Contextualized Privacy Filters in Video Surveillance Using Crowd Density Maps

Hajer Fradi, Andrea Melle, and Jean-Luc Dugelay
Multimedia Communications Dept.
EURECOM
Sophia Antipolis, France
Email: {fradi, melle, dugelay}@eurecom.fr

***Abstract*—The widespread growth in the adoption of digital video surveillance systems emphasizes the need for privacy-preservation video analytics techniques. While these privacy aspects have shown big interest in recent years, little importance has been given to the concept of context-aware privacy protection filters. In this paper, we specifically focus on the dependency between privacy preservation and crowd density. We show that additional information about the crowd density in the scene can be used in order to adjust the level of privacy protection according to the local needs. This additional information cue consists of modeling time-varying dynamics of the crowd density using local features as an observation of a probabilistic crowd function. It also involves a feature tracking step which enables excluding feature points on the background. This process is favourable for the later density function estimation since the influence of features irrelevant to the underlying crowd density is removed. Then, the protection level of personal privacy in videos is adapted according to the crowd density. Afterwards, a framework for objective evaluation of the contextualized protection filters is proposed. The effectiveness of the proposed context-aware privacy filters has been demonstrated by assessing the intelligibility vs. privacy trade-off using videos from different crowd datasets.**

***Keywords*-Privacy, Protection Filters, crowd density, local features, intelligibility, detection**

I. INTRODUCTION

In recent years, a widespread growth in the adoption of digital video surveillance systems for monitoring of buildings and public spaces has been observed. In this context, several concerns have been raised related to the possibility of infringing the privacy rights of the subjects being monitored [1]. At the same time, the adoption of automated methods for the analysis of video surveillance data has raised additional concerns, since algorithms such as face recognition or people re-identification could potentially expose the identity of any individual under video surveillance at any time [2].

Current video surveillance systems either do not implement any mechanism for privacy protection, or they use naïve approaches, for instance uniformly applying simple filters (e.g. masking, Gaussian blur, and pixelization) to some regions of the image which contain privacy sensitive information, such as faces or license plates. The lack of specific methods to detect privacy sensitive regions of interest and to evaluate the amount of privacy protection required in a specific scenario often causes failure in either minimizing the intrusion of the surveillance system or goes against the purpose of the surveillance itself.

One big challenge in defining privacy protection policies for video surveillance applications is the identification of the correct trade-off between intelligibility of the video, which should be adequate to the monitoring tasks, and privacy protection itself. Consequently, a number of recent studies have been conducted to propose more adequate systems for privacy protection.

In this perspective, the concept of context-aware privacy protection has emerged, as the notion that the amount of privacy protection required is deeply linked to the context of the scene and purpose of the monitoring activity. A context-dependent approach to privacy protection is described in [3], where image processing and scene understanding techniques are employed to automatically evaluate the context in which video surveillance takes place, in order to apply context-specific privacy rules. This approach is based on scene and object detection algorithms such as bag-of-visual-words, people tracking and gait analysis in order to recognize specific sub-contexts which require the application of different privacy protection rules. In [4], the authors propose another context-aware surveillance system, where the situation within an environment is interpreted by combining a number of contextual information, which are then used to determine an appropriate level of privacy. Six levels of privacy protection ranging from high to low are proposed, and their application is based on the analysis of visual features such as global motion in the scene and detection-based crowd size estimation.

As employed in [4], the crowd size (or more precisely the number of people in the scene) can be an important indication of which events are expected and therefore which privacy level is suitable in the scene. If we take crowd management as an exemplary standard task within the field of video surveillance, video operators need clear visual information in crowded regions. Mainly in case of abnormal events such as potential overcrowding or dangerous motion patterns, a video operator should be able to perceive the maximal information for early detection of unusual situations in large scale crowd to ensure assistance and emergency contingency plan and to decide if an intervention by security forces is needed. At the same time, the more people are present around a site, the less perceivable and identifiable is a single individual. It is therefore reasonable in many applications to reduce the privacy filtering level in crowded areas compared to spaces composed of isolated individuals.

In this paper, we propose a system which is able to choose a suitable level of privacy according to a crowd density measure. In the simplest form, the used crowd density measures could be the number of persons [5], [6] or the crowd level [7], [8]. However, these measures have the limitation of giving only global information for the entire image and discarding local information about the crowd.

We therefore resort to another crowd density measure, in which local information at pixel level substitutes a global number of people or a crowd level per frame. The alternative solution based on computing crowd density maps is indeed more appropriate as it enables both the detection and the location of potentially crowded areas. It is typically based on using local features as an observation of a probabilistic crowd function. A feature tracking step is also involved in the process to alleviate the effects of feature components irrelevant to the underlying crowd density. Our following objective is then to use these results in order to build adaptive privacy protection filters, in which the privacy level gradually decreases with the crowd density.

