# IMPACT OF MINI-DRONE BASED VIDEO SURVEILLANCE ON INVASION OF PRIVACY

*Pavel Korshunov*[1]*, Margherita Bonetto*[2]*, Touradj Ebrahimi*[1]*, and Giovanni Ramponi*[2]

[1] Multimedia Signal Processing Group, EPFL, Lausanne, Switzerland
[2] Image Processing Laboratory, DIA, University of Trieste, Italy

## ABSTRACT

An increase in adoption of video surveillance, affecting many aspects of daily lives, raises public concern about an intrusion into individual privacy. New sensing and surveillance technologies, such as mini-drones, threaten to eradicate boundaries of private space even more. Therefore, it is important to study the effect of mini-drones on privacy intrusion and to understand how existing protection privacy filters perform on a video captured by a mini-drone. To this end, we have built a publicly available video dataset of typical drone-based surveillance sequences in a car parking. Using the sequences from this dataset, we assessed five privacy protection filters at different strength levels 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 protection provided by the filters.

***Index Terms—*** Video surveillance, mini-drones, dataset, privacy, crowdsourcing evaluation

## 1. INTRODUCTION

In recent years, the advances in microelectronics, signal processing, and aerodynamics led to the popularity of unmanned aerial vehicles and mini-drones in particular. Drones are commonly used in applications such as military, surveillance, photography, cinema, entertainment, and agriculture. They can operate at different heights, can capture the same scene at different angles, and can get closer to targets. As a consequence, they can be used to spy on individuals and collect sensitive surveillance data, which adds a new threat to privacy and calls for appropriate privacy protection solutions [1, 2].

This paper addresses the privacy issues in drone-based video surveillance systems. To better understand the implications of using drone-based surveillance, a publicly available video dataset[1] was created with a DJI Phantom 2 Vision+ mini-drone. The dataset is specifically designed for the analysis and evaluation of privacy issues. It consists of 38 different sequences that depict a typical surveillance scenario in a parking lot exposing different levels of privacy intrusiveness.

To understand and quantify the balance between privacy of people under surveillance and security-related features, a subjective evaluation approach proposed in [3] was adapted to evaluate several privacy filters, such as blurring, pixelization, masking, warping [4], and morphing [5]. The filters were applied with different degrees of strength on video sequences from the created mini-drone dataset. The performance of each privacy filter was subjectively evaluated using a crowdsourcing method, since this approach was shown as a viable alternative to lab-based subjective assessments [6, 7]. For this purpose, the open-source framework QualityCrowd2 [8] was adapted and crowdsourcing workers were employed from Microworkers[2] platform. The workers 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 allowed us to investigate the balance each filter can offer between intelligibility and privacy protection.

## 2. DATASET

The created dataset consists of 38 different contents captured in full HD, with a duration of 16 to 24 seconds each, shot with DJI Phantom 2 Vision+ mini-drone in a parking lot (see the sample frame in Figure 1a). The dataset shows different people and accessories in various situations, viewing angles, and lighting conditions (day and evening). The contents can be clustered in three categories: normal, suspicious, and illicit behaviors. Normal behavior includes people walking, getting in and parking their cars. The suspicious content, *a priori* nothing wrong happens but people display questionable behavior, such as loitering or taking pictures of parked cars. Illicit behaviors shows people mis-parking their vehicles, stealing items and cars, or fighting. The impact on the privacy of those under surveillance is variable, because of the different ways the drone is able to capture video. To emphasize the differences with respect to conventional CCTV surveillance, the drone maneuvered by hovering at different heights, following, getting closer to, or rotating around an object of interest.

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 and help in the dataset and evaluation criteria in tests.

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

[2] http://microworkers.com/

(a) Annotated original (b) Blurring (40) (c) Pixelization (20) (d) Masking (0.8) (e) Morphing (0.95) (f) Warping (5)

**Fig. 1**: Original frame of stealing bag video from the mini-drone dataset and a filtered cropped region (strength is in brackets).

**Table 1**: 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 do not know |
| 5. What is the ETHNICITY of the person in the red box? | White, African, Asian, I do not 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 |

The contents in the dataset provide a variety of personal visual information. The sensitive regions in each content, including faces, body silhouettes, accessories, cars bodies, and license plates, are manually annotated and recorded in an XML format.

## 3. VISUAL PRIVACY FILTERS

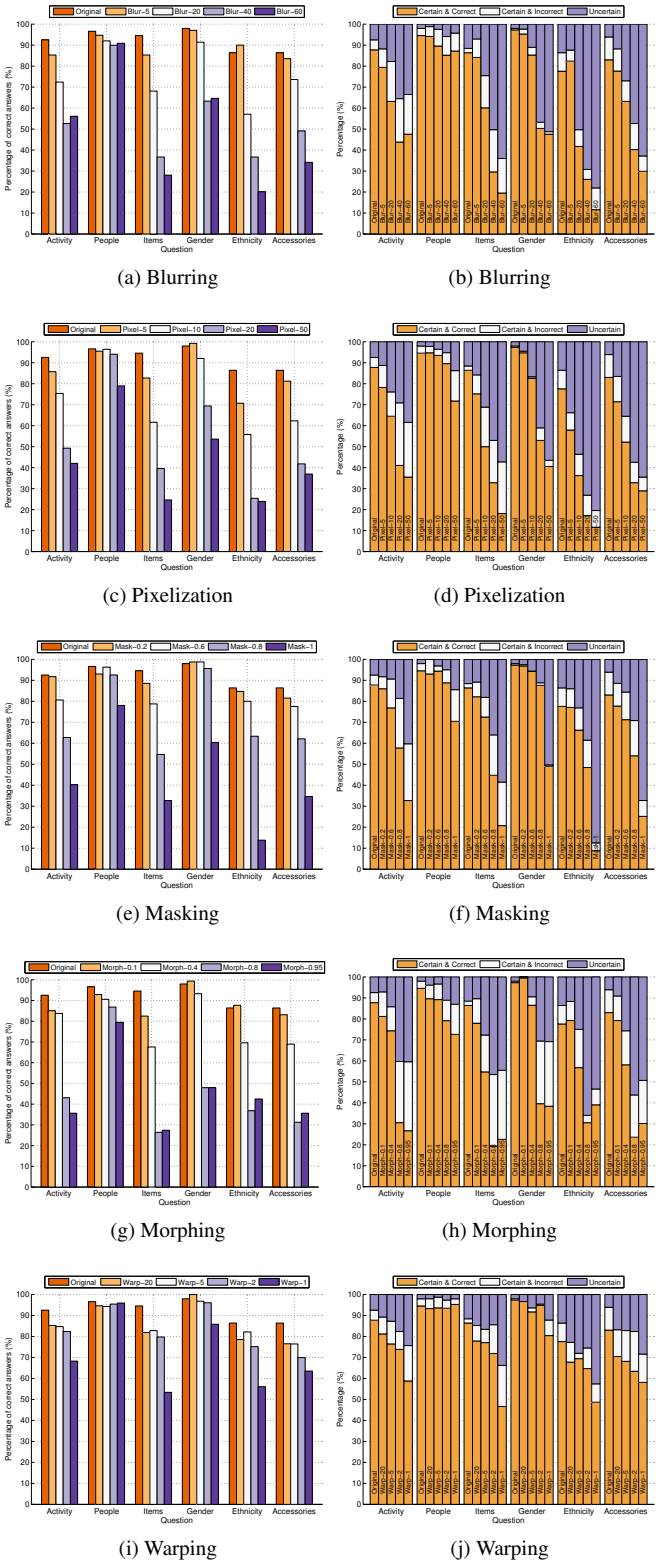
Several state-of-the-art tools for privacy protection have been previously applied to surveillance-related video datasets [3, 6]. Based on these studies, a number of protection tools were chosen for testing on mini-drone videos, including blurring, pixelization, and masking filters, as well as more complex reversible filters such as warping [4] and morphing [5]. To investigate their performance, four levels of strength were selected for each filter. The choice of the filters parameters is a challenging issue by itself, because sudden perspective changes of the on-board camera result in a change of size for privacy sensitive regions. The approach suggested in [9] was adopted, which focuses on the performance of recognition algorithms in privacy evaluation. The strength levels were selected to fit into the following four categories: (i) mild, when the filter is hard to notice, (ii) noticeable, when the filtered region is generally visible but some minor details such as license plates are unclear, (iii) obfuscating, when most of the protected objects are visually concealed, and (iv) completely obfuscating when the filter yields its maximum protection.

Based on the above considerations, the strength for blurring filter was adjusted by changing the Gaussian kernel size to values 5, 20, 40, and 60; for pixelization filter, the size of the averaging block to values 5, 10, 20, and 50; for masking filter, the opacity to values 0.2, 0.6, 0.8, and 1.0; for warping filter, the distance for the shifted points to values 1, 2, 5, and 20; and for morphing filter, the weight of the pixel intensities to values 0.1, 0.4, 0.8, and 0.95. Examples of filtered video frames are shown in Figure 1.

For the crowdsourcing evaluations, privacy protection filters were applied to body silhouettes and cars, which, consequently, also obfuscated faces, license plates, and accessories.

## 4. EVALUATION FRAMEWORK

The goal of the subjective assessment based on the crowdsourcing approach is to understand whether a given surveillance task can be performed or an individual's behavior can be detected, even after the privacy protection filters are applied. At the same time, the effectiveness of privacy protection filters is assessed by determining the degree by which the individual identities in the sequences remain hidden. For this purpose, each crowdsourcing worker was asked to watch a video sequence and to answer to one of the questions in Table 1, as per the approach proposed in [3]. Also, each question had an accompanied question about certainty ('How sure are you?'). A red box was drawn in the video to avoid confusion regarding the person to which questions 4, 5 and 6 in Table 1 referred to. The Microworkers[2] platform allows employers to choose the location of workers, which was selected in countries where English is a dominant language.




(a) Blurring (b) Blurring

(c) Pixelization (d) Pixelization

(e) Masking (f) Masking

(g) Morphing (h) Morphing

(i) Warping (j) Warping

**Fig. 2**: Results of crowdsourcing evaluation for different filters and their strength. Left column shows the correct answers by workers. Right column shows the 'certain & correct', 'certain & incorrect', and 'uncertain' answers.

Seven different contents were selected from the dataset to evaluate the performance of the privacy tools. They show a variety of sensitive regions and individuals behavior. Original sequences in $1920 \times 1080$ resolution were compressed in MPEG-4, converted to Flash Video format, and played back at a resolution of $960 \times 540$ to make sure the video could be properly viewed using most common browsers and 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, and considering the presence of unreliable workers (about 50% in a typical crowdsourcing evaluation), 40 workers evaluated each sequence, with a total of 840 workers performing the corwdsourcing task.

The sequences corresponding to the same content were randomly distributed among different tasks with special care devoted to guaranteeing that a particular content was used only once in every task, i.e., that each worker assessed only one version of a given content.

## 5. EVALUATION RESULTS

### 5.1. Reliable workers detection

A major shortcoming of the crowdsourcing-based subjective evaluation is that the employer is unable to supervise the behavior of online workers and to ensure the similarity of test conditions in all evaluations. This leads to a risk of including unreliable data into the analysis, for instance, due to the different environment, lighting conditions, or untrusted workers who submit low quality work to reduce their effort while maximizing their compensation [10].

By following recommendations in [10], we first detected unreliable workers using two 'Honeypot' questions inserted in each task. Honeypot questions are obvious easy-to-answer questions for detecting people who did not pay attention. In addition, for each worker, the following time-based metrics were used: (i) task completion time, (ii) mean time spent on each question in a task, and (iii) standard deviation of the time spent on each question. These metrics allow removing the results from workers who are very different from average, for instance, if they take too long to finish the task or time to answer each question varies drastically. In total, 456 out of 840 (54% of total) workers were found to be reliable with 19 to 24 reliable workers for each tested video sequence, which insures the statistical significance of the evaluation results.

### 5.2. Evaluation Analysis

Figure 2 demonstrates the crowdsourcing evaluation results for each privacy protection filter and their different strength levels. In the figure, the bars are grouped according to the questions from Table 1. Each plot also shows the results for original video sequence for the ease of comparison.

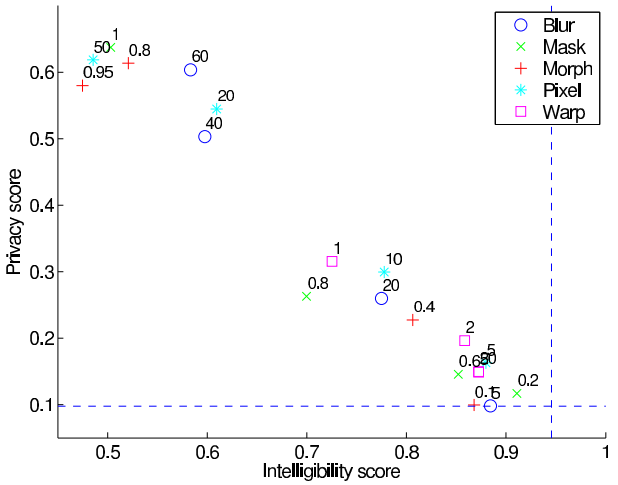



**Fig. 3**: Privacy vs. intelligibility tradeoff for each filter.

The left column of Figure 2 shows how each filter affects the visibility of different regions when applied at different strengths. The bars represent the average percentage of correct answers across different video contents (7 were used in the evaluation). The average standard deviation of correct answers is about 18% with less than 10% for original 'unfiltered' video and from about 10% (minimal strength levels) to 28% (high strength levels) for filtered video. Such high deviation values are probably due to the significantly varying video content used in the evaluation with scenes ranging from those with high motion, when the mini-drone was moving and circling at high speed, to static close-up scenes. It means that the number of correct answers, for both privacy and intelligibility, not only depends on filter and its strength but also on the type of the content.

The right column of Figure 2 illustrates the effects of filters strength levels on certainty, with which workers responded to accompanied questions about certainty of their answers to the main questions. Hence, the total number of answers are split into those that were 'certain & correct', 'certain & incorrect', and 'uncertain'. An ideal privacy protection filter should lead to high uncertainty but very low number of 'certain & incorrect' answers, because surveillance related judgements based on wrong information are undesirable, for instance, when a guard decides to ignore an activity, because of wrongly attributing it to a normal behavior, when it is not.

Figure 2 demonstrates a general trend: application of filters with high strength levels decreases the number of correct answers to all questions. The least affected are questions about the number of people and gender. Such decrease in correct answers is desirable for questions related to privacy but not for questions related to intelligibility, because it makes filtered video to be less useful for surveillance purposes.

To understand the effect of the privacy tools on the tradeoff between privacy and intelligibility, the answers to the first 3 questions were averaged to compute the intelligibility score, while privacy score was computed from the last 3 questions as a difference between total and correct answers (the less number of people answer correctly to privacy question, the better is the protection). The resulted scores are presented as a scatter plot in Figure 3, where the different point markers correspond to different filters with the respective strength level shown above each point. The scores for original unfiltered video are indicated by dashed horizontal and vertical lines. The figure demonstrates a tradeoff similar to that reported in [6] with filters able to either achieve high intelligibility with sacrifice in privacy, for low filter strength levels, or high privacy but with low intelligibility, i.e., usefulness in terms of surveillance, for high filter strength levels.

From Figure 3, it can be noted that basic filters such as blurring and pixelization lead to more suitable privacy-intelligibility tradeoffs in practical applications, since their values cluster around the middle of the plot. Also, from Figure 2b, blurring leads to lesser number of 'certain & incorrect' answers. Morphing, especially at high strength levels, leads to a lot of 'certain & incorrect' answers for questions about activity, items, and gender (see Figure 2h), but it is probably due to the type of morphing filter applied (a cartoonish 'Queen of England') and may lead to a different conclusion if a different type of morphing target is used.

## 6. CONCLUSION AND FUTURE WORK

In this paper, we have investigated for the first time the performance of privacy protection filters in drone-based video surveillance. Five typical privacy protection tools were applied 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.

The subjective assessment with crowdsourcing approach suggests that the pixelization and blurring filters can be effectively controlled: the intelligibility decreases and the privacy protection smoothly increases by increasing the strength parameters. The masking tool with opacity 1 shows the lowest privacy intrusiveness; unexpectedly, the ability to perform the surveillance task was preserved since the drone's perspective view permits to exploit the shadows of people and objects. Morphing with the highest strength value is similar to masking, while the warping tool did not show a large variability.

As future work, the evaluations should also include other advanced privacy protection filters such as scrambling [11] or encryption-based [12] tools. Crowdsourcing results could also be compared with lab-based evaluations. Different questions could also be selected, for example, related to the age and the expression of the person. Also, since drone-based videos are significantly more challenging for video analytics than typical CCTV video, specialized new approaches can be developed for an automated detection of privacy-sensitive objects, e.g., person tracking or license plate recognition.

## 7. REFERENCES


[1] A. Villasenor, "Observations from above: unmanned aircraft systems and privacy," *Harvard Journal of Law Public Policy*, vol. 36, no. 2, 2013.

[2] D. Wright and R. L. Finn, "Unmanned aircraft systems: surveillance, ethics and privacy in civil applications," *Computer Law Security Review*, vol. 28, pp. 184–194, 2012.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

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

[10] 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.

[11] 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.

[12] Paula Carrillo, Hari Kalva, and Spyros Magliveras, "Compression independent reversible encryption for privacy in video surveillance," *EURASIP J. Inf. Secur.*, vol. 2009, pp. 5:1–5:13, Jan. 2009.