Effective evaluation of privacy protection techniques in visible and thermal imagery

Tahir Nawaz[a,*], **Amanda Berg**[b,c], **James Ferryman**[a], **Jörgen Ahlberg**[b,c], **Michael Felsberg**[c]
[a]Computational Vision Group, Department of Computer Science, University of Reading, Reading, RG6 6AF, United Kingdom
[b]Termisk Systemteknik AB, SE-583 35 Linköping, Sweden
[c]Computer Vision Laboratory, Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden

Abstract. Privacy protection may be defined as replacing the original content in an image region with a new (less intrusive) content having modified target appearance information to make it less recognizable by applying a privacy protection technique. Indeed the development of privacy protection techniques needs also to be complemented with an established objective evaluation method to facilitate their assessment and comparison. Generally, existing evaluation methods rely on the use of subjective judgements or assume a specific target type in image data and use target detection and recognition accuracies to assess privacy protection. This paper proposes a new annotation-free evaluation method that is neither subjective nor assumes a specific target type. It assesses two key aspects of privacy protection: protection and utility. *Protection* is quantified as an appearance similarity and *utility* is measured as a structural similarity between original and privacy-protected image regions. We performed an extensive experimentation using six challenging datasets (having 12 video sequences) including a new dataset (having six sequences) that contains visible and thermal imagery. The new dataset is made available online for community. We demonstrate effectiveness of proposed method by evaluating six image-based privacy protection techniques, and also show comparisons of proposed method over existing methods.

Keywords: Privacy protection, evaluation, visible imagery, thermal imagery.

*Tahir Nawaz, t.h.nawaz@reading.ac.uk

1 Introduction

Recently, countries around the world have seen a rapid growth in the use of surveillance applications in public places.[1] According to an estimate, in 2014 there were 245 million operational surveillance cameras in the world[1]. Indeed, this surge in surveillance applications has led to an increasing need for the privacy protection of individuals.[2,3]

[1]An estimate by IHS, a global information company. https://technology.ihs.com/532501/245-million-video-surveillance-cameras-installed-globally-in-2014.

[PRE-PRINT] Journal of Electronic Imaging, Vol. 26, Issue 5, 051408, 2017. DOI: 10.1117/1.JEI.26.5.051408

Privacy protection refers, in this work, to replacing the original content of an image (or a region thereof) with a new content having modified appearance information of target(s) so as to make it(them) less recognizable. The new content may be a result of (i) transforming the original content, or (ii) perturbing the original content, or (iii) using a different image capturing modality. The first case may involve hiding image regions by applying image processing and filtering operations[4–11] to provide different levels of privacy protection to targets under consideration.[3] In the second case, perturbations may be added to the original content[12,13] in the form of displacing the state of the image patch thus obscuring target's recognition by motion. The third case involves employing a different imaging device that is assumed to be privacy protecting *per se*. In this work, we use thermal infrared (TIR) camera that has been considered privacy preserving for years[14,15] due to low resolution, high noise levels, and difficulty for a human to interpret thermal imagery and recognize targets; see[16,17] for details on infrared imaging. It would therefore be interesting and desirable to quantitatively evaluate the achieved privacy protection with the thermal modality that, to the best of our knowledge, has not been attempted before. Other approaches also exist that analyze a scene holistically (instead of considering explicit object models) for the sake of protecting individuals' privacy;[18–20] however they are out of the scope of this paper.

The development of privacy protection techniques need also to be complemented by an effective means to evaluate them in order to facilitate their fair comparison. Indeed the two key aspects to consider for objectively[2] evaluating a privacy protection technique are protection and utility.[6,21] *Protection* refers to a quantification of the extent of appearance information (that would make a target recognizable) modified in an image region by a privacy protection method. Completely hiding out image regions may however not be desirable as there may be a need to preserve structural

[2]Here the term 'objective' means non reliance of an evaluation method on subjective judgements.

information for performing a higher-level behavioral analysis in a surveillance application. *Utility* is therefore also computed as a quantification of the preservation of structural information in the image region by a privacy protection technique. An ideal privacy protection technique may aim to maximize both protection and utility.

1.1 Contributions

In this paper we present a new objective evaluation method for assessing protection and utility aspects of image-based privacy protection techniques. The evaluation method is target independent (i.e. it does not assume presence of specific target type in image data) and annotation free. Protection is assessed in terms of measuring the appearance similarity between original and privacy-protected image regions. Utility is measured by quantifying the structural similarity between the original and privacy-protected image regions. A preliminary version of this work appeared earlier.[22] Unlike the work[22] this paper provides an improved method (in terms of the normalization to make the bounds of the protection score well defined) with more justifications. Moreover, this paper presents an extensive experimentation with a detailed validation and analysis using more privacy protection techniques and a much larger number of challenging sequences (including several new ones) containing visible as well as thermal imagery with varying target types (face, full person body, vehicle). We make available online[3] the new dataset containing thermal and visual sequences to facilitate community in the evaluation of privacy protection techniques in particular and other tasks (e.g. detection, tracking) in general. We demonstrate the effectiveness of the proposed evaluation method by assessing and comparing six privacy protection techniques in the context of video tracking. We also show comparisons of the proposed method over objective and

[3] http://www.cvl.isy.liu.se/research/datasets/TST-Priv/.

subjective evaluation approaches.

This paper is organized as follows. Sec. 2 reviews the related work and highlights the novelty of the proposed evaluation method with respect to the existing methods. Sec. 3 defines the problem that is followed by a description of the proposed method in Sec. 4. A detailed experimental validation and analysis is presented in Sec. 5. Sec. 6 concludes the paper.

2 Related work

Methods exist in the literature that were aimed at evaluating privacy protection techniques based on the use of the judgements of a set of human subjects, *subjective evaluation*,[21,23–27] or using objective measures that do not rely on subjective responses, *objective evaluation*,.[28–30] Next, we provide a review of the existing subjective evaluation (Sec. 2.1) and objective evaluation methods (Sec. 2.2).

2.1 Subjective evaluation

Zhao and Stasko[27] performed a study that evaluated privacy protection filters by showing the filtered video streams to a set of human subjects. Boyle *et al.*[25] presented a methodology that involved applying global (full-frame) privacy protection on a set of video sequences and showing them to subjects, and in turn assessing the privacy protection techniques based on the collected subjects' responses using questionnaires. Saini *et al.*[21] and Korshunov *et al.*[26] also used a similar subjective methodology except that they applied privacy protection locally (only on sensitive image regions) in video sequences. Some more examples of works that also used subjective criteria for evaluating privacy protection include a study by Babaguchi *et al.*[23] and a more recent one by Birnstill and Ren *et al.*[24] The above evaluation methods rely on subjective judgements thus

Table 1 State-of-the-art evaluation methods for privacy protection. (Key. NTDR: Non reliance of a method on target detection and recognition accuracies.)

Reference	Objective/Subjective	Target independence	NTDR
[21,23–27]	Subjective	✓	✓
[28–30,32]	Objective		
Proposed	Objective	✓	✓

resulting in an inclusion of a possible bias in the assessment. Moreover, in a subjective evaluation study the test data under consideration, the choice and number of subjects, and the setting-up of questionnaire(s) may need to be analyzed and backed up by the statistical significance testing.[31]

2.2 Objective evaluation

An evaluation framework was proposed that did not rely on subjective judgement and used the face detection and face recognition accuracies on the privacy-protected data as measures of privacy protection.[28,32] Saini *et al.*[29] computed privacy loss using a model that incorporated the face detection and face recognition measures in addition to the scene contextual knowledge and some implicit identity inference information. In another work[30] privacy loss was modeled using only the scene contextual knowledge and implicit identity inference information. While interesting contributions, the above works are target dependent as they assume presence of a specific target type ('face') in the image data and rely also on the performance of detection and recognition algorithms used.

2.3 Discussion

Table 1 provides a summary of the existing privacy protection evaluation methods. Unlike existing methods[21,23–27] the proposed method in this paper does not rely on subjective judgements. Additionally, unlike the methods[28–30,32] the proposed method is target independent as it does not require application of privacy protection methods on image data with a particular target type. Moreover,

the proposed method does not require target detection and recognition accuracies in the evaluation procedure as in existing methods[28–30,32] and is therefore annotation free. Furthermore, while evaluation criteria exist for assessing methods in other computer vision areas including optical flow estimation,[33] stereo correspondence estimation[34] and video tracking,[35,36] there is an absence of an established method for the evaluation of different aspects of privacy protection methods. An initiative was made in the form of a challenge for assessing privacy protection techniques under the MediaEval workshops that however used an evaluation that mainly relied on subjective judgements.[37]

3 Problem definition

Consider a video sequence V consisting of K frames:

$$V = (f_k)_{k=1}^{K}, \tag{1}$$

where f_k denotes the frame k. Let $\mathscr{X}$ be a set of trajectories (or tracks) estimated by a tracker in V:

$$\mathscr{X} = \{\mathfrak{X}_j\}_{j=1}^{J}, \tag{2}$$

where J is the total number of estimated trajectories. $\mathfrak{X}_j$ is the estimated trajectory for target j:

$$\mathfrak{X}_j = (X_{k,j})_{k=k_{start}^j}^{k_{end}^j}, \tag{3}$$

where k^j_{start} and k^j_{end} are the first and final frame numbers of $\mathfrak{X}_j$, respectively.

$$X_{k,j} = (x_{k,j}, y_{k,j}, A_{k,j}, l_j), \tag{4}$$

where $(x_{k,j}, y_{k,j})$ and $A_{k,j}$ denote the position and the occupied area information of target j on the image plane and l_j defines its ID. Here, $A_{k,j}$ is considered in the form of a bounding box in which case $X_{k,j}$ can be re-written as:

$$X_{k,j} = (x_{k,j}, y_{k,j}, w_{k,j}, h_{k,j}, l_j), \tag{5}$$

where $w_{k,j}$ and $h_{k,j}$ denote width and height of the bounding box for target j at f_k. The number of estimated targets at f_k is denoted as n_k, which are defined as $\{X_{k,1}, \ldots, X_{k,j}, \ldots, X_{k,n_k}\}$. Let $B_{k,j}$ denote the image region within the bounding box $X_{k,j}$ containing the recognizable appearance information for target j. B_k is the set of image regions within all the bounding boxes in f_k:

$$\mathrm{B}_k = \{B_{k,1}, \ldots, B_{k,j}, \ldots, B_{k,n_k}\}. \tag{6}$$

Note that in this work we consider the tracking problem to be already reliably solved.

Let $B'_{k,j}$ denote the privacy-protected image region obtained by applying a privacy protection technique to hide or obscure the target recognizable appearance information in $B_{k,j}$. Therefore, B'_k is the set containing the privacy-protected image regions within all the bounding boxes in f_k:

$$\mathrm{B}'_k = \{B'_{k,1}, \ldots, B'_{k,j}, \ldots, B'_{k,n_k}\}. \tag{7}$$

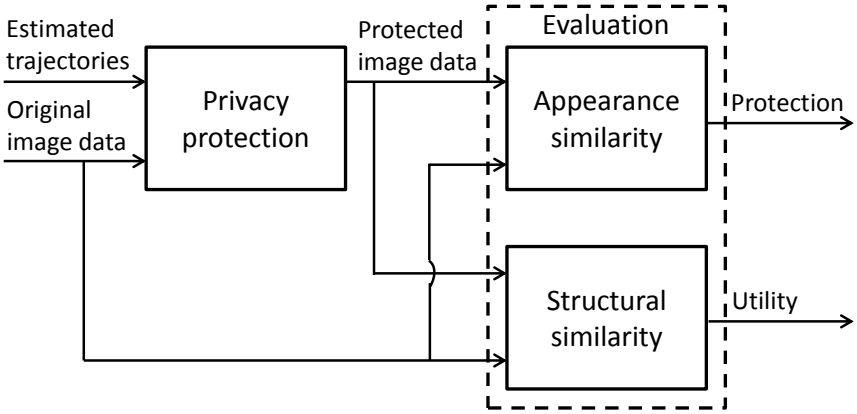


Fig 1 Proposed method for objectively evaluating a privacy protection technique in the context of video tracking by quantifying the *protection* and *utility* aspects.

The evaluation procedure compares B'_k with respect to B_k, the original unprotected region that acts as a reference, to assess the candidate privacy protection technique in the form of a score, S_k, at f_k without the need of any annotated ground-truth information.

