PEViD: Privacy Evaluation Video Dataset

Pavel Korshunov and Touradj Ebrahimi

Multimedia Signal Processing Group, EPFL, Switzerland

ABSTRACT

Visual privacy protection, i.e., obfuscation of personal visual information in video surveillance is an important and increasingly popular research topic. However, while many datasets are available for testing performance of various video analytics, little to nothing exists for evaluation of visual privacy tools. Since surveillance and privacy protection have contradictory objectives, the design principles of corresponding evaluation datasets should differ too. In this paper, we outline principles that need to be considered when building a dataset for privacy evaluation. Following these principles, we present new, and the first to our knowledge, **P**rivacy **E**valuation **Vi**deo **D**ataset (PEViD). With the dataset, we provide XML-based annotations of various privacy regions, including face, accessories, skin regions, hair, body silhouette, and other personal information, and their descriptions. Via preliminary subjective tests, we demonstrate the flexibility and suitability of the dataset for privacy evaluations. The evaluation results also show the importance of secondary privacy regions that contain non-facial personal information for privacy-intelligibility tradeoff. We believe that PEViD dataset is equally suitable for evaluations of privacy protection tools using objective metrics and subjective assessments.

Keywords: Dataset, video surveillance, privacy protection tools, privacy evaluations

1. INTRODUCTION

Recent adoption of digital video surveillance systems, especially in public spaces and communities, has significantly increased the concern for protection of individual privacy, demanding development and deployment of privacy protection tools. Many visual privacy protection techniques are already available: from primitive ones, like blurring or masking, to more advanced methods, like scrambling,[1] encryption,[2] and geometrical warping.[3] Most approaches aim to remove or distort sensitive regions in video or images that contain certain identifiable personal information.

While significant efforts have been made for advancing the surveillance accuracy and robustness, including development of standard datasets, evaluation methodologies, validation techniques, accuracy metrics, etc., research in privacy protection is still in a relatively immature state. Although there are recent advances in subjective evaluation methodologies of privacy protection tools,[4,5] no commonly used dataset for evaluation of such tools exists yet. However, such dataset is necessary for establishing a standard approach to privacy evaluation.

Existing datasets for testing performance of video analytics are often designed for specific tasks, e.g., detection, recognition, or tracking, such as well-known series of datasets for PETS workshop*, and, therefore, are not suitable for evaluation of privacy protection tools. Another reason is the contradictory but dual goals of surveillance task and privacy protection, when an increase in privacy protection typically leads to the decrease in intelligibility of surveillance (making it harder to do surveillance task) and visa versa.[4] Also, visual privacy has many aspects[6] and is highly context- and application-dependent.

The notion of visual privacy is also not very well defined. Different visual regions may contain different types of personal information. For instance, face is important for human identification, and therefore is the visual region that is often obscured to protect privacy. Hence, it can be classified as *primary privacy region*. However, there are other types of visual regions, which we can term as *secondary privacy regions*, that arguably are as important to protect as the face. For instance, such personal items like bags, glasses, umbrellas or personal features like hair or skin color, can also reveal important identifiable information to an attentive observer, even though they are used less often for identification purposes than a face. In surveillance scenarios, when a closed

pavel.korshunov@epfl.ch, touradj.ebrahimi@epfl.ch

*IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS) space (e.g., an office or corridor) is surveyed for a long time by a guard, he or she can learn to know very well almost every observed person without even meeting them personally. The guard would then be able to identify some people easily without looking at their faces but just by looking at their secondary privacy regions.

Hence, a non-trivial privacy evaluation dataset should be general enough to reflect these different facets of privacy and to allow evaluating protection tools for face, as well as, for other visual regions that can reveal information about race, gender, height, personal identifiable items and accessories, gait, etc. Dataset should contain realistic scenarios, in which the privacy aspects are emphasized. Also, video in the dataset should be of such quality and have such clearly visible regions of personal information that both human and video analytics are able to observe and process this information easily. This last requirement is in contrast with the design principles of currently available surveillance datasets, which are usually designed to challenge the state of the art video analytics.

In this paper, we present our privacy evaluation video dataset, abbreviated PEViD, that follows the above guidelines. It is also, to the authors' knowledge, the first public dataset for privacy evaluation (for access, please, contact the authors directly). The dataset consists of 65 video sequences (16 seconds each) of full HD resolution covering different video surveillance scenarios: walking, fighting, stealing, and dropping bag, in outdoor and indoor environments, as well as during day and night conditions. Most of the scenes are captured by two Canon HD cameras simultaneously from different angles to provide a clearer view of people in video sequences. Participants are of various gender, race, dressed differently, and carrying various personal accessories. They have also read and signed a consent form, allowing free usage of these video sequences for research purposes.

