Privacy in Mini-drone Based Video Surveillance

Margherita Bonetto[1], Pavel Korshunov[2], Giovanni Ramponi[1], and Touradj Ebrahimi[2]
[1] Image Processing Laboratory, DIA, University of Trieste, Italy
[2] Multimedia Signal Processing Group, EPFL, Lausanne, Switzerland

***Abstract*— Mini-drones are increasingly used in video surveillance. Their areal mobility and ability to carry video cameras provide new perspectives in visual surveillance which can impact privacy in ways that have not been considered in a typical surveillance scenario. To better understand and analyze them, we have created a publicly available video dataset of typical drone-based surveillance sequences in a car parking. Using the sequences from this dataset, we have assessed five privacy protection filters via a crowdsourcing evaluation. We asked crowdsourcing workers several privacy- and surveillance-related questions to determine the tradeoff between intelligibility of the scene and privacy, and we present conclusions of this evaluation in this paper.**

***Index Terms*— Mini-drones, video surveillance, dataset, privacy, crowdsourcing evaluation**

I. INTRODUCTION

Recently, mini-drones became widely available due to affordable prices and stable flight performance. They are also able to carry sophisticated video acquisition devices. One of their main weak points is their short autonomy, a problem that will be progressively solved, since battery technology is also improving rapidly. Mini-drones can capture the same scene from different points of view, and can get close to targets. As a consequence, they can collect sensitive personal data, which adds a new dimension to issues around privacy and calls for appropriate privacy protection solutions.

In order to better understand the implications of such novel devices, a publicly available video dataset[1] was created with a DJI Phantom 2 Vision+ mini-drone. The dataset is designed for the analysis and evaluation of privacy concerns. It consists of 38 different contents that depict a typical surveillance scenario in a parking lot exposing different levels of privacy intrusiveness. Participants appearing in the video have various gender and ethnicity, are dressed differently, and carry personal items and accessories in order to emphasize visual privacy, i.e., personal visual information. The sequences were processed and various privacy-sensitive regions, including body silhouettes, faces, cars, accessories, and license plates, were manually annotated, and stored in an XML format.

Several state-of-the-art privacy filters were applied with different degrees of strength to each content shot with the mini-drone, in order to understand if a balance can be found between privacy issues and surveillance effectiveness. Privacy filters included simple filters such as blurring, pixelization and masking, as well as more advanced reversible warping [1] and morphing [2] filters. The performance of each tool was subjectively evaluated using a crowdsourcing approach. Test subjects were asked to answer carefully selected questions related to visual privacy and typical surveillance tasks, in order to assess performance of visual privacy protection filters. The results of this investigation allowed us to find the right balance each filter can offer between intelligibility and privacy protection. The evaluation results are also included in the created dataset to help researchers in the analysis of privacy in mini-drones and as an example of how the dataset can be used.

This work was conducted in the framework of Network of Excellence VideoSense and COST Action IC1206. Special thanks to Dr. Jens Hälterlein and Dr. Leon Hempel for the valuable discussions about ethical problems in surveillance, their help in the dataset and evaluation test creation.

[1] http://mmspg.epfl.ch/mini-drone

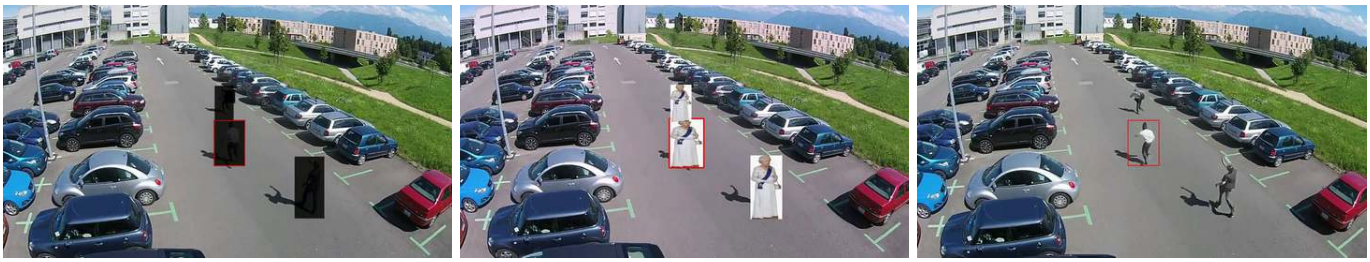
II. BACKGROUND AND RELATED WORK

The new features implemented in drone-based surveillance affect visual privacy, as already observed by several researchers [3], [4], [5], [6], [7], [8], [9]. However, there is a notable lack of adequate datasets that can be used to analyze these new surveillance devices. Many datasets exist for the evaluation of video analytics, such as various detection, recognition, and tracking algorithms; for instance VIRAT, CAVIAR, ChokePoint, and PETS 2007. A few datasets were recently created for privacy evaluations in video surveillance when using different types of visual sensors [10], [11], but none of them includes footage from mini-drones.

