Multimedia Tools and Applications (2019) 78:17437–17459
https://doi.org/10.1007/s11042-018-7065-3

Scalable and flexible wireless distributed architecture for intelligent video surveillance systems

Isaac Martín de Diego[1] · Ignacio San Román[1] · Javier Cano Montero[1] · Cristina Conde[1] · Enrique Cabello[1]

Received: 22 December 2015 / Revised: 30 November 2018 / Accepted: 11 December 2018 /
Published online: 8 January 2019

Abstract

This paper presents a novel distributed intelligent video surveillance architecture based on Wireless Multimedia Sensor Networks (WMSNs). This architecture is part of a video surveillance project and has been built using the Robot Operating System (ROS). ROS allows to develop and connect (through wireless TCP-IP) several modules to process and manage multimedia data in an easy way. The real-time intelligent surveillance system has been trained for detecting, tracking and monitoring people and vehicles in an indoor-outdoor real environment. The test process shows the reliability of the developed system as a tool for the identification of security incidents. Besides, using wireless connections and a distributed architecture together, we have achieved a really flexible, easy to install and lower-maintenance system that supports many different devices. Thus, the proposed architecture can be applied in distributed locations such as smart cities.

Keywords Intelligent video surveillance architecture · Distributed system · Robot operating system · Wireless multimedia sensor network

✉ Isaac Martín de Diego
isaac.martin@urjc.es

Ignacio San Román
ignacio.sanroman@urjc.es

Javier Cano Montero
javier.cano.motero@urjc.es

Cristina Conde
cristina.conde@urjc.es

Enrique Cabello
enrique.cabello@urjc.es

[1] Face Recognition and Artificial Vision group, University Rey Juan Carlos, Móstoles, Spain

1 Introduction

The development of training technologies for deployment of automated systems, capable of interpreting and understanding the dynamics of objects and events in multimedia environments, is intended to become a valuable tool for modern society and smart world. New systems incorporating Wireless Multimedia Sensor Networks (WMSNs) are being increasingly applied to smart cities. Thus, the smart cities concept is becoming a reality due to the deployment of a variety of those technologies. One of the main research areas where WMSNs can be widely applied is Video surveillance since this kind of technologies can provide easier installations and lower maintenance systems. In Video surveillance environments, a huge volume of information is generated. Distributed architectures allow to manage this amount of information in different locations and to share only relevant information. Industry reports predict that the global Video surveillance market will grow at an annual rate of 16.97% in the coming future [34], specially in the case of smart cities. Latest reports indicate that in the future, surveillance solutions efforts will focus on the introduction of video analysis and the concept of cloud computing. In fact, intelligent video surveillance is one of the most active research areas in computer vision. Many different types of video surveillance systems have been developed [20, 22, 24, 37]. A video surveillance system is a combination of hardware and software components that are used to capture and analyze video. The primary aim of these systems is to monitor the behaviour of objects (usually people or vehicles) in order to check for suspicious or abnormal behaviours, using the extracted information of those objects (physical features, trajectories, speed…) from a variety of sensors (such as surveillance cameras). The sensor sensitivities from a variety of sensors are decisive phenomena for extracting accurate information based on the size in pixels for high resolution 2D or 3D cameras, the temperature in degrees obtained from thermal information cameras, and time in seconds for any activities in the scene. In sensor's operating range, an ideal sensor has a constant sensitivity. In saturation state, sensor can no longer respond to any changes. For real-world applications, the scene composition (e.g., occluded or abandoned object(s)) is an important issue for activities detection based on an accurate information extraction process that a video surveillance system's performance is highly affected from actual scene composition and sensor sensitivities (see, for instance, [11]).

A review of computer vision and pattern recognition technologies used in intelligent multi-camera video surveillance is presented in [35]. A summary of surveillance systems generations evolutions as well as previous developments on object detection, recognition, tracking, behaviour analysis and storage can be found in [33]. However, there is not still a standard solution that supports the entire life cycle of these information systems: acquisition of information, video processing, categorization and management, including automatic detection of risk situations and decision support, all of them based on context information according to the specific characteristics of the monitored environment. In addition, most of these systems do not allow automatic re-configuration based on previous behaviours in the system. For instance, to switch a part of the system on (one or several sensors) given an alarm generated in other part of the system based on the combined information from one or several sensors.

