HOLISTIC PRIVACY IMPACT ASSESSMENT FRAMEWORK FOR VIDEO PRIVACY FILTERING TECHNOLOGIES

Atta Badii[1], Ahmed Al-Obaidi[1], Mathieu Einig[1], Aurélien Ducournau[1]

[1] University of Reading
Intelligent Systems Research Laboratory,
School of Systems Engineering
United Kingdom

ABSTRACT

In this paper, we present a novel Holistic Framework for Privacy Protection Level Performance Evaluation and Impact Assessment (H-PIA) to support the design and deployment of privacy-preserving filtering techniques as may be co-evolved for video surveillance through user-centred participative engagement and collectively negotiated solution seeking for privacy protection. The proposed framework is based on the UI-REF normative ethno-methodological framework for Privacy-by-Co-Design which is based on collective-interpretivist and socio-psycho-cognitively rooted Human Judgment and Decision Making (JDM) theory including Pleasure-Pain-Recall (PPR)-theoretic opinion elicitation and analysis. This supports not only the socio-ethically reflective conflicts resolution, prioritisation and traceability of privacy-preserving requirements evolving through user-centred co-design but also the integration of Key Holistic Performance Indicators (KPIs) comprising a number of objective and subjective evaluation metrics for the design and operational deployment of surveillance data/-video-analytics from a system-of-system-scale context-aware accountability engineering perspective. For the objective tests, we have proposed five crucial criteria to be evaluated to assess the optimality of the balance of privacy protection and security assurance as may be negotiated with end-users through co-design of a privacy filtering solution. This evaluation is supported by a process of quantitative assessment of some of the KPIs through an automated objective measurement of the functional performance of the given filter. Additionally, a subjective qualitative user study has been conducted to correlate with, and cross-validate, the results obtained from the objective assessment of the KPIs. The simulation results have confirmed the sufficiency, necessity and efficacy of the UI-REF-based methodologically-guided framework for Privacy Protection evaluation to enable optimally balanced Privacy Filtering of the video frame whilst retaining the minimum of the information as negotiated per agreed process logic. Insights from this study have served the co-design and deployment optimisation of privacy-preserving video filtering solutions. This UI-REF-based framework has been successfully applied to the evaluation of MediaEval 2012-2013 Privacy Filtering and as such has served to motivates further innovation in co-design and multi-level, multi-modal impact assessment of multimedia privacy-security-balancing risk mitigation technologies.

KEYWORDS

Privacy Preserving, Privacy Protection, Video Analytics, UI-REF, Privacy-by-Co-Design, Filtering, Evaluation, Visual Surveillance, Holistic Privacy Impact Assessment (H-PIA), Human Judgement and Decision Making Theory (JDM), Pleasure-Pain-Recall Theory l(PPR), Context-aware Privacy Filtering.

1. INTRODUCTION

The installation of the traditional CCTV cameras, if appropriately deployed, may prove helpful as part of a crime reduction strategy particularly in the urban environment. The increasing awareness

of security threats and the need for efficient and effective means of security monitoring and crime reduction management has led to an exponential growth in the development of fully automated video surveillance systems. However much effort is needed to ensure that the potentially harmful impacts of surveillance technologies are avoided and this requires socially negotiated, context-sensitive, co-design and evaluation of the performance of such systems including the assessment of the societal and legal framework for their adoption, and, operational deployment responsibility. Accordingly our approach is based on a system-of-systems scale (citizen-centred, socio-ethical, societal, legal and technical) perspective for (re)-negotiated evolutionary reflective co-design. This is to serve the transparency-accountability engineering and evaluation of video surveillance solution systems as supported by the UI-REF [1] Privacy-by-Co-Design framework. This is a normative ethno-methodology which supports the situated elicitation, interpretation, prioritisation and context-specific conflicts resolution of both the requirements, and, the criteria for the multi-perspective impacts assessment of the levels of privacy protection actually delivered by a surveillance system and the various social and normative consequences of its adoption.

Computer vision technologies have been designed and customised to automate CCTV operations through video-analytics capabilities such as object detection, tracking, and behaviour recognition. Irrespective of whether these implementations are to take place at the back-end such as in the centralised server system or in the "edge-smart" mode at the camera-end, they can potentially lead to invasion of citizens' privacy with various adverse personal and social impacts. These can arise not just from how securely the data protection is managed but also due to the uncertainty/inaccuracy of the data as captured e.g. noisy images that could cause mis-classifications with no human over-ride in the decisional framework thus propagating false positive/negative classifications based on simplistic behavioural stereotypes as may have been applied to noisy data.

Often referred to as 'Big Brother', such public privacy concerns for video surveillance have been raised since the early days of deployment of passive CCTV. With the increasing maturity of the more advanced active CCTVs, the privacy of the citizen can be potentially at higher risk unless protected through an inclusivist negotiation approach to the design and development of such systems. Privacy risks mitigation technologies need to be deployed but for this to be well-motivated, acceptable, and effective an actionable framework is needed for shared sense-making and co-evolution of privacy-preserving surveillance solutions.

In recent years the computer vision research community has started to contribute to such solutions but from a largely technological perspective and a variety of image processing techniques, for example [2], have been proposed to mitigate privacy protection failure risks.

These approaches tend to apply some privacy filtering to obscure the privacy-sensitive parts of the captured video; in much the same way as is the established practice in the film and television sector. However, the naïve application of such privacy filters could lead to video surveillance systems that are potentially ineffective in either adequately protecting the privacy of the citizen (also referred to as the "data-subject") or in retaining the essential information as justified and established to be absolutely necessary for the intended security monitoring in the given situation.

Accordingly, a Privacy-by-Co-Design requirements elicitation and evaluation approach is crucially needed to enable collective-participative judgements to be negotiated amongst all the implicated stakeholders about: i) the level of surveillance information that can be agreed to be justifiably essential for the situated security context and security protection purpose; ii) the context-specific criteria for attribution of "suspect-ness" to a citizen whose image and behaviour may be captured by a video surveillance camera, and, thus iii) the level and scope of situated (context-dependent) privacy filtering to be afforded to the citizen(s).. This is to take into

consideration a negotiated ethically-compliant category-theoretic judgement about what constitutes "suspect"-ness and thus the respective privacy boundaries to be afforded to a citizen whose image may be captured in a surveillance video-frame. It is acknowledged that the perceived privacy boundaries of a citizen cannot be divorced from the contexts in which the privacy is valued by that citizen and thus needs protecting. A citizen may have a variety of roles, responsibilities and relationship sets; each associated with a particular persona of the citizen as activated in a given context within their everyday life-style e.g. husband, father, employee, boss; and each such persona of the same citizen is commensurate with a particular privacy boundary linked to a certain (type of) context as may be viewed and defined by that citizen.

