8

SuperVision: Video Content Analysis Engine for Videosurveillance Applications

Lisa Usai, Francesco Pantisano,
Leonardo G. Vaccaro and Franco Selvaggi
Elsag Datamat S.p.A.,
Italy

1. Introduction

Greater challenges in the security area and declining cost of technology have promoted the development of ever more sophisticated video surveillance systems. Such systems are widely employed both in the public sector, to support police activities for example, and in the private sector, in banks, shopping centres security, etc, working 24 hours a day, 7 days a week.
Beside security applications, video surveillance is successfully employed in other fields, such as monitoring traffic or studying people's behaviour or consumer's preferences.
The increasing extent of the areas to be monitored requires the use of a large number of cameras. Video-streams flow to central control room and are displayed in real time to operators. The large amount of data makes the task of security staff demanding but also very tedious. Although security operators are trained, it's impossible they maintain high levels of attention when confronted with multiple inputs for more than a few minutes, (also because most of the time video streams show ordinary behaviour). Furthermore sociological researches (McCahill & Norris, 2003; Smith, 2004) have proven that often it is the operator who decides on which camera to focus his attention, basing the decision on the appearance rather than the behaviour of people in the scene.
Video content analysis represents a solution to these problems. Its main purpose is to analyze video streams and alert the operator only when relevant events are detected. This will help solve the problem of operators discontinuous attention.
Even the European community is paying close attention to these issues and in recent years several funded projects were launched to develop the most appropriate technologies to solve specific problems. The main goal of ISCAPS, for example, has been to reinforce security for European citizens and to try to reduce terrorist threats. The aim of SAMURAI is to develop and integrate an innovative intelligent surveillance system to monitor people and vehicle activities in critical public infrastructures and their surrounding areas. SUBITO addresses the problem of automated real time detection of abandoned luggage, fast identification of the owner and his/her subsequent path and current location.
The organization of intelligent video surveillance systems is hierarchical and generally starts with object detection, estimates the position of the detected object over time (object tracking) and describes what happens in the scene (event recognition).

The Elsag Datamat SuperVision system (SV) is a set of software technologies, whose aim is to support video surveillance systems development. It consists of a set of modules:

- A server that includes scene analysis, scene interpretation and alarm generation (it is the core of the system);
- A client that allows interaction with the user through streams viewer;
- A digital recorder of video streams, alarms and metadata.

The core algorithm of the SuperVision system uses Video Content Analysis technologies to describe the scene symbolically and to produce alarms when potentially hazardous actions occur: for example unattended baggage or crossing of not allowed areas are detected. The SV exploits special cameras, like the omnidirectional with fisheye or catadioptric optics and can drive a Pan/Tilt/Zoom (PTZ) camera to follow specific targets in the monitored area and it is capable of providing the operator with high resolution images of the area of interest.

The scene is described in world coordinates that consider the actual size of the monitored area and of the detected objects.

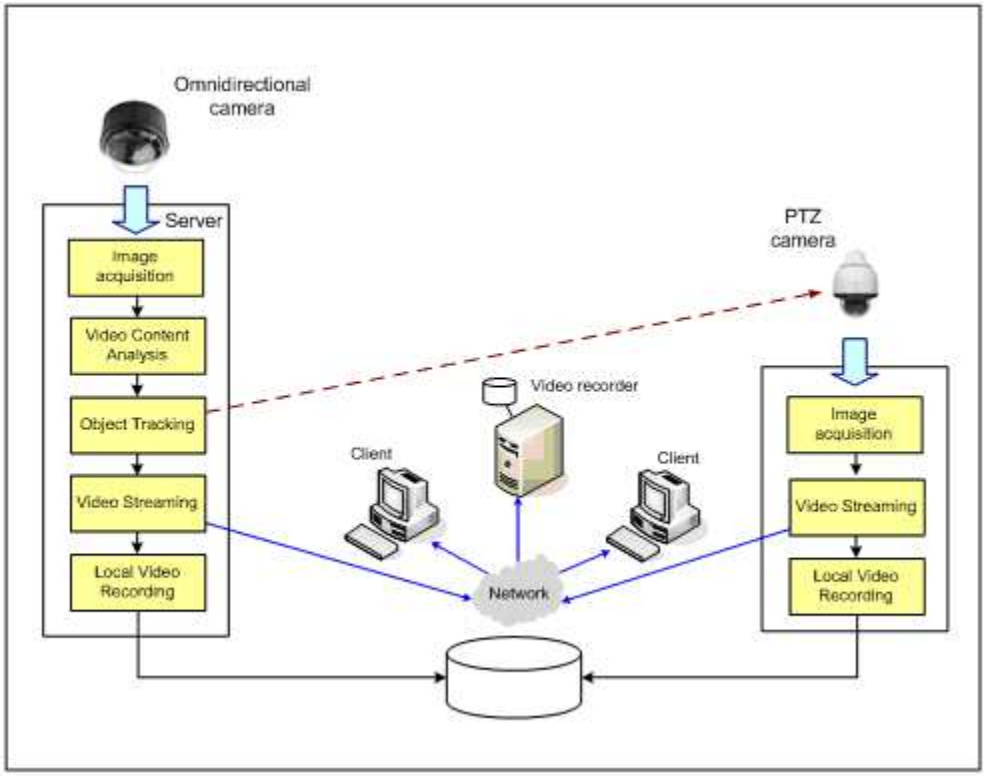


Fig. 1. Process flow

2. Features of SuperVision

The SuperVision system includes many video analysis functions that can be grouped into the following classes:

1. *General Video Content Analysis*. To extract synthetic information from video streams and describe moving objects in the scene. For each detected target, position, dimensions, type (person, vehicle, baggage), trajectory and radiometric features (intensity, colour) are extracted. (This information is called "metadata" and is collated to the video stream).
2. *Video Analytics*. Used, together with the analysis of metadata, to recognize preconfigured dynamic events and to report them as warnings or alarms. Typical filters are tripwire, unattended baggage, loitering, speed, abnormal direction detection. Video analytic applications are expected to grow meeting new requirements.
3. *Graphic Features*. Graphic features modify, with suitable transformations, images that are presented to the operator. The most common transformations are projection on the ground and correction of panoramic images (cylindrical projection and virtual PTZ).
4. *Support Features*. To help set up the operating environment. They include camera calibration (required for the system precision) and diagnostic functions (required to guarantee continuous operations).
5. *Automatic drive of PTZ camera*. It can be useful to capture high resolution images in the most interesting areas of the scene, although it isn't properly a video content analysis feature. Typically the PTZ is controlled automatically by the video analytic applications output, but it can also be controlled by Video Content Analysis metadata.