INVISUM (Intelligent VideoSurveillance System) is a project funded by the Spanish Ministry of Economy and Competitivity focused on the development of an advanced and complete security system [8]. The goal of the project is the development of an intelligent video surveillance system that addresses the limitations of scalability and flexibility of current video surveillance systems incorporating new compression techniques, pattern detection, decision support, and advanced architectures to maximize the efficiency of the system. Thus, a modular platform that facilitates the integration of modules from different manufactures and based on various technologies, is desired. In this paper we present a context-aware wireless distributed architecture for a real-time intelligent surveillance system that is able to detect and track multiple people and vehicles, monitoring their behaviours in a combined indoor-outdoor environment. The system has been designed thinking on many different types of devices (sensors and computers) should be able to be connected among them in an easy-way. Thus, the architecture described below is aimed at allowing wireless (as well as wired) connections and a flexible processes distribution in a cloud and highly modular environment.

The paper is organized as follows. Section 2 details the wireless distributed architecture of the proposed system. Section 3 describes some developed high level components that use this architecture. A case of study is presented in Section 4. Section 5 summarizes a discussion of the main results. Section 6 provides the concluding remarks.

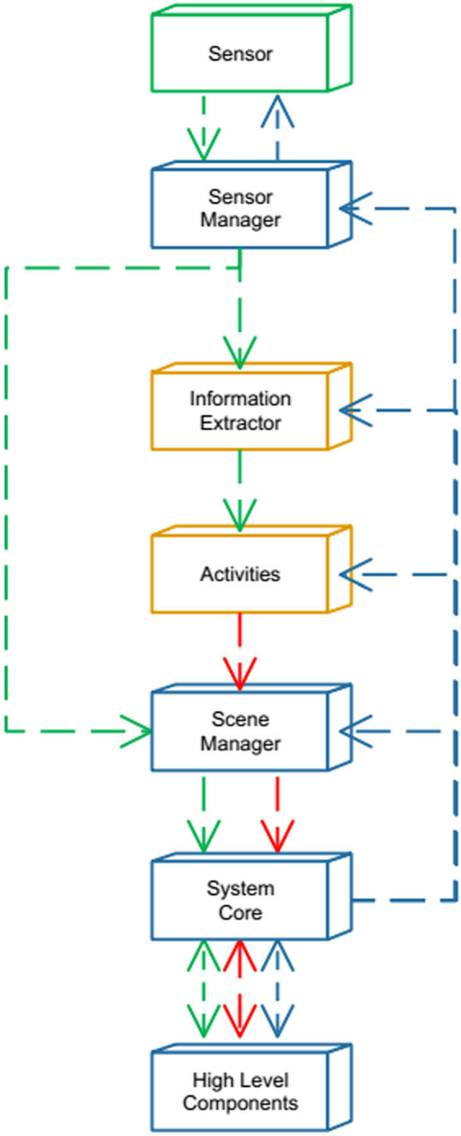
2 System architecture

The proposed intelligent video surveillance architecture has been designed to simultaneously manage a variety of scenarios by using several scenes with several wireless interconnected sensors in each scene. That is, a WMSN is designed. The data extracted from each multimedia sensor in the network are processed to detect abnormal activities. In such a case, an alarm is generated and all the relevant information regarding this alarm is stored for later analysis. By fusing the alarms information, more complex alarms could be generated later in the system.

Figure 1 presents the proposed system architecture based on six modules: Sensor Manager, Information Extractor, Activities, Scene Managers, System Core and High Level Components. Each module includes processes, and the different modules cooperate to achieve the system aims. On the one hand, the modules in orange correspond to processing data modules. On the other hand, the modules in blue correspond to managing data modules. To transmit the information, there are a data flow (green lines), an alarm signals flow (red lines), and a control signals flow (blue lines).

In a nutshell, the performance of the system is as follows. First, the Sensor Manager configures the sensors with the set-up provide by the Scene Manager. Then, the Sensor Manager reads the data from the sensors and send these data to the Information Extractors in the proper format. The Information Extractor extracts the relevant information from the data given by the Sensor Manager. This information is sent to the Activities modules and to its corresponding Scene Manager. The Activities modules deal with the defined security activities in the proposed scenarios. The paper focuses on the security issues regarding the activities considered in the INVISUM project, i.e., a combined indoor-outdoor video surveillance environment. The main task of the Activities is to detect abnormal activities and to generate alarms when needed. When an alarm is generated, it is sent to the corresponding Scene Manager which processes it and transmits the result to the System Core. The Scene Manager modules supervise data, configurations and alarms of each scene. Notice that there are as many Scene manager modules as possible scenes in the whole system. The Scene Managers configure the Sensor Managers, the Information Extractors and the Activities modules with the configuration given by the System core for each scene. The System Core receives all the data and alarms from the Scene Managers, and manage them before transmitting global alarms and data to the High Level Components modules. In addition, the System Core could receive

Fig. 1 Proposed Architecture for the INVISUM project



external configurations from the High Level Components. The System Core could also take some decisions regarding the system configuration. When a global alarm is generated the System Core generates additional data and alarms which are sent to the High Level Components modules in order to enrich the current information available in these modules.