It is the explication of such privacy boundaries in each context for each persona of each citizen and the subjective ascribing of a meaning, a value/sensitivity for the data in the respective context as perceived by the citizen that accordingly specifies the boundaries of privacy filtering that must be maintained whilst retaining the minimum required data-intelligence (the "privacy-security-balancing" optimisation challenge; sometimes also referred to as the privacy-intelligibility "trade-off"). Here, a pre-requisite to the notion of a "trade-off" between privacy and data intelligence (i.e. the retained personal information post filtering) is the existence of an agreed context-specific framework of meaning-value system mappings across two seemingly incongruent worlds. Thus the challenge of privacy-security optimisation implies that the Privacy-by-Co-Design approach must support a negotiated resolution of such so-called "trade-offs". Clearly the system must also safeguard against any discriminatory, inequitable and stereotyping classification of any citizens as "Suspects" arising from mere video-analytics based attributions of suspicious behaviour.

These mandatory capabilities of the Privacy-by-Co-Design system call for context-aware privacy filtering and evaluation to assess the situated optimality of privacy. Although some advances have been recently reported in the application of video privacy filtering techniques, for example as in [3]; the negotiation-centric co-innovation, co-evaluation and thus optimisation of situated privacy filtering has only just begun. This requires participative definition of both video capture context and security context for responsive, trace-able and accountable video-content category judgments supported by a framework such as UI-REF to provide high resolution requirements prioritisation and forensic designation of the Holistic Privacy Evaluation and Impact Assessment Metrics [4, 5].

In practice, video analytics techniques can be deployed to detect the privacy-sensitive information of the data-subjects as featured in video-frames (e.g. faces). The privacy-sensitive elements of a user's profile or any part of their picture need to be identified based on negotiation, and explicitly stated user-led preferences-in-context. For some elements in some contexts a default rule consistent with the applicable privacy-regimes-in-context may provide a pointer to the user's privacy preferences subject to explicit prior agreement of the user. Afterwards, image processing techniques could be deployed to obscure the detected sensitive elements.

The main contributions of this paper could be summarised as the following:

- An Integrated Holistic Evaluation and Impact Assessment Framework is proposed for privacy filtering using novel evaluation criteria.
- A comparison of several conventional privacy filters is presented; these can be used as reference for future works.
- Both objective and subjective evaluation are performed and combined for cross-validation and conclusive results – an integrated quantitative and qualitative approach.
- Numerical metrics are proposed to assess the efficacy of the performance evaluation and impact assessment criteria.

2. RELATED WORK

The literature includes many privacy filtering techniques that can be broadly classified as reversible and irreversible methods. The former approach includes scrambling and encryption methods, e.g. [6], which confer the benefit of a data recovery option but also, have the disadvantage of exposure to privacy protection risks arising from possible hacking. The latter, on the other hand, includes image processing and filtering techniques such as Gaussian blurring and image pixilation, e.g. [7], which require more careful deployment to exclusively conceal the privacy-sensitive elements within given video-frames. As a result, the widely-shared view of the stakeholders, as expressed by Andrew Senior [3], for example, has been to call for advances in effectiveness measurement of the privacy protection level afforded by video surveillance solutions; this is the challenge to which the work reported in this and relevant other work [1, 4, 5] have responded through the pioneering methodologically-guided UI-REF-enabled approach to negotiation-centric co-design as may be applied to context-aware socio-technical personalisation of requirements e.g. for privacy and its evaluation and holistic impact analysis.

However, to-date, only a few attempts have been reported which aim to assess and evaluate the performance and impact of privacy-protecting systems within a relatively limited analysis perspective.

Dufaux [8] has proposed a privacy filtering evaluation framework with validated pixilation, blurring, and scrambling filtering methods based on: i) objective image quality measures using PSNR and SSIM; ii) Face recognition performance using PCA and LDA.

Zhao and Stasko [9] have attempted to assess the relative accuracy of various filtering approaches by conducting a comparative study of image filtering techniques that seek to mask the object identity and activity in videos. Boyle et al. [10] have performed a subjective evaluation of blurred and pixelated video filtering techniques. In their study, the filters had been tested at different levels of fidelity and examined by human observers.

Accordingly the research study reported in this paper has responded to the need for a more inclusive, holistic and high resolution assessment of privacy filtering requirements as well as the evaluation of the effectiveness of the resulting privacy filtering solutions. We have addressed the need for more exhaustive evaluation criteria to critically assess the privacy-preserving and security-informative-ness balance of the situated filtering methods. Therefore, the primary contribution of this work is the extended and integrated framework of context-aware requirements prioritisation, and, objective and subjective evaluations to automatically validate the effectiveness and impact of privacy filters based on user-centred preferences and metrics.

The rest of this manuscript is structured as follows: section 2 defines the proposed evaluation criteria which form the basis of the proposed framework as described in Section 3. Section 4 discusses the simulation results. Finally Section 5 sets out the conclusions and briefly indicates the scope for future work in responding to the outstanding societal challenges of optimising the effectiveness of the protection of the privacy of the citizens as well as the security monitoring.

3. PRIVACY-PRESERVING EVALUATION CRITERIA

The UI-REF is an established normative ethno-methodology [1, 5], providing a framework for integrative high resolution context-aware requirements ranking and usability-relationships-based metrics for evaluation. This includes the Effects, Side-Effects and Cross-Effects, Affects criteria set (The ESA Matrix). This confers several advantages as a means of holistic socio-psycho-

cognitive impact assessment based on the context-aware, and, relationships-aware measurement of Effects (direct impacts of a proposed privacy protection solution), Side-Effects and Cross-Effects (secondary effects including un-intended/unforeseen/indirect effects), and Affects. The *Affects* are the multi-level multi-modal emotional, psycho-social-and sentimental effects arising from the primary/direct effects and/or secondary effects or side-effects of the adoption of a proposed solution for privacy-preserving surveillance video-monitoring and its associated video-analytics including privacy filtering. Therefore the privacy filtering evaluation framework as studied in our recent work, as reported here, builds on and extends the above UI-REF based evaluation and impact assessment criteria to derive a new set of contextualised metrics that can be used for both objective and subjective evaluation of privacy-protecting filters compliant with UI-REF-based Privacy-by-Co-Design.

In this context the privacy filtering algorithm design has to achieve the optimum balance of privacy protection with minimal loss of necessary information as deemed essential to the approved mission of the surveillance. Thus the key UI-REF based framework metrics for Privacy-by-Co-Design that constitute the focus of this study include the following:

- **Efficacy**
  The main objective of any visual privacy-protecting filter is the ability to effectively obscure the privacy-sensitive elements. In other words, the system should have the capacity to de-identify the "data-subject" whose image is captured in the video-frames. The measure of this criterion could be obtained using a combination of tests that for instance could separately examine the face region as well as the full body. To fulfil this criterion, the privacy-protecting filter should prevent the human evaluator from being able to detect any faces and/or re-identify the person whose image is privacy-filtered.

- **Consistency**
  Object tracking is an essential functionality for the majority of video surveillance applications. In order to successfully and continuously track the moving subject in the field of view of a single camera or over a network of cameras, the short-term visual appearance of the subject is required to support this task. Therefore, the privacy-protecting filter needs to maintain a reasonable and consistent level of detail of the person's body shape and appearance. Successful cross-frame object tracking of a filtered subject would fulfil this criterion.

