Visual Content Privacy Protection: A Survey

Ruoyu Zhao, Yushu Zhang, Tao Wang, Wenying Wen, Yong Xiang, and Xiaochun Cao

Abstract—Vision is the most important sense for people, and it is also one of the main ways of cognition. As a result, people tend to utilize visual content to capture and share their life experiences, which greatly facilitates the transfer of information. Meanwhile, it also increases the risk of privacy violations, e.g., an image or video can reveal different kinds of privacy-sensitive information. Researchers have been working continuously to develop targeted privacy protection solutions, and there are several surveys to summarize them from certain perspectives. However, these surveys are either problem-driven, scenario-specific, or technology-specific, making it difficult for them to summarize the existing solutions in a macroscopic way. In this survey, a framework that encompasses various concerns and solutions for visual privacy is proposed, which allows for a macro understanding of privacy concerns from a comprehensive level. It is based on the fact that privacy concerns have corresponding adversaries, and divides privacy protection into three categories, based on computer vision (CV) adversary, based on human vision (HV) adversary, and based on CV & HV adversary. For each category, we analyze the characteristics of the main approaches to privacy protection, and then systematically review representative solutions. Open challenges and future directions for visual privacy protection are also discussed.

Index Terms—Security and privacy, privacy protection, Image and video, usability.

I. INTRODUCTION

Vision is the most important and complex sense of people, which has become a textbook answer in a sense [1]. For example, there are usually a large number of chapters about vision rather than others in textbooks (e.g., [2]) on cognitive science. The importance of vision has also been pointed out in some works. In [2], it is stated, "vision is the most widely recognized and the most widely studied perceptual modality"; In [3], it is said, "vision is the most complex, highly developed, and important sense for humans and most other mobile creatures". For a reasonable extension, humans prefer to intuitively obtain information from visual content compared with other ways[1]. An example that illustrates the intuition of visual content can be found in the ancient practice of early humans drawing on rock walls to convey information, and even today, young children learn about the world through picture books.

Visual content is better for expressing information than text. this is an era of rough reading; a considerable number of

R. Zhao, Y. Zhang, and T. Wang are with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China (e-mail: zhaoruoyu@nuaa.edu.cn; yushu@nuaa.edu.cn; wang-tao21@nuaa.edu.cn).

W. Wen is with the School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330032, China (e-mail: wenying-wen@sina.cn).

Y. Xiang is with the School of Information Technology, Deakin University, Burwood, Victoria 3125, Australia (e-mail: yxiang@deakin.edu.au).

X. Cao is with School of Cyber Science and Technology, Sun Yat-sen University, Shenzhen, China (e-mail: caoxch5@sysu.edu.cn).

[1]https://blog.hubspot.com/marketing/visual-content-marketing-strategy

people will not carefully read text, which is a tedious process. For example, research has shown that most of the reading on Twitter occurs only roughly on the first line [4]. In fact, from a human perspective, the response and processing effect of visual content is better than any other type of data, e.g., the human brain processes images about 60k times faster than text, and about 90% of the information transmitted to the brain is visual [4].

People apply visual content data to record and share daily life, important things, and the world, which greatly promotes information exchange and highlights effective information [5]. Since the first photo was taken, i.e., Nipce Heliograph was made in 1827 [6], people have entered a new era of recording life with visual content. According to some data, more than 1 trillion photos were taken in 2015 [7], which is expected to be 1.81 trillion in 2023 [8]. Meanwhile, social networks (OSNs) have occupied most people's lives in the Internet era. Sharing visual content in OSNs is becoming increasingly popular, and people can express themselves and interact with others at any time with a simple click [9]. According to data, 6.9 billion and 1.3 billion photos are shared on Whatsapp and Instagram every day, respectively [8]. 72 hours of video content will be added to Youtube every minute [5].

These captured visual content may reveal privacy information. Traditionally, privacy include but are not limited to personal identity, habits, preferences, and social relations [10]. Meanwhile, with the continuous development of technology, computer vision can detect more information from visual content (e.g., [11], [12], [13]), which has brought great benefits to mankind but also exacerbated the privacy risk. Like the first act of digging a window on the dark wall of the cave, the caveman will be divided into two groups: one group is pleased with the extra sunshine brought by it, and the other group is shocked by the fact that it makes their lives spied [14]. In order to obtain sunlight and prevent prying, a ground glass can be added to the window. The purpose of privacy protection is similar to that of ground glass, that is, to balance usability and privacy security in visual content.

A. Characteristics of Visual Content Privacy

In the information age, privacy has received widespread attention due to frequent data leakage, e.g., iCloud leakage event [15], and the introduction of relevant laws and regulations, e.g., EU' General Data Protection Regulation (GDPR) [16], US' California Consumer Privacy Act [17], and UN' Declaration of Human Rights [18]. Privacy is considered a personal right in a sense, that is, an individual can control the extent to which relevant data are used and disclosed [19]. If an issue is concerned with the rights of individuals, it is considered a privacy issue. It should be pointed out that the terms 'privacy' and 'confidiality' are all response

data application guarantees, and they have subtle differences, although they seem to be similar in perception. First, privacy is closely related to a person. Broadly speaking, it refers to the personal right not to be infringed, i.e., the right to protect own personal data is more emphasized [20]. Second, confidentiality emphasizes that unauthorized persons cannot access data. It emphasizes more on an objective right and a general data protection [19].

The purpose of privacy protection is that individuals want to prevent data information that is considered privacy sensitive from entering the public domain. When the data are images and videos, this is called **visual privacy protection** [21]. In this survey, unless otherwise specified, the terms 'privacy' and 'visual privacy' are the same.

Visual content is a rich and subjective source of information as a widely circulated proverb, '*A picture is worth a thousand words*' [22]; and similar words can be found in different cultures and language backgrounds. Different readers will extract different meanings from visual content, and readers' reactions will also change with time and environment [23]. This makes that using visual content to record information fashionable and will become increasingly popular. However, it also leads to a large amount of privacy information carried by visual content, which is difficult to handle. Meanwhile, even seemingly non-private and mediocre visual content may accidentally reveal privacy.

There are three fundamental difficulties in the privacy of visual content, which makes the privacy protection task very challenging.

The new technology of computer vision (CV) is constantly proposed. CV is designed to recognize and understand the visual content to extract information. It can bring a lot of benefits but also poses a great threat to privacy. Meanwhile, with the development of CV new technology, the recognition and understanding of visual content by CV have been greatly improved in both depth and breadth. Depth refers to the information that CV is already able to recognize and understand, but it is more accurate with the help of new technology (i.e., fine-grained). Breadth means that information that CV would not have been able to recognize and understand can be done by new technology; i.e., valid information that could not originally be extracted is made by the new technology (diversity). This continuous fine-grained and diversified improvement has broadened the boundaries of privacy caused by CV. In a way, this makes privacy protection and CV a never-ending arms race.

Human vision (HV) is a sophisticated perception. Vision is one of the most important senses for human beings. It is the process of deriving meaning from what is seen, and it is a complex function involving multiple skills. Meanwhile, HV is also not simply a visual input but a combined neurological reasoning process [24]. For example, a study pointed out that with only a small fragment of the original content, people can also reason about the complete content [25]. Despite centuries of research, humans still do not fully understand this process [26], and HV has been identified as having a subjective nature in a sense [27]. Thus, although the ability of HV does not seem to have been improving in recent years, new privacy protection schemes for HV are still being proposed.

Usability requirements for visual content are complex and diverse. The source of privacy concerns lies in the information carried by visual content. There is no doubt that privacy can be well protected if some methods, e.g., image [28] and video [29] encryption, are used to discard all information in visual content without distinction [30]. On the other hand, this also abandons the original purpose of visual content, which is to be used to express information, i.e., usability. Meanwhile, different visual content will have various usability requirements, and even the same visual content in different application scenarios will have different requirements, leading to the complexity and diversion of usability requirements, and the expansion of CV technology in depth and breadth has led to expanding usability requirements.

Case study 1: How much information can visual content (for example, an image) reveal?

As shown in Fig. 1, this is an image that is often captured in everyday life, and it will be often seen in OSNs, cloud storage platforms, chats, etc. Although it may seem mundane, it can reveal a lot of information. CV and HV can cope well with some direct tasks, such as recognizing identity, gender, beauty, etc, based on the face. On the other hand, CV and HV also have their own unique abilities. The reason for this is that when humans process visual content with a preference for using semantic information, in which CV cannot reliably use, to reason on the whole [31], while CV prefers low-level information such as texture to process on the data space [32]. For those who know the lady in the image, they may be able to identify her simply by her clothes or body outline, which is a high-level and subjective reasoning process and CV is difficult to do [33]. For CV, it can be well captured and analyzed some subtle changes beyond HV since it acts in the data space. For example, the heartbeat can cause subtle changes in skin gloss, which can hardly be detected by HV but can be accurately captured by CV. Further, information such as heart rate [34], [35] and blood oxygen [36] be obtained through analysis.

Case study 2: Do privacy/usability requirements differ in different scenarios?

For just this one image, there can be complex and even completely opposite privacy and usability requirements. For example, if a user uploads this image to OSNs, chances are she wants to showcase herself to friends but is worried about being probed by CV, e.g., ClearView AI event [37], to find out who is in the image. On the other hand, an example is that the image owner wants to apply CV for recognition, but does not want HV to get the information by browsing. For the former, usability is HV to identify people and privacy is blocking CV to identify; For the latter, the opposite is true.

It is important to note that this is a simple case, but the actual one is much more complex and diverse than the above. For example, the researchers [22] analyzed Facebook profile photos on a tiny scale and came up with a lot of surprising information, e.g., constructing online identity and personality.

B. Analysis of Existing Surveys

There are currently some relevant surveys covering this topic, and they can be divided into three main categories:
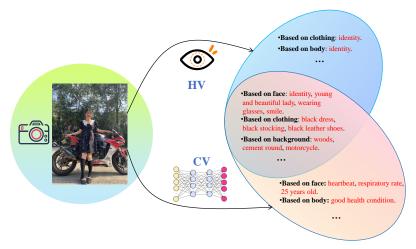
Fig. 1: A simple case of visual content revealing information. For a visual content, CV and HV are able to extract a large amount of information, some of which is unique and some of which has overlap.

- *Surveys on specific privacy concerns.* They only care about specific concerns without considering the scenarios and technologies used. In [38], privacy concerns arising from identification in multimedia are focused on, and a review summarizes the de-identification mechanism of biometric, non-biometric, and physiologic features in multimedia. In [39], the privacy of face images is concerned, and meanwhile, the main concepts, features, and challenges of protection technologies are introduced, and existing schemes are classified.

- *Surveys on specific scenarios.* They consider privacy concerns in a specific scenario. In [40], the major security and privacy concerns brought by multimedia content sharing OSNs are discussed, and some possible countermeasures are probed. In view of the complex privacy threats faced by image sharing in OSNs, taking the process of image sharing as a clue, the survey [9] proposes a lifecycle privacy protection framework, and classifies and reviews the existing schemes. The popularity of surveillance cameras has recorded a large amount of visual content with privacy information, and the popularity of CV has led to a rapid increase in the ability to automate the processing of the information, which in turn poses a huge threat to privacy. In [21], a comprehensive review of the image privacy is presented for surveillance scenarios, and classification and summary of schemes are presented. The deployment of automated face detection/recognition systems has been a significant threat to personal privacy In [41], a summary survey of works addressing such concerns is presented, discussing their limitations and future possibilities.

- *Surveys on specific technologies.* They consider specific technologies to deal with some privacy concerns. In [42], privacy attacks, as well as protections based on machine learning, are reviewed, which deals with attacks on and protection of visual content. In [43], it is reviewed for protection schemes for differential privacy applications to unstructured content, which involves (but are not limited to) image and video data.

A quantitative comparison of this survey with the above surveys is shown in Table I, in which the meaning of symbols are as follows:

- ●: The survey explicitly mentions this item;
- ◐: Although some (very small) percentages in the surveys deal with non-biometric features, their goal is still to protect biometric ones;
- ○: The survey explicitly does not include this item;
- N/A: No relevant items are mentioned.

In all, previous surveys often focused on a specific privacy (e.g., de-identification [38]) or on a specific scenario (e.g., OSNs [9]) or on the impact that a particular technology may have on the privacy protection of visual content (e.g., machine learning [42]). They achieve the intended goal well but are inadequate in terms of privacy protection of visual content. For example, de-identification is indeed an important topic in privacy protection, but concerns in visual content go far beyond personal identifiers (e.g., health status); OSNs are indeed one of the significant streams of visual content, and also generate a great of privacy concerns as a result (e.g., ClearView AI event [37]), but the application scenarios for visual content are not limited to OSNs either (e.g., cloud storage); machine learning is an important source of privacy threats to visual content and an influential enabler of privacy protection, but HV is also a major threat to privacy, and non-machine learning methods can also be good for privacy protection (e.g., blur and mosaic [21]).

The fundamental difference between this survey and previous pioneering works is that it contains all privacy concerns and countermeasures related to visual content, rather than a limited set of preferences. The key differences are as follows:

- Previous surveys have often been conducted from a problem-driven perspective, i.e., presupposing a specific concern or scenario. Instead, this survey tries to build a more comprehensive overview of the field by focusing on the visual content itself.
- Previous surveys often focused on the privacy concerns posed by CV, and even when HV is considered, they do not place CV and HV at the same level of narrative on an equal footing. In contrary, this survey considers the threat to privacy posed by CV and HV to be equally important in terms of for visual content.

TABLE I: Comparison of this survey with existing surveys.

