Overview of the MediaEval 2015 Drone Protect Task

Atta Badii[1], Pavel Koshunov[2], Hamid Oudi[3] , Touradj Ebrahimi[4] , Tomas Piatrik[5] , Volker Eiselein[6] ,

Natacha Ruchaud[7], Christian Fedorczak[8], Jean-Luc Dugelay[9], Diego Fernandez Vazquez[10]

1,3. {atta.badii, h.oudi}@reading.ac.uk, **2 , 4.** {pavel.korshunov, touradj.ebrahimi}@epfl.ch,.**5**. t.piatrik@qmul.ac.uk,

6. eiselein@nue.tu-berlin.de, **7**.Natacha.Ruchaud@eurecom.fr , **8**. christian.fedorczak@thalesgroup.com,

9. jean-luc.dugelay@eurecom.fr , **10**. dfvazquez@isdefe.es

ABSTRACT

This paper presents an overview of the Drone Protect Task (DPT) of MediaEval 2015, its objectives, related dataset, and evaluation approach. Participants in this task were required to implement a privacy filter or a combination of filters to protect various personal information regions in the video sequences provided. The challenge was to achieve an adequate balance between the degree of privacy protection, intelligibility (how much useful information is retained after privacy filtering), and pleasantness (how minimal were the adverse effects of filtering on the appearance of the video frames). The evaluation methods for this task include subjective evaluation by those working in the video surveillance sector and also by naïve viewers.

1. INTRODUCTION

The number of drones deployed for civil applications and other non-military uses such as journalism, recreation, public safety, and precision agriculture is increasing. In particular, the deployment of the highly mobile and versatile drones for aerial surveillance in urban policing and crowd management gives rise to new challenges for civil liberties, privacy and safety. 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 Drone Protect Task (VPT) of MediaEval 2015 has thus provided an opportunity for experimentation to explore how video-analytic techniques may arrive at enhanced solutions to some visual privacy problems. This task focuses on privacy protection techniques that are responsive to the context-specific needs of persons for privacy. The DroneProtect performance evaluation involves three distinct user studies aimed at developing a deeper understanding of users' perceptions of the *effects* and *side-effects* of privacy filtering for a user-centered evaluation of the privacy solutions offered.

2. DPT 2015 DATASET

The DPT dataset has provided 38 video clips of about 20 seconds each, in full HD resolution with sufficient number of examples of video images depicting different typical scenarios in a car park [3]. The bounding boxes for persons and cars are annotated. However, the detection of the face-head area as a region of interest and detecting a "person-entering-a-car" event are regarded as task to be solved as would be the cased in a real-life Car Park Security use-case scenario; this will provide an appropriate level of challenge in the present task, especially as the region-specific privacy filtering element has been previously benchmarked within the MediaEval 2014 Visual Privacy Task [1].

The video data included various scenarios featuring one or several human subjects walking and interacting with vehicles in a car park. The contents of the videos were grouped into three categories: Normal, Suspicious and Illicit behaviour. The videos in the Normal category involved subjects performing common social behaviours in the car park such as entering or leaving a car. The Suspicious category included loitering, taking a picture of parked cars and other questionable behaviours. On the other hand, Illicit behaviour included Actors stealing a car, leaving a car unattended, parking the car in forbidden areas or fighting.

The actors in the videos, carry specific items and so 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, stealing, loitering, or simply walking. Actors may be at a distance from the camera or near the camera, making their faces appear with varying size and quality. Despite the use of advanced stabilisation techniques for the Camera on board the drone, the drone maneuvering and the variable conditions outdoors still led to some jitter effects in some video segments. The ground truth data set has been created manually by the task organisers and consists of annotations of the bounding boxes containing the regions of High (H), Medium (M), or, Low (L) Personally Identifiable Information (PII) including vehicles, persons' faces and accessories, and, unusual events such as fighting, stealing and bag dropping.

The data included such annotations that distinguished the relative privacy sensitivity of PIIs; namely for License Plates(H), Skin (M), Face (H), Hair (L), Accessories (M), and for Person's body (L). The dataset has been provided in accordance with the European Data Protection and ethical compliance guidelines including informed consent and access control as required. Figure 1, shows an example of a video illustrating Illicit behaviour.

Figure 1 Sample of the video in the dataset [3].

3. AIM AND OBJECTIVES

The objective of the DroneProtect: Mini-drone Video Privacy Task is to benchmark privacy filtering solutions for drone videos

MediaEval 2015 Workshop, Sept. 14-15, 2015, Wurzen, Germany.

related to public safety. The performance of solutions is judged by their ability to retain sufficient (frame-level) semantic information about activities and situations, while at the same time providing the required level of privacy for people appearing in the videos. Task participants are to implement a combination of privacy filters to protect various personal information regions **in a set of drone videos** as **had been** provided. Privacy filtering is to be optimised for the fulfillment of both the following criteria:

- Obscure personally identifying information effectively
- Preserving **the information** needed by a human viewer in order to interpret the video at the level required to maintain security in the area monitored by the drone.