3. Omnidirectional optics

The omnidirectional optics can capture the scene with a 360° horizontal angle. Several (correlated) traditional cameras would be necessary to obtain an equivalent view. Common used optics are fisheye and catadioptric optics. The choice depends on applications. The former cover the half plane from the vertical to the horizon and are suited for fixed and dominant positions (high above the ground). The latter can look over the horizon but have a blind zone and they are suited for mobile applications and low height. The SuperVision system can manage both. In other words it's capable to convert their image coordinates in world coordinates (camera calibration is required).

3.1 Fisheye optics

The field of view of the fisheye optics is from 0° to about 95°. They cover the half plane from the vertical up to the horizon and thus they don't have a central blind zone. Their drawback is that the angular resolution goes down rapidly when moving away from the optical centre. The fisheye lens projects the image on the sensor with a transformation:

$$r = f \cdot \theta \tag{1}$$

Where r is the distance on the sensor from the optical centre, expressed in pixel, θ is the paraxial angle expressed in radiant and f is the focal length (in pixels) of the fisheye.

Fig. 2. Fisheye optic. On the left the optic's schema, on the right a fisheye lens.

3.2 Catadioptric optics

The catadioptric optics is formed coupling refractive lenses and reflective surfaces. In this way it is possible to achieve a field of view that extends above the horizon. The drawback of these types of optics is the presence of a blind zone at the optical centre that corresponds to the minimum tilt angle visible by the mirror. In other words the field of view is a ring that covers 360 degrees horizontally but only about from -10° to + 10° degrees vertically.

A convex mirror coupled with a traditional lens is the simplest configuration of catadioptric optics. The constraint is that the axis of symmetry of the mirror must coincide with the optical axis of the camera. This way the centre of projection is unique (Baker & Nayar, 1999) and it is quite simple to rectify the distortion due to the mirror and to produce geometrically correct planar prospective images. The profile of the mirror is typically hyperbolic because it's the only shape capable of removing the astigmatism, achieving a homocentric system. Unfortunately this solution has a strong drawback: it considerably reduces the resolution. In other words it's impossible to bring into focus the entire image.

Fig. 3. Single mirror catadioptric optic. On the left the optic's schema, on the right a catadioptric lens**.**

An alternative consists of two mirrors, one convex and one concave, coupled with a traditional lens. The profiles of the mirrors are (in general) parabolic and thus the system is homocentric. The presence of the two mirrors increases the degrees of freedom, attaining a partial correction of the curvature.

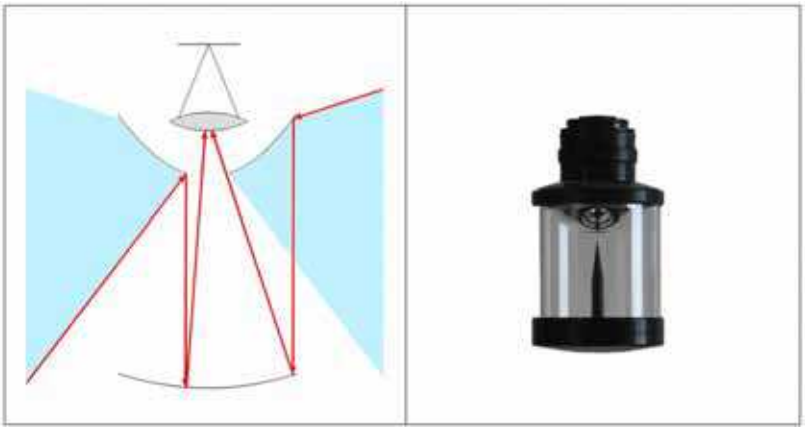


Fig. 4. Double mirror catadioptric optics. On the left the optics design principle, on the right an example of catadioptric lens.

A third solution, more expensive to realize, is a special refractive surface coupled with two reflective surfaces and a traditional lens. The refractive surface and the reflective one are realized with a single piece of optical glass and the reflection occurs inside the glass. The refractive surface is homocentric and its profile is described with a forth order polynomial. One of the reflective surfaces has a concave elliptical profile and the other a convex parabolic profile. Thus the whole system is homocentric. In addition the three surfaces increase the degrees of freedom and it is possible to accurately correct the curvature.

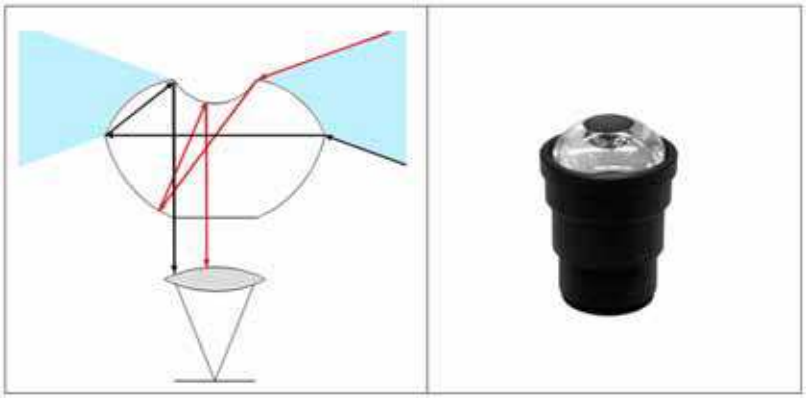


Fig. 5. Catadioptric optic with three surfaces. On the left the optic's schema, on the right a catadioptric lens.