- **Disambiguity**
  This is the degree by which a privacy filter does not introduce additional ambiguity in cross-frame trackability of same persons/objects. The intra-object-class variations are the visual cues on which the majority of object classification and tracking algorithms depend. Thus, a privacy filter should not alter the subject to the point that the subject could not be distinguishable from other subjects within the same object class as may be encountered inter/intra frame. This criterion could be tested using a person re-identification procedure and applying a one-vs.-all strategy.

- **Intelligibility**
  This criterion examines the ability of the privacy filtering system 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. Dufaux [8] has proposed an approach for this assessment using the structural Similarity Index (SSIM) quality metrics. However, we propose to use an analytics-based approach similar to that used for pedestrian detection and recognition applications.

- **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. This would depend on how stylistically coherent and suitably blended such visual effects might appear after applying the filtering techniques. Methods for colour histogram comparison of before-and-after effects in addition to the Structural Similarity Index (SSIM) could provide an indication of filtering performance to meet this requirement.

The above evaluation criteria can be similarly applied to privacy-filtered audio-segments if an audio-track is present. These criteria constitute a sub-set of the requirements and evaluation framework as set out within the UI-REF Privacy-by Co-Design Methodology for user-centred, negotiable, context-aware co-evolution and holistic assessment of the impacts of privacy-preserving video surveillance as set out in Badii 2012 [4]; these include:

1. Negotiated ontological framework for innovation and deployment of mitigation technologies for privacy preserving video surveillance co-design.
2. Accordingly the establishment of prototypical templates to delineate an agreed context hierarchy within which various video-analytics operations can be co-specified for specific agreed contexts-objectives; for example, Low-Level Video-Analytics (LLDA), Shallow Vide-Analytics (SVA), Deep Video-analytics (DVA), and, Context-Aware Video-Analytics (CAVA) as sub-spaces of deployment of surveillance technologies [4].
3. A decisional framework for ethically, legally and normatively coherent category judgements as to the classification of persons whose images are captured by video surveillance cameras and accordingly their *tentative attribution*, to be (dis)confirmed, from a privacy-security viewpoint e.g. as Suspects or Non-Suspects based on types and levels of evidence agreed to be un-mistake-ably indicative. This will follow a *least-and-latest-commitment* hierarchical evidence-based decisional framework for trace-able and reversible *man-in-the-loop* classifications.
4. Key Holistic Performance Indicators (KPIs) such as Efficacy, Consistency, Intelligibility and Disambiguity as defined above.
5. Quality-of-Experience: UI-REF-based use-context-dependent Effects, SideEffects and Cross-Effects, Affects (ESA) as part of the Holistic Impact Assessment of a proposed Privacy-preserving socio-technical system-of-systems [1].
6. User-centred perceived aesthetics (structural, textural, tonal, harmony and symmetry maps).
7. Selectivity, sensitivity, of the Privacy Filtering solution system.
8. Computational efficiency of the Privacy Filtering solution system.
9. Scalability (for real-time web-scale) of the Privacy Filtering solution deployment.
10. The level of vulnerability to attack within a proposed video filtering technique and the additional computational cost of protecting it against unauthorised attempts at privacy protection reversal.

4. EVALUATION FRAMEWORK

The evaluation framework includes subjective and objective measurements of which the sub-set defined in the previous section has been implemented for the present study. The aim was to test the privacy filtering method based on user-perceive visual effects and preferences (Human

Vision) as well as automatic video analysis (Computer Vision) requirements. The evaluation framework was designed to encourage intelligent application of image filtering algorithms over the different parts of the object featured in the video. The following sets out the description of the framework components.

4.1. Subjective metrics

The subjective evaluation metrics were used to cross-validate the objective metrics by measuring the performance of the privacy filter with respect to the same defined criteria. These were deployed in two passes in order to test the filtered video in terms of the ability of person recognition and the perception of the privacy filter separately.

4.1.1. Person recognition from pictures

In the first pass, participants were asked to recognise a given unfiltered image of a person in a given set of filtered pictures. The filtered images were uniformly filtered using the same filter type. To avoid bias due to the background or illumination, the images from the test set were extracted from a different camera view point. The selected images were extracted from PEViD dataset [11]. A sample of this test is shown in Figure1. The score resulting from this test contributed to the assessment of the Disambiguity and Efficacy criteria through the level of the recognisability.

Figure 1: Original-filtered, matching for person recognition

4.1.2. Questionnaire on videos

In the second stage, the participants were asked to watch a set of privacy-filtered videos from the PEViD dataset [11] representing different indoor and outdoor scenarios and featuring single and multiple persons, and had to answer a questionnaire relating to different aspects of the filter.
From a meta-descriptive and interpretivist perspective, a set of questions were designed to examine the aesthetics i.e. pleasantness of the visual effects resulting from the privacy filter and the level of distraction that might be consequently experienced by the viewer. Another question was devised to assess the stylistic congruence or suitability of the blended visual effects in the privacy-filtered region of the video-frame. The scores of the answers for these questions were collectively considered to reflect the aesthetic criterion.

Other questions were designed to detect whether the filter could trick the viewer into thinking some features were still obvious, such as the gender and the ethnicity. These questions were given in pairs, with the first one being “Is it possible to guess the person’s gender?”, and the second: “What is the gender of the person?” The ability of the viewer to answer this pair of questions correctly would indicate that the filter as deployed was inadequately effective. However, answering “Yes” to the first question but failing to answer the second one correctly

would mean that both questions as answered by the viewer were wrong thus indicating effective performance by the filter as deployed in masking the gender of the person in the image.

The criterion of Consistency was examined through questions designed to assess the ability of the viewers to track privacy-filtered persons/objects as well as measure the stability of the privacy-filtering effects across the video frames of the dataset and from different viewpoints. On the other hand, the Disambiguity criterion was assessed through questions seeking to examine if the viewers were able to distinguish multiple persons and identify their distinctive accessories and (re)appearances. A final set of questions were incorporated to evaluate whether the filtered video-frames still retained sufficient informative-ness to support the expected surveillance purposes in terms of the information preserved by the privacy filter and whether the viewers could still discern the main event shown in the video and how these were initiated.

4.2. Objective metrics

In this section we described the techniques used to produce the objective metrics followed by the criteria mapping strategies. Figure 2 illustrates an overview of the proposed objective evaluation for video privacy protection solutions.