As an additional contribution of this paper, we identify a framework for objective evaluation, which enables assessing the intelligibility vs. privacy balance based on the performances of state-of-art video surveillance analysis algorithms. In our experiments, we intend to demonstrate that the proposed contextualized privacy protection filters are resistant to local features-based person matching algorithms, which potentially threaten one's individual privacy, while still preserving those visual features which are fundamental for automated crowd analysis tasks such as people detection and counting.

The remainder of the paper is organized as follows: we introduce our proposed approach for crowd density map estimation in Section II. Section III shows then how the crowd density information is incorporated into a privacy protection framework which alters the data protection level accordingly. The objective evaluation framework and results using two privacy filters and different video sequences are given in Section IV. Finally, we briefly conclude in Section V.

II. Crowd Density Estimation

Crowd density analysis has been studied as a major component for crowd monitoring and management in visual surveillance systems. In this paper, we explore a new promising application of crowd density measures in privacy context. An illustration of the proposed crowd density map modules [9] is shown in Figure 1. The remainder of this section describes each of these system components.

A. *Extraction of local features*

One of the key aspects of crowd density measurements is crowd feature extraction. Under the assumption that regions of low density crowd tend to present less dense local features compared to a high-density crowd, we propose to use local feature as a description of the crowd by relating dense or sparse local features to the crowd size. For this purpose, the crowd density map is estimated by measuring how close local features are.

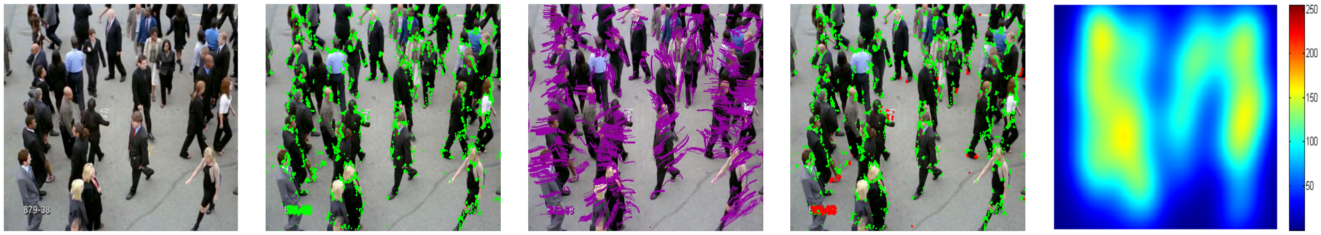
For local features, we assess Features from Accelerated Segment Test (FAST) [10], Scale-Invariant Feature Transform (SIFT) [11], and Good Features to Track (GFT) [12]. The reason behind selecting these features for crowd measurement is as follows: FAST was proposed for corner detection in a reliable way. It has the advantage of being able to find small regions which are outstandingly different from their surrounding pixels. Besides in [13], FAST was used to detect dense crowds from aerial images and the derived results demonstrate a reliable detection of crowded regions using FAST. SIFT is another well-known texture descriptor, for which interest point locations are defined as maxima/minima of the difference of Gaussians in scale-space. Under this respect, SIFT is rather independent of the perceived scale of the considered object which is appropriate for crowd measurements. These two aforementioned features are compared to the classic feature detector GFT, which is based on the detection of corners containing high frequency information in two dimensions and typically persist in an image despite object variations.

B. *Local features tracking*

Using the extracted features directly to estimate the crowd density map without a feature selection process might incur at least two problems: Firstly the high number of local features increases the computation time of the crowd density. As a second and more important effect, the local features contain components irrelevant to the crowd density. Thus, we need to add a separation step between foreground and background entities to our system. This is done by assigning motion information to the detected local features in order to distinguish between moving and static ones. Based on the assumption that only persons are moving in the scene, these can then be differentiated from background by their non-zero motion vectors.

Motion estimation is performed using the Robust Local Optical Flow (RLOF) [14], which computes accurate sparse motion fields by means of a robust norm[1]. The motion vector $\mathbf{d}$ is computed by a minimization of the shrinked Hampel norm with the parameters σ_1, σ_2 defining the treatment of

[1] www.nue.tu-berlin.de/menue/forschung/projekte/rlof



(a) (b) (c) (d) (e)

Figure 1. Illustration of the proposed crowd density map estimation using local features extraction: (a) Exemplary frame, (b) FAST Local features (c) Feature tracks, (d) Distinction between moving (green) and static (red) features - red features at the lower left corner are due to text overlay in the video, (e) Estimated crowd density map

outliers:

$$\rho(y, \boldsymbol{\sigma}) = \begin{cases} y^2 & , |y| \leq \sigma_1 \\ \sigma_1\sigma_2 & , |y| \geq \sigma_2 \\ \frac{\sigma_1}{\sigma_1 - \sigma_2}(|y| - \sigma_2)^2 + \sigma_1\sigma_2 & , \text{else} \end{cases} . \quad (1)$$

A common problem in local optical flow estimation is the choice of feature points to be tracked. Depending on texture and local gradient information, these points often do not lie on the center of an object but rather at its borders and can thus be easily affected by other motion patterns or by occlusion. While RLOF handles these noise effects better than the standard Kanade-Lucas-Tomasi (KLT) feature tracker from [15], it is still not prone against all errors. This is why we establish a forward-backward verification scheme where the resulting position of a point is used as input to the same motion estimation step from the second frame towards the first one. Points for which this “reverse motion” does not result in their respective initial position are discarded. For all other points, motion information is aggregated to form longterm trajectories.

