Multimedia Tools and Applications manuscript No.
(will be inserted by the editor)

Privacy Protection vs. Utility in Visual Data

An Objective Evaluation Framework

Ádám Erdélyi · Thomas Winkler · Bernhard Rinner

Received: date / Accepted: date

Abstract Ubiquitous and networked sensors impose a huge challenge for privacy protection which has become an emerging problem of modern society. Protecting the privacy of visual data is particularly important due to the omnipresence of cameras, and various protection mechanisms for captured images and videos have been proposed. This paper introduces an objective evaluation framework in order to assess such protection methods. Visual privacy protection is typically realised by obfuscating sensitive image regions which often results in some loss of utility. Our evaluation framework assesses the achieved privacy protection and utility by comparing the performance of standard computer vision tasks, such as object recognition, detection and tracking on protected and unprotected visual data. The proposed framework extends the traditional frame-by-frame evaluation approach by introducing two new approaches based on aggregated and fused frames. We demonstrate our framework on eight differently protected video-sets and measure the trade-off between the improved privacy protection due to obfuscating captured image data and the degraded utility of the visual data. Results provided by our objective evaluation method are compared with an available state-of-the-art subjective study of these eight protection techniques.

Keywords visual privacy · video surveillance · privacy evaluation framework · privacy/utility trade-off

1 Introduction

Privacy concerns have been raised by the rapidly increasing number of visual data capturing devices. Not only surveillance cameras threaten privacy but also other video-capable multimedia devices such as smart phones, tablets and wearable smart technology including Google Glass and Microsoft HoloLens when used in public areas. Web cameras also pose privacy threats—especially when abused

Ádám Erdélyi · Thomas Winkler · Bernhard Rinner
Institute of Networked and Embedded Systems,
Alpen-Adria-Universität Klagenfurt and Lakeside Labs, Austria
E-mail: adam.erdelyi@aau.at, thomas.winkler@gmail.com, bernhard.rinner@aau.at

through spy-ware. Domestic IP cameras designed for home surveillance can also lead to privacy loss due to careless installation [2]. Furthermore, an emerging privacy threat is posed by camera-equipped unmanned aerial vehicles (UAVs) also known as drones [18, 4, 12, 49]. Traditional CCTV (closed-circuit television) and other old-fashioned surveillance camera systems are continuously replaced recently by visual sensor networks (VSNs) which consist of smart cameras [45, 44]. Due to networking and on-board processing capabilities of the above mentioned visual data capturing devices, sophisticated artificial vision tasks can be performed. Therefore, privacy is at an even higher risk nowadays.

So-called *privacy filters* are often applied to protect visual data by obfuscating the sensitive parts of the captured data or replacing them with a de-identified representation—both of which entails some loss of utility. By the term *utility* we refer to certain system properties (e.g., the operating speed of a filter) and to intelligibility which represents the amount of useful information that can be extracted from the visual data. For example in case of a retail surveillance camera, privacy protection means that the identity of monitored people cannot be disclosed, and utility refers to the ability of still being capable to recognise the behaviour of monitored people such as detecting shoplifting. The privacy protection performance and the utility of the protected visual data represent two important (and inter-dependent) design aspects of various video applications. Finding an acceptable trade-off between privacy protection and utility is therefore an essential issue in the development and deployment of privacy protection methods. Therefore, it is essential to have a tool by which privacy filters can be evaluated and compared in terms of privacy and utility. Furthermore, privacy is scenario dependent and an ideal privacy-preserving method should be able to adapt to various scenarios by automatically selecting the most useful protection filter and hence selecting a Pareto-optimal point in the privacy-utility trade-off [20]. In order to support such automatic protection selection, the ability to evaluate the actual effectiveness of the privacy protection filters in use is essential. Such evaluation can be realised by subjective or objective methods. This paper focuses on an objective evaluation method due to its advantages over a subjective evaluation such as the support for automatic operation (no human assessment required), the reduced costs of implementation, and the increased reproducibility. Many techniques have been proposed for visual privacy protection [37, 17, 24, 5, 59, 38, 31, 20, 32, 22, 7, 41, 21, 36, 50, 42], but only a few papers have been published on how to evaluate, assess or compare these techniques [13, 53, 46, 29, 19, 8, 52, 33]. The main motivation behind this work was therefore to comprehensively explore the objective evaluation of the privacy-utility design space for visual privacy filters. Exploiting sequences of frames or the fusion of frames can reveal significant identifying information, however this aspect has not intensively been studied in related evaluation approaches so far (e.g., in [19, 8, 33]).

The contribution of this paper includes (1) a formal definition of privacy protection and utility in visual data based on the performance of standard computer vision tasks, (2) the introduction of aggregated and fused frames based evaluation approaches, (3) a concrete implementation to realise an objective evaluation framework, and (4) an extensive comparison of the results of our framework prototype with the results of a recent *subjective* study [9] on privacy protection mechanisms.

The remainder of this paper is structured as follows. Section 2 discusses related work in the area of privacy protection methods and their evaluation. In Section 3 we introduce our proposed objective evaluation framework and a formal definition is provided in Section 4. Section 5 presents implementation details and the evaluation results of eight different privacy protection filters. Section 6 concludes this paper with a summary and a brief discussion of future work.

2 Related Work

We start our discussion of related work with highly abstracted and multidisciplinary aspects of privacy in general and continue then with the evaluation of visual privacy protection methods.

A traditional approach of protecting privacy is called privacy enhancing technologies (PET) meaning that already existing systems are patched with protective mechanisms retroactively. Privacy by design (PbD) on the other hand pursues that privacy should be considered as an indispensable part of system design. PbD is built upon seven foundational principles [14]. According to these principles, privacy should be protected in a proactive instead of a reactive manner, and a default protection level should always be provided without any extra intervention. The protection of privacy should not restrict the original functionality of a system and make unnecessary trade-offs. Furthermore, privacy protection should be extended throughout the entire life-cycle of the data involved from start to finish. It has to be done transparently so that all stakeholders can be assured that the stated promises and objectives are actually kept. A privacy-preserving system should also respect user-privacy by being user-centric and keeping the interests of individuals uppermost. Cavoukian [15] also stated that privacy does not equal secrecy, but privacy equals control. The problem with this statement regarding visual privacy is that most people do not even know they are being observed by visual surveillance devices. If they are unaware of the existence of these devices, how could they have control over the captured data. Furthermore, people do not really feel the value of privacy until they have problems as a consequence of privacy loss. In addition, people usually do not live up to their self-reported privacy preferences and they regularly share sensitive information. This is called privacy paradox. More details about the issues around awareness and the so-called privacy paradox can be found in [39].

A multidisciplinary framework to include privacy in the design of video surveillance systems is described in [35]. It covers the field of privacy from political science to video technologies and points out that there are grey areas posing serious privacy risks. Furthermore, it raises the question of the definition of personal and sensitive information. Table 1 summarises a possible answer to this question with regard to visual privacy. Chaaraoui *et al.* [16] also describe a new approach called privacy by context (PbC) which supports the idea that privacy is scenario/context dependent.

Over the last decade various methods have been developed to protect visual privacy. These privacy-preserving techniques basically rely on image processing algorithms such as scrambling by JPEG-masking [37], in-painting [17], pixelation [24], blanking [5], replacement with silhouettes [59], blurring [38], warping or morphing [31]. In a recent workshop dedicated protection methods have been proposed in order to solve the specified visual privacy task [32, 22, 7, 41, 21, 36, 50, 42].

Information	**Related Visual Clues**
Who is the person? (identity)	Face, hair, skin, height, clothes, gait
How is the person displayed? (appearance)	Face expressions, hair (e.g., colour, hairstyle, etc.), body (e.g., nudity), posture, shape, colour
Where is the person? (location)	Room, spatial position (e.g., on the floor, on the bed, etc.), room signs
What is the person doing? (activity)	Behaviour (i.e., movement, gesture, action, activity), gaze, spatial position, objects and interactions
When is the activity taking place? (time)	Temporal clues (e.g., a wall clock, weather)

Table 1: The types of information that can be extracted out of image sequences and the related visual clues that can provide this information [16].

A comprehensive discussion on the state of the art in this field can be found in the surveys of Winkler *et al.* [58] and Padilla-López *et al.* [40].

Due to the steadily increasing number of protection approaches as well as high variability of visual tasks and scenes, an evaluation methodology for comparing the approaches is urgently needed. Privacy impact assessments (PIAs) are an integral part of the above mentioned privacy by design approach [28]. Existing evaluation methods usually consider two aspects, namely privacy and utility. The levels of privacy protection and utility can be assessed by subjective and objective evaluation methods. Subjective methods are quite common and include techniques such as questionnaires and user studies [13, 53, 46, 29, 12, 11]. Naturally, they are tedious and expensive to implement, and the assessment may depend on the study group.

Objective evaluation of privacy-preserving techniques in the field of visual surveillance is a challenging issue because privacy is highly subjective and depends on various aspects such as culture, location, time and situation. Nevertheless, a couple of techniques have been developed which are mostly based on computer vision algorithms. Dufaux and Ebrahimi [19] proposed an evaluation method that uses the face identification evaluation system (FIES) of Colorado State University (CSU), which provides standard face recognition algorithms and standard statistical methods for assessment. Principal components analysis (PCA) [54] and linear discriminant analysis (LDA) [10] are used as face recognition algorithms together with the grey-scale facial recognition technology (FERET) dataset. A more comprehensive evaluation framework is described in [8], where Badii *et al.* carried out both subjective and objective evaluation along the following five crucial categories.

- *Efficacy* – The ability to effectively obscure privacy-sensitive elements.
- *Consistency* – In order to successfully and continuously track a moving subject, the details of its shape and appearance have to be maintained on a reasonable and consistent level.
- *Disambiguity* – The degree by which a privacy filter does not introduce additional ambiguity in cross-frame trackability of same persons/objects.
- *Intelligibility* – The ability to only protect the privacy-sensitive attributes and retain all other features / information in the video-frame(s) in order not to detract from the purpose of the surveillance system.
- *Aesthetics* – To avoid viewers' distraction and unnecessary fixation of their attention on the region of the video-frame to be obscured by the privacy filter,

it is important for the privacy filter to maintain the perceived quality of the visual effects of the video-frame.

Subjective and objective evaluations are cross-validated and the authors claim that the results indicate the same trend. Unfortunately, this paper does not provide sufficient details of the study.

Sohn *et al.* [52] have also carried out objective and subjective evaluations together. They assessed their JPEG XR based privacy filter in four aspects: spatial resolution, visual quality, replacement attack and non-scrambled colour information. In their objective evaluation Sohn *et al.* [52] used various face recognisers and the subjective evaluation was conducted with 35 participants whose task was to match 45 privacy protected face images against the 12 original ones. Privacy evaluation was exclusively focused on face recognition.

