# Crowd Context-Dependent Privacy Protection Filters

Hajer Fradi [1#], Volker Eiselein [2*], Ivo Keller [*], Jean-Luc Dugelay [#], Thomas Sikora [*]

[#] *Multimedia Communications Dept., EURECOM, Sophia Antipolis, France*
[1] E-mail: fradi@eurecom.fr

[*] *Communication Systems Group, Technische Universität Berlin, Germany*
[2] E-mail: eiselein@nue.tu-berlin.de

***Abstract*—While various privacy protection filters have been proposed in the literature, little importance has been given to the context relevance of these filters. In this paper, we specifically focus on the dependency between privacy preservation and crowd density. We show that information about the crowd density in a scene can be used in order to adjust the level of privacy protection according to the local needs. For the estimation of density maps, we use an approach based on FAST feature extraction and local optical flow computation which allow excluding feature points on the background. This process is favorable for the later density function estimation since the influence of features irrelevant to the crowd density is removed. Afterwards, we adapt the protection level of personal privacy in videos according to the crowd density. The effectiveness of the proposed framework is evaluated with videos from different crowd datasets.**

***Index Terms*—Crowd density, local features, tracking, privacy filters**

## I. Introduction

In recent times, video surveillance has developed from a rather abstract, research-related topic to a key technology for modern society. Camera prices are dropping while their performance is increasing steadily. Analytics algorithms are becoming more intelligent while CCTV is ubiquitous in almost all public areas. On the one hand, this poses a number of interesting technical challenges, but then it also emphasizes the need for privacy-preserving video analytics techniques.

Privacy aspects in video surveillance systems have already been discussed in different approaches. In [1], [2] and [3] extensive overviews of general requirements such as the need for integrity, confidentiality or access authorization are given. In [4], Cavallaro points out how the ongoing changes towards digital CCTV footage leads to easier storing, transmission and analysis of video data compared to earlier years. This also enables CCTV network operators to choose which analysis tasks have to be run in real-time and which can be done on stored video data as not all tasks have to be carried out in all scenes and contexts. Consequently, [4] proposes to use a privacy-by-design approach in which smart cameras split the recorded data into a behavioral part and a part containing personal data. From this splitting point on, a video operator can only access the behavioral part while personal data is maintained confidential and only stored in a video archive in order to allow a later access for police and law enforcement agencies (if this is needed and permitted by jurisdiction).

A fundamentally new approach has been given by the concept of scene-dependent privacy levels. It is a natural and intuitive idea that a specific human action in a video has to be considered according to the scene context. As a simple example, detection of fireworks in a train station on a normal day would be an unusual and potentially dangerous event but can be mostly considered normal in an outdoor scene on New Year's Eve. While in this case, simply time and site information is taken into account, in [5], the authors propose to adapt the privacy level according to the nature of the detected events. Taking an example of crowd management, normally only the number of people in the scene and their motion is of interest to a video operator. However, in case of severely abnormal events such as potential overcrowding or dangerous motion patterns, the operator has to decide if an intervention by security forces is needed and should thus be able to perceive the maximal information possible. Badii *et al.* show in [6] how this model can be extended even further by inclusion of more dynamic information such as gait analysis or human tracking.

The number of people in the scene can be an important indication of what events are to be expected and therefore also 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 which areas of the scene overcrowding or potentially dangerous crowd movements occur. Also crimes such as pickpocketing or violence in demonstrations are more likely to happen when a scene is crowded. 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 level in crowded areas compared to spaces with isolated individuals.

In the following sections, 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 [7], [8], [9] or the crowd level [10], [11], [12]. 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 a crowd information at a local level by computing crowd density maps. This alternative solution is indeed more appropriate because it enables both the detection and the location of potentially crowded areas. The estimation of crowd density maps is typically based on extracting local features.

In our work, this approach is extended to feature tracking as well and enables us to identify which objects in the scene have undergone a sufficient motion to be considered as a person. Consequently, the effort of computation is reduced to the features relevant for 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.

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 we incorporate the crowd density information into a privacy protection framework which alters the data protection level accordingly. Experimental results for two privacy filters and several video sequences are given in Section IV. Finally, we briefly conclude and give some potential for future works in Section V.