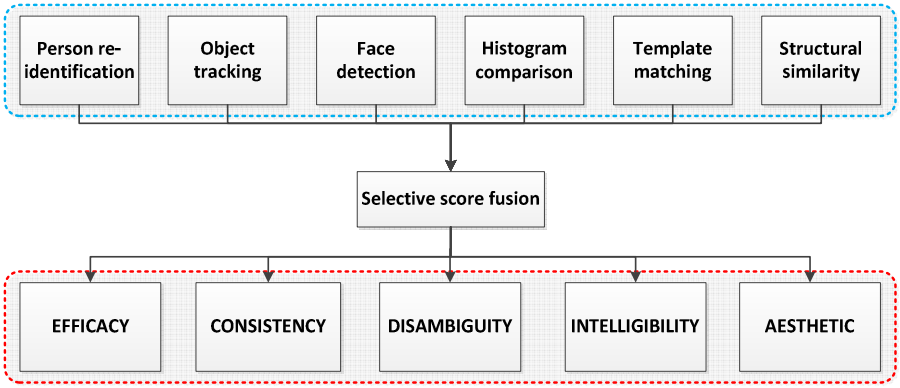


Figure 2: UI-REF-based objective privacy filters evaluation toolkit

4.2.1. Person re-identification

Originally, person re-identification referred to the process of determining whether a given detected or tracked individual has already appeared in a different image or camera view. This concept was exploited to assist in the evaluation of the proposed evaluation criteria using an appearance-based model with a supervised learning approach. In order to build a discriminative object model, a pyramid dense feature detector was implemented to provide a sufficient number of key points at different scales followed by a SURF descriptor [12] which is known to be invariant to the pose, viewpoint and illumination changes. A bag-of-words representation [17] was then generated from the extracted descriptors to form a holistic model of the object appearance. For the machine learning of the resulting model, we used a supervised SVM classifier with a non-linear Radial Basis Function (RBF) kernel.

For each person, we first created a training dataset using selected frames featuring the same subject from different viewpoints as extracted from PEViD training set. Thereafter we built an object classifier using the original unfiltered data and a second classifier based on the respective filtered frames.

To evaluate the privacy-protected videos, three (3) main testing strategies were used: i) query the image of a filtered person against the unfiltered classifier, ii) query the image of an unfiltered person against the filtered classifier, and iii) query the image of a filtered person against the filtered classifier. The Anonymity score was obtained by applying the first (i) and the second (ii) strategies using a one-vs.-one scheme (i.e. matched against a single object classifier). The Consistency score was produced through the third strategy also using a one-vs.-one scheme; whereas, the F-score (Eq.1) was computed using the third strategy with a one-vs.-all scheme (i.e. matched against all objects classifiers); as:

$$F - score = 2 \cdot \frac{precision\,.\,recall}{precision + recall} \qquad \text{Eq.1}$$

A high Anonymity score would indicate a higher Efficacy privacy protection afforded by the privacy filtering techniques as deployed in each case. In the case of a re-identification Consistency score, a high score indicated that the filtered video still carried sufficient information to enable an observer to perform tasks such as person tracking across images from the CCTV network without finding out the person's identity. Finally, a higher F-score indicated a higher Disambiguity capability arising from the deployment of the filtering technique whose Efficacy was being thus evaluated.

4.2.2. Object tracking

The human detection and tracking based on the Histogram of Oriented Gradient HOG [13] was deployed for object tracking. The ability to successfully detect humans after the application of a privacy filter signifies that the resulting video could still carry sufficient visible clues to serve the surveillance purposes of video analytics including object tracking. The baseline was defined as the detection and tracking rate when performed on the original unfiltered videos. The final Trackability score was defined in (Eq.2) as a minimum and maximum area ratio of the overlapped region between the ground truth bounding box (i.e. the one detected in the original frame) and the one detected in the filtered frame.

$$Trackability = \frac{\min(A_{overlap}, A_{GT})}{\max(A_{overlap}, A_{GT})} \qquad \text{Eq.2}$$

Where:

$A_{overlap} = area(detected \cap ground\ truth)$
$and\ A_{GT} = area(ground\ truth)$

The calculated Trackability score was compared against the baseline following the procedure as described in Algorithm 1. The final value of this metric was used in the Disambiguity as well as the Intelligibility criteria.

Algorithm 1

```
while frame do
  perform detection on unfiltered video
  if detected then
    calculate metric* on filtered video
  end if
end while
```

* Trackability or FaceDetection

4.2.3. Face detection

The Viola-Jones cascade classifier [14] was trained to detect faces and was then performed on the privacy-protected videos. Ideally, no faces should be found, since they all should be obscured. The faces found by the face detection algorithm were matched against the ground truth following the same procedure as in Algorithm 1 to avoid taking into account any false positives as may be output by the detection algorithm. The FaceDetection score was calculated based on (Eq.3).

$$FaceDetection = \frac{\min(A_{overlap}, A_{GT})}{\max(A_{overlap}, A_{GT})} \quad \text{Eq.3}$$

Where: $A_{overlap} = area(detected \cap ground\ truth)$

and $A_{GT} = area(ground\ truth)$

A higher FaceDetection rate implies higher Intelligibility, whereas a lower detection rate would be consistent with a higher level of Efficacy in privacy protection.

3.2.4. Histogram comparison

The 2D Hue-Saturation histograms of the filtered and unfiltered face region were computed. The Correlation metric of the two histograms was calculated to produce numerical parameters that expressed how well the two histograms H1 and H2 matched each other following (Eq.4).

$$Correlation(H_1, H_2) = \frac{\sum_I \left(H_1(I) - \overline{H_1}(I)\right)\left(H_2(I) - \overline{H_2}(I)\right)}{\sqrt{\sum_I \left(H_1(I) - \overline{H_1}(I)\right)^2 \left(H_2(I) - \overline{H_2}(I)\right)^2}} \quad \text{Eq.4}$$

Whereby:

$H1 = Filtered\ Face\ Histogram$

and $H2 = Original\ Face\ Histogram$

The higher the metric, the more accurate the matching scores. This metric was used to evaluate the Aesthetics after-effects arising from the deployment of a given privacy filter.

4.2.5. Template matching

This technique was exploited to find the areas of an image that could be matched to a template image. A cropped image of an unfiltered face was used as a template *T(x',y')* and matched by sliding it over the corresponding filtered frame *I(x,y)*. Using the squared difference method (Eq.5), a 2D result matrix *R* was generated with each value standing for a match metric.

$$R(x,y) = \sum_{x',y'} \left(T(x',y') - I(x+x', y+y')\right)^2 \quad \text{Eq.5}$$

The best match could be found as global minimum in the *R* matrix which represented the corner of the candidate detection bounding box *C* with dimensions equivalent to the template image *T*. The Matching score was later given by the area of the intersection between the detection *C* and template *T* bounding boxes divided by the template area. The Matching score was obtained using (Eq.6).

$$Matching = \frac{area(C \cap T)}{area(T)} \quad \text{Eq.6}$$

Ideally, after applying a privacy protection filter there should be no matching between *before* and *after* effects frames and this could be reflected on the Efficacy, Intelligibility, and Aesthetics criteria.

