Security and safety assessment for cultural heritage: enhancing security profiles by wireless video-surveillance

G. Mondini[1], S. Olivero[2], R. M. Scopigno[3], S. Belfiore[3] & M. Martines[4]
[1]*Politecnico di Torino, Turin, Italy*
[2]*SiTI, Turin, Italy*
[3]*ISMB, Turin, Italy*
[4]*Dylog S.p.A, Turin, Italy*

Abstract

Museums, monuments, picture galleries and works of art represent major assets that need security and safety measures against both anthropic (thefts, vandalism, terrorism) and environmental threats. On behalf of the Compagnia di San Paolo, SiTI designed an assessment tool to score security and safety levels of cultural/artistic site. The project, called "Security and Safety Assessment of Cultural Heritage" (SESASS), provides a "protocol" to evaluate security and safety profiles through a process involving checklists in all relevant areas.

The lack of a robust video-surveillance system clearly emerged as the most frequent weakness of a number of assessments: as a fact, the ability to recognize (and classify) objects, persons, human activities and behaviours "automatically", managing alarms criteria, is fundamental to (help) perform security tasks in a scalable way, overcoming typical limitations of a "pure human" approach. For this reason SiTI and ISMB decided to jointly focus on video-surveillance of cultural heritage, taking into account both technological aspects and organisational requirements (operative procedures). In this paper both the available technological solutions and the involved organizational and legal aspects are shortly overviewed.

Keywords: intelligent video-surveillance, security, cultural heritage, wireless transmission.

1 Introduction

Cultural/artistic sites which (museums, monuments, picture galleries and works of art) are recognized to be major assets needing protection against anthropic and environmental threats. For this reason, on behalf of the Compagnia di San Paolo, SiTI designed an assessment tool to score security and safety levels of such sites: the project was called "Security and Safety Assessment of Cultural Heritage" (SESASS) and was meant to provide a "protocol" to assess security and safety profiles through a detailed check-list.

The main areas of investigation considered are: boundaries protection, inventory of objects in the exposition rooms, protection of objects in warehouses/expositions/restoration laboratories, alarm systems against unauthorized access, video surveillance, management of physical keys, personnel training and qualifying, and managers' approach to security systems.

The checklists (of each area) contain elements concerning the security against thefts and vandalism, as well as installation requirements (energy supply, lighting system, fire alarm) and safety specifications (i.e. the environmental indoor characteristics that enable a good preservation of the works shown in the museums). A trade-off among the assessed weaknesses, criticalities and technical requirements leads to security and safety scores, and enables the definition of the actions needed for the enhancement of the security and safety levels of the site as well as the priority among them.

A sample of five museums was chosen to tune the methodology, and the revised approach to security assessment was eventually tested on a major museum institution. The lack of a robust video surveillance system clearly emerged as the most frequent weakness of a number of assessments. As a fact, the ability to recognize (and classify) objects, persons, human activities and behaviours "automatically", managing alarms criteria, is fundamental to (help) perform security tasks in a scalable way, overcoming typical limitations of a "pure human" approach. For this reason SiTI and ISMB decided to jointly focus on video surveillance of cultural heritage, taking into account both technological aspects and organisational requirements (operative procedures).

In this paper the results of this study are presented: the remainder of the structured as follows: in sect.2 a short overview of video surveillance state-of-art is provided; sect.3 infers research directions and needs from the assessment of security performed according to methodology of SESASS project; sect.4 is focused on privacy implications of video surveillance technology.

2 State of the art of video-surveillance

Due to the application envisaged, i.e. the video-surveillance of a museum, it is legitimate presuming that the automatic system will be supervised by an operator who is responsible for the control of the system. This context allows for a more precise tuning, for example setting two different thresholds: exceeding the first one only a warning on the video could be displayed, while the surmounting of the second threshold would produce a loud alarm, with human intervention. Afterwards, we will use "automatic" meaning an intelligent system supervised or

unsupervised, indiscriminately. This on one hand relaxes requirements on the automatic surveillance system and, on the other hand preserves or, we could say, amplifies traditional human approach. Effectively all the automatic video surveillance systems, more or less, try to mimic human surveillance: for example all of them include a core block which performs motion analysis; such block usually includes the following logical steps: motion detection; segmentation and object recognition; object tracking. The first extracts temporal changes; the second merges pixels with similar displacement vectors; the third builds the moving objects trajectories: not all these steps are strictly necessary.

