ADDING PRIVACY CONSTRAINTS TO VIDEO-BASED APPLICATIONS

A. CAVALLARO

Multimedia and Vision Lab
Queen Mary University of London
Mile End Road, London E1 4NS, UK
E-mail: andrea.cavallaro@elec.qmul.ac.uk

Remote accessibility to video cameras, face recognition software, and searchable image and video databases enable automated video surveillance. The same technologies and features enabling advanced functionalities facilitate as well the misuse of a surveillance system and can lead to the infringement of the privacy of an individual. Examples of misuse are voyeurism and the unauthorized collection of data on activities or behaviours of an individual. To overcome this problem, we propose a system architecture for privacy-preserving video-based applications which uses image and video segmentation. Image and video segmentation are inherently ill-posed problems. In this sense, no unique segmentation of a scene exists: a priori knowledge about the nature of the problem is required in most situations. To this end, we propose a segmentation framework that embeds such a priori knowledge in the form of privacy constraints. The proposed framework is composed of two levels of semantic segmentation and subsequent description. The first level extracts video objects whereas the second extracts parts of a scene that can identify an individual, such as faces and license plates. The rationale behind this framework is that the data that one needs to hide for preserving the privacy of individuals and the data one needs to process to understand a scene may be divided into two non-overlapping sets. These two sets can be stored and processed separately thus enabling the use of automated surveillance in public places while protecting privacy.

1. Introduction

Video cameras are becoming ubiquitous and are used in a number of applications ranging from video-based human-machine interaction to video-based behaviour modelling, and from interactive games to video surveillance. Moreover, the increasing processing power of low-cost microprocessors, the decreasing cost of storage devices and the shift from analogue to digital video cameras are enabling the automation of tasks that otherwise would require the intervention of a human operator. An example is automated video surveillance, where digitalization and automation result in more efficient and cost effective monitoring. However, digitalization and automation also facilitate the collection of information about an individual that can be easily searched and copied. For

this reason, there is an increasing risk of misuse of surveillance information that can lead to the infringement of the privacy of an individual. In particular, data mining techniques and face recognition software have amplified the problem of privacy in video-based applications and especially in video-based surveillance.

Video-based surveillance systems are in rapid expansion and are already widely present in our everyday life: there are more than 2.5 million CCTV cameras in operation in the United Kingdom and 25 million worldwide. The average citizen is caught on CCTV cameras 300 times a day [1]. Video surveillance was initially adopted as a tool to improve safety and security, and to foster the perception by the citizen or the customer that assistance is available. However, with the development of technologies discussed above, video surveillance may generate the opposite effect: people do not perceive it as a security tool but as a threat to their private sphere [2]. In fact, with the progress in video technologies the scenario of what we could call the *people google* is becoming possible. A person could be searched through a large number of surveillance and web cameras by querying him or her by example, i.e. by giving as input a picture of his or her face. The output of the search would be the last time that person was caught by a web cam and where, and possibly the history of his or her displacements. Current solutions to the privacy issue include time limitations for the storage of recorded material. For instance, videos recorded by cameras covering town centres and streets may not need to be stored for longer than one month, whereas video recorded by cameras protecting individuals' safety at ATMs could be retained for a period of three months in order to resolve customer disputes about cash withdrawals. Once the retention period has expired, the video should be removed or erased. This solution is effective for analogue technology, but can be insufficient with digital video and real-time image processing and data mining techniques. For this reason new solutions should be searched for coping with the privacy issue.

We propose to include the privacy requirement in the design of a surveillance system, as opposed to the current solution of adding this requirement to an existing system. To this end, smart cameras are used that automatically separate the video data into two classes, namely behavioural video data and personal video data. *Behavioural video data* are used for monitoring and prevention. *Personal video data* are sent to a separate location where an authority can allow their use for law enforcement purposes only (Figure 1).

The paper is organized as follows. Section 2 introduces the concept of privacy preserving video camera and describes the first level of semantic segmentation, i.e. the techniques for extracting objects of interest from video. Section 3 describes second level of semantic segmentation and description, i.e. the algorithm for separating video information into personal and behavioural data. Furthermore, it explains how data are searched and accessed remotely with different level of authorization. Finally, Section 4 concludes the paper and describes future directions.