Currently, a representative subset of 20 video sequences has been annotated using Open Source ViPER-GT[†] annotation tool. Various privacy sensitive regions are annotated, including primary and secondary privacy regions. Additional personal information, such as gender, race, hair color, etc., is also recorded in annotation files, which are stored in XML format and can be easily processed in an automatic way.

In the rest of the paper, we discuss in details the process of building privacy evaluation video dataset, the guidelines we followed, the annotations obtained, and preliminary subjective evaluation of masking privacy protection filter, which demonstrates the usage of the dataset.

2. RELATED WORK

Numerous methods for privacy protection were proposed earlier. 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 Ref. 7 people's identities are protected by obscuring their face with a colored ellipse. Other naïve approaches also include blurring, pixelization, or masking for hiding the faces of people in video. More complex methods include technique for obscuring the whole body silhouettes,[8] 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.[9] Such approaches often rely on RFID tags for pinpointing the people locations in space. Another way to protect a sensitive region securely is to encrypt it. In privacy through Invertible Cryptographic Obscuration (PICO)[2] 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 Ref. 10 and Ref. 11, where the focus is on the compression based encryption mechanism. Secure methods based on geometrical transformations[3] were also developed recently boasting the independence of compression encoders.

All these methods for privacy protection rely on the fact that regions needing protection are known and available. Therefore, a dataset that emulates practical surveillance scenarios and provides various privacy sensitive regions is necessary for a correct evaluation and comparison of different privacy filters.

There are many datasets for evaluation of video analytics, such as various detection, recognition, and tracking algorithms. The most notable datasets include FERET dataset[‡] and Labeled Faces in the Wild (LWF)[§] for evaluation of face detection and recognition algorithms, as well as several datasets representing different video

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

[‡] http://www.itl.nist.gov/iad/humanid/feret/feret_master.html

[§] http://vis-www.cs.umass.edu/lfw/

Table 1: Summary of the PEViD video dataset for evaluation of privacy protection tools.

Environment	Scenario	Gender & Race	Accessories	Videos
indoor, day	walking	woman: white, asian; man: white, asian	phone, backpack, glasses, scarf	14
	fighting	woman: white, asian: man: white, asian	backpack, bag, umbrella, glasses, bottle	8
	dropping bag	woman: white; man: white	bag, glasses	4
	stealing	woman: white, asian; man: white, black	backpack	6
indoor, night	dropping bag	woman: black; man: indian	bag, scarf	4
outdoor, day	walking	woman: white; man: white, asian	backpack, glasses, scarf	8
	fighting	woman: white, man: white, asian	backpack, scarf	6
	exchanging bag	woman: white; man: white	bag	4
outdoor, night	walking	woman: white, black	bag, scarf	5
	stealing	woman: white; man: white, indian	backpack, scarf	4
	running	woman: black; man: white	scarf	2

surveillance scenarios, such as VIRAT[¶], CAVIAR[‖], ChokePoint[**] and PETS 2007[††]. But since these datasets were not designed with privacy issues in mind, they are not suitable for evaluation of privacy filters or testing other privacy related aspects.

3. DATASET DESIGN PRINCIPLES

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 in addition 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 race, gender, 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.