4.2.6. Structural similarity (SSIM)

The SSIM index [15] is designed to measure the structural similarity and the alteration between two images. In a full reference fashion, the image quality of the filtered video frame was measured based on the corresponding unfiltered copy as a reference. SSIM is considered as an improvement on the traditional peak signal-to-noise ratio (PSNR); in terms of compatibility with the human visual perception system. A successful privacy filtering system should have a minimal impact on the global quality of the image with modifications occurring only on the privacy-sensitive areas which should be anonymised. This metric was used to evaluate the Aesthetics criterion.

4.2.7. Selective fusion

The final scores for the evaluation criteria resulting from the objective metrics were calculated by selectively combining the most relevant subset of the metrics as mapped in Figure 3. Regarding the Efficacy, we were interested in how effectively the privacy filters performed in obscuring the privacy-sensitive information. Accordingly, we proposed to combine the FaceDetection rate in addition to the anonymity score of the re-identification system as well as the template matching output as given in the following equation.

$$Efficacy = \ Avg[(1 - FaceDetection) + Anonymity + (1 - Matching)] \quad \text{Eq.7}$$

The re-identification score of Consistency was sufficient to judge the Consistency criteria of the privacy-protected video-frames.

$$Consistency = Consistency\ (reid) \quad \text{Eq.8}$$

To obtain the overall Disambiguity score of the examined filter, an average combination of the object tracking performance and the F-score measure for the person re-identification system was used; as follows.

$$Disambiguity = \ Avg[Trackability + \ F - Score] \quad \text{Eq.9}$$

The Intelligibility of the privacy filter was inferred from the scores as previously computed for the face detector, the object tracker, and, the template matching metrics; as follows.

$$Intelligibility = \ Avg[FaceDetection + Trackability + Matching \quad \text{Eq.10}$$

Finally, the level of the aesthetic appeal of the visual effects arising from the application of a privacy filter was determined based on the correlation of the histogram of *before* and *after* effects combined with the template matching and SSIM index.

$$Aesthetics = \ Avg[Correlation + Matching + SSIM] \quad \text{Eq.11}$$

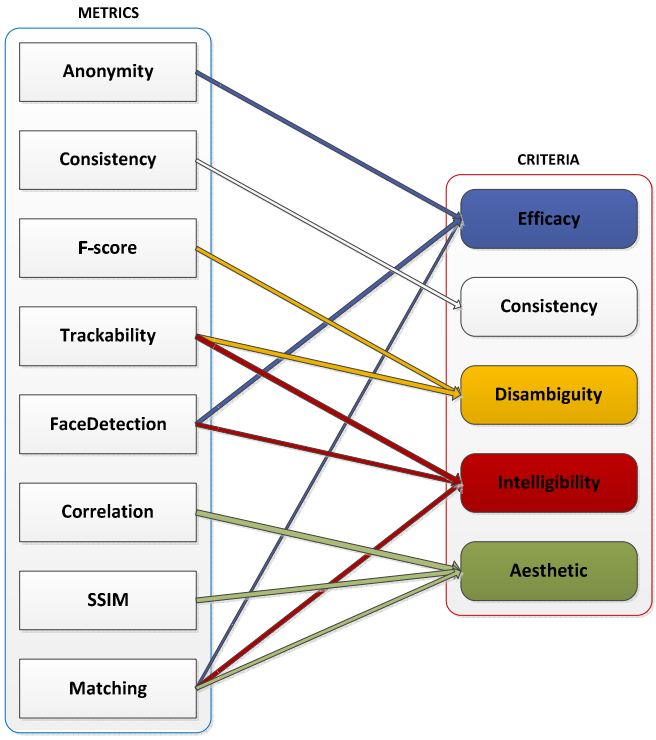


Figure 3: UI-REF-based privacy filters performance evaluation metrics mapping

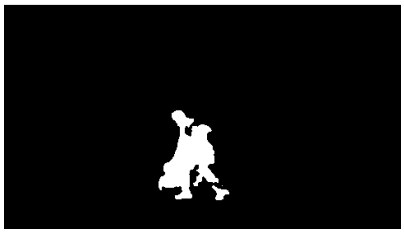
5. EXPERIMENTS AND RESULTS

In this section we discuss the procedure for conducting the survey, and, the selection of videos from the dataset for the testing and the simulation of privacy filters to which the objective and subjective evaluation processes were consistently applied to arrive at the results that have are reported in this paper.

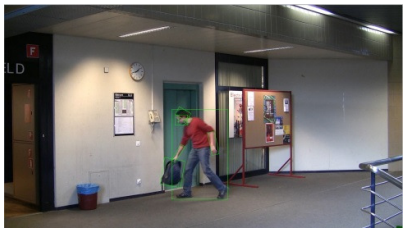
5.1. Dataset

The PEViD dataset [11] was specifically created for impact assessment of the privacy protection technologies. The dataset consists of two subsets (training and testing) of videos collected from a range of standard and high resolution cameras and contains clips of different scenarios showing one or several persons walking or interacting in front of the cameras. The actors may also carry specific items which could potentially reveal their identity and may therefore need to be filtered appropriately. The actors are featured carrying backpacks, umbrellas, wearing scarves, and, can be seen fighting, pickpocketing or simply walking around. Actors may be at a distance from the camera or near the camera, making their faces vary considerably in pixel size and quality. The ambient lighting conditions of the videos vary widely as half of the clips were recorded at night. The used dataset contained (20) video clips and associated ground truth in xml format as well as a foreground mask. The ground truth consisted of annotations of persons, faces, skin regions, and personal accessories. The videos included indoor, outdoor, day-time and night-time environments, showing people interacting or performing various actions. The video-clips were made available in MPEG format with a resolution of 1920x1080 pixels at a rate of (25) frames per second. Figure 4 depicts a sample frame from the described dataset.

(a)



(b)



(c)

Figure 4: Sample data from the PEViD dataset: (a) original frame, (b) foreground mask, (c) annotated frame

5.2. Filters

In order to validate the framework, several filters which covered a range of image manipulation techniques for privacy protection were implemented and evaluated; as follows:

- **Disc**: variable size discs were drawn on top of the original image on a regular grid. The colour of each disc was the colour of the pixel underneath its centre; Figure 5, top middle.
- **Median**: Median blur was applied to the image; Figure 5, top right.
- **Pixelate**: The image was simply down-sampled and up-sampled back to its original size with a nearest neighbour interpolation; Figure 5, bottom left.
- **Resample**: The resample method which first down-sampled the image before up-sampling it back to its original size using the Lanczos re-sampling method [16]; Figure 5, bottom middle.
- **Scramble-blur**: in the first pass, we computed the DCT transform of each 8x8 block and switched the sign of 4 coefficients selected randomly within the first 5 DC coefficients of any colour channel; in the second pass, we applied a median blur; Figure 5, bottom right.