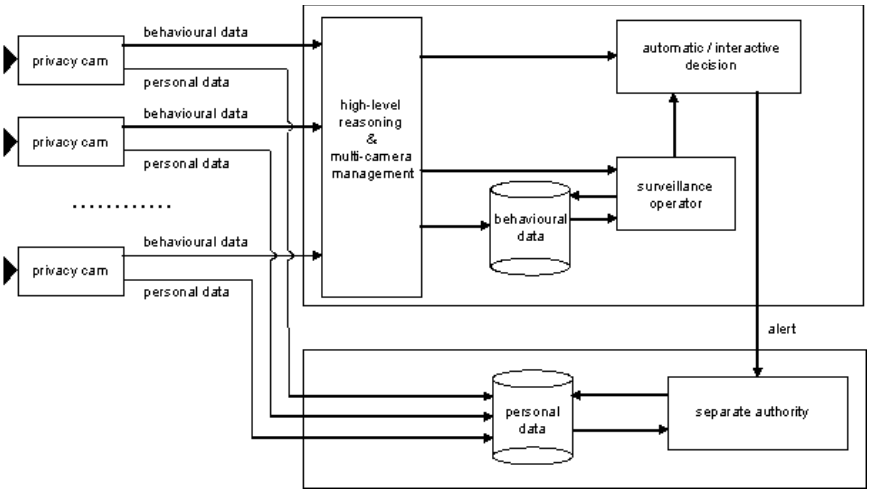

Figure 1. Flow diagram of the proposed privacy-preserving surveillance system. A privacy camera (privacy cam) automatically separates the video data into behavioural video data and personal video data.

2. Privacy-preserving cameras

The decreasing cost of processing power enables to perform video analysis close to the sensor, as opposed to the traditional centralised processing. Well-defined tasks can therefore be performed directly in the video camera. For example, motion information can be used by means of change detection algorithms to automatically detect intruders or moving objects.

A change detection algorithm is ideally expected to extract the precise contours of moving objects (spatial accuracy). An accurate extraction is desired when the objects are analyzed for recognition and for automatically collecting statistics about the scene. Change detection identifies temporal changes that are used to segment the input video into moving objects. However, temporal changes may be generated not only by moving objects, but also by noise components. The main sources of noise are illumination variations, camera noise, uncovered background and texture similarity between objects and background.

The illumination of a scene is not always constant, and changes in viewing conditions may occur. We refer to these variations as global illumination changes. Global illumination changes can be classified into two major groups: slow and fast. Slow illumination variations generally occur due to natural causes, such as the changing daylight. Fast illumination variations are caused by

artificial phenomena, such as automatic gain control of the camera, lights being switched on or off, and by natural causes, such as passing clouds. A global illumination change introduces apparent motion in the entire image. The apparent motion introduced by the variations in the global illumination masks the apparent motion introduced by the camera noise, and therefore an illumination invariant change detection technique is desired to cope with this problem. For this reason, an adaptive change detection algorithm for the extraction of multiple moving objects is used for the surveillance scenario [4]. The technique makes use of a colour edge detection scheme applied to the difference between a current and a reference image. Colour edge detection is performed by applying a Sobel operator on each colour component and by fusing the results by means of a logical *or* (V). Let I_r and I_t represent the reference and the current frame, respectively, and R, G, and B the 3 colour components of each frame. Then, the colour edge change detection operation between the current and reference frame, $C(I_r, I_t)$, can be defined as

$$C(I_r, I_t) = \varepsilon(|R_r - R_t|) \vee \varepsilon(|G_r - G_t|) \vee \varepsilon(|B_r - B_t|)$$

where ε is the Sobel operator. The Sobel operator provides thick edges which are useful for application in change detection since they allow lightening the post-processing for contour filling. The filling procedure is preceded by a morphological filtering composed of two successive eight connected dilations and erosions.

Figure 2 shows typical results from the colour edge change detection technique compared with well-known change detection algorithms [3]. A sample frame of the test sequence Hall Monitor from the MPEG-4 data set is shown in Fig.2 (a). The change detection masks computed in controlled light conditions with a statistical change detector, a change detector based on the shading model, and the colour edge detector are shown in Fig.2 (b), (c), and (d) respectively. Fig.2 (e) and (f) shows the results of the shading model and the colour edge detector when varying light conditions are simulated on the same sequence. Results of the statistical change detector are not shown in this case since the entire frame is classified as object due to the global illumination change. It is possible to notice the relative accuracy of the colour change detector with respect to the state-of-the-art techniques both in controlled and varying light conditions.

Since the result of change detection is the classification of the pixels into two classes, namely foreground and background, a change detection algorithm provides no information about different objects in the scene. The separation of the objects is achieved by linking through tracking different time instances of the same object [4]. Once the sequence is divided into objects (first level of semantic segmentation), behavioural and personal video data are generated as described in the following section.