In every temporal step, the overall mean motion m_t of a trajectory t is compared to a certain threshold β which is set according to image resolution and camera perspective. Moving features are then identified by the relation $m_t > \beta$ while the others are considered as part of the static background.

The advantage of using trajectories in this system instead of computing the motion vectors only between two consecutive frames is that outliers are filtered out and the overall motion information is less affected by noise. As a result, the separation between foreground and background entities is improved and the number and position of the tracked features undergo an implicit temporal filtering step which makes them smoother.

C. *Kernel density estimation*

After generating trajectories to filter out static features, we define the crowd density map as a kernel density estimate based on the positions of local features. Starting from the assumption of a similar distribution of feature points on the objects, the observation can be made that the more local features come towards each other, the higher crowd density is perceived. For this purpose, a probability density function (pdf) is estimated using a Gaussian kernel density.

For a given video sequence of N frames $\{I_1, I_2, ..., I_N\}$, if we consider a set of m_k local features extracted from a frame I_k at their respective locations $\{(x_i, y_i), 1 \leq i \leq m_k\}$, the corresponding density map C_k is defined as follows:

$$C_k(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=1}^{m_k} \exp -(\frac{(x - x_i)^2 + (y - y_i)^2}{2\sigma^2}) \quad (2)$$

where σ is the bandwidth of the 2D Gaussian kernel.

The resulting crowd density map characterizes the spatial and temporal variations of the crowd which convey rich information about the distributions of pedestrians in the scene.

III. Incorporation of Crowd Density Measure in a Privacy-Preservation Framework

In this Section, we propose to apply crowd density information for context-aware privacy purposes. In particular, the proposed crowd density measure described in Section II is employed to adjust the level of privacy protection according to the local needs. The reason behind that is to hide personal information to the video operator without preventing him to be able to identify potential dangerous areas and events. A simple way for that could be to just use crowd density directly as an input to a privacy filter in such way that the obfuscation level depends directly on the density of a given region. This method could substantially decrease the visibility of potentially important information since all crowded areas would be obscured. Because of that, we restrict the application of privacy preservation filers to some regions of interest, i.e. only regions that contain personal information are obfuscated. These could include face, clothing, skin/hair color or even gait depending on the scene context. Given this variety and considering that these information is not perceivable under all circumstances (e.g. heavy crowding, different lighting conditions, motion blur, low contrast, low resolution...), in our work we consider head obfuscation as the most visible part of a human in a crowd.

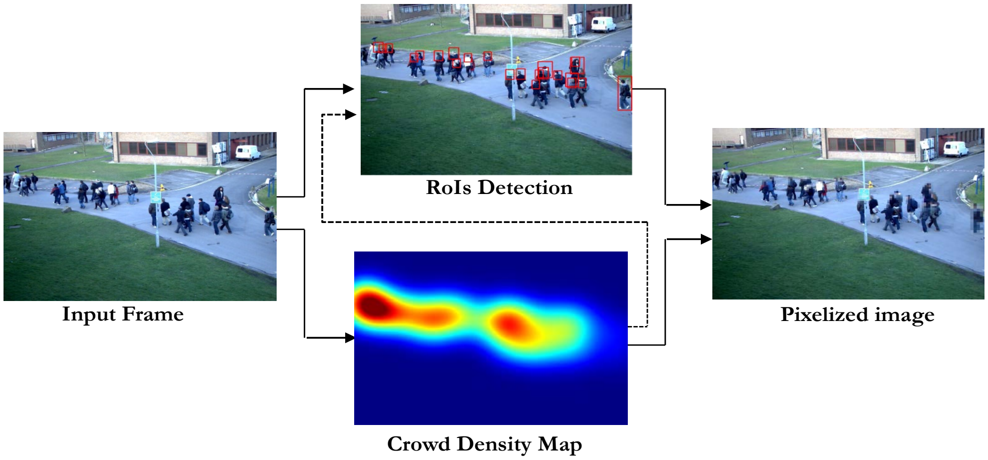


Figure 2. Flowchart of the proposed contextualized privacy preservation filters using an examplary frame from PETS 2009 [16], the dotted line in this figure shows that the crowd density map can also be used to improve the robustness of the detection in crowded scenes

However, once a person has left the crowd and is perceived as an isolated subject, more information has to be hidden. This is why in these cases we extend the obfuscated region to the whole body in order to hide details such as clothing or skin color from the viewer.