4 Evaluation method for privacy protection

Given B_k and B'_k the proposed evaluation method is aimed to assess the two key aspects of privacy protection: protection and utility (Fig. 1). Next we describe the computation of the protection both at frame level (P_k) and sequence level (P) in Sec. 4.1 followed by a description of the quantification of the utility both at frame level (U_k) and sequence level (U) in Sec. 4.2 for a privacy protection technique. Sec. 4.3 highlights the advantages of the proposed method as compared to an existing method.

4.1 Protection

Protection is assessed in terms of the appearance similarity between B_k and B'_k. A smaller appearance similarity between B_k and B'_k alludes to a greater impact of the applied privacy protection. Some well-known similarity measures include Bhattacharyya distance, Kullback-Leibler

divergence, Mahalanobis distance, Chi-squared similarity and Earth Mover's distance. We use the Bhattacharyya distance as it has the following advantages: it is a metric unlike the Kullback-Leibler divergence that is non-symmetric and hence not a metric; unlike the Mahalanobis distance it does not assume the same variance for $B_{k,j}$ and $B'_{k,j}$; it avoids the singularity problems of Chi-squared similarity;[38] compared to the Earth Mover's distance[39] it is computationally more efficient.[40] Indeed, Bhattacharyya distance has been very widely used in the community for decades.[40–42]

At frame k we therefore compute the amount of achieved protection, P_k, in B'_k as follows:

$$\mathrm{P}_k = \frac{1}{n_k} \sum_{j=1}^{n_k} D(q^{B_{k,j}}, q^{B'_{k,j}}), \tag{8}$$

where $D(q^{B_{k,j}}, q^{B'_{k,j}})$ is the Bhattacharyya distance that provides a similarity at a frame k between the probability density functions (PDFs) of $B_{k,j}$, $q^{B_{k,j}}$, and $B'_{k,j}$, $q^{B'_{k,j}}$:

$$D(q^{B_{k,j}}, q^{B'_{k,j}}) \;=\; \sqrt{1 - BC(q^{B_{k,j}}, q^{B'_{k,j}})}, \tag{9}$$

which is also termed as Hellinger distance. A PDF is computed as a normalized histogram that is calculated by counting the occurences of each intensity level (bin) divided by the sum of the counts of all bins. In Eq. 9, $BC(q^{B_{k,j}}, q^{B'_{k,j}})$ is the Bhattacharyya coefficient and is given as follows:[43]

$$BC(q^{B_{k,j}}, q^{B'_{k,j}}) \;=\; \sum_{z=0}^{Z} \sqrt{q^{B_{k,j}}(z) q^{B'_{k,j}}(z)}. \tag{10}$$

$Z = 255$ as we use 256 bins (which is equal to the number of intensity levels) in computing the

normalized histograms for $B_{k,j}$ and $B'_{k,j}$. In the case of a RGB image

$$D(\cdot) \;=\; \frac{\sqrt{D^2_{red}(\cdot)+D^2_{green}(\cdot)+D^2_{blue}(\cdot)}}{\sqrt{3}}, \tag{11}$$

where $D_{red}(\cdot)$, $D_{green}(\cdot)$ and $D_{blue}(\cdot)$ are respectively the Bhattacharyya distances between the corresponding PDFs of the red, green and blue channels of $B_{k,j}$ and $B'_{k,j}$. Likewise, $BC_{red}(q^{B_{k,j}}, q^{B'_{k,j}})$, $BC_{green}(q^{B_{k,j}}, q^{B'_{k,j}})$ and $BC_{blue}(q^{B_{k,j}}, q^{B'_{k,j}})$ are respectively the Bhattacharyya coefficients for the corresponding PDFs of the red, green and blue channels of $B_{k,j}$ and $B'_{k,j}$. The denominator in (11), $\sqrt{3}$, is a normalization factor in order to numerically bound $D(\cdot) \in [0,1]$ as defined below.

Lower bound ($D(\cdot) = 0$): When $B_{k,j}$ and $B'_{k,j}$ are completely similar (i.e. $B_{k,j} = B'_{k,j}$), from (10) $BC_{red}(\cdot) = BC_{green}(\cdot) = BC_{blue}(\cdot) = 1$. Thus, using (9), $D_{red}(\cdot) = D_{green}(\cdot) = D_{blue}(\cdot) = 0$, which implies $D(\cdot) = 0$ using (11).

Upper bound ($D(\cdot) = 1$): When $B_{k,j}$ and $B'_{k,j}$ are completely dissimilar, their corresponding PDFs $(q^{B_{k,j}}, q^{B'_{k,j}})$ do not overlap, in which case $BC_{red}(\cdot) = BC_{green}(\cdot) = BC_{blue}(\cdot) = 0$ using (10). Thus, using (9), $D_{red}(\cdot) = D_{green}(\cdot) = D_{blue}(\cdot) = 1$, which implies $D(\cdot) = 1$ using (11); hence the need for normalization by $\sqrt{3}$ in (11).

Note that for the case of grey-scale data $D(\cdot)$ can simply be computed using (9) that is anyway numerically bounded i.e. $D(\cdot) \in [0,1]$. Therefore, $\mathrm{P}_k \in [0,1]$: the higher P_k the greater the amount of achieved protection.

While the computation of P_k enables analyzing the achieved protection at each frame, to facilitate the performance comparison between different privacy protection methods we provide the following two statistics to compute the overall achieved protection, P, across a sequence in the

form of a single score:

$$\mathrm{P} = \frac{1}{K}\sum_{k=1}^{K} \mathrm{P}_k; \tag{12}$$

or,

$$\mathrm{P} = \min\{\mathrm{P}_k\}_{k=1}^{K}. \tag{13}$$

(12) provides P by computing *mean* of P_k across the frames of a sequence, which might be biased towards outliers, if any. (13) presents an alternative solution that provides P as the *minimum* protection across all the frames of a sequence. We would experimentally analyze the effect of these two statistics on the outcome of evaluation in Sec. 4.3.

4.2 Utility

Utility is quantified in terms of the structural similarity between B_k and B'_k. A smaller structural similarity refers to a lower preservation of structural information. For computing the structural similarity we use the well-known Structural Similarity Index (SSIM),[44] which was employed in earlier related works.[6,45–47] SSIM has advantages of being a general-purpose measure and offers an assessment that was shown to be perceptually closer to the human visual system.[44] It provides a similarity assessment between $B_{k,j}$ and $B'_{k,j}$ by encapsulating their structural comparison (local intensity patterns of pixels) together also with luminance and contrast comparisons. Unlike the existing works,[6,45–47] the proposed SSIM-based formulation quantifies utility for local (region-based) privacy protection across a sequence that is directly applicable to (single-target as well as multi-target) tracking applications.

At frame k the utility, U_k, is therefore computed as follows:

$$\mathrm{U}_k = \frac{1}{n_k} \sum_{j=1}^{n_k} \mathrm{MSSIM}(B_{k,j}, B'_{k,j}), \tag{14}$$

where $\mathrm{MSSIM}(B_{k,j}, B'_{k,j})$ is the mean SSIM value between $B_{k,j}$ and $B'_{k,j}$ for a variation of local windows.[44]

$$\mathrm{MSSIM}(B_{k,j}, B'_{k,j}) \;=\; \frac{1}{M} \sum_{m=1}^{M} \mathrm{SSIM}^m(B_{k,j}, B'_{k,j}), \tag{15}$$

where $\mathrm{SSIM}^m(B_{k,j}, B'_{k,j})$ is the SSIM value for mth window and is given as follows:[44]

$$\mathrm{SSIM}^m(B_{k,j}, B'_{k,j}) \;=\; \frac{(2\,{}^m\mu_{B_{k,j}}\,{}^m\mu_{B'_{k,j}} + C_1)(2\,{}^m\sigma_{B_{k,j}B'_{k,j}} + C_2)}{({}^m\mu^2_{B_{k,j}} + {}^m\mu^2_{B'_{k,j}} + C_1)({}^m\sigma^2_{B_{k,j}} + {}^m\sigma^2_{B'_{k,j}} + C_1)}. \tag{16}$$

SSIM is computed on grey-scale data[44] such that ${}^m\mu_{B_{k,j}}$ and ${}^m\mu_{B'_{k,j}}$ are the mean intensity values, and ${}^m\sigma_{B_{k,j}}$ and ${}^m\sigma_{B'_{k,j}}$ are the standard deviations in $B_{k,j}$ and $B'_{k,j}$, respectively, for the local window m; ${}^m\sigma_{B_{k,j}B'_{k,j}}$ is the correlation coefficient; and C_1 and C_2 are constants. $\mathrm{U}_k \in [0,1]$: the higher U_k the larger the utility retained. As done for the case of P (Sec. 4.1) we provide the following two statistics to compute the overall retained utility, U, in the form of a single score:

$$\mathrm{U} = \frac{1}{K} \sum_{k=1}^{K} \mathrm{U}_k; \tag{17}$$

or,

$$\mathrm{U} = \min\{\mathrm{U}_k\}_{k=1}^{K}. \tag{18}$$

We would experimentally analyze the effect of these two statistics on the outcome of evaluation in Sec. 4.3.

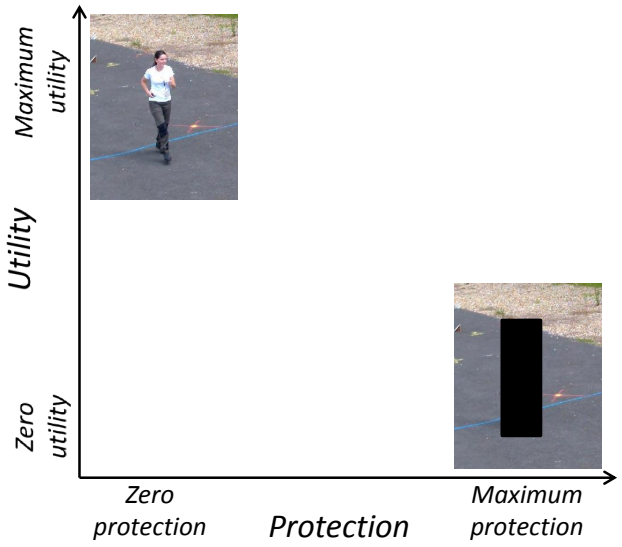


Fig 2 Trade off between protection and utility. Completely masking out the object information in an image would provide a maximum protection but at the expense of a zero utility. Likewise, leaving the object information unaffected would provide a maximum utility but at the expense of a total loss of protection.

Indeed, it boils down to the determining an appropriate trade off between protection and utility to choose a privacy protection technique for a specific application. A total masking out of object information in an image could result in the achievement of a large protection score but at the expense of a total loss of utility. Similarly, leaving the image region unaffected could result in obtaining a maximum utility but at the expense of a zero achieved protection (Fig. 2). Ideally, a privacy protection technique would be expected to maximize both protection and utility. Fig. 3 provides numercial examples of the computed protection and utility scores for different privacy protection techniques on sample tracked object patches, which is further discussed in Sec. 5.