Figure 5: The privacy filters considered: Original image, Disc, Median, Pixelate, Resample, and Scramble-blur

5.3. Survey procedure

Ten (10) male adults with normal vision participated in this survey. Participants' ages ranged between (21-35) years old. None of the participants had seen the test data previously or knew the actors. To prepare the data, each of the above described filters (in section 5.2) was applied to five different video-clips from the PEViD dataset representing different scenarios. The privacy filter aimed to cover the region of the video frame where a person was located. Each participant was asked to view five videos which were filtered using two different filters. The participant was required to answer a separate set of questions for each video. The questionnaire had been carefully designed to probe for the content of the examined videos and was focused on the proposed evaluation criteria. The questions had been prioritised by considering the precedence and sequentiality effects due to any incremental observations that might arising from the ordering of the questions and the staging of the overall subjective evaluation. The participants performed their evaluations alone and independently.

5.4. Simulation results

This section reports and discusses the results obtained from the objective and the subjective evaluation and the corresponding validation process.

5.4.1. Objective evaluation

The five types of filters as described in the previous section were evaluated using the proposed objective metrics separately. The filters were applied to the region of the video frame where a person was located corresponding to the foreground mask. The average scores of each filter as a response to the metrics performed on the PEViD dataset are listed in Table 1.

	Disc	*Median*	*Pixelate*	*Resample*	*Scramble*
FaceDetection	*0.05*	*0.26*	*0.41*	*0.16*	*0.06*
Anonymity	*0.49*	*0.46*	*0.46*	*0.48*	*0.49*
Consistency	*0.57*	*0.58*	*0.56*	*0.54*	*0.54*
F-Score	*0.54*	*0.55*	*0.59*	*0.58*	*0.53*
Trackability	*0.15*	*0.63*	*0.46*	*0.55*	*0.34*
Correlation	*0.65*	*0.64*	*0.65*	*0.65*	*0.64*
Matching	*0.59*	*0.96*	*0.96*	*0.90*	*0.52*
SSIM	*0.74*	*0.86*	*0.82*	*0.79*	*0.68*

Table 1: Average metrics score of the tested privacy filters

A comparison of the filters performance is shown in Figure6. The overall scores have been normalised and represented such that the higher the score shown, the better the performance of the privacy filter; except for the FaceDetection in which no detection is the most desirable outcome.

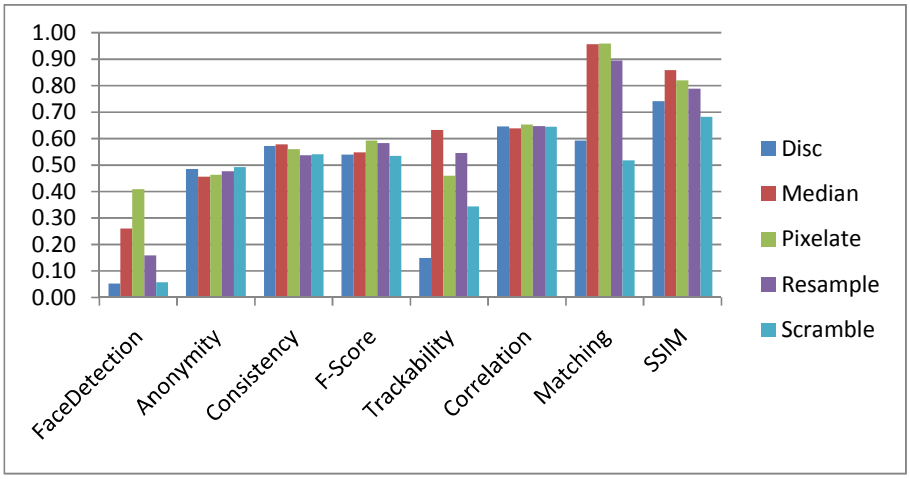


Figure 6: Privacy filter performance based on different quality metrics

The FaceDetection metric was applied to the privacy filtered videos in order to examine if it was capable of detecting the faces after the filtering effect had been applied. Ideally no faces should be detected. In this test, Pixelate had the highest success rate as it preserved the spatial arrangement of the facial features. The Median filter was ranked next due to the fact that it only blurred the face region with some detail still visible.

Although Resample shared the same approach as Pixelate, the use of the Lanczos method instead of linear up-sampling evidently led to a smoother filter output and thus made FaceDetection more challenging. Finally, the deployment Disc and the Scramble caused the highest failure rate for the face detector due to significant image deformation in the case of Disc and spatial re-arrangement as a result of the Scramble filter.

The re-identification metrics namely, Anonymity, Consistency, and F-score provided similar scores for the tested filters which ranged around mid-point. This was due to the fact that our re-identification scheme relied heavily on colours, which were not changed significantly by any of the filters.

Object tracking using HOG was relatively successful in tracking the Median filtered person as it still carried sufficient appearance information for object tracking. The Resample and Pixelate filters had a lower performance than the Median while the Disc and Scramble filters had the lowest trackability scores as would be expected given their level of image modifications.
The histogram correlation metric scores were similar for all filter types; which could be expected due to the nature of the filters in use.

On the other hand, Median and Pixelate filters achieved an equally high score for the template matching metric followed by Resample. The score for Disc and Scramble was significantly lower than the rest of the filters.

With respect to the Structural Similarity (SSIM) metric, Disc and Scramble approaches to filtering scored slightly lower than the other techniques. This was due to their trackability score

being the lowest. However, all the tested filters produced relatively high scores indicating an acceptable filtering effect and moderate level of modification had been achieved in comparison to more severe obscuring methods such as a solid coloured mask or virtual object placement.

Based on the above, we can conclude that the stand alone individual scores of the metrics are not sufficiently informative. However, the selective fused scores as proposed to meet the evaluation criteria proved to be an efficient method with more meaningful measures for critical and comparative analysis of the privacy-protective performance of alternative privacy filtering solutions.

Table 2 summarises the overall merit score for the evaluated filters as calculated using the selective fusion methods as described in section4.2.7. A comparison of the performance of the evaluated filters in terms of evaluation criteria is shown in Figure7.

	Disc	***Median***	***Pixelate***	***Resample***	***Scramble***
Efficacy	*0.32*	*0.44*	*0.39*	*0.49*	*0.64*
Consistency	*0.57*	*0.58*	*0.56*	*0.54*	*0.54*
Intelligibility	*0.27*	*0.62*	*0.61*	*0.53*	*0.31*
Aesthetics	*0.66*	*0.82*	*0.81*	*0.78*	*0.61*
Disambiguity	*0.34*	*0.59*	*0.53*	*0.56*	*0.44*
SCORE	*0.39*	*0.49*	*0.50*	*0.44*	*0.40*

Table 2: Overall evaluation score of the objectively tested privacy filters

The effectiveness of the tested filters was generally low suggesting the need to explore more advanced filtering methods. The only exception was the Scramble method as it achieved the highest level of image modification.

Overall, the filters were assessed to be of average value for the Consistency criterion; this could allow a reasonable tracking performance. Comparatively, the filters with the highest score for the Intelligibility criterion were the Median and the Pixelate closely followed by the Resample filter while the Disc and Scramble scored significantly lower.