As a measure for privacy protection, the level of obfuscation is adapted according to the crowd density for the following reasons: Crowds are usually interesting to video operators as they are a common place for crimes or for dangerous overcrowding events. At the same time, people in a crowd exhibit a smaller amount of information to a video operator, thus they do not have to be filtered to the same degree as for isolated people who are entirely visible. We therefore propose to lower the level of privacy protection within a crowded area. The flowchart of the proposed contextualized privacy protection filers is shown in Figure 2. In the following, we describe the RoIs detection and adaptive filters system components.

A. *RoIs detection*

To obfuscate people in the scene, we apply an additional RoI detection step using the deformable part-based models [17]. Firstly proposed in [18], Histograms of Oriented Gradients (HOG) extracts gradient information from a detection window, derives a feature vector from it and compares it against annotated samples. Then, HOG is extended to the deformable part-based models which achieves much more accurate results than the original HOG and marks the state-of-the-art. The detector uses a feature vector over multiple scales and a number of smaller parts to get additional information about an object.

While human detection using the deformable part-based models has become a quite popular technique, its extension to crowded scenes has a limited success. In fact, the density of people substantially affects their appearance in video sequences. Especially in dense crowds, people occlude each other and only some parts of each individual's body are visible. Therefore, robust human detection in such scenarios with frequent occlusions and high interactions among the targets remains a challenge.

Out of the scope of this paper, the crowd density map can also be used to improve the robustness of the detection in crowded scenes (dotted line in Figure 2). This improvement has been proposed in [19], where a scene-adaptive dynamic parametrization using this crowd density measure is applied. In addition, the proposed extension of deformable part models to crowds includes a self-adaptive geometrical filtering in order to reduce false positive detections in crowded areas. As demonstrated in [19], by integrating the crowd density and geometrical constraints, the detection results are enhanced considerably.

In our framework, we employ this improved deformable part models proposed in [19] to detect people in crowded scenes. Then, for people obfuscation, we apply adaptive privacy preservation filters to only the head part or to the whole body depending if the target is isolated or within the crowd. More details about the adaptive protection filters in given in the next paragraph.

B. *Adaptive privacy filters*

After applying person detection, we get a set of RoIs $\mathcal{D}_k = \{d_1^k, ..., d_{n_k}^k\}$ at frame I_k, d_j^k denotes the j^{th} detection at this frame and is defined as $d_j^k = \{x_j^k, y_j^k, w_j^k, h_j^k\}$, where (x_j^k, y_j^k) denotes the top left corner position and w_j^k, h_j^k the respective width and height. Given also the crowd density map $C_k(x, y)$ that shows information about the crowd size and the crowd location as well, our goal is to adapt the level of the privacy protection filters according to the crowd density. More precisely, as explained before we intend to use high privacy protection in less crowded areas while reducing the level of privacy protection in areas with many people. For this purpose, given a set of filter parameters representing different obfuscation levels $P = \{P_{min}, ..., P_{max}\}$, we quantify the crowd density values into $c = |P|$ crowd levels. Then, for a given detection d_j^k, its average crowd density value $\hat{C}_k(d_j^k)$ is used to choose the respective filter parameter that has to be applied to the bounding box d_j^k:

$$\hat{C}_k(d_j^k) = \frac{\sum\limits_{p=0}^{h_j^k-1} \sum\limits_{q=0}^{w_j^k-1} C_k(x_j^k + p, y_j^k + q)}{w_j^k \cdot h_j^k} \tag{3}$$

In addition to the crowd density, the visibility of a person in the scene is also sensitive to his distance from the camera because of perspective effects. The perspective distortions can be explained by the fact that persons far away from the camera appear smaller than the closest ones. Thus, the distance from the camera is another parameter that has to be taken into account to choose the suitable obfuscation level. To achieve that, the range of obfuscation levels given by the lower and upper boundary P_{min}/P_{max} is adapted according to the distance from the camera. A simple method to interpret the distance from the camera is to use the size of the detected bounding box. Since this information could be subject to errors, a better method consists of computing the aspect ratio and the perceived height of a person from all accepted detections. Following [20], we assume the relationship between a person's position and his perceived height to be:

$$\widetilde{h}_j^k = \alpha_{k-1} \cdot y_j^k + \beta_{k-1}, j \in \{1...n_k\} \tag{4}$$

where α_{k-1} and β_{k-1} parameters are learned using a standard regression. And the aspect ratio is defined as:

$$\gamma_{k-1} = median \left\{ \frac{w_j^i}{h_j^i} \right\}_{1 \leq i \leq (k-1), 1 \leq j \leq n_i} \tag{5}$$

γ_{k-1}, α_{k-1}, and β_{k-1} parameters are computed from all accepted detections $\{\mathcal{D}_1, ..., \mathcal{D}_{k-1}\}$ and updated at each frame. Using this method, we are able to predict the height and the ratio of a detection from the previous detections. Thus, the estimated size of a bounding box d_j^k is $\widetilde{S}_j^k = (\widetilde{h}_j^k)^2 * \gamma_{k-1}$ which is more robust than $w_j^k * h_j^k$.