4.3 Advantages and comparisons

We show advantages and comparisons of the proposed evaluation method with an existing objective evaluation method (which is widely used including the MediaEval workshops[6,28,32,48]) that employs face recognition accuracy as a measure of protection and face detection accuracy as a

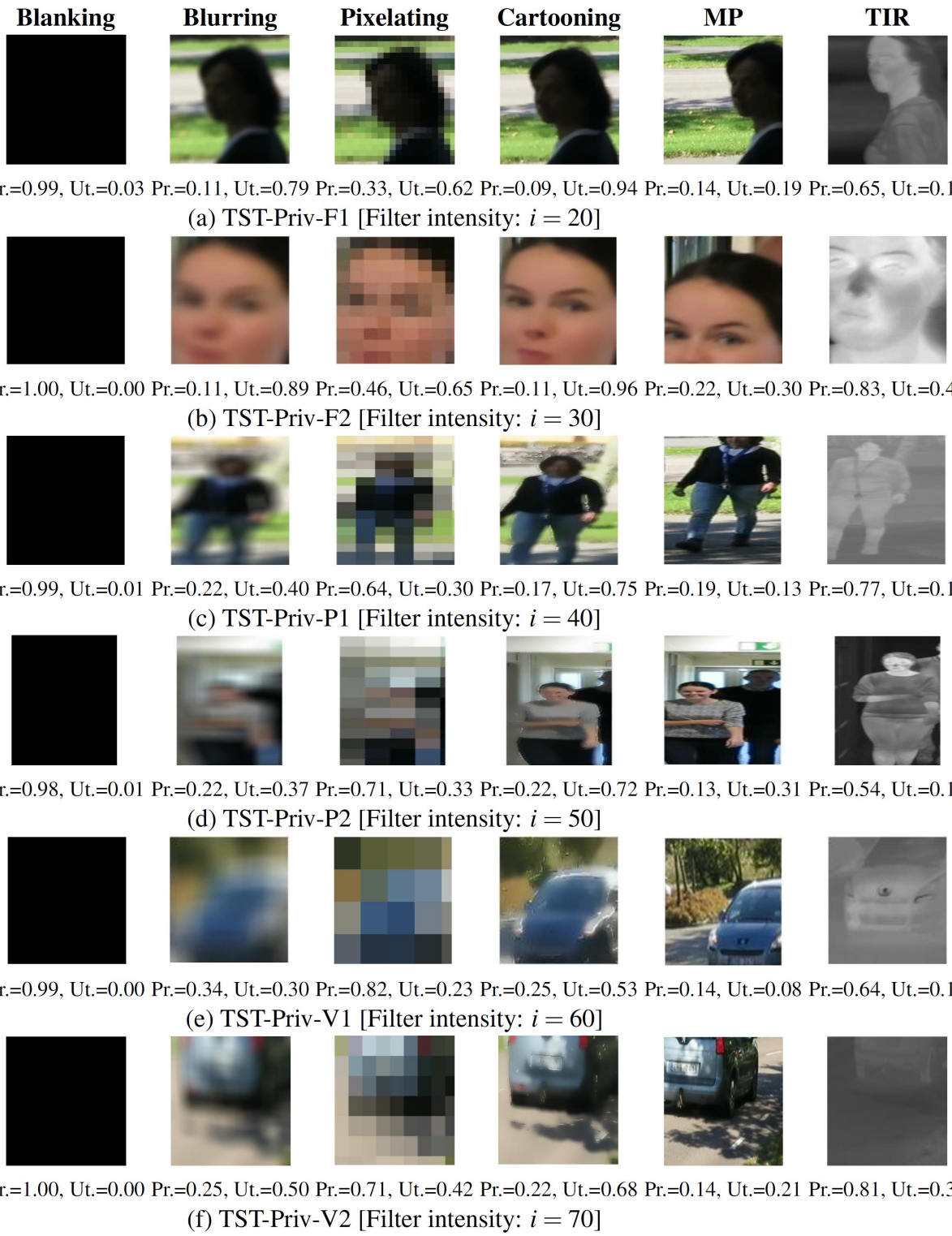


Fig 3 Sample qualitative results for different privacy protection techniques applied on tracked patches with an increasing filter intensity ($i = 20, 30, \ldots, 70$). The corresponding protection (Pr.) and utility (Ut.) scores are also listed under each patch. Key. MP: motion perturbation; TIR: thermal infrared.

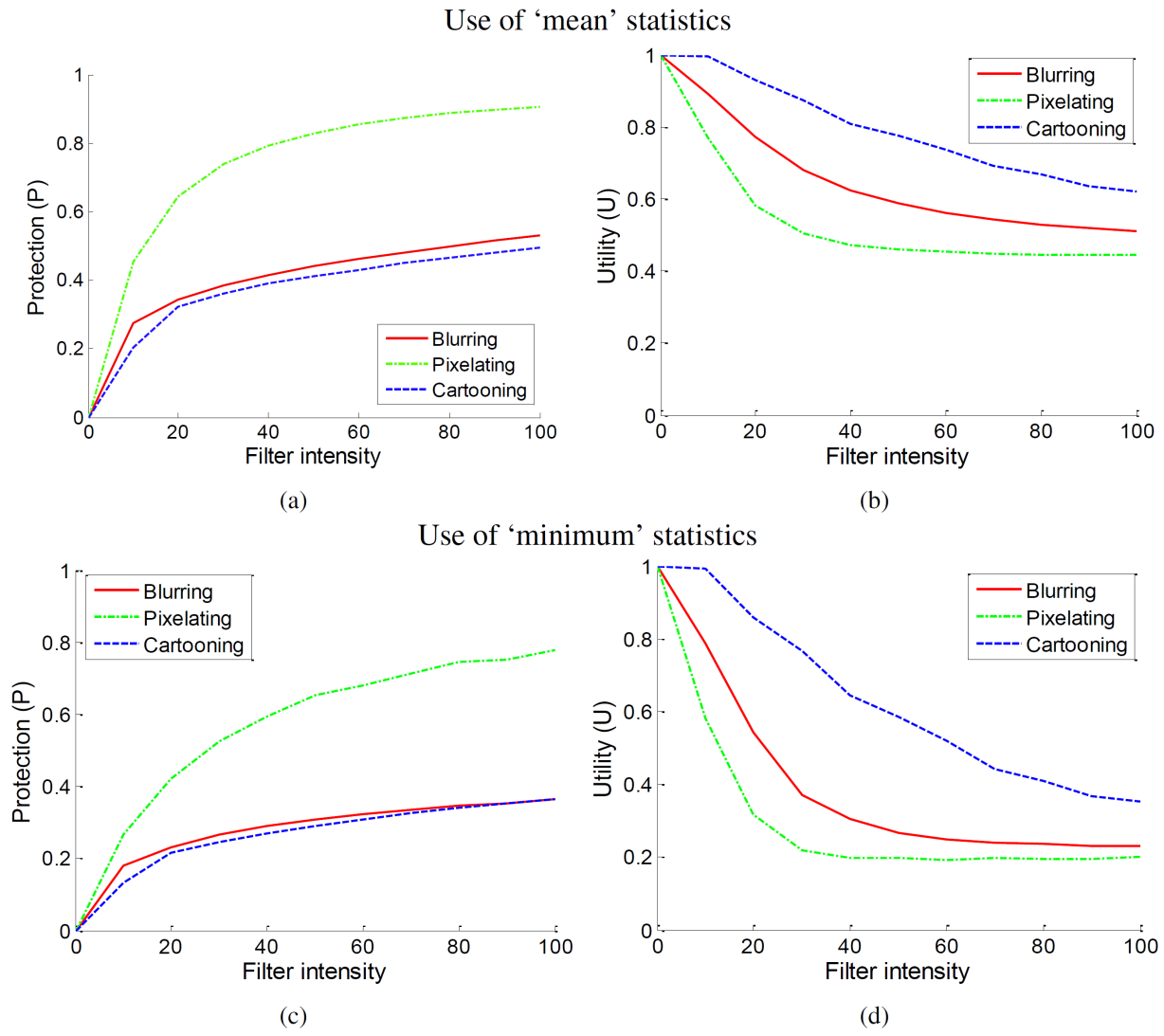


Fig 4 Average protection score (a,c) and average utility score (b,d) obtained by blurring, pixelating and cartooning on all sequences of PEViD dataset for a variation of filter intensity. (a,b) P and U are computed using 'mean' statistics (i.e. Eq. 12, 17); (c,d) P and U are computed using 'minimum' statistics (i.e. Eq. 13, 18).

measure of utility, as well as with respect to a web-based subjective assessment approach. In this regard we use the well-known PEViD dataset[49] as it provides face annotations that are needed for computing face detection and recognition accuracies. PEViD dataset has a total of 21 video sequences each containing 400 frames.

We first plot in Fig. 4 the average protection (P) and average utility (U) scores on all the PEViD sequences using the proposed method separately based on 'mean' statistics (Eq. 12, 17) as well as

'minimum' statistics (Eq. 13, 18) for privacy protection techniques (blurring, pixelating, cartooning) over a variation of filter intensity; see Sec. 5.2 for details on privacy protection techniques. It is interesting to note that the evaluation outcomes produced by mean statistics (Fig. 4(a,b)) as well as minimum statistics (Fig. 4(c,d)) are the same in terms of the trends of rankings of privacy protection techniques both with protection (P) and utility (U) scores. Additionally, unlike the minimum statistics that use only the minimum protection/utility value to provide an estimate, the mean statistics could provide a more representative estimate by incorporating contributions from the protection/utility values of all the frames in a sequence. Moreover, the potential bias towards outliers in mean statistics is expected to be minimized because filter strength is kept constant across all the frames of a sequence. Furthermore, mean statistics is often employed in literature for computing the overall protection/utility across a sequence.[6,28,45–47] We therefore employ mean statistics (Eq. 12, 17) for quantifying the overall protection (P) and utility (U) scores in the rest of this paper.

We now compare and show advantages of the proposed P and U scores (Fig. 4(a,b)) with the reported average face recognition and average detection accuracies on blurring, pixelating and cartooning for a variation of filter intensity in the existing work[6] on the PEViD dataset. The authors in the work[6] used PCA-based, LDA-based and LBP-based face recognition methods, and a Viola-Jones face detection method. The following observations and points could therefore be highlighted here. *First*, the average P scores obtained using the proposed method show that pixelating is consistently the best followed by blurring and cartooning (Fig. 4(a)), which is interestingly similar to the trends of the results of average face recognition accuracies reported in the work[6] (the higher the face recognition accuracy, the lower the protection). Likewise, the average U scores obtained using the proposed method show cartooning to be the best followed by blurring and pixelating (Fig. 4(b)), which are also in line with the trends of the results of average face detection accuracy

reported in the work[6] (the higher the face detection accuracy, the higher the utility). Therefore, the results show a strong correlation between the rankings obtained using the proposed P/U scores and the rankings obtained using face recognition/face detection accuracies with the computed "Spearman's rank correlation coefficient = 1". *Second*, the proposed method has an advantage that it does not require any annotation for the computation of protection and utility scores as compared to the evaluation based on face detection and recognition accuracies that rely on annotation. *Third*, another advantage of the proposed method is that it is target independent and could be applied irrespective of the target type present in the image. Indeed, in Sec. 5.3 and Sec. 5.4 that describe the detailed experimental results and analysis, we show the results of evaluation of privacy protection techniques using the proposed method with varying target types (face, full body, vehicle). The evaluation based on face detection and recognition accuracies has a limitation that it could obviously be used with the face targets only. For other target types, one would need to use the dedicated detection and recognition methods. *Last*, unlike the evaluation based on face detection and recognition accuracies that depends on the performance of the detection and recognition methods used, the proposed evaluation method does not have any such constraints.

For an enhanced validation of the proposed method, we also adopted a subjective approach that involves an assessment of privacy protection techniques based on the judgements of humans. In the subjective assessment, we use all of the 21 sequences of the PEViD dataset with the above mentioned privacy-protection techniques (blurring, pixelating, cartooning). The assessment has been conducted using a website that, before the start of assessment, provides a uniform set of written instructions to a participating human subject. We use a representative sample of 11 human subjects including people that are *skilled*, *semi-skilled* and *unskilled* in privacy protection in videos. The written instructions then follows showing 21 video clips, one by one, to subjects. Each clip

contains four sub videos (all played simultaneously) that are embedded in the top-left, top-right, bottom-left and bottom-right quadrants of the video screen. The top-left video contains the original video data of a sequence, whereas the other three videos contain the corresponding processed results generated by applying three privacy-protection techniques (blurring, pixelating, cartooning) with uniform filter strengths (see Fig. 5). As per ITU recommendation we choose a gray color (red=green=blue=130) for the background that gives a relaxing effect to human eyes.[50] For each clip, the subjects are asked to perform assessment by rating the three processed videos based on the level of achieved privacy protection by assigning a score between 0 to 5 into a questionnaire form: '0' corresponds to no (zero) privacy protection achieved; and '5' corresponds to the maximum privacy protection achieved. The subjects can watch each clip multiple times, if needed, to reach decision. The results show that, across all of the clips, pixelating is consistently ranked the best in terms of privacy protection followed by blurring and cartooning (Fig. 6), which is of course similar to the rankings obtained using the proposed method for these techniques (Fig. 4(a)), thus showing a strong correlation between the results two approaches: "Spearman's rank correlation coefficient= 1".

5 Experimental results and analysis

This section provides in detail the experimental results and analysis. We first describe the datasets (Sec. 5.1) and the privacy protection techniques (Sec. 5.2) used in the experimentation. The results are described in Sec. 5.3 and Sec. 5.4, which are followed by a discussion in Sec. 5.5.