The catadioptric optics projects the images with a transformation quite similar to the conformal projection:

$$r = 2 \cdot f \cdot tan\frac{\theta}{2} \tag{2}$$

where r is the distance on the sensor from the optical centre expressed in pixel, θ is the paraxial angle expressed in radiant and f is the focal length (expressed in pixels).
The peculiarity of the conformal projection is that both the angular resolutions (sagittal and tangential) are the same. Thus the image is not locally deformed. Furthermore the angular resolution increases with the distance from the optical centre.

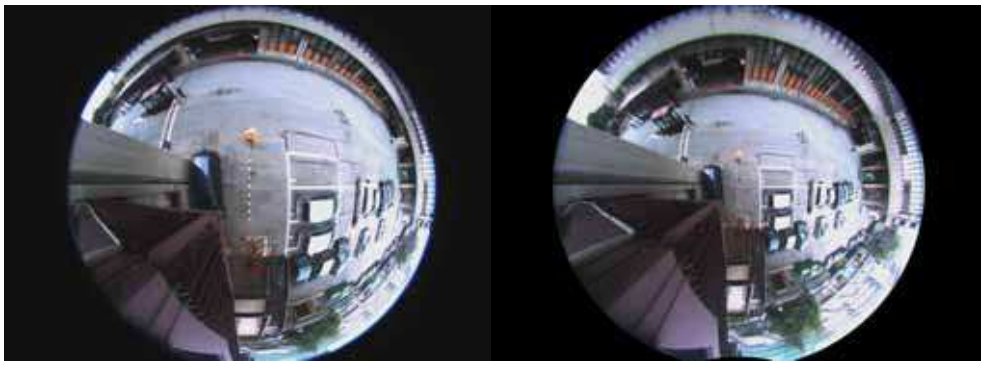


Fig. 6. Comparison between the fisheye projection (on the left) and the catadioptric projection (on the right). In the catadioptric projection the local proportions are respected and the distortion at the centre of the image is lower than in the fisheye projection.

4. Video Content Analysis features

The Video Content Analysis is a technology that evaluates the contents of a video stream to extract specific information. The SuperVision system is a modular system that implements this kind of features according to the diagram in figure 7. Each module operates at different levels of abstraction. The first level deals with individual pixels (treated as individual entities). The output of these modules shows only the variation of the pixel appearing in different frames. The background updating module and the foreground detection module belong to this level. The purpose of these modules is to identify motion areas in the scene.
The second level does not consider pixel individually but as groups. At this level the clusterization of the foreground takes place, as well as the absorption of still groups in the background and clusters classification.
Finally the third level considers frame sequences and introduces temporal correlation among clusters. The tracking module belongs to this level.

4.1 Background updating

Moving object detection is a low level task necessary for the comprehension of high level events. There are different approaches (Elhabian et al., 2008). The SuperVision system

implements a statistical approach and considers each pixel independently basing its processing on the pixel temporal history.
First of all in a mobile temporal window the module calculates the average and the standard deviation of the pixel intensity. There is a state machine for each pixel that decides to update or not the tolerance and the reference. These four variables and the decision to update or not the background are stored in the "status of background" to be used in the next frame. Updating of background is based on the difference between the standard deviation and the tolerance. If the first is greater than the second the pixel is considered as belonging to a moving object and its reference and tolerance value are not updated.
A counter is initialized to measure the length of the lock, i.e. a measure of how long the background has been maintained unchanged. If the counter exceeds a predefined threshold a background update is forced (action performed by the "absorption of still clusters" module).

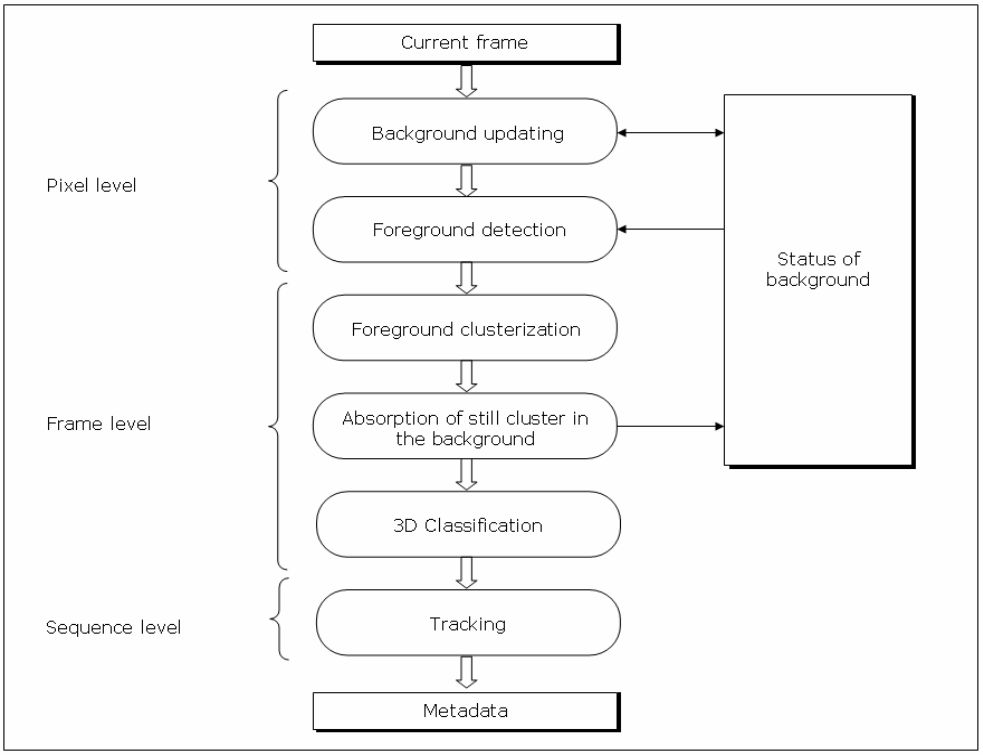


Fig. 7. SuperVision core algorithm flow diagram

4.2 Foreground detection

The foreground pixels are those with a considerable distance from the reference value of the background. The module decides if a pixel belongs to the foreground by comparing its intensity value with the reference and tolerance values. The module shows a noise tolerance as it is based on the signal standard deviation.