In this paper, we show results for two typical privacy protection filters which are:

Gaussian Blurring: This privacy filter consists essentially of removing details in a region of interest by applying Gaussian low pass filtering.

$$I_{blur}^k(x, y) = I_k(x, y) * \frac{1}{2\pi {\sigma_{k,j}}^2} e^{\frac{(x^2+y^2)}{2{\sigma_{k,j}}^2}} \tag{6}$$

For this technique, the bandwidth $\sigma_{k,j}$ of the used Gaussian is adapted according to the crowd density level and the predicted size.

Pixelization: This filter is based on decreasing the resolution of any region of interest by replacing each block of pixels in this area with its respective average. The pixelization of frame I_k corresponding to d_j^k detection is given by:

$$I_{pix}^k(x, y) = \frac{1}{b_{k,j}^2} \sum_{i=0}^{b_{k,j}-1} \sum_{j=0}^{b_{k,j}-1} I\left(\left\lfloor \frac{x}{b_{k,j}} \right\rfloor + i, \left\lfloor \frac{y}{b_{k,j}} \right\rfloor + j \right) \tag{7}$$

As for the blurring process, the filter size $b_{k,j} \propto (\hat{C}_k(d_j^k), \widetilde{S}_j^k)$.

For both pixelization and Gaussian blurring, the region of interest in restricted to head part only if the person is moving inside the crowd, i.e: if $\hat{C}_k(d_j^k) \leq \tau$, $x \in [x_j^k...x_j^k + w_j^k - 1]$ and $y \in [y_j^k...y_j^k + h_j^k - 1]$, otherwise $x \in [x_j^k + \Delta_x...x_j^k + w_j^k - \Delta_x - 1]$ and $y \in [y_j^k...y_j^k + h_j^k - \Delta_y - 1]$, where Δ_x and Δ_y parameters are used to crop the head part from the detected bounding box d_j^k.

IV. Experimental Results

A. Datasets and Experiments

The proposed framework is evaluated within challenging crowd scenes from multiple video datasets. In particular, we selected some videos from PETS 2009 [16], UCF [21] and Data Driven Crowd Analysis [22] public datasets. To evaluate our proposed context-dependent privacy protection, we adopt an objective evaluation framework, by studying the variation in performances of the state-of-the-art algorithms commonly used in video surveillance analytics before and after applying the proposed privacy protection filters. We recall, as mentioned in the introduction, that one of the major challenges in defining privacy protection policies lies in identifying the correct balance between the two axis of intelligibility and privacy protection of the surveillance data. Therefore, our evaluation framework will consider both axis and model each of them based on the performance scores of an appropriate algorithm.

We model the impact of privacy filters on intelligibility by evaluating the performances of a people counting-by-detection algorithm before and after applying the protection filters. We motivate our choice observing that privacy protected video surveillance footage must at least retain those visual features necessary to perform very basic monitoring tasks such as people detection and counting.

To evaluate the amount of privacy guaranteed by our method, we model privacy as inverse score of a person matching algorithm based on local features. Such algorithm tries to identify an individual among a set of other subjects by extracting and matching local features between a gallery and a probe set. This algorithm represents a common step for higher level tasks such as person re-identification, recognition or tracking, which could potentially reveal information on the identity of a subject. In our implementation, we use Hessian-Laplace interest point detector together with the SIFT descriptor and nearest neighbor matching, based on the efficient approximate implementation of [23]. Details of the people matching algorithm, together with an extensive evaluation of the different feature extraction and description approaches suitable to the task can be found in [24]. Based on such premise, a good privacy filter should prevent the person matching algorithm to correctly detect and describe local features.

In both cases of intelligibility and privacy, we are only interested in the relative change of performances from the original unprotected images, which constitutes the baseline for privacy filter evaluation. We adopt people counting score as a measure of intelligibility, and one minus person matching score as a measure of privacy protection.

B. Results and Analysis

In Figure 3, the results using three frames from different videos are shown. In the first and the second columns we show the results of RoIs detection, and the estimated crowd density maps. These two sources of information are combined for adaptive protection filters. For this purpose, two privacy protection tools (blurring, and pixelization) are employed to show different ways to protect personal privacy in video sequences. In this figure, it is visible that the block size in the pixelization filter and the bandwidth of the Gaussian blurring are changed by our system according to the crowd density value and perceived size of the person. Comparing e.g. the woman in the lower right corner of the first image row, to the persons walking in the crowd, it is well perceivable that the privacy level is reduced within the crowd by a smaller block size or a smaller bandwidth respectively. At the same time, it can be seen that this woman compared to groups of people walking in the crowd does not generate such a high density measure and is consequently obfuscated to a higher degree on the whole body. We also note that the estimated crowd density is lower for the second scene (second row), compared to the first one.That justifies why people in the second scene show rather higher protection levels.