4. DATASET DESCRIPTION

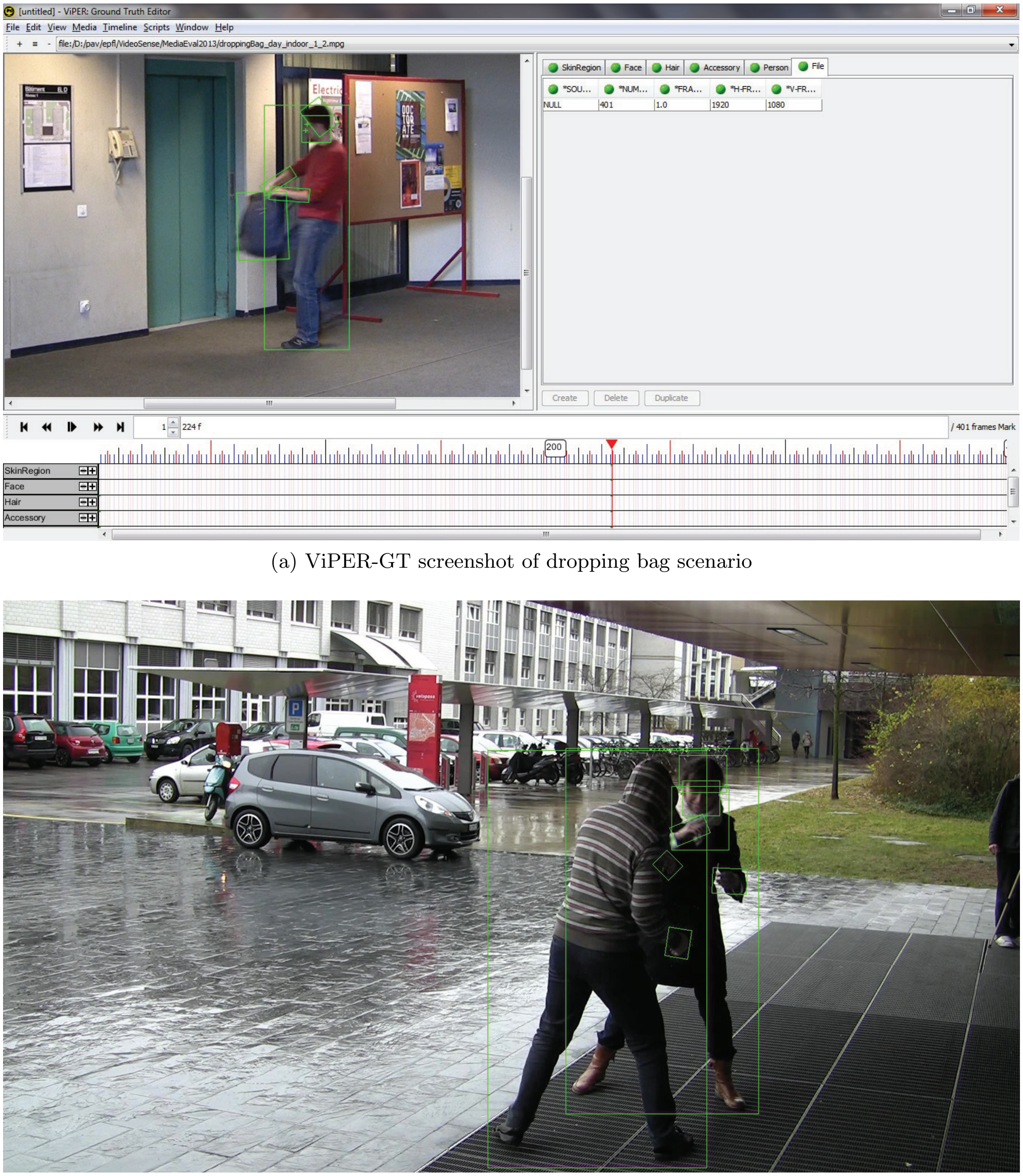
Following the above principles, we have built privacy evaluation video dataset PEViD (see Figure 1 and Figure 2 for screenshot examples). In total, 65 video sequences (full HD resolution 1080p, 25 fps), of 16 seconds each, were recorded using Canon HD HG21 camera.

[¶] http://www.viratdata.org/

[‖] http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/

[**] http://itee.uq.edu.au/ uqywong6/chokepoint.html

[††] http://pets2007.net/



[untitled] - ViPER: Ground Truth Editor

file:/D:/pav/epfl/VideoSense/MediaEval2013/droppingBag_day_indoor_1_2.mpg

*SOU...	*NUM...	*FRA...	*H-FR...	*V-FR...
NULL	401	1.0	1920	1080

1 224 f / 401 frames Mark

(a) ViPER-GT screenshot of dropping bag scenario

(b) Dropping bag

Figure 1: Examples of the dataset screenshots with annotated regions.

Several typical indoor and outdoor video surveillance scenarios were considered, such as simple walking (1 participant), stealing (2 participants), exchanging a bag (2 participants), dropping a bag (1 participant), fighting (2 participants), and running (1 participant). Several scenes were recorded for each scenario from two different angles by two cameras. Most of the participants in dataset recording were students from EPFL campus. We specifically made an effort 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.

PEViD dataset is created in such a way that allows evaluating different aspects and definitions of privacy (racial, gender, facial information, accessories- and gait-based) independently, as well as jointly, by using either objective or subjective tests. In accordance to European and Swiss laws and best practices, each participant of the recordings read and signed a disclaimer, which allows the obtained recordings to be freely used for research purposes.