To build this architecture the Robot Operating System (ROS) was used [23]. ROS is a set of software libraries and tools that provide a structured communications layer above the host operating systems of a heterogenous computers set. Communications provided by ROS are established through a TCP-based network. Thus, all the communications in the proposed achitecture are wireless ones. Since the chosen communication protocol just transmits the useful information, fast inter-modules communication is expected. Notice that all modules in the system run concurrently and asynchronously. Given the distributed nature of the designed architecture each module could run on different computers.

Given that one of the main tasks of the Scene Managers and System Core is to fuse all the relevant information, power and capability requirements for the WMSN should be greater here than in the other modules. A typical distribution could be the following. The different Sensor Managers could be integrated on the sensors, or in small capability computers (such as, for instance, a Raspberry Pi). The Information Extractors, the Activities, and the High Level Components could be distributed in several standard computers. However, the Scenes Managers and the System Core should be in a more poweful and reliable dedicated computer (such as a server).

In Fig. 2, a more detailed representation of the system architecture is presented. This figure includes the components of each module, the interfaces and the conexions. The proposed architecture has been designed for the INVISUM project [8]. One of the main aims of the project is to build a global, flexible and scalable video surveillance sytem to operate in a multiple-sensor environment in several indoor-outdoor scenarios. Modules, processes and signals flows presented in Fig. 2 respond to such complex video surveillance environment. In the next sections, these modules are detailed.

2.1 Sensor manager

The Sensor Manager modules read the data from the sensors through a Reader process, and configure the sensors using a Configuration process. Notice that, there are a Sensor Manager for each kind of sensor in the system. The idea behind these modules is to achieve a plug-and-play sensor connection. One of the tasks of the Sensor Managers is to homogenize data formats. For instance, if there are several kinds of video-cameras in the system, the Sensor Manager module of each camera translates each video format into a unique format established by the Sensor Manager configuration. Each Sensor Manager provides data to the corresponding Information Extractor. For example, when the sensor is a camera, the Information Extractor processes the frames in order to extract the relevant information. Notice that, when the Sensor Manager is accompanied by algorithms capable of automatically extracting and processing useful sensor information, then no Information Extractor action is needed. In such a case, the useful information is directly sent from the Sensor Manager to the Activities module. For instance, when we are dealing with Fingerprint sensors, these usually provide a vector of features that describes the individual pattern of interest (see [12] for a complete description of Fingerprint recognition algorithms). These patterns will be send to the corresponding activity (Access Control). Another example are the Thermal sensors, that provide the temperature of several points in the corresponding scene (see [7] for a complete description of Thermal Sensors). As before, these features will be send to the corresponding activity (Fire Detector).

Furthermore, the Configuration process translates the system configuration given by the Scence manager into the proper sensor configuration format.

2.2 Information extractor

Each Information Extractor module receive data, in the proper format, from the corresponding Sensor Manager. The configurations for the Information Extractors are set by the Scenes Managers through the control signals flow. For the purposes of this project, the following processes have considered: Objects Information, Vehicle Features, Motion, and Faces for 2D

