin Heidelberg, pp 151–163
87. McPherson R, Shokri R, Shmatikov V (2016) Defeating image obfuscation with deep learning
88. Oh SJ, Benenson R, Fritz M, Schiele B (2016) Faceless person recognition: Privacy implications in social media. In Leibe B, Matas J, Sebe N, Welling M (eds) Computer vision - ECCV 2016. Springer Cham , pp 19–35. https://doi.org/10.1007/978-3-319-46487-92
89. Zhou J, Pun C-M (2021) Personal privacy protection via irrelevant faces tracking and pixelation in video live streaming, vol 16. pp 1088–1103. https://doi.org/10.1109/TIFS.2020.3029913
90. Menon S, Damian A, Hu S, Ravi N, Rudin C (2020) Pulse: Self-supervised photo upsampling via latent space exploration of generative models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr42600.2020.00251
91. Wang H, Wang Y, Zhou Z, Ji X, Gong D, Zhou J, Li Z, Liu W (2018) Cosface: Large margin cosine loss for deep face recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) . https://doi.org/10.1109/cvpr.2018.00552
92. Newton EM, Sweeney L, Malin B (2005) Preserving privacy by deidentifying face images 17:232–243 . https://doi.org/10.1109/TKDE.2005.32
93. Korshunov P, Ooi WT (2011) Video quality for face detection, recognition, and tracking, vol 7. Association for computing machinery, New York, NY USA . https://doi.org/10.1145/2000486.2000488
94. Menon S, Damian A, Hu S, Ravi N, Rudin C (2020) Pulse: Self-supervised photo upsampling via latent space exploration of generative models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR) . https://doi.org/10.1109/cvpr42600.2020.00251
95. Rozumnyi D, Oswald MR, Ferrari V, Matas J, Pollefeys M (2021) DeFMO: Deblurring and shape recovery of fast moving objects. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp 3456–3465 . https://doi.org/10.1109/cvpr46437.2021.00346
96. Kupyn O, Martyniuk T, Wu J, Wang Z (2019) Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In Proceedings of the IEEE/CVF international conference on computer vision (ICCV). https://doi.org/10.1109/iccv.2019.00897
97. Neverova N, Guler RA, Kokkinos I (2018) Dense pose transfer. In Proceedings of the european conference on computer vision (ECCV). https://doi.org/10.1007/978-3-030-01219-98
98. Knight W (2021) Clearview AI has new tools to identify you in photos. Conde nast. Accessed: 11 Mar 2021
99. Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2018) DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs, vol 40, pp 834–848. https://doi.org/10.1109/TPAMI.2017.2699184
100. Ni Z, Shi Y-Q, Ansari N, Su W (2006) Reversible data hiding. IEEE Trans Circ Syst Video Tech 16(3):354–362. https://doi.org/10.1109/tcsvt.2006.869964
101. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol 1, pp 886–8931. https://doi.org/10.1109/CVPR.2005.177
102. Oh SW, Lee S, Lee J-Y, Kim SJ (2019) Onion-peel networks for deep video completion. In Proceedings of the IEEE/CVF international conference on computer vision (ICCV). https://doi.org/10.1109/iccv.2019.00450
103. Wright CV, Feng W-c, Liu F (2015) Thumbnail-preserving encryption for jpeg. In Proceedings of the 3rd ACM workshop on information hiding and multimedia security. IH&amp;MMSec '15, Association for Computing Machinery, New York, NY USA, pp 141–146. https://doi.org/10.1145/2756601.2756618
104. Gross R, Sweeney L, de la Torre F, Baker S (2006) Model-based face De-Identification. In 2006 conference on computer vision and pattern recognition workshop (CVPRW' 06), pp 161–161. https://doi.org/10.1109/CVPRW.2006.125

105. Gross R, Airoldi E, Malin B, Sweeney L (2006) Integrating utility into face De-identification. In Danezis G, Martin D (eds) Privacy enhancing technologies. Springer, Berlin Heidelberg, pp 227–242. https://doi.org/10.1007/11767831
106. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)
107. Pavlakos G, Choutas V, Ghorbani N, Bolkart T, Osman AAA, Tzionas D, Black MJ (2019) Expressive body capture: 3D Hands, face, and body from a single image. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2019.01123
108. Choi Y, Choi M, Kim M, Ha J-W, Kim S, Choo J (2018) StarGAN: Unified generative adversarial networks for multi-domain image-to- image translation. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)