Korshunov *et al.* [33] evaluated privacy protection methods by measuring the amount of visual details (such as facial features) in the sample images as a metric of privacy and the overall shape of faces as a metric for intelligibility. In their demonstration they used three different datasets with various resolutions and face sizes, and three different privacy filter methods, namely blurring, pixelation and blanking. For measuring the level of privacy, the failure rates of automatic face recognition methods (PCA [54], LDA [10], LBP [3]) were considered, while the accuracy rate of a face detector (Viola-Jones [55]) were used to measure intelligibility. In these experiments only faces were considered, which is insufficient for proper privacy protection taking into account the above mentioned privacy by context approach or the secondary (implicit) privacy channels described by Saini *et al.* [47].

Our paper focuses on establishing an objective evaluation framework by exploiting various evaluator functions to measure privacy and utility in various aspects. The main difference to the related work lies in its generalisation and flexibility. Our framework does not restrict the evaluation to a particular algorithm (e.g., a face detector) but rather uses a set of evaluator functions which can be easily adapted to the specific application.[1] It further does not impose constraints to the visual data and the privacy filters. The privacy and utility evaluation is based on the performance of the evaluator functions on the provided visual data. While state-of-the-art evaluation frameworks [19, 8, 33] usually assess only individual frames, we also consider aggregated and fused frames for the evaluation.

3 Objective Evaluation Framework

Our primary goal is to provide a framework that enables the evaluation of privacy protection techniques along two inter-dependent dimensions: (i) the achieved privacy protection level and (ii) the utility of the technique and the overall system. A particular protection technique (or a particular strength of a protection filter) will therefore result in specific values for privacy and utility when using our framework. Figure 1 presents an overview of the proposed framework. The "privacy protection filter" represents a computer vision algorithm which transforms

[1] The output of the evaluator function is basically derived by comparing the performance of a specific computer vision algorithm on the protected visual data with a "reference" performance. Such reference can be provided either as (manually generated) ground truth data or as the output of the computer vision algorithm on the unprotected visual data.

an input video into an output video stream where privacy sensitive elements are protected.

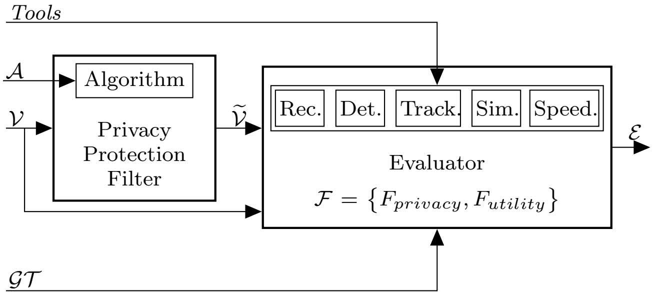


Fig. 1: Our objective evaluation framework.

The evaluator tools (*Tools*), the privacy-preserving algorithm under test $\mathcal{A}$ and the unprotected visual data $\mathcal{V}$ together with the ground-truth $\mathcal{GT}$ serve as input to our framework. The visual data is preferably captured in heterogeneous scenes such as indoor and outdoor, day and night, empty and crowded environments in order to achieve a comprehensive evaluation. The unprotected visual data $\mathcal{V}$ is processed by the privacy protection filter which is the algorithm under test. The unprotected $\mathcal{V}$ and the protected $\widetilde{\mathcal{V}}$ visual data along with the ground-truth $\mathcal{GT}$ are then fed into the main component of the framework, namely the evaluator. This evaluator relies on two major sets of evaluator functions $F_{privacy}$ and $F_{utility}$ that are used to evaluate the examined privacy protection filter from the perspectives of privacy and utility. Each evaluation function provides a real number between zero and one as a result. The implementation of these functions depends on the selected tools which are based on computer vision algorithms. The output of the evaluation framework is given by the set $\mathcal{E}$ which is determined by the results of the evaluation functions.

3.1 Notation

In this section we describe the notation used in our framework for the unprotected videos, the protected videos, the ground-truth data and the evaluator functions.

3.1.1 Unprotected Visual Data

The unprotected visual data is specified by a set of video clips

$$\mathcal{V} = \left\{ {}^{1}V, \ldots, {}^{N}V \right\} \tag{1}$$

$${}^{n}V = \left\{ {}^{n}v^{1}, \ldots, {}^{n}v^{{}^{n}L} \right\} |_{n=1\ldots N} \tag{2}$$

where

${}^nV$ represents the n^{th} unprotected video clip composed by a set of image frames,
N is the number of all video clips being used in the evaluation process,
${}^nv^i$ is a single image frame with index i from the unprotected video clip ${}^nV$, and
${}^nL$ is the length of the n^{th} video clip in $\mathcal{V}$.

3.1.2 Privacy Protected Visual Data

The unprotected visual data is processed by the protection algorithm under test and is transformed into the protected visual data. The protected visual data is thus given by the set of video clips derived running the protection filter on $\mathcal{V}$

$$\widetilde{\mathcal{V}} = \left\{{}^1\widetilde{V}, \dots, {}^N\widetilde{V}\right\} \tag{3}$$

$${}^n\widetilde{V} = \left\{{}^n\widetilde{v}^1, \dots, {}^n\widetilde{v}^{{}^nL}\right\}|_{n=1\dots N} \tag{4}$$

where

${}^n\widetilde{V}$ represents the n^{th} privacy protected video clip which is a set of filtered image frames, and
${}^n\widetilde{v}^i$ is an image frame from the protected video clip ${}^n\widetilde{V}$.

3.1.3 Ground-Truth Data

The ground-truth data contains the position and size of the objects of interest in form of bounding boxes along with their classification in form of descriptors. Furthermore, each object of interest has an identity in form of a globally unique number. The ground-truth data for all input video clips is available as

$$\mathcal{GT} = \left\{{}^1\mathcal{O}_{gt}, \dots, {}^N\mathcal{O}_{gt}\right\} \tag{5}$$

$${}^n\mathcal{O}_{gt} = \left\{{}^nO^1_{gt}, \dots, {}^nO^{{}^nL}_{gt}\right\}|_{n=1\dots N} \tag{6}$$

$${}^nO^i_{gt} = \left\{{}^no^i_{gt_1}, \dots, {}^no^i_{gt_J}\right\}|_{n=1\dots N, i=1\dots {}^nL} \tag{7}$$

$${}^no^i_{gt_j} = ({}^nb^i_{gt_j}, {}^nd^i_{gt_j})|_{n=1\dots N, i=1\dots L_n, j=1\dots J} \tag{8}$$

where

${}^n\mathcal{O}_{gt}$ is a set that contains the ground-truth data for each frame of the n^{th} video clip,
${}^nO^i_{gt}$ is the ground-truth of frame i in the n^{th} video clip,
J is the number of distinct objects in $\mathcal{GT}$,
j is a globally unique identifier of an object running from 1 to J,
${}^no^i_{gt_j}$ is a pair (b, d) for each object in frame i of the n^{th} video clip,
${}^nb^i_{gt_j}$ is the bounding box of object j in frame i of the n^{th} video clip, and
${}^nd^i_{gt_j}$ is the descriptor of object j in frame i of the n^{th} video clip.

$\mathcal{GT}$ can be explicitly given (e.g., by manual video annotation) or derived by running various computer vision algorithms such as object recognisers, detectors, or trackers on the visual data.

3.1.4 Evaluator Functions

The evaluation is based on comparing the results of selected algorithms on the protected visual data with the ground truth or the performance on the unprotected visual data, respectively. The set of evaluation functions is given as

$$\mathcal{F} = \{F_{privacy}, F_{utility}\} \tag{9}$$

where

$$F_{privacy} = \left\{f_{id_{ind}}, f_{id_{aggr}}, f_{id_{fused}}\right\} \tag{10}$$

and

$$F_{utility} = \left\{f_{det_{ind}}, f_{det_{aggr}}, f_{det_{fused}}, f_{track}, f_{sim}, f_{speed}\right\} . \tag{11}$$

The subscripts $_{id}$, $_{det}$, $_{track}$, $_{sim}$, and $_{speed}$ mark evaluation functions that are based on object identification, detection, tracking, image similarity and the processing speed of the privacy protection filter, respectively. Functions for object identification correspond to functions for measuring the privacy protection performance. The other functions represent examples for measuring the utility. The subscripts $_{ind}$, $_{aggr}$ and $_{fused}$ refer to *independent*, *aggregated* and *fused* frames. More details about these functions and the different classes of frames are described in Section 4.

The output of the evaluator is the set of results

$$\mathcal{E} = \{E_{privacy}, E_{utility}\} \tag{12}$$

where

$$E_{privacy} = \left\{e_{id_{ind}}, e_{id_{aggr}}, e_{id_{fused}}\right\} \tag{13}$$

and

$$E_{utility} = \left\{e_{det_{ind}}, e_{det_{aggr}}, e_{det_{fused}}, e_{track}, e_{sim}, e_{speed}\right\} . \tag{14}$$

These results are constituted by the outputs of the evaluator functions where $e_{id_{ind}}$, $e_{id_{aggr}}$, $\ldots$, e_{speed} represent the output values of the functions $f_{id_{ind}}$, $f_{id_{aggr}}$, $\ldots$, f_{speed}, respectively and $\forall e \in \mathbb{R} \mid 0 \leq e \leq 1$. The set $\mathcal{E}$ can be considered as a "signature" of the evaluated privacy protecting method along the privacy and utility dimensions. The evaluator functions represent different aspects of the privacy-utility design space and were chosen based on the most commonly used approaches of the related work and our own experience in the field. It is important to note that these evaluator functions are examples, and our framework be can easily adapted to functions covering different utility aspects.

4 Definition of the Evaluation Framework

State-of-the-art privacy evaluation frameworks [19, 8, 33] usually work on the basis of individual frames. This means that the effect of a privacy protection filter is evaluated by assessing the evaluator functions for each image frame independently. Such frame-by-frame evaluation methods have limitations in revealing a privacy loss caused by the exploitation of aggregated or fused frames from different time instances and/or multiple cameras looking at the same object. In our framework definition we attempt to overcome these deficiencies. For each evaluator function f, if applicable, we will provide various measurement methods that take

1. *independent* frames,
2. *aggregated* frames of the same visual data from different time instances or from multiple capturing devices, and
3. *fused* frames of the same visual data from different time instances or from multiple capturing devices

into account. Aggregated and fused frames may provide more information about the objects of interest than individual frames. Thus, it might be helpful to consider this additional information for the privacy evaluation. In case of *aggregated* frames an evaluator function f has access to a set of frames and carries out the measurements jointly for this set (i.e., multiple frames are used simultaneously during the evaluation). The performance of a privacy protection filter might deteriorate using aggregated frames despite its good frame-by-frame performance. For example, if there is at least one insufficiently protected frame in the visual data where an object of interest can be recognised, this object may lose its privacy in other frames as well due to successful object tracking even if the object's identity is well protected in all other frames. In case of *fused* frames, multiple frames from the same or different cameras are analysed and combined in order to construct a new set of abstracted visual data. It is possible that fused frames constructed from multiple frames from different time instances or view angles may provide a better view on an object. Examples for fusion methods include image stitching, super-resolution or de-filtering. The fused information may lead to privacy loss as well.

