MediaEval 2013 Visual Privacy Task: Using Adaptive Edge Detection for Privacy in Surveillance Videos

Volker Eiselein
Communications Systems Group
Technische Universität Berlin
eiselein@nue.tu-berlin.de

Tobias Senst
Communications Systems Group
Technische Universität Berlin
senst@nue.tu-berlin.de

Ivo Keller
Communications Systems Group
Technische Universität Berlin
keller@nue.tu-berlin.de

Thomas Sikora
Communications Systems Group
Technische Universität Berlin
sikora@nue.tu-berlin.de

ABSTRACT

In this paper we present a system for preserving the privacy of individuals in a video surveillance scenario. While a person's privacy should not be revealed to a viewer of the video without special needs, it is still important that the action in a scene as the semantic content of a video remain perceivable by a human observer.

The proposed system uses edge detection and adaptive thresholding in order to estimate the persons' silhouettes in a video scene and thus rendering most of their actions visible, while hiding sensitive personal information. In order to obtain a more complete contour around a person, an adaptive thresholding scheme using edge histograms is used as well as background subtraction which limits the edge extraction to foreground masks and thus avoids distraction of the viewer's eyes to background structures.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues

Keywords

Privacy preservation, video analysis, obfuscation, adaptive edge detection

1. INTRODUCTION

With the increasing usage of Video Surveillance systems and a continuously growing amount of CCTV data being recorded and looked at, the need to apply privacy protection techniques in this area rises as well. In order to increase acceptance of CCTV-cameras among people being observed, it is of special importance to ensure their personal rights be not violated, while the surveillance system is still able to work and security staff is able to identify critical events in a video stream. The MediaEval 2013 Visual Privacy Task [1] addresses this issue and provides an evaluation based on the PEViD Dataset [5].

For obfuscation of personal data, we propose to show only the contour of a person in the video. In case of a low number of persons in the scene, their movements and actions could be identified while it is still impossible to identify personal details such as the person's face, color of clothes or skin color. In this paper, we show a system which uses adaptive edge detection to determine the contour of a person and in the privacy-filtered result video replaces the interior of a person's silhouette by background information.

Figure 1: original PEViD frame (left) and privacy-filtered result (right)

2. SYSTEM DESCRIPTION

The proposed system uses a background model in order to exchange the information within a person's contour by recent background information (see Fig. 1). The bounding boxes which are provided as manually-annotated ground truth for the purpose of MediaEval are used in order to identify where the privacy filter needs to be applied in the image. In these regions, edge detection is performed, and the result is blended into the background image. As the bounding boxes for objects could also be provided automatically (e.g. by human detection algorithms or trackers such as [3]), our system does not necessarily depend on manual annotations.

2.1 Double Background model

Our algorithm uses two background models: a) a standard Gaussian-Mixture Model (GMM) similar to [4] in order to obtain accurate foreground masks and b) a very simplified background model which essentially consists of only a single frame. The latter is used to maintain a very recent background information. This allows the viewer (or an analytics

algorithm) to identify e.g. objects left in the scene or graffiti sprayed on the walls. As one of the aims of our system is to give the user a clear impression of what is going on in the scene, it is crucial to provide very recent and fast background information. We therefore just use a simple on-line learning method which adapts the background pixels for the regions without objects as follows:

$$BG_i(x, y) = (1 - \alpha) \cdot BG_{i-1}(x, y) + \alpha \cdot I(x, y) \quad (1)$$

In order to adapt the background model quickly to the new frame, we use rather high values for α (e.g. $\alpha = 0.35$) and update this model every 5th frame. This also has the advantage that the system can cope quickly with slight camera movements. In some of the PEViD videos, the camera shakes a bit and the image is shifted for some pixels. While first the background image in our system gets a bit blurry in these situations, after a few frames our system maintains the current background and is not disturbed by older values anymore. The GMM uses a much slower learning rate and is only used to mask out background parts within the given region of interest (ROI) - a task for which the previously described background model b) alone would be to simple.