## II. Crowd Density Map 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 application of crowd density measures in privacy context. From this perspective, generating locally accurate crowd density maps is more helpful than computing only an overall density or the number of people in a whole frame. Using our approach, local information at pixel level substitutes global, per-frame information. To achieve this goal, we compute crowd density maps using FAST local features as an observation of a probabilistic crowd function.

In the following, our proposed approach for crowd density estimation is presented. First, local features are extracted to infer the contents of each frame under analysis. Then, we perform local feature tracking using the Robust Local Optical Flow algorithm from [13] and a point rejection step using forward-background projection. Building trajectories from these data enables us also to remove static features which can be considered background. Finally, crowd density maps are estimated from the feature points using a symmetric Gaussian kernel function.

An illustration of the density map modules is shown in Figure 1. The remainder of this section describes each of these system components.

### A. Features from accelerated segment test

For local features, we extract *features from accelerated segment test* (FAST) [14]. FAST is proposed for corner detection in a fast and a reliable way. It depends on a wedge model style corner detection. Also, it uses machine learning techniques to find automatically optimal segment test heuristics. The segment test criterion considers 16 surrounding pixels of each corner candidate $P$. Then, $P$ is labeled as corner if there exist $n$ contiguous pixels that are all brighter or darker than the candidate pixel intensity.

The reason behind applying FAST as local features for crowd measurement is its ability to find small regions which are outstandingly different from their surrounding pixels. The selection of this feature is also motivated by the work in [15], where FAST is used to detect dense crowds from aerial images. The derived results in [15] demonstrate a reliable detection of crowded regions using FAST.

The extracted FAST features will be further used as observations of the probability density function. But since the probability density function should correspond to the density of crowds, a feature selection process is required to remove features which are not relevant to the crowd.

### 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 in our system a separation step between foreground and background entities. It 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) algorithm from [13] which computes very accurate sparse motion fields by means of a robust norm[1].

However, 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 KLT feature tracker from [16], it still is 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 into 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 time step, the overall motion $m_t$ of a trajectory $t$ is compared to a certain threshold $\beta$ 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 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.

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

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

Fig. 1. Illustration of the proposed crowd density map estimation using FAST as local features: (a) Exemplary frame, (b) Local feature points by FAST algorithm, (c) Feature tracks, (d) Distinction of 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

### C. Kernel density estimation

After generating feature tracks to filter out static points, 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 when more local features come close to each other, higher crowd density is perceived. For this purpose, the probability density function (pdf) is estimated using a Gaussian kernel density.

If we consider a set of $K$ FAST local features extracted from a given image at their respective locations $\{(x_i, y_i), 1 <= i <= K\}$, the density in $(x, y)$ position is defined as follows:

$$D(x,y) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=1}^{K} \exp -(\frac{(x-xi)^2 + (y-yi)^2}{2\sigma^2})$$

where $\sigma$ is the bandwidth of the 2D Gaussian kernel.

## III. Application of Crowd Density Estimation on Privacy Context

In order to apply crowd density information for privacy purposes, we want to hide personal information to the video operator while still allowing him to identify potentially dangerous areas and events in the scene. A simple approach for this could be to just use crowd density directly as an input to a privacy filter. In this case, the obfuscation level would depend directly on the number of people present in a given region. However, this approach would decrease the visibility of potentially important information as just all crowded areas would be obscured.

In our work we follow a different approach and focus on obfuscating only the regions which contain personal information. Generally these are depending on the scene context and could include face, clothing, skin/hair color or even gait. Given this variety and considering that these information are not perceivable under all circumstances (e.g. heavy crowding, different lighting conditions, motion blur, low contrast, low resolution...), in our work we consider only head obfuscation as the most visible part of a human in a crowd.

As a measure for privacy protection, the level of obfuscation is adapted to the crowd density for the following reasons: Crowds are usually interesting to video operators as they are a common place for crimes such as pickpocketing or for dangerous events as e.g. overcrowding. At the same time, as said above, people in a crowd exhibit a smaller amount of information to a video operator than isolated people who are completely visible. We therefore propose to lower the level of privacy protection within a crowded area because the visible information is also smaller and does not have to be filtered to the same degree as for isolated people.