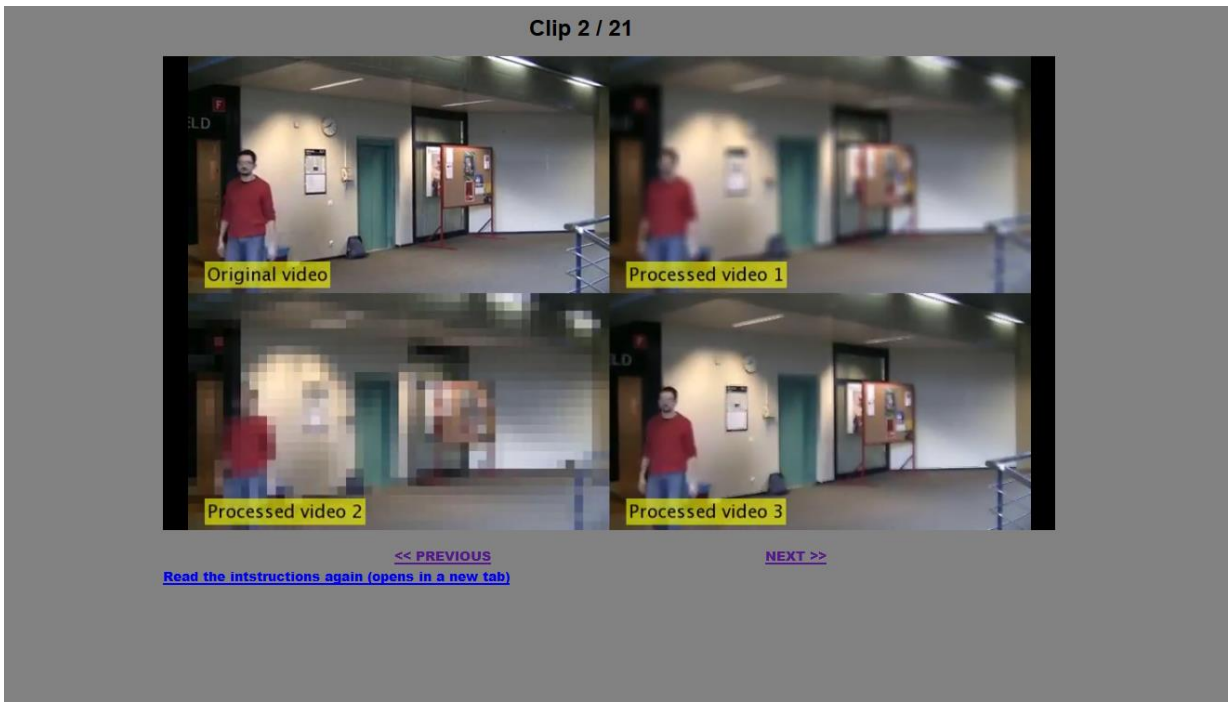


Fig 5 A snap shot of the subjective assessment website showing the original and the three processed (privacy-protected) videos; all played simultaneously in a synchronised manner.

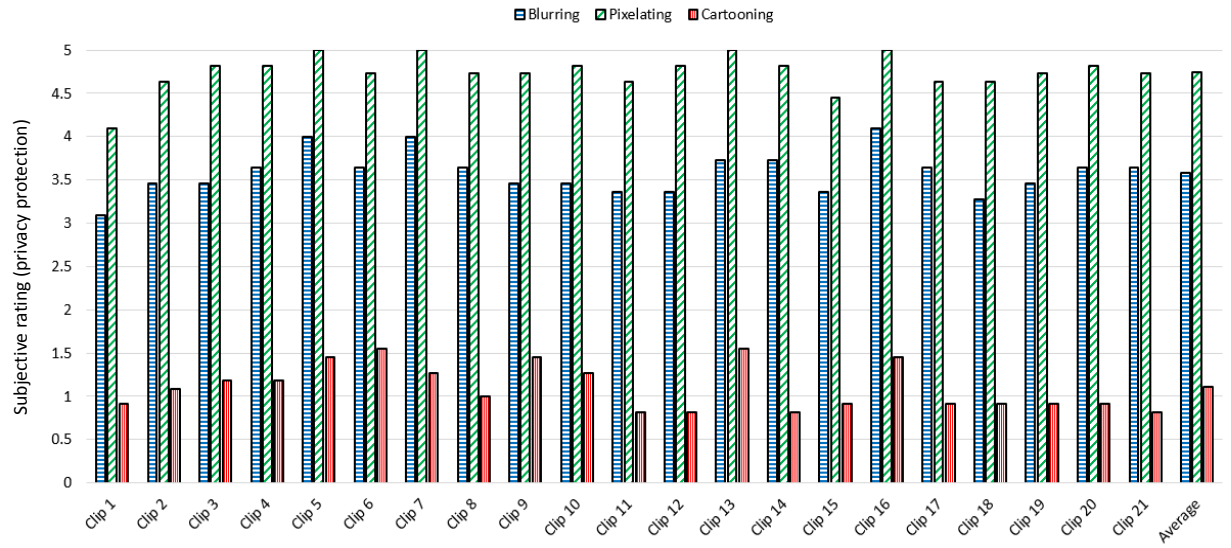


Fig 6 The results of the subjective assessment of privacy-protection techniques (blurring, pixelating, cartooning) in terms of the mean of the ratings of all subjects on each clip.

Table 2 Summary of the datasets. Key. K: number of frames; VIS: visual sequence; TH: thermal-infrared sequence; Occ: occlusion; SC: scale changes; IC: illumination changes; Cr: crowdedness; PC: pose changes. Note that '(-)' means that the corresponding thermal-infrared sequence is not available.

Sequence	K **VIS (TH)**	**Frame size VIS (TH)**	**Target type**	**Challenges**
TST-Priv-F1	420 (341)	1080×1920 (480×640)	Face	Occ., SC, PC
TST-Priv-F2	497 (386)	1080×1920 (480×640)	Face	Occ., SC, PC
TST-Priv-P1	319 (261)	1080×1920 (480×640)	Full body	Occ., SC, PC
TST-Priv-P2	539 (450)	1080×1920 (480×640)	Full body	Occ., SC, PC
TST-Priv-V1	820 (685)	1080×1920 (480×640)	Vehicle	Occ., SC, PC
TST-Priv-V2	1090 (885)	1080×1920 (480×640)	Vehicle	Occ., SC
ETH Bahnhof	999 (-)	480×640 (-)	Full body	Occ, SC, IC, Cr
ETH Sunnyday	354 (-)	480×640 (-)	Full body	Occ, SC, IC, Cr
iLids Easy	5220 (-)	576×720 (-)	Full body	Occ, SC, IC
PETS 2000	160 (-)	576×768 (-)	Vehicle	SC, PC
PETS 2015	243 (-)	960×1280 (-)	Vehicle	SC, PC
P5-UK	150 (-)	960×1280 (-)	Vehicle	SC, PC

5.1 Datasets

We use 12 sequences belonging to six challenging datasets in experiments (Table 2). Among the datasets, we introduce a new dataset, called TST-Priv, whereas the remaining are existing datasets.

TST-Priv contains six sequences (TST-Priv-F1, TST-Priv-F2, TST-Priv-P1, TST-Priv-P2, TST-Priv-V1, TST-Priv-V2). The dataset provides synchronized visual as well as thermal sequences recorded simultaneously from a visual camera (Canon PowerShot G16) and a thermal camera (FLIR SC655). The choice of cameras is made so as to provide (what is currently regarded as) a high resolution in their respective domains (1080×1920 and 480×640, respectively). In the setup, both cameras are mounted very close to each other with a very small disparity in their corresponding images. Additionally, the relative position and orientation of both cameras is the same for all recordings. TST-Priv is made publicly available online for the community.

Among the *existing datasets*, four are well known and publicly available including ETH,[51] iLids Easy,[52] PETS 2015[53] and PETS 2000.[54] The fifth one (P5-UK) is recorded in a UK site under the EU project P5.[55] We use two sequences from ETH dataset (ETH Bahnhof, ETH Sunnyday)

and a sequence each from iLids Easy, PETS 2015, P5-UK and PETS 2000 datasets.

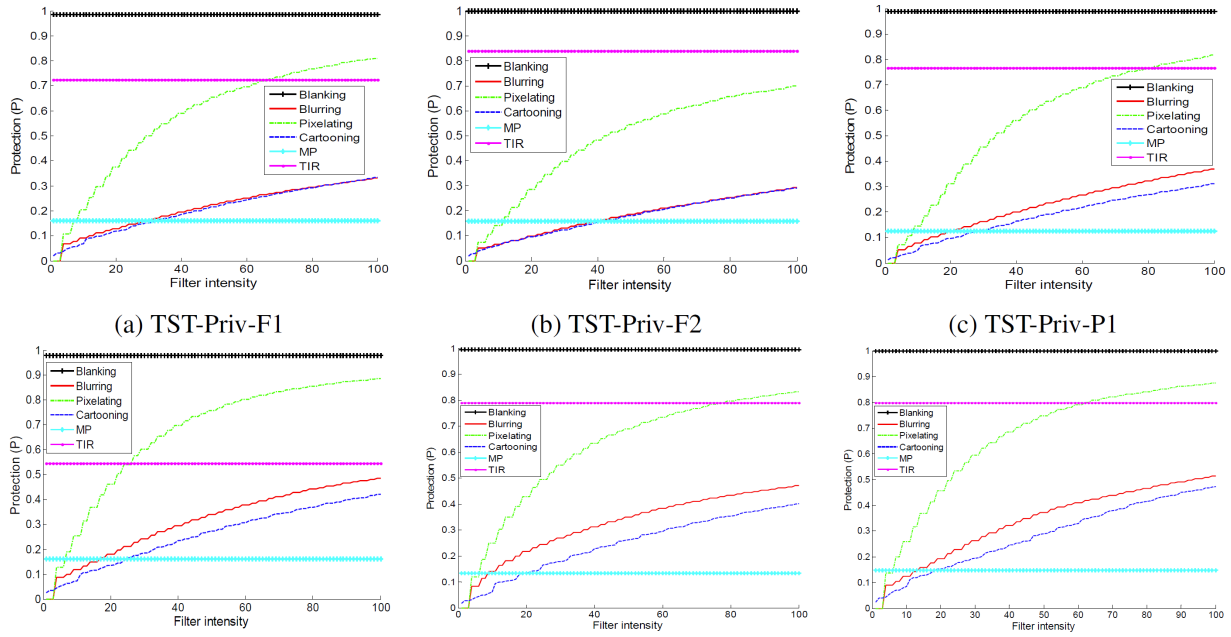
The datasets contain *full-person bodies*, *faces*, and *vehicles* as target types. TST-Priv sequences contain both a single target (TST-Priv-V1) as well as multiple targets (TST-Priv-F1, TST-Priv-F2, TST-Priv-P1, TST-Priv-P2, TST-Priv-V2). Tracking results on these sequences are generated using a single target tracker[56] on visual imagery. In the case of multiple targets, the tracker is run separately for each target to obtain its track. To obtain the corresponding tracks on thermal imagery for each scenario, a homography between the corresponding visual and thermal views is estimated. We employ a well-known method that uses (six) point matches in a plane to compute homography.[57] The computed homography is used to map the corner points of the bounding boxes from visual to thermal domain. Note that the estimated tracks on thermal imagery could contain some discrepancies (data offsets) due to possible inaccuracies in the homography computation (see sample TIR patches in Fig. 3). ETH Bahnhof, ETH Sunnyday and iLids Easy contain multiple targets and we use tracking results from a multi-target tracker[58] in these scenes. PETS 2015, P5-UK and PETS 2000 contain a single target and we use tracking results from a single-target tracker[59] in these scenes. Note that this work does not focus on the choice of tracker *per se* and is aimed at evaluating different privacy protection techniques on the tracked bounding boxes (of targets) that, in principle, could be generated from any tracker (that uses the same target model).