Paper	Year	Adversary CV	Adversary HV	Image	Video	Non-biometric feature	Focus
[38]	2016	●	●	●	●	◐	De-identification
[39]	2021	●	○	●	○	○	Face biometrics
[40]	2015	N/A	N/A	●	●	N/A	Cyber and system security in OSNs
[9]	2023	●	●	●	○	●	Image content in OSNs
[21]	2015	●	●	●	○	◐	Image content in surveillance
[41]	2023	●	○	●	●	○	Biometric facial recognition system
[42]	2021	●	○	●	●	N/A	Machine learning privacy
[43]	2022	N/A	N/A	●	●	N/A	Differential privacy for unstructured data
This survey	-	●	●	●	●	●	Macro understanding of visual content privacy

- The privacy threats considered in previous surveys are mainly brought by biometrics, and they do not consider non-biometrics, or consider them also for de-biometrics. Conversely, in this survey, non-biometric features are also an important part of privacy protection, especially for HV.

In brief, this survey provides a comprehensive understanding of visual content privacy through the following contributions: 1) the characteristics and taxonomy of visual content privacy; 2) a novel privacy protection framework for visual content; 3) any visual privacy protection scheme can find its place in this framework; 4) any privacy concerns regarding visual content can find potential solutions here; 5) an in-depth review of current advances in visual privacy protection solutions; 6) a discussion of the challenges and future directions of visual privacy protection.

The remainder of this survey is organized as follows. Section II presents a taxonomy based on adversaries, along with a description of each adversary and their relationships and differences. Sections III to V are specific to each category, review representative solutions, and discuss common principles. Section VI discusses the challenges and future directions of visual privacy protection. Section VII is a brief conclusion.

II. VISUAL CONTENT PRIVACY: AN OVERVIEW

A. Adversary-Based Taxonomy

The field of information security is essentially an offensive and defensive confrontation, proposing the targeted defense solution for a specific security threat (i.e., adversary). In other words, the role of an adversary can be found in any given work. Privacy protection, in a sense, also belongs to the field of information security, and any specific scheme can also find a privacy adversary. The taxonomy by adversaries can include all privacy threats and protection solutions for visual content. There are three adversaries of privacy for visual content.

CV adversary: It refers to an adversary that uses only artificial intelligence to reason or associate sensitive information implied in visual content. Essentially, CV is an attempt to replicate the HV system in order to understand visual content. Therefore, it can directly recognize some observable information of visual content, such as face identity, color, license plate number, etc., just like HV. Meanwhile, for certain information that is difficult for humans to perceive, such as soft biometric attributes [51], CV is good at it which can be accurately reasoned in latent feature space.

HV adversary: It refers to an adversary that uses only the human eyes to directly view visual content to obtain sensitive information such as the face, signature, license plate number, etc. This is the most intuitive privacy adversary, since the original purpose of visual content is to convey information to people through browsing. This is also the easiest adversary to achieve privacy breaches since once the visual content is accessed maliciously, the HV adversary will immediately gain access to privacy-sensitive information.

CV&HV adversary: It refers to an adversary that uses both CV and HV to infringe on sensitive information in visual content. As shown in Fig. 1, some information can be recognized and extracted by both CV and HV, such as face identity, gender, and license plate number, etc. The difference between this adversary and the above two is that the above only considers pure CV and HV threats to privacy. However, this part considers a combination of the both privacy threats and takes into account the usability of a specific CV.

The differences and connections between these three are as follows:

- Protection against CV adversaries considers only attacks on CV, with visual content changes as small as possible, with no or minimal impact on HV. This is because CV is far less robust than HV and only tiny changes are required to achieve protection against CV.
- Protection against HV tends to work for CV, and the opposite does not hold [39]. The protection of HV should significantly change the visual content since HV is robust and tends to learn the visual content in its entirety. An example is shown in Fig. 2. One of the protection against HV adversaries is to erase all text, such as license plate numbers, which naturally also has a decisive effect on CV. The protection against HV adversaries is added with noise that has no effect on the text extracted by HV.
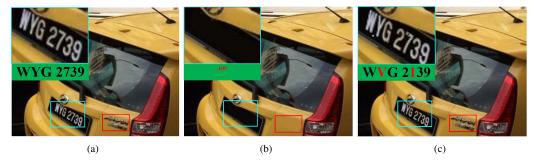
Fig. 2: Examples (from [44]) against HV adversaries VS. against CV adversaries, in which text is the privacy content. (a) original image; (b) against HV adversaries; (c) against CV adversaries. As shown in Fig. 2(b), if the protected content (e.g., license plate number area) has a feature that can be detected by a specific CV for the text 'WYG 2739', then it is said to be protected against HV&CV adversaries.
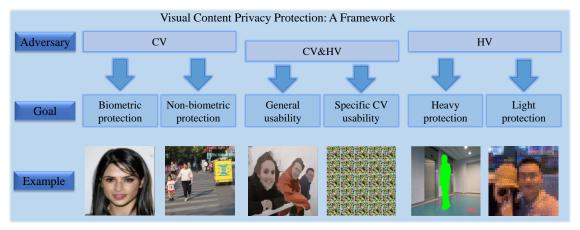

Fig. 3: Visual content privacy protection framework, including three adversaries: CV, HV, and CV&HV. Examples from left to right are from, respectively, [45], and [46], [47], [48], [49], and [50].

- If the solution for HV adversaries considers the usability of the specific CV, then CV is divided into usability CV and adversary CV, and it becomes a protection against HV&CV adversary. Specifically, the difference between protection against HV&CV and against HV only is that the former takes into account the usability of the protected content to a specific CV (model or task), while the latter does not. As shown in Fig. 2(b), an example is given: if the protected content (e.g., license plate number area) has a feature that can be detected by a specific CV for the text 'WYG 2739', then it is said to be protected against HV&CV.

- In all, the ranking of the privacy protection ability of solutions is HV&CV > HV > CV. Specifically, privacy protection schemes for HV&CV adversaries also work for CV adversaries; and for CV adversaries also work for HV ones. However, the opposite is not valid.

B. Privacy Protection Framework

We design an adversary-based privacy protection framework to help identify different privacy concerns in visual content and investigate corresponding countermeasures. The framework covers all privacy adversaries that visual content can may encounter, forming a closed loop for privacy analysis, and any scheme of visual privacy protection can find a suitable classification in this framework, as shown in Fig. 3.

1) CV adversary:

- **Biometric protection:** It prevents CV from extracting biometric features from visual content (mainly faces), which is one of the main directions of interest to the privacy community today.
- **Non-biometric protection:** It prevents CV from learning about non-biological features from the visual content, e.g., license plate number and location.

2) HV adversary:

- **Local heavy protection:** It targets sensitive areas in visual content and completely eliminates the visual effects therein.
- **Global light protection:** Its goal is to eliminate all the refined visual content, so that the naked eye can only observe the rough content.

3) CV&HV adversary:

- **Partial usability preserved:** It prevents CV and HV from identifying specific or generalized private information, while preserving certain visuals for CV and HV to perform cursory tasks, e.g., counting.
- **Specific CV usability:** It blocks CV and HV from identifying private information while not preserving any visuals, but allows specific CV models to perform a limited task.
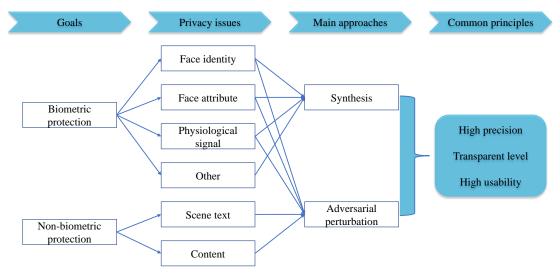
Fig. 4: Overview of the privacy protection analysis for CV adversary.

III. PRIVACY PROTECTION FOR CV ADVERSARY

This section provides a comprehensive analysis of privacy concerns posed by CV-based adversaries to visual content, including privacy issues, main approaches, and common principles, as shown in Fig. 4.

A. Privacy Issues in CV Adversary

1) Issues associated with biometric protection: Extracting biometrics from visual content, especially face, using CV has always been an important interest in the field of artificial intelligence due to its practicality. A rich variety of biometric features are extracted from the visual content represented by face, and the accuracy and variety continue to expand, have also led to serious privacy concerns. Depending on the focus of biometrics, the corresponding protection can be divided into four categories.

- **Face identity:** One of the major purposes of taking photos or videos is to show oneself, which inevitably records the person's face. One of the most mature, highly regarded, and widely used technologies in the CV field is facial identity recognition [52]. Meanwhile, face identity is also one of the hardest hit areas for privacy issues from visual content, e.g., ClearView AI event [37], and accordingly one of the focuses of research in the privacy community.
- **Face attribute:** As CV's ability to analyze and process visual content has increased, CV' analysis of faces has expanded beyond identity to include attribute features such as age, expression, race, and gender. These attributes may help the adversary to spy on people more accurately, and in a sense, attributes are more sensitive than identity, which can give rise to bias and discrimination (e.g., race and gender) [53].
- **Physiological signal:** In the video content, not only the information about a person's identity and attributes exist, but also the physiological signals of the person are captured [54]. The study points to the feasibility of remote photoplethysmography (rPPG) based on visual content via CV, e.g., heart rate [55] and respiration rate

[56]. These signals may be secretly analyzed and used by intentional people, which may pose a threat to privacy, such as gaining advantages in negotiation and analyzing health status.

- **Other:** The studied described above are biometric protections of great concern to the privacy community since they pose intuitive privacy concerns that ordinary people are often aware of in their daily lives. On the other hand, CV is also capable of extracting other biometrics that pose privacy threats in addition to the types mentioned above, e.g., gait [57] and eye gaze [58], and there is some sporadic but equally important work that considers this privacy.

2) Issues associated with non-biometric protection: CV work on the non-biometric content is not as focused as biometric features (often for faces), focusing mainly on scene text and content analysis.

- **Scene text:** Text is one of the basic media for transmitting information, and it is ubiquitous in daily life, such as street nameplates and license plate numbers. This kind of text in the natural environment is called scene text [59], which is one of the focuses of CV attention. On the other hand, scene text can contain all kinds of sensitive information (e.g., personal names and addresses), and with the advancement of CV, the resulting privacy risks are greater.
- **Content:** CV is able to analyze visual content and infer the presence of sensitive non-biometric information such as scene and clothing, which can also adversely affect privacy if used by adversaries [60].

B. Main Approaches

Synthesis: It refers to privacy protection by replacing the sensitive content in the original visual content with the privacy-preserving one that can be pre-existing or generated. With the advent of deep generative models, such as Generative Adversarial Networks (GAN), great progress has been made in building high quality generative models. For face privacy protection, privacy-sensitive features are usually extracted from

the original face, modified (e.g., eliminated or obfuscated), and then the protected features are combined with other necessary features are fed together into a generator to reconstruct the privacy-preserving face.

Adversarial perturbation: Research has shown that CV can perceive features that are not perceived by humans but are critical to CV to recognize objectives [61], making CV particularly sensitive to some perturbations that are not perceptible to humans. Therefore, corruption of these features, i.e., using carefully crafted perturbations added directly to the original visual content, can effectively prevent CV from recognizing the according sensitive features by CV.

C. Solutions for Biometric Protection

1) Face identity:

Synthesis. Bitouk *et al.* [62] created a large face database by crawling web images, protecting identities by comparing the face in the source image with faces in the database and selecting a face with the highest scores to swap. Afterward, the swapped faces are re-illuminated and recolored to blend into the image as much as possible and produce visually convincing results. Mosaddegh *et al.* [63] achieved identity protection by dividing a face into some components and then using the different components from multiple faces in the database for corresponding swaps. It somehow ensures that the protected faces do not resemble any human face and protects the privacy of the donor. 2D face swapping requires the face in the source image to have similar pose and appearance to the face in the target one, which affects its application. Lin *et al.* [64] proposed a 3D model-based face swapping scheme that is capable of rendering any pose of the head, eliminating restrictions on pose and appearance similarity, and enabling the swapping of front or side faces.

GAN has been shown to be successful in generating realistic face regions and is therefore heavily used in existing schemes to protect face privacy [65]. DeepFake [66], a deep forgery technique using GAN, is proposed for face swapping, which is robust to different facial expressions, lighting, and poses, and is also applicable to video. Zhu *et al.* [67] proposed the use of DeepFake applied to medical videos to protect the privacy of patient's identity while applying to the medical examination of Parkinson. Nirkin *et al.* [65] proposed a face swap scheme based on GAN that can manipulate pose, expression, and change identity simultaneously.

Face swapping schemes protect identity privacy in the source content by transferring the face from the target content to the source one. However, there are some problems. First, this often requires a target face or even a large database of faces. Second, although the identity of the source face is protected, the information of the target one is compromised.

Some GAN-based identity protection schemes that directly suppress or eliminate the original identity features of faces have also been proposed. Maximov *et al.* [45] proposed a framework for face anonymization based Conditional GAN for images and videos, where faces are anonymized based on Siamese networks to provide guidance for identity signals. Wen *et al.* [68] combined differential privacy and GAN to propose an image anonymization scheme that can adjust the balance of privacy and usability through privacy budget parameters, while explicitly proposing to maintain the attributes unchanged. Zhai *et al.* [69] pointed out that considering anonymization purely in terms of identity characteristics may lead to uncontrollable changes in some attributes. Therefore, they proposed a scheme to change identity by controlling changes in attributes, so as to achieve the three requirements of anonymity, realism, and controllability. In addition, Gu *et al.* [70] combined passwords with GAN to propose reversible face anonymization schemes that recover the original face when the user gives the correct password and generate faces with different identities when different passwords are used.

The GAN-based scheme to synthesize facial content is indeed good at ensuring the quality of the generated visual content, while it does not preserve the face of the original face, and thus not only prevents CV from recognizing identity, but may also mislead HV (although it is not intended to do so).

Adversarial perturbation. Chatzikyriakidis *et al.* [71] proposed the use of penalized fast gradient value method to generate adversarial perturbations to achieve face de-identification, which is guaranteed to be visually very similar to the original image. Cherepanova *et al.* [72] addressed the privacy problem of capturing images on OSNs to build face databases, and proposed the first tool to block commercial face API to evade face recognition by adding adversarial perturbation before uploading images to OSNs. Zhong *et al.* [73] proposed a custom cloak face privacy protection using adversarial perturbation, in which each identity has a personalized cloak and all images about this identity can apply the clock.