4.1 Evaluation of Privacy

In this section we define the evaluator functions used for privacy evaluation.

$$F_{privacy} = \left\{ f_{id_{ind}}, f_{id_{aggr}}, f_{id_{fused}} \right\} \tag{15}$$

In our framework we measure the privacy protection level of visual data by the *de-identification rate* of protected objects as a successful identification of the object of interest is the primary cause of privacy loss. The level of privacy is considered to be low if objects can be clearly identified and high if the identification is not possible.

1. *Independent frames*
   Frame-by-frame evaluation of de-identification is performed by object recognition algorithms trained for the specific objects of interest. Object recognisers are trained based on the unprotected visual data. Object recognition is carried out within each annotated bounding box ${}^{n}b^{i}_{gt_j}$ of each privacy protected frame ${}^{n}\tilde{v}^{i}$ in each video ${}^{n}\widetilde{V}$ from $\widetilde{\mathcal{V}}$ where object ${}^{n}o^{i}_{gt_j}$ actually appears. If the output of the recogniser does not match the ground-truth then the de-identification was considered successful and hence privacy is protected. The privacy level provided by the protection algorithm can be calculated depending on how often the object's identity has been successfully recognized. Therefore, the final output of the function $f_{id_{ind}}$ is defined as the ratio between the number of unrecognised objects in $\widetilde{\mathcal{V}}$ and the total number of occurrences of all objects

in $\mathcal{GT}$ which can be calculated as the inverse of the average hit-rate of the recognitions.

$$f_{id_{ind}}\left(\widetilde{\mathcal{V}},\mathcal{GT}\right)=1-\frac{h_{id_{ind}}}{\sum\limits_{j=1}^{J}\texttt{occurrences}\left(o_{gt_j}\right)} \tag{16}$$

The function $\texttt{occurrences}\,()$ returns the total number of occurrences of the object o_{gt_j} in $\mathcal{GT}$, i.e., the number of frames where the object is visible. $h_{id_{ind}}$ represents the number of successful object recognitions (hit-rate) in $\widetilde{\mathcal{V}}$ and is calculated as follows:

```
$h_{id_{ind}} := 0$;
for $j := 1$ to $J$ do
  for $\forall {}^n\tilde{v}^i|_{n:=1\ldots N, i:=1\ldots {}^nL}$ where $o_{gt_j} \in {}^n\tilde{v}^i$ do
    $j_{rec} := \texttt{recognise}\left({}^nb^i_{gt_j}\right)$;
    if $j = j_{rec}$ then
      $h_{id_{ind}}$++;
    end if
  end for
end for
```

where the function $\texttt{recognise}\,()$ performs object recognition within the bounding box of a given object and returns the identifier of the top ranked object. This is then stored in j_{rec} and compared to the object's true identifier. In our framework, $\texttt{recognise}\,()$ is not bound to any specific object recognition algorithm. Any suitable algorithm that fits the purpose and the object type can be used for the concrete framework implementation.

2. *Aggregated frames*

When using multiple frames simultaneously the de-identification rate can be computed as follows. Object recognition is carried out within each annotated bounding box ${}^nb^i_{gt_j}$ of each protected frame ${}^n\tilde{v}^i$ in each video ${}^n\widetilde{V}$ from $\widetilde{\mathcal{V}}$ where ${}^no^i_{gt_j}$ actually appears. If a particular object ${}^no^i_{gt_j}$ can be recognised at least once in the input data-set, then all the occurrences of that object are considered as successfully recognised. This severe loss of privacy is due to the perfect object tracking assumption among all aggregated frames. Although tracking does not reveal the identity per se, the identity of a successfully recognised objected can be propagated among all aggregated frames. The final output of the function $f_{id_{aggr}}$ is the ratio between the number of unrecognised objects in $\widetilde{\mathcal{V}}$ and the total number of occurrences of all objects in $\mathcal{GT}$ which can be calculated as follows:

$$f_{id_{aggr}}\left(\widetilde{\mathcal{V}},\mathcal{GT}\right)=1-\frac{h_{id_{aggr}}}{\sum\limits_{j=1}^{J}\texttt{occurences}\left(o_{gt_j}\right)} \tag{17}$$

where the function $\texttt{occurrences}\,()$ returns the total number of occurrences of the object o_{gt_j} in $\mathcal{GT}$. $h_{id_{aggr}}$ stands for the number of successful object recognitions (hit-rate) in $\widetilde{\mathcal{V}}$ and is calculated as follows:

```
$h_{id_{aggr}} := 0$;
for $j := 1$ to $J$ do
  for $\forall {}^n\tilde{v}^i|_{n:=1\ldots N, i:=1\ldots {}^nL}$ where $o_{gt_j} \in {}^n\tilde{v}^i$ do
    $j_{rec} := \texttt{recognise}\left({}^nb^i_{gt_j}\right)$;
```

```
            if $j = j_{rec}$ then
                $h_{id_{aggr}}$ += occurrences $(o_{gt_j})$;
                break;
            end if
        end for
    end for
```

where the function `recognise` () performs object recognition within the bounding box of a given object and returns the identifier of the top ranked object. This is then stored in j_{rec} and compared to the object's true identifier. As previously mentioned, the recognition algorithm can be chosen arbitrarily.

3. *Fused frames*
   If frames are fused in order to get abstracted information of the objects, de-identification is measured as follows. A set of fused images is created, and object recognition is carried out on each fused image. If an object can be recognised based on fused images, then by assuming perfect object tracking all occurrences of that object in $\mathcal{GT}$ are considered to be recognised in the data-set. The final output of the function $f_{id_{fused}}$ is the ratio between the number of unrecognised objects in $\widetilde{\mathcal{V}}$ and the total number of occurrences of all objects in $\mathcal{GT}$.

$$f_{id_{fused}}\left(\widetilde{\mathcal{V}}, \mathcal{GT}\right) = 1 - \frac{h_{id_{fused}}}{\sum\limits_{j=1}^{J} \texttt{occurences}\left(o_{gt_j}\right)} \tag{18}$$

The function `occurrences` () returns the total number of occurrences of a certain object based on the ground-truth. $h_{id_{fused}}$ is the hit-rate of object recognition and is calculated as follows:

```
$h_{id_{fused}} := 0$;
$\mathcal{FI}$ := { set of fused images }
for $j := 1$ to $J$ do
    for $\forall o_{fused} \in \mathcal{FI}$ do
        $j_{rec}$ := recognise $(b_{fused})$;
        if $j = j_{rec}$ then
            $h_{id_{fused}}$ += occurences $(o_{gt_j})$;
            break;
        end if
    end for
end for
```

where the function `recognise` () performs object recognition in a fused frame within the bounding box of a given object and returns the identifier of the top ranked object. This is then stored in j_{rec} and compared to the object's true identifier. As previously mentioned, the recognition algorithm can be chosen arbitrarily.

4.2 Evaluation of Utility

In our framework we measure utility by the performance ratio of various functions on the protected and unprotected visual data. The utility of visual data includes various aspects such as the capability of detecting specific objects or activities, the fidelity of the protected data or the complexity/efficiency of the protection filters. We propose the following evaluator functions for utility evaluation.

$$F_{utility} = \left\{ f_{det_{ind}}, f_{det_{aggr}}, f_{det_{fused}}, f_{track}, f_{sim}, f_{speed} \right\} \tag{19}$$

For the detection capability, we focus on object detection in terms of independent, aggregated and fused frames as well as on object tracking algorithms. For the fidelity aspect, we measure the similarity between unprotected and protected visual data, and we use the processing speed of privacy protection filters as a measure for efficiency. In the following subsections we explain in detail how these evaluator functions are determined.[2]

4.2.1 Utility Evaluation by Object Detection

One way of measuring utility is by the detection rate of privacy protected objects. If the position and type of objects can be well detected, the utility level of visual data is considered to be higher than in case of insufficiently detected objects. For example, if an unattended baggage at an airport can be clearly localised in privacy protected visual data, then the utility level is not decreased significantly due to privacy protection. Below, we provide a detailed explanation on how to evaluate utility in visual data based on independent, aggregated and fused frames.

1. *Independent frames*
   Calculating the detection rate on a frame-by-frame basis can be done by comparing the detected objects to the ground-truth in each frame ${}^{n}\tilde{v}^{i}$ of each video ${}^{n}\widetilde{V}$ from $\widetilde{\mathcal{V}}$. If the bounding box ${}^{n}b^{i}_{det_{j_d}}$ of the detected object is sufficiently close to the annotated object ${}^{n}b^{i}_{gt_j}$ and their description is the same ${}^{n}d^{i}_{gt_j} = {}^{n}d^{i}_{det_{j_d}}$, the detection is considered to be successful. The output of the function $f_{det_{ind}}$ is the ratio between the number of successfully detected objects in $\widetilde{\mathcal{V}}$ and the number of all annotated objects in $\mathcal{GT}$.

$$f_{det_{ind}}\left(\widetilde{\mathcal{V}}, \mathcal{GT}\right) = \frac{1}{N \cdot {}^{n}L} \sum_{n=1}^{N} \sum_{i=1}^{{}^{n}L} h_{{}^{n}\tilde{v}^{i}} \tag{20}$$

   $h_{{}^{n}\tilde{v}^{i}}$ represents the number of successful detections (hits) in ${}^{n}\tilde{v}^{i}$ and is calculated by the following algorithm.

```
$h := 0$;
${}^{n}O^{i}_{det} := \texttt{detect}\left({}^{n}\tilde{v}^{i}\right)$;
for $\forall {}^{n}o^{i}_{det_{j_d}}|_{j_d:=1\ldots J_d} \in {}^{n}O^{i}_{det}$ do
  if $\exists {}^{n}o^{i}_{gt_j}$ where ${}^{n}b^{i}_{gt_j} \approx {}^{n}b^{i}_{det_{j_d}}$ and ${}^{n}d^{i}_{gt_j} = {}^{n}d^{i}_{det_{j_d}}$ then
    $h{+}{+}$;
  end if
end for
if $J_{{}^{n}\tilde{v}^{i}} = 0$ then
  $h_{{}^{n}\tilde{v}^{i}} := 1$;
else
```

[2] The evaluator functions can be easily modified/extended to represent different utility aspects such as pleasantness or intelligibility of visual data (e.g., [9]).

```
    $h_{{}^n\tilde{v}^i} := \frac{h}{J_{{}^n\tilde{v}^i}}$;
  end if
```

The function `detect()` performs object detection on a given frame and returns a set of object annotations about the detected objects, namely their bounding boxes and descriptions. As previously explained for the `recognise()` function, the `detect()` function is not bound to any specific algorithm. Any suitable detection algorithm for the object type and the requirements of the evaluation can be used for the framework implementation. For example, the Viola-Jones face detector [55] is widely used if faces are the objects of interest. J_d is the number of objects detected by the detector and $J_{{}^n\tilde{v}^i}$ is the number of objects actually appearing in frame ${}^n\tilde{v}^i$ according to the ground-truth ${}^nO^i_{gt}$.