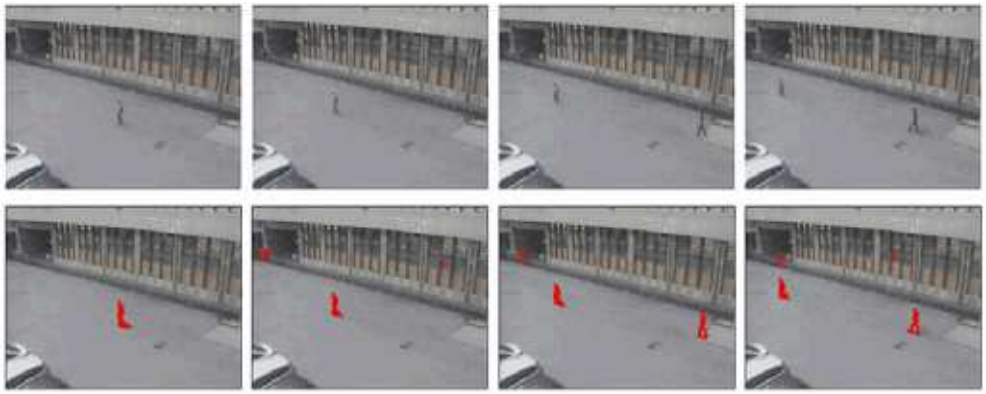


Fig. 8. Output of the foreground detection.

4.3 Foreground clusterization

The foreground clusterization module action is at frame level. It works aggregating neighbouring pixel. Pixels that are closer to each others have high probability of belonging to the same object. These clusters represent the area of interest for the next process steps.

4.4 Absorption of the cluster in the background

For each cluster of interest every pixel is under the control of the algorithm. When a considerable number of pixels are in a locked status, a background update is forced. All the information about the background and the foreground of the cluster are stored on a dedicated buffer. This way it is possible to control the absorption of stationary targets for a long time, by preserving their state in memory. Stationary targets do not interfere with moving objects and, if they start to move, they can be recognized thanks to the information saved in the status buffer.
This mechanism allows a multilayer management of the background.

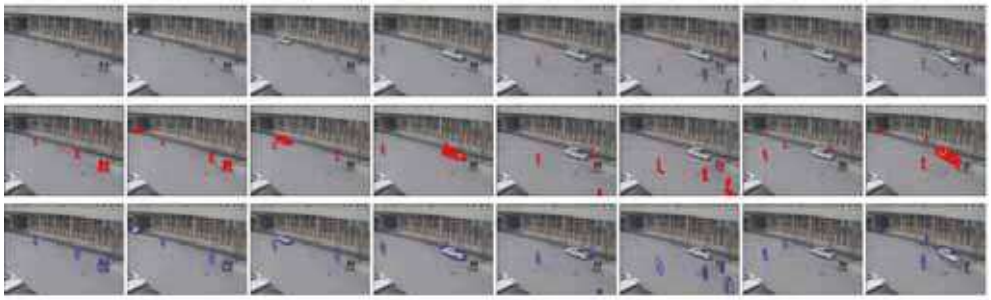


Fig. 9. Output of the foreground clusterization. In the first row some frame of the original sequence, in the middle the output of the foreground detection and in the bottom, circled in blue, the clusters. We can see how still cluster (like the two persons in the third frame and the car in the fifth frame) are reabsorbed into the background and they don't interfere with the other moving targets.

4.5 3D classification

The 3D classification module works with real world coordinates. For each cluster of interest it projects, on the image, different 3D models of the target. For each 3D model the "silhouette" is calculated and is compared with the cluster. The model with the greatest likeness is chosen and its position and its orientation in world coordinate are calculated.
For the next step, the tracking module, the real position of the cluster and its radiometric information are also required. Finally this module can be used to segment clusters into different targets if they are too close. Alternatively the same module can fuse more clusters into a single target if they are fragmented. Camera calibration (to convert image coordinates into world coordinates) is needed to project the model on the image.

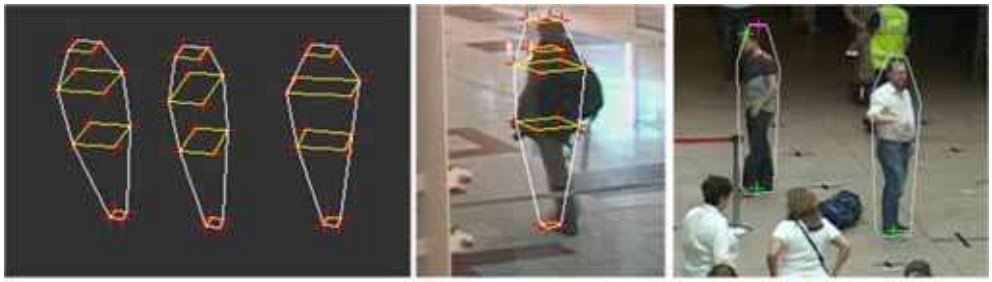


Fig. 10. On the left different 3D model of a person. In the middle and on the right the output of the 3D classification.

4.6 Tracking

The task of the tracking module is to estimate the trajectory of a moving object in the scene (Yilmaz et al. 2006). This module analyzes the target behavior in time, through the re-identification of the detected targets on a frame by frame basis.

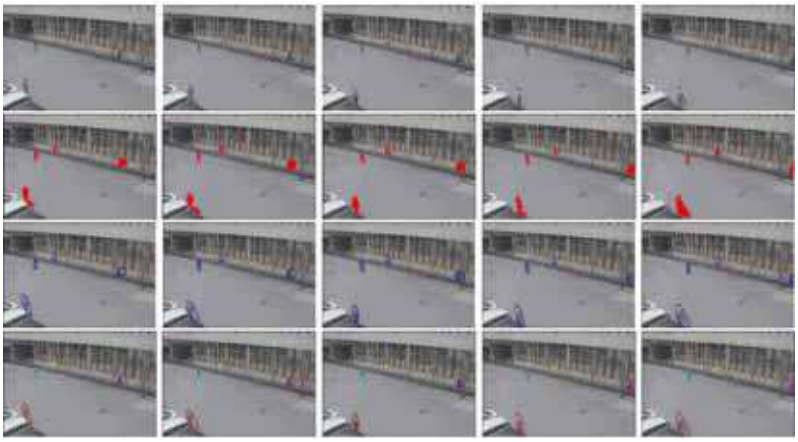


Fig. 11. Output of tracking module. In the first row the original sequence, in the second row the motion detection output, in the third one the foreground clusterization output and in the last row the tracking output. Each person is identified with a different colour, coherently with time.