Little has been done to better understand privacy issues in practical multimedia applications. But recently the impact of privacy protection tools has been analyzed in video surveillance and effective evaluation methodologies have been developed to take into account both the context and the content. The objective evaluation of several primitive privacy filters was first performed by Newton *et al.* [12]: the authors demonstrated that such filters cannot adequately protect from successful face recognition, because recognition algorithms are robust. The robustness of face recognition and detection algorithms to primitive distortions is also reported in [13]. Further, in a work by Dufaux *et al.* [14], a framework is defined to evaluate the performance of face recognition algorithms applied to images altered by various obfuscation methods.

Crowdsourcing has shown to be a viable alternative to conventional laboratory-based subjective assessments, especially for cognitive tasks [11]. Crowdsourcing-based evaluation of privacy tradeoff in video surveillance has shown good consistency with laboratory-based studies [15]. The crowdsourcing methodology benefits from a large number of participants and can be performed efficiently and at a relatively low cost

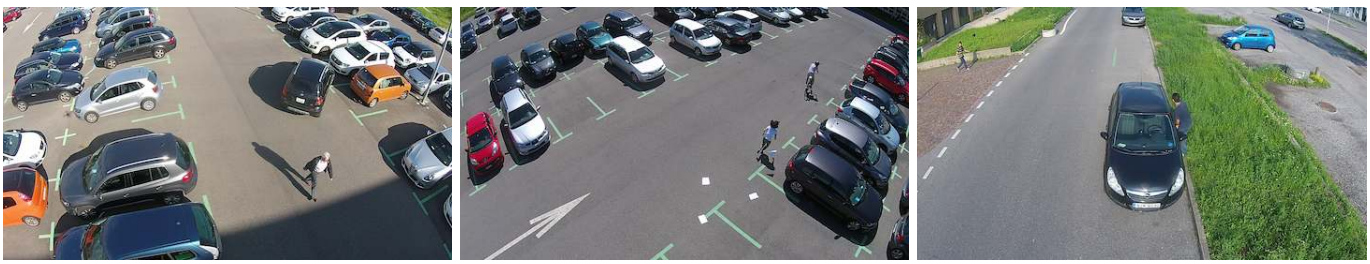
(a) Annotated original (b) Blurring, strength 40 (c) Pixelization, strength 20



(d) Masking, strength 0.8 (e) Morphing, strength 0.95 (f) Warping, strength 5

Fig. 1: Original and filtered frames of stealing bag video from the mini-drone dataset.

(a) Normal behavior: Walking (b) Suspicious behavior: Taking pictures of the cars (c) Illicit behavior: Stealing an item



(d) Illicit behavior: Mis-parking a car (e) Illicit behavior: Crashing (f) Illicit behavior: Stealing a car

Fig. 2: Examples of dataset scenarios.

without requiring a significant commitment from subjects, which are called workers in the crowdsourcing terminology. Workers accept to undertake a task (usually a short 5-20 minutes task) and are grouped in larger units, called batches. When the evaluation experiment is over, workers submit their answers. Unlike laboratory-based experiments, crowdsourcing cannot impose specific displays or controlled illumination of surroundings in which assessments take place. However, since standard environment and equipment conditions for surveillance operators have not been established, typical monitors even with different resolutions and color settings are considered as appropriate in this study.

To display video sequences to different workers and to collect evaluation results, we selected QualityCrowd2[2] framework [16] and the Microworkers[3] crowdsourcing platform that provide online workers from around the world. QualityCrowd2 is an open-source framework designed for QoE evaluation with crowdsourcing. This framework was selected because it is easy to modify for our privacy evaluation task using the provided simple scripting language for batch creation, training sessions, and control questions.