The face protection scheme using adversarial perturbation ensures that the information perceived visually by humans is consistent with the original visual content. However, it is often computationally expensive and needs to be retrained for each face or each identity to gain perturbations.

2) Face attribute:

Synthesis. Yang *et al.* [74] proposed a tensor-based 3D face geometry reconstruction method for modifying face expressions in videos. He *et al.* [75] proposed to use attribute classification constraints to guarantee the correct change of the required attributes, while introducing reconstruction learning to preserve attribute-excluding details to ensure that irrelevant attributes do not change. Wang *et al.* [76] ensured that the sensitive attributes of faces in images cannot be inferred by CV through inverting corresponding attributes or maximizing the uncertainty of the attributes, rather than removing them. Mirjalili *et al.* proposed a new model for protecting multiple attribute privacy, such as gender, race, and age, while maintaining the matching performance of attribute features [77].

Adversarial perturbation. Mirjalili *et al.* [78] proposed to use adversarial perturbation to enable gender attributes to be assessed by CV as flipped, i.e., females as judged as males and vice versa, and meanwhile, the face matcher is still available. On the other hand, the perturbation it produces is clearly visible to HV. Low *et al.* [79] focused on the privacy implications of face expression detection in video, proposing a framework seeking a universal perturbation that allows automatic micro-expression recognition to be confused,

TABLE II: A summary of the solutions in biometric protection.

Sub class	Paper	Year	Key technology	HV		Reversi-bility	Type		Note
				Reality	Invariance		I	V	
Face identity	[62]	2008	face swapping	☐	☒	☒	☑	☒	Infringed the identity of others
	[63]	2015	face swapping	☐	☒	☒	☑	☒	Multi-facial component splicing
	[64]	2012	face swapping	☐	☒	☒	☑	☒	Arbitrary face pose
	[67]	2020	DeepFake	☑	☒	☒	☒	☑	Medical applications
	[65]	2019	GAN	☑	☒	☒	☑	☑	Arbitrary face pose
	[45]	2020	GAN	☑	☒	☒	☑	☑	Identity vector guidance
	[68]	2022	GAN	☑	☒	☒	☑	☒	Differential privacy
	[69]	2022	GAN	☑	☒	☒	☑	☒	Control attributes to change identity
	[70]	2020	GAN	☑	☒	☑	☑	☒	New identity used for wrong password
	[71]	2019	Perturbation	☑	☑	☒	☑	☒	-
	[72]	2021	Perturbation	☑	☑	☒	☑	☒	Against commercial APIs
	[73]	2023	Perturbation	☑	☑	☒	☑	☑	One identity one cloak
Face attribute	[74]	2012	3D face reconstruction	☑	☒	☒	☒	☑	Focused on expression
	[75]	2019	GAN	☑	☒	☒	☑	☒	Only change target attributes
	[76]	2021	GAN	☑	☒	☒	☑	☒	Attribute obfuscation
	[77]	2020	GAN	☑	☒	☒	☑	☒	Multi-attribute
	[78]	2017	Perturbation	☒	☑	☒	☑	☒	Focused on gender
	[79]	2022	Perturbation	☑	☑	☒	☒	☑	Focused on emotion
Physio-logical signal	[80]	2017	Elimination	☐	☑	☒	☒	☑	Unrealistic videos may appear
	[54]	2022	Perturbation	☑	☑	☒	☒	☑	Designating rPPG
	[81]	2022	Deep learning	☑	☑	☒	☒	☑	Real time
Other	[82]	2015	Reshaping	☑	☒	☒	☒	☑	Preserving posture
	[83]	2022	GAN	☑	☒	☒	☑	☒	pedestrain anonymination

which is then injected into the video to protect the expressions in the video.

3) Physiological signal:

Chen *et al.* [80] proposed to eliminate the physiological signal from facial videos using Laplace pyramid representation so that the corresponding physiological signals can no longer be accurately measured, but the scheme may produce unrealistic videos in order to eliminate signals [91]. PulseEdit [54] is proposed to protect rPPG signals in facial video by adversarial perturbation. It first generates a target rPPG signal, and then combines the target signal with the original one to solve an optimization problem and obtain the perturbation, which is added uniformly to the facial region. However, it does not run fast enough to support real-time processing, and is poorly defended against deep learning rPPG methods. Sun *et al.* [81] proposed to use a deep learning method to modify the rPPG signal in facial videos, achieving a faster and more effective protection method compared to PulseEdit.

4) Other:

Xu *et al.* [82] considered the identity privacy issue posed by the body and proposed to modify some body features, e.g., height, weight, and skin color, by putting in place while preserving the usability of body postures. Kuang *et al.* [83] proposed to encode information such as the pose to protect the pedestrian image privacy, which can be visually strict to the original image, but looks natural.

One of the biggest threats of CV to visual content is the detection and extraction of biometric features, especially for faces. Therefore, the study of privacy protection schemes for biometrics has been the focus of the fight against CV adversaries and is able to be subdivided into more directions as described above compared to others. However, it is also possible that the source of privacy for visual content is non-biometric, such as text and content themselves.

D. Solutions for Non-Biometric Protection

1) Scene text:

TABLE III: A summary of the solutions in non-biometric protection.

Sub class	Paper	Year	Key technology	HV		Reversi-bility	Type		Note
				Reality	Invarance		I	V	
Scene text	[84]	2020	Perturbation	☑	☑	☒	☑	☒	White-box
	[85]	2021	Perturbation	☑	☒	☒	☑	☒	White-box
	[86]	2020	Perturbation	☑	☑	☒	☑	☒	White-box
	[44]	2023	Perturbation	☑	☑	☒	☑	☑	Black-box
Content	[87]	2017	Perturbation	☑	☑	☒	☑	☒	Content all protected
	[81]	2020	Perturbation	☑	☑	☒	☑	☒	Optional protected content
	[88]	2021	Perturbation	☑	☑	☒	☑	☒	Dropping information
	[89]	2019	Perturbation	☑	☑	☒	☒	☑	Black-box
	[90]	2021	Perturbation	☒	☑	☑	☑	☒	Superimposing multiple images

Adversarial perturbation for non-sequential visual tasks (e.g., recognition [89] and detection [46]) has been extensively studied in recent years. However, research on adversarial perturbation for sequential tasks (e.g., scene text recognition) is limited.

Yuan *et al.* [84] proposed an adversarial attack scheme based on multi-task learning to sequential tasks for preventing optical character recognition (OCR). Chen *et al.* [85] proposed to disguise the adversarial perturbation as a watermark to attack the OCR model. However, both schemes are for OCR, but optical characters are hard to see in everyday life, and often appear as text contained in natural images. Xu *et al.* [86] proposed an efficient and generalized adversarial method for scene text in the natural image, and attacked 5 state-of-art CV models, which are connectionist-based and attention-based temporal classification, to show the effectiveness of the method. However, the above methods are a white-box attack, which requires a complete understanding of the adversary's CV model, including (but not limited to) structure, parameters, and gradients, which is difficult in practice. Recently, a novel black-box perturbation generation method for scene text is proposed that requires only a prior knowledge of the adversary model output [44].

2) Content:

Liu *et al.* [87] proposed to apply adversarial perturbations to all contents in the image that can be detected by the CV model to make them invisible to the CV, while the scheme allows adjusting parameters to regulate the perturbation strength. However, the scheme indiscriminately applies permutations to all detectable content, reducing the usability of the image. Xue *et al.* [46] proposed a framework to first determine the sensitive information, then determine the location of the sensitive content in the image, and finally use adversarial perturbation to protect the privacy of the sensitive content. Duan *et al.* [88] proposed a novel perspective on adversarial perturbation, i.e., making perturbation by dropping information rather than adding, and experimental results showed successful blocking of CV detection to image content. Jiang *et al.* [89] proposed a block-box adversarial perturbation method for video, which first generates perturbations experimentally

by training the model on an image to reduce the number of queries to the video recognition model in order to increase efficiency. However, as mentioned above, most adversarial perturbation schemes, both white-box and block-box, require a large computational cost, which is not practical for the average user. Rajabi *et al.* [90] proposed a semantic adversarial perturbation scheme that blocks the detection of CV adversaries by superimposing multiple images together using cryptographic ideas, which is almost computationally costless and supports reversibility.

Existing non-biometric protection against CV adversaries is often based on adversarial perturbations, which ensures visual realism and imperceptibility. However, it is often computationally expensive and may be specific to a specific CV model and not valid for other models, i.e., it is not transferable. Although Rajabi *et al.* [90] proposed an extremely low-overhead and transferable adversarial perturbations, the results in terms of visual effects are poor. Future work may require further research on low overhead, transferability and high visual quality to further improve the feasibility of protection.

E. Common Principles

Tables II and III provide the breakdown of the reviewed solutions in biometric and non-biometric protection, respectively, in which HV-reality means that there is no obvious unnatural or false visual content, HV-invariance denotes that the information obtained from original visual content and protected one are the same for HV, and Type-I and V mean image and video, respectively. ☑ represents that this item is full compliant with; ☒ means that this item is not met; ☐ implies that in some cases it may occur. It is should be noted that the fact that HV-invanriance is ☒ does not mean that this is a defense against a HV adversary. An example is presented in Fig. 5, in which HV does perceive that the protected image has changed, but it does not affect HV's ability to correctly perceive that the two identities are the same person.

By reviewing and summarizing the above solutions, 3 common principles can be identified.

High precision. Although CV adversaries can automate the execution of visual content understanding and extraction, they

Fig. 5: Examples (from [68]) about the HV-invanriance is x, not equal to the ability to defend against HV adversaries. Left: the original image; eight: the image that identity is protected. For HV, the perceived content of the protected image does change somewhat compared to the original one, but not enough to affect HV's perception that it is the same identity.

are often designed for specific tasks, i.e., they can only perform a single task using fixed features/contents associated with privacy in the visual content. Based on this, for the privacy protection tasks targeting CV, only the fixed features/contents need to be targeted. Other content unrelated to them can be maintained as is, thus achieving high-precision privacy protection.

Transparent level. It refers to the ability of visual content to be viewed naturally, i.e., without creating an unpleasant experience like blurring, while at the same preventing attacks on privacy by CV adversaries. This is because CV is still not comparable to HV, and its robustness is far worse than HV, and a small change (even imperceptible to the HV) can make CV fail.

High usability. Privacy protection schemes for CV adversaries tend to maintain high usability for CV and HV. This is a natural conclusion that can be drawn based on two principles mentioned above. For high precision, it ensures that non-privacy-related content does not change (or small change), and thus other content is often still understood and detected by CV. For the transparent level, it makes changes difficult to perceive for HV, ensuring that non-privacy content is correctly understood by HV, even for the content that is private to the CV adversary.

IV. PRIVACY PROTECTION FOR HV ADVERSARY

This section provides a comprehensive analysis of privacy concerns posed by HV-based adversaries to visual content, including characteristics, main approaches, and common principles, as shown in Fig. 6. It should be noted that the HV is often very robust compared to CV perception, and it is difficult to change a feature or add noise to ht HV as in the case of CV adversary defense. In this part of the protection solution, it is often performed on objects or contents (which are collectively referred to areas), and thus is not classified according to privacy issues as in the case of CV adversary.

A. Characteristics in HV Adversary Protection

1) Characteristics associated with heavy protection: For heavy protection against the HV adversary, it means that the protected area does not reveal visual information. It is all local protection, as shown in Fig. 7, since if it is a global one, then it translates into confidentiality rather than privacy protection. Meanwhile, heavy protection is generally taken to two extremes, either HV clearly senses that a place has undergone severe abstraction (or distortion), or HV cannot detect it.

- *Imperceptibility.* It is a kind of visual privacy protection against HV adversaries that is imperceptible to the HV or

difficult to detect. In general, this protection is refined and the content in the area is natural to HV and has semantic integrity (at the cost of misleading HV), as shown in the fourth column in Fig. 7.

- *Perceptibility.* It refers to a very obvious perceptible treatment of a content or object, and although the HV knows that it is treated, no information can be extracted from the protected area, as shown in the second and third columns in Fig. 7.

2) Characteristics associated with light protection: For light protection against the HV adversary, it means that the protected area still conveys the general properties of the original content, but the visual details have been erased. Also, how much of this general property is preserved in the protected content can often be adjusted by parameters as shown in Fig. 8. This protection can be localized not only for an area as in heavy protection, but also for the visual content as a whole.

- *Locality.* It is a kind of processing containing rough content for the specific area of visual content, while the rest is maintained, which often occurs in every day, such as online chats and OSNs.
- *Globality.* It mainly considers that the whole of the visual content is related to privacy, and therefore degrades all the content, which means that the protected content is a lower resolution version of the original one.

B. Main Approaches

Synthesis: It has the same basic idea as 'synthesis' in Section III-B, which is to replace privacy-sensitive content. However, the above section tends to be cautious about content changes is not sufficient to form a defense against HV. In this section, synthesis often apply object removal, replacement with abstract content, and regeneration that changes the semantic-level understanding.

Filtering: It is one of the most common methods of visual content privacy protection in daily life due to its simplicity, efficiency, and effectiveness. As shown in Fig. 9, two common filters, blur and pixelation, effects are demonstrated. Blurring is the application of a Gaussian function to the visual content, which uses adjacent pixels to modify each pixel. Pixelation is the process of dividing the visual content into square grids, calculating the average color of the pixels inside each grid, and assigning the color to all pixels in that grid.