The re-identification is based on the neighbourhood, the speed in world coordinates and the geometric and radiometric features of the target. It is then possible to associate a label coherent with time and thus it is possible to describe not only statically but also dynamically all the detected targets.
Each target is defined by its metadata: information about position, velocity (trajectory), dimension, radiometric features and classification (person, car, etc.)

5. Video Analytic applications

The effectiveness of the surveillance system operation can be considerably increased by Video Analytics. These are specialized applications that perform the task of real-time event detection as well as post-event analysis, saving manpower costs and providing an immediate alert upon detection of a relevant event. These applications are also called filters and they must be customized by parameterization. The parameterization is a procedure to customize the events to detect. Instead of general functions which produce output frame by frame, filters produce asynchronous events.
Typical filters are target tracking, tripwire, abnormal direction, speed, unattended object, loitering, graffiti and vandalism acts, falls, climbing, people counting and crowding detection. In the following readers can find a brief description of the most used filters.

1. ***Object detection.*** It's used to detect targets in one or more regions of the area under surveillance.
2. ***Tripwire detection.*** It's used to detect targets entering a restricted area. When a target passes on a tripwire, an alarm is raised.

Fig. 12. Example of tripwire detection. The detected target is circled in red.

3. ***Abnormal direction detection.*** The abnormal direction detection filter is used to detect objects moving in opposite direction with respect to the expected direction.
4. ***Detection of fast moving objects.*** This filter is used to detect moving targets travelling above a pre-defined speed.

5. ***Unattended/Removed object detection.*** Used to detect still targets in the scene. It can be used to detect unattended or removed objects.
6. ***Loitering detection***. Unlike the speed detection filter, it is used to detect moving targets that remain in an area for longer than a predefined time.
7. ***Graffiti and vandalism acts detection.*** With this filter it is possible to detect relevant and permanent background changes. This event can be detected only with significant delay.
8. ***Detection of people laying (fall detection).*** It can be used to detect people laying, in particular to detect falling people.
9. ***Climbing detection***. Similar to fall detection filter, climbing detection is used to detect people assuming abnormal position and pose. Generally is used to detect people that climb turnstiles or barriers.
10. ***People counting passing over a tripwire.*** This filter is used to define how many people pass across a tripwire and the direction of transit. The output of the filter is a series of event at the predefined frequency supplied with the following information: the tripwire label, the time and the number of people crossing the tripwire during the last interval of time. In this case the events are not random but at regular intervals.
11. ***People counting in a region of interest.*** Similar to counting people passing across a tripwire filter, it is used to evaluate how many people are in a predefined area. The output of the filter is a series of events at the predefined frequency with the following information: the area of interest, the time and the number of people in the area. As in the previous case the events are not random but at regular intervals.
12. ***Crowding detection.*** It can be used to detect the presence of excessive crowding in one or more predefined areas.

The parameters that can be customized are summarized in the following table:

	Area or Tripwire of interest (*)	Class of the object	Availability of a "SV controlled" PTZ	General parameters (i.e. inertia and radiometric sensitivity)	Motion or crossing direction	Minimum time of persistence in the detected status (**)	Dimensions of the object to detect	Threshold (***)	Counting frequency
Object detection	•	•	•			•			
Tripwire detection	•	•	•		•				
Abnormal direction detection	•	•	•		•	•			
Overspeed detection	•	•	•	•		•		•	
Unattended/ Removed object detection	•		•	•		•	•		

	Area or Tripwire of interest (*)	Class of the object	Availability of a "SV controlled" PTZ	General parameters (i.e. inertia and radiometric sensitivity)	Motion or crossing direction	Minimum time of persistence in the detected status (**)	Dimensions of the object to detect	Threshold (***)	Counting frequency
Loitering detection	•	•	•	•		•			
Graffiti and vandalism acts detection	•							•	
Detection of people laying (fall detection)	•		•	•		•		•	
Climbing detection	•		•	•		•		•	
People counting passing over a tripwire	•				•				•
People counting in a region of interest	•								•
Crowding detection	•		•	•		•		•	

Table 1. Customizable parameters

(*) The area or tripwire is defined by drawing a polygonal over the image or a ground projection of the area under surveillance.
(**) This parameter has different interpretation according to the filter. Its meaning is reported in the following table:

Filter	**Minimum time of persistence in the detected status**
Object detection	Minimum time the object must remain in the area to be detected
Abnormal direction detection	Minimum time the object must remain in the area to be detected
Overspeed detection	Minimum time of persistence of the object above the speed limit
Unattended/Removed object	Minimum time of persistence in the motionless

Filter	Minimum time of persistence in the detected status
detection	status
Loitering detection	Minimum time of persistence in the predefined area
Detection of people laying (fall detection)	Minimum time of persistence in the position
Climbing detection	Minimum time of persistence in the abnormal position
Crowding detection	Minimum time of persistence in the crowding status

Table 2. Meaning of the minimum time of persistence parameter

(***) This parameter has different interpretation according to the filter. Its meaning is reported in the following table:

Filter	Threshold
Overspeed detection	Speed limit
Graffiti and vandalism acts detection	Extension (in percentage) of the change to detect
Detection of people laying (fall detection)	Laying pose, expressed as percentage of the height of the object
Climbing detection	Abnormal pose, expressed as percentage of the height and the barycentre value
Crowding detection	Crowding (number of people)

Table 3. Meaning of the threshold parameter

6. Coordinate transformation support

A class of functions and tools that convert point's coordinates into different reference systems and to rectify images from omni-directional devices.
These operations are required to generate images that are more easily understood by surveillance operators and to generate ground projections of the images. The latter case is useful to define the tripwire and thus to generate alarms when these barriers are crossed. By coupling these features with the camera calibration, virtual PTZs can be implemented.