Aesthetically, all the filters scored above the average; although the Median outscored the rest of filters. The Disc filter scored the lowest for the Disambiguity criterion followed by the Scramble filter, whereas the Median, Pixelate, and Resample were performed above the mid-point.

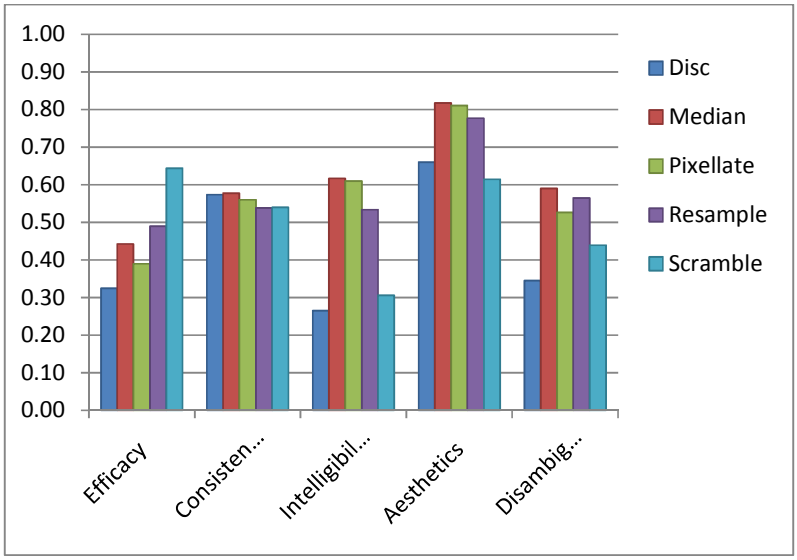


Figure 7: Comparison of the objective performance of the privacy filters in terms of the evaluation criteria

The final scores of the tested filters in terms of the evaluation criteria demonstrated the effectiveness of the proposed UI-REF-based framework to critically and objectively evaluate privacy-protecting filters. As expected, the simplicity of the evaluated filters has led to a below average overall score emphasising the need for more effective filters.

5.4.2. Subjective evaluation

To ensure that the subjective evaluation was conducted systematically, each volunteer participating in the subjective evaluation was provided with the same set and number of video-clips as had been privacy filtered using each of the same set of 5 privacy Filters. In this way the evaluators all received identical sets. Figure 8 depicts the averaged responses of the participants to score against each of the evaluation criteria. In terms of Efficacy, Scramble and Disc filters outperformed the other filters as most of the participants found them more effective in hiding the features of the person's appearance. The Consistency values were relatively similar amongst the tested filters; this showed the same trend as the results for the equivalent objective evaluation.

Expectedly, the score for the Intelligibility criterion was significantly higher for all the filters in the subjective evaluation as compared with the objective measures; this highlighted the difference between the human visual perception and the computer vision performance. In terms of Aesthetic appeal, only the subjective results for the Pixilate filter were found to be inconsistent with the corresponding results from the objective evaluation; this was considered to be due to the fact that the Pixelate filters box up and block off the filtered areas. Although the average Disambiguity scores from the subjective tests were higher than the ones arrived at through the objective evaluation, the two sets of results indicated the same trend.

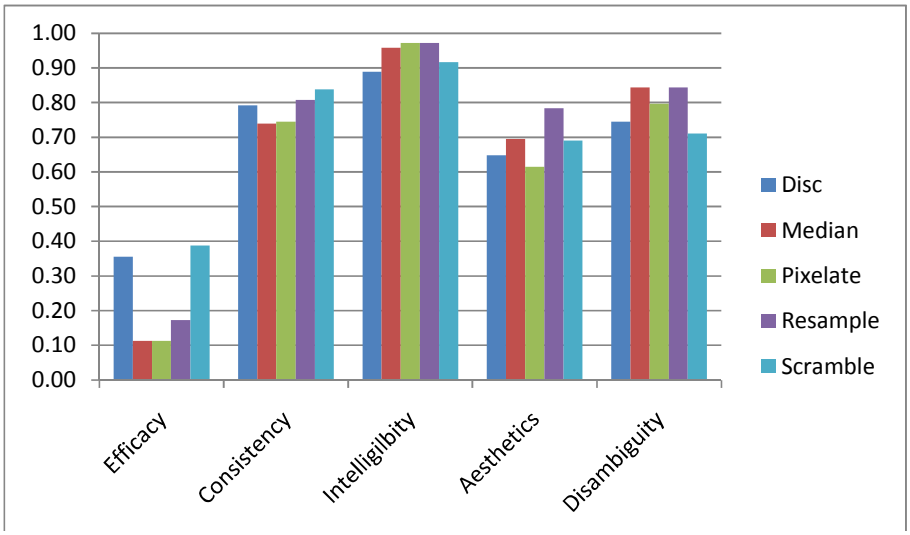


Figure 8: Comparison of the subjective performance of privacy filters in terms of evaluation criteria

Table 3 summarises the averaged values of all the responses obtained from the questionnaire. The overall scores were noticeability higher than the corresponding scores from the objective evaluation.

	Disc	*Median*	*Pixelate*	*Resample*	*Scramble*
Efficacy	*0.36*	*0.11*	*0.11*	*0.17*	*0.39*
Consistency	*0.79*	*0.74*	*0.74*	*0.81*	*0.84*
Intelligibility	*0.89*	*0.96*	*0.97*	*0.97*	*0.92*
Aesthetics	*0.65*	*0.70*	*0.61*	*0.78*	*0.69*
Disambiguity	*0.74*	*0.84*	*0.80*	*0.84*	*0.71*
SCORE	*0.69*	*0.67*	*0.65*	*0.72*	*0.71*

Table 3: Overall evaluation score of the subjectively tested privacy filters

6. CONCLUSION

This paper reports the results of a research study which has applied the UI-REF Requirements Prioritisation and Holistic system-of-systems-scale Impact Assessment (H-PIA) Methodology for Privacy-by Co-design. This has been supported by integrated high resolution objective and subjective evaluation of the performance and impacts of various privacy filtering techniques. We have outlined the UI-REF ontological commitment to the negotiation of situated context-content-purpose and the analysis of surveillance solution as underpinned by the UI-REF decisional framework. This is to support the socio-ethically reflective and normatively adaptive category judgments and attributions, e.g. as to the *Suspect-ness* of an individual whose image has been captured in some surveillance video dataset. As a precursor to subsequent studies that would incorporate additional criteria applied to a wider set of contexts (as set out in Badii 2013 [5]) we have implemented a sub-set of the UI-REF-based Holistic Privacy Impact Assessment (H-PIA) metrics as a set of five evaluation criteria that enable the integrative objective and subjective evaluation of multi-modal multimedia privacy protection solutions. Accordingly a number of filtering techniques have been evaluated and the simulation results have demonstrated the consistency and efficiency of the framework. The scores of the implemented objective metrics have been mapped to reflect each evaluation criterion as well as the subjective findings. Future