2. *Aggregated frames*

In case of independent frames we used only the information available at the given frame. Here we use the information from all available frames together for the detection. The performance of a generally trained object detector can be increased by adapting its model specifically to the input data. Thus, before we perform the evaluation, we further train the detector with aggregated frames using the following algorithm.

```
for $\forall {}^n\tilde{v}^i|_{n:=1\ldots N, i:=1\ldots {}^nL}$ do
  ${}^nO^i_{det}$ := detect $({}^n\tilde{v}^i)$;
  for $\forall {}^no^i_{det_{j_d}}|_{j_d:=1\ldots J_d} \in {}^nO_{det}$ do
    if $\exists {}^no^i_{gt_j}$ where ${}^nb^i_{gt_j} \approx {}^nb^i_{det_{j_d}}$ and ${}^nd^i_{gt_j} = {}^nd^i_{det_{j_d}}$ then
      update $(detector)$;
    end if
  end for
end for
```

J_d is the number of objects detected by the detector in the current frame $({}^n\tilde{v}^i)$ and the `update()` function is responsible for updating the detector's model. This process requires stored visual data. If the evaluation framework would be used in an on-line manner, the detector's model could only be updated on the fly. After adapting the detector to the input data, the measurement can be done similarly to independent frames.

$$f_{det_{aggr}}\left(\tilde{\mathcal{V}}, \mathcal{GT}\right) = \frac{1}{N \cdot {}^nL} \sum_{n=1}^{N} \sum_{i=1}^{{}^nL} h_{{}^n\tilde{v}^i} \tag{21}$$

$h_{{}^n\tilde{v}^i}$ is the hit-rate of the detector in the privacy protected frame ${}^n\tilde{v}^i$ and is calculated by the following algorithm.

```
h := 0;
${}^nO^i_{det}$ := detect $({}^n\tilde{v}^i)$;
for $\forall {}^no^i_{det_{j_d}}|_{j_d:=1\ldots J_d} \in {}^nO^i_{det}$ do
  if $\exists {}^no^i_{gt_j}$ where ${}^nb^i_{gt_j} \approx {}^nb^i_{det_{j_d}}$ and ${}^nd^i_{gt_j} = {}^nd^i_{det_{j_d}}$ then
    $h$++;
  end if
end for
if $J_{{}^n\tilde{v}^i} = 0$ then
  $h_{{}^n\tilde{v}^i} := 1$;
  else
    $h_{^n\tilde{v}^i} := \frac{h}{J_{^n\tilde{v}^i}}$;
  end if
```

The function detect() performs object detection on a given frame and returns a set of object annotations about the detected objects, namely their bounding boxes and descriptions. J_d is the number of objects detected by the detector and $J_{^n\tilde{v}^i}$ is the number of objects actually appearing in frame ${}^n\tilde{v}^i$ based on the ground-truth ${}^nO^i_{gt}$.

3. *Fused frames*

   Frames constructed by combining multiple independent frames can also be used to enhance the detector. Before performing the evaluation, the detector is further trained as in case of aggregated frames. However, fused frames are used instead of multiple independent frames. The preliminary detector training can be performed by the following algorithm.

```
  $\mathcal{FI}$ := { set of fused images }
  for $\forall \tilde{v}_{FI} \in \mathcal{FI}$ do
    $O_{det_{FI}}$ := detect ($\tilde{v}_{FI}$);
    for $\forall o_{det_{j_d}}|_{j_d:=1...J_d} \in O_{det_{FI}}$ do
      if $\exists o_{gt_j}$ where $b_{gt_j} \approx b_{det_{j_d}}$ and $d_{gt_j} = d_{det_{j_d}}$ then
        update (detector);
      end if
    end for
  end for
```

   J_d is the number of objects detected by the detector in the current fused frame $\tilde{v}_{FI}$ and the update() function is responsible for updating the detector's model. After the detector has been adapted to the input data, the measurement can be done as described below.

$$f_{det_{fused}}\left(\widetilde{\mathcal{V}}, \mathcal{GT}\right) = \frac{1}{N \cdot L_n} \sum_{n=1}^{N} \sum_{i=1}^{L_n} h_{^n\tilde{v}^i} \tag{22}$$

   $h_{^n\tilde{v}_i}$ is calculated by the same algorithm as for Equation 21.

Utility Evaluation by Object Tracking

Another way of utility evaluation is to apply tracking algorithms to the privacy protected input data. For instance in retail surveillance, the customers' traces in the shop is a very useful information. However, tracking should only be performed on the protected visual data in order not to reveal the customers' identities. We only consider aggregated frames in terms of tracking. Aggregated frames can originate either from a single camera or from multiple cameras. The task of a tracking algorithm is basically to detect and "‘follow" selected objects across various frames over time in a video sequence or over different videos from multiple cameras. Trackers usually rely on a model that stores all knowledge about objects that are initially handed over to the tracker. This model is continuously updated after each processed frame and used to estimate the objects' positions in the next frame. Measuring the accuracy of a tracking algorithm can be performed by comparing the trackers output with the ground-truth [48]. Tracking is considered to be successful if an object's location and description provided by the tracker matches the ground-truth data. The function f_{track} can be defined as follows:

$$f_{track}\left(\widetilde{\mathcal{V}},\mathcal{GT},\mathcal{M}\right)=\frac{1}{N\cdot {}^{n}L}\sum_{n=1}^{N}\sum_{i=1}^{{}^{n}L}h_{{}^{n}\tilde{v}^{i}} \tag{23}$$

where $\mathcal{M}$ is the model of the tracker. $h_{{}^{n}\tilde{v}^{i}}$ stands for the hit-rate of the tracker and is calculated with the algorithm below.

```
$h := 0;$
${}^{n}O^{i}_{track} :=$ track $\left({}^{n}\tilde{v}^{i}\right);$
for $\forall {}^{n}o^{i}_{track_{j_t}} |_{j_t:=1\ldots J_t} \in {}^{n}O^{i}_{track}$ do
  if $\exists {}^{n}o^{i}_{gt_j}$ where ${}^{n}b^{i}_{gt_j} \approx {}^{n}b^{i}_{track_{j_t}}$ and ${}^{n}d^{i}_{gt_j} = {}^{n}d^{i}_{track_{j_t}}$ then
    $h{+}{+};$
  end if
end for
if $J_{{}^{n}\tilde{v}^{i}} = 0$ then
  $h_{{}^{n}\tilde{v}^{i}} := 1;$
else
  $h_{{}^{n}\tilde{v}^{i}} := \frac{h}{J_{{}^{n}\tilde{v}^{i}}};$
end if
update $(\mathcal{M});$
```

The function `track`() performs object detection in the current frame based on object information in $\mathcal{M}$ and the previous frame, and returns a set of annotations about the tracked objects. The `track`() function is not bound to any specific tracking algorithm. Any suitable algorithm that fits the requirements of the evaluation scenario can be used for the concrete framework implementation (e.g., [34]). J_t is the number of objects tracked by the tracker and $J_{{}^{n}\tilde{v}^{i}}$ is the number of objects actually appearing in the protected frame ${}^{n}\tilde{v}_{i}$ according to the ground-truth ${}^{n}O^{i}_{gt}$ while the `update`() function is responsible for updating the tracker's model $\mathcal{M}$.

Utility Evaluation by Image Similarity

Another utility measurement is to visually compare the privacy protected video to the unprotected video by using image similarity metrics. The similarity corresponds to the deviation of the unprotected from the protected data. Such deviation can be measured by the differences in pixel intensities or the mean and variance values of intensity values in specific image regions. The output of the function f_{sim} is basically the average of the similarities between each unprotected ${}^{n}v^{i}$ and protected ${}^{n}\tilde{v}^{i}$ frame in each video ${}^{n}V$ and ${}^{n}\widetilde{V}$ from $\mathcal{V}$ and $\widetilde{\mathcal{V}}$ respectively. These metrics work solely on a frame-by-frame basis, and therefore aggregated and fused frames are not discussed here.

$$f_{sim}\left(\mathcal{V},\widetilde{\mathcal{V}}\right)=\frac{1}{N\cdot {}^{n}L}\sum_{n=1}^{N}\sum_{i=1}^{{}^{n}L}\texttt{similarity}\left({}^{n}v^{i},{}^{n}\tilde{v}^{i}\right) \tag{24}$$

For the function `similarity`(), a specific similarity metric which returns the extent of similarity between two given image frames must be chosen (e.g., the structural similarity index SSIM [56]).

Utility Evaluation by Processing Speed

Some privacy protection filters are computationally expensive and cannot be applied in real time. In terms of utility this can be an important issue because online protection of visual data is often required or the protection should be performed onboard of the cameras.. We measure the processing speed of privacy protection filters in order to make our evaluation framework as comprehensive as possible. This speed does not only depend on the computational complexity of the filter's algorithm, but also on the image resolution and the computing power of the underlying hardware. Depending on the requirements of the surveillance scenario a target speed (τ) can be chosen arbitrarily. The processing speed of privacy protection filters can be measured for example in frames per second (FPS). The function f_{speed} can therefore be calculated as follows:

$$f_{speed}\left(\widetilde{\mathcal{V}}\right) = \frac{1}{N \cdot {}^{n}L} \sum_{n=1}^{N} \sum_{i=1}^{{}^{n}L} max\left(\frac{1}{\tau \cdot \left(t\left({}^{n}\tilde{v}^{i}\right) - t\left({}^{n}\tilde{v}^{i-1}\right)\right)}, 1\right) \tag{25}$$

where τ is the arbitrary target speed of the filter. The function $t()$ returns the time when the processing of a given image frame was finished.

5 Implementation and Test of the Framework Prototype

We have developed one possible implementation of the previously defined evaluation framework using standard algorithms for object recognition, detection and tracking from OpenCV [25]. With this prototype implementation we demonstrate the capabilities of our approach and compare objective and subjective evaluation techniques. In the following subsections we describe implementation details of our prototype and present measurement results.

5.1 Framework Implementation

The goal of our implementation is to present objective measurement results based on various state-of-the-art privacy protection algorithms. Therefore, we have implemented the following functions (as described in Sections 3 and 4):

- $f_{id_{ind}}$, $f_{id_{aggr}}$, and $f_{id_{fused}}$ by using the PCA [54], LDA [10] and LBP [3] based face recognisers,
- $f_{det_{ind}}$, $f_{det_{aggr}}$, and $f_{det_{fused}}$ by using the cascade classifier based face detection module and the histogram of oriented gradients (HOG) based person detector,
- f_{track} by using the MIL, Boosting, MedianFlow and TLD object trackers, and
- f_{sim} by calculating MSE (mean squared error) and SSIM (structural similarity) index.