6.1 Coordinates conversion

As already stated, the SuperVision system works in real world coordinates, so it is necessary to convert the various coordinate reference systems.
The reference systems types are:

- Image coordinates, expressed in pixel, represent the point coordinates on the stored image buffer. They are indicated with the couple (U,V);
- Focal coordinates, expressed in pixel, represent the point coordinates on the focal plane. They are indicated with the vector (Q_1, 0, Q_3);
- Local world coordinates, expressed in meters, represent the point coordinates on the local system of reference of the camera. They are indicated with the vector (P_1, P_2, P_3).

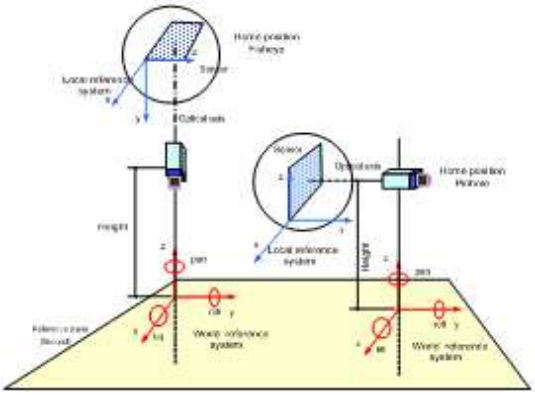


Fig. 13. Camera reference system for local world coordinates (ground projection of the camera position)

In the following sections the conversions between different systems of reference are described.

6.1.1 Image coordinates to focal coordinates

This conversion takes into account the optical centre position on the image (U_C, V_C), the anisotropy of the sensor (C_A) and the (possible) distortion. If the distortion is positive the warped image has distances from its centre greater than the original image (pin cushion distortion) and the focal coordinates are calculated as follows:

$$Q_1 = \frac{x}{y} \cdot \hat{Q}_1 \qquad Q_3 = \frac{x}{y} \cdot \hat{Q}_3 \tag{3}$$

where:

$$\hat{Q}_1 = U - U_C$$
$$\hat{Q}_3 = \frac{V}{C_A} - V_C$$
$$\hat{Q} = \sqrt{\hat{Q}_1^2 + \hat{Q}_3^2}$$
$$y = \frac{\hat{Q}}{R_{MAX}}$$

x is obtained inverting the polynomial $y = x + ax^3 + bx^5$, $a \geq 0$, $b \geq 0$.

a is the third order distortion coefficient, b is the fifth order distortion coefficient and R_{MAX} is the half diagonal of the image expressed in pixel.

If the distortion is negative ($a \leq 0$, $b \leq 0$) the warped image has distances from the centre less than the original (barrel distortion) and the focal coordinates are calculated as follows:

$$Q_1 = \hat{Q}_1 \cdot \left[1 - a \cdot \left(\frac{\hat{Q}}{R_{MAX}} \right)^2 - b \cdot \left(\frac{\hat{Q}}{R_{MAX}} \right)^4 \right] \quad Q_3 = \hat{Q}_3 \cdot \left[1 - a \cdot \left(\frac{\hat{Q}}{R_{MAX}} \right)^2 - b \cdot \left(\frac{\hat{Q}}{R_{MAX}} \right)^4 \right] \tag{4}$$

6.1.2 Focal coordinates to local world coordinates

This conversion varies with the camera type.
For a pin-hole camera the conversion from focal coordinates to world coordinates is obtained using the following formulas:

$$P_1 = Q_1 \qquad P_2 = F \qquad P_3 = Q_3 \tag{5}$$

where F is the focal of the camera.
For a fish-eye camera, instead, the conversion is obtained using the following formulas:

$$P_1 = Q_1 \qquad P_2 = E \qquad P_3 = Q_3 \tag{6}$$

where

$$E = \begin{cases} \dfrac{Q}{tan\,\theta} & per \quad \theta \geq \dfrac{1}{500} \\ F & per \quad \theta < \dfrac{1}{500} \end{cases}$$

e

$$\theta = \frac{Q}{F}$$

6.2 Projections

The rotational symmetric lenses can be represented by a single general model. This model, in polar coordinates, is described by the equations:

$$\begin{cases} u = r(\theta) \cdot cos\,\phi \\ v = r(\theta) \cdot sin\,\phi \end{cases} \tag{7}$$

where
u,v are the coordinates of the point in the focal plane, θ is the paraxial angle of the field of view, *r* is a function of θ, and ϕ is the polar angle.
The local sagittal focal length is expressed by the formula:

$$F_S = \frac{\partial r}{\partial \theta} \tag{8}$$

while the local tangential focal length is the quantity:

$$F_T = \frac{r}{sin\,\theta} \tag{9}$$

Different values of r(θ) produce different projections. The most common are:

- *Perspective projection or Gnomonic projection*. The pinhole camera is the simplest device to capture the geometry of perspective projection. The relationship between a point on the image plane and a point on the focal plane is:

$$r = F \cdot \tan\theta \tag{10}$$

with $\theta < \frac{\pi}{2}$.

Homogeneous coordinates handle the problem in a linear way. In this kind of projection the sagittal and the tangential focal length are different, that is the local object proportions in the projected image are not maintained, as can be inferred by the following formulas:

$$F_S = \frac{F}{\cos^2\theta} \tag{11}$$

$$F_T = \frac{F}{\cos\theta} \tag{12}$$

- *Conformal cylindrical projection*. It can be used to represent cylindrically a panoramic image. In this kind of projection the sagittal and the tangential resolution are equal, thus the local proportions of the object are maintained.

$$\begin{cases} U = F \cdot \phi \\ V = F \cdot ln\left[\tan\frac{\theta}{2}\right] \end{cases} \tag{13}$$

$$F_S = F_T = \frac{F}{\sin\theta} \tag{14}$$

θ cannot reach the limit value 0 and π.