2.1 Motion detection

Limiting this survey to the most common 2-D models, a first discrimination has to be made on the basis of the available camera type: panning and mobile cameras need a pre-processing stage consisting of a motion compensation process in order to construct a consistent set of background images. In order to allow the estimation of motion, four main methods are used to determine the displacement vector field from a sequence of images.

1. The first family of methods includes all differential techniques, or "optical flow" techniques [5]: these techniques are based on the hypothesis that the brightness of a point does not change during the time and are quite noise sensitive. The most common and simple method foresees the use of an image as a reference the currently acquired image is compared to: said $R(x,y)$ the reference picture, $I(x,y)$ the current one, $D(x,y)$ the difference between the two images, the evaluation of the binary motion map ($M(x,y)$) is obtained by comparison as follows in the paradigm

$$D_i(x,y) = |I_i(x,y) - R_i(x,y)|$$

$$M(x,y) = \begin{cases} 0 \; if \; AND/OR \underset{i}{\{}D_i(x,y) < TH_i\} \\ 1 \; if \; AND/OR \underset{i}{\{}D_i(x,y) > TH_i\} \end{cases}$$

   where i indicates the chromatic component (in RGB or YUV colour spaces) and the logical operator AND/OR depends on the desired sensitivity to changes (this approach is suitable for indoor scenes, where a proper updating of the reference image can effectively face slow changes, while this model have difficult in counteracting the quick changes typical of the outdoor environment (e.g. light changes, leaf movements, etc.). As an option a modelled background may be used as reference with pixels preventively classified in background and foreground pixels attaining the motion map $M(x,y)$: the method suggested in [3] better suites this case and foresees to build a statistical model of the background, using three values (the three values used are: the minimum and the maximum values and the maximum difference between consecutive frames over a predefined time interval in a training period). In order to counteract the effects of possible illumination variations or scene changes permanent and not interesting, the reference image has to be periodically updated.

2. The second group of solutions is based on primitives matching: image primitives extracted at time [t] are matched with those extracted at time [t – 1]. A similitude criteria (such as distance measure, correlation level, etc.) is

used for selecting the right associations. Also in this case background models can be employed.

3. The third set of methods is based on motion characterisation in frequency space. The basis principle is the elaboration of a set of filters, each of them adapted to a specific motion direction. In the last category spatio-temporal analysis techniques can be grouped. These solutions are quite complex to be shortly explained and make use of a double approach: on one hand there is an integration, in single image, of all moving objects detected in the sequence of images; on the other hand, moving objects are still detected in a sequence of images, with accumulation of images along a time axis. The analysis is performed inside the spatio-temporal volume and direction of motion is obtained as a result of a cut inside this volume.

2.2   Segmentation and object classification

In order to successfully track objects and analyze their behaviour, it is essential to correctly detect and separate moving objects. Moving regions can be characterized in many ways, including two-dimensional (2-D) contour, silhouette, skeleton, and bulk box (or blobs): blobs in particular allow to restrict the complex tracking algorithms to only simple regions. In order to obtain regular blobs, two pre-processing steps are needed: in the first phase a density map is evaluated on the basis of the binary motion map $M(x,y)$, giving greater weights to pixels belonging to compact areas; in the second phase a threshold is applied to the density map, yielding another binary map. In order to improve the robustness to noise interference and the detection capabilities of the blob extraction process, a post-processing stage is performed: the blobs smaller than a defined area are discarded, distance between blobs is evaluated and the blobs nearer than a defined threshold are merged together to form a unique bulk box.

Once the blobs have been detected, they are identified by some parameters such as position, dimension (height, width, area of both the object and the bulk box), dispersion, aspect ratio, average colour and, in case, classified on the basis of these measures. Moving objects are then classified according to shape, colour and motion, into four classes isolated: human beings, vehicles, human groups and clutter using a three-layer neural network classifier. Classification is carried out using a set of features including a mixture of image-based and scene-based object parameters such as blob area, aspect ratio of the bounding box, the camera zoom information and so on.

2.3 Object tracking

In video-surveillance applications, it is important to study the evolution of detected objects (or persons) over a certain time interval. There are different tracking techniques (region-based, contour-based, feature-based and model-based) depending on the blob characteristic they are based on.

o Region-based tracking [5, 10] algorithms track the objects frame-by-frame according to variations of moving image regions detected, typically, by subtracting a dynamic background from the current image. Once the moving regions (e.g. bulk boxes) are detected in consecutive frames, a dependency

graph is build through a blob matching algorithm (five relations can be defined). This algorithm family is quite simple and is particularly suitable in case of low motion scenes or in case of few objects; on the other hand, it cannot reliably handle occlusions and it is not suitable for noisy and cluttered environments.