109. Moosavi-Dezfooli S-M, Fawzi A, Frossard P (2016) DeepFool: A simple and accurate method to fool deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)
110. Climent-Pérez P, Florez-Revuelta F (2021) Protection of visual privacy in videos acquired with RGB cameras for active and assisted living applications. Springer, pp 1–16. https://doi.org/10.1007/s11042-020-10249-1
111. Mondéjar-Guerra VM, Rouco J, Novo J, Ortega M (2019) An end-to-end deep learning approach for simultaneous background modeling and subtraction. In British machine vision conference, p 266
112. Phillips PJ, Wechsler H, Huang J (1998) Rauss PJ. The FERET database and evaluation procedure for face-recognition algorithms 16:295–306. https://doi.org/10.1016/S0262-8856(97)00070-X
113. Pilu M (2007) Detector for use with data encoding pattern. Google patents. US patent App. 11(491):174
114. Pavlakos G, Choutas V, Ghorbani N, Bolkart T, Osman AAA, Tzionas D, Black MJ (2019) Expressive body capture: 3D Hands, face, and body from a single image. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2019.01123
115. Osman AAA, Bolkart T, Black MJ (2020) STAR: A sparse trained articulated human body regressor. In European conference on computer vision (ECCV), pp 598–613 . https://doi.org/10.1007/978-3-030-58539-636. https://star.is.tue.mpg.de
116. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL (2014) Microsoft COCO: Common objects in context. In Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds) Computer vision - ECCV 2014, pp 740–755. Springer, Cham .https://doi.org/10.1007/978-3-319-10602-148
117. He K, Gkioxari G, Dollar P, Girshick R (2017) Mask r-CNN. In 2017 IEEE International conference on computer vision (ICCV). IEEE. https://doi.org/10.1109/iccv.2017.322
118. Alp Guler R, Trigeorgis G, Antonakos E, Snape P, Zafeiriou S, Kokkinos I (2017) DenseReg: Fully convolutional dense shape regression In-The-Wild. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) . https://doi.org/10.1109/cvpr.2017.280
119. Remagnino P, Foresti GL, Ellis T (2004) Ambient intelligence: A novel paradigm. Springer, New York USA
120. Rezaei B, Farnoosh A, Ostadabbas S (2020) G-LBM: Generative lowdimensional background model estimation from video sequences. In ECCV. https://doi.org/10.1007/978-3-030-58610-218
121. Bashir K, Xiang T, Gong S (2010) Gait recognition without subject cooperation, vol 31. Meta-heuristic intelligence based image processing, pp 2052–2060. https://doi.org/10.1016/j.patrec.2010.05.027. https://www.sciencedirect.com/science/article/pii/S0167865510001844
122. Liu Z, Sarkar S (2006) Improved gait recognition by gait dynamics normalization, vol 28. pp 863–876 . https://doi.org/10.1109/TPAMI.2006.122
123. Rong Y, Shiratori T, Joo H (2021) FrankMocap: A monocular 3D whole-body pose estimation system via regression and integration. In IEEE international conference on computer vision workshops. https://doi.org/10.1109/iccvw54120.2021.00201
124. Makihara Y, Sagawa R, Mukaigawa Y, Echigo T, Yagi Y (2006) Gait recognition using a view transformation model in the frequency domain. In Leonardis A, Bischof H, Pinz A (eds) Computer vision - ECCV 2006, Springer, Berlin Heidelberg , pp 151–163
125. Bobick AF, Johnson AY (2001) Gait recognition using static, activity-specific parameters. In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, vol 1, p . https://doi.org/10.1109/CVPR.2001.990506
126. Wan C, Wang L, Phoha VV (2018) A Survey on gait recognition, vol 51. Association for computing machinery, New York, NY USA. https://doi.org/10.1145/3230633
127. Agrawal P, Narayanan PJ (2011) Person De-identification in videos, vol 21. pp 299–310 . https://doi.org/10.1109/TCSVT.2011.2105551

128. Shahroudy A, Liu J, Ng T-T, Wang G (2016) Ntu rgb+d: A large scale dataset for 3d human activity analysis. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2016.115
129. Das S, Dai R, Koperski M, Minciullo L, Garattoni L, Bremond F, Francesca G (2019) Toyota smarthome: Real-world activities of daily living. In Proceedings of the IEEE/CVF international conference on computer vision (ICCV) . https://doi.org/10.1109/iccv.2019.00092