- *Stereographic projection*. It can be used for polar representation of a panoramic image (projection of a sphere onto a plane). Like the cylindrical projection it is a conformal transformation, that is it preserves angles and thus the local proportions of the objects. The r(θ) function for this projection is:

$$r = 2F \cdot \tan\frac{\theta}{2} \tag{15}$$

and the sagittal and the tangential focal length are:

$$F_S = F_T = F \cdot \cos^{-2}\frac{\theta}{2} \tag{16}$$

In the SuperVision system it is used for the virtual PTZ.

6.3 Virtual PTZ

The virtual PTZ function simulates the behaviour of a PTZ camera using as source the panoramic images. Coupling this feature with a high resolution sensor it's possible to produce detailed views of any part of an image, based on virtual parameters of pan, tilt and zoom. This allows the monitoring of areas that normally are not controlled by traditional camera and, thanks to the high resolution, to extract features of particular interest such as detail of the action occurring, plate numbers or faces. An example is shown in figure 15. The camera is mounted on a car. This allows to control what happens around it. The virtual PTZ can be activated in an automatic way, choosing on the rectified image the target of interest and requesting tracking, or manually when only the prospective projection of a portion of image is required.

Fig. 14. Example of a virtual PTZ. On the left a fisheye image. In the middle and on the right the prospective projections of the circled areas, based on the pan, tilt and zoom parameters.

Fig. 15. Example of a virtual PTZ. On the left an image from a catadioptric lens mounted over a car. In the middle the projection of the area circled in red and on the right the projection of the area circled in green.

7. Support features

Camera calibration and diagnostic functions are the support features of the SuperVision system. Camera calibration evaluates the real dimensions of the targets in the scene to track

their position and to classify them. It is also required to put in relation real points and the remote control parameters of the PTZ used to track objects.
The diagnostic functions, instead, allow the detection of tampering or faults in the system, such as the obscuring, shift of the camera, malfunctioning, etc.

7.1 Camera calibration

The target of the calibration is the evaluation of a set of parameters of the camera and the position of the camera in the world (camera coordinates and camera orientation) with respect to a reference plane (typically the ground plane). This process is necessary in those applications where metric information of the environment derives from images. In order to do the evaluation, the calibration process needs information about the image positions and world measures on a set of "calibration points".
The user specifies a set of segments anchored to the reference plane in two possible ways: either the segments lie on the plane (horizontal segments) or on a line normal to the plane with one end touching the horizontal plane (vertical segment). The only world information needed is the actual segment length.
Starting from the knowledge of some parameters such as the dimensions in pixel of the sensor, the position of the optical centre and, for a fisheye camera, the maximum radius of the image, it is possible to estimate the intrinsic parameters, that describe the internal geometry of the camera (focal expressed in pixel) and the extrinsic parameters, that define its position and its orientation (roll angle, tilt angle, height).
In this way the camera is calibrated on a reference system on the ground. The origin of this reference system corresponds to the on-ground camera projection, X and Y axes lay on the ground with Y direction parallel to the on-ground optical axis projection. In this reference system three camera extrinsic parameters (X, Y and pan angle) are always null.

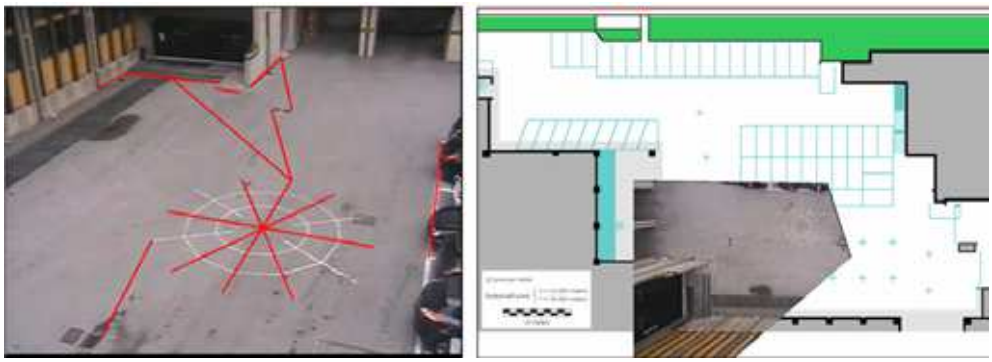


Fig. 16. On the left the image from the uncalibrated camera with the known length segments. On the right the ground projection of the image based on the calibration parameters (X=61.18 m, Y=18.52 m, Z=10.40 m, Pan=109.59°, Tilt=-30.75°, Roll=0.78°)

When two or more cameras have been calibrated, it is possible to link them to a common reference system. After selecting one of the cameras as the reference camera, the three null extrinsic parameters (X, Y and pan) of the other cameras are computed.

This process is called registration. For each camera the registration is equivalent to a rototranslation of the on-ground projection image. In order to evaluate the correct rototranslation, a set of corresponding points (on the ground) are selected on the two on-ground projection images, and a least-square method is applied. In this way it's possible to define not only the real dimensions of the objects in the scene, but also their position with respect to a reference system shared from all cameras (e.g. necessary for multicamera tracking). If a site map is available, it can be used as a substitute of the reference camera. Cameras position and orientation are anchored to any given reference system. After calibration and registration it is possible to generate a composite image using portions of the ground plane projection images.