5.2 Test Data

We used our evaluator prototype to objectively evaluate eight privacy protection filters proposed at the MediaEval 2014 Workshop [9]. Figure 2 demonstrates the visual effects of the eight different protection filters. The key objective of these protection filters was to protect the privacy of the persons but still keep the "intelligibility" and "visual appearance" high. In order to evaluate the performance among these categories the Visual Privacy Task organisers of the MediaEval 2014 Workshop carried out a user study. In this paper we compare our objective and their subjective evaluation results in order to demonstrate the pertinence of our proposed framework.

The subjective evaluation was based on a subset of the PEViD dataset [30] which originally contains 65 full HD (1920×1080, 25 fps, 16 seconds each) video sequences covering a broad range of surveillance scenarios. The video clips are annotated by the ViPER GT tool [1] which produces XML files containing the ground-truth and general information about the surveillance scenario (walking, fighting, etc.). The Visual Privacy Task organisers selected six particular video clips from the PEViD dataset [30] for their subjective evaluation including day/night, indoor/outdoor and close-up/wide area scenarios. The dataset further included the ground-truth for every image frame, i.e., bounding boxes around faces, hair regions, skin regions, body regions and accessories.

The user study was conducted on the submitted privacy protected videos of eight research teams evaluating and investigated aspects such as privacy, intelligibility and pleasantness by means of questionnaires [9]. The protected videos were evaluated by three different user groups: (i) an online, crowd-sourced evaluation by the general public, (ii) an evaluation by security system manufacturers and video-analysis technology and privacy protection solutions developers, and (iii) an on-line evaluation by a target group comprising trained CCTV monitoring professionals and law enforcement personnel.

Our objective evaluation is based on the following setting.

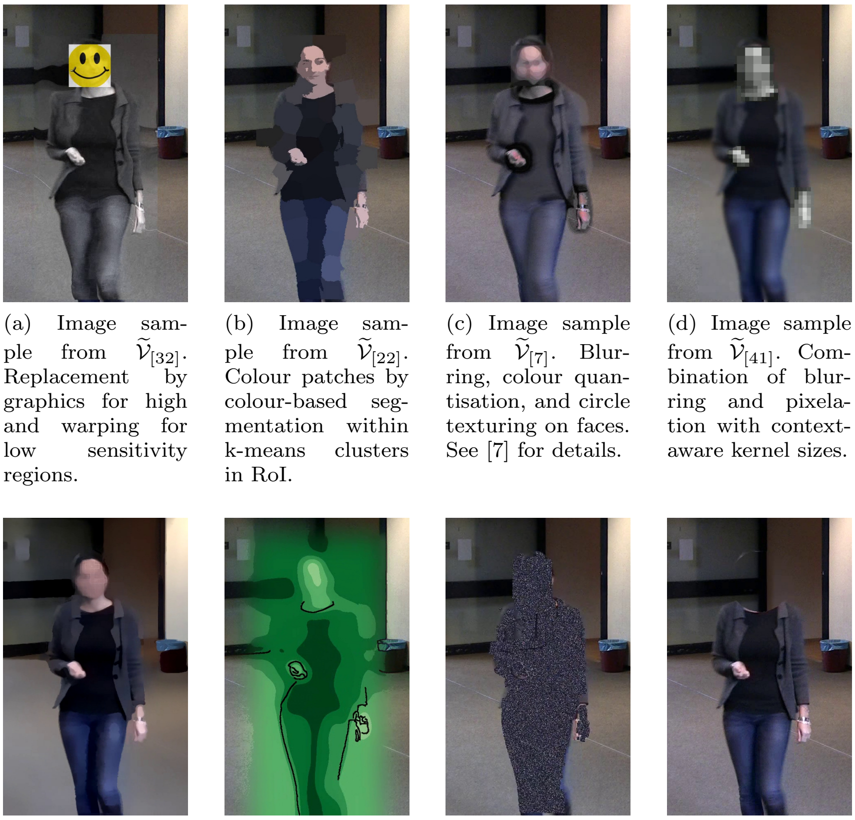
Input:

- The same six selected video clips from the PEViD dataset [30] served as unprotected input videos. Each clip is in full HD resolution (1920×1080) and contains 400 image frames.
  $\mathcal{V} = \left\{{}^{1}V, {}^{2}V, {}^{3}V, {}^{4}V, {}^{5}V, {}^{6}V\right\}$ where ${}^{i}L = 400|_{i=1,\ldots,6}$
- Ground-truth data was also used in the evaluation process. It is provided by the PEViD dataset [30] for each video clip in ViPER XML [1] format.
  $\mathcal{GT} = \left\{{}^{1}\mathcal{O}_{gt}, {}^{2}\mathcal{O}_{gt}, {}^{3}\mathcal{O}_{gt}, {}^{4}\mathcal{O}_{gt}, {}^{5}\mathcal{O}_{gt}, {}^{6}\mathcal{O}_{gt}\right\}$
- Furthermore, we used the privacy protected version of each video clip filtered by the privacy-preserving methods [32, 22, 7, 41, 21, 36, 50, 42] proposed at the MediaEval 2014 Workshop [9].

Output:

- A set of real numbers between $[0, 1]$ provided by the evaluator functions, where zero represents the worst and one the best result.
  $\mathcal{E} = \{e_{id_{ind}}, e_{id_{aggr}}, e_{id_{fused}}, e_{det_{ind}}, e_{det_{aggr}}, e_{det_{fused}}, e_{track}, e_{sim}, e_{speed}\}$ where $\forall e \in \mathbb{R}$ and $0 \leq e \leq 1$.

In the following subsections we describe the details of each implemented function and discuss the produced results.



(a) Image sample from $\widetilde{\mathcal{V}}_{[32]}$. Replacement by graphics for high and warping for low sensitivity regions.

(b) Image sample from $\widetilde{\mathcal{V}}_{[22]}$. Colour patches by colour-based segmentation within k-means clusters in RoI.

(c) Image sample from $\widetilde{\mathcal{V}}_{[7]}$. Blurring, colour quantisation, and circle texturing on faces. See [7] for details.

(d) Image sample from $\widetilde{\mathcal{V}}_{[41]}$. Combination of blurring and pixelation with context-aware kernel sizes.

(e) Image sample from $\widetilde{\mathcal{V}}_{[21]}$. Global and local multi-level cartooning with extra pixelation on faces.

(f) Image sample from $\widetilde{\mathcal{V}}_{[36]}$. Blurring and colour remapping with silhouettes and special colours for various events.

(g) Image sample from $\widetilde{\mathcal{V}}_{[50]}$. Pseudo-randomly scrambled pixels within foreground masks.

(h) Image sample from $\widetilde{\mathcal{V}}_{[42]}$. Inpainting of faces with background estimated by median filtering.

Fig. 2: Image samples of each privacy filter proposed at the MediaEval 2014 Workshop [9].

5.3 Evaluation of Privacy

In Section 4 we have defined our evaluation framework by using general object recognisers. The most critical objects are however faces in terms of privacy. Therefore, in our current prototype we focused on faces when evaluating visual privacy and used the PCA [54], LDA [10], and LBP [3] based face recogniser functions from OpenCV.

In our current prototype we have implemented de-identification evaluator functions for independent ($f_{id_{ind}}$), aggregated ($f_{id_{aggr}}$), and fused ($f_{id_{fused}}$) frames by using the above mentioned face recogniser tools. We have used all valid faces from the six unprotected input videos (${}^{1}V, \ldots, {}^{6}V$) as a training set for the face recognisers. By valid faces we mean those 766 faces from the 2400 video frames where both eyes are visible. We need both eyes in order to correctly align and resize faces because OpenCV's face recognisers require aligned faces and equal input image sizes. The position of faces and eyes were taken from the ground-truth data and the output of the face recognisers were also compared with the ground-truth during the evaluation process.

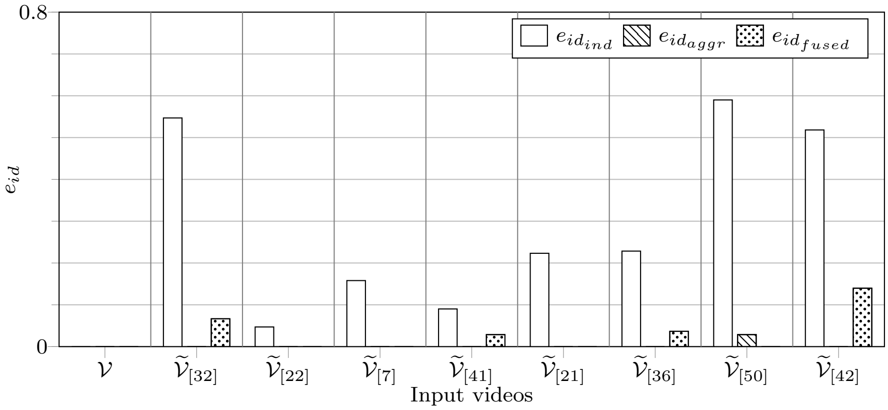


Fig. 3: Privacy evaluation results for independent, aggregated, and fused frames. No privacy protection can be observed for the unprotected videos $\mathcal{V}$ and only $\widetilde{\mathcal{V}}_{[50]}$ provides some protection when using aggregated frames while protection levels remain zero for all the other videos. Results for fused frames are also significantly lower than for independent frames.

After training the three face recognisers we tested them on the same 766 valid face regions of the privacy filtered videos from $\widetilde{\mathcal{V}}_{[32,\ 22,\ 7,\ 41,\ 21,\ 36,\ 50,\ 42]}$. At each frame we chose the best-performing recogniser. This measurement provided the results for independent frames. In case of aggregated frames we performed further calculations according to the rules defined by Equation 17 in Section 4.1. Namely, we considered all the occurrences of a certain face as recognised when it was successfully recognised at least once during the evaluation. When following the fused frames approach, again, we carried out our calculations based on the algorithm defined under Equation 18 in Section 4.1. The set of fused frames were created as follows. We grouped the 766 valid face images per person based on the structural similarity (SSIM) index. Those face images got placed in one group which were at least 70% similar to each other (i.e., SSIM ≥ 0.7). Within each group we created image pairs in every possible combination and fused them pair-wise based on two-level discrete stationary wavelet transform [43]. These fused images constituted the set of fused frames ($\mathcal{FI}$). Figure 3 shows the calculated privacy evaluation results for independent, aggregated, and fused frames. When evaluating the unprotected videos the results are $e_{id_{ind}} = 0$, $e_{id_{aggr}} = 0$, and $e_{id_{fused}} = 0$, which refers to no privacy protection. That is expected since we used the faces from these unprotected videos to train the face recognisers and thereby those faces can be recognised with 100% accuracy. The privacy filter from Paralic *et al.* [42] inpaints all faces with the background, therefore it is somewhat surprising that $e_{id_{ind}} = 0.52$, $e_{id_{aggr}} = 0$, and $e_{id_{fused}} = 0.14$ only while these values are expected to be close to 1 as there are no faces to recognise at all. A possible explanation is that the face recognisers we used always provide an output and with a certain probability they may still guess the right face identity. Furthermore, the inpainted background may also contain face-like structures that are similar to the face to be recognised from the face recognisers' point of view. Another interesting observation about the evaluation results is that $\widetilde{\mathcal{V}}_{[50]}$ is the only one providing some low-level privacy protection in case of aggregated frames while all the others provide no protection. Furthermore, results in terms of fused frames are significantly lower than in case of individual frames and they are very close or equal to zero several times.