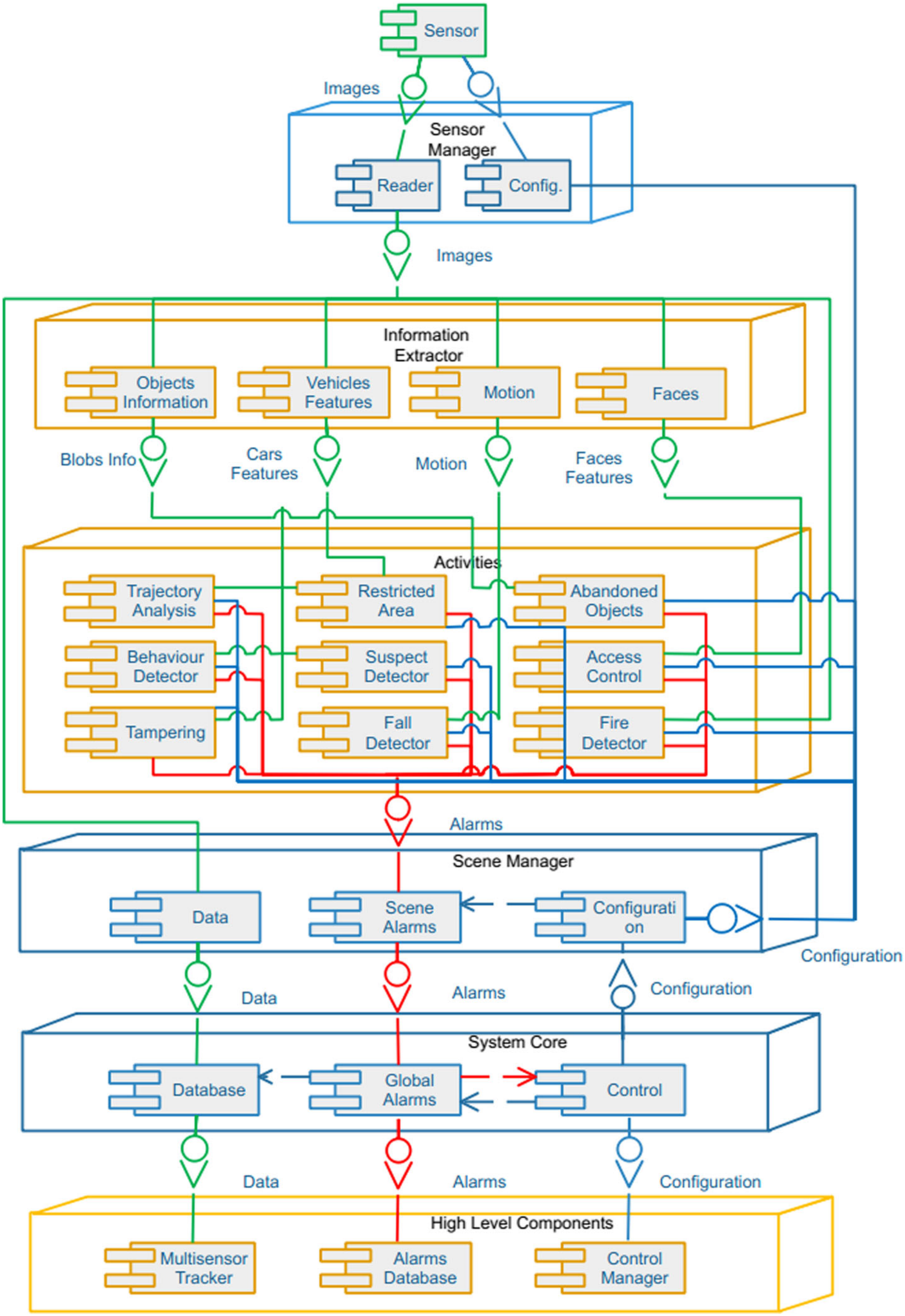


Fig. 2 Detailed Proposed Architecture for the INVISUM project

and 3D camera sensors. Several state-of-the-art methods to process information have been used. Given a sensor, and the corresponding Sensor Manager, one of these processes is included in the corresponding Information Extractor module. For instance, if the sensor is a 3D camera for a control access, the process included in the corresponding Information Extractor is the Faces process. However, if the sensor is a 2D RGB camera in a car access control, the process included in the Information Extractor is the Vehicle Features process. Then, these processes work on the data and the resulting information is sent to the Activities and corresponding Scene Manager. Next, each process is explained in detail.

2.2.1 Objects information

This process is aimed at the extraction of information associated with 2D camera sensors. Given a sequence of images, the Objects Information process extracts objects, their features and their trajectories. We use the background substraction method based on Gaussian Mixture of Models (GMM) for object detection. After background substraction, a bag-of-soft-biometric features is extracted from each detected object. These features are related to the RGB color space, grayscale statistics and histograms, geometry, HSV color space, co-occurrence matrix and Local Binary Patterns [17]. Other features, such as size of the object, are considered. The Kalman filter algorithm is used for trajectory prediction. The input parameters of the Objects Information process are the minimum object size to be considered, and the background update rate, that is, the speed or frequency at which the background gets updated. These parameters are fixed by the corresponding Scene Manager Configuration process. For instance, given two different scenes, this process allows that the minimum object size parameter could be different in each scene.