Again, different filter sizes can be seen also using UCF frame (third row). However, in general, the blurring filter seems to be better suited for our application as in general already small block sizes are sufficient in the pixelization filter to render it completely unrecognizable to humans. Nonetheless, our results indicate clearly that crowd density maps are well-suited to improve the crowd context-specific privacy protection in CCTV systems and thus offer a lot of options for further applications.

Following the described evaluation procedure, we tested counting and matching on original and privacy protected sequences of PETS, Inria, and UCF datasets. Figure 4 reports the people counting results for blurring and pixelization protection techniques for the different types of features used in crowd density estimation respectively. Since we are interested in evaluating the task of counting people before and after applying a privacy filter, rather than the effectiveness of the counting algorithm itself, a simple evaluation score is chosen, i.e. the percentage $p \in [0, 1]$ of correctly detected individuals with respect to the annotated ones in the ground truth. The red horizontal line represents the counting score when no protection filter has been applied. As a general trend, we can observe that the counting results do not decrease significantly after applying the protection filters. On average, the score drop is 0.10, with 0.03 representing the minimum and 0.18 the maximum loss observed respectively for the blur filter with the SIFT feature and the pixelization filter with the GFT feature. As a consequence, we are still able to correctly perform people counting within a 10% error margin. We also notice that the pixelization algorithm causes the counting to perform worse than the blurring algorithm.

Matching results are displayed in Figure 5, following a convention similar to the previous one for detection. In this case, the red horizontal line represents the baseline matching result when no filter is applied. We can clearly observe a dramatic drop in performances of the person matching algorithm. On average, the drop in matching score is 0.41, with 0.42 and 0.39 being the minimum and maximum observed loss respectively for the pixelation filter with the FAST feature and for the blur filter with the GFT feature.

These results confirm that our approach to privacy protection behaves in accordance to requirements we mentioned in the introduction, in terms of preservation of intelligibility and privacy of the original source. Our privacy protection filters cause a relatively small loss in people counting score, and therefore in intelligibility, compared to the drop in performances of the matching step, and thus the gain in privacy protection level.

We notice as well how in both the counting and the matching experiment, the exact choice of feature does not influence the results significantly, while it is rather the choice of protection filter which causes variations in the results. Observing that the counting scores, and therefore the intelligibility, is significantly worse in the case of pixelization, and at the same time pixelization offers slightly less privacy protection (higher matching results), the choice of filter falls back on the specific application scenario, according to the desired privacy-intelligibility trade off.

V. Conclusion

In this paper, we show how it is possible to include crowd density information into a privacy-preserving framework. Using an additional RoIs detection step, we adapt the degree of data obfuscation for privacy according to the crowd level. By doing that, it is possible to preserve an acceptable level of privacy for the people in a scene while still allowing the operator to view the data relevant for him. As an additional contribution, we proposed an objective evaluation of privacy and intelligibility trade-off offered by our proposed contextualized privacy protection filters. By leveraging the state-of-the-art video surveillance analysis algorithms, such as people counting and matching, we demonstrate that our privacy filters retain good performances on common intelligibility tasks such as people counting and detection. At the same time, such privacy filters are able to significantly lower the performances of person matching algorithms based on local features, which potentially can expose identity information of the subject being monitored, therefore threatening their privacy. Furthermore, our evaluation shows that the choice of blur over pixelization as the preferred obfuscation method leads to a better privacy-intelligibility balance.

Acknowledgment

This work has received funding under the VideoSense project which is co-funded by the European Commission under the 7th Framework Programme Grant Agreement Number 261743.

References

[1] V. Norris, M. McCahill, and D. Wood, "Editorial: The growth of CCTV: a global perspective on the international diffusion of video surveillance in publicly accessible space," *Surveillance and Society*, vol. 2(2/3), pp. 110–135, 2004.

[2] A. Senior, S. Pankanti, A. Hampapur, L. Brown, Y.-L. Tian, A. Ekin, J. Connell, C.-F. Shu, and M. Lu, "Enabling video privacy through computer vision," *Security Privacy, IEEE*, vol. 3, no. 3, pp. 50–57, 2005.

[3] A. Badii, M. Einig, M. Tiemann, D. Thiemert, and C. Lallah, "Visual context identification for privacy-respecting video analytics," in *Multimedia Signal Processing (MMSP), 2012 IEEE 14th International Workshop on*, 2012, pp. 366–371.

[4] S. Moncrieff, S. Venkatesh, and G. West, "Context aware privacy in visual surveillance," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, 2008, pp. 1–4.