[2]https://github.com/ldvpublic/QualityCrowd2
[3]http://microworkers.com/

TABLE I: Questions asked in the crowdsourcing study (left column) and the choice of the answers (right column).

Question	**Choice of answers**
1. What is the main activity happening in the video?	Stealing a car, attacking a driver, stealing an item, walking, parking a car, taking pictures, I do not know
2. How many people do you see?	One, two, three, four, five, I do not know
3. Is there any of the following items? (select all that apply)	Backpack, umbrella, photo camera, papers, wallet, none, I do not know
4. What is the GENDER of the person in the red box?	Male, female, I dont know
5. What is the ETHNICITY of the person in the red box?	White, African, Asian, I dont know
6. Which accessories does the person in the red box wear? (select all that apply)	Jacket, sunglasses, glasses, helmet, shorts, hat, hoodie, none of the above, I do not know

TABLE II: Scenarios depicted in the video dataset.

Type of scenario	**Main action**	**Gender, Age and Ethnicity**
Normal	A person tries to fix his broken down car;	Caucasian man;
	People walk in the car parking;	Caucasian men and women, Asian guy;
	A driver parks his car and leaves on foot;	Caucasian man;
	A driver gets into his car and leaves;	Caucasian man;
Suspicious	A person falls down and asks for help;	Asian girl;
	A driver parks his car and join two people, they start talking stealthily;	Caucasian men and woman;
	Loitering people in the car parking	Caucasian guys and girl, Asian guy;
	A person takes pictures of the parked cars;	Asian guy;
Illicit	A person pushes the driver outside his car and steals the vehicle;	Caucasian men;
	Two people start arguing and fighting;	Asian and Caucasian guys;
	A driver parks his car in the middle of the road and leaves;	Caucasian man;
	A woman parks his car and takes up two lots;	Caucasian woman;
	A driver parks his car in the forbidden area and leaves;	Caucasian man;
	A cyclist crashes with a pedestrian;	Caucasian and Asian girls;
	A person puts some bottles into a parking lot;	Asian girl;
	A person steals or tries to steal a car;	Caucasian girl and guy;
	A person steals a wallet, a bag or a backpack;	Caucasian and Asian guys and girls;
	A person approaches a car and steals it, two people are on lookout;	Caucasian guys;

III. DATASET CREATION

Drone-based surveillance is particularly advantageous when it is not possible to set up a full-fledged surveillance system, for example, when a temporary major event such as a concert or a marathon is organized. Mini-drones can be used for monitoring the area, helping in managing parking spaces, controlling crowds and reporting useful information such as suspicious behaviors, mis-parked cars, number of free parking spots, etc.

A video dataset suitable for privacy inspection in drone-based video surveillance should have appropriate features:

- Practical scenarios: since many vehicles are left unattended, theft and vandalism are common. Therefore, most of the dataset videos show suspicious people and criminal behaviors;
- Different levels of privacy intrusiveness: the impact on the privacy of those under surveillance is variable, because the drone can remain still or move, it can follow, get closer or rotate around a person or a vehicle;
- Emphasis on people' visual privacy: the recorded videos should not only include facial information but also ethnicity, age, gender, personal items, and accessories;
- Emphasis on vehicles visual privacy: the recorded videos should include information about the license plate, the model, and the color;
- Varying environment and illumination conditions: in order to thoroughly evaluate the performance of privacy protection;
- Video of high quality: the sensitive privacy regions should be clearly visible if unprotected.

The created dataset consists of 38 different contents captured in full HD resolution, with a duration of 16 to 24 seconds each, shot with the mini-drone Phantom 2 Vision+ in a parking lot. The dataset contents can be clustered in three categories: normal, suspicious, and illicit behaviors. The scenarios are reported in Table II and examples are shown in Figure 2. Normal content depicts people walking, getting in their cars and parking their vehicles. In suspicious content, nothing *a priori* wrong happens but people act in a questionable way. Contents with illicit behaviors show people mis-parking their vehicles, stealing items and cars, or fighting. All participants read and signed a consent form, stating they agree to appear with their vehicles in the video.