2.2.2 Vehicle features

The Vehicle Features process uses information extracted from 2D camera sensors. Typically, this sensor would be placed in a parking area for access control and parking applications (see [5] for a complete state-of-the-art review). The proposed process recognizes a vehicles license plate number from an image or images using domain information. In order to locate the region of interest the shape and the size of a car plate are considered. For the character recognition, Neural Networks were trained. In addition, a vehicle brand recognition algorithm is used [10]. This is a complex task since each car has a unique logo, but it could vary in size, color, texture, etc.

2.2.3 Motion

This is a process for the extraction of information from 2D and 3D camera sensors. The Motion process extracts the well-known Motion History Image (MHI) and the Motion Energy Image (MEI) (see [1] for a complete description). The MHI gives the temporal information of the motion at the image plane, and the MEI indicates where the motion has occurred in the image plane [3]. Vehicles and people extracted information results quite useful for the system funcionallity and to improve its perfomance.

2.2.4 Faces

The Faces process uses information from 2D and 3D camera sensors. For each image representation (2D, 3D), Parallel Gabor Principal Component Analysis is used for feature extraction (see [15, 29] for a complete description). These filters have the ability to extract the most significant features in a visual scene. The main challenge in the implementation of this module is the proper combination of 2D and 3D information. In [15] linear combinations of several sources of information are used. However, more complex methods could be considered [13].

2.3 Activities

Activities modules process information and generate alarms when abnormal situations are detected. The configuration for the Activities is set by the Scenes Managers. The generated alarms are sent, through the alarm signals flow, to the corresponding Scene Manager to be processed. For the purposes of the INVISUM project, we consider several activity processes: Trajectory Analysis, Restricted Area, Behaviour Detector, Fire Detector, Suspect Detector, Abandoned Objects, Tampering, Fall Detector and Access Control. These processes require some configuration parameters to define when an alarm must be set. These parameters are fixed in the Scene Manager Configuration process. Given the design of the proposed system, all the activities have the same interface. In addition, all the activities interact in the same way with the Information Extractors and Scene Managers.

2.3.1 Trajectory analysis

The Objects Information process of the Information Extractor module obtains the trajectory of an object as described in Section 2.2.1. In any anomaly trajectory method, one of the main challenges is to define what an anomaly (outlier) is. In the proposed system, the configuration from the Scene Manager defines what is an anomalous trajectory. This information could come as a result of a previous training stage, or as a direct definition from an user [19, 32]. Given the scene, the security user could define a trajectory as anomalous (for instance, to wander around the area of interest). Given a new trajectory, this process returns the maximum similarity to a predefined normal trajectory [30].

2.3.2 Restricted area

This process has been designed to detect objects in a restricted area of a scene. Such an area is defined by the Scene Manager configuration. Object features and trajectory are obtained in the Information Extractor module as described in Section 2.2.1. Given these features (specially the object size) and the trajectory (specially motionlessness), it is possible to detect illegal parking, or unwanted human presence in the predefined area.

2.3.3 Behaviour detector

Crowd behaviour analysis is a recent area of interest in computer vision (see [36] for a complete survey). However, the real crowd motion exhibits complex behaviors that are difficult to model. We build a Behaviour Detector from objects trajectories obtained in the Information Extractor module. The Behaviour Detector provides a crowd density measure, and generates an alarm when this measure is higher than a fixed threshold. Furthermore, an alarm is generated when a higher correlation among a large number of trajectories in the crowd appears. As before, these parameters are defined in the Configuration process of the Scene Manager.

2.3.4 Fire detector

In this process, the input information is the temperature from Thermal Information cameras. In this case, the information arrives directly from the Sensor Manager. Thus, no Information Extractor process in needed. This thermal imaging system uses infrared imaging technology that detects infrared radiation or heat. In each frame of the video, the pixels with temperature higher than a threshold are detected. Those pixels define a blob. When the size of the blob is high enough, and this blob is detected for a sufficient number of seconds, an alarm is generated and transmitted to the Scene Manager. Thus, the minimum size for the blob, and the minimum number of seconds are the two parameters to be fixed.

2.3.5 Suspect detector

The Suspect Detector process receives the blobs and features from the Objects Information process of the Information Extractor module. In addition, it receives the configuration signal from the Scene Manager that informs about the reference features values that the Suspect Detector process should look for. The normalized (from 0 to 1) Euclidean distance between those reference values and the features values in each new blob, are calculated (as indicated in [17, 18]). If the distance is lower than a threshold, an alarm is generated and transmitted to the Scene Manager.