Encryption: In cryptography, encryption is the encoding of data in such a way that the original data becomes incomprehensible. Classical image and video encryption also aim to completely eliminate the comprehensibility of the visual content, that is, to guarantee confidentiality. For privacy protection, the complete elimination of comprehensibility is not
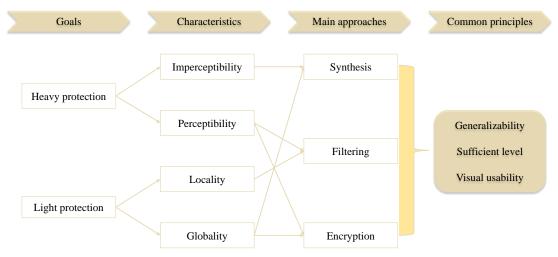
Fig. 6: Overview of the privacy protection analysis for HV adversary.

necessary. It is possible to preserve some comprehensibility of visual content by selecting to encrypt some areas or by visual encryption.

C. Solutions for Heavy Protection

1) Imperceptibility:

Object (or people) removal is aims to protect privacy by hiding objects so that no trace of them appears in the protected visual content [94]. After an object is removed, a gap appears, which needs to be inpainted. Inpainting means reconstructing the missing part using an imperceptible way, i.e., using information from the surrounding area to fill in the missing part.

Barnes et al. [95] proposed to search for similar patches in known regions to synthesize the missing content parts step by step, where the authors designed an approximate nearest-neighbor search to find the closest patches to match. Although it is good at filling high-frequency textures, it cannot get the support of the overall structure of the visual content. Recently, deep convolutional networks (CNNs) have shown great potential for visual content inpainting [94]. CNN-based inpainting techniques can predict the missing parts based on the surrounding environment, and thanks to the increase in computing power, large-scale data training is possible, making it possible to produce semantically sound results. However, the result may lack surprising texture details and be blurry. Yan et al. [96] combined the advantages of patch-based and CNN-based methods, and proposed to apply U-Net [97] as the backbone network and add the shift-connection layer to U-Net. Then, the features of known parts are used to shift to the missing parts, thus generating semantically sound and finely textured inpainting results.

Uittenbogaard et al. [98] are concerned about the millions of images hosted in Street View maps, which contain sensitive objects such faces and license plates. They suggested that the use of methods such as blurring might not allow for adequate protection of privacy, and argued that sensitive objects should be removed. They proposed a privacy-preserving scheme that automatically segments and removes the moving objects in street view maps, and impaints the missing parts. Nakamura et

al. [99] proposed that the text appearing in images may carry private information, and to protect privacy by erasing them. They suggested that instead of extracting the text precisely, only need to know the approximate location of the text, and then apply the CNN model to fill the text area with the color of the surrounding pixels. However, this scheme is an indiscriminate removal of all text from the image and is not flexible. Tursun et al. [100] proposed a conditional GAN with an auxiliary mask, capable of removing all or user-specified text.

More recently, for the protection of scene text against HV adversaries, researchers have proposed to replace text. Roy et al. [101] proposed a font adaptive network for modifying individual characters in the scene text directly on the image, achieving the same structure as the source font and preserving the original color. Then, a work to modify the scene text on a video is proposed, which works by selecting a frame as a reference in which the text is modified, and then transferring the text to other frames using a newly designed network [102]. Yu et al. [103] proposed to apply differential privacy combined with GAN to protect privacy in IoT, where the protection for license plate numbers is mentioned. That is, the original license plate number is extracted as a feature, Laplace noise is added to it, and the rule-compliant license plate number is regenerated.

Imperceptible privacy protection is difficult for adversaries to distinguish between what visual content is protected and what is not. Meanwhile, it also ensures the comfortable viewing of visual content. However, this is a double-edged sword for users, making them equally unaware of whether the content has changed and whether their understanding of the content is correct.

2) Perceptibility:

Filtering. Chinomi et al. [49] proposed a privacy scheme using visual abstraction to protect specific objects in video surveillance, where the relationship between viewer and object is considered. Different viewers will view different levels of visual abstraction, e.g., silhouette and edge, which is a kind of adaptive privacy protection. Orekondy et al. [60] proposed an image redaction scheme for privacy protection. It is able to label multiple privacy regions, which can be considered

Fig. 7: Examples (from [92]) of heavy protection for HV adversaries. From left to right: privacy area, filtering, encryption, and synthesis.

as an auto-tagging task for multiple privacy attributes, and segment them automatically. Then, these regions are completely masked using a heavy filter (like the second column in Fig. 7). Meanwhile, they are then labeled with text in order to make it easier for the user to know what is in them to prevent misinterpretation. For example, the face area is masked and labeled with 'face'. Bonetto *et al.* [105] used filters for surveillance videos captured by small UAVs to process the privacy regions in them, where masking filters were used.

Encryption. Concerned about the invasion of human privacy by surveillance video, Boult [106] proposed an encryption scheme for the human face area, which allows surveillance of a general nature while improving privacy issues, and full access by key in the event of an accident. Dufaux *et al.* [107] considered the format compatibility, and proposed a region-of-interest (ROI) video encryption scheme, based on transform-domain or codestream-domain, suitable for MPEG-4. Users expect different levels of privacy for viewing the images they share in OSNs for different people. Aribilola *et al.* [108] considered the privacy impact of video captured by dynamic surveillance cameras and proposed a low computationally intensive unsupervised learning for accurate motion detection in dynamic surveillance video. Then, the detected sensitive content (i.e., ROI) is encrypted using lightweight Chacha 20 cipher that is specifically used in IoT devices. In addition, He *et al.* [109] proposed to be able to encrypt one or more ROIs of

an image while distributing different keys to different viewers, enabling them to see the image with different decryption levels.

Perceptible protection ensures that HV adversaries are completely unaware of the content of sensitive areas, while ensuring that users know that the areas are protected. However, the effect of heavy protection is very obvious, and viewing is often very unattractive. Meanwhile, the sensitive areas do not reveal any visual information to the user, which sometimes can be a nuisance.

D. Solutions for Light Protection

1) Locality:

Filtering. The large amount of realistic images collected in Google Street View has the huge side effect of potentially carrying private information such as faces and license plates. Therefore, Google applies the blurring filter method to blur the privacy area detected therein to protect privacy [110]. Brkic *et al.* [111] considered the privacy problem due to the presence of the human body in the image and proposed a scheme to blur the face and body, which not only protects biometric but also non-biometric features. Irrelevant faces in live streaming often inevitably appear in the footage, damaging their privacy. Zhou *et al.* [112] proposed a pixelated protection scheme for irrelevant faces in live streaming. The scheme first detects faces on a single frame and generates the corresponding
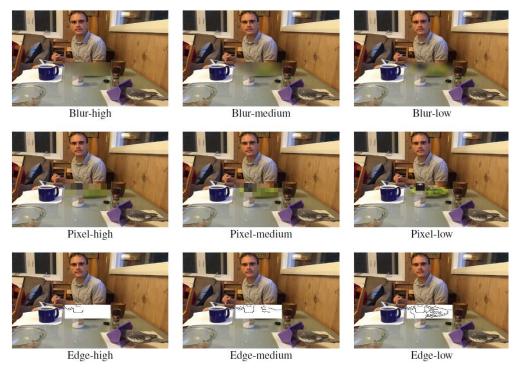
Fig. 8: Examples (from [93]) of light protection adjustable (which is local).
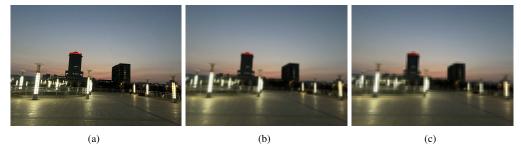

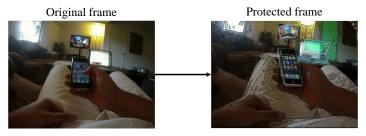Fig. 9: Examples of typical filters. (a) original image; (b) blur; (c) pixelation.

Fig. 10: An example (from [104]) of cartooning to protect privacy.

face vectors, then quickly associates the same faces between frames to generate face track vectors and refine them, and finally pixelates the faces. Sun *et al.* [113] considered the privacy issues associated with videoconferencing, such as videoconferencing recordings to record a face. They proposed the use of the blurring filter to protect privacy.

2) Globality:

Synthesis. Erdlyi *et al.* [114] considered images (e.g., captured by surveillance cameras) as a whole to be private and proposed transforming the original image into an abstract cartoon in order to erase the sensitive details, while this scheme allows the intensity of protection to be changed according to different parameters. Note that the authors claimed to be

able to use it in videos, but it is actually only done on one frame. Subsequently, the authors [115] further considered that different contents in an image have different privacy levels and proposed that the protection intensity is greater for highly privacy-sensitive contents and becomes lower for the rest. Hassan *et al.* [104] proposed a content replacement scheme on video streams to protect privacy, i.e., using cartoonish objects and backgrounds (which erase the details of the original content and are an abstract representation) to replace the original contents. As shown in Fig. 10, everything in the original visual content is transformed into cartoon form, erasing the real details.

Encryption. Traditional visual content encryption often

Fig. 11: An example (from [50]) of tunable balance between privacy and visual usability in TPE schemes.

transforms an image or video into a form without any visual meaning, making it difficult to view and lacking in visual usability. Recently, some new types of encryption with visual effects have been proposed. Bao *et al.* [116] proposed the concept of visually meaningful image encryption, that is, encrypting an image using a traditional scheme and then hiding it within a carrier image using data hiding. As a result, the original image is protected, but the final result is visually appealing. However, visual effects have no relationship to the original image, leaving visual usability still missing. To address this, Zhao *et al.* [30] proposed the primitively visually meaningful image encryption, which means that the encrypted image carries the rough visual information of the original image, without revealing the fine content, to ensure visual usability. Wright *et al.* [117] proposed the idea of thumbnail-preserving encryption (TPE), which means that the thumbnail of the encrypted image is the same or nearly the same as the original thumbnail. Meanwhile, the information preserved in the encrypted image can be easily adjusted according to the parameters as shown in Fig. 11, thus balancing privacy with visual usability. Then, they designed the first TPE scheme by utilizing the permutation-only encryption. However, some cryptographic studies [118] show that permutation-only encryption is not unbreakable. Tajik *et al.* [119] proposed the first TPE scheme with theoretical security support, which applies the substitution-permutation encryption framework commonly used in traditional cryptography, and proved the corresponding security. Based on this, Zhang *et al.* [50] proposed to utilize the idea of data hiding to construct a TPE scheme with security guarantee.

Light protection is relatively better visually and has a better browsing experience compared to heavy protection. However, it also has the obvious side effect that adversaries may also infer the content from the rough information preserved in the protected visual content.

E. Common Principles

Tables IV and V provide the breakdown of the reviewed solutions in heavy and light protection, respectively, in which adjustability means that the extent of the protected area, the content shown can change according to different parameters, or different people can have differentiated visual effect.

By reviewing and summarizing the above solutions, 3 common principles can be identified.

Generalization. In contrast to CV adversaries, protection against HV adversaries tends to specify an area to modify, rather than first extracting features from the visual content, modifying the features, and then regenerating according to the features. This makes protection schemes often have general-

ized characteristics, making that they protect all the details of an entire area, rather than being finely targeted.

Sufficient level. Sufficient protection of visible information about sensitive content is required since HV is far more complex and robust than CV. That is, fine information about sensitive content requires to be erased, and even rough information cannot be preserved if necessary. In general, there are two very different ways to implement this principle: one is to erase sensitive content and generate realistic fake content; the other is to use a visually distinctive method, such as a filter, to obscure sensitive content.

Visual usability. Protection against HV adversaries, whether heavy or light, it preserves a degree of visual usability that otherwise translates into a form of confidentiality protection. For heavy protection, it is all a localized protection, which means that the visual content of the unprotected area can be used normally. For light protection, the protected private content itself comes with some visual usability, but it is the fine content is erased.

V. PRIVACY PROTECTION FOR CV&HV ADVERSARY

This section provides a comprehensive analysis of privacy concerns posed by CV&HV-based adversaries to visual content, including characteristics, main approaches, and common principles, as shown in Fig. 12. Compared to the above two adversaries, this section has relatively little work.

A. Characteristics in HV&CV Adversary Protection

1) Characteristics associated with general usability: For general usability protection, it means that the protected visual content also preserves a certain visual effect to ensure the HV usability and that is not weaker than the CV usability. Meanwhile, CV models may be able to perform specific tasks, e.g., detection, on protected visual content.

- *Target protection.* It refers to finding the privacy area in the visual content to be protected, processing it to remove private information, while preserving certain features to ensure that the CV can perform a specific task.
- *Generalized protection.* It refers to the processing of the visual content as a whole, rather than some specific areas, to remove the maximum amount of private information. It tends to ensure that neither the HV nor CV can access or even reason about the private information, while still supporting other effective analyses and perception.

2) Characteristics associated with specific CV usability: For Specific CV usability protection, it implies that the visual content contains almost no information usable to HV, only usable for a limited CV task or a specific CV model, while this usability often requires prior adversarial training of

TABLE IV: A summary of the solutions in heavy protection.

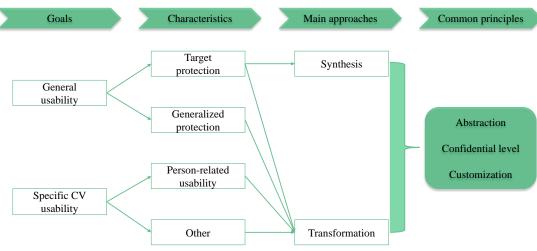| Sub class	Paper	Year	Key technology	Adjustability	Reversibility	Type		Note
						I	V	
Imperceptibility	[95]	2009	Inpainting	✗	✗	✓	✗	Based on patches
	[96]	2018	Inpainting	✗	✗	✓	✗	Based on CNNs
	[98]	2019	Object removal	✗	✗	✓	✗	Focused on view street
	[99]	2017	Text removal	✗	✗	✓	✗	Based on CNNs
	[100]	2019	Text removal	✓	✗	✓	✗	Focused on user-specified text
	[101]	2020	Text modification	✓	✗	✓	✗	Characteristic-level modification
	[102]	2021	Text modification	✓	✗	✗	✓	Frame-level modification
	[103]	2021	GAN	✗	✗	✓	✗	Differential privacy
Perceptibility	[49]	2008	Visual abstraction	✓	✗	✗	✓	Different effects for different users
	[60]	2018	Redaction	✗	✗	✓	✗	Multiple privacy attributes protection
	[105]	2015	Visual abstraction	✓	✗	✗	✓	Focused on drone video
	[106]	2005	Encryption	✗	✓	✗	✓	Focused on face
	[107]	2008	Encryption	✗	✓	✗	✓	Compatible with MPEG-4
	[108]	2022	Encryption	✗	✓	✗	✓	Based on Chacha20 cipher
	[109]	2016	Encryption	✓	✓	✓	✗	Different effects for different viewers

Fig. 12: Overview of the privacy protection analysis for CV&HV adversary.