IV. DATASET ANNOTATION

The sensitive data, also referred to as regions of interest, ROIs, were manually annotated using the open source ViPER-GT tool[4] and provided in flexible XML format. For every video, frame-by-frame annotations for each person and vehicle were performed manually. The following privacy-related regions were annotated:

- Body silhouette: Rectangle around the body region with recorded information about gender, ethnicity and age. Stored information about the surveillance scenario: the

[4]http://viper-toolkit.sourceforge.net/

main action, such as stealing or parking, and the role, such as thief or driver;
- Facial region: Rectangle around the face;
- Accessories. Rectangle around each personal item such as bag, backpack, sunglasses, hat, wallet or bottle;
- Vehicle: Rotated rectangle around the vehicle body, car or bicycle.
- License plate: Rotated rectangle around the license plate. Number of license plate recorded;
- Video capture: Information about video format, including resolution, frame rate, and the total number of frames.

Since our dataset was created to evaluate different aspects and definitions of privacy, the attribute 'level of privacy content' is reported for each ROI and defined as low (L), medium (M), or high (H). It is related to the distance between the region and the drone and to the amount of visible details. The ROIs face and license plate are more sensitive than the others. A person can be recognized more easily if his face is visible and the number of the license plate can help to identify the owner of the vehicle. Therefore, their default value for 'level of privacy content' attribute is H while for other regions of interest it is set to L.

V. VISUAL PRIVACY FILTERS

Privacy protection tools have been already applied to surveillance-related video datasets [15], [17]. Based on these studies, a number of popular protection tools including blurring, pixelization, and masking filters, but also more complex filters such as warping [1] and morphing [2] were applied. The choice of the filters parameters is a challenging issue by itself, because the perspective of the on-board camera can suddenly change and result in a change of size for privacy sensitive regions in a video sequence. The approach suggested in [18] was adopted, which focused on the performance of recognition algorithms in privacy evaluation. The strength levels were selected according to the following four categories: (i) mild, when the filter is almost imperceptible, (ii) noticeable, when the filter is clearly visible and leads to obfuscation of some minor details such as license plates, (iii) clearly visible, when most of the protected objects in the video are obfuscated, and (iv) completely obfuscating when the filter yields its maximum protection. The strength was adjusted by changing the Gaussian kernel size for the blurring filter to values 5, 20, 40, and 60, the size of the averaging block for the pixelization filter to values 5, 10, 20, and 50, the opacity for the rectangular masking filter to values 0.2, 0.6, 0.8, and 1.0, the value added to the shifted points for the warping filter to values 1, 2, 5, and 20, and the weights of the pixel intensities for the morphing filter to values 0.1, 0.4, 0.8, and 0.95. Figure 1 illustrates the original annotated content and the results of applying the filters to a sample video frame.

Privacy protection filters were applied to body silhouettes and cars. In this way, faces, license plates, and accessories were also filtered at the same time.

VI. EVALUATION FRAMEWORK

The crowdsourcing assessment aims to check whether a given surveillance task can be performed or an individual's behavior can be detected, even after the privacy protection filters are applied. For this purpose, each crowdsourcing worker was asked to watch a video sequence and to answer to one of the questions in Table I, as per the approach proposed in [17] and [18].

The first three questions were created to measure the amount of intelligibility. The answers to the last three questions permit instead to determine how much privacy sensitive information such as ethnicity and gender or other privacy related details still remain visible after filtering. It should be noted that a red box was drawn in the video to avoid confusion regarding the person to which questions 4, 5 and 6 in Table I referred to. The 3^{rd} and 6^{th} questions were multiple-choice. The answer "I don't know" could also be selected for all the questions posed. The Microworkers platform provides online workers with the ability to choose the location of workers, which was selected in countries where English is a dominant language.