work could usefully apply the framework to more video analytics algorithms in order to evaluate the videos at a finer level of granularity. Furthermore, the fusion of results with low-level metrics could be enhanced by introducing a weighted selective fusion strategy and experimenting with an extended set of privacy filters. This work has been motivated by the commitment to provide rigorous benchmarking mechanisms to support Privacy Impact Assessment as an integral process within evolutionary socio-ethical co-design of surveillance systems. The full engagement of citizens in holistic societal impact assessment of privacy failure risks is likely to remain an unattainable ideal unless the interlocutors of the Privacy-by- Co-Design negotiations are symmetrically empowered with technological enablers to make transparent to *all* just how citizens' personal data are handled and used in what contexts by whom and for what purpose. This must include mechanisms for robust forensic assessment of the performance efficacy of the proposed privacy protection solutions in responding to citizens' most deeply valued needs in each of the evolving multi-level multi-modal situated contexts of their lifestyle as can be dynamically (re)defined and (re) negotiated by them with all implicated stakeholders. The proposed UI-REF-based methodological framework is accordingly fully equipped to support such operationalisation of Privacy-by-Co-design to break away from the abundant rhetoric of the oft-avowed privacy-caring aspirations and idealism of the past towards enablers to support the full practical realisation of an inclusivist Privacy-by-Co-Design framework enriched by collectively actionable engagement, and reflective practice as empowered by transparent accountability.

ACKNOWLEDGEMENTS

This work was supported by the European Commission under contracts FP7-261743 VideoSense project.

REFERENCES

[1] Badii A, "User-Intimate Requirements Hierarchy Resolution Framework (UI-REF): Methodology for Capturing Ambient Assisted Living Needs", Proceedings of the Research Workshop, Int. Ambient Intelligence Systems Conference (AmI'08), Nuremberg, Germany November 2008

[2] Senior, Andrew, et al. "Blinkering surveillance: Enabling video privacy through computer vision." IBM Technical Paper, RC22886 (W0308-109) (2003).

[3] Senior, Andrew. "Privacy enablement in a surveillance system." Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on. IEEE, 2008.

[4] Badii, A., Einig, M., Tiemann, M., Thiemert, D. and Lallah, C. (2012) *Visual context identification for privacy-respecting video analytics.* In: IEEE 14th International Workshop on Multimedia Signal Processing (MMSP 2012), 17-19 Sep 2012, Banff, Canada, pp. 366-371.

[5] Badii, A, Framework for Requirements Prioritisation, Usability Evaluation and Holistic Impact Assessment of Privacy-preserving Video Analytics by co-Design, Working Paper UoR-ISR-VS-2013-4, September 2013.

[6] Dufaux, Frederic, and Ebrahimi, Touradj. "Scrambling for privacy protection in video surveillance systems." Circuits and Systems for Video Technology, IEEE Transactions on 18.8 (2008): 1168-1174

[7] Boyle, Michael, Christopher Edwards, and Saul Greenberg. "The effects of filtered video on awareness and privacy." Proceedings of the 2000 ACM conference on Computer supported cooperative work. ACM, 2000

[8] Dufaux, Frederic. "Video scrambling for privacy protection in video surveillance: recent results and validation framework." SPIE Defense, Security, and Sensing. International Society for Optics and Photonics, 2011.

[9] Zhao, Qiang Alex, and John T. Stasko. "Evaluating image filtering based techniques in media space applications." Proceedings of the 1998 ACM conference on Computer supported cooperative work. ACM, 1998.

[10] Boyle, Michael, Christopher Edwards, and Saul Greenberg. "The effects of filtered video on awareness and privacy." Proceedings of the 2000 ACM conference on Computer supported cooperative work. ACM, 2000.

[11] Korshunov, Pavel, and Ebrahimi, Touradj. "PEViD: privacy evaluation video dataset". Applications of Digital Image Processing XXXVI, San Diego, California, USA, August 25-29, 2013.

[12] Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. "Surf: Speeded up robust features." Computer Vision–ECCV 2006. Springer Berlin Heidelberg, 2006. 404-417.

[13] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE, 2005.

[14] Viola, Paul, and Michael J. Jones. "Robust real-time face detection." International journal of computer vision 57.2 (2004): 137-154.

[15] Wang, Zhou, et al. "Image quality assessment: From error visibility to structural similarity." Image Processing, IEEE Transactions on 13.4 (2004): 600-612.

[16] Turkowski, Ken. "Filters for common resampling tasks." Graphics gems. Academic Press Professional, Inc., 1990.

[17] Fei-Fei, Li, and Pietro Perona. "A bayesian hierarchical model for learning natural scene categories." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 2. IEEE, 2005.

Authors

Atta Badii

Senior Professor, Secure Pervasive Technologies, Founding Director of Intelligent Systems Research Laboratory (www.isr.reading.ac.uk), the University of Reading, School of Systems Engineering and Distinguished Professor of Systems Engineering and Digital Innovation, University of Cordoba., Atta has a multi-disciplinary academic and industrial research experience in the fields of Distributed Intelligent and Multi-modal Interactive Systems, Pattern Recognition, Security, Trust and Privacy-Preserving Architectures and Semantic Media Technologies. Atta's research stands at the confluence of intelligent interactive systems, and, human agent modelling as informed by the well-established principles of psycho-cognitively-based requirements elicitation, user-centred prioritisation, dynamic usability-relationship-based evaluation and co-design and innovation of socially responsible systems-of-systems; He has pioneered such approaches since 1997 and the UI-REF-based application of context-aware video privacy protection and its evaluation since 2011.

Ahmed Al-Obaidi

Ahmed Al-Obaidi received his BSc degree in computer engineering from the University of Mosul in 2002 and the MSc degree in computer systems engineering from Putra University in 2008. His R&D career began at MIMOS Bhd., Malaysia in 2007 where he was involved in projects for intelligent video surveillance. Ahmed joined the ISR Laboratory at the University of Reading, in April 2013 within the Joint Research Activity Programme of the European Centre for Video-Analytics Research (VideoSense) led by Professor Badii. Ahmed's research interests include computer vision, machine learning, and context-aware video privacy protection.

Mathieu Einig

In 2008 Mathieu Einig completed his BSc in Computer Science from Université "A" Paul Sabatier, Toulouse, France. This achievement was followed by Mathieu's BSc (Hon) in Computer Graphics Science, Teesside University, 2010. Mathieu's research work at ISR focused on Semantic Media Technology projects as an accomplished software engineer with keen research interests in Computer Vision, Computer Graphics and Augmented Reality.

Aurélien Ducournau

In 2009, Aurelien Ducournau received his MSc in Computer Science and Image Analysis with Honors from Universite de Caen Basse-Normandie, France; followed by a 6-months research internship at the Telecom ParisTech Paris, France and a PhD in 2012 (with Distinction) in Computer Science, Signal, Image and Vision at the National Engineering School of Saint-Etienne, France. His research work at ISR, University of Reading mainly focused on Computer Vision, Image and Video Processing and Analysis, Image Segmentation and Graph/Hypergraph theory.