CV using protected data. This part considers the dilemma of wanting a CV to be able to perform an important task while preventing it from gaining unnecessary information that could lead to a privacy breach.

- *Person-related usability.* It considers the usability associated with people in visual content, in which CV performs recognition of human body behavior, faces, etc. The visual content in this sub-part is often considered surveillance video, which needless to say also contains a lot of private information, and also requires that they also have the need to analyze the necessary information about in it, such as action recognition.
- *Other.* It is protects not only the privacy in a way, but also the security of the data itself, preventing the data from being used by unauthorized CV models or from being to perform unintended tasks.

B. Main Approaches

Synthesis: Compared to the 'synthesis' in the two sections above, the synthesis in this section is more radical, like an abstraction, i.e., all unnecessary information is removed, e.g., color, texture, and shape, and only the information necessary to perform a certain task is preserved in the re-generated content. It is important to note that the necessary information preserved is not only usable to CV, but can also be used by HV.

Transformation: It is meant to convert data from one format to another one. In this part, the original visual content is often converted into a different style to erase all real details

TABLE V: A summary of the solutions in light protection.

Sub class	Paper	Year	Key technology	Adjustability	Reversibility	Type		Note
						I	V	
Locality	[110]	2009	Blur	☒	☒	☑	☒	Focused on view street
	[111]	2017	Blur	☑	☒	☑	☒	Focused on body and face
	[112]	2021	Pixelation	☒	☒	☒	☑	Focused on live streaming
	[113]	2022	Blur	☒	☒	☒	☑	Focused on video conference
Globality	[114]	2014	Cartoonli-zation	☑	☒	☑	☒	Homogeneous protection
	[115]	2014	Cartoonli-zation	☑	☒	☑	☒	Differentiated protection
	[104]	2017	Cartoonli-zation	☑	☒	☒	☑	Objection replacement
	[116]	2015	Encryption	☒	☑	☑	☒	Visual effect irrelevant
	[30]	2022	Encryption	☑	☑	☑	☒	Visual effect relevant
	[117]	2015	Encryption	☑	☑	☑	☒	Insecurity
	[119]	2019	Encryption	☑	☑	☑	☒	Security guarantee
	[50]	2022	Encryption	☑	☑	☑	☒	Based on data hiding

or directly into a form that is not perceptible to HV while preserving the necessary CV usable information.

C. Solutions for General Usability

1) Target protection:

Synthesis. Zou *et al.* [120] considered the importance of pedestrian action in videos, but also carry unnecessary information such as identity. They proposed to protect their privacy while supporting action recognition by replacing human bodies with their pose models. Kunchala *et al.* [121] likewise noted that the large amount of pedestrian privacy contained in the video and proposed using avatars to replace real bodies. The avatar was synthesized using a 3D body while integrated with digital inpainting, erasing all personal information while retaining each person's presence, pose, shape, and background information. Yang *et al.* [122] proposed the use of digital masks to ensure patient face privacy while preserving the necessary information needed for diagnosis. Experiments in disease diagnosis have shown comparable accuracy using digital masks and raw images.

Transformation. Brki *et al.* [123] proposed a scheme for de-identifying humans in surveillance videos, while supporting CV detection and segmentation to humans. Specifically, human locations are first detected using background subtraction based on Gaussian mixture modes, and then rendered using a deep neural network to remove biometric and non-biometric features about the person (including face, clothes, hair, etc.). This makes face recognition and identity detection no longer feasible, but retains the role of overall human body detection and segmentation.

Targeted protection is suitable for scenarios where a clear protection target is known, and where best efforts are made to remove the privacy of the target while preserving its specific usability. However, in real-world scenarios, privacy is often unexpected, meaning that it is difficult to say that there is no privacy threat to something outside a specific target, and targeted protection to deal with such threats is difficult.

2) Generalized protection:

Ryoo *et al.* [126] proposed that extreme low-resolution can be used to anonymize video and prevent CV and HV to recognize the identity of the person appeared in the video. Meanwhile, the paradigm of inverse super resolution was proposed to ensure the usability of video for activity recognition. Specifically, the aim of the paradigm is to perform a learning transformation from a set of high-resolution images such that the low-resolution images which is generated maintain a comparable amount of information as the high-resolution ones. The experimentation verified that the scheme is successful in activity recognition. Hinojosa *et al.* [127] considered the direct design of privacy-preserving cameras that capture only useful information to sense people in the scene while hiding other privacy-sensitive information, making the captured image also an extreme low resolution.

Wu *et al.* [125] proposed a novel framework of the adversarial training to balance privacy and usability of video content, achieving privacy-preserving action recognition. However, this adversarial training requires training data containing privacy labels, which requires a huge cost to annotate the raw data. Meanwhile, the scheme based on privacy label training are difficult to generalize since privacy may still be implied in other information. Dave *et al.* [124] proposed a self-supervised learning to remove privacy information from videos without privacy labels, and meanwhile, it also supports action recognition. A general overview of the two schemes is shown in Fig. 13, where it can be seen that both can achieve privacy-preserving action recognition, but the scheme [124] does not require privacy labels.

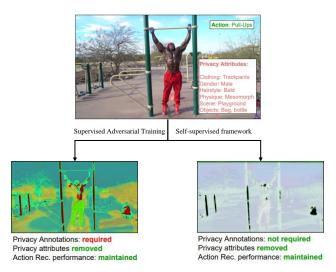Wu *et al.* [47] designed a reversible transformation for

Fig. 13: Overview (from [124]) of the privacy-preserving action recognition scheme [125] (left) and [124] (right). Both schemes enable action recognition, but [125] requires labels and [124] does not.

video privacy protection supporting analytics, in which a combination of CycleGAN [128] and the key is utilized to ensure secure reversion. The experiment demonstrated that the scheme supports real-time operation, which enables the video recorded by the camera to be processed and transmitted to the analysis end in real time.

Generalized protection is nearly as powerful for privacy protection compared to target protection, erasing almost all visual information except that related to usability. On the other hand, in contrast to specific protection, this part of the scheme mostly requires adversarial training to enable CV models to perform specific tasks using the preserved information.

D. Solutions for Specific CV Usability

1) Person-related usability:

Wu *et al.* [129] proposed an adversarial training framework to achieve privacy-preserving CV recognition, in which the original video is transformed into a visually almost imperceptibly degraded version. The framework explicitly learns the degraded transform for the original video inputs to optimize the trade-off between the target performance and the privacy budget (introducing differential privacy) on the degraded video. Experiments showed that the risk of privacy leakage is effectively suppressed while action recognition is achieved. Wang *et al.* [130] proposed a lens-free coded aperture camera system for privacy protection while enabling the human action recognition, where the protected visual content is unintelligible and does not reveal any information. This scheme applies the coded aperture principle [136], which is capable of erasing visual features, while the authors utilized deep neural networks to use the coded aperture data for action recognition. Liu *et al.* [131] proposed a fall detection system with visual shielding, which can effectively detect the occurrence of a fall while achieving a home camera that cannot record additional information. It implements visual shielding by multi-layer compressed sensing, then proposes a corresponding feature extraction algorithm to extract human behavior features, and finally implements fall detection based on GAN.

Facial regions have been the focus of research on privacy-preserving and usability recognition, and there is also some work supporting specific CV usability that focuses on facial regions. Gupta *et al.* [132] proposed that when performing a specific CV task, only locally necessary information relevant to the task is required, while all other information can be erased. Experiments showed that it can successfully perform facial emotion and attribute recognition after erasing all non-essential information. Ji *et al.* [133] transformed the original image to the frequency domain and then removed the DC coefficients to add random perturbation in the framework of differential privacy to provide strong protection for facial privacy. Meanwhile, adversarial training is performed to achieve face recognition in the privacy-preserving case.

Person-related usability has been the focus of CV research. The above studies have tried their best to remove visual information while preserving person-related CV usability. However, it is important to note that while person-related usability/privacy is a hot topic, there is a lot of other visual content that is also integral to CV usability.

2) Other:

Outsourcing visual content data for AI-related detection and inference is one of the hot directions, but it also raises the risk of data privacy, e.g., data may be used for unintended tasks. Wu *et al.* [134] proposed to minimize the unnecessary pixel entropy before delivering the image data for outsourcing. They used a neural network to train a model, and the trained model can transform the input image to remove irrelevant information on demand, while ensuring good inference accuracy for a specific CV task. Ye *et al.* [48] proposed a protective perturbation generator to protect the privacy of outsourced images, which can effectively eliminate the visual information in the images and prevent privacy leakage without affecting the target recognition model. Meanwhile, it does not require the outsourced model to be retrained. In addition, Aprilpyone *et al.* [135] proposed three image transformation algorithms based on cryptographic ideas. These protected images are utilized for adversarial training and the resulting model is able to classify the images effectively. Meanwhile, the visual

TABLE VI: A summary of the solutions in general usability.

Sub class	Paper	Year	Key technology	Adversarial training for task	Reversibility	Type		CV task
						I	V	
Target protection	[120]	2022	Postural replacement	☒	☒	☒	☑	Person segmentation
	[121]	2023	3D replacement	☒	☒	☒	☑	Pedestrian analysis
	[122]	2022	Digital mask	☒	☒	☑	☒	Disease diagnosis
	[123]	2017	Gaussian mixture modes	☒	☒	☒	☑	Person segmentation
Generalized protection	[126]	2017	Inverse super resolution	☑	☒	☒	☑	Action recognition
	[127]	2021	Optical encoder	☑	☒	☑	☒	Human pose estimation
	[125]	2022	Supervised learning	☑	☒	☒	☑	Action recognition
	[124]	2022	Self-supervised learning	☑	☒	☒	☑	Action recognition
	[47]	2021	CycleGAN	☒	☑	☒	☑	Detection

TABLE VII: A summary of the solutions in specific CV usability.

Sub class	Paper	Year	Key technology	Adversarial training for task	Reversibility	Type		CV task
						I	V	
Person-related usability	[129]	2018	Degradation	☑	☒	☒	☑	Action recognition
	[130]	2019	Coded aperture	☑	☒	☒	☑	Action recognition
	[131]	2021	Compressed sensing	☑	☒	☒	☑	Fall detection
	[132]	2021	Masking	☑	☒	☑	☒	Face recognition
	[133]	2022	Frequency domain transform	☑	☒	☑	☒	Face recognition
Other	[134]	2021	Reducing pixel entropy	☑	☒	☑	☒	Inference
	[48]	2022	Perturbation	☒	☒	☑	☒	Recognition
	[135]	2021	Encryption	☑	☒	☑	☒	Classification

information contained in the image is effectively protected or even erased.

Overall, this part mitigates the privacy risks associated with outsourcing visual content data while preserving the usability of specific tasks (e.g., recognition and classification). This may make people more willing to share visual content with third-party services, and to some extent may help solve the data silo problem. However, they are often targeted at coarse-grained CV tasks and may be powerless for more detailed tasks, such as face recognition.

E. Common Principles

Tables VI and VII provide the breakdown of the reviewed solutions in general usability protection and specific CV usability, respectively, in which adversarial training for tasks means that the model needs to be trained on protected visual content in order to perform a specific CV task.

By reviewing and summarizing the above solutions, 3 common principles can be identified.

Abstraction. Since this part considers not only the defense against HV and CV adversaries, but also the usability of CV, it is often necessary to extract the valid information from the visual content while other unnecessary information is removed. It can be seen as a high level of abstraction of visual content.

Confidential level. As mentioned above, this part of the protection of the visual content tends to remove all unnecessary information, leaving only the key information necessary to perform the CV task. Therefore, for almost protection schemes, it is difficult to obtain much useful information about privacy content from HV alone, let alone for CV tasks and models that are not expected.

Customization. CV usability in this part is often a customization, that is, a customized protection of the visual content in order to take a specific task, ensuring that the information left is sufficient to perform that task. Even the CV models used need to be trained adversarially based on the protected visual content to obtain a customized model for performing that task.

VI. CHALLENGES AND FUTURE DIRECTIONS

There is no doubt that privacy has been taken very seriously across all communities, from legal [16] to engineering [137], from politics [138] to science [139], and a number of visual

privacy solutions are rapidly being proposed to address or mitigate various privacy risks. However, there are several foundational or important problems about privacy protection that need to be considered as priorities. In this section, we summarize what we see as the main challenges related to the privacy protection of visual content.

A. What is privacy?

All privacy protection schemes talk about privacy, but they tend to assume that something is private, such as faces, and that processing it is the same as completing privacy protection. However, is this really a privacy protection? Or is it just an objective visual content processing task? These schemes do not care about defining the concept of privacy, but go straight to talking about privacy protection, and thus the protection of privacy is often difficult to put in place.

In fact, there is still no universal conclusion about the concept of privacy. Just as Thomson said [20] that "*Perhaps the most striking thing about the right to privacy is that nobody seems to have any clear idea of what it is*". Privacy is significantly cross-disciplinary and interdisciplinary [140]. Research on privacy involves sociology, law, political science, philosophy, ethics, economics, etc [14]. These studies and discourses differ significantly in their definitions of the concept of privacy, making the findings vary widely and even fundamentally opposed.