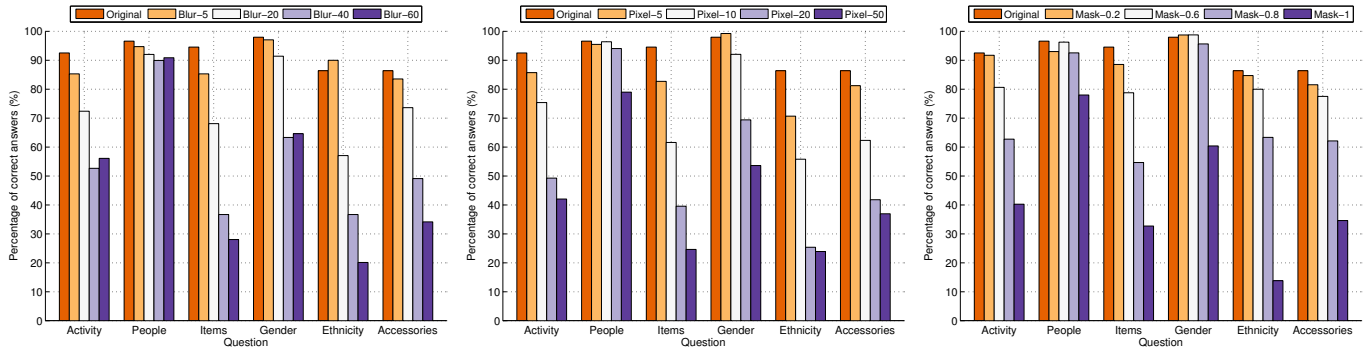
Seven different contents were selected from the dataset to evaluate the performance of the privacy tools. The contents depict the scenarios: attacking a driver, mis-parking a car, stealing a wallet, stealing a backpack, people talking, stealing a car, and taking pictures of the cars. They show a variety of sensitive regions and individuals' behavior. Original sequences in 1920×1080 resolution were compressed in MPEG-4, converted to Flash Video format, and played back at a resolution of 960×540 to make sure the video could be properly decoded and represented by most common browsers and on typical monitors.

In total, 21 different video sequences were created for each content (the original, plus 20 filtered versions) for assessment. To ensure a statistically significant number of evaluations for each sequence, 40 subjects were assigned to each sequence, with a total of 840 subjects participating in the evaluations.

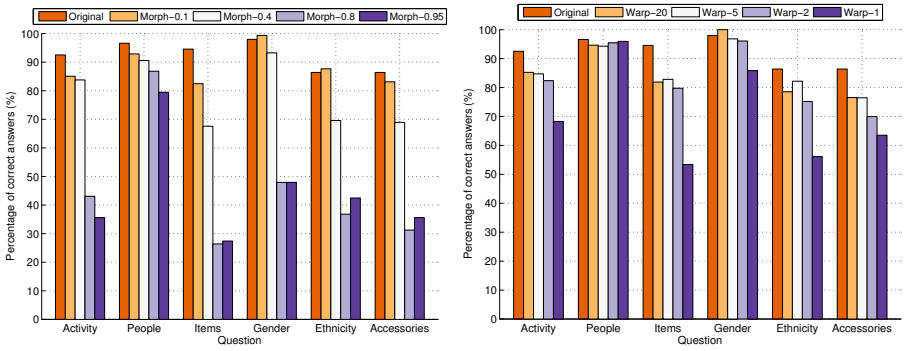
The sequences corresponding to the same content were randomly distributed among the batches; special care was devoted to guarantee that each subject assessed only one version of a given content. Since every subject has to answer six questions for seven contents, 42 steps should be performed to complete a batch. After each question, subjects were asked to report also how certain they were about their answer. Each batch starts with a training session describing the evaluation procedure. A display brightness test is performed using a method similar to that described in [19].

VII. EVALUATION RESULTS

Since the major shortcoming of the crowdsourcing-based subjective evaluation is the inability to supervise participants behavior and to restrict their test conditions, there are several techniques to exclude unreliable workers [19]. To identify a worker as 'trustworthy', the following four approaches were used in our crowdsourcing evaluation:

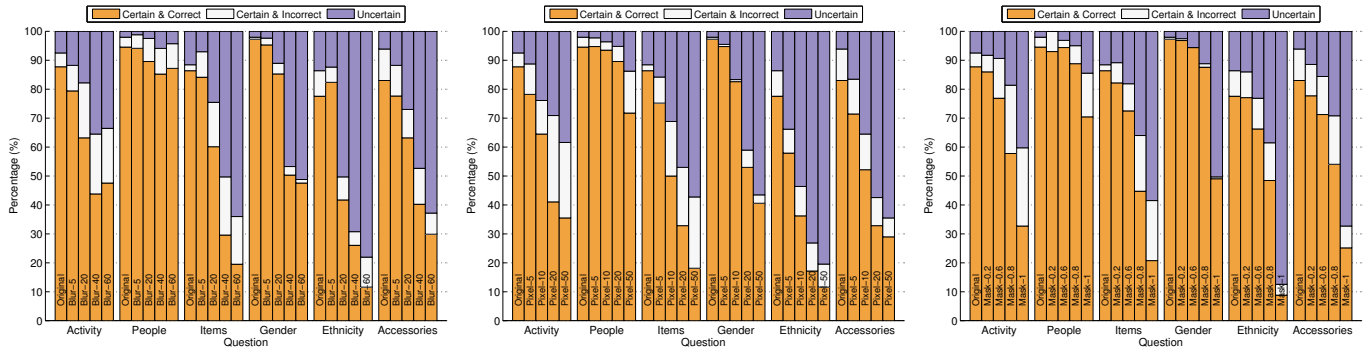


(a) Blurring (b) Pixelization (c) Masking

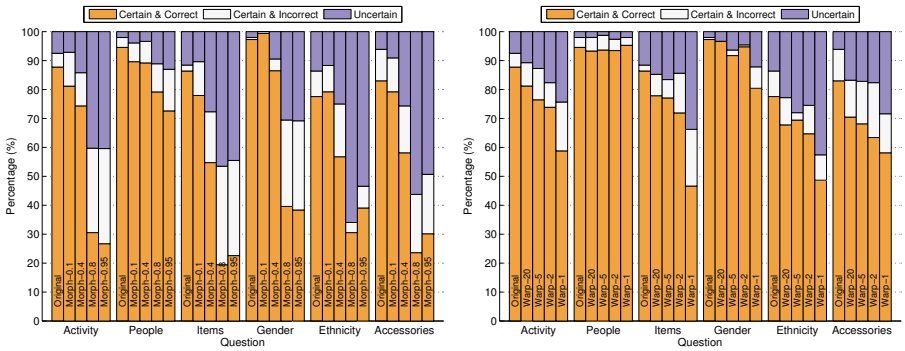


(d) Morphing (e) Warping

Fig. 3: Correct answers for different filters and their strength by workers from crowdsourcing evaluation.



(a) Blurring (b) Pixelization (c) Masking



(d) Morphing (e) Warping

Fig. 4: Certain & correct, certain & incorrect, and uncertain answers for different filters and their strength by workers from crowdsourcing evaluation.

- Two 'Honeypot' questions were inserted in each batch. These are the obvious easy-to-answer questions to detect people who do not pay attention;
- Task completion time of the worker;
- Mean time spent on each question by the worker;
- Deviation of the time spent on each question by the worker.

Based on these factors, 456 out of 840 (54% of total) workers were found to be reliable with 19 to 24 reliable workers for each tested video sequences, which ensures the statistical significance of the evaluation results.

Figure 3 demonstrates the crowdsourcing evaluation results for each privacy protection filter and their different strength levels. The figure shows how each filter affects the visibility of different regions when applied at different strength. The bars represent the average across different video contents of correct answers grouped according to the questions from Table I, as shown on the x-axis. Each plot in the figure also shows the results for original 'unfiltered' video sequence for the ease of comparison. The average deviation of the correct answers across different contents is about 18% with less than 10% for original video, about 10% for the minimal levels of strength, and up to 28% for high strength levels.

Figure 4 illustrates the effects of filters strength levels on certainty with which workers answered the questions. The total number of answers are split into those that were certain and correct, certain and incorrect, and uncertain. An ideal privacy protection filter would lead to high uncertainty but very low number of certain and incorrect answers, because surveillance related judgements based on wrong information are not desirable.

Figure 3 demonstrates a general trend of filters able to decrease the number of correct answers to all questions when high strength levels are used. The least affected are questions about the number of people (question 2 in Table I) and gender (question 4 in Table I).

From the presented figures, it can be noted that basic filters such as blurring and pixelization are able to achieve the better tradeoffs, as they cluster towards intermediate values of both privacy and intelligibility. Also, they lead to less certain and incorrect answers, as shown in Figure 4a and Figure 4b.

VIII. CONCLUSION AND FUTURE WORK

We have presented a publicly available dataset designed for the analysis and evaluation of privacy concerns in mini-drone video surveillance. Using this dataset, we have investigated for the first time the performance of privacy protection filters in drone-based video surveillance. We have applied five typical privacy protection tools with four different levels of strength. The filtered sequences have been evaluated by the workers of a crowdsourcing platform, and the results have been analyzed to investigate the balance between intelligibility and privacy protection of different privacy filters.

More advanced privacy protection filters like scrambling [20] or encryption-based tools will be exploited in the continuation of the present study. Different questions could also be selected, for example related to the age and the expression of the person. Other privacy-related features could be studied too, such as those related to the identification of the vehicles (the license plate, but also the details of the wheels, or stickers that may be present). Video content from CCTV should be compared to the mini-drone video dataset to highlight the versatility of the latter.