5.2 Privacy protection techniques

We demonstrate the effectiveness of the proposed evaluation method by evaluating and comparing six image-based privacy protection techniques including cartooning, blurring, pixelating, blanking, motion perturbation (MP) and thermal infrared imaging (TIR). Cartooning involves applying an initial blurring on the input image region followed by mean-shift filtering and edge recovery

using the already generated gradient mask with sobel edge detector.[6] The kernel size at the initial blurring stage (A) and the spatial radius (sp) and color radius (sr) at the mean-shift filtering stage are given as follows:[6] $A_i = [i \cdot A_{orig}/50]$; $sp_i = [i \cdot sp_{orig}/50]$; $sr_i = [i \cdot sr_{orig}/50]$; where i is the filter intensity: $i \in [1, 100]$ and the parameters $A_{orig} = 7$, $sp_{orig} = 20$ and $sr_{orig} = 40$.[6] Additionally, as done in the work,[6] for establishing some correspondence and a fair comparison the kernel size used in the case of blurring and pixelating for a particular filter intensity, i, is equal to sp_i as defined above for cartooning. We therefore apply cartooning, blurring and pixelating on the tracking results in all datasets for a full variation of filter intensity, i. Blanking completely masks out the privacy-sensitive information in an image region. MP involves perturbing the motion state of a patch ($B_{k,j}$) thus obscuring the identification of an object by motion. For a given $B_{k,j}$, a perturbation is added to its position $(x_{k,j}, y_{k,j})$ by displacing x and y coordinates. The amount of perturbation is randomly chosen while ensuring that there still remains a reasonable amount of overlap (set to be atleast 50%[60]) between $B_{k,j}$ and the perturbed patch ($B'_{k,j}$). TIR, of course, achieves privacy protection through the use of thermal imagery and is included in the comparison only on TST-Priv dataset because thermal imagery is available only for this dataset. Blanking, MP and TIR obviously remain unaffected over a variation of i.

Therefore, on TST-Priv we provide the results of the comparison of cartooning, blurring, pixelating, blanking, MP and TIR using the proposed evaluation method (Sec. 5.3). Note that to compute privacy score for TIR bounding boxes (that are used in 8-bit form), the corresponding RGB bounding boxes are converted into grayscale (8-bit) format. On the remaining datasets, we provide the results of the comparison of cartooning, blurring, pixelating, blanking and MP using the proposed method (Sec. 5.4).



(a) TST-Priv-F1 (b) TST-Priv-F2 (c) TST-Priv-P1

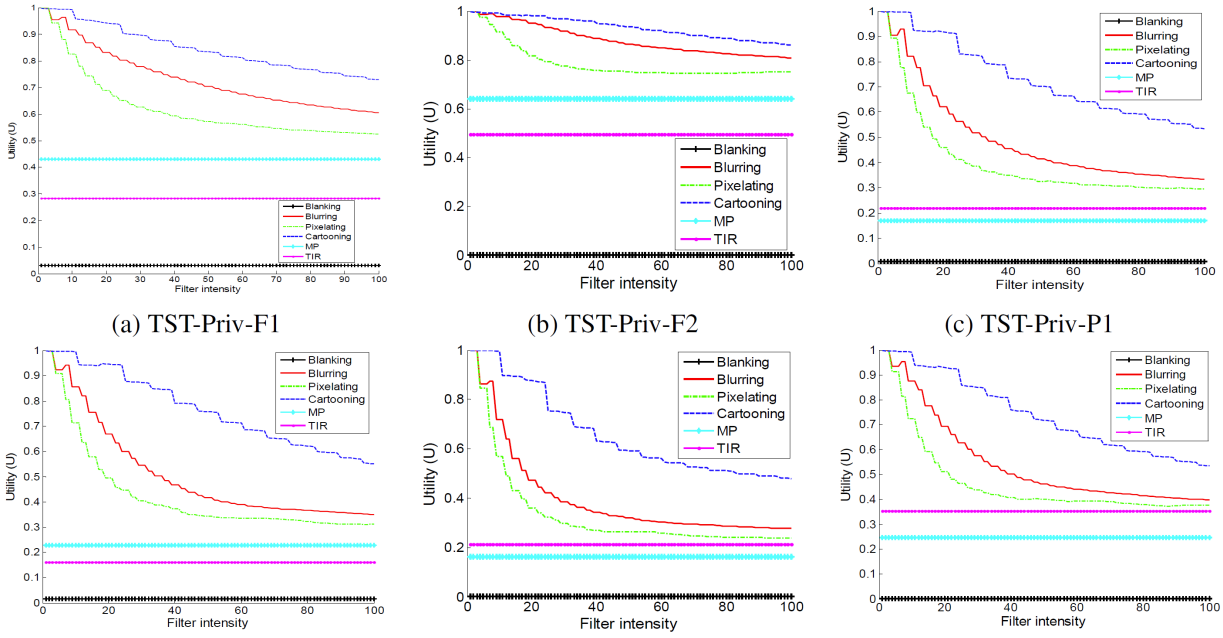
(d) TST-Priv-P2 (e) TST-Priv-V1 (f) TST-Priv-V2

Fig 7 Protection score (P) obtained by different privacy protection techniques for a variation of filter intensity on (a) TST-Priv-F1, (b) TST-Priv-F2, (c) TST-Priv-P1, (d) TST-Priv-P2, (e) TST-Priv-V1 and (f) TST-Priv-V2.

5.3 Results on TST-Priv dataset

Fig. 7 and Fig. 8 plot the protection (P) and utility (U) scores, respectively, of the six privacy protection techniques (blanking, blurring, pixelating, cartooning, MP, TIR) for a variation of i on all TST-Priv sequences. Fig. 3 shows sample qualitative results for these techniques with an increasing i.

Expectedly, on all sequences blanking provides the highest P (Fig. 7(a-f)) as it masks out the entire information in the image data (Fig. 3). Among TIR, cartooning, blurring and MP, TIR consistently obtains the highest P on all sequences for the entire variation of i as shown in Fig. 7(a-f). Between TIR and pixelating, the former obtains a higher P on TST-Priv-F2 (Fig. 7(b)). On the remaining sequences TIR does not achieve a higher P than pixelating for an entire variation of i: indeed the former outperforms the latter on TST-Priv-F1 for $i \leq 66$ (Fig. 7(a)), on TST-Priv-P1 for $i \leq 81$ (Fig. 7(c)), on TST-Priv-P2 for $i \leq 26$ (Fig. 7(d)), on TST-Priv-V1 for $i \leq 78$ (Fig. 7(e)),



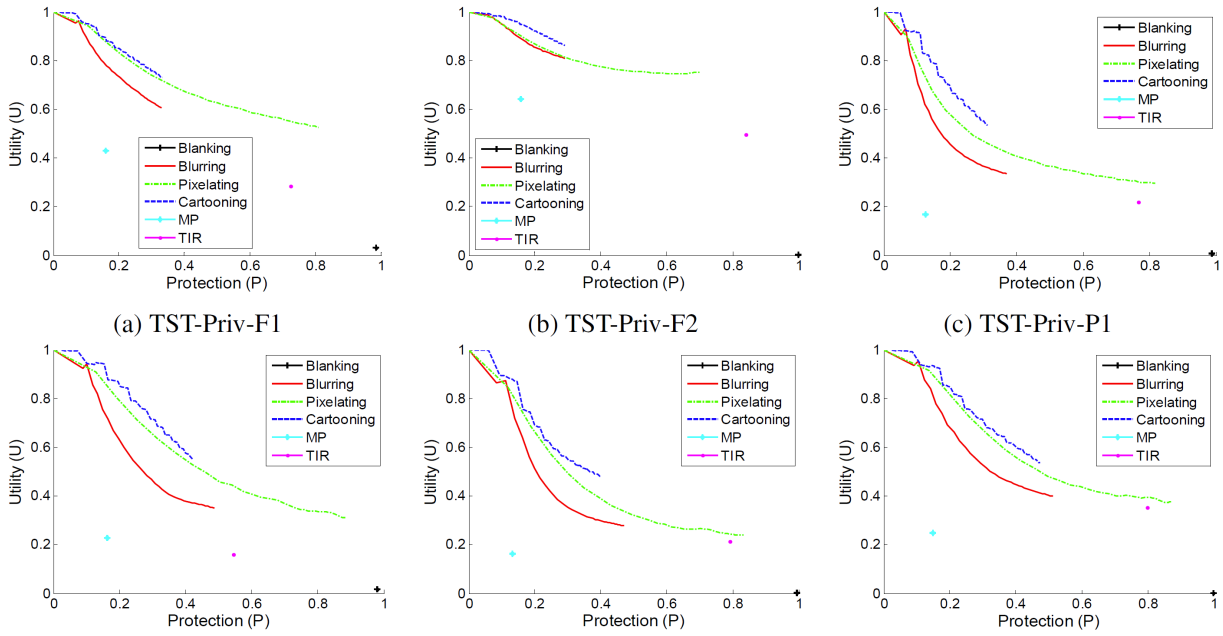
(a) TST-Priv-F1 (b) TST-Priv-F2 (c) TST-Priv-P1

(d) TST-Priv-P2 (e) TST-Priv-V1 (f) TST-Priv-V2

Fig 8 Utility score (U) obtained by different privacy protection techniques for a variation of filter intensity on (a) TST-Priv-F1, (b) TST-Priv-F2, (c) TST-Priv-P1, (d) TST-Priv-P2, (e) TST-Priv-V1 and (f) TST-Priv-V2.

and on TST-Priv-V2 for $i \leq 61$ (Fig. 7(f)). Among pixelating, blurring and cartooning, pixelating is consistently the best in terms of P followed by blurring and cartooning over a variation of i. Note that for $i \in [1,3]$, $\mathrm{P} = 0$ for blurring and pixelating because according to the equation of sp_i their kernel size is 1×1 thus leaving the image regions unaltered by these two techniques. Finally, MP is generally found to achieve the lowest P for most of the variation of i (Fig. 7(a-f)), which is in line with the conclusions of an earlier study[61] that data perturbation may not always be an effective means for protecting privacy.

The utility scores (U) obtained by six techniques over a variation of i across all TST-Priv sequences are plotted in Fig. 8(a-f). Cartooning preserves the structural information better than all methods (Fig. 3). Indeed, on all sequences cartooning shows the best U for the entire variation of i with blurring and pixelating obtaining the second-best and third-best U. The fourth-best U is either by MP (on TST-Priv-F1, TST-Priv-F2, TST-Priv-P2) or TIR (on TST-Priv-P1, TST-Priv-V1,



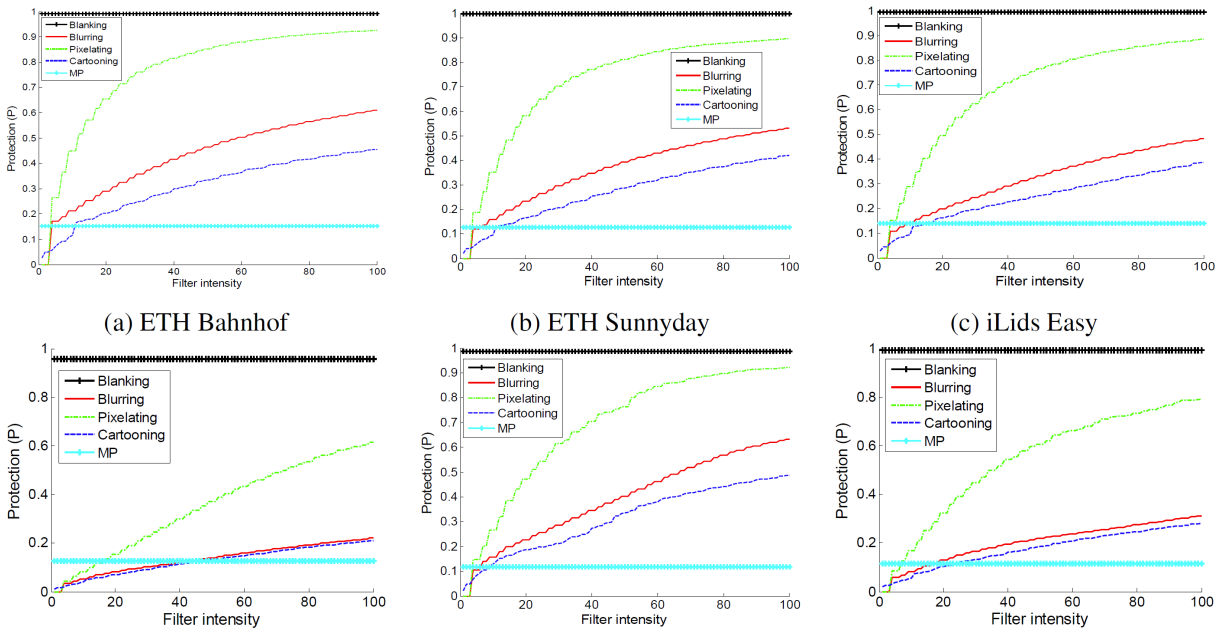
(a) TST-Priv-F1 (b) TST-Priv-F2 (c) TST-Priv-P1

(d) TST-Priv-P2 (e) TST-Priv-V1 (f) TST-Priv-V2

Fig 9 Utility score (U) plotted *vs.* protection score (P) obtained by different privacy protection techniques for a variation of filter intensity on (a) TST-Priv-F1, (b) TST-Priv-F2, (c) TST-Priv-P1, (d) TST-Priv-P2, (e) TST-Priv-V1 and (f) TST-Priv-V2.