A subjective evaluation described in Sections 3.1 and 3.2 of [9] has been carried out as part of the MediaEval 2014 Workshop. The privacy-preserving methods from [32, 22, 7, 41, 21, 36, 50, 42] have been evaluated in three distinct user studies. The *first* study followed a crowd-sourcing approach targeting naïve subjects from online communities. The *second* study targeted the trained video surveillance staff of Thales, France. A focus group comprising video-analytics technology and privacy protection solution developers was the target of the *third* study. Hereinafter, we refer to the privacy protection level results of these three studies as p_{crowd}, p_{thales}, and p_{focus}, respectively, while i_{crowd}, i_{thales}, and i_{focus} refer to the intelligibility levels. In the following we compare our measurement results with the outcome of the MediaEval study to see if our *objective* method complies with their *subjective* approach.

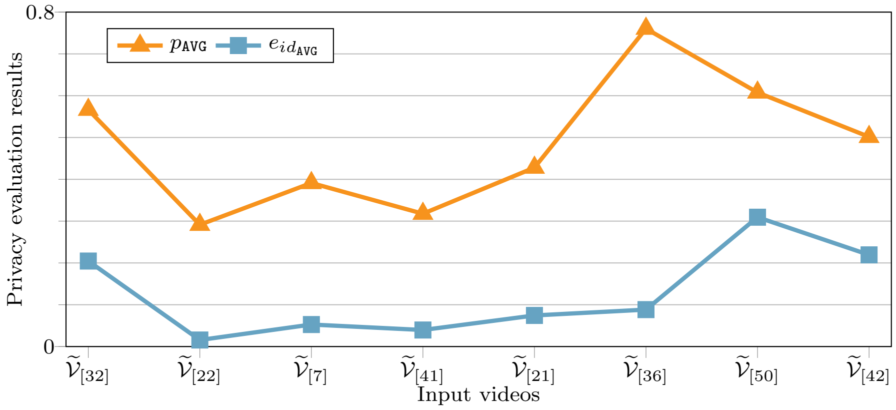


Fig. 4: Comparison of objective and subjective privacy evaluation results where $p_{\texttt{AVG}} = \texttt{AVERAGE}(p_{crowd}, p_{thales}, p_{focus})$ and $e_{id_{\texttt{AVG}}} = \texttt{AVERAGE}(e_{id_{ind}}, e_{id_{aggr}}, e_{id_{fused}})$.

In order to compare our objective ($e_{id_{ind}}$, $e_{id_{aggr}}$, $e_{id_{fused}}$) and the subjective privacy evaluation results (p_{crowd}, p_{thales}, p_{focus}) from [9], we plotted the average values together in a single chart which can be seen in Figure 4. It is clearly visible that objective and subjective results follow the same trend except one deviation at $\widetilde{\mathcal{V}}_{[36]}$. The privacy filter from [36] replaces the whole body of each person with a blurry colour blob which obscures original shapes as well. While our objective method considered only faces, human viewers usually watch the entire body. They may find privacy protection better in this case because there is not even any secondary information (e.g., body shape or clothes) available to identify people. Our result for $\widetilde{\mathcal{V}}_{[36]}$ is lower because the face recognisers achieved a higher recognition rate. This is due to the already mentioned fact that the recognisers always provide an output and with a certain probability they can still guess the identities properly, especially in case of such a small population (10 people in the dataset). Although the plots are following the same trend, a certain offset between objective and subjective results can be observed. This is due to the differences in the nature of measurements and in the scaling of the extracted data. The Pearson product-moment correlation coefficient [51][3] for the subjective and objective privacy evaluation results results in a value of 0.563 which indicates a rather strong positive correlation. If we exclude the above described outlier case of $\widetilde{\mathcal{V}}_{[36]}$, the coefficient value increase to 0.95 which indicates a very strong positive correlation.

Table 2 compares the ranking of the subjective evaluation conducted by [9] and the ranking achieved by our objective evaluation framework. The rankings are based on the average privacy metrics $p_{\texttt{AVG}}$ and $e_{id_{\texttt{AVG}}}$, respectively (cp. Figure 4). As can be clearly seen, the subjective and our objective evaluation methods achieve highly correlated results for the used MediaEval 2014 test data. The strong positive correlation of both rankings are also indicated by the Spearman and the Kendall rank correlation coefficients [51] which are given as $\rho = 0.850$ and $\tau = 0.764$, respectively.

Protection Filter	$\widetilde{\mathcal{V}}_{[22]}$	$\widetilde{\mathcal{V}}_{[41]}$	$\widetilde{\mathcal{V}}_{[7]}$	$\widetilde{\mathcal{V}}_{[21]}$	$\widetilde{\mathcal{V}}_{[42]}$	$\widetilde{\mathcal{V}}_{[32]}$	$\widetilde{\mathcal{V}}_{[50]}$	$\widetilde{\mathcal{V}}_{[36]}$
Subjective Ranking	1.	2.	3.	4.	5.	6.	7.	8.
Protection Filter	$\widetilde{\mathcal{V}}_{[22]}$	$\widetilde{\mathcal{V}}_{[41]}$	$\widetilde{\mathcal{V}}_{[7]}$	$\widetilde{\mathcal{V}}_{[21]}$	$\widetilde{\mathcal{V}}_{[36]}$	$\widetilde{\mathcal{V}}_{[32]}$	$\widetilde{\mathcal{V}}_{[42]}$	$\widetilde{\mathcal{V}}_{[50]}$
Objective Ranking	1.	2.	3.	4.	8.	6.	5.	7.
Spearman coefficient ρ	0.850							
Kendall coefficient τ	0.764							

Table 2: Ranking of protection methods based on the subjective privacy evaluation results presented in [9] and our objective evaluation results produced by our prototype together with their Spearman's and Kendall's [51] rank correlation coefficient.

5.4 Evaluation of Utility

Implementation details and measurement results are discussed in the following subsections. Similarly to the above described privacy evaluation, instead of using objects in general we specified certain object types for each evaluation function to keep our first prototype simple.

[3] The Pearson product-moment correlation coefficient is a measure of the linear dependence between two variables in the range of $[-1 \ldots +1]$, where $+1$ represents a total positive linear correlation, 0 no linear correlation, and -1 a total negative linear correlation.

5.4.1 Detection

For utility evaluation by object detection we chose faces and bodies as target objects. We used the face detection functionality of OpenCV [25] which is based on Haar-cascades. For person detection, we used the histogram of oriented gradients (HOG) based detector from OpenCV [25]. We used all six videos protected by the eight privacy-preserving methods [32, 22, 7, 41, 21, 36, 50, 42] along with their unprotected version as an input for the above mentioned detectors. Similarly to privacy evaluation, here we also compared the output of the detectors with the ground-truth data. If a bounding box of a detected face or person was sufficiently overlapping with the annotated bounding box from the ground-truth data, we counted that detection as a hit. We call two bounding boxes sufficiently overlapping if their Sørensen-Dice coefficient is greater than 0.5. This criteria can be formulated as follows:

$$\frac{2 \cdot A_{b_{det} \cap b_{gt}}}{A_{b_{det}} + A_{b_{gt}}} > 0.5 \tag{26}$$

where A refers to the area of a bounding box while b_{det} and b_{gt} represent detected and annotated bounding boxes, respectively. Then, we calculated the evaluation results for independent ($e_{detF_{ind}}$), aggregated ($e_{detF_{aggr}}$), and fused ($e_{detF_{fused}}$) frames in terms of face detection which are depicted in Figure 5. Figure 6 shows evaluation results for independent frames ($e_{detP_{ind}}$) based on the HOG person detector. Here we only considered independent frames because OpenCV [25] does not have an option for updating or retraining the HOG detector's model.

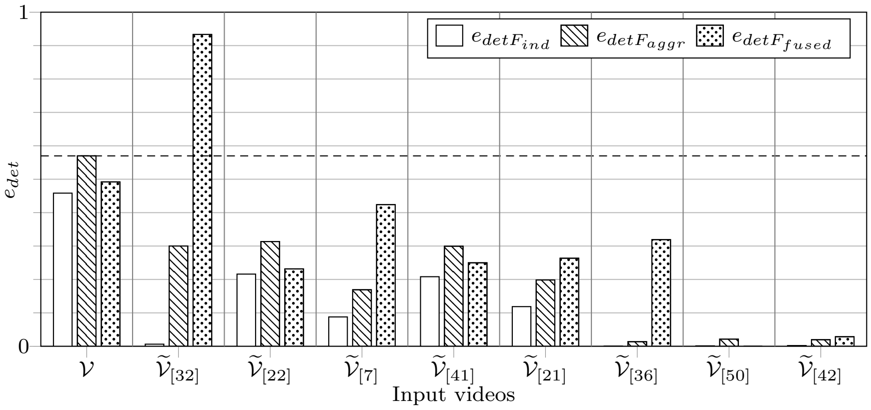


Fig. 5: Results of utility evaluation by *face detection* for independent, aggregated, and fused frames. The dashed line marks the highest utility level of the unprotected videos.

The overall best utility in terms of face detection is obviously provided by the unprotected videos ($\mathcal{V}$). The privacy filters from [32], [36], [50], and [42] totally replace faces, thereby providing the worst utility levels in terms of independent frames. In case of $\widetilde{\mathcal{V}}_{[22]}$, $\widetilde{\mathcal{V}}_{[7]}$, $\widetilde{\mathcal{V}}_{[41]}$, and $\widetilde{\mathcal{V}}_{[21]}$ a certain utility level can still be achieved along privacy protection. Furthermore, face detection performance and hence the utility level is always higher when considering aggregated frames and even higher for fused frames. This is expected because in case of aggregated and fused frames the face detector's model is extended by using specific training samples from the relevant protected video clips. An outstanding result can be observed at $\widetilde{\mathcal{V}}_{[32]}$ where $e_{detF_{fused}}$ is significantly higher than the results for the unprotected videos ($\mathcal{V}$). This suggests that despite the information loss caused by the application of privacy protection methods, the utility level can even be increased. Both for aggregated and fused frames we followed the algorithms defined in Section 4.2.1 and the set of fused frames were created exactly the same way as described above for privacy evaluation. The detector's model was updated by using the `opencv_traincascade` utility from OpenCV [25].