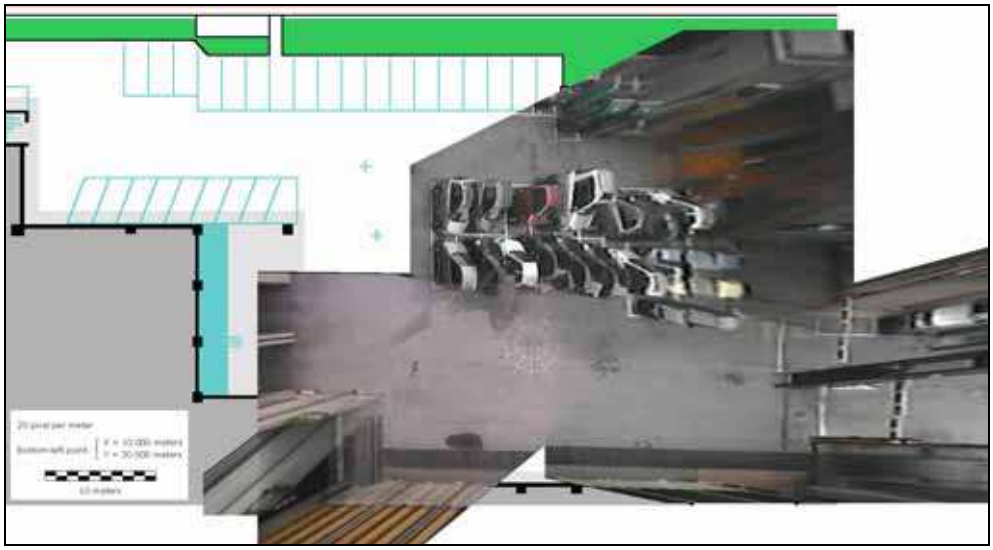


Fig. 17. Map obtained with the multicamera calibration.

In the SuperVision system, the "controlled" PTZs are also calibrated. This is necessary in order to control these cameras based on the detected events referenced to world coordinates. The PTZ calibration is required to obtain correspondence between the PTZ camera and images from the traditional or omnidirectional camera. When the features of the camera (focal and dot pitch), the pan and the tilt angle and the target position in the image are known the system returns to the PTZ the relative angular shifts.

7.2 Diagnostic functions

The diagnostic functions provide alarms when the effectiveness of a camera is compromised. Lack of signal, voluntary or accidentally shift of the camera from its setup position, bad quality of the signal, due for example to the obscuring of the lens or to the loss of focus, are the detected events.

The operator is immediately alerted and solicited to take corrective actions.

Camera shift is detected comparing the background, which contains fixed elements of the scene, with the video stream. The background is periodically updated to take into account the environmental changes. The gradient of these images is calculated and then a binary threshold is applied. The Hausdorff's distance between the gradients is used to decide if a change in the background occurs or not. In the following picture the output of the comparison of the two gradients is reported. When a change in the scene occurs, the background superimposed on the new image tends to rapidly disappear (fig. 18 e) and the distance between the two images is high.

Fig. 18. Example of diagnostic function. The figure a) shows a "standard" scene, that contains its invariant elements (in this case the upper part of the image). The figure b) shows the gradient of the background after applying the binary threshold. Figures c) and d) show the scene before and after the shift of the camera and in figure e) the result of the comparison.

8. Automatic control of PTZ cameras

The PTZ camera is automatically controlled by the SuperVision system, based on the world coordinates received from the fixed system. This allows high resolution tracking of a specific target, extraction of interesting features and continuous pointing of the camera only to those areas where action is detected, avoiding the storage of video of limited interest.
When an event is detected, for instance the crossing of a tripwire, the PTZ camera sets itself on the target which has produced the event and it moves according to the target position upon the image plane. The quantity of shift is known thanks to the camera calibration. Once the tilt angle of the first position is known, the shifts are calculated whenever the target is too close to the image boundaries. In this way the object is always into the camera field of view.

Fig. 19. Examples of PTZ automatic control.

9. Conclusions and future works

The aim of all video surveillance automatic systems is to avoid tedious chores for the operators. In this paper we described Elsag Datamat approach to satisfy these requirements. The SuperVision system can analyze video streams from different types of camera, in particular omnidirectional, and to set alarms when preconfigured events are detected. The types of events detected are numerous, and can be composed according to the context and needs of applications. Future improvements will include new modules, such as face detection and recognition, for automatic check of people's identity.

10. References

Baker, S. & Nayar, S. K. (1999). A theory of single-viewpoint catadioptric image formation, *International Journal of Computer Vision (IJCV)*, Vol. 35, No. 2, pp. 175–196

Elhabian, S. Y., El-Sayed, K. M. & Ahmed, S. H. (2008). Moving object detection in spatial domain using background removal techniques – State-of-art, *Recent Patents on Computer Science*, Vol. 1, No. 1, January 2008, pp. 32-54, ISSN: 1874-4796

ISCAPS project (2006). www.iscaps.reading.ac.uk

McCahill, M. & Norris, C. (2003). CCTV systems in London: their structures and practices, In: *On the threshold to Urban Panopticon?: Analysing the Employment of CCTV in*

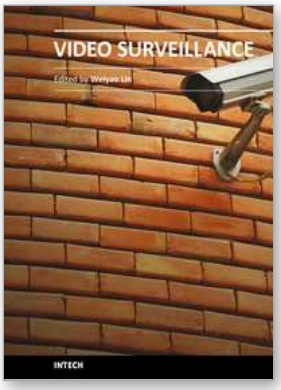
European Cities and Assessing its Social and Political Impacts, Technical University, Berlin.

SAMURAI project (FP7/2008-2011). www.samurai-eu.org

Smith, G.J.D. (2004). Behind the screens: examining constructions of deviance and informal practices among CCTV control room operators in the UK, *Surveillance & Society. CCTV Special*, Vol. 2 Issue 2/3, pp. 376–395.

SUBITO project (FP7/2009-2011). www.subito-project.eu

Yilmaz, A., Javed, O. & Shah, M. (2006). Object tracking: A survey. *ACM Computing Surveys,* Vol. 38, No. 4, December 2006, pp. 1-45, ISSN:0360-0300



Video Surveillance
Edited by Prof. Weiyao Lin

ISBN 978-953-307-436-8
Hard cover, 486 pages
Publisher InTech
Published online 03, February, 2011
Published in print edition February, 2011

This book presents the latest achievements and developments in the field of video surveillance. The chapters selected for this book comprise a cross-section of topics that reflect a variety of perspectives and disciplinary backgrounds. Besides the introduction of new achievements in video surveillance, this book also presents some good overviews of the state-of-the-art technologies as well as some interesting advanced topics related to video surveillance. Summing up the wide range of issues presented in the book, it can be addressed to a quite broad audience, including both academic researchers and practitioners in halls of industries interested in scheduling theory and its applications. I believe this book can provide a clear picture of the current research status in the area of video surveillance and can also encourage the development of new achievements in this field.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821