o Contour-based techniques [11, 15] track moving objects by extracting frame by frame object boundaries and shapes and by updating them through time. This sort of description provide a more detailed information of the moving objects in comparison with region-based methods, even if it is known that edge detection algorithms are generally noise sensitive. Furthermore, they partially solve the occlusion problem ant they are suitable for the porting of 2-D data into 3-D environments.

o Feature-based algorithms rely on two modules performing the extraction of elements and their segmentation through a clustering process; this way more than one feature can be identified in each relevant area and the features can be tracked independently until their centroid movements have distinguishable velocity (even in case of occlusions).

o The last group is called model-based [14] and relies on the attempt of finding a match between the identified object and a model, based on a prior knowledge. Due to the envisaged application, we restrict our description to human models and tracking.

Model-based techniques applies a predict-match-update approach: first, a prediction of the model pose in the next frame is performed, thanks to the prior knowledge an to the motion history, then a similarity measure is evaluated between the image data and the prediction and, in the end, chosen the beast matching pose among a number of candidates, the model history is updated. The most used models are the one based on a human body description ant the one based on a description of a number of possible movements. Standard tracking algorithms are based on statistical methods: various human movements represented by a sequence of stick movements are modelled by Hidden Markov Models (HMMs) for classification.

The tracked moving objects can be eventually classified according to several algorithms, such as: Kalman filtering based methods, neural networks or Hidden Markov models.

3 Research directions after security assessment experience

Considering the state of the art of video surveillance solutions, surveillance systems available in museums were assessed, paying great attention to the service available, rather than on technical solutions. The following areas were investigated:

o coverage capability: boundaries, exposition rooms, warehouses and restoration laboratories (when the laboratory is external to the museum, the presence of operational security procedures was investigated);

o rate between number of monitors and video cameras;

o quality of vision: daily and during the night; model, technology and year;

o   presence of warning message on the existence of the video surveillance system, for both security and privacy purposes;
o   presence of motion detection and anomalous behaviour recognition;
o   capability of security operators to continuously check the monitor (ergonomy, absence of disturbing parallel activities, external control rooms);
o   period of video recording and times of storage;
o   tools for image retrieval and advanced surveillance operations;
o   existence of a maintenance program and contract;
o   compliance of video surveillance installation with law and standards.

The assessment resulted in a "desired" security profile, that enabled the development of innovation paths, the main relevant described below. The desired innovations can be mainly grouped into two sets: features for advanced service layer (motion classifications in crowded [13] and ever changing areas, abandoned and removed objects recognition, anomalous behaviour recognition, selective recording, automatic rating and tagging of movies, etc.) and features involving infrastructure (wireless transmission, multi-spectral cameras). The desired overall architecture is depicted below.
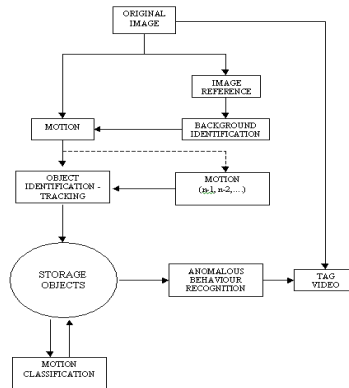
Figure 1:   Block scheme of a general video-surveillance system.

3.1 Advanced motion classification

As we mentioned, motion detection is the first elementary step in image sequence analysis. Therefore it is the most critical element of the chain: if it fails, the whole process gets ruined; unfortunately in crowded spaces or areas subject to continuous changes a flat motion detection algorithm would lead to strong errors. Some solutions to overcome this limit are available but could still be improved: the system should decide automatically (self-learning and with the minimum number of false errors) whether a motion event is interesting or not (crowded background or periodic motion). In particular tracking of individuals and groups will be similar to an ensemble of particles: tracking particles over time can give an indication of how such particle move within an event of interest. Basic states for a small group of individuals can be defined by the dynamics of particle (for instance a spatial expansion, compression, detachment

of particles), while an entropy measure can be assigned measuring the state of order of a group or its irregularity and could be used as an indication of the state of alert of a situation. The solution is likely to strongly depend on the video surveillance environment.

3.2 Advanced background identification

Since images in motion are more likely to be observed by human operators, the automatic detection of background changes should complete the alerting system. A Statistical Background Analysis can build a dynamic model of the background being observed, able to absorb usual changes and alert in case of non-usual ones. Changes considered "usual" are those due to light fluctuations, for example, in outdoor environments: background analysis is not affected by such events, which, otherwise, would cause false alarms. Potential dangerous events may be presence of objects of doubtful origin, such as bag or luggage, left alone for an excessive time, or absence of object retained part of the scene (vandalism).