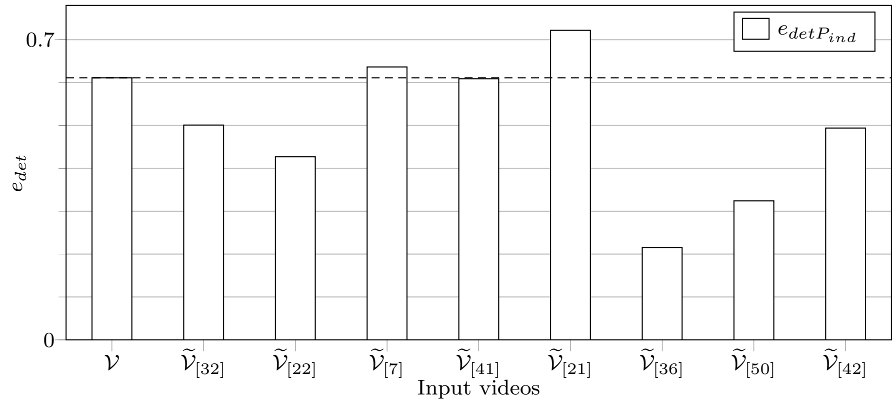


Fig. 6: Results of utility evaluation by *person detection* for independent frames. The dashed line marks the utility level of the unprotected videos.

As for person detection, the results are higher for $\widetilde{\mathcal{V}}_{[7]}$ and $\widetilde{\mathcal{V}}_{[21]}$ than for the unprotected video ($\mathcal{V}$). This means that the utility level in visual data can not only be maintained but can even be further increased while protecting privacy. We find this a quite important message for privacy protection filter developers. The lowest result is provided by $\widetilde{\mathcal{V}}_{[36]}$ which is not surprising at all considering the large amount of changes in terms of both colour and visual structure (see Figure 2f).

5.4.2 Tracking

When evaluating utility by object tracking we used the whole bodies of people as target objects. We used the following 4 trackers that are implemented in OpenCV [25]: MIL [6], Boosting [23], MedianFlow [26], and TLD [27]. We fused the results of these trackers by always choosing the best performing tracker per frame similarly to our approach regarding face recognisers. Tracking is considered to be successful in a frame if the output of a tracker is sufficiently overlapping with the annotated bounding box from the ground-truth data. Again, we consider an overlapping sufficient if the Sørensen-Dice coefficient is greater than 0.5. Figure 7 shows our results for utility evaluation through object tracking (e_{track}). Several privacy protected videos achieved slightly better utility results than the unprotected videos which further supports the fact that utility can be improved even when obfuscating the unprotected visual data for the sake of privacy protection. However, differences are not too significant between the protection techniques and there is no outstanding result. Tracking performance is almost equal in each case.

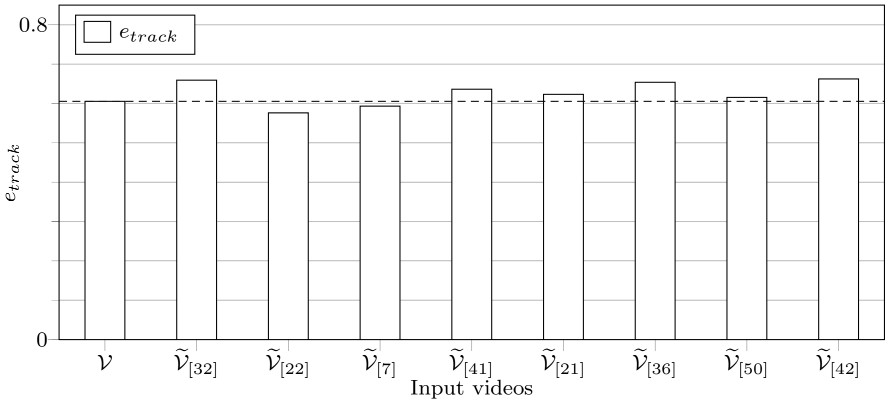


Fig. 7: Results of utility evaluation by *object tracking*. The level of the unprotected videos is marked with the dashed line.

5.4.3 Similarity

As part of utility evaluation we measured visual similarity by calculating the mean squared error (MSE) and the structural similarity (SSIM) index for the protected videos ($\widetilde{\mathcal{V}}_{[32,\ 22,\ 7,\ 41,\ 21,\ 36,\ 50,\ 42]}$) compared to the unprotected videos ($\mathcal{V}$). Differences between the privacy protected videos in terms of mean squared error are very small. All similarity results are within the $[0.99, 1]$ interval. Therefore, all types of protected videos are considered to be very similar to the unprotected videos based on this metric. This fact suggests that MSE is not a suitable metric when comparing privacy protection filters.

Our measurement results regarding structural similarity are depicted in Figure 8. $\widetilde{\mathcal{V}}_{[21]}$ shows the most substantial difference from the unprotected videos ($\mathcal{V}$). The global modifications carried out by the privacy filter cause notably large changes in the image structure which explains the extent of dissimilarity.

Our definition of *utility* and the way Badii *et al.* [9] define *intelligibility* is rather different. We measure quite different things by using computer vision methods than they do with their questionnaires. Thus, while comparing objective and subjective evaluation results for utility in Figure 9, it is not surprising that no correlation can be observed between objective and subjective results. Figure 9 depicts the average of the intelligibility results i_{crowd}, i_{thales}, and i_{focus} together with our objective evaluation results regarding utility (e_{detF}, e_{detP}, e_{track}, $e_{sim_{\mathrm{SSIM}}}$).

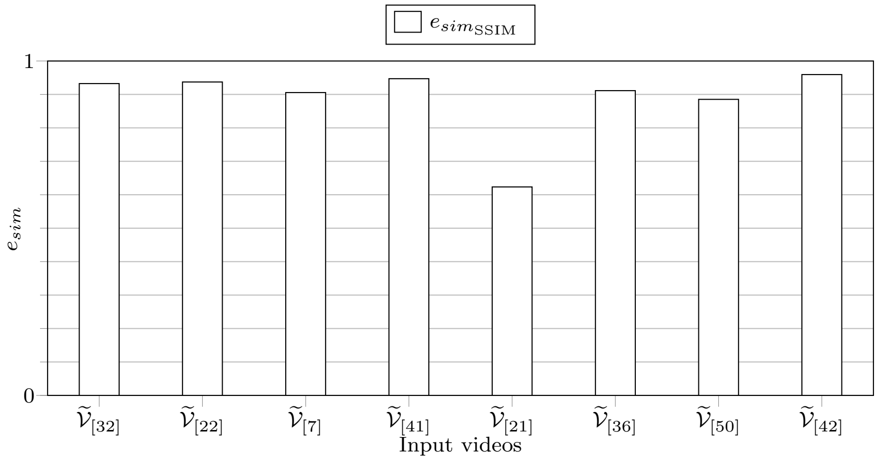


Fig. 8: Results of utility evaluation by measuring visual similarity to the unprotected video $\mathcal{V}$ when using the structural similarity index ($e_{sim_{\mathrm{SSIM}}}$).

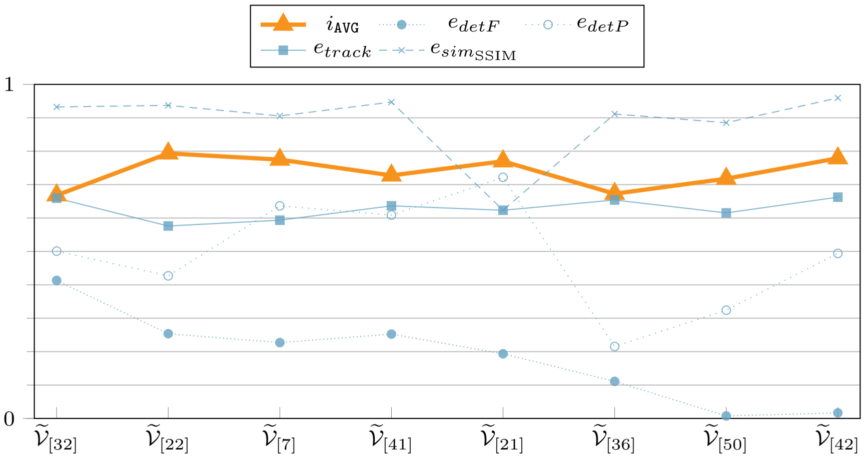


Fig. 9: Comparison of objective and subjective utility evaluation results where $i_{\mathtt{AVG}} = \mathtt{AVERAGE}(i_{crowd}, i_{thales}, i_{focus})$ and $e_{detF} = \mathtt{AVERAGE}(e_{detF_{ind}}, e_{detF_{aggr}}, e_{detF_{fused}})$.

6 Conclusions and Future Work

We have proposed an objective visual privacy evaluation framework that considers a rather wide variety of aspects including the use of aggregated and fused frames as opposed to traditional frame-by-frame assessment methods. A formal definition has been provided by which reproducible results can be measured. This framework is based on a general definition of privacy protection and utility, and can be used to benchmark various protection techniques. Thus, our framework may serve as a useful tool for developers of visual privacy-preserving techniques. We have applied this framework to state-of-the-art privacy protection methods and compared our results to a recently conducted subjective evaluation. For privacy protection, subjective and objective evaluation results show a high correlation.

A possibility for future work is to conduct a survey with an even larger number of participants and compare these subjective results with the output of the proposed objective framework. Then the definitions of the measured aspects within the framework could also be fine-tuned in order to better approximate subjective results. Another possible task for the future is to create a more comprehensive implementation of our objective evaluation framework in form of an on-line API which would make our work useful to the research community.

Acknowledgements This work was performed as part of the *TrustEYE: Trustworthy Sensing and Cooperation in Visual Sensor Networks* project [57] and received funding from the European Regional Development Fund (ERDF) and the Carinthian Economic Promotion Fund (KWF) under grant KWF-3520/23312/35521.

