UHD VIDEO DATASET FOR EVALUATION OF PRIVACY

Pavel Korshunov and Touradj Ebrahimi

Multimedia Signal Processing Group, EPFL, Lausanne, Switzerland

ABSTRACT

Ultra High Definition (UHD) is one of the emerging immersive video technologies already available to public, as even some of the smartphones are capable of capturing UHD video. The increasingly widespread availability of UHD capable recording devices has important implications on privacy. This paper addresses the problem by proposing a publicly available UHD video dataset designed for evaluation of privacy issues. The dataset depicts typical surveillance scenarios of people fighting, exchanging bags, walking, and stealing, in an indoor and outdoor environment. The dataset also includes the data from the subjective assessment, which evaluated the impact of UHD on privacy compared to a currently common High Definition (HD) video and declining Standard Definition (SD) video. The results of the assessment not only demonstrate that UHD is a significantly more privacy intrusive technology when compared to HD and SD used today, but they also quantify the impact of intrusiveness.

Index Terms— Dataset, UHD video, video surveillance, privacy evaluations, subjective study

1. INTRODUCTION

Recent adoption of digital video surveillance systems, especially in public spaces and communities, has significantly increased the concern for intrusion into individual privacy. New visual sensing technologies, such as Ultra High Definition (UHD) video capture, threaten to eradicate boundaries of private space even more. UHD TV sets and cameras are already available in stores and even some smartphones, e.g., Samsung Galaxy Note 3, are capable of recording UHD video. Therefore, it is normal to assume that UHD is likely to be adopted by video surveillance systems in the nearest future. However, the degree to which UHD exposes privacy is little studied and is not well understood.

To understand the implications of using UHD video in video surveillance, a dataset suitable for privacy evaluation is necessary. Existing public UHD datasets are mostly designed for evaluation of video compression and do not carry privacy sensitive information. And datasets dedicated to the evaluation of privacy issues, such as PEViD [1], contain at most HD video sequences.

To remedy this problem, we created a publicly available UHD video dataset PEViD-UHD[1] designed specifically for evaluation of privacy. The dataset consists of 26 video sequences, each 13 seconds long, with a frame resolution of 3840×2160 pixels and captured at 30 fps using a Samsung Galaxy Note 3 smartphone. The design principles given in [1] were applied when building the dataset, which makes it general enough to reflect different aspects of privacy and allows the evaluation of protection tools for facial, as well as, for other visual features that can reveal information about race, gender, height, personal identifiable items and accessories, gait, etc.

Video sequences in the proposed dataset depict several typical surveillance scenarios: walking, exchanging bags, fighting, and stealing, which were shot in outdoor and indoor environments. Participants appearing in the video have various gender and race, they are dressed differently and carry various personal items and accessories. Their silhouettes were manually annotated and the annotations are provided in XML format. All participants have read and signed a consent form, allowing free usage of these video sequences for research purposes.

A specific property of this dataset that highlights, in terms of privacy, the difference between UHD and a typical HD video, is that the action scenes were shot from far away (see snapshot examples in Figure 1). The aim is to demonstrate that an important privacy-related details remain visible at distances that are much greater than what is typically assumed as a safe distance from a person recording with a smartphone. It can potentially imply a serious threat to privacy if UHD-capable surveillance cameras and smartphones become commonly used in public spaces.

To better understand and quantify the degree to which a UHD camera is privacy-threatening, we conducted subjective experiments evaluating privacy intrusiveness of UHD and compared it with HD and SD video. The sequences from the UHD dataset were resized to HD (1920×1080 pixels) and SD (720×404 pixels) resolutions. Privacy evaluation methodology proposed in [2] was adapted to the specific surveillance scenarios of the dataset. Subjective evaluations were con-

This work has been conducted in the framework of EC funded Network of Excellence VideoSense and COST IC1003 European Network on Quality of Experience in Multimedia Systems and Services QUALINET. Special thanks for the help in creating dataset and running subjective experiment to Fatma Belghith.