5. ANNOTATIONS

A subset containing 20 video sequences of PEViD video dataset was annotated using ViPER-GT annotation tool (see Figure 1a for a screenshot of ViPER tool with annotated video). For every video, frame-by-frame annotations for each person was done manually for different privacy sensitive regions. The following privacy regions and other information was annotated and recorded (see an example annotated frame in Figure 1b):

- Facial region. Rectangle around the face, as well as points for left and right eyes and nose.
- Hair region. Rotated rectangle around the hair.
- Skin regions. Skin regions such as neck or arms are annotated separately as rotated rectangles. A color of the hair, when visible, was recorded too.
- Accessories. Each personal item such as bag, backpack, scarf, umbrella, hat, wallet, or bottle were recorded and annotated as rotated rectangles.
- Body region. A rectangle around the body silhouette with recorded information about gender and race.
- Information about surveillance scenario, such as fighting, dropping bag, walking, running, or stealing.
- Technical information about frame rate and resolution of the video.

All annotations are stored in flexible XML format as presented in Listing 1. All details about a privacy sensitive region are recorded and stored for every frame where the region is visible. Each annotation file also stores information about video format, including resolution, frame rate, and the total number of frames. The flexibility of XML allows using only those annotation regions that are necessary for a specific evaluation scenario, as we demonstrate in the next section.

Listing 1: Excerpt from XML video annotation file

```
<viper xmlns:data="http://lamp.cfar.umd.edu/viperdata#">
  <config>
    ...
    <descriptor name="Face" type="OBJECT">
        <attribute dynamic="true" name="box"
          type="http://lamp.cfar.umd.edu/viperdata#bbox"/>
        <attribute dynamic="true" name="left-eye"
          type="http://lamp.cfar.umd.edu/viperdata#point"/>
        <attribute dynamic="true" name="right-eye"
          type="http://lamp.cfar.umd.edu/viperdata#point"/>
        <attribute dynamic="true" name="nose"
          type="http://lamp.cfar.umd.edu/viperdata#point"/>
        <attribute dynamic="false" name="person_id"
          type="http://lamp.cfar.umd.edu/viperdata#dvalue"/>
    </descriptor>
      ...
  </config>
  <data>
    <sourcefile filename="fighting_day_indoor_1_1.mpg">
      <file id="0" name="Information">
        <attribute name="SOURCETYPE"/>
        <attribute name="NUMFRAMES">
            <data:dvalue value="401"/>
        </attribute>
        <attribute name="FRAMERATE">
            <data:fvalue value="25.0"/>
        </attribute>
        <attribute name="H-FRAME-SIZE">
            <data:dvalue value="1920"/>
        </attribute>
        <attribute name="V-FRAME-SIZE">
            <data:dvalue value="1080"/>
        </attribute>
      </file>
        ...
      <object framespan="1:29 38:167 198:372" id="0" name="Face">
        <attribute name="box">
          <data:bbox framespan="1:9" height="27" width="33"
            x="335" y="566"/>
          <data:bbox framespan="10:10" height="27" width="33"
            x="337" y="566"/>
          <data:bbox framespan="12:12" height="26" width="33"
            x="343" y="567"/>
          ...
      </object>
    </sourcefile>
  </data>
</viper>
```

6. SUBJECTIVE EVALUATIONS

To demonstrate different influences of various types of privacy sensitive regions, we have considered three privacy protection scenarios:

- Scenario I: only protect faces
- Scenario II: protect faces and other personal sensitive regions: accessories, hair, and skin regions
- Scenario III: protect whole body silhouettes

The first scenario corresponds to a conventional way of privacy protection used by many TV broadcasters and media in general, and, intuitively, it offers the least privacy protection out of the three scenarios. The last scenario, when a whole body is protected, corresponds to an extreme case when every information about a person, related or not to privacy, is redacted. This scenario is used in some highly privacy sensitive applications. A less conventional but more moderate in terms of privacy protection is the second scenario, when only the minimum required privacy related information is protected and the rest is visible.

To make the demonstration simple, we have applied a trivial yet popular privacy filter – masking, which is essentially a black box placed on top of the sensitive visual region. Figure 2 illustrates three different privacy protection scenarios using one of the video from our dataset.