References

1. ViPER XML: A Video Description Format. `http://viper-toolkit.sourceforge.net/docs/file/`. Last accessed: November 2016
2. Online IP netsurveillance cameras of the world. `http://www.insecam.org/` (2014). Last accessed: November 2016
3. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(12), 2037–2041 (2006)
4. Anderson, S.: Privacy by design: An assessment of law enforcement drones. Ph.D. thesis, Georgetown University (2014)
5. Aved, A.J., Hua, K.A.: A general framework for managing and processing live video data with privacy protection. Multimedia systems **18**(2), 123–143 (2012)
6. Babenko, B., Yang, M.H., Belongie, S.: Visual tracking with online multiple instance learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 983–990 (2009)
7. Badii, A., Al-Obaidi, A.: MediaEval 2014 Visual Privacy Task: Context-Aware Visual Privacy Protection. In: Working Notes Proceedings of the MediaEval Workshop (2014)
8. Badii, A., Al-Obaidi, A., Einig, M., Ducournau, A.: Holistic privacy impact assessment framework for video privacy filtering technologies. Signal and Image Processing: An International Journal **4**(6), 13–32 (2013)
9. Badii, A., Ebrahimi, T., Fedorczak, C., Korshunov, P., Piatrik, T., Eiselein, V., Al-Obaidi, A.: Overview of the MediaEval 2014 Visual Privacy Task. In: Proceedings of the MediaEval Workshop. Barcelona, Spain (2014)
10. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence **19**(7), 711–720 (1997)
11. Birnstill, P., Ren, D., Beyerer, J.: A User Study on Anonymization Techniques for Smart Video Surveillance. In: Proceedings of the IEEE Conference on Advanced Video and Signal-based Surveillance, pp. 1–6 (2015)

12. Bonetto, M., Korshunov, P., Ramponi, G., Ebrahimi, T.: Privacy in Mini-drone Based Video Surveillance. In: Proceedings of the Workshop on De-identification for Privacy Protection in Multimedia, p. 6 (2015)
13. Boyle, M., Edwards, C., Greenberg, S.: The effects of filtered video on awareness and privacy. In: Proceedings of the Conference on Computer Supported Cooperative Work, pp. 1–10 (2000)
14. Cavoukian, A.: Privacy by Design – The 7 Foundational Principles. `http://www.privacybydesign.ca/content/uploads/2009/08/7foundationalprinciples.pdf` (2011). Last accessed: November 2016
15. Cavoukian, A.: Surveillance, Then and Now: Securing Privacy in Public Spaces. `http://www.ipc.on.ca/images/Resources/pbd-surveillance.pdf` (2013). Last accessed: November 2016
16. Chaaraoui, A.A., Padilla-López, J.R., Ferrández-Pastor, F.J., Nieto-Hidalgo, M., Flórez-Revuelta, F.: A Vision-Based System for Intelligent Monitoring: Human Behaviour Analysis and Privacy by Context. Sensors (MDPI) **14**(5), 8895–8925 (2014)
17. Cheung, S.C.S., Venkatesh, M.V., Paruchuri, J.K., Zhao, J., Nguyen, T.: Protecting and Managing Privacy Information in Video Surveillance Systems. In: Protecting Privacy in Video Surveillance, pp. 11–33. Springer (2009)
18. Clarke, R.: The regulation of civilian drones' impacts on behavioural privacy. Computer Law & Security Review **30**(3), 286 – 305 (2014)
19. Dufaux, F., Ebrahimi, T.: A framework for the validation of privacy protection solutions in video surveillance. In: Proceedings of International Conference on Multimedia and Expo, pp. 66–71 (2010)
20. Erdélyi, A., Barát, T., Valet, P., Winkler, T., Rinner, B.: Adaptive cartooning for privacy protection in camera networks. In: Proceedings of the International Conference on Advanced Video and Signal Based Surveillance, pp. 44–49 (2014)
21. Erdélyi, Á., Winkler, T., Rinner, B.: Multi-Level Cartooning for Context-Aware Privacy Protection in Visual Sensor Networks. In: Working Notes Proceedings of the MediaEval Workshop (2014)
22. Fradi, H., Yan, Y., Dugelay, J.L.: Privacy Protection Filter Using Shape and Color Cues. In: Working Notes Proceedings of the MediaEval Workshop (2014)
23. Grabner, H., Grabner, M., Bischof, H.: Real-Time Tracking via On-line Boosting. In: Proceedings of the British Machine Vision Conference, vol. I, pp. 47–56 (2006)
24. Han, B.J., Jeong, H., Won, Y.J.: The privacy protection framework for biometric information in network based CCTV environment. In: Proceedings of the Conference on Open Systems, pp. 86–90 (2011)
25. itseez: OpenCV – Open Source Computer Vision. `http://opencv.org` (2014). Last accessed: November 2016
26. Kalal, Z., Mikolajczyk, K., Matas, J.: Forward-backward error: Automatic detection of tracking failures. In: Proceedings of the International Conference on Pattern Recognition, pp. 2756–2759 (2010)
27. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**(7), 1409–1422 (2012)
28. Korff, D., Brown, I., Blume, P., Greenleaf, G., Hoofnagle, C., Mitrou, L., Pospisil, F., Svatosova, H., Tichy, M., Anderson, R., Bowden, C.,

Nyman-Metcalf, K., Whitehouse, P.: Comparative study on different approaches to new privacy challenges, in particular in the light of technological developments. http://ec.europa.eu/justice/policies/privacy/docs/studies/new_privacy_challenges/final_report_en.pdf (2010). Last accessed: November 2016

29. Korshunov, P., Araimo, C., Simone, F., Velardo, C., Dugelay, J.L., Ebrahimi, T.: Subjective study of privacy filters in video surveillance. In: Proceedings of the International Workshop on Multimedia Signal Processing, pp. 378–382 (2012)
30. Korshunov, P., Ebrahimi, T.: PEViD: Privacy Evaluation Video Dataset. In: Proceedings of SPIE, vol. 8856 (2013)
31. Korshunov, P., Ebrahimi, T.: Using face morphing to protect privacy. In: Proceedings of the 10th International Conference on Advanced Video and Signal Based Surveillance, pp. 208–213 (2013)
32. Korshunov, P., Ebrahimi, T.: MediaEval 2014 Visual Privacy Task: Geometrical Privacy Protection Tool. In: Working Notes Proceedings of the MediaEval Workshop (2014)
33. Korshunov, P., Melle, A., Dugelay, J.L., Ebrahimi, T.: Framework for objective evaluation of privacy filters. In: Proceedings of SPIE Optical Engineering+ Applications, pp. 1–12 (2013)
34. Kristan, M., Pflugfelder, R., Leonardis, A., et al.: The visual object tracking VOT2014 challenge results. In: Proceedings of the European Conference on Computer Vision, pp. 191–217 (2014)
35. Ma, Z., Butin, D., Jaime, F., Coudert, F., Kung, A., Gayrel, C., Maña, A., Jouvray, C., Trussart, N., Grandjean, N., et al.: Towards a Multidisciplinary Framework to Include Privacy in the Design of Video Surveillance Systems. In: Privacy Technologies and Policy, pp. 101–116. Springer (2014)
36. Maniry, D., Acar, E., Albayrak, S.: TUB-IRML at MediaEval 2014 Visual Privacy Task: Privacy Filtering through Blurring and Color Remapping. In: Working Notes Proceedings of the MediaEval Workshop (2014)
37. Martin, K., Plataniotis, K.N.: Privacy protected surveillance using secure visual object coding. Transactions on Circuits and Systems for Video Technology **18**(8), 1152–1162 (2008)
38. Martinez-Balleste, A., Rashwan, H.A., Puig, D., Fullana, A.P.: Towards a trustworthy privacy in pervasive video surveillance systems. In: Proceedings of the Pervasive Computing and Communications Workshops, pp. 914–919 (2012)
39. Morando, F., Iemma, R., Raiteri, E.: Privacy evaluation: what empirical research on users valuation of personal data tells us. http://policyreview.info/articles/analysis/ (2014). Last accessed: November 2016
40. Padilla-López, J.R., Chaaraoui, A.A., Flórez-Revuelta, F.: Visual privacy protection methods: A survey. Expert Systems with Applications **42**(9), 4177–4195 (2015)
41. Pantoja, C., Izquierdo, E.: MediaEval 2014 Visual Privacy Task: De-identification and Re-identification of Subjects in CCTV. In: Working Notes Proceedings of the MediaEval Workshop (2014)
42. Paralic, M., Jarina, R.: UNIZA@ Mediaeval 2014 Visual Privacy Task: Object Transparency Approach. In: Working Notes Proceedings of the MediaEval Workshop (2014)

43. Pradnya P., M., D. Ruikar, S.: Image Fusion based on Stationary Wavelet Transform. International Journal of Advanced Engineering Research and Studies **2**(4), 99–101 (2013)
44. Reisslein, M., Rinner, B., Roy-Chowdhury, A.: Smart Camera Networks [guest editors' introduction]. Computer **47**(5), 23–25 (2014)
45. Rinner, B., Wolf, W.: An Introduction to Distributed Smart Cameras. Proceedings of the IEEE **96**(10), 1565–1575 (2008)
46. Saini, M., Atrey, P., Mehrotra, S., Kankanhalli, M.: Anonymous surveillance. In: Proceedings of the International Conference on Multimedia and Expo, pp. 1–6 (2011)
47. Saini, M., Atrey, P.K., Mehrotra, S., Kankanhalli, M.: W3-privacy: understanding what, when, and where inference channels in multi-camera surveillance video. Multimedia Tools and Applications **68**(1), 135–158 (2014)
48. SanMiguel, J.C., Cavallaro, A., Martnez, J.M.: Adaptive Online Performance Evaluation of Video Trackers. IEEE Transactions on Image Processing **21**(5), 2812–2823 (2012)
49. Sarwar, O., Rinner, B., Cavallaro, A.: Design Space Exploration for Adaptive Privacy Protection in Airborne Images. In: Proceedings of the IEEE Conference on Advanced Video and Signal-based Surveillance, pp. 1–7 (2016)
50. Schmiedeke, S., Kelm, P., Goldmann, L., Sikora, T.: TUB@ MediaEval 2014 Visual Privacy Task: Reversible Scrambling on Foreground Masks. In: Working Notes Proceedings of the MediaEval Workshop (2014)
51. Sheskin, D.J.: Handbook of Parametric and Nonparametric Statistical Procedures. CRC Press (2011)
52. Sohn, H., Lee, D., Neve, W.D., Plataniotis, K.N., Ro, Y.M.: An objective and subjective evaluation of content-based privacy protection of face images in video surveillance systems using JPEG XR. Effective Surveillance for Homeland Security: Balancing Technology and Social Issues **3**, 111–140 (2013)
53. Tansuriyavong, S., Hanaki, S.i.: Privacy protection by concealing persons in circumstantial video image. In: Proceedings of the Workshop on Perceptive User Interfaces, pp. 1–4 (2001)
54. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, pp. 586–591 (1991)
55. Viola, P., Jones, M.J.: Robust real-time face detection. International journal of computer vision **57**(2), 137–154 (2004)
56. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004)
57. Winkler, T., Ádám Erdélyi, Rinner, B.: TrustEYE – Trustworthy Sensing and Cooperation in Visual Sensor Networks. `http://trusteye.aau.at` (2012). Last accessed: November 2016
58. Winkler, T., Rinner, B.: Security and Privacy Protection in Visual Sensor Networks: A Survey. ACM Computing Surveys **47**(1), 42 (2014)
59. Zhang, C., Tian, Y., Capezuti, E.: Privacy preserving automatic fall detection for elderly using RGBD cameras. In: Proceedings of the International Conference on Computers Helping People with Special Needs, pp. 625–633 (2012)