TST-Priv-V2). Blanking, which consistently achieves the highest protection (P), provides the least utility (smallest U) on all sequences due to a total loss of structural information.

Indeed, the aim for a privacy protection technique would be to provide an appropriate trade off between U and P. To this end in Fig. 9 we also plot U (as computed in Fig. 8) *vs.* P (as computed in Fig. 7) on all sequences. U *vs.* P plot would be desirable for choosing among different privacy protection techniques for a specific scenario. For example, on TST-Priv-F1, TST-Priv-F2, TST-Priv-P1, TST-Priv-P2, TST-Priv-V1 and TST-Priv-V2 (Fig. 9(a-f)), for a desired $P = 0.25$ cartooning would be the best choice as it provides the highest U. Likewise, for a desired $P = 0.65$ pixelating would be the best choice due to the highest U. Generally, cartooning appears more desirable than the remaining techniques in terms of providing an appropriate U-P trade off with all target types (*face* (Fig. 9(a-b)), *full body* (Fig. 9(c-d)), *vehicle* (Fig. 9(e-f))).



(a) ETH Bahnhof (b) ETH Sunnyday (c) iLids Easy

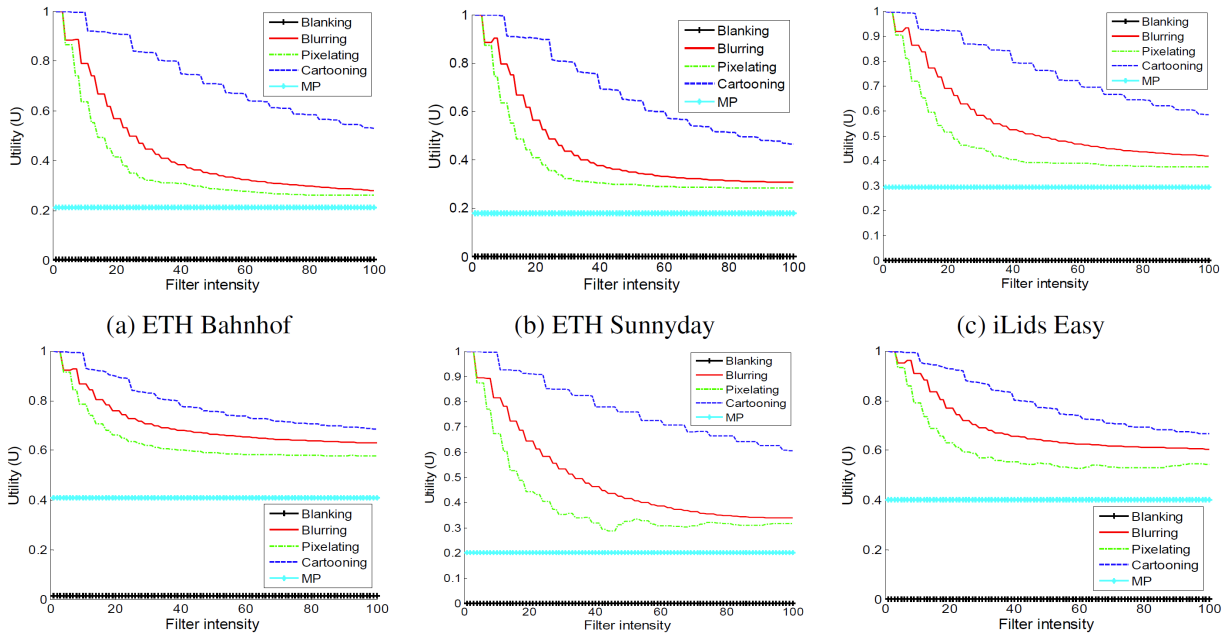
(d) PETS 2015 (e) P5-UK (f) PETS 2000

Fig 10 Protection score (P) obtained by different privacy protection techniques for a variation of filter intensity on (a) ETH Bahnhof, (b) ETH Sunnyday, (c) iLids Easy, (d) PETS 2015, (e) P5-UK and (f) PETS 2000.

5.4 Results on existing datasets

Fig. 10 and Fig. 11 plot the protection (P) and utility (U) scores, respectively, of the five privacy protection techniques (blanking, blurring, pixelating, cartooning, MP) for a variation of i on ETH Bahnhof, ETH Sunnyday, iLids Easy, PETS 2015, P5-UK and PETS 2000.

The trends of the protection scores (P) obtained by five techniques over a variation of i are generally the same across all sequences (Fig. 10(a-f)) and similar to the results reported on TST-Priv sequences (Sec. 5.3). As in the case of TST-Priv, on all sequences blanking achieves the highest P (Fig. 10(a-f)). After blanking, pixelating consistently obtains the highest P followed by blurring and cartooning on all sequences for the entire variation of i. MP has generally shown the lowest P among all techniques for most of the variation of i on all sequences. Moreover, the trends of the utility scores (U) obtained by the five methods over a variation of i are alike across all sequences (Fig. 11(a-f)). The highest U is obtained by cartooning followed by blurring, pixelating,



(a) ETH Bahnhof (b) ETH Sunnyday (c) iLids Easy

(d) PETS 2015 (e) P5-UK (f) PETS 2000

Fig 11 Utility score (U) obtained by different privacy protection techniques for a variation of filter intensity on (a) ETH Bahnhof, (b) ETH Sunnyday, (c) iLids Easy, (d) PETS 2015, (e) P5-UK and (f) PETS 2000.

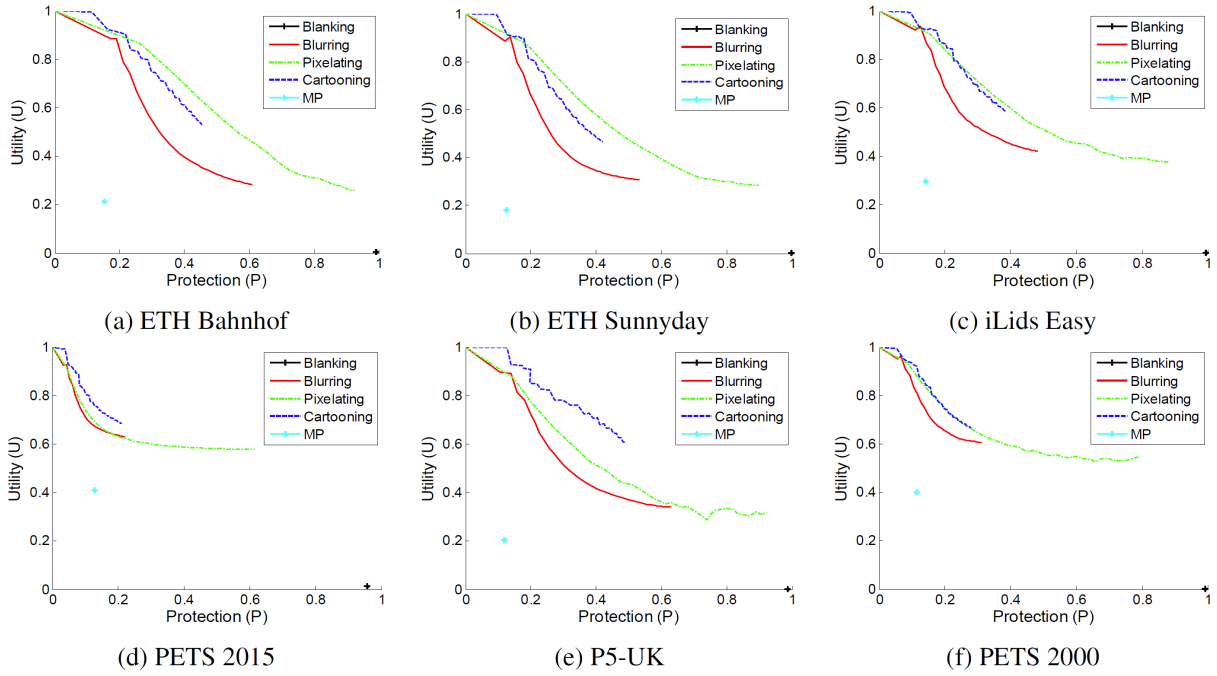
MP and blanking, which also corresponds to the results obtained on TST-Priv sequences (Sec. 5.3).

As done in the case of TST-Priv sequences, we also plot U *vs.* P for ETH Bahnhof, ETH Sunnyday, iLids Easy, PETS 2015, P5-UK and PETS 2000 (Fig. 12(a-f)). In general pixelating is found to provide a better trade off between U and P on datasets with *full body* target (Fig. 12(a-c)), and cartooning provides a better trade off on datasets with *vehicle* target (Fig. 12(d-f)).

5.5 Discussion

We checked the statistical significance of the P and U scores obtained by the privacy protection techniques on each of the 12 sequences using the Welch ANOVA test.[62] Statistical significance is achieved at the standard 5% significance level both for the case of P and U scores on each sequence.

Table 3 provides the overall performance of privacy protection techniques computed in the form of mean P (P_μ) and mean U (U_μ) over the entire variation of filter intensity (i) on all TST-Priv



(a) ETH Bahnhof (b) ETH Sunnyday (c) iLids Easy

(d) PETS 2015 (e) P5-UK (f) PETS 2000

Fig 12 Utility score (U) plotted *vs.* privacy score (P) obtained by different privacy protection techniques for a variation of filter intensity on (a) ETH Bahnhof, (b) ETH Sunnyday, (c) iLids Easy, (d) PETS 2015, (e) P5-UK and (f) PETS 2000.

sequences. Blanking expectedly achieves the highest P_μ on all sequences. Among the remaining techniques, TIR shows the best P_μ on every sequence except on TST-Priv-P2 where pixelating is the best. However, based on visual judgement, TIR data seems to possess features (see Fig. 3, 'TIR' column) that might enable target recognition thus alluding it might not be as privacy-protecting as reflected by it's high protection scores. Additionally, cartooning is the best in terms of U_μ on all sequences. Moreover, we also computed the cumulative protection and utility scores that are the average scores over all sequences (Table 3). The results show that (except blanking) TIR has the best cumulative protection score and cartooning has the best cumulative utility score.

Similarly, Table 4 lists the mean P (P_μ) and mean U (U_μ) over the variation of i on ETH Bahnhof, ETH Sunnyday, iLids Easy, PETS 2015, P5-UK and PETS 2000. Except blanking that again has the best P_μ, pixelating outperforms blurring, cartooning and MP in terms of P_μ on all

Table 3 Overall performance of privacy protection techniques in terms of mean P (P_μ) and mean U (U_μ) over a variation of filter intensity (i) on all TST-Priv sequences. Note that blanking, MP and TIR of course remain unaffected over a variation of i. Cumulative (Cum.) protection and utility scores are also listed that are the average scores over all sequences. On each sequence the best two techniques with the highest P_μ are highlighted in yellow and the best two techniques with the highest U_μ are highlighted in gray; the darker the color the better the performance.