[1]The dataset can be downloaded here: http://mmspg.epfl.ch/pevid-uhd

Fig. 1: Video frame examples from the PEViD-UHD dataset.

ducted using a professional reference UHD Sony Trimaster SRM-L560 monitor and 20 naïve subjects took part in the evaluations.

In summary, the contribution of this paper is twofold: (i) it proposes the first UHD video dataset for evaluation of privacy issues in video surveillance, and (ii) it presents subjective study of the quantitative impact on privacy of UHD video when compared to HD and SD video.

In the rest of the paper, the UHD video dataset for privacy evaluations is described in details, including the scenarios, capture process, and annotations, followed by the presentation of subjective evaluation and discussion of the results on how much UHD video is more privacy intrusive when compared to HD and SD video.

2. BACKGROUND AND RELATED WORK

There are many datasets for evaluation of video analytics, such as various detection, recognition, and tracking algorithms. The most notable datasets include FERET dataset[2] and Labeled Faces in the Wild (LWF)[3] for evaluation of face detection and recognition algorithms, as well as several datasets representing different video surveillance scenarios, such as VIRAT[4], CAVIAR[5], ChokePoint[6] and PETS 2007[7]. But as these datasets were not designed with privacy issues in mind, they are not suitable for the evaluation of privacy protection filters or testing other privacy related aspects.

There is one existing video dataset specifically designed for evaluation of privacy and privacy protection tools, called PEViD [1]. This dataset is used in Visual Privacy Task of MediaEval benchmarking initiative[8] in 2013 and 2014 for evaluation of visual privacy protection tools. The proposed PEViD-UHD dataset is created following similar design principles that were used to build PEViD dataset with the main difference that PEViD contains only HD video sequences.

One of the main purposes of a dataset for privacy evaluations is to assess visual privacy protection tools. There are many methods for privacy protection. The simplest ones rely on visual distortion of the pixels of sensitive regions or on replacement of faces in a video frame with some simple shapes. For instance, in [3] people's identities are protected by obscuring their faces with a colored ellipse. Other naïve approaches also include blurring, pixelization, or masking for hiding the faces of people in video. More complex distortion-based methods include technique for obscuring the whole body silhouettes [4], which is based on edge and motion models, or a complete removal of the silhouette of the moving person from the scene to hide identity [5]. Aiming to avoid constraints of the distortion-based methods, more advanced reversible and secure scrambling-based privacy filters are proposed in [6] and [7]. These techniques are based on randomized (seeded with a secret key) modifications of the

[2] http://www.itl.nist.gov/iad/humanid/feret/feret_master.html
[3] http://vis-www.cs.umass.edu/lfw/
[4] http://www.viratdata.org/
[5] http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/
[6] http://itee.uq.edu.au/ uqywong6/chokepoint.html
[7] http://pets2007.net/
[8] http://www.multimediaeval.org/

Table 1: Summary of the different video sequences in PEViD-UHD dataset.

Environment	Scenario	Gender & Race	Accessories	Videos
indoor, day	walking	woman: white; man: asian	bracelet, scarf	2
	fighting	woman: white; man: white	scarf, bag, necklace, glasses	3
	exchanging bag	woman: white; man: white, asian	glasses, paper bag, backpack	3
	stealing	woman: white; man: white, asian	student card, glasses, phone, money	4
outdoor, day	walking	woman: white; man: white, asian	glasses, beard, phone	5
	fighting	man: white	none	1
	exchanging bag	woman: white; man: white, asian	glasses, paper bag, backpack	3
	stealing	woman: white; man: white	glasses, phone, money	5

compressed video stream that reversibly obscure privacy sensitive regions. In privacy through Invertible Cryptographic Obscuration (PICO) [8] facial pixels are encrypted in order to conceal identity. The process is reversible for authorized users in possession of a secret encryption key. The idea of encrypting or scrambling face regions has also been proposed in [9] and [10], where the focus is on the compression based encryption mechanism. Secure methods based on geometrical transformations [11, 12] were also developed recently boasting the independence of compression encoders.