As part of a preliminary study, we asked six different naïve observers (2 females and 4 males) several questions related to privacy and to intelligibility (accuracy of surveillance task). The questions and the answers given for each type of the evaluation scenario are presented in Table 2. We asked the subjects to answer questions on a scale from 1 to 5, with 1 corresponding to the least confident answer (very unsure about the answer) and 5 corresponding to the most confident answer (very sure about the answer). Three different video sequences were chosen for this preliminary evaluation and three versions of each sequence were obtained by applying masking filter according to the selected evaluation scenarios (see Figure 2 for screenshot examples). All sequences demonstrate fighting surveillance scenario with one video showing two people fighting outdoor and two videos showing different views of two people fighting indoor. A short training session preceded each evaluation session, explaining the scenarios, questions, and the answering scale.

From the evaluation results given in Table 2, we can note that Scenario III (protection of full body as in Figure 2c) is the most protective of privacy but it does not allow any assessment of what is going on in the video (see Question 3 in Table 2), which degrades the quality of surveillance. On the other hand, obscuring only a face is not enough for privacy protection, since such personal information as gender and race are easily identifiable via other means (e.g., skin regions). The best balance is achieved in Scenario II, when only privacy sensitive regions are identified and protected leaving all other information visible. This approach allows performing surveillance task with high accuracy, as demonstrated by the high score 4.4 given to Question 3 for Scenario II. And it also protects privacy relatively well, as indicated by the low rating for the first two questions and by the high rating for Question 4. The main difficulty with Scenario II, though, is that it requires the most effort to implement in practice, compared to other scenarios. The reason is that accurate automatic detection of secondary privacy regions (hair, bags, etc.) remains to be a hard problem, leaving only manual or semi-manual annotation as accurate enough methods for identification of such regions. On the other hand, many face and body detection and tracking algorithms are available with acceptably high accuracy, hence, the annotation of such regions is a relatively easy task, making Scenarios I and III more feasible to be used in practice compared to Scenario II.

7. CONCLUSION

In this paper, we described the guiding principles for building a dataset for evaluation of visual privacy protection tools. Based on these principles, we have created a public video dataset, allowing flexibility of objective and subjective evaluations for different types of privacy protection tools. The dataset was annotated with primary (face and body silhouette) and secondary (personal items and features, such as hair, skin regions, accessories, etc.) privacy sensitive regions. The annotations are stored in XML format. PEViD dataset is publicly available for research. Interested parties, please, contact the authors for the access instructions.

Table 2: Questions asked during the assessment (first column). For three evaluation scenarios, subjects gave answers ranged from 1 to 5, with 1 being the least confident and 5 being the most confident.

Question	Scenario I	Scenario II	Scenario III
1. Can you recognize the gender of the people?	4.7	3.8	1
2. Can you recognize the race of the people?	2.7	1.2	1
3. Can you recognize what are they doing?	5	4.4	1
4. Their privacy is well protected	1.9	3.5	5

Our preliminary subjective evaluation shows that a protection of both primary and secondary classes of personal information regions results in the best tradeoff between privacy and intelligibility. It also means that it is very important to identify the context and content of the surveillance video when applying privacy protection. However, a proper lab-based or crowdsourcing-based subjective evaluation is required in order to confirm these preliminary findings.

The proposed dataset is used for evaluations of privacy filters in Privacy Task of MediaEval benchmarking initiative.

ACKNOWLEDGMENTS

This work was conducted in the framework of the EC funded Network of Excellence VideoSense. Special thanks for the hard work to all partners of VideoSense project, especially to University of Reading (UoR), Queen Mary University of London (QMUL), Thales, Eurécom, and Technical University of Berlin (TUB) teams, who helped annotating videos in the dataset.

REFERENCES

[1] Dufaux, F. and Ebrahimi, T., "Video surveillance using JPEG 2000," in [*proc. SPIE Applications of Digital Image Processing XXVII*], **5588**, 268–275 (Aug 2004).

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

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

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

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

[6] Senior, A., Pankanti, S., Hampapur, A., Brown, L., Tian, Y.-L., Ekin, A., Connell, J., Shu, C., and Lu, M., "Enabling video privacy through computer vision," *IEEE Security and Privacy* **3**, 50–57 (May 2005).