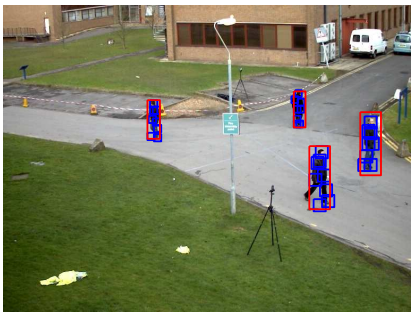


Fig. 2. Exemplary result of the used detector [17] on a frame of PETS2009 public dataset. By building a combination of multiple object parts, detection results are increased compared to the standard HOG detector from [18].

### A. Head detection

For obfuscation of people's heads in a scene, we apply a detection step using Histograms of Oriented Gradients (HOG) in order to find the head positions in the scene. Firstly proposed in [18], this algorithm takes gradient information from a detection window, derives a feature vector from it and compares it against manually annotated samples. We use a state-of-the-art detector [17] which is an extended form of the original algorithm [18] using multiple scales and resolutions. By means of a part-based model (see Figure 2 for details), the detection accuracy of the HOG algorithm is enhanced considerably. We will see in the results section that the HOG detector gives mostly promising results for person detection especially when the face is visible. As it is our goal to enhance the privacy of the people in the scene, this detector adds valuable information to our system. However, the performance of HOG detector can be affected by many factors such pose of the person, camera view, resolution of the image. This is also the reason why we will use a FAST-based approach for the estimation of crowd density maps.

### B. Adaptive privacy filters

After performing head detection, we obtain $F_n(x,y), n \in [1,2,...,N]$, where $n$ denotes the image index in a video sequence and $F_n(x,y) = 1$ if a detection is found at $(x,y)$ position, otherwise $F_n(x,y) = 0$. On the other hand, we get a crowd density map $D_n(x,y)$ which gives information about the crowd size and the crowd location as well.

At this stage, we intend to adapt the level of the privacy protection filters according to the crowd density. More precisely, as explained before we propose to allow 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 $D_n(x,y)$ into $d = |P|$ crowd levels and apply the respective filter parameter to the region of interest (ROI).

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

- Blurring: This privacy filter consists essentially of removing details in a region of interest by applying Gaussian low pass filtering. For this technique, the bandwidth of the used Gaussian is adapted according to the crowd density level.
- Pixelization: This filter is based on decreasing the resolution of any region of interest by replacing each block of $p$ pixels in this area with its respective average. As for the blurring process, the filter size $p$ is chosen with respect to the estimated crowd density level.

## IV. Experimental evaluation

The proposed framework is evaluated within challenging crowd scenes from multiple video datasets. In particular, we selected some videos from PETS 2009 [2], UCF [19], and Data Driven Crowd Analysis[20] public datasets.

As described in Section II, FAST local features are extracted and tracked in each frame under analysis. The moving local features are further used for estimating the crowd density map. On the other hand, head detections are performed using the state-of-the-art detector. The goal is to combine these two sources of information (crowd density map and head detections) for adaptive protection filters. For this purpose, two privacy protection tools (blurring, pixelization) are employed to show different ways to protect personal privacy in video sequences. The results using four frames from different videos are shown in Figure 3. 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 density values of the crowd density function and both are not affected by the density scaling problem. Comparing e.g. the man dressed in black (lower left corner of first image row) to the man walking behind the three blond ladies in the lower middle of the image, it is well perceivable that the privacy level is reduced within the crowd by a smaller block size or a smaller bandwidth respectively. As one would expect, the overall filter performance can be limited to maximal and minimal values for both privacy filters but for presentation purposes we show rather high protection levels.

We also note that the estimated crowd density is lower for the second scene (second row), compared to to the first one. As in both scenes people are rather uniformly dressed, in the second example there is not much texture available which reduces the number of FAST features. However, as in the first video faces of many people are visible, there is a higher contrast which results in more feature points and thus also a higher overall crowd density.