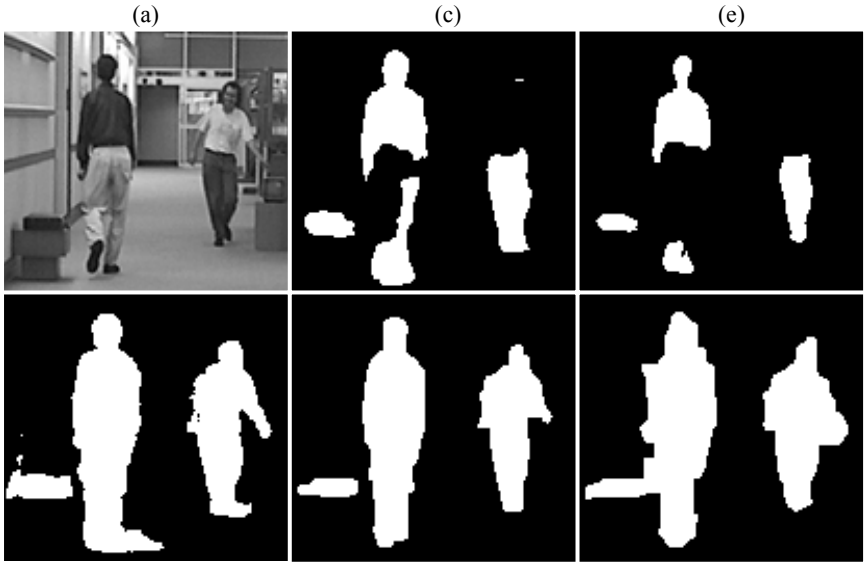

Figure 2. Change detection results in controlled and varying light conditions. (a) Original image. Controlled light conditions: (b) statistical change detection; (c) shading model; (d) colour edge change detection. Varying light conditions: (e) shading model; (f) colour edge change detection

3. Behavioural and personal video data

After change detection and tracking, video data are separated into two classes, namely personal video data and behavioural video data. Examples of personal video data are data representing the face of a person or the licence plate of a vehicle, whereas examples of behavioural video data are tracks of walking people or the trajectory of a vehicle [5] (Figure 3).

Personal video data are identified throught the second step of semantic segmentation. In order to segment faces, colour-based segmentation [6] or a cascade of classifiers [7] can be used. A number of relatively robust algorithms for face segmentation are based on the fact that human skin colours are confined to a narrow region in the chrominance plane, and their distribution is quite stable [6]. Colour-based segmentation combined with change detection is used to select faces, even when their size in the video is very small. When the size of the area covered by a face is larger than 24-by-24 pixels a cascade of simple classifiers is used to reiforce the detection [7]. Each classifier is trained to detect a specific face feature, such as the intensity difference between the eye region

and the upper cheek or between the eye region and the bridge of the nose. Personal video data identified with the second step of semantic segmentation contain information about the identity of an individual and are separated from behavioural data and sent to a separate, secure database (Fig. 1). Personal video data are linked to behavioural data.

Behavioural data are computed directly in the video camera and sent to the control center. Behavioural data do not convey any information about a particular individual and therefore are accessible to any surveillance operators. Behavioural data are formatted according to the MPEG-7 standard in order to allow for interoperability between different devices. The descriptors are object identifier, polygonal region locator and dominant colour. Behavioural data are rendered for the surveillance operator and are used for generating statistics about the scene and for triggering alarms in case of anomalous situations. Sending MPEG-7 BiM descriptors of a scene instead of the entire scene has also the benefit of reducing tenfold the bandwith requirements while mantaining and enhancing relevant information. This in turn facilitates the use of a wide range of mobile devices for surveillance.

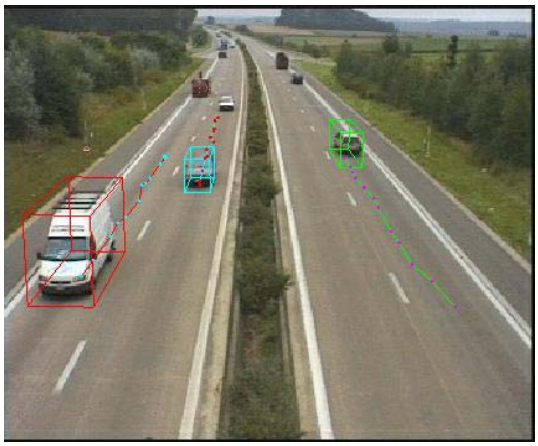

Figure 3 Example of behavioural data. The information provided by the bounding boxes of the vehicles and their trajectories over time is sufficient for traffic monitoring purposes. Instances of the real objects are stored and accesses only by authorized users.