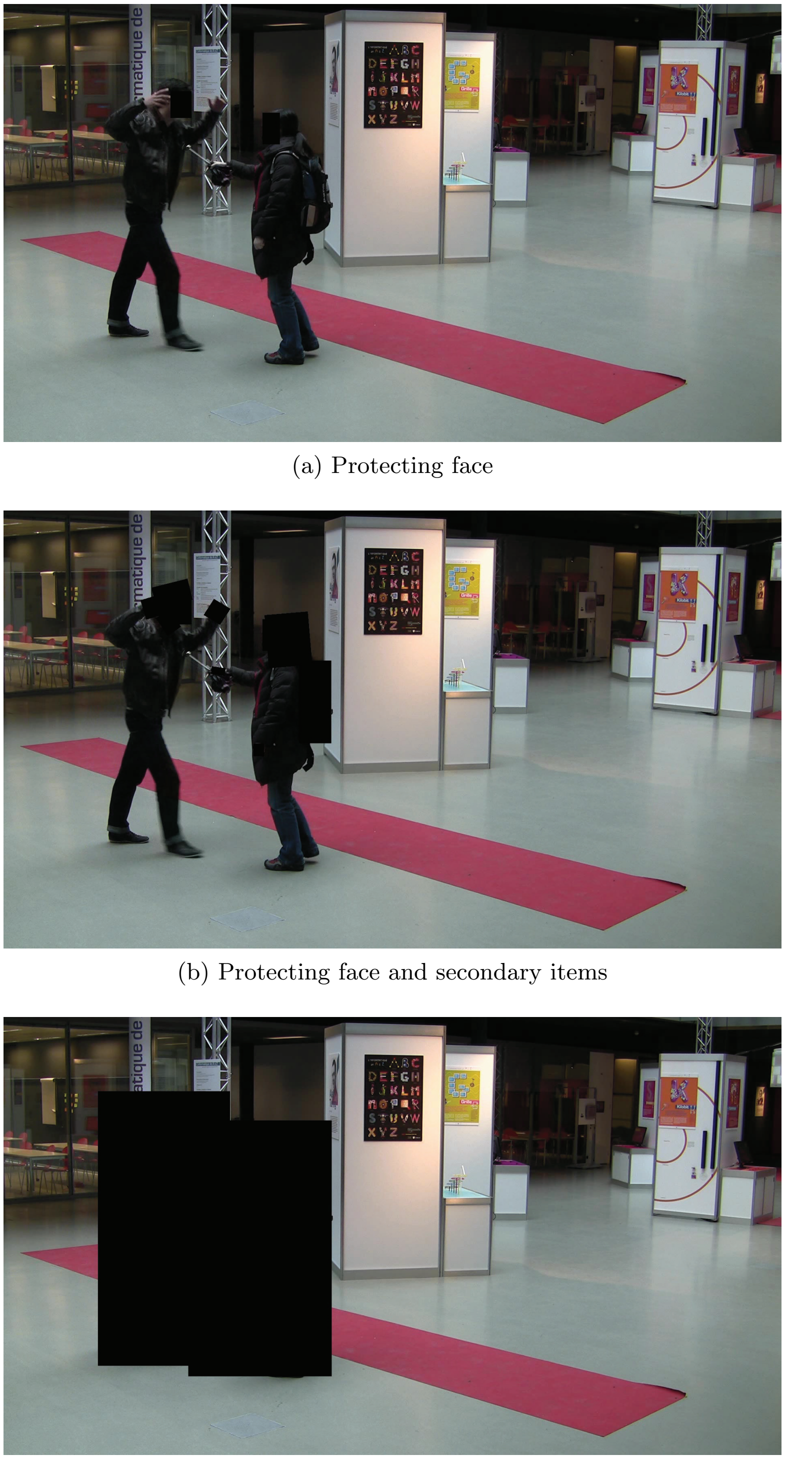
[7] Schiff, J., Meingast, M., Mulligan, D. K., Sastry, S., and Goldberg, K., "Respectful cameras: detecting visual markers in real-time to address privacy concerns," in [*International Conference on Intelligent Robots and Systems (IROS 2007)*], 971–978 (Oct 2007).

[8] Chen, D., Chang, Y., Yan, R., and Yang, J., [*Protecting privacy in video surveillance*], ch. Protecting Personal Identification in Video, 115–128, Springer-Verlag (2009).

[9] Wickramasuriya, J., Datt, M., Mehrotra, S., and Venkatasubramanian, N., "Privacy protecting data collection in media spaces," in [*Proceedings of the 12th annual ACM international conference on Multimedia*], 48–55 (Oct. 2004).

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

[11] Pande, A., Mohapatra, P., and Zambreno, J., "Securing multimedia content using joint compression and encryption," *IEEE Multimedia* **PP**(99), 1 (2012).



(a) Protecting face

(b) Protecting face and secondary items

(c) Protecting the whole body

Figure 2: Screenshot examples from the dataset with various regions protected by masking filter.