The PETS sequences shows more variance in crowd density and are therefore better suited to show the different levels of privacy protection. It can be seen that the single person (3rd row) and the smaller group of people (4th row) walking behind the crowd do not generate such a high density measure and are consequently obfuscated to a higher degree than people walking in the big crowd. Again, the different filter sizes can be seen also in these sequences. However, if no other constraints are given, 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.

## V. Conclusion

In this paper we show how it is possible to include crowd density information into a privacy-preserving framework. Using FAST local features as observations, we computed a probability density function for the feature points in each frame. Separation of foreground and background objects in the scene is performed using Robust Local Optical Flow which also implicitly introduces a temporal filtering step to improve the smoothness of the density maps between separate frames.

The computed crowd density maps are shown to be a valuable information source for privacy preservation. Using an additional head detection step, we adapt the degree of data obfuscation for privacy according to the crowd level and are thus able to show how it is possible to achieve an acceptable level of privacy for the people in a scene while still allowing the operator to view the data relevant for him.

For future works, we are working on ways of complementing person detection by crowd information in order to obtain better people detection results also in highly crowded situations. While this would primarily provide better detections, it could also improve privacy protection for cases as shown in this paper.

## 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.

[2] http://www.cvg.rdg.ac.uk/PETS2009/

## REFERENCES


[1] A. Senior, S. Pankanti, A. Hampapur, L. Brown, Ying li Tian, A. Ekin, J. Connell, Chiao Fe Shu, and M. Lu, "Enabling video privacy through computer vision," *Security Privacy, IEEE*, vol. 3, no. 3, pp. 50 – 57, may-june 2005.

[2] D.N. Serpanos and A. Papalambrou, "Security and privacy in distributed smart cameras," *Proceedings of the IEEE*, vol. 96, no. 10, pp. 1678 – 1687, oct. 2008.

[3] Thomas Winkler and Bernhard Rinner, "A systematic approach towards user-centric privacy and security for smart camera networks," in *Proceedings of the International Conference on Distributed Smart Cameras (ICDSC) 2010*, 2010.

[4] Andrea Cavallaro, "Privacy in video surveillance," *IEEE SIGNAL PROCESSING MAGAZINE*, vol. March, 2007.

[5] Simon Moncrieff, Svetha Venkatesh, and Geoff West, "Context aware privacy in visual surveillance," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, 2008.

[6] Atta Badii, Mathieu Einig, Marco Tiemann, Daniel Thiemert, and Chattun Lallah, "Visual context identification for privacy-respecting video analytics," in *Multimedia Signal Processing (MMSP), 2012 IEEE 14th International Workshop on*, 2012.

[7] A. B. Chan, Z. S. J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–7.

[8] D. Conte, P. Foggia, G. Percannella, F. Tufano, and M. Vento, "A method for counting people in crowded scenes," in *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2010.

[9] 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.

[10] W. Ma, L. Huang, and Ch. Liu, "Estimation of crowd density using image processing," *Computer Sciences and Convergence Information Technology*, 2010.

[11] 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.

[12] 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.

[13] Tobias Senst, Volker Eiselein, and Thomas Sikora, "Robust local optical flow for feature tracking," *Transactions on Circuits and Systems for Video Technology (TCSVT 2012)*, vol. PP, no. 99, 2012.

[14] 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.

[15] 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.