[5] H. Fradi and J.-L. Dugelay, "People counting system in crowded scenes based on feature regression," in *EUSIPCO 2012, European Signal Processing Conference, August, 27-31, 2012*, 08 2012.

[6] H. Fradi and J. L. Dugelay, "Low level crowd analysis using frame-wise normalized feature for people counting," in *IEEE International Workshop on Information Forensics and Security*, December 2012.

[7] H. Yang, H. Su, S. Zheng, S. Wei, and Y. Fan, "The large-scale crowd density estimation based on sparse spatiotemporal local binary pattern," *IEEE International Conference on Multimedia and Expo*, pp. 1–6, 2011.

[8] H. Fradi, X. Zhao, and J. L. Dugelay, "Crowd density analysis using subspace learning on local binary pattern," in *ICME 2013, IEEE International Workshop on Advances in Automated Multimedia Surveillance for Public Safety*, July 2013.

[9] H. Fradi and J.-L. Dugelay, "Crowd density map estimation based on feature tracks," in *MMSP 2013, 15th International Workshop on Multimedia Signal Processing, September 30-October 2, 2013*, 09 2013.

[10] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, pp. 105–119, 2010.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," in *Int. J. Comput. Vision*, 2004, pp. 91–110.

[12] J. Shi and C. Tomasi., "Good features to track," in *CVPR*, 1994, pp. 593 –600.

[13] M. Butenuth, F. Burkert, F. Schmidt, S. Hinz, D. Hartmann, A. Kneidl, A. Borrmann, and B. Sirmacek, "Integrating pedestrian simulation, tracking and event detection for crowd analysis," pp. 150–157, 2011.

[14] T. Senst, V. Eiselein, and T. Sikora, "Robust local optical flow for feature tracking," *Transactions on Circuits and Systems for Video Technology*, vol. 09, no. 99, 2012.

[15] C. Tomasi and T. Kanade, "Detection and tracking of point features," CMU, Technical Report CMU-CS-91-132, 1991.

[16] J. Ferryman and A. Shahrokni, "Pets2009: Dataset and challenge," in *PETS*, 2009, pp. 1–6.

[17] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.

[18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *CVPR*, vol. 2, 2005, pp. 886–893.

[19] V. Eiselein, H. Fradi, I. Keller, T. Sikora, and J.-L. Dugelay, "Enhancing human detection using crowd density measures and an adaptive correction filter," in *IEEE International Conference on Advanced Video and Signal-Based Surveillance*, 2013.

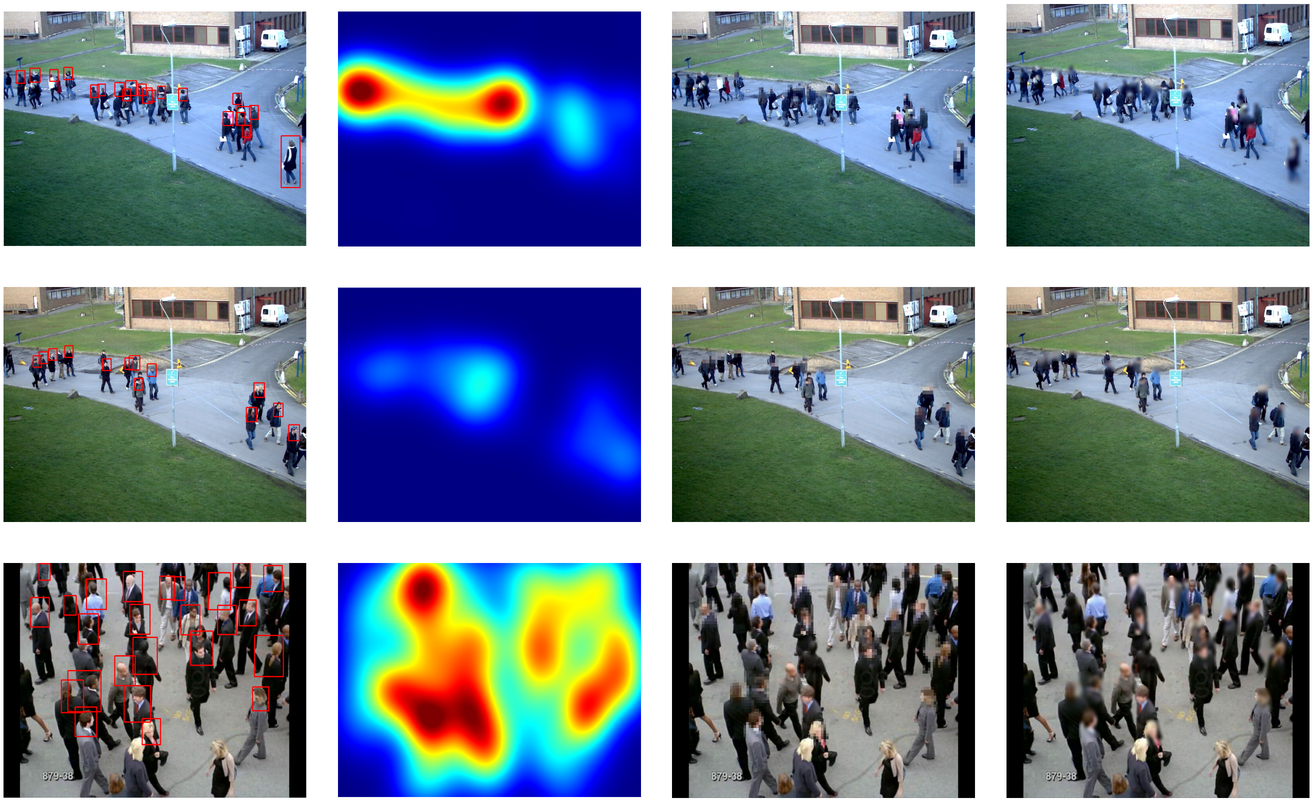
[20] D. Hoiem, A. A. Efros, and M. Hebert, "Putting objects in perspective," *International Journal of Computer Vision*, vol. 80, no. 1, pp. 3–15, 2008.