3.3 Behaviour understanding

This is a very challenging features: a behaviour understanding module should be sophisticated enough to allow the classification and recognition of time varying data in order to identify suspicious behaviours. This involves studying the evolution of the evolving extracted features or objects and comparing them to a book of pre-classified events. The reference book can be built through training test sequences of known behaviours, either using deterministic (e.g. Dynamic Time Warping - DTW-, Finite-State Machine, etc.) or stochastic (e.g. Hidden Markov Model, self organizing neural network) processes. Those methods can be fed up with static measures (i.e. position, dimension - height, width, area -, dispersion, aspect ratio, average colour, distance from fixed points, etc.), and/or dynamic measures (i.e. speed, acceleration, direction and direction changes, etc.).

A major innovation could be the use of the behaviour models combined with the geometric and semantic model of the site: this would allow the location of a detected person in a simulated 3D environment, even while the person is moving from the field of view of one camera to another. Extended, multi-camera scenarios could enable also 3D motion analysis and refine the behaviour models.

3.4 Selective recording, tagging and rating of movies

Retrieving a specific event record as fast as possible could help human operators in the analysis of the emergency situation in order to take the right decision immediately: a smart marking/indexing solution would be a key enabler of this.

In particular video processing (and recording) may be "continuous" or driven by "events" (frames recorded are those carrying information of the occurring event).

Since the former case would generate huge videos, it is not always possible and does not always "scale" with the needs: event-driven actions could strongly improve the overall scenario. More in details, any actions aimed at distinguishing the most relevant frames in the movie, during its recording, could be applied for several purposes, such as: for selective recording; to tag the frames for retrieval

purposes; for automatic rating of the movie (to get the right piece of video in order to understand a cause/effect relationship)

3.5 Multi-spectral cameras

Additional benefits could also come from the introduction of multi-spectral cameras. In fact it could be interesting to study and (hopefully) demonstrate the possibility to integrate different types of information, such as images from visible range with information coming from the elaboration of image in different spectral ranges (*e.g.* UV and IR). This could be helpful to activate warning messages to be confirmed by using multi-spectral analysis.

3.6 Wireless coverage

The most common request that rose from the security assessments within museums was the possibility to collect video data and transmit it over a wireless network infrastructure. This issue is not strictly related to surveillance but could strongly enforce it (due to the cabling restrictions in historical buildings) and, on the other hand, should be carefully evaluated, due to several reasons, such as: a wide range of wireless consumer solutions are available: each has different costs and benefits; wireless transmission is flexible but could be easily obscured; interference scenarios should be preventively studied; data should be encrypted, being broadcast available, in order to avoid privacy violations (as discussed in the following section); wireless could contemporarily facilitate the interconnection of a large range of heterogeneous devices (intelligent equipments designed to help the visit, consumer devices for a scalable incident management).

4 Privacy constraints

Whenever video surveillance is studied, privacy constraints must be taken into great consideration. In other words, one must be aware that the over-proliferation of image acquisition systems in public and private areas should not result in placing unjustified restrictions on citizens' rights and fundamental freedoms.

For this reason the European Community voted Directives (Dir. 95/46/3c, Art. 29) to set a Working Party to study Privacy matters in general, and implications from video surveillance in particular. The most relevant result is the Working Document of the Data Protection Working Party [11] where the following international legal instruments are recalled:

o   the Convention of Human Rights and Fundamental Freedoms (the protection of privacy is ensured by Article 8 of the Convention of Human Rights);

o   the Council of Europe Convention No. 108/1981 for the protection of individuals with regard to automatic processing of personal data;

o   the Charter of Fundamental Rights of the European Union (it provides in Article 7 for the protection of private and family life, home and communication and in Article 8 for the protection of personal data);