130. Mihailidis A, Colonna L (2020) A methodological approach to privacy by design within the context of lifelogging technologies, vol 46. p 1. HeinOnline
131. Konecný J, McMahan H, Yu F, Richtárik P, Suresh A, Bacon D (2016) Federated learning: strategies for improving communication efficiency, vol. abs/1610.05492 . https://doi.org/10.48550/arXiv:1610.05492
132. Council of the European Union, European Parliament (2018) Article 7 GDPR -conditions for consent. https://gdpr-info.eu/art-7-gdpr/. Accessed: 10 Aug 2021
133. Boulemtafes A, Derhab A, Challal Y (2020) A review of privacy-preserving techniques for deep learning, vol 384. pp 21–45 . https://doi.org/10.1016/j.neucom.2019.11.041. https://www.sciencedirect.com/science/article/pii/S0925231219316431
134. Sun Q, Ma L, Oh SJ, Van Gool L, Schiele B, Fritz M (2018) Natural and effective obfuscation by head inpainting. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2018.00530
135. Wagner I, Eckhoff D (2018) Technical privacy metrics: A systematic survey, vol 51. Association for computing machinery, New York, NY USA . https://doi.org/10.1145/3168389
136. Dwork C, Roth A, et al. (2014) The Algorithmic foundations of differential privacy, vol 9. pp 211–407
137. Amazon Web Services (2021) Amazon rekognition API. https://aws.amazon.com/rekognition/. Accessed: 30 June 2021
138. Sweeney L (2002) k-Anonymity: A model For protecting privacy, vol 10. pp 557–570 . https://doi.org/10.1142/S0218488502001648
139. Machanavajjhala A, Kifer D, Gehrke J, Venkitasubramaniam M (2007) l-Diversity: Privacy Beyond k-Anonymity, vol 1. Association for computing machinery, New York, NY USA . p 3 https://doi.org/10.1145/1217299.1217302
140. Li N, Li T, Venkatasubramanian S (2007) t-Closeness: Privacy beyond k-Anonymity and l-Diversity. In 2007 IEEE 23rd international conference on data engineering, pp 106–115. https://doi.org/10.1109/ICDE.2007.367856
141. Google (2008) Cloud vision API. https://cloud.google.com/vision Accessed: 13 Oct 2021
142. Microsoft Azure (2021) Facial recognition–microsoft azure. https://azure.microsoft.com/en-us/services/cognitive-services/face/. Accessed: 30 June 2021
143. Wagner I, Eckhoff D (2018) Technical privacy metrics: A systematic survey, vol 51. Association for computing machinery, New York, NY USA. https://doi.org/10.1145/3168389
144. Wan C, Wang L, Phoha VV (2018) A Survey on gait recognition, vol 51. Association for computing machinery, New York, NY USA. https://doi.org/10.1145/3230633
145. Xie S, Girshick R, Dollar P, Tu Z, He K (2017) Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) . https://doi.org/10.1109/cvpr.2017.634
146. Wang H, Wang Y, Zhou Z, Ji X, Gong D, Zhou J, Li Z, Liu W (2018) Cosface: Large margin cosine loss for deep face recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2018.00552
147. Prolific (2021) Prolific. https://prolific.co/. Accessed: 30 June 2021
148. Wei L, Huang Q, Ceylan D, Vouga E, Li H (2016) Dense human body correspondences using convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2016.171
149. Wilkowska W, Heek JO-v, Florez-Revuelta F, Ziefle M (2021) Video cameras for lifelogging at home: Preferred visualization modes, acceptance, and privacy perceptions among German and Turkish participants, vol 37, Taylor & Francis. pp 1436–1454 https://doi.org/10.1080/10447318.2021.1888487
150. Phillips PJ, Wechsler H, Huang J, Rauss PJ (1998) The FERET database and evaluation procedure for face-recognition algorithms, vol 16, pp 295–306 . https://doi.org/10.1016/S0262-8856(97)00070-X
151. Zhang N, Paluri M, Taigman Y, Fergus R, Bourdev L (2015) Beyond frontal faces: Improving person recognition using multiple cues. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) . https://doi.org/10.1109/cvpr.2015.7299113
152. AT&T Laboratories Cambridge (2002). https://git-disl.github.io/GTDLBench/datasets/attfacedataset/
153. Fan L (2018) Image pixelization with differential privacy. In Kerschbaum F, Paraboschi S (eds) Data and applications security and privacy XXXII. Springer Cham, pp 148–162. https://doi.org/10.1007/978-3-319-95729-610

154. Xie S, Girshick R, Dollar P, Tu Z, He K (2017) Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2017.634