Sequence	**Blanking**		**Blurring**		**Pixelating**		**Cartooning**		**MP**		**TIR**	
	$\mathbf{P}_\mu$	$\mathbf{U}_\mu$	$\mathbf{P}_\mu$	$\mathbf{U}_\mu$	$\mathbf{P}_\mu$	$\mathbf{U}_\mu$	$\mathbf{P}_\mu$	$\mathbf{U}_\mu$	$\mathbf{P}_\mu$	$\mathbf{U}_\mu$	$\mathbf{P}_\mu$	$\mathbf{U}_\mu$
TST-Priv-F1	0.99	0.03	0.21	0.73	0.58	0.62	0.21	0.85	0.16	0.43	0.72	0.28
TST-Priv-F2	1.00	0.00	0.17	0.88	0.48	0.79	0.17	0.93	0.16	0.64	0.84	0.50
TST-Priv-P1	0.99	0.01	0.22	0.49	0.55	0.40	0.18	0.73	0.12	0.17	0.76	0.22
TST-Priv-P2	0.98	0.02	0.31	0.51	0.66	0.42	0.26	0.77	0.16	0.23	0.54	0.16
TST-Priv-V1	1.00	0.00	0.32	0.40	0.61	0.34	0.25	0.66	0.13	0.16	0.79	0.21
TST-Priv-V2	1.00	0.00	0.34	0.55	0.65	0.47	0.28	0.74	0.15	0.25	0.80	0.35
Cum. score	0.99	0.01	0.26	0.59	0.59	0.51	0.22	0.78	0.15	0.31	0.74	0.29

sequences. As in the case of TST-Priv sequences, cartooning again shows the best U_μ on all sequences. In terms of cumulative protection score blanking and pixelating are the best, and in terms of cumulative utility score cartooning and blurring are the best (Table 4).

6 Conclusions

This paper presented a new annotation-free and target-independent objective evaluation method for image-based privacy protection techniques. The proposed method evaluates *protection* by quantifying the Bhattacharyya distance-based appearance similarity and *utility* by measuring the

Table 4 Overall performance of privacy protection techniques in terms of mean P (P_μ) and mean U (U_μ) over a variation of filter intensity (i) on ETH Bahnhof, ETH Sunnyday, iLids Easy, PETS 2015, P5-UK and PETS 2000. Note that blanking and MP of course remains unaffected over a variation of i. Cumulative (Cum.) protection and utility scores are also listed that are the average scores over all sequences. On each sequence the best two techniques with the highest P_μ are highlighted in yellow and the best two techniques with the highest U_μ are highlighted in gray; the darker the color the better the performance.

Sequence	**Blanking**		**Blurring**		**Pixelating**		**Cartooning**		**MP**	
	$\mathbf{P}_\mu$	$\mathbf{U}_\mu$	$\mathbf{P}_\mu$	$\mathbf{U}_\mu$	$\mathbf{P}_\mu$	$\mathbf{U}_\mu$	$\mathbf{P}_\mu$	$\mathbf{U}_\mu$	$\mathbf{P}_\mu$	$\mathbf{U}_\mu$
ETH-Bahnhof	0.99	0.00	0.43	0.44	0.77	0.36	0.31	0.73	0.15	0.21
ETH Sunnyday	1.00	0.00	0.36	0.44	0.72	0.37	0.27	0.68	0.13	0.18
iLids Easy	0.99	0.00	0.32	0.56	0.67	0.47	0.24	0.77	0.14	0.30
PETS 2015	0.96	0.01	0.14	0.70	0.35	0.64	0.13	0.79	0.13	0.41
P5-UK	0.99	0.00	0.39	0.50	0.69	0.40	0.31	0.78	0.12	0.20
PETS 2000	0.99	0.00	0.20	0.69	0.54	0.60	0.17	0.80	0.11	0.40
Cum. score	0.99	0.01	0.31	0.56	0.62	0.47	0.24	0.76	0.13	0.28

SSIM-based structural similarity between the original and the privacy-protected image regions. We showed the effectiveness and advantages of the proposed method using the PEViD dataset containing visible imagery over an existing objective evaluation method, which is widely used and employs the face recognition accuracy as a measure of protection and the face detection accuracy as a measure of utility. We highlighted that the proposed method interestingly offered similar results (in terms of ranking of different privacy protection techniques) to those of that existing method, thus showing a strong correlation: Spearman's coefficient=1. We underlined the advantages of the proposed method of being annotation free and target independent over that existing method. Moreover, using the PEViD dataset we also adopted a web-based subjective assessment approach to further validate the effectiveness of the proposed method. Again, in this case, a strong correlation (Spearman's coefficient=1) has been reported between the rankings of privacy-protection techniques obtain using the proposed method and those obtained based on the judgements of a set of human subjects that include people who are skilled, semi-skilled and unskilled in privacy protection in videos.

We also conducted an extensive experimentation on six challenging datasets (including a new one) containing 12 sequences with face, full person body and vehicle as target types. The new dataset, called TST-Priv, contains six sequences with both visible and thermal imagery, and is made available online for the community. We demonstrated a statistically-significant comparison of a diverse set of six privacy protection techniques using the proposed evaluation method. While some techniques (blurring, pixelating, cartooning) involve the use of filter intensity to provide varying protection strengths, others (blanking, motion perturbation (MP), thermal infrared imaging (TIR)) do not rely on the use of filter intensity. Blanking is expectedly the best in terms of protection score. It, however, causes a total loss of the information that is likely to be not desir-

able in general. Among the remaining techniques, TIR or pixelating showed the highest protection score. It is, however, important to highlight that TIR visually seems not to preserve the privacy well because of an apparent presence of recognizable target features in the imagery (see Fig. 3, 'TIR' column), which of course does not correspond to the conclusions drawn based on its quantitative evaluation using the proposed method. Therefore, a need remains to subjectively assess the proposed evaluation method for the case of TIR imagery as well. On the other hand, in terms of utility score, cartooning is consistently the best and blurring is the second best. Although TIR visually seems to preserve the structural information well, it has obtained a lower utility than expected in the experiments. This is likely to be due to the offset caused in the estimated bounding box patches on the thermal imagery as a result of possible inaccuracies in the homography computation. Finally, in terms of providing an appropriate trade off between protection and utility, cartooning and pixelating are generally found to be desirable.

The proposed evaluation method also has some limitations. First, there exist works that are aimed at identification of persons by analyzing body structure (e.g. anthropometric biometrics)[63,64] and gait.[65,66] Often these approaches rely on *temporal* measurements by analyzing recorded imagery over a period of time. In such a case, the problem of privacy protection and its evaluation becomes different, and is expected to account for *temporal* information of targets. Our work instead focuses on *image-based* measurements and the above-mentioned *temporal* aspects are out of the scope and hence not considered. Second, the proposed evaluation method may not provide completely plausible results for a privacy protection that involves showing only the target contours. In such a case, the proposed evaluation method is expected to give a higher (plausible) protection score but a lower (nonplausible) utility score. Third, the use of the proposed evaluation method (and, potentially, some of the existing related methods) is limited mainly to the visible imaging

modality and their application to other imaging modalities (e.g. TIR) needs to be further experimentally analyzed and validated based on, for example, the subjective assessment. Future work could focus in addressing the above limitations. Moreover, in this work, while the usefulness of the proposed evaluation method is shown in the context of video tracking by applying privacy protection techniques locally on estimated (target) image regions by trackers, the method is actually not constrained by tracking. In fact, the proposed method is generic and could be employed for evaluating privacy protection techniques whether applied on image regions generated by a tracker or otherwise, or even on full frames.

Disclosures

The financial conflicts of interest are stated under the acknowledgements below. There are no other known conflicts of interest.

Acknowledgments

The research was funded by the Swedish Research Council through the project, Learning Systems for Remote Thermography, grant no. D0570301, as well as by the European Community Framework Programme 7, Privacy Preserving Perimeter Protection Project (P5), grant agreement no. 312784.

References

1 S. Fleck and W. Strasser, "Smart camera based monitoring system and its application to assisted living," *Proceedings of IEEE* **96**(10), 1698–1714 (2008).

2 A. Cavallaro, "Privacy in video surveillance," *IEEE SPM* **24**(2), 165–166 (2007).

3 T. Winkler and B. Rinner, "Security and privacy protection in visual sensor networks: A survey," *ACM Comp. Surv.* **47**(1) (2014).

4 P. Agrawal and P. J. Narayanan, "Person de-identification in videos," *IEEE Trans. CSVT* **21**(3), 299–310 (2011).

5 A. J. Aved and K. A. Hua, "A general framework for managing and processing live video data with privacy protection," *Multimedia Systems* **18**(2), 123–143 (2012).

6 A. Erdelyi, T. Barat, P. Valet, *et al.*, "Adaptive cartooning for privacy protection in camera networks," in *Proc. of IEEE AVSS*, (Seoul) (2014).

7 B.-J. Han, H. Jeong, and Y.-J. Won, "The privacy protection framework for biometric information in network based CCTV environment," in *Proc. of ICOS*, (Langkawi) (2011).

8 S. E. Hudson and I. Smith, "Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems," in *Proc. of ACM CSCW*, (Boston) (1996).

9 A. Martinez-Balleste, H. A. Rashwan, D. Puig, *et al.*, "Towards a trustworthy privacy in pervasive video surveillance systems," in *Proc. of IEEE PERCOM Workshops*, (Lugano) (2012).

10 E. M. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying face images," *IEEE Trans. KDE* **17**(2), 232–243 (2005).

11 F. Z. Qureshi, "Object-video streams for preserving privacy in video surveillance," in *Proc. of IEEE AVSS*, (Genova) (2009).

12 K. Liu, C. Giannella, and H. Kargupta, *Privacy-Preserving Data Mining: Models and Algorithms*, ch. A Survey of Attack Techniques on Privacy-Preserving Data Perturbation Methods, 359–381. Springer US (2008).

13 F. Zhang, L. He, W. He, *et al.*, "Data perturbation with state-dependent noise for participatory sensing," in *Proc. of IEEE INFOCOM*, (Orlando, FL) (2012).

14 E. Aspelin and B. Almebäck, Eds., *Optisk och elektronisk övervakning (Optical and electronic surveillance)*, Statens Offentliga Utredningar 1987:74, Ministry of Justice, Stockholm, Sweden (1987).

15 Y. Zhang, Y. Lu, H. Nagahara, *et al.*, "Anonymous camera for privacy protection," in *Proc. of IEEE ICPR*, (Stockholm) (2014).

16 X. Maldague, *Theory and Practice of Infrared Technology for Nondestructive Testing*, Wiley-Interscience (2001).

17 R. S. Ghiass, O. Arandjelovic, H. Bendada, *et al.*, "Infrared face recognition: A literature review," in *Proc. of IJCNN*, (Dallas, TX) (2013).

18 A. B. Chan, Z.-S. J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in *Proc. of IEEE CVPR*, (Anchorage, AK) (2008).

19 D. Conte, P. Foggia, G. Percannella, *et al.*, "A method for counting people in crowded scenes," in *Proc. of IEEE AVSS*, (Boston, MA) (2010).

20 H. Fradi, V. Eiselein, I. Keller, *et al.*, "Crowd context-dependent privacy protection filters," in *Proc. of ICDSP*, (Fira) (2013).

21 M. Saini, P. Atrey, S. Mehrotra, *et al.*, "Anonymous surveillance," in *Proc. of IEEE ICME*, (Barcelona) (2011).

22 T. Nawaz and J. Ferryman, "An annotation-free method for evaluating privacy protection techniques in videos," in *Proc. of AVSS*, (Karlsruhe) (2015).

23 N. Babaguchi, T. Koshimizu, I. Umata, *et al.*, *Protecting Privacy in Video Surveillance*, ch. Psychological Study for Designing Privacy Protected Video Surveillance System: PriSurv, 147–164. Springer London (2009).

24 P. Birnstill and D. Ren, "A user study on anonymization techniques for smart video surveillance," in *Proc. of AVSS*, (Karlsruhe) (2015).

25 M. Boyle, C. Edwards, and S. Greenberg, "The effects of filtered video on awareness and privacy," in *Proc. of CSCW*, (Philadelphia, PA) (2000).

26 P. Korshunov, C. Araimo, F. D. Simone, *et al.*, "Subjective study of privacy filters in video surveillance," in *Proc. of IEEE Work. MMSP*, (BANFF) (2012).

27 Q. A. Zhao and J. T. Stasko, "Evaluating image filtering based techniques in media space applications," in *Proc. of ACM CSCW*, (Seattle, WA) (1998).

28 P. Korshunov, A. Melle, J.-L. Dugelay, *et al.*, "Framework for objective evaluation of privacy filters," in *Proc. of SPIE 8856, Applications of Digital Image Processing XXXVI*, (San Diego, California) (2013).

29 M. Saini., P. K. Atrey, S. Mehrotra, *et al.*, "Privacy modeling for video data publication," in *Proc. of IEEE ICME*, (Suntec City) (2010).