While some scholars have attempted to provide some definitions of privacy in different dimensions, it tends to be highly generalized. For example, a scholar [141] argued that "*Interest that individuals have in sustaining personal space, free from interference by other people and organizations*". This may seem right, but it is difficult to use it to clarify what is privacy in practice.

If we can clarify what privacy is, it can bring a huge improvement to privacy protection in at least 3 ways. *1)* Privacy can be fully, automatically, and adaptively protected, rather than sporadically for a particular attribute or object, as it is today. *2)* Privacy protection can be measured, quantified, and even made into a science, rather than an art that can only be verified using extensive experiments. *3)* A theoretically optimal balance between privacy protection and usability can be further approached without worrying that too much protection will spill over into irrelevant usability and too little protection will not eliminate privacy.

B. Privacy and Usability

Some scholars pointed out that privacy should be a personal interest [142], [141]. This means that privacy needs to be a trade-off against other interests, which may be personal, political, social, or technological [39]. In fact, privacy protection is not often the primary goal (which is usability) of the user, but a secondary one [143]. Privacy protection for visual content is meaningless if it completely compromises usability. For example, the purpose of a user posting an image on an OSN is for a certain visual viewability for others. If this viewability is completely erased, it ensures that the visual content does not compromise privacy, but then does the user still need to post the image?

The trade-off between data privacy and usability is always an open challenge for privacy protection [9], especially for visual content because of the diversity and complexity of information it contains. In general, the greater the privacy, the worse the usability, and vice versa, as shown in Fig. 11. Ideally, privacy protection and usability should be separated from each other. That is, only private information is processed, without affecting any extraneous information to provide usability.

There are two difficulties in this separation of privacy and usability. *1)* In theory, what privacy is is still inconclusive as mentioned above, and it is difficult to separate things that cannot be defined. *2)* In technology, it is difficult to extract information accurately as well as to process the target information without affecting others.

It should be noted that researchers have been advancing in terms of technology. Especially on the face, precise and controlled extraction, processing, and synthesis techniques have been investigated [144], [145], [146]. However, they still inevitably change some attributes irrelevant to the target (privacy), but of course, they are more precise than previous work.

C. Recognition of Protection Area

Visual content privacy protection either modifies all the areas or circles only one (or more) ROI(s) to modify. Almost all ROI protection reviewed in this survey either assumes that this is related to a particular object, such as face, or is directly hand-drawn. In fact, the privacy objects in different visual contents may not be the same and are very diverse. Meanwhile, humans' perception of privacy in visual content is often not that comprehensive, thus leading to making privacy decisions about ROI that are often prone to errors. It is also very difficult and time-consuming for humans to manually label privacy for all visual content [147].

Some privacy prediction methods have also been proposed to automate the process of helping humans to identify the privacy of visual content [147], [148] or to give privacy labels [149]. However, these methods often still have a huge database marked with what is private content and what is otherwise [149]. Alternatively, privacy is assessed based on pre-specified privacy attributes [**?**], [148], [147]. This means that if it is not in databases, or if a certain type of information is not pre-specified, then it cannot be correctly predicted either.

It is also important to note that the existing non-manual privacy recognition may have a certain error rate. If privacy is not properly recognized as well as circled, then privacy protection will fail. Meanwhile, for video, if one frame is not circled, then the rest of the effect is greatly diminished. Therefore, how to reduce the probability of CV misidentifying privacy areas and improve the robustness of detection, while recognizing as many kinds of privacy as possible is crucial to automate the recognition of protected areas.

D. People-Oriented Protection

The topic of privacy has always been of great interest, and images and videos in particular, because of the rich and visual

information they can carry, have led to a constant stream of advanced schemes for visual privacy protection. On the other hand, in real life, it may be rare to see so many advanced schemes in use, and the vast majority of ordinary people (non-academics) are not even aware of them. They still utilize seemingly backward ways (pixelation and blurring) to protect visual privacy. Even, many of the more recently proposed advanced schemes (e.g., [68], [69], [83]) still choose these age-old methods as a comparison since they are simply too common.

Compared to advanced schemes, pixelation/blurring has two huge advantages: 1) no need for the user to master any expertise, anyone can simply use it to get the obvious protection effect; 2) no high computational cost, ordinary devices can be executed in real time [90]. These two points can be described as a kind of people-oriented protection. Many advanced schemes do not consider the people-oriented characteristics, especially for CV adversaries. For example, adversarial perturbation schemes can be divided into white-box and black-box. The white-box schemes require the user to have the full knowledge of the adversarial CV model, including structure, parameters, and output classes. This is not to mention the ordinary people, professionals may also be difficult to fully know the knowledge of the adversary. The black-box schemes require extensive testing, which entails a large computational cost that is difficult to support with the ordinary devices that the user has, and they also require the user to have expertise in training the model.

If this people-oriented characteristic is implemented in a protection scheme, it may help the scheme to become rapidly popular in daily life. Meanwhile, it should be noted that privacy protection itself should be people-oriented, since it is intended to help the individual to protect privacy. If the scheme is not easily used by individuals but requires expertise and a lot of effort, it is undesirable and even difficult to accept in practice, even if it has excellent results in the laboratory.

VII. CONCLUSION

Vision is an important source of human perception of the world, always faster and more intuitive than other ways of obtaining information. Therefore, visual content always conveys information quickly and accurately, and thus people are willing to use visual content data to record and share their lives. However, the convenience of visual content comes with serious privacy concerns and is only expected to increase over time. A great deal of research has been done to address or mitigate these concerns, making the benefits of visual content usable to individuals while also taking steps to minimize the privacy impact of visual content. That is, these researches attempt to strike a balance between privacy and usability of visual content.

This survey aims to provide a comprehensive overview of research related to the privacy protection of visual content. Visual content privacy protection is the collective term for solutions that address or mitigate the privacy concerns that arise from the visual effects of image and video content. First, the characteristics of visual content privacy and the difficulty

and challenge of protection are introduced. Then, for the criterion that privacy protections have adversaries, a high-level privacy protection framework based on the type of adversary is proposed. Any privacy protection scheme will be able to find a corresponding classification in this framework. For each of these classifications, we reviewed the corresponding solutions, summarized and compared them, and also proposed their common principles. Finally, the open challenges and future directions for visual privacy protection are discussed.

This survey explores a comprehensive picture of visual privacy protection, and privacy adversaries and solutions are included in this survey. We hope that our work will provide a macro perspective on current research, contribute to the development of the visual privacy protection community, and provide assistance for individual privacy protection needs.

Finally, we also suggest to service providers (e.g., cloud, surveillance, and CV's API) who process visual content that they take privacy protection seriously. For the public, their awareness of privacy continues to grow, and they are increasingly intolerant of a service with privacy concerns. For example, in 2021, Apple tried to promote the detection of images of child sexual violence on iCloud [150], which sounded reasonable. However, many people immediately attacked it since it was, in fact, a kind of surveillance system on the device, which seriously violated the privacy of users. For this reason, the system was announced to be canceled by Apple before it had a chance to go live. This means that if service providers do not take privacy seriously, their products may be boycotted by users before they even have a chance to get distributed. Meanwhile, laws, regulations, and standards, e.g., EU's GDPR [16], Illinois's Biometric Information Privacy Act [151], and ISO/IEC 30137-1:2019 [152], are also becoming clearer and stricter on privacy issues, with less and less ambiguity. This also means that privacy violations can lead to investigations, questioning, and large fines in some countries. In short, privacy should now be rapidly gaining ground in service providers, and while it may not make a mediocre product successful, it can make a good product discarded by both the public and the government.

REFERENCES

[1] F. Hutmacher, "Why is there so much more research on vision than on any other sensory modality?" *Front. Psychol.*, vol. 10, 2019.

[2] R. L. Solso, M. K. MacLin, and O. H. MacLin, *Cognitive psychology*. Pearson Education New Zealand, 2005.

[3] R. J. Gerrig, P. G. Zimbardo, A. J. Campbell, S. R. Cumming, and F. J. Wilkes, *Psychology and life*. Pearson Higher Education AU, 2015.

[4] A. Reed, "Visual content vs text content — epic face-off with obvious winner," 2021, https://www.motocms.com/blog/en/visual-content-vs-text-content/.

[5] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, 2016.

[6] H. Gernsheim, "The 150th anniversary of photography," *Hist. Photogr.*, vol. 1, no. 1, pp. 3–8, 1977.

[7] S. J. Nightingale, K. A. Wade, and D. G. Watson, "Can people identify original and manipulated photos of real-world scenes?" *Cogn. Res.*, vol. 2, no. 1, pp. 1–21, 2017.

[8] M. Broz, "Number of photos (2023): Statistics, facts, & predictions," 2023, https://photutorial.com/photos-statistics/.

[9] C. Liu, T. Zhu, J. Zhang, and W. Zhou, "Privacy intelligence: A survey on image privacy in online social networks," *ACM Comput. Surv.*, vol. 55, no. 8, pp. 1–35, 2023.

[10] T. Winkler and B. Rinner, "Security and privacy protection in visual sensor networks: A survey," *ACM Comput. Surv.*, vol. 47, no. 1, 2014.

[11] D. McDuff, "Camera measurement of physiological vital signs," *ACM Comput. Surv.*, vol. 55, no. 9, 2023. [Online]. Available: https://doi.org/10.1145/3558518

[12] W. Qi and H. Su, "A cybertwin based multimodal network for ecg patterns monitoring using deep learning," *IEEE Trans. Ind. Inform.*, vol. 18, no. 10, pp. 6663–6670, 2022.

[13] M. Kim, A. K. Jain, and X. Liu, "Adaface: Quality adaptive margin for face recognition," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn.*, June 2022, pp. 18 750–18 759.

[14] K. O'Hara, "The seven veils of privacy," *IEEE Internet Comput.*, vol. 20, no. 2, pp. 86–91, 2016.

[15] T. Wang, J. Zhou, X. Chen, G. Wang, A. Liu, and Y. Liu, "A three-layer privacy preserving cloud storage scheme based on computational intelligence in fog computing," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 2, no. 1, pp. 3–12, 2018.

[16] P. Voigt and A. Von dem Bussche, "The EU general data protection regulation (GDPR)," *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, vol. 10, no. 3152676, pp. 10–5555, 2017.

[17] S. L. Pardau, "The california consumer privacy act: Towards a european-style privacy regime in the united states," *J. Tech. L. & Pol'y*, vol. 23, p. 68, 2018.

[18] U. G. Assembly *et al.*, "Universal declaration of human rights," *UN General Assembly*, vol. 302, no. 2, pp. 14–25, 1948.

[19] G. Elkoumy, S. A. Fahrenkrog-Petersen, M. F. Sani, A. Koschmider, F. Mannhardt, S. N. n. Von Voigt, M. Rafiei, and L. V. Waldthausen, "Privacy and confidentiality in process mining: Threats and research challenges," *ACM Trans. Manage. Inf. Syst.*, vol. 13, no. 1, 2021.

[20] J. J. Thomson, "The right to privacy," *Philos. Public Aff.*, vol. 4, no. 4, pp. 295–314, 1975. [Online]. Available: http://www.jstor.org/stable/2265075

[21] J. R. Padilla-Lpez, A. A. Chaaraoui, and F. Flrez-Revuelta, "Visual privacy protection methods: A survey," *Expert Syst. Appl.*, vol. 42, no. 9, pp. 4177–4195, 2015.

[22] N. J. Hum, P. E. Chamberlin, B. L. Hambright, A. C. Portwood, A. C. Schat, and J. L. Bevan, "A picture is worth a thousand words: A content analysis of facebook profile photographs," *Comput. Hum. Behav.*, vol. 27, no. 5, pp. 1828–1833, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563211000690

[23] S.-F. Chang, J. R. Smith, M. Beigi, and A. Benitez, "Visual information retrieval from large distributed online repositories," *Commun. ACM*, vol. 40, no. 12, pp. 63–71, 1997.

[24] P. Kok, G. J. Brouwer, M. A. van Gerven, and F. P. de Lange, "Prior expectations bias sensory representations in visual cortex," *J. Neurosci.*, vol. 33, no. 41, pp. 16 275–16 284, 2013. [Online]. Available: https://www.jneurosci.org/content/33/41/16275

[25] T. Denning, K. Bowers, M. van Dijk, and A. Juels, "Exploring implicit memory for painless password recovery," in *Conf. Hum. Fact. Comput. Syst. Proc.*, 2011, p. 26152618. [Online]. Available: https://doi.org/10.1145/1978942.1979323

[26] D. R. Hilbert, *Color and Color Perception: A Study in Anthropocentric Realism*. Csli Press, 1987.

[27] T. I. Panagiotaropoulos, V. Kapoor, and N. K. Logothetis, "Subjective visual perception: from local processing to emergent phenomena of brain activity," *Philos. Trans. R. Soc. B-Biol. Sci.*, vol. 369, no. 1641, p. 20130534, 2014.

[28] P. Liu, X. Wang, and Y. Su, "Image encryption via complementary embedding algorithm and new spatiotemporal chaotic system," *IEEE Trans. Circuits Syst. Video Technol., in press, doi: 10.1109/TCSVT.2022.3222559*, 2022.

[29] B. Tang, C. Yang, and Y. Zhang, "A format compliant framework for hevc selective encryption after encoding," *IEEE Trans. Circuits Syst. Video Technol., in press, doi: 10.1109/TCSVT.2022.3212865*, 2022.

