Overview of the MediaEval 2014 Visual Privacy Task

Atta Badii[1], Touradj Ebrahimi[2], Christian Fedorczak[3], Pavel Korshunov[4],
Tomas Piatrik[5], Volker Eiselein[6], Ahmed Al-Obaidi[7]

1 & 7 {atta.badii, a.al-obaidi}@reading.ac.uk
2 & 4 {touradj.ebrahimi, pavel.korshunov}@epfl.ch
3 christian.fedorczak@thalesgroup.com
5 t.piatrik@qmul.ac.uk
6 eiselein@nue.tu-berlin.de

ABSTRACT

This paper presents an overview of the Visual Privacy Task (VPT) of MediaEval 2014, its objectives, related dataset, and evaluation approaches. Participants in this task were required to implement a privacy filter or a combination of filters to protect various personal information regions in video sequences as provided. The challenge was to achieve an adequate balance between the degree of privacy protection, intelligibility (how much useful information is retained post privacy filtering), and pleasantness (how minimal were the adverse effects of filtering on the appearance of the video frames). The submissions from the eight (8) teams who participated in this task were evaluated subjectively by surveillance experts, practitioners, data protection experts and by naïve viewers using a crowdsourcing approach.

1. INTRODUCTION

Advances in artificial intelligence and video surveillance have led to increasingly complex surveillance systems of rising scale and capabilities. The ubiquity and enhanced capability of such surveillance can pose significant threats to citizens' privacy and therefore new mitigation technologies are needed to ensure an appropriate level of privacy protection. The Visual Privacy Task (VPT) of MediaEval 2014 thus provided an opportunity for experimentation to explore how video-analytic techniques may arrive at enhanced solutions to some visual privacy problems [1]. This task focuses on privacy protection techniques that are responsive to the context-specific needs of persons for privacy. The evaluation was performed using three distinct user studies aimed at developing a deeper understanding of users' perceptions of the *effects* and *side-effects* of privacy filtering to ensure the validity and user-acceptability of the evaluation results.

2. VPT 2014 DATASET

The PEViD dataset [2] was specifically created for impact assessment of the privacy protection technologies. The dataset consists of two subsets, namely the training and testing sets; comprising of (21) videos as captured with both standard and high resolution cameras. The video clips are in MPEG format in full HD resolution of (1920x1080) pixels at a rate of (25) frames per second and approximately (16) second each.

The video data includes various scenarios featuring one or several human subjects walking or interacting. The actors may also carry specific items, which could potentially reveal their identity and may therefore need to be privacy-filtered appropriately. For example, the actors are featured carrying backpacks, umbrellas, wearing scarves, and performing various actions, such as fighting, pickpocketing, dropping-a-bag, or simply walking. Actors may be at a distance from the camera or near the camera, making their faces appear with varying pixel size and quality. The ambient lighting conditions of the videos also varied widely as they recorded a range of indoors, outdoors, day/night-time scenes. The ground truth was created manually by the task organisers and consisted of annotations of the bounding boxes containing the regions of High (H), Medium (M), or Low (L) Personally Identifiable Information elements (PIIs) including persons' faces and accessories. In order to simulate context–aware privacy protection solutions [3], unusual events occurring within the video datatset, such as fighting, stealing and dropping-a-bag were also annotated. The annotations were provided in XML format alongside a foreground mask in the form of binary sequences. These included such annotations that distinguished the relative privacy sensitivity of PIIs; namely for Skin (M), Face (H), Hair (L), Accessories (M), and for Person's body (L). The dataset was provided in accordance with the European Data Protection and ethical compliance guidelines including *informed consent* and *access control* as required. Figure 1 depicts a sample frame from the dataset with annotated regions as rectangles.

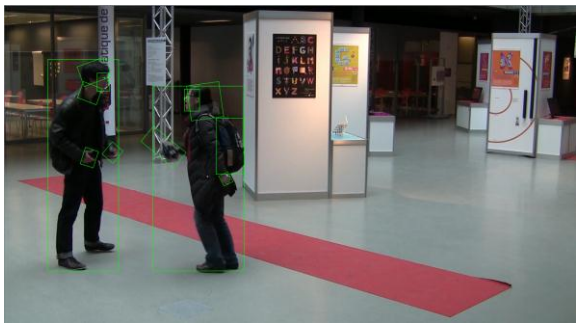

Figure 1: Sample annotated frame from the VPT Dataset

3. MOTIVATION AND OBJECTIVES

The MediaEval 2014 Visual Privacy Task was motivated by application domains such as video privacy filtering of videos taken in public spaces, by smart phones, web-cams, surveillance CCTVs, and, videos stored in social websites. For this task, the participants were encouraged to implement a combination of several privacy filters to protect various personal information regions in videos, by optimising the privacy filtering so as to: **i)** obscure such personal information effectively whilst, **ii)** keeping as much as possible of the 'useful' information that would enable a human viewer to form some 'useful' interpretation of the obscured video frame at some level of abstraction without compromising the privacy protection level as required by the person(s) featured in the video-frame. Personal visual information is subjective human-perceived information that can expose a person's identity to a human viewer. This can include richly detailed image regions such as distinctive facial features or personal jewellery as well as less rich uniform regions e.g. skin

MediaEval 2014 Workshop, October 16-17, 2014, Barcelona, Spain

regions (that expose racial identity) or body silhouette showing a person's gait (that generally helps to differentiate women from men and in some cases it may even enable a close friend or a spouse to identify the person). Therefore, to satisfy both of the above-mentioned criteria, i), and, ii) above, privacy protection solutions were required to take into account different types of visual personal information. . The participants were encouraged to exploit the annotation information to achieve the appropriate level of privacy filtering for each person, object, and Low/Medium/High information regions and accordingly select the best-fit filtering. It was anticipated that a single privacy filter applied to all parts of an image would result in a sub-optimal solution and a combination of several privacy filters would provide more effective filtering.