[21] S. Ali and M. Shah, "A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis," in *CVPR 07*, 2007, pp. 1–6.

[22] M. Rodriguez, J. Sivic, I. Laptev, and J.-Y. Audibert, "Data-driven crowd analysis in videos," in *ICCV*, 2011.

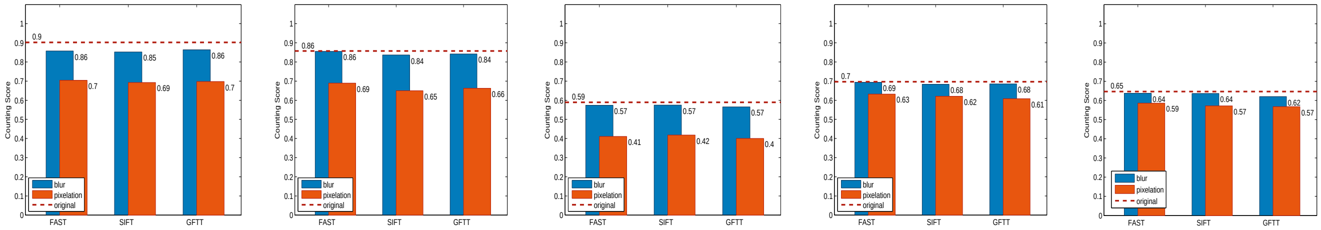
[23] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *In VISAPP International Conference on Computer Vision Theory and Applications*, 2009, pp. 331–340.

[24] M. Bauml and R. Stiefelhagen, "Evaluation of local features for person re-identification in image sequences," in *AVSS*, 2011, pp. 291–296.



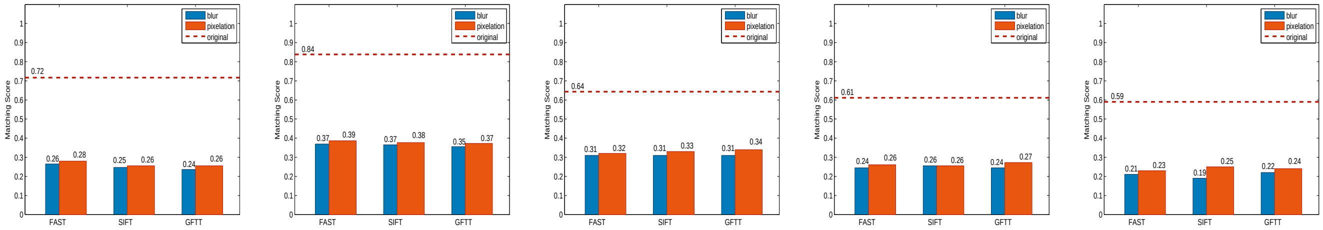
(a) Head detections (b) Crowd density map (c) Pixelized image (d) Blurred image

Figure 3. Results of adaptive protection filters using three frames from different test videos. From top to down order: PETS2009 S1.L1 1357.V1, PETS2009 S1.L1 1359.V1, and UCF 879. From left to right order: RoIs detection, estimated crowd density map (the color map Jet is used so red values represent higher density where blue values represent low density), application of pixelization filter, application of blurring filter



(a) PETS S1.L1 1359 (b) PETS S1.L1 1357 (c) PETS S2.L3 1431 (d) UCF 879 (e) INRIA 879 42

Figure 4. Counting scores on sequences protected by blur and pixelization, compared to original results



(a) PETS S1.L1 1359 (b) PETS S1.L1 1357 (c) PETS S2.L3 1431 (d) UCF 879 (e) INRIA 879 42

Figure 5. Matching scores on sequences protected by blur and pixelization, compared to original results