[30] R. Zhao, Y. Zhang, Y. Nan, W. Wen, X. Chai, and R. Lan, "Primitively visually meaningful image encryption: A new paradigm," *Inf. Sci.*, vol. 613, pp. 628–648, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025522009100

[31] T. Fang, Y. Qi, and G. Pan, "Reconstructing perceptive images from brain activity by shape-semantic gan," in *Adv. neural inf. proces. syst.*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 13 038–13 048. [Online]. Available: https://proceedings.neurips.cc/paper/2020/file/9813b270ed0288e7c0388f0fd4ec68f5-Paper.pdf

[32] B. Zhang, "Computer vision vs. human vision," in *Proc. IEEE Int. Conf. Cogn. Informatics Cogn. Comput.*, 2010, pp. 3–3.

[33] W. J. Scheirer, S. E. Anthony, K. Nakayama, and D. D. Cox, "Perceptual annotation: Measuring human vision to improve computer vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1679–1686, 2014.

[34] D. Botina-Monsalve, Y. Benezeth, and J. Miteran, "RTrPPG: An ultra light 3DCNN for real-time remote photoplethysmography," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn. Workshops*, June 2022, pp. 2146–2154.

[35] Z. Yang, H. Wang, and F. Lu, "Assessment of deep learning-based heart rate estimation using remote photoplethysmography under different illuminations," *IEEE T. Hum.-Mach. Syst.*, vol. 52, no. 6, pp. 1236–1246, 2022.

[36] C. . Casado, M. L. Caellas, and M. B. Lpez, "Depression recognition using remote photoplethysmography from facial videos," *IEEE Trans. Affect. Comput., in press, doi: 10.1109/TAFFC.2023.3238641*, 2023.

[37] P. Marks, "Can the biases in facial recognition be fixed; also, should they?" *Commun. ACM*, vol. 64, no. 3, p. 2022, 2021. [Online]. Available: https://doi.org/10.1145/3446877

[38] S. Ribaric, A. Ariyaeeinia, and N. Pavesic, "De-identification for privacy protection in multimedia content: A survey," *Signal Process.-Image Commun.*, vol. 47, pp. 131–151, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0923596516300856

[39] B. Meden, P. Rot, P. Terhrst, N. Damer, A. Kuijper, W. J. Scheirer, A. Ross, P. Peer, and V. truc, "Privacyenhancing face biometrics: A comprehensive survey," *IEEE Trans. Inf. Forensic Secur.*, vol. 16, pp. 4147–4183, 2021.

[40] C. Patsakis, A. Zigomitros, A. Papageorgiou, and A. Solanas, "Privacy and security for multimedia content shared on osns: Issues and countermeasures," *The Computer Journal*, vol. 58, no. 4, pp. 518–535, 2015.

[41] M. R. Hasan, R. Guest, and F. Deravi, "Presentation-level privacy protection techniques for automated face recognition - a survey," *ACM Comput. Surv.*, feb 2023, just Accepted. [Online]. Available: https://doi.org/10.1145/3583135

[42] B. Liu, M. Ding, S. Shaham, W. Rahayu, F. Farokhi, and Z. Lin, "When machine learning meets privacy: A survey and outlook," *ACM Comput. Surv.*, vol. 54, no. 2, 2021. [Online]. Available: https://doi.org/10.1145/3436755

[43] Y. Zhao and J. Chen, "A survey on differential privacy for unstructured data content," *ACM Comput. Surv.*, vol. 54, no. 10s, sep 2022. [Online]. Available: https://doi.org/10.1145/3490237

[44] Y. Xu, P. Dai, Z. Li, H. Wang, and X. Cao, "The best protection is attack: Fooling scene text recognition with minimal pixels," *IEEE Trans. Inf. Forensic Secur., in press, doi: 10.1109/TIFS.2023.3245984*, 2023.

[45] M. Maximov, I. Elezi, and L. Leal-Taixe, "CIAGAN: Conditional identity anonymization generative adversarial networks," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn.*, June 2020.

[46] H. Xue, B. Liu, M. Din, L. Song, and T. Zhu, "Hiding private information in images from ai," in *IEEE Int. Conf. Commun.*, 2020, pp. 1–6.

[47] H. Wu, X. Tian, M. Li, Y. Liu, G. Ananthanarayanan, F. Xu, and S. Zhong, "Pecam: Privacy-enhanced video streaming and analytics via securely-reversible transformation," in *Proc. Annu. Int. Conf. Mobile Comput. Networking*, 2021, p. 229241. [Online]. Available: https://doi.org/10.1145/3447993.3448618

[48] M. Ye, Z. Tang, H. Phan, Y. Xie, B. Yuan, and S. Wei, "Visual privacy protection in mobile image recognition using protective perturbation," in *Proc. ACM Multimed. Syst. Conf.*, 2022, p. 164176. [Online]. Available: https://doi.org/10.1145/3524273.3528189

[49] K. Chinomi, N. Nitta, Y. Ito, and N. Babaguchi, "Prisurv: Privacy protected video surveillance system using adaptive visual abstraction," in *Advances in Multimedia Modeling*, S. Satoh, F. Nack, and M. Etoh, Eds., 2008, pp. 144–154.

[50] Y. Zhang, R. Zhao, X. Xiao, R. Lan, Z. Liu, and X. Zhang, "HF-TPE: High-fidelity thumbnail- preserving encryption," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 947–961, 2022.

[51] D. Osorio-Roig, C. Rathgeb, P. Drozdowski, P. Terhrst, V. truc, and C. Busch, "An attack on facial soft-biometric privacy enhancement," *IEEE trans. biom. behav. identity sci.*, vol. 4, no. 2, pp. 263–275, 2022.

[52] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep face recognition: A survey," in *SIBGRAPI Conf. Graph., Patterns Images*, 2018, pp. 471–478.

[53] P. Dhar, J. Gleason, A. Roy, C. D. Castillo, and R. Chellappa, "PASS: Protected attribute suppression system for mitigating bias in face recognition," in *Proc. IEEE Int. Conf. Comput. Vision*, October 2021, pp. 15 087–15 096.

[54] M. Chen, X. Liao, and M. Wu, "PulseEdit: Editing physiological signals in facial videos for privacy protection," *IEEE Trans. Inf. Forensic Secur.*, vol. 17, pp. 457–471, 2022.

[55] Q. Zhu, M. Chen, C.-W. Wong, and M. Wu, "Adaptive multi-trace carving for robust frequency tracking in forensic applications," *IEEE Trans. Inf. Forensic Secur.*, vol. 16, pp. 1174–1189, 2021.

[56] J. Du, S.-Q. Liu, B. Zhang, and P. C. Yuen, "Weakly supervised rppg estimation for respiratory rate estimation," in *Proc. IEEE Int. Conf. Comput. Vision Workshops*, October 2021, pp. 2391–2397.

[57] Z. Zhu, X. Guo, T. Yang, J. Huang, J. Deng, G. Huang, D. Du, J. Lu, and J. Zhou, "Gait recognition in the wild: A benchmark," in *Proc. IEEE Int. Conf. Comput. Vision*, October 2021, pp. 14 789–14 799.

[58] M. Tran, T. Sen, K. Haut, M. R. Ali, and E. Hoque, "Are you really looking at me? a feature-extraction framework for estimating interpersonal eye gaze from conventional video," *IEEE Trans. Affect. Comput.*, vol. 13, no. 2, pp. 912–925, 2022.

[59] X. Wang, Y. Jiang, Z. Luo, C.-L. Liu, H. Choi, and S. Kim, "Arbitrary shape scene text detection with adaptive text region representation," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn.*, June 2019.

[60] T. Orekondy, M. Fritz, and B. Schiele, "Connecting pixels to privacy and utility: Automatic redaction of private information in images," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn.*, June 2018.

[61] A. Ilyas, S. Santurkar, D. Tsipras, L. Engstrom, B. Tran, and A. Madry, "Adversarial examples are not bugs, they are features," in *Adv. neural inf. proces. syst.*, vol. 32, 2019. [Online]. Available: https://proceedings.neurips.cc/paper/2019/file/e2c420d928d4bf8ce0ff2ec19b371514-Paper.pdf

[62] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar, "Face swapping: Automatically replacing faces in photographs," in *ACM SIGGRAPH 2008 Papers*, 2008. [Online]. Available: https://doi.org/10.1145/1399504.1360638

[63] S. Mosaddegh, L. Simon, and F. Jurie, "Photorealistic face de-identification by aggregating donors' face components," in *Computer Vision – ACCV 2014*, Cham, 2015, pp. 159–174.

[64] Y. Lin, S. Wang, Q. Lin, and F. Tang, "Face swapping under large pose variations: A 3d model based approach," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2012, pp. 333–338.

[65] Y. Nirkin, Y. Keller, and T. Hassner, "FSGAN: Subject agnostic face swapping and reenactment," in *Proc. IEEE Int. Conf. Comput. Vision*, October 2019.

[66] Deepfakes, "Faceswap," 2017, .https://github.com/deepfakes/faceswap.

[67] B. Zhu, H. Fang, Y. Sui, and L. Li, "Deepfakes for medical video de-identification: Privacy protection and diagnostic information preservation," in *Proc. AAAI/ACM Conf. AI, Ethics, Soc.*, 2020, p. 414420. [Online]. Available: https://doi.org/10.1145/3375627.3375849

[68] Y. Wen, B. Liu, M. Ding, R. Xie, and L. Song, "Identitydp: Differential private identification protection for face images," *Neurocomputing*, vol. 501, pp. 197–211, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231222007597

[69] L. Zhai, Q. Guo, X. Xie, L. Ma, Y. E. Wang, and Y. Liu, "A3gan: Attribute-aware anonymization networks for face de-identification," in *Proc. ACM Int. Conf. Multimed.*, 2022, p. 53035313. [Online]. Available: https://doi.org/10.1145/3503161.3547757

[70] X. Gu, W. Luo, M. S. Ryoo, and Y. J. Lee, "Password-conditioned anonymization and deanonymization with face identity transformers," in *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds., 2020, pp. 727–743.

[71] E. Chatzikyriakidis, C. Papaioannidis, and I. Pitas, "Adversarial face de-identification," in *Proc. Int. Conf. Image Process.*, 2019, pp. 684–688.

[72] V. Cherepanova, M. Goldblum, H. Foley, S. Duan, J. P. Dickerson, G. Taylor, and T. Goldstein, "Lowkey: Leveraging adversarial attacks to protect social media users from facial recognition," in *Int. Conf. Learn. Represent.*, 2021. [Online]. Available: https://openreview.net/forum?id=hJmtwocEqzc

[73] Y. Zhong and W. Deng, "Opom: Customized invisible cloak towards face privacy protection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3590–3603, 2023.

[74] F. Yang, L. Bourdev, E. Shechtman, J. Wang, and D. Metaxas, "Facial expression editing in video using a temporally-smooth factorization," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn.*, 2012, pp. 861–868.

[75] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, "Attgan: Facial attribute editing by only changing what you want," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5464–5478, 2019.

[76] H.-P. Wang, T. Orekondy, and M. Fritz, "Infoscrub: Towards attribute privacy by targeted obfuscation," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn. Workshops*, June 2021, pp. 3281–3289.

[77] V. Mirjalili, S. Raschka, and A. Ross, "Privacynet: Semi-adversarial networks for multi-attribute face privacy," *IEEE Trans. Image Process.*, vol. 29, pp. 9400–9412, 2020.

[78] V. Mirjalili and A. Ross, "Soft biometric privacy: Retaining biometric utility of face images while perturbing gender," in *IEEE Int. Jt. Conf. Biom.*, 2017, pp. 564–573.

[79] Y.-Y. Low, A. Tanvy, R. C.-W. Phan, and X. Chang, "Adverfacial: Privacy-preserving universal adversarial perturbation against facial micro-expression leakages," in *IEEE Int. Conf. Acoust. Speech Signal Process. Proc.*, 2022, pp. 2754–2758.

[80] W. Chen and R. W. Picard, "Eliminating physiological information from facial videos," in *Proc. - IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2017, pp. 48–55.

[81] Z. Sun and X. Li, "Privacy-phys: Facial video-based physiological modification for privacy protection," *IEEE Signal Process. Lett.*, vol. 29, pp. 1507–1511, 2022.

[82] W. Xu, S.-c. S. Cheung, and N. Soares, "Affect-preserving privacy protection of video," in *Proc. Int. Conf. Image Process.*, 2015, pp. 158–162.

[83] Z. Kuang, L. Teng, Z. Yu, J. Yu, J. Fan, and M. Xu, "Delegate-based utility preserving synthesis for pedestrian image anonymization," in *Proc. ACM Int. Conf. Multimed.*, 2022, p. 23142323. [Online]. Available: https://doi.org/10.1145/3503161.3548235

[84] X. Yuan, P. He, X. Lit, and D. Wu, "Adaptive adversarial attack on scene text recognition," in *IEEE Conf. Comput. Commun. Workshops*, 2020, pp. 358–363.

[85] L. Chen, J. Sun, and W. Xu, "Fawa: Fast adversarial watermark attack on optical character recognition (ocr) systems," in *Machine Learning and Knowledge Discovery in Databases*, 2021, pp. 547–563.

[86] X. Xu, J. Chen, J. Xiao, L. Gao, F. Shen, and H. T. Shen, "What machines see is not what they get: Fooling scene text recognition models with adversarial text images," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn.*, June 2020.

[87] Y. Liu, W. Zhang, and N. Yu, "Protecting privacy in shared photos via adversarial examples based stealth," *Secur. Commun. Netw.*, vol. 2017, 2017.

[88] R. Duan, Y. Chen, D. Niu, Y. Yang, A. K. Qin, and Y. He, "Advdrop: Adversarial attack to dnns by dropping information," in *Proc. IEEE Int. Conf.Comput. Vision*, October 2021, pp. 7506–7515.

[89] L. Jiang, X. Ma, S. Chen, J. Bailey, and Y.-G. Jiang, "Black-box adversarial attacks on video recognition models," in *Proc. ACM Int. Conf. Multimed.*, ser. MM '19, 2019, p. 864872. [Online]. Available: https://doi.org/10.1145/3343031.3351088

[90] A. Rajabi, R. B. Bobba, M. Rosulek, C. Wright, and W.-c. Feng, "On the (im) practicality of adversarial perturbation for image privacy," *Proceedings on Privacy Enhancing Technologies*, 2021.