30 M. Saini, P. K. Atrey, S. Mehrotra, *et al.*, "W^3-privacy: understanding what, when, and where inference channels in multi-camera surveillance video," *MTAP* **68**(1), 135–158 (2014).

31 T. Nawaz, F. Poiesi, and A. Cavallaro, "Assessing tracking assessment measures," in *Proc. of ICIP*, (Paris) (2014).

32 F. Dufaux and T. Ebrahimi, "A framework for the validation of privacy protection solutions in video surveillance," in *Proc. of IEEE ICME*, (Singapore) (2010).

33 S. Baker, D. Scharstein, J. Lewis, *et al.*, "A database and evaluation methodology for optical flow," *IJCV* **92**(1), 1–31 (2011).

34 D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IJCV* **47**(1/2/3), 7–42 (2002).

35 M. Kristan, R. Pflugfelder, A. Leonardis, *et al.*, "The vot2013 challenge: overview and additional results," in *Proc. of CVWW*, (Krtiny) (2014).

36 T. Nawaz, F. Poiesi, and A. Cavallaro, "Measures of effective video tracking," *IEEE Trans. IP* **23**(1), 376–388 (2014).

37 "http://www.multimediaeval.org/mediaeval2015/. Accessed September 2016."

38 F. J. Aherne, N. A. Thacker, and P. I. Rockett, "The Bhattacharyya metric as an absolute similarity measure for frequency coded data," *Kybernetika* **34**(4), 363–368 (1998).

39 Y. Rubner, C. Tomasi, and L. J. Guibas, "A metric for distributions with applications to image databases," in *Proc. of ICCV*, (Bombay) (1998).

40 B. Huet and E. R. Hancock, "Line pattern retrieval using relational histograms," *IEEE Trans. PAMI* **21**(12), 1363–1370 (1999).

41 T. Kailath, "The divergence and Bhattacharyya distance measures in signal selection," *IEEE Trans. on Commun. Technol.* **15**(1), 52–60 (1967).

42 I. B. Ayed, K. Punithakumar, and S. Li, "Distribution matching with the Bhattacharyya similarity: A bound optimization framework," *IEEE Trans. PAMI* **37**(9), 1777–1791 (2015).

43 A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bulletin of the Calcutta Mathematical Society* **35**, 99–109 (1943).

44 Z. Wang, A. C. Bovik, H. R. Sheikh, *et al.*, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. IP* **13**(4), 600–612 (2004).

45 F. Dufaux, "Video scrambling for privacy protection in video surveillance: recent results and validation framework," in *Proc. SPIE 8063, Mobile Multimedia/Image Processing, Security, and Applications*, (Orlando, Florida) (2011).

46 O. Sawar, B. Rinner, and A. Cavallaro, "Design space exploration for adaptive privacy protection in airborne images," in *Proc. of IEEE AVSS*, (Colorado Springs, CO) (2016).

47 T. Nawaz, B. Rinner, and J. Ferryman, "User-centric, embedded vision-based human monitoring: A concept and a healthcare use case," in *Proc. of ACM ICDSC*, (Paris) (2016).

48 "http://multimediaeval.org/mediaeval2013/. Accessed September 2016."

49 P. Korshunov and T. Ebrahimi, "Pevid: privacy evaluation video dataset," in *Proc. SPIE 8856, Applications of Digital Image Processing XXXVI*, (San Diego, California) (2013).

50 "Subjective video quality assessment methods for multimedia applications. http://videoclarity.com/pdf/t-rec-p.910-199909-ipdf-e1.pdf," (last accessed on May 2017).

51 "ETH Bahnhof and Sunnyday Datasets. http://www.vision.ee.ethz.ch/~aess/iccv2007/. Accessed March 2015."

52 "http://www.eecs.qmul.ac.uk/~andrea/avss2007_d.html. Accessed March 2015."

53 L. Li, T. Nawaz, and J. Ferryman, "PETS 2015: Datasets and challenge," in *Proc. of AVSS*, (Karlsruhe) (2015).

54 "PETS 2000 dataset. ftp://ftp.cs.rdg.ac.uk/pub/PETS2000/. Accessed March 2015."

55 "EU project P5 (Privacy Preserving Perimeter Protection Project). http://www.p5-fp7.eu. Accessed June 2015."

56 M. Felsberg, "The Thermal Infrared Visual Object Tracking VOT-TIR2015 Challenge Results," in *Proc. of ICCV Workshops*, (Santiago) (2015).

57 R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2nd ed. (2004).

58 H. Pirsiavash, D. Ramanan, and C. C. Fowlkes, "Globally-optimal greedy algorithms for tracking a variable number of objects," in *Proc. of IEEE CVPR*, (Colorado Springs, CO) (2011).

59 J. Ning, L. Zhang, D. Zhang, *et al.*, "Robust mean-shift tracking with corrected background-weighted histogram," *IET Computer Vision* **6**(1), 62–69 (2012).

60 T. Nawaz and A. Cavallaro, "A protocol for evaluating video trackers under real-world conditions," *IEEE Trans. IP* **22**(4), 1354–1361 (2013).

61 H. Kargupta, S. Datta, Q. Wang, *et al.*, "On the privacy preserving properties of random data perturbation techniques," in *Proc. of IEEE ICDM*, (Melbourne, Florida) (2003).

62 B. L. Welch, "On the comparison of several mean values: An alternative approach," *Biometrika* **38**(3-4), 330–336 (1951).

63 B. C. Munsell, A. Temlyakov, C. Qu, *et al.*, "Person identification using full-body motion and anthropometric biometrics from kinect videos," in *Proc. of ECCV Workshops*, (Firenze) (2012).

64 V. O. Andersson and R. M. Araujo, "Person identification using anthropometric and gait data from kinect sensor," in *Proc. of AAAI*, (Austin, Texas) (2015).

65 Z. Wu, Y. Huang, L. Wang, *et al.*, "A comprehensive study on cross-view gait based human identification with deep cnns," *IEEE Trans. PAMI* **39**(2), 209–226 (2017).

66 W. Kusakunniran, Q. Wu, J. Zhang, *et al.*, "A new view-invariant feature for cross-view gait recognition," *IEEE Trans. IFS* **8**(10), 1642–1653 (2013).

Tahir Nawaz received a PhD in 2014 jointly from Queen Mary University of London, UK, and Alpen-Adria University of Klagenfurt, Austria, and an MSc in 2009 jointly from Heriot-Watt University, UK, University of Girona, Spain, and University of Burgundy, France. Since 2014, he is working as a Post-Doctoral Research Assistant at University of Reading, UK. He has published several journal and conference papers, is a reviewer of well-known journals, and was Co-organizer of PETS 2015.

Amanda Berg received her M.Sc. degree in Applied Physics and Electrical Engineering in 2013 from Linköping University, Sweden. She is currently an industrial Ph.D. student at the company Termisk Systemteknik and the Computer Vision Laboratory, Linköping University.

James Ferryman is Professor of Computational Vision and leads CVG housed within the Department of Computer Science, University of Reading, UK. He has over 20 years of experience in image and video analysis and has co-authored over 100 papers related to computer vision. He has been a PI on several EPSRC and EC projects. He also leads the IEEE International series of PETS workshops and has been a Director of both BMVA and SITC.

Jörgen Ahlberg received his M.Sc. degree in Computer Science and Engineering in 1996 and his Ph.D. in Electrical Engineering in 2002, both from Linköping University, Sweden. He then held positions as scientist and research leader at FOI, the Swedish Defence Research Agency for nine years. He is currently an Adjunct Senior Lecturer at Linköping University and runs R&D projects at the companies Visage Technologies, Termisk Systemteknik, and Glana Sensors.

Michael Felsberg received his PhD degree in engineering from Kiel University in 2002. Since 2008, he is Full Professor and Head of the Computer Vision Laboratory, Linköping University. He has published more than 120 reviewed conference papers, journal articles, and book contributions. He has won several visual tracking challenges, received several paper awards, and the Olympus award in 2005. He is General Chair of CAIP 2017 and Associate Editor for JMIV, IMAVIS, and JRTIP.

List of Figures

1 Proposed method for objectively evaluating a privacy protection technique in the context of video tracking by quantifying the *protection* and *utility* aspects.

2 Trade off between protection and utility. Completely masking out the object information in an image would provide a maximum protection but at the expense of a zero utility. Likewise, leaving the object information unaffected would provide a maximum utility but at the expense of a total loss of protection.

3 Sample qualitative results for different privacy protection techniques applied on tracked patches with an increasing filter intensity ($i = 20, 30, \ldots, 70$). The corresponding protection (Pr.) and utility (Ut.) scores are also listed under each patch. Key. MP: motion perturbation; TIR: thermal infrared.

4 Average protection score (a,c) and average utility score (b,d) obtained by blurring, pixelating and cartooning on all sequences of PEViD dataset for a variation of filter intensity. (a,b) P and U are computed using 'mean' statistics (i.e. Eq. 12, 17); (c,d) P and U are computed using 'minimum' statistics (i.e. Eq. 13, 18).

5 A snap shot of the subjective assessment website showing the original and the three processed (privacy-protected) videos; all played simultaneously in a synchronised manner.

6 The results of the subjective assessment of privacy-protection techniques (blurring, pixelating, cartooning) in terms of the mean of the ratings of all subjects on each clip.

7 Protection score (P) obtained by different privacy protection techniques for a variation of filter intensity on (a) TST-Priv-F1, (b) TST-Priv-F2, (c) TST-Priv-P1, (d) TST-Priv-P2, (e) TST-Priv-V1 and (f) TST-Priv-V2.

8 Utility score (U) obtained by different privacy protection techniques for a variation of filter intensity on (a) TST-Priv-F1, (b) TST-Priv-F2, (c) TST-Priv-P1, (d) TST-Priv-P2, (e) TST-Priv-V1 and (f) TST-Priv-V2.

9 Utility score (U) plotted *vs.* protection score (P) obtained by different privacy protection techniques for a variation of filter intensity on (a) TST-Priv-F1, (b) TST-Priv-F2, (c) TST-Priv-P1, (d) TST-Priv-P2, (e) TST-Priv-V1 and (f) TST-Priv-V2.

10 Protection score (P) obtained by different privacy protection techniques for a variation of filter intensity on (a) ETH Bahnhof, (b) ETH Sunnyday, (c) iLids Easy, (d) PETS 2015, (e) P5-UK and (f) PETS 2000.

11 Utility score (U) obtained by different privacy protection techniques for a variation of filter intensity on (a) ETH Bahnhof, (b) ETH Sunnyday, (c) iLids Easy, (d) PETS 2015, (e) P5-UK and (f) PETS 2000.

12 Utility score (U) plotted *vs.* privacy score (P) obtained by different privacy protection techniques for a variation of filter intensity on (a) ETH Bahnhof, (b) ETH Sunnyday, (c) iLids Easy, (d) PETS 2015, (e) P5-UK and (f) PETS 2000.

List of Tables

1 State-of-the-art evaluation methods for privacy protection. (Key. NTDR: Non reliance of a method on target detection and recognition accuracies.)

2 Summary of the datasets. Key. K: number of frames; VIS: visual sequence; TH: thermal-infrared sequence; Occ: occlusion; SC: scale changes; IC: illumination changes; Cr: crowdedness; PC: pose changes. Note that '(-)' means that the corresponding thermal-infrared sequence is not available.

3 Overall performance of privacy protection techniques in terms of mean P (P_μ) and mean U (U_μ) over a variation of filter intensity (i) on all TST-Priv sequences. Note that blanking, MP and TIR of course remain unaffected over a variation of i. Cumulative (Cum.) protection and utility scores are also listed that are the average scores over all sequences. On each sequence the best two techniques with the highest P_μ are highlighted in yellow and the best two techniques with the highest U_μ are highlighted in gray; the darker the color the better the performance.

4 Overall performance of privacy protection techniques in terms of mean P (P_μ) and mean U (U_μ) over a variation of filter intensity (i) on ETH Bahnhof, ETH Sunny-day, iLids Easy, PETS 2015, P5-UK and PETS 2000. Note that blanking and MP of course remains unaffected over a variation of i. Cumulative (Cum.) protection and utility scores are also listed that are the average scores over all sequences. On each sequence the best two techniques with the highest P_μ are highlighted in yellow and the best two techniques with the highest U_μ are highlighted in gray; the darker the color the better the performance.