Overview of the MediaEval 2013 Visual Privacy Task

Atta Badii[1], Mathieu Einig[1], Tomas Piatrik[2]

[1] University of Reading
Intelligent Systems Research Laboratory,
School of Systems Engineering
United Kingdom
{atta.badii; m.l.einig}@reading.ac.uk

[2] Queen Mary,
University of London
Multimedia & Vision Research Group
United Kingdom
tomas.piatrik@elec.qmul.ac.uk

ABSTRACT

This paper describes the Visual Privacy Task (VPT) 2013, its scope and objectives, related dataset and evaluation approach.

1. INTRODUCTION

Advances in artificial intelligence and video surveillance have led to increasingly complex surveillance systems of rising scale and capabilities. This ubiquity and intelligence poses great threats to privacy, and new mitigation technologies must be found to ensure an appropriate level of privacy protection. The Visual Privacy Task (VPT) aims at exploring how image processing, computer vision and scrambling techniques can deliver technological solutions to some visual privacy problems. The evaluation is performed using both video analytics algorithms and user studies so as to provide both subjective and objective evaluation of privacy protection techniques. The manner in which the privacy of individual actors appearing in a video scene may need to be protected can, at least partially, depend on their context-specific preferences for privacy. The context can include their behaviour and interaction with each other and/or with any objects in the scene. Effective privacy protection must model any such context-dependent personal privacy preferences and this is a challenging extension of this VPT for the future.

2. THE VPT 2013 DATASET

The data set consists 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. People may also carry specific items which could potentially reveal their identity and may therefore need to be filtered appropriately. For this year, people can carry backpacks, umbrellas, wear scarves, and can be seen fighting, pickpocketing or simply walking around. People may be at a distance from the camera or near the camera, making their faces vary considerably in pixel size and quality. The videos have variable ambient lighting with half of the clips recorded at night. The dataset contains 22 video clips and associated annotations in xml form. The videos include indoor, outdoor, day-time and night-time environments, showing people interacting or performing various actions. The clips are in the mpeg format with a resolution of 1920x1080 pixels at 25 frames per second. Publications arising from experiments performed using PEViD must acknowledge its publishers [1].

MediaEval 2013 Workshop, October 18-19, 2013, Barcelona, Spain

3. VISUAL PRIVACY TASK

This task explores how image processing, computer vision and scrambling techniques can deliver technological solutions to some visual privacy problems [2] [3] [4]. The goal of privacy protection is to prevent potential access to information, the divulgement of which can amount to a (perceived) intrusion of an individual's privacy. The extent of such a (perceived) loss of privacy depends on the individual as well as the context and as such can only be determined by reference to the user ("data-subject") in each case. Context-specific privacy protection constitutes an interesting extension of this VPT task which is planned to be included in future challenges. The goal of this VPT is to propose methods whereby persons featured in digital imagery can be obscured so as to render them unrecognisable. Privacy level variations may also be triggered by detected anomalies, critical events, and alerts etc. or be based on prior official permission granted by higher authorities to suspend the masking of the identity of an individual in specific cases. Since the resulting partially obscured videos would still have to convey some video information to be worth viewing, an optimal balance should be struck so that despite the extent of such masking of the facial identity as may be necessary, the categorical identity of any masked actors e.g. humans can still be recognisable to the viewer. Thus identity obscuring techniques should not result in artefacts that are 'socially inappropriate/offensive' and unacceptable to the human users. The participants should also demonstrate that their choice of obscuring technique is such that the resulting obscured (e.g. pixelated) faces do not tend to fixate a viewers' attention thus distracting the viewer and/or adversely impacting the acceptability-usability of any obscured/scrambled images, from the perspective of both the data-subject as well as other viewers. Participants are provided with videos containing faces from different camera angles. The ground truth consists of annotations of persons' images, including face, hair, visible skin regions, as well as their personal accessories.

3.1 Objective metrics

The objective metrics are computed automatically with a mixture of object detection and matching in order to evaluate the impact of the filtering on the privacy and intelligibility. Some additional image quality measures will be taken into account in order to give credit to filters resulting in visually-pleasant masking.

3.1.1 Face Detection

A face detection algorithm will be run on the obscured videos submitted for the evaluation using the Viola-Jones face detection

from OpenCV library. Ideally, no faces should be found, since they all should be obscured. The faces found by the face detection algorithm are matched against the ground truth to avoid including the false positives of the detection algorithm.