2.2 Adaptive Edge Detection

We use Canny edge detection [2] in order to extract the silhouettes of the persons in the image. After a noise reduction step, this algorithms computes the gradients in the image and performs thresholding with hysteresis. A common problem for this algorithm is the choice of the two thresholds T_1, T_2 used in the hysteresis step. While T_1 sets the minimum edge level accepted for the starting points, T_2 determines which edge level must be kept during the hysteresis. Usually, it is hard to set general values which give satisfying results for arbitrary video sequences.

In our system which has to work under varying conditions (indoor, outdoor, different lighting, changing weather...), we propose to adapt the threshold according to the gradient histogram in the ROI. For the given ROI, we compute the absolute values of the gradient information and set up a histogram over every ROI. Using the assumption, that most of a human's silhouette can be recovered by the highest 15 % of the gradient information, we can thus find T_1 for the Canny algorithm. In order to close holes in the silhouette, T_2 is set to $T_2 = 0.9 \cdot T_1$. This automatic choice of T_1 and T_2 adapts well to most scenes in the data set and gives a good contour information of a walking person in most cases.

3. EVALUATION RESULTS

The evaluation of the proposed filter is based on objective and subjective metrics described in [1]. The results are shown in Table 1. The proposed filter score has been compared with the average score of all 9 participants of the MediaEval Visual Privacy Task challenge. The results are very promising in terms of the privacy behavior where the adaptive edge filter is competitive with other participants.

In regards to the intelligibility and appropriateness the results of the subjective and objective evaluations related to the average of all participants are contrary. While for these categories the score of the objective results are less than the average, the score of the subjective results is higher than the average score. As the proposed filter extracts and reprojects the edge information into the image and fills the interior of

Objective evaluation		
	Adaptive edge filter	Average (9)
Intelligibility	0.358355	**0.502378**
Privacy	**0.693835**	0.664903
Appropriateness	0.368627	**0.56048**

Subjective evaluation		
	Adaptive edge filter	Average (9)
Intelligibility	**0.678333**	0.655741
Privacy	0.683750	0.683843
Appropriateness	**0.532500**	0.492130

Table 1: Evaluation results for the proposed method compared to 9 other methods in the workshop.

a person with background information, it is obvious that a standard person detection and tracking method which is usually based on extracting edges by HoG or wavelet filtering and is trained on normal videos cannot have the same performance as in unfiltered images. To apply the filtered videos for automated video analytics the respective preprocessing of the methods has to be adapted and we thus recommend a new training step based on feature extraction on the filtered images. However, based on the results of the subjective evaluation, the filter is especially appropriate for display in video surveillance systems that are guided and evaluated by human operators.

4. CONCLUSIONS

In this paper we showed how adaptive edge detection can be used to preserve the privacy of people in CCTV videos. The subjective test demonstrates that the proposed filter outperforms most of the other privacy filters proposed in the MediaEval 2013 challenge. Thus, our method is of special interest for semi-automatic video surveillance systems which need to hide sensitive personal information while preserving the context of actions.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] A. Badii, M. Einig, and T. Piatrik. Overview of the mediaeval 2013 visual privacy task. 2013.

[2] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698, 1986.

[3] V. Eiselein, T. Senst, I. Keller, and T. Sikora. A motion-enhanced hybrid probability hypothesis density filter for real-time multi-human tracking in video surveillance scenarios. In *15th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, pages 6–13, 2013.

[4] R. Heras Evangelio, T. Senst, and T. Sikora. Detection of static objects for the task of video surveillance. In *IEEE Workshop on Applications of Computer Vision (WACV)*, pages 534–540, 2011.

[5] P. Korshunov and T. Ebrahimi. PEViD: privacy evaluation video dataset. In *Proceedings of SPIE Volume 8856*, volume 8856, 2013.