REFERENCES

[1] P. Korshunov and T. Ebrahimi, "Using warping for privacy protection in video surveillance," in *18th International Conference on Digital Signal Processing (DSP)*, Santorini, Greece, June 2013, DSP'13.

[2] P. Korshunov and T. Ebrahimi, "Using face morphing to protect privacy," in *IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*, Krakow, Poland, Aug. 2013.

[3] A. Villasenor, "Observations from above: unmanned aircraft systems and privacy," *Harvard Journal of Law Public Policy*, vol. 36, pp. 458–517, 2013.

[4] D. Wright and R. L. Finn, "Unmanned aircraft systems: surveillance, ethics and privacy in civil applications," *Computer Law Security Review*, vol. 28.

[5] R. Clarke, "The regulation of civilian drones impacts on behavioral privacy," *Computer Law Security Review*, vol. 30.

[6] L. B. Moses and R. Clarke, "The regulation of civilian drones impacts on public safety," *Computer Law Security Review*, vol. 30.

[7] J. Villasenor, "Drones and the future of domestic aviation [point of view]," in *Proceedings of IEEE*.

[8] R.L. Wilson, "Ethical issues with use of drone aircraft," in *Ethics in Science, Technology and Engineering, 2014 IEEE International Symposium on*, May 2014, pp. 1–4.

[9] J. Pitt, C. Perakslis, and K. Michael, "Drones humanus [introduction to the special issue]," *Technology and Society Magazine, IEEE*, vol. 33, no. 2, pp. 38–39, Summer 2014.

[10] P. Korshunov and T. Ebrahimi, "PEViD: privacy evaluation video dataset," in *SPIE Applications of Digital Image Processing XXXVI*, San Diego, California, USA, Aug. 2013, vol. 8856.

[11] P. Korshunov, H. Nemoto, A. Skodras, and T. Ebrahimi, "Crowdsourcing-based evaluation of privacy in HDR images," in *SPIE Photonics Europe 2014, Optics, Photonics and Digital Technologies for Multimedia Applications*, Brussels, Belgium, Apr. 2014.

[12] E. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying face images," *IEEE Trans. on Knowledge and Data Engineering*, vol. 17, no. 2, pp. 232–243, Feb. 2005.

[13] P. Korshunov and W. T. Ooi, "Video quality for face detection, recognition, and tracking," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 7, no. 3, pp. 14:1–14:21, Sept. 2011.

[14] F. Dufaux and T. Ebrahimi, "A framework for the validation of privacy protection solutions in video surveillance," in *Proceedings of IEEE International Conference on Multimedia & Expo (ICME 2010)*, Singapore, July 2010.

[15] P. Korshunov, S. Cai, and T. Ebrahimi, "Crowdsourcing approach for evaluation of privacy filters in video surveillance," in *Proceedings of the ACM Multimedia 2012 Workshop on Crowdsourcing for Multimedia*, Nara, Japan, Oct. 2012, CrowdMM'12, pp. 35–40.

[16] Christian Keimel, Julian Habigt, Clemens Horch, and Klaus Diepold, "Qualitycrowd — a framework for crowd-based quality evaluation," in *Picture Coding Symposium 2012 (PCS2012)*, May 2012, pp. 245–248.

[17] P. Korshunov, C. Araimo, F. De Simone, C. Velardo, J. Dugelay, and T. Ebrahimi, "Subjective study of privacy filters in video surveillance," in *2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*, Sept. 2012, pp. 378–382.

[18] P. Korshunov and T. Ebrahimi, "Towards optimal distortion-based visual privacy filters," in *IEEE International Conference on Image Processing, ICIP'2014*, Paris, France, 2014.

[19] Tobias Hossfeld, Christian Keimel, Matthias Hirth, Bruno Gardlo, Julian Habigt, Klaus Diepold, and Phuoc Tran-Gia, "Best practices for QoE crowdtesting: QoE assessment with crowdsourcing," *IEEE Transactions on Multimedia*, vol. PP, no. 99, pp. 1–1, 2013.

[20] F. Dufaux and T. Ebrahimi, "Scrambling for privacy protection in video surveillance systems," *Circuits and systems for video technology, IEEE Transactions on*, vol. 18, no. 8, pp. 1168–1174, July 2008.