155. Kumar N, Berg AC, Belhumeur PN, Nayar SK (2009) Attribute and simile classifiers for face verification. In 2009 IEEE 12th international conference on computer vision, pp. 365–372. https://doi.org/10.1109/ICCV.2009.5459250
156. Yang W, Luo P, Lin L (2014) Clothing Co-Parsing by joint image segmentation and labeling. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2014.407
157. Yeh RA, Chen C, Yian Lim T, Schwing AG, Hasegawa-Johnson M, Do MN (2017) Semantic image inpainting with deep generative models. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2017.728
158. Yonetani R, Naresh Boddeti V, Kitani KM, Sato Y (2017) Privacy-preserving visual learning using doubly permuted homomorphic encryption. In Proceedings of the IEEE international conference on computer vision (ICCV). https://doi.org/10.1109/iccv.2017.225
159. Yu J, Lin Z, Yang J, Shen X, Lu X, Huang TS (2018) Generative image inpainting with contextual attention. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2018.00577
160. LeCun Y (1998) The MNIST database of handwritten digits. http://yann.lecun.com/exdb/mnist/. Accessed: 13 Sept 2021
161. Abadi M, Chu A, Goodfellow I, McMahan HB, Mironov I, Talwar K, Zhang L (2016) Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. CCS '16, pp 308–318. Association for Computing Machinery, New York, NY USA .https://doi.org/10.1145/2976749.2978318
162. Zhang N, Paluri M, Taigman Y, Fergus R, Bourdev L (2015) Beyond frontal faces: Improving person recognition using multiple cues. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2015.7299113
163. Zhang H, Mai L, Xu N, Wang Z, Collomosse J, Jin H (2019) An internal learning approach to video inpainting. In Proceedings of the IEEE/CVF international conference on computer vision (ICCV). https://doi.org/10.1109/iccv.2019.00281
164. Zhang Y, Zhao R, Zhang Y, Lan R (2022) Chai X. High-efficiency and visual-usability image encryption based on thumbnail preserving and chaotic system 34:2993–3010. https://doi.org/10.1016/j.jksuci.2022.04.001 www.sciencedirect.com/science/article/pii/
165. Yu S, Tan D, Tan T (2006) A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In 18th international conference on pattern recognition (ICPR'06), vol 4, pp 441–444. https://doi.org/10.1109/ICPR.2006.67
166. Yang W, Luo P, Lin L (2014) Clothing Co-Parsing by joint image segmentation and labeling. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) . https://doi.org/10.1109/cvpr.2014.407
167. Ionescu C, Papava D, Olaru V, Sminchisescu C (2014) Human3.6M: Large scale datasets and predictive methods for 3D human sensing in natural environments, vol 36, pp 1325–1339 . https://doi.org/10.1109/TPAMI.2013.248
168. Brkić K, Hrkać T, Kalafatić Z (2017) Protecting the privacy of humans in video sequences using a computer vision-based De-identification pipeline, vol 87, pp 41–55. https://doi.org/10.1016/j.eswa.2017.05.067. https://www.sciencedirect.com/science/article/pii/S0957417417303986
169. Shahroudy A, Liu J, Ng T-T, Wang G (2016) Ntu rgb+d: A large scale dataset for 3d human activity analysis. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) .https://doi.org/10.1109/cvpr.2016.115
170. Wang ZW, Vineet V, Pittaluga F, Sinha SN, Cossairt O, Bing Kang S (2019) Privacy-preserving action recognition using coded aperture videos. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR) workshops . https://doi.org/10.1109/cvprw.2019.00007
171. Zhou T, Krahenbuhl P, Aubry M, Huang Q, Efros AA (2016) Learning dense correspondence via 3D-guided cycle consistency. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/cvpr.2016.20
172. Gurrin C, Albatal R, Joho H, Ishii K (2014) A privacy by design approach to lifelogging. In Digital enlightenment yearbook 2014, IOS Press pp 49–73
173. Sightengine (2021) Text moderation in image/videos. https://sightengine.com/docs/ocr-text-moderation-in-images
174. Barth S, de Jong MDT (2017) The privacy paradox -Investigating discrepancies between expressed privacy concerns and actual online behavior. A Systematic literature review, vol 34 pp

1038–1058. https://doi.org/10.1016/j.tele.2017.04.013. https://www.sciencedirect.com/science/article/pii/S0736585317302022

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.