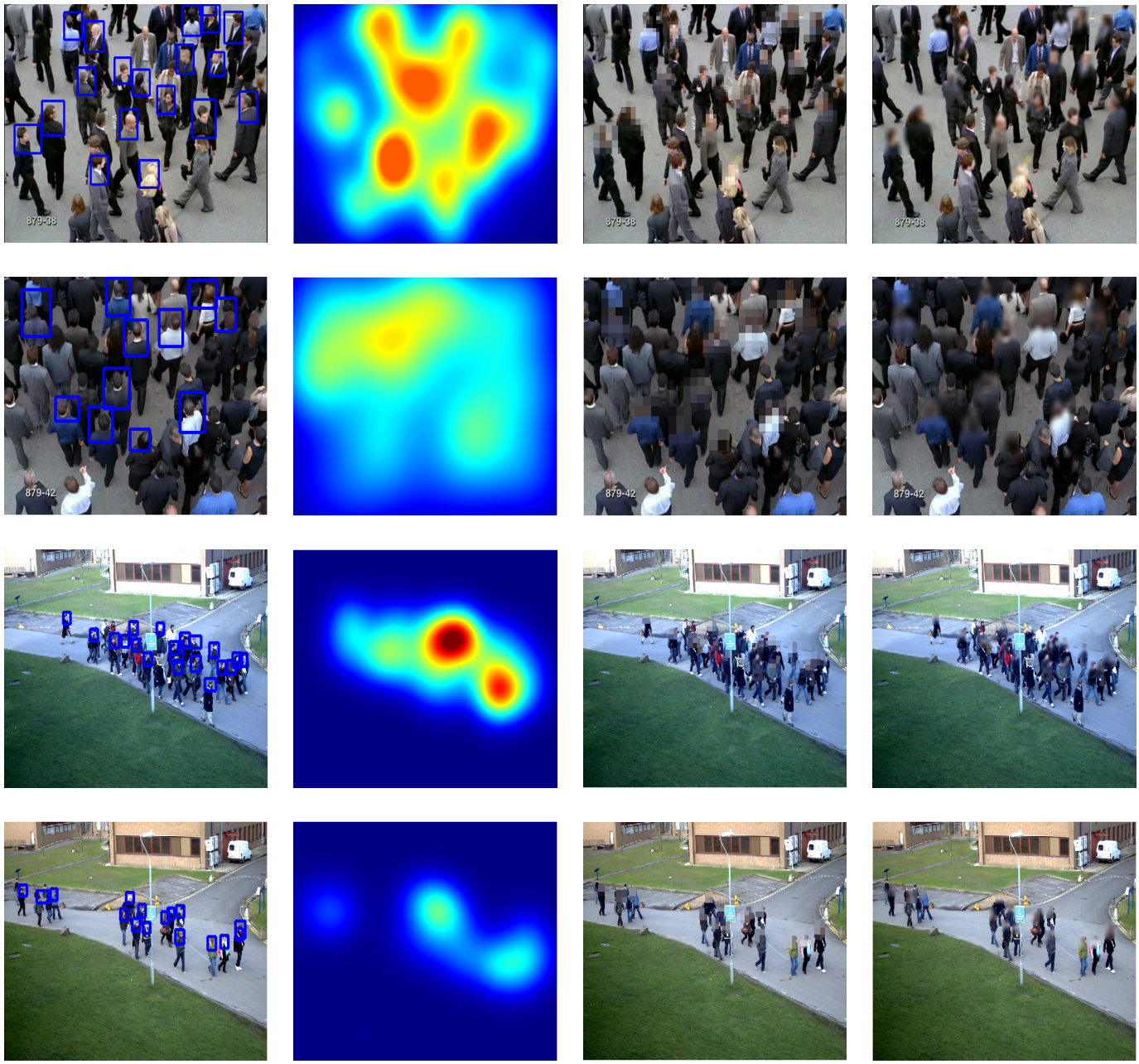
[16] C. Tomasi and T. Kanade, "Detection and tracking of point features," Technical report CMU-CS-91-132, CMU, 1991.

[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] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *International Conference on Computer Vision & Pattern Recognition*, INRIA Rhône-Alpes, ZIRST-655, av. de l'Europe, Montbonnot-38334, June 2005, vol. 2, pp. 886–893.

[19] Saad Ali and Mubarak Shah, "A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis," *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[20] M. Rodriguez, J. Sivic, I. Laptev, and J.Y. Audibert, "Data-driven crowd analysis in videos," *Proceedings of the IEEE International Conference on Computer Vision*, 2011.



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

Fig. 3. Results of adaptive protection filters using four frames from different test videos. From top to down order: UCF 879, INRIA 879, PETS2009 S1.L2.14.06.V1, and PETS2009 S1.L1.13.59.V1. From left to right order: Head detections according to [17], Estimated crowd density map, Application of pixelization filter, Application of blurring filter