3.1.2 Object Tracking

The intelligibility is measured by applying the Histogram of Oriented Gradient as a human detector taking the video images as input. Successful detections of a human means that even although the sensitive areas may have been obscured, the resulting video could still carry sufficient visible clues for Video Analytics including tracking. These detections are compared against the detections from the raw video.

3.1.3 Person Re-identification

A visual model of the un-filtered images of persons as featured in the video set will be developed and matched against the privacy-filtered versions of the images of the same persons as selected from the submission set. The matching process will be implemented in two ways so as to provide the basis for a Merit Criterion Framework for Privacy Impact Assessment based on Efficacy, Consistency, Disambiguity and Intelligibility PIAF[5]; as follows: **i)** by building a visual model from the original unfiltered image in each case and then attempting to match this against the respective filtered image, and, **ii)** by building the model from the filtered image and attempting to match it against the respective unfiltered original set. A low re-identification score arising from the above matching cycles would indicate a higher *Efficacy* privacy protection afforded by the privacy filtering techniques as deployed in each case. *Consistency, Disambiguity and Intelligibility* properties of the deployed Privacy Filtering approach will also be assessed by comparing the filtered visual model to the filtered instances of the target person(s) in the image set. A high score would indicate that the filtered video still carries 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. The framework has been extended to enable the video-context-sensitive thresholding of the Merit Criteria. This provides a powerful benchmarking mechanism for the spectrum of possible privacy filtering techniques, in terms of their optimisation of the trade-offs (identity maskability /trackability) across the specific criteria to suit the objectives of the video processing with privacy protection and surveillance by best balancing the resulting Efficacy, Consistency, Disambiguity and Intelligebility impacts of particular privacy filtering techniques as deployed in arbitrary *situated* video-contexts and UI-REF based privacy requirements.

3.1.4 Metric for Visual Appropriateness

Obscuring of the image of persons and their accessories will be evaluated using SSIM and PSNR metrics for image quality based on the human eye perception of salience in the image. A successful privacy filtering system should have a minimal impact on the global quality of the image with modifications occurring only on the sensitive areas which should be thus anonymised.

3.2 User Study for Assessment of Appropriateness of Visual Privacy Filtering

A random subset of videos from the submitted runs will also be evaluated through a user study aimed at developing a deeper understanding of user perceptions of appropriateness in terms of UI-REF based privacy protection requirements. This subjective evaluation will take into account three main aspects of any obscured (element of) image, namely intelligibility, privacy, and appropriateness. In the context of surveillance scenarios, questions related to whether a person wears personal items that can be used for identification e.g. (branded) backpack, scarf, etc, will be considered as relevant to privacy and intelligibility.

Figure 1. Sample frame from the VPT Data Set [1]

The visual appropriateness of the obscured images will be evaluated based on the various aspects such as pleasantness, distraction, and user acceptance for video surveillance, etc. The visual appropriateness criterion will essentially follow a UI-REF based evaluation methodology [6]. This metricates: **i)** the categoric "recognisability" of an obscured image as a member of a particular species, and, **ii)** the obscuring *Effects*, and *Side-Effects* on the perception of the image by a viewer, and, **iii)** the extent of any resulting negative or positive emotions or *Affects* or distraction in the mind of the viewer of an image that has been subjected to such obscuring (indignity/stigma). Insights from this user study will serve as a baseline for refining the metrics and shall inform the design of the future privacy tasks.

4. ACKNOWLEDGMENTS

This Visual Privacy MediaEval task was supported by the European Commission under contracts FP7-261743 VideoSense.

5. REFERENCES

[1] Korshunov P. & Ebrahimi T., "PEViD: privacy evaluation video dataset". Applications of Digital Image Processing XXXVI, San Diego, California, USA, August 25-29, 2013.

[2] Dufaux, F. & Ebrahimi, T., "Scrambling for Privacy Protection in Video Surveillance Systems," IEEE Transaction on Circuits and Systems for Video Technology, Vol. 18, Nr. 8 (2008), p. 1168-1174

[3] Dufaux, F. & Ebrahimi, T., "A framework for the validation of privacy protection solutions in video surveillance," 2010 IEEE International Conference on Multimedia and Expo (ICME), pp.66-71, 19-23 July 2010.

[4] Senior, A., "Privacy Protection in a Video Surveillance System," Privacy Protection in Video Surveillance, Springer, 2009

[5] Badii, A, Einig, M, Al-Obaidi, Ducournau, A, "The Merit Criteria Framework for Impact Assessment of Privacy Filtering Technologies, based on Efficacy, Consistency, Disambiguity and Intelligibility of Privacy Protection", Working Paper UoR-ISR-VS-2013-3, May2013.

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