2.3.6 Abandoned objects

The most popular abandoned object detection algorithms are base on background subtraction, due to their superior robustness in complex real world scenario as those proposed in this project. Following [31], we consider an abandoned object to be a nomotion object that has not been in the images before. The parameters for this process are the size of the object and the time lapse to label an object as abandoned. The main challenge in this module is the proper context-aware configuration of these parameters. To perform this task, human domain expert is mandatory.

2.3.7 Tampering

In this paper, camera tampering is defined as any relevant event which dramatically alters the image seen by the camera. For instance, the camera is moved, partially obscured, severely defocused, covered or sprayed, etc. This situations imply big objects in the image. Thus, numerous methods for tampering detection use image difference calculation (see, for instance, [25]). In the proposed system tampering is detecting by the size of the detected objects. If the size is larger than a threshold fixed by the Scene Manager configuration, an alarm is generated.

2.3.8 Fall detector

Recently some works exploiting computer vision for detecting falls have been presented [3, 4, 26, 38]. The Motion process of the Information Extractor module obtains the motion information of an object as described in Section 2.2.3. In this process we detect Humans fall based on the MHI and MEI.

2.3.9 Access control

Biometric characteristics are widely used in access controls. For example, automated border controls (ABC) systems have been installed in different worldwide airport entries in the last years (see, for instance, [21, 28]). In this paper, the Access Control process uses information from the Face process in the Information Extractor module (Section 2.2.4), and information from fingerprint sensors. To perform fingerprint biometric acquisition, a traditional touch-based fingerprint recognition system is used [12]. In this case, the information arrives to the Access Control process directly from the Fingerprint sensors (no Information Extractor is needed). The methods proposed in [13] have been used to combine these two sources of information.

The process compares the biometric data with the models fixed by the Scene Manager configuration. Typically, this means a search of a person in a list. The process output is the identification or not of the person. In the last case, an alarm is generated and transmitted to the Scene Manager.

2.3.10 Additional activities

The considered activities correspond to that included in the scope of the INVISUM project. That is, an indoor-outdoor multisensor video surveillance problem. However, given that all the activities share the same interface, adding new activities requires little effort, if necessary. That is, our system is flexible and easily customizable.

2.4 Scene manager

As shown in Fig. 1, there are three processes in the Scene Manager modules: Data, Scene Alarms and Configuration. The wireless and distributed system has been designed to manage several scenes with several sensors in each scene. The Scene Manager configures each Sensor Manager, each Information Extractor and each Activity related to the scene. The Data process receives the data from the Information Extractors in order to synchronize them. After that, the synchronized data are sent to the System Core. Besides, the Scene Manager receives all the alarms generated in the Activities modules. In the Scene Alarms process, all these alarms are synchronized and combined. On the one hand, this combination could be very simple, for instance using a rule such as “to generate an scene alarm when a single activity alarm is generated”. On the other hand, this combination could be based on a more complex algorithm. For instance, a method to build an alarm buffer for the measurement of risk according to the system activities could be developed (see, [16] for a similar approach in the traffic safety domain).

The alarms from all the Scene Managers are sent to the System Core for further processing.

2.5 System core

The System Core is the responsible to monitor the system for proper performance. There are three processes in the System Core: DataBase, Global Alarm and Control (see Fig. 1). The System Core receives all the synchronized data, and alarms from the Scene Managers. In the Global Alarms process, all alarms received are synchronized and combined following the configuration parameters given by the Control process. This process generates an unique global alarm for the system.

In the DataBase process all received data are synchronized using the timestamp given by the Sensors Managers. When a global alarm is generated, these data are sent to the High Level Components following the configuration from the Control process. This DataBase process works as a buffer, that performs with a temporal memory allocation scheme.

The Control process generates the signals for the Scene Manager Configuration processes. These signals can be fixed by the Control Manager process in the High Level Components as a set of preconfigured rules, or they can be learned to respond to specific alarm situations. Notice that the security user manages the system through this Control process. For instance, when the system generates an alarm from the information of sensor A, the set of rules states the other sensors B,C,.., etc., to collect as much information as possible about that alarm.

3 High level components

In this Section, three high level components that use the previous architecture are detailed: Alarms DataBase, Multisensor Tracker and Control Manager. These processes receive/send data, alarms and configuration signals from/to the System Core.

3.1 Alarms database

This process manages the DataBase of the system's alarms. When an alarm is generated in the System Core, all the data related to that alarm are stored in the DataBase.

3.2 