o   the Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

The specific features of the processing of personal information included in sound and image data have been expressly highlighted by Directive 95/46/EC, which refers to them in several points. A considerable portion of the information collected by means of video surveillance concerns identified and/or identifiable people, who have been recorded as they moved in public and/or publicly accessible premises: the video recording must not interfere with personal rights and freedoms, as also related to anyone's own private sphere and image. Following this rationale in [11] the right to free movement of individuals is considered: this freedom may only be subject to such restrictions as are necessary in a democratic society and proportionate to the achievement of specific purposes. Consequently images must be processed fairly and lawfully as well as for specified, explicit and legitimate purposes. Based on these criteria, it is necessary for the processing of personal data by means of video surveillance to be grounded on at least one of the prerequisites referred to in Article 7 – unambiguous consent, necessity for contractual obligations, for compliance with a legal obligation, for the protection of the data subject's vital interests, for the performance of a task carried out in the public interest or in the exercise of official authority, balancing of interests. The data controller should ensure that the purposes sought are neither unclear nor ambiguous, also in order to be provided with a precise criterion when assessing compatibility of the purposes aimed at by the processing. The sensitive nature of the processing operations requires careful consideration of the scope of the tasks, powers and legitimate interests concerning the data controller. Superficiality and the groundless extension of the scope of such tasks and powers should be absolutely banned in carrying out this analysis. Finally, it is fundamental for the operators concretely involved in video surveillance activities to be adequately trained in and made aware of the steps to be taken to fully comply with the relevant requirements.

5   Conclusions

The enforcement of video surveillance solutions seems to be a common request of all museums and historical places: a wide range of hand-key solutions is available and many further developments have been proposed as research issues. Consequently the optimization of the design of the system has several degrees of freedom, due to the possible practical solutions, but also to the different requirements and constraints (including those due to privacy), and to the specific museum characteristics. As a result, it does not seem possible to have a flat and completely automatic approach for a generic video surveillance design, but it is rather possible to obtain a sort of "protocol" which help make appropriate design choices, giving evidence to strong points and weaknesses of the solution.

References

[1]   Hu W., Tan T., Wang L., Maybank S., *A Survey on Visual Surveillance of Object Motion and Behaviors*, IEEE Trans. on Systems, Man and Cybernetics, Vol. 34, No.3, Aug. 2004.

[2]   Haritaoglu I., Harwood D., Davis L. S., *W⁴: Real-Time Surveillance of People and Their Activities,* IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 22, No. 4 Aug. 2000.

[3]   Teschioni A., Oberti F., Regazzoni C. S., *ROC Curves for Performances Evaluation of Video Sequences Processing Systems for Surveillance Applications*, IEEE International Conference on Image Processing, ICIP99, Kobe, Japan, Oct. 1999.

[4]   Meyer D., Denzler J., Niemann H., *Model based extraction of articulated objects in image sequences for gait analysis*, IEEE International Conference on Image Processing, ICIP98.

[5]   Marcenaro L., Gera G., Regazzoni C., *Adaptive Change Detection Approach for Object Detection in Outdoor Scenes Under Variable Speed Illumination Changes,* European Signal Processing Conference, Eusipco 2000, Tampere, Finland.

[6]   Cupillard F., Avanzi A., Bremond F., Thonnat M., *Video understanding for metro surveillance*, IEEE Int. Conf. on Networking, Sensing and Control 2004.

[7]   Sacchi C., Regazzoni C., Vernazza G., *A neural-network-based image processing system for Detection of vandal acts in unmanned railway environments*, International Conference on Image Analysis and Processing, ICIAP01, Palermo, Italy, 2001.

[8]   Xu F., Fujimura K., *Human detection using depth and gray images*, Proc. IEEE Conf. on Advanced Video and Signal Based Surveillance (AVSS'03).

[9]   Serra J., *Image analysis and mathematical morphology,* Academic Press 1982.

[10]   Fuentes L. M., Velastin S. A., *People tracking in surveillance applications*, 2nd IEEE Int. Workshop on Performance Evaluation on Tracking and Surveillance, PETS 2001, Kauai (Hawaii-USA), 2001.

[11]   Baumberg A., Hogg D. C., *Learning deformable models for tracking the human body*, in *Motion-Based Recognition*, Shah M., Jain R., Eds., Norwell, MA: Kluwer, 1996, pp. 39–60.

[12]   Article 29 Data Protection Working Party (Art.29 of Directive 95/46/EC) – European Community - Rodota S. and oth., *Working Document on the Processing of Personal Data by means of Video Surveillance*, 11750/02/EN WP67 – Adopted on 25 November 2002

[13]   Anthony C. Davies, Jia Hong Yin and Sergio A. Velastin, *Crowd Monitoring Using Image Processing*, IEE Electronic and Communications Engineering Journal, Vol. 7, No. 1 (Feb), pp.37-47.

[14]   Jonathan Owens, Andrew Hunter & Eric Fletcher, *A Fast Model-Free Morphology-Based Object Tracking Algorithm,* British Machine Vision Conference 2002 (BMVC2002)

[15]   Bohyung Han, Changjiang Yang, Ramani Duraiswami, Larry Davis. *Bayesian Filtering and Integral Image for Visual Tracking*, University of Maryland, USA.