4. SUBMISSIONS EVALUATIONS

The submitted video clips were methodologically evaluated using UI-REF based privacy protection requirements. Accordingly the evaluations attempted to assess the perceived *efficacy*, as well as *side-effects* and *affects* arising from a proposed privacy filtering solution -as described in [4,5]. Eight (8) research teams submitted privacy filtered video sequences for the evaluation. In the context of surveillance scenarios, three distinct user studies were conducted to ensure the validity of the evaluation results. The subjective evaluations comprised:

a) Stream 1: crowdsourcing evaluations by the general public from online communities ("naïve subjects") in accordance with the methodology in [6];

b) Stream 2: subjective evaluations by security system manufacturers and video-analytics technology and privacy protection solutions developers;

c) Stream 3: online subjective evaluations by a focus group comprising trained CCTV monitoring professionals, and law enforcement personnel.

For consistency in the analysis of evaluation results from all streams for all participants' solutions, the same six (6) video clips were pre-selected from each submission and evaluated using the three (3) evaluation streams. A questionnaire consisting of 12 questions had been carefully designed to examine aspects related to privacy, intelligibility, and pleasantness; this was used in stream 2 and 3. The first (5) questions were aimed at eliciting the opinions of the evaluators re the Contents of the viewed videos. The responses to these questions were considered with respect to the ground truth. The rest of the questions were aimed at eliciting the Subjective Opinions of the evaluators re the viewed videos.

Stream 1 used a shortened version of the questionnaire with (7) questions in total due to crowdsourcing constraints. Some 290 workers responded to the crowd-sourced evaluations. In the design of the crowdsourcing campaign, special care was taken so that a worker would not see the same content with different filters (only one filter per content) and would not see different contents with the same filter (only one content per filter). Also, only the answers from reliable crowd-sourcing workers were taken into account. The reliability was ensured via honeypots, mean and deviation metrics of time per response to a question, and total time per campaign. Out of the total 290 workers, 230 were found to have provided reliable responses to all the 8 evaluation batches, which resulted in 230/8=29 sets of workers' evaluations for each filter submitted by each participant.

In Stream 2 evaluations, the focus group consisted of (65) participants, (15) of them were females; staff from Thales, France took part in this evaluation. The majority of the participants were from the R&D departments, while the rest were from Management, Security, and other departments. The submissions were evaluated via paper-based responses to the questions.

In Stream 3 evaluations, the focus group comprised of (59) participants including (22) females. This group included some key stakeholder types such as people from R&D, data protection, and law enforcement, who took part in this study from around the world. The participants streamed the videos and answered the questionnaire using online forms. As as results of the described evaluations, VPT participants received a set of 3 by 3 matrices comprising the results of each participant for each tier of evaluation; quantified in terms of the following criteria:

1) The **Privacy Protection Level** – an average level of privacy protection across all testing video clips.

2) The **Level of Intelligibility –** the amount of 'useful' information that was *retained* in the video frames after privacy filtering had been applied.

3) The **Pleasantness** of the resulting privacy filtered video frames in terms of their 'aesthetic' perceptual appeal to human viewers.

Figure 2 depicts an overview of the results from the three (3) evaluation streams represented by the median values of the submissions for each criterion.

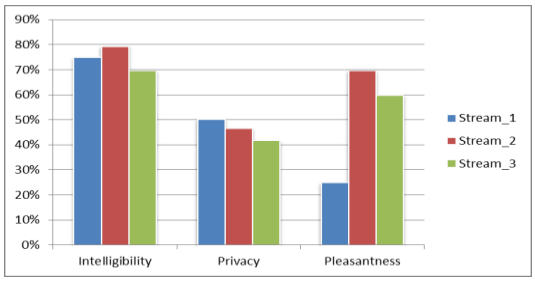

Figure 2: Median values of the results from the 3 streams

5. ACKNOWLEDGMENTS

The Visual Privacy Task at MediaEval 2014 was supported by the European Commission under contracts FP7-261743 VideoSense.

6. REFERENCES

[1] Senior, A., Pankanti, S., Hampapur, A., Brown, L., Tian, Y.L., Ekin, A., Connell, J., Shu, C.F., and Lu, M., Enabling video privacy through computer vision; IEEE Security and Privacy 3, no. 3, pp. 50-57, 2005.

[2] Korshunov, P., and Ebrahimi, T., PEViD: privacy evaluation video dataset applications of digital image processing; XXXVI, San Diego, USA, August 2013.

[3] Badii, A., Einig, M., Tiemann, M., Thiemert, D. and Lallah, C., Visual context identification for privacy-respecting video analytics; in IEEE 14th International Workshop on Multimedia Signal Processing (MMSP 2012), pp. 366-371, Banff, Canada, September 2012.

[4] Badii, A, "UI-REF Methodology", articles online.

[5] Badii, A., Al-Obaidi, A., and Einig, M., MediaEval 2013 Visual Privacy Task: Holistic Evaluation Framework for Privacy by Co-Design Impact Assessment. Proceedings of the MediaEval 2013 Workshop, Barcelona, Spain, 18-19 October 2013

[6] Korshunov, P., Nemoto, H., Skodras, A., and Ebrahimi, T. Crowdsourcing-based Evaluation of Privacy in HDR Images. SPIE Photonics Europe 2014, Optics, Photonics and Digital Technologies for Multimedia Applications, Brussels, Belgium, Proceedings of SPIE, 2014.