[91] D. McDuff and E. M. Nowara, "warm bodies: A post-processing technique for animating dynamic blood flow on photos and avatars," in *Conf. Hum. Fact. Comput. Syst. Proc.*, 2021. [Online]. Available: https://doi.org/10.1145/3411764.3445719

[92] J. Paruchuri, S.-c. Cheung, and M. Hail, "Video data hiding for managing privacy information in surveillance systems," *EURASIP J. Inf. Secur.*, vol. 2009, no. 1, pp. 1–18, 2009.

[93] R. Hasan, E. Hassan, Y. Li, K. Caine, D. J. Crandall, R. Hoyle, and A. Kapadia, "Viewer experience of obscuring scene elements in photos to enhance privacy," in *Conf. Hum. Fact. Comput. Syst. Proc.*, 2018, p. 113. [Online]. Available: https://doi.org/10.1145/3173574.3173621

[94] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, and Y. Akbari, "Image inpainting: A review," *Neural Process. Lett.*, vol. 51, pp. 2007–2028, 2020.

[95] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "Patchmatch: A randomized correspondence algorithm for structural image editing," *ACM Trans. Graph.*, vol. 28, no. 3, jul 2009. [Online]. Available: https://doi.org/10.1145/1531326.1531330

[96] Z. Yan, X. Li, M. Li, W. Zuo, and S. Shan, "Shift-net: Image inpainting via deep feature rearrangement," in *ECCV*, September 2018.

[97] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241.

[98] R. Uittenbogaard, C. Sebastian, J. Vijverberg, B. Boom, D. M. Gavrila, and P. H. d. With, "Privacy protection in street-view panoramas using depth and multi-view imagery," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn.*, June 2019.

[99] T. Nakamura, A. Zhu, K. Yanai, and S. Uchida, "Scene text eraser," in *Proc. Int. Conf. Doc. Anal. Recognit.*, vol. 01, 2017, pp. 832–837.

[100] O. Tursun, R. Zeng, S. Denman, S. Sivapalan, S. Sridharan, and C. Fookes, "Mtrnet: A generic scene text eraser," in *Proc. Int. Conf. Doc. Anal. Recognit.*, 2019, pp. 39–44.

[101] P. Roy, S. Bhattacharya, S. Ghosh, and U. Pal, "Stefann: Scene text editor using font adaptive neural network," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn.*, June 2020.

[102] V. K. B. G, J. Subramanian, V. Chordia, E. Bart, S. Fang, K. Guan, and R. Bala, "Strive: Scene text replacement in videos," in *Proc. IEEE Int. Conf. Comput. Vision*, October 2021, pp. 14 549–14 558.

[103] J. Yu, H. Xue, B. Liu, Y. Wang, S. Zhu, and M. Ding, "Gan-based differential private image privacy protection framework for the internet of multimedia things," *Sensors*, vol. 21, no. 1, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/1/58

[104] E. T. Hassan; Rakibul Hasan; Patrick Shaffer; David Crandall; Apu Kapadia, "Cartooning for enhanced privacy in lifelogging and streaming videos," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn. Workshops*, July 2017.

[105] M. Bonetto, P. Korshunov, G. Ramponi, and T. Ebrahimi, "Privacy in mini-drone based video surveillance," in *IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, vol. 04, 2015, pp. 1–6.

[106] T. Boult, "Pico: Privacy through invertible cryptographic obscuration," in *Computer Vision for Interactive and Intelligent Environment (CVIIE'05)*, 2005, pp. 27–38.

[107] F. Dufaux and T. Ebrahimi, "Scrambling for privacy protection in video surveillance systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 8, pp. 1168–1174, 2008.

[108] I. Aribilola, M. N. Asghar, N. Kanwal, M. Fleury, and B. Lee, "Securecam: Selective detection and encryption enabled application for dynamic camera surveillance videos," *IEEE Trans. Consum. Electron., in press, doi: 10.1109/TCE.2022.3228679*, 2022.

[109] J. He, B. Liu, D. Kong, X. Bao, N. Wang, H. Jin, and G. Kesidis, "Puppies: Transformation-supported personalized privacy preserving partial image sharing," in *Proc. - Annu. IEEE/IFIP Int. Conf. Dependable Syst. Networks*, 2016, pp. 359–370.

[110] A. Frome, G. Cheung, A. Abdulkader, M. Zennaro, B. Wu, A. Bissacco, H. Adam, H. Neven, and L. Vincent, "Large-scale privacy protection in google street view," in *Proc. IEEE Int. Conf. Comput. Vision*, 2009, pp. 2373–2380.

[111] K. Brkic, I. Sikiric, T. Hrkac, and Z. Kalafatic, "I know that person: Generative full body and face de-identification of people in images," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn. Workshops*, 2017, pp. 1319–1328.

[112] J. Zhou and C.-M. Pun, "Personal privacy protection via irrelevant faces tracking and pixelation in video live streaming," *IEEE Trans. Inf. Forensic Secur.*, vol. 16, pp. 1088–1103, 2021.

[113] Y. Sun, S. Zhu, and Y. Chen, "Zoomp 3: Privacy-preserving publishing of online video conference recordings," in *Proceedings on Privacy Enhancing Technologies*, 2022, pp. 630–649.

[114] . Erdlyi, T. Bart, P. Valet, T. Winkler, and B. Rinner, "Adaptive cartooning for privacy protection in camera networks," in *IEEE Int. Conf. Adv. Video Signal-Based Surveill.*, 2014, pp. 44–49.

[115] Á. Erdélyi, T. Winkler, and B. Rinner, "Multi-level cartooning for context-aware privacy protection in visual sensor networks." in *MediaEval*, 2014.

[116] L. Bao and Y. Zhou, "Image encryption: Generating visually meaningful encrypted images," *Inf. Sci.*, vol. 324, pp. 197–207, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025515004818

[117] C. V. Wright, W.-c. Feng, and F. Liu, "Thumbnail-preserving encryption for jpeg," in *Proc. ACM Workshop Inf. Hiding Multimedia Secur.*, 2015, p. 141146. [Online]. Available: https://doi.org/10.1145/2756601.2756618

[118] A. Jolfaei, X.-W. Wu, and V. Muthukkumarasamy, "On the security of permutation-only image encryption schemes," *IEEE Trans. Inf. Forensic Secur.*, vol. 11, no. 2, pp. 235–246, 2016.

[119] K. Tajik, A. Gunasekaran, R. Dutta, B. Ellis, R. B. Bobba, M. Rosulek, C. V. Wright, and W.-c. Feng, "Balancing image privacy and usability with thumbnail-preserving encryption," in *Proc. Symp. Netw. Distrib. Syst. Secur.,*, 2019.

[120] C. Zou, D. Yuan, L. Lan, and H. Chi, "Privacy-preserving action recognition," in *IEEE Int. Conf. Acoust. Speech. Signal Process. Proc.*, 2022, pp. 2175–2179.

[121] A. Kunchala, M. Bouroche, and B. Schoen-Phelan, "Towards a framework for privacy-preserving pedestrian analysis," in *Proc. - IEEE/CVF Winter Conf. Appl. Comput. Vis.*, January 2023, pp. 4370–4380.

[122] Y. Yang, J. Lyu, R. Wang, Q. Wen, L. Zhao, W. Chen, S. Bi, J. Meng, K. Mao, Y. Xiao *et al.*, "A digital mask to safeguard patient privacy," *Nat. Med.*, vol. 28, no. 9, pp. 1883–1892, 2022.

[123] K. Brki, T. Hrka, and Z. Kalafati, "Protecting the privacy of humans in video sequences using a computer vision-based de-identification pipeline," *Expert Syst. Appl.*, vol. 87, pp. 41–55, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417417303986

[124] I. R. Dave, C. Chen, and M. Shah, "Spact: Self-supervised privacy preservation for action recognition," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn.*, June 2022, pp. 20 164–20 173.

[125] Z. Wu, H. Wang, Z. Wang, H. Jin, and Z. Wang, "Privacy-preserving deep action recognition: An adversarial learning framework and a new dataset," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 4, pp. 2126–2139, 2022.

[126] M. Ryoo, B. Rothrock, C. Fleming, and H. J. Yang, "Privacy-preserving human activity recognition from extreme low resolution," *Proc. AAAI Conf. Artif. Intell.*, vol. 31, no. 1, Feb. 2017. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/11233

[127] C. Hinojosa, J. C. Niebles, and H. Arguello, "Learning privacy-preserving optics for human pose estimation," in *Proc. IEEE Int. Conf. Comput. Vision*, October 2021, pp. 2573–2582.

[128] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vision*, Oct 2017.

[129] Z. Wu, Z. Wang, Z. Wang, and H. Jin, "Towards privacy-preserving visual recognition via adversarial training: A pilot study," in *ECCV*, September 2018.

[130] Z. W. Wang, V. Vineet, F. Pittaluga, S. N. Sinha, O. Cossairt, and S. Bing Kang, "Privacy-preserving action recognition using coded aperture videos," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn. Workshops*, June 2019.

[131] J. Liu, R. Tan, G. Han, N. Sun, and S. Kwong, "Privacy-preserving in-home fall detection using visual shielding sensing and private information-embedding," *IEEE Trans. Multimedia*, vol. 23, pp. 3684–3699, 2021.

[132] A. Gupta, A. Jaiswal, Y. Wu, V. Yadav, and P. Natarajan, "Adversarial mask generation for preserving visual privacy," in *Proc. - IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2021, pp. 1–5.

[133] J. Ji, H. Wang, Y. Huang, J. Wu, X. Xu, S. Ding, S. Zhang, L. Cao, and R. Ji, "Privacy-preserving face recognition withlearnable privacy budgets infrequency domain," in *ECCV*, 2022, pp. 475–491.

[134] H. Wu, X. Tian, Y. Gong, X. Su, M. Li, and F. Xu, "Dapter: Preventing user data abuse in deep learning inference services," in *Companion Proc. Web Conf.*, 2021, p. 10171028. [Online]. Available: https://doi.org/10.1145/3442381.3449907

[135] M. Aprilpyone and H. Kiya, "Block-wise image transformation with secret key for adversarially robust defense," *IEEE Trans. Inf. Forensic Secur.*, vol. 16, pp. 2709–2723, 2021.

[136] T. M. Cannon and E. E. Fenimore, "Coded aperture imaging: many holes make light work," *Opt. Eng.*, vol. 19, no. 3, pp. 283–289, 1980.

[137] S. Spiekermann and L. F. Cranor, "Engineering privacy," *IEEE Trans. Softw. Eng.*, vol. 35, no. 1, pp. 67–82, 2009.

[138] D. Mokrosinska, "Privacy and autonomy: On some misconceptions concerning the political dimensions of privacy," *Law Philos.*, vol. 37, no. 2, pp. 117–143, 2018.

[139] M. R. Anderlik and M. A. Rothstein, "Privacy and confidentiality of genetic information: what rules for the new science?" *Annu. Rev. Genomics Hum. Genet.*, vol. 2, no. 1, pp. 401–433, 2001.

[140] A. Dehghantanha and K. Franke, "Privacy-respecting digital investigation," in *Annu. Int. Conf. Priv., Secur. Trust*, 2014, pp. 129–138.

[141] R. Clarke, "Internet privacy concerns confirm the case for intervention," *Commun. ACM*, vol. 42, no. 2, p. 6067, feb 1999. [Online]. Available: https://doi.org/10.1145/293411.293475

[142] J. E. Cohen, "What privacy is for," *Harv. Law Rev.*, vol. 126, no. 7, pp. 1904–1933, 2013. [Online]. Available: http://www.jstor.org/stable/23415061

[143] L. F. Cranor and S. Garfinkel, *Security and usability: designing secure systems that people can use.* " O'Reilly Media, Inc.", 2005.

[144] Y. Deng, J. Yang, D. Chen, F. Wen, and X. Tong, "Disentangled and controllable face image generation via 3d imitative-contrastive learning," in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn.*, June 2020.

[145] Y. Ren, G. Li, Y. Chen, T. H. Li, and S. Liu, "Pirenderer: Controllable portrait image generation via semantic neural rendering," in *Proc. IEEE Int. Conf. Comput. Vision*, October 2021, pp. 13 759–13 768.

[146] J. Sun, X. Wang, Y. Shi, L. Wang, J. Wang, and Y. Liu, "Ide-3d: Interactive disentangled editing for high-resolution 3d-aware portrait synthesis," *ACM Trans. Graph.*, vol. 41, no. 6, nov 2022. [Online]. Available: https://doi.org/10.1145/3550454.3555506

[147] A. Tonge and C. Caragea, "Image privacy prediction using deep neural networks," *ACM Trans. Web*, vol. 14, no. 2, apr 2020. [Online]. Available: https://doi.org/10.1145/3386082

[148] T. Orekondy, B. Schiele, and M. Fritz, "Towards a visual privacy advisor: Understanding and predicting privacy risks in images," in *Proc. IEEE Int. Conf. Comput. Vision*, Oct 2017.

[149] A. Tonge and C. Caragea, "Privacy-aware tag recommendation for accurate image privacy prediction," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 4, aug 2019. [Online]. Available: https://doi.org/10.1145/3335054

[150] T. Gernand, "Scanning iphones to save children: Apples on-device hashing algorithm should survive a fourth amendment challenge," *Dickinson Law Review (2017-Present)*, vol. 127, no. 1, p. 307, 2022.

[151] C. N. Insler, "How to ride the litigation rollercoaster driven by the biometric information privacy act," *S. Ill. ULJ*, vol. 43, p. 819, 2018.

[152] S. I. 30137-1, "Information technology use of biometrics in video surveillance systems part 1: System design and specification," *International Organization for Standardization*, 2019.