The access to the different data types is defined by appropriate privileges (Figure 4). An access privileges table is defined, composed of the following fields: LoginID, Password, and Level. The LoginID and Password allow the user to have access to the system, while the Level field gives different access permissions for different types of users. The level of authority is assigned by an authorized administrator and only he/she can input data into this table. The

instances of a scene selected by the smart cameras are sent to the remote database and organized according to the following indices: camera number, date and time. The authorized user can search (Figure 3 - b) for specific instances (Figure 3 - a) of objects that are related to a particular event identified by analyzing the behavioural data. A user with limited access privileges such as the surveillance operator will not be authorized to access to the identity of the objects (Figure 3 – c) and will only be able to access statistical data (Figure 3 - d). When requested by the surveillance operator, personal data can finally be accessed by an authorized authority that can browse the personal data for law enforcement purposes.

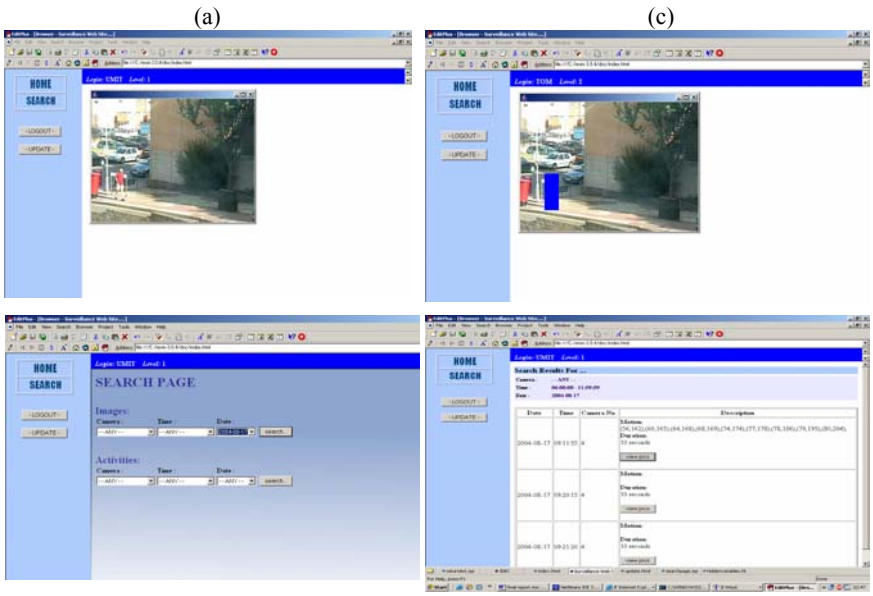

Figure 4. Interface to access personal and behavioural data. (a) Full image of a relevant event accessible by an authorized user. (b) Search page allowing retrieving a specific event in the surveillance database. (c) Image with personal data hidden accessible by user monitoring the site. (d) Example of behavioural data.

4. Conclusions

In this paper, we discussed the privacy issues in video-based applications and we proposed an architecture of a privacy-preserving video surveillance system.

The proposed architecture can be easily extended to other video-based applications, such as human-computer interaction and behaviour modelling in shopping areas. We showed that it is possible to operate surveillance in public places while protecting privacy.

The proposed system represents an example of the adoption of societal needs (the privacy) in the design of an engineered system in order to gain its acceptance by the general public. The system is based on the concept of smart camera and uses MPEG-7 description and JAVA technologies.

Future work includes the extension of the system in order to store and retrieve not only images but also video clips corresponding to relevant events and a formal evaluation of the approach based on a questionnaire to validate the acceptance of a privacy preserving-surveillance system by the general public.

References

1. Jane Wakefield, *BBC news*, 7 February 2002.
2. David Brin. The Transparent Society: will technology force us to choose between privacy and freedom. *Perseus Publishing*, 1999.
3. A. Cavallaro and T. Ebrahimi. Change detection based on color edges. *Proc. of IEEE International Symposium on Circuits and Systems*, Sydney (Australia), 6-9 May 2001.
4. A. Cavallaro, O.Steiger, T. Ebrahimi. Tracking video objects in cluttered background. *IEEE Transactions on Circuits and Systems for Video Technology* (in press)
5. B. Abreu, L. Botelho, A. Cavallaro, et al.. Video-Based Multi-Agent Traffic Surveillance System. *Proc. of IEEE Intelligent Vehicles Symposium*, Detroit (USA), pp. 457-462, 3-5 October 2000.
6. J. Yang, W. Lu, A. Waibel. Skin-color modeling and adaptation. *Proc. Asian Conf. Computer Vision*, vol. 2, pp. 687-694, Hong Kong, 1998.
7. P. Viola, M. Jones. Robust Real-time Object Detection. *Proc. of Int. Workshop on Statistical Learning and Computational Theories of Vision Modeling, Learning, Computing and Sampling*, Vancouver, Canada, 2002.