Solutions attempted to preserve the overall visual acceptability-attractiveness of the resulting privacy filtered video-frames, since these factors had potential impact on interpretability and on the quality of the work experience for humans interpreting the videos. As a secondary goal, the task aimed to investigate mixtures of reversible and irreversible privacy filters.

For this task, the use-case scenario was Car Park Security and so the typical objectives of such a scenario would determine how much of which type of information must be retained in the video to support the goal of maintaining security. The video input for the privacy filtering process consisted of drone video clips showing examples of: Persons walking, running, or fighting in the car park area, Persons attacking a driver, loitering, entering or leaving a particular car in the car park, wrongly parked cars, and collision with cyclists.

The output of the privacy filtering process was to preserve sufficient semantics for recognition of specific security-relevant events unfolding in the car park scenes whilst reversibly masking the following aspects:

- Person's face and silhouette
- Person's gender and race (note this does not entail *gender/race* recognition but rendering *un-classifiable*)
- Personal accessories
- Vehicle make and model
- Vehicle license plate (if zoomed-in on)

The face and the car body have high personal identification potential, whereas the human body outline, particularly one that has been rendered gender-unclassifiable, has a low personal identification potential. Note that gait analysis has been excluded in the formulation of the task. Accordingly all image regions as listed above needed to be masked respectively with corresponding filter strength, High (H), Low (L), Medium (M) so as to maintain the appropriate privacy protection, intelligibility and attractiveness-acceptability of the resulting privacy filtered video frame. Thus this privacy filtering task required the detection of the human face-and-head zone within each bounding box that has already delineated a person.

As a secondary goal the task invited solutions that deployed an appropriately managed mix of reversible and irreversible privacy filters. Such filters are typically optimised responsively to the context of the events and persons' behaviours occurring in the video. Such filtering must also allow the car park staff to reverse the filtering to investigate any activities as deemed possibly relevant to the investigation of any security incidents within a specific time frame as set by the regulations; e.g., within 7-30 days of any video-recording after which all videos are usually deleted.

As an additional challenge a set of 5 un-annotated videos were provided for the participants, optionally to attempt blind privacy filtering; a separate evaluation process has been planned for the results of this additional element of the task.

4. SUBMISSIONS EVALUATIONS

The ground truth consisted of video frames with annotations of the bounding boxes containing description of entities in the video images of persons and cars plus examples of alternative filtering approaches and questionnaires used by the human viewers who had evaluated them, and, the final rankings achieved.

Privacy Solutions Evaluation: Participants have submitted privacy protected video clips using the testing subset. The evaluation of the submitted privacy solutions is based on the human-perceived level of privacy filtering i.e., the level by which the High/Low regions of personally identifiable information, as previously annotated in the dataset, have been responsively obscured by appropriate filtering techniques. Thus the evaluation is essentially based on the overall human perception and interpretation of the resulting privacy filtered image in terms of the level of retained information i.e., intelligibility, and, appropriateness (acceptability-attractiveness) of the privacy filtered image (also defined in the MediaEval 2012, MediaEval 2013, and, MediaEval 2014 Privacy Task descriptions [1,2]).

Participants will each receive the results of the evaluations of their submission as well as the overall results and rankings for all the submitted entries. The rankings will be based on the application of different weightings to the results for each of the above three criteria (privacy protection level, intelligibility, appropriateness) as calculated from the evaluation results arising from evaluations by the surveillance security practitioners and naïve evaluators.

The weightings will be agreed by the participants so as to reflect the relative importance of each of the above three evaluation criteria as perceived by each of the human evaluator groups. 6 participants from the security practitioner's category and 11 from the naïve category will be asked to complete a survey with 13 questions after viewing each of 3 randomly selected and distinct videos of results of privacy filtering as submitted by each team. The 13 question will evaluate all three criteria. The score given to each team will consist of the average score for each of the criteria mentioned above for each evaluation category.

5. ACKNOWLEDGEMENTS

The Drone Protect Task at MediaEval 2015 was supported by the European Commission under contract FP7-261743 VideoSense.

6. REFERENCES

[1] Badii, A., Al-Obaidi, A., and Einig, M., MediaEval 2013 Visual Privacy Task: Holistic Evaluation Framework for Privacy by Co-Design Impact Assessment. MediaEval 2013 Workshop. CEUR-WS.org, 1043, Barcelona, Spain, October 2013.

[2] A. Badii, T. Ebrahimi, C. Fedorczak, P. Korshunov, T. Piatrik, V. Eiselein, and A. Al-Obaidi. Overview of the MediaEval 2014 visual privacy task, In MediaEval 2014 Workshop, Barcelona, Spain, October 2014.

[3] Bonetto, M., Korshunov, P., Ramponi, G., and Ebrahimi, T., Privacy in Mini-drone Based Video Surveillance, Workshop on De-identification for privacy protection in multimedia, May 2015.

[4] 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.