Proposed UHD dataset together with the results of subjective evaluations can be used for benchmarking of privacy protection tools and filters in the context of UHD, HD, and SD video. Therefore, the proposed dataset can be of a great interest to the research community, and it provides a quantifiable assessment of the impact of privacy as a function of video resolution.

3. DATASET DETAILS

Datasets for evaluation of video analytics in video surveillance and datasets for evaluation of visual privacy protection tools should both contain typical video surveillance scenarios. However, there are few differences from typical surveillance dataset that a dataset for privacy protection should further include:

- Wide range of practical surveillance scenarios. This is as opposed to typical surveillance datasets where some specific conditions are assumed for evaluation of a particular video analysis algorithm, such as face recognition;
- Emphasis on personal visual information and its variety. It should not include just a facial information but also sufficient content on different races, genders, personal items and accessories, etc.;
- The means to select different privacy regions for different evaluation scenarios.
- Video of high quality, so the sensitive privacy regions are clearly visible if unprotected. Video analytics are expected to perform well under such conditions, which challenges privacy protection tools.

In this paper, we focus on event detection scenarios when one or two people are engaged in an activity (fighting, walking, stealing, etc.). Crowd analysis and crowd surveillance is out of the scope of the paper.

3.1. Dataset description

Following the above principles, we created a UHD video dataset for privacy evaluation (see Figure 1 for screenshot examples). In total, 26 video sequences of 4K UHD frame resolution 3840×2160 pixels were recorded at 30 fps. Each sequence was cut to 13 seconds to keep it long enough to show a particular scenario but short enough, so it can be easily used in subjective evaluations. All video sequences were recorded using a Samsung Galaxy Note 3 smartphone. The goal was not to record a high quality video, as it is done for evaluation of compression algorithms, but to obtain realistic video that can be potentially used in surveillance. The resulted video was therefore degraded with some blurring and color correction artifacts due to the significant post-processing and compression performed by the acquisition system.

Several typical indoor and outdoor video surveillance scenarios were considered, such as simple walking (1 participant), stealing (2 participants), exchanging a bag (2 participants), and fighting (2 participants). Most of the participants in dataset recording were students from EPFL campus. A specific effort was made to ensure the variety of gender, race, and different personal accessories that people carried or wore. All various scenarios contained in the dataset are summarized in Table 1.

A subset containing 20 video sequences of the dataset was annotated using ViPER-GT annotation tool[9]. For every video, frame-by-frame annotations for each person was performed manually to record the silhouette of the moving person. All annotations were stored in flexible XML format. Each annotation file also included information about video format, such as resolution, frame rate, and the total number of frames.

[9] http://viper-toolkit.sourceforge.net/docs/gt/

Table 2: Questions asked during the assessment (left column) and the choice of answers (right column).

Question	Choice of answers
1. Which accessory does the person in the red box wear?	Glasses, Beard, Bracelet, Sunglasses, Watch, Scarf, Necklace, None of the above
2. How certain are you about what the person wears?	Very sure, Sure, Neutral, Not so sure, Unsure
3. Which item do you see in the video?	Student Card, Credit card, Phone, Paper bag, Plastic bag, Backpack, Money, None of the above
4. How certain are you about the item you saw?	Very sure, Sure, Neutral, Not so sure, Unsure
5. What is the main action shown in the video?	Walking, Running, Stealing, Dancing, Exchanging bags, Fighting, No action
6. How certain are you about the main action?	Very sure, Sure, Neutral, Not so sure, Unsure
7. What is the gender of the person in the red box?	Male, Female, I don't know
8. What is the race of the person in the red box?	White, African, Asian, I don't know

Besides following general principles of designing a video dataset for privacy evaluation, UHD dataset should highlight and emphasize the difference that UHD brings in terms of privacy intrusiveness when compared to HD or SD resolution. Therefore, special care was devoted to record UHD video sequences in such a way that demonstrates this difference. Therefore, Modulation Transfer Function (MTF) was computed for the screenshot of ISO 12233 Test Chart defined in [13] in UHD, HD, and SD resolutions. The obtained values of MTF function, which is 1317 LW/PH for UHD, 852 LW/PH for HD, and 660 LW/PH for SD when recording with Samsung Galaxy Note 3, allowed us to estimate ranges of frequencies that are visible in UHD but not in HD or SD. Based on these estimations, we computed the distances range at which the scene should be shot, so the important details are still visible in UHD but not necessarily visible in other resolutions.

The UHD video dataset was created in such a way to allow the evaluation of different aspects and definitions of privacy (race, gender, face information, accessories- and gait-based personal infromation) independently, as well as jointly, by using either objective or subjective tests. In accordance to European and Swiss laws and best practices, each participant in the shooting of the video sequences have read and signed a consent, allowing for the obtained sequences to be freely used for research purposes.

4. SUBJECTIVE EVALUATIONS

To better understand the degree to which an UHD video impacts privacy intrusiveness when compared to lower resolutions, we conducted subjective experiments comparing privacy intrusiveness of UHD to HD and SD video. The sequences from the UHD dataset were resized to HD (1920×1080 pixels) and SD (720×404 pixels) resolutions and they were evaluated using the methodology proposed in [2] and extended in [14].

Fig. 2: Experimental setup in the lab.

4.1. Evaluation Methodology

An important issue to resolve when evaluating the same type of content in different resolutions was the memory effect during viewing. If an observer first watches UHD video with the highest level of details, it will affect the evaluation of the same content in HD resolution, even if not all details are visible in HD. Therefore, to avoid the memory effect, subjects first viewed all video sequences in SD (the least number of visible details), followed by HD, and finally by UHD. Hence, the test was divided into three separate sessions corresponding to their resolutions.

A total of 20 naïve subjects took part in the experiments, with 25% being female and overall age ranging from 18 to 27 years old. The test was planned over one entire day, with 7 time slots. Each time slot was attended by 3 subjects. A short training of the subjects of each group was conducted before the first test session to explain the experiment and how to use the evaluation software. Subjects were provided with iPads to answer the evaluation questions for each video. All subjects were tested for correct visual acuity and color vision using Snellen and Ishihara charts.

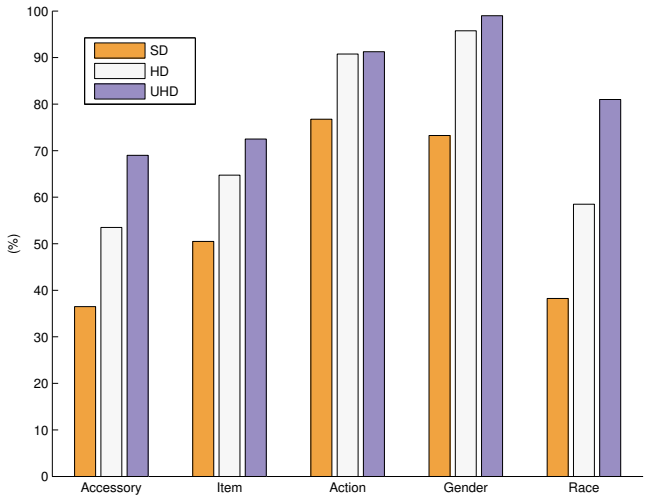

Fig. 3: Correct answers to the questions.

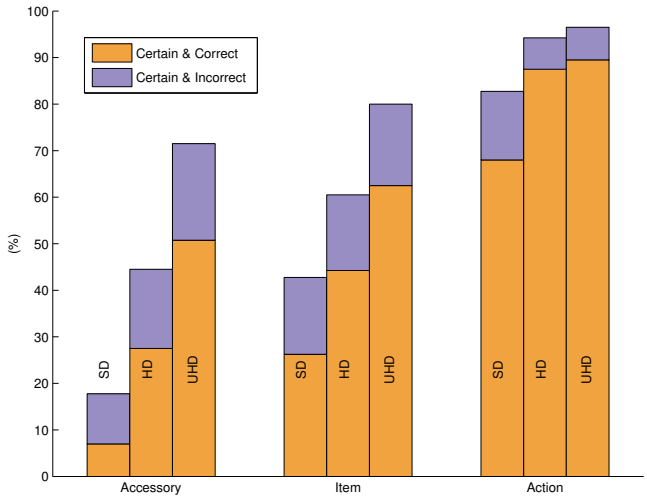

Fig. 4: Answers, for which subjects were certain.

Test consisted of three sessions of 15 minutes each (SD, HD, and UHD content) with a few minutes breaks in-between. During each session, subjects assessed 20 video sequences of 13 in duration. Test subjects were then given 30 seconds to answer questions reported in Table 2. The questions can be viewed as measuring privacy and intelligibility (or usefulness) of each video. Questions about personal accessories, gender, and race are privacy related questions, since they expose personal information. Questions about carried items and the main action are related to intelligibility. In the 'useful' video content where privacy is protected, privacy related questions should yield small number of correct answers, whereas high number of correct answers is expected for intelligibility questions.

For each group and session, video sequences were presented in a randomized order. Video was played on a 56-inch professional high performance Sony Trimaster SRM-L560 4K/QFHD LCD reference monitor. Viewing conditions for the participating subjects were set according to recommendation ITU-R BT.2022 [15] and the evaluations were performed in the testing laboratory, which fulfills the recommendations for subjective evaluation of visual data issued by ITU-R [16]. Figure 2 shows the testing laboratory and how the subjective evaluations were performed. The high quality monitor was used for the subjective evaluation, because the monitor can accurately reproduce the captured video and, therefore, the experimental setup does not have influence on the evaluation results. This is a reasonable approach, since the study does not focus on the influence of display technology on privacy, but on the influence of the increased capturing resolution.

5. RESULTS

Figure 3 shows percentages of correct answers that subjects gave to the content-related questions, i.e., questions 1 ('Accessory'), 3 ('Item'), 5 ('Action'), 7 ('Gender'), and 8 ('Race') presented in Table 2. Answers for different video resolutions are grouped together for each question. From the figure, it is clear that the number of correct answers for the same question is different for UHD, HD, and SD video. Subjects had easier time recognizing actions shown in the video and gender of the specified person, but it was harder for them to identify accessories and items. These results are expected, since accessories and items are small in size and it is more difficult to see them, while human visual system is highly sensitive to motion and distinguishing body silhouettes, which is how gender can be judged from the far distance. It is interesting to note that race was one of the hardest to identify, especially in SD video. It shows that race recognition depends on the visibility of finer details.

Considering privacy implications, Figure 3 shows that UHD video is much more privacy intrusive, even when a smartphone with a low quality sensor was used to record the video. This is especially the case for the questions that need more visible details to answer correctly.

The certainty of the subjects when answering different content-related questions (see questions 1, 2, and 3 in Table 2) was also checked and the results are illustrated by Figure 4. The figure shows only those answers, for which the subjects were certain, i.e., they answered 'Sure' or 'Very sure'. The answers are grouped according to the question to highlight the differences between SD, HD, and UHD. Some of the certain answers were incorrect and some were correct as indicated by different colors of the bars.

From the Figure 4, it can be noted that subjects tend to make a 'leap of faith' and guess the answer, hence the high number of incorrect answers were given when there were not enough visible details to make a clear judgement. For instance, in the figure, 'Accessory' bar corresponding to SD resolution shows that the number of people who are certain and

incorrect is higher than the number of people who are certain and correct. It means that more people were sure that they see some accessory (e.g., a phone or a wallet), while, in fact, they saw it wrongly. The reason for such bold judgment is a very little number of visible details in SD resolution, so many people 'guess' incorrectly. Basically, in an uncertain situation, more people tend to make a certain guess anyway. This finding implies that in practical surveillance systems, which often capture video footage at low quality, a human guard in an uncertain situation may incorrectly mistreat a person based on personal feelings even if video does not clearly implicate the person.

6. CONCLUSION

This paper presents the first public UHD video dataset designed for evaluation of visual privacy. In addition to UHD, the dataset also contains HD and SD versions of the same content and annotations of body silhouettes. Data from the subjective evaluation of the impact of UHD on privacy is also included in the dataset. The evaluation results show that UHD is a privacy intrusive technology and public needs to be more aware of the capability of this technology to capture finer visual details at large distances.

The presented dataset allows flexibility to perform objective and subjective evaluations for different types of privacy protection tools. It can be used for deeper analysis of UHD related privacy intrusiveness or for benchmarking of privacy protection filters.

7. REFERENCES

[1] 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, pp. 88561S–88561S–9.

[2] P. Korshunov, C. Araimo, F. De Simone, C. Velardo, J. Dugelay, and T. Ebrahimi, "Evaluation of visual privacy filters impact on video surveillance intelligibility," in *International Workshop on Quality of Multimedia Experience (QoMEX)*, Yarra Valley, Australia, July 2012, pp. 150–151.

[3] J. Schiff, M. Meingast, D. K. Mulligan, S. Sastry, and K. Goldberg, "Respectful cameras: detecting visual markers in real-time to address privacy concerns," in *International Conference on Intelligent Robots and Systems (IROS)*, Oct 2007, pp. 971–978.

[4] D. Chen, Y. Chang, R. Yan, and J. Yang, "Protecting personal identification in video," in *Protecting privacy in video surveillance*, A. Senior, Ed., pp. 115–128. Springer-Verlag, 2009.

[5] J. Wickramasuriya, M. Datt, S. Mehrotra, and N. Venkatasubramanian, "Privacy protecting data collection in media spaces," in *ACM international conference on Multimedia (ACM MM)*, New York, NY, USA, Oct. 2004, pp. 48–55.

[6] F. Dufaux and T. Ebrahimi, "Video surveillance using JPEG 2000," in *SPIE Applications of Digital Image Processing XXVII*, Denver, CO, Aug 2004, vol. 5588, pp. 268–275.

[7] F. Dufaux and T. Ebrahimi, "H.264/AVC video scrambling for privacy protection," in *IEEE International Conference on Image Processing (ICIP)*, San Diego, CA, Oct. 2008, pp. 1688–1691.

[8] T. E. Boult, "PICO: Privacy through invertible cryptographic obscuration," in *IEEE Workshop on Computer Vision for Interactive and Intelligent Environments*, Lexington, KY, Nov 2005, pp. 27–38.

[9] R. Grosbois, P. Gerbelot, and T. Ebrahimi, "Authentication and access control in the JPEG 2000 compressed domain," in *SPIE 46th Annual Meeting - Applications of Digital Image Processing*, 2001, pp. 95–104.

[10] A. Pande, P. Mohapatra, and J. Zambreno, "Securing multimedia content using joint compression and encryption," *IEEE Multimedia*, vol. 20, no. 4, pp. 50–61, 2013.

[11] P. Korshunov and T. Ebrahimi, "Using warping for privacy protection in video surveillance," in *18th International Conference on Digital Signal Processing (DSP)*, Santorini, Greece, July 2013, pp. 1–6.

[12] 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, pp. 208–213.

[13] ISO 12233:2014, "Photography – electronic still picture imaging – resolution and spatial frequency responses," International Organization for Standardization, Feb. 2014.

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

[15] ITU-R BT.2022, "General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays," International Telecommunication Union, August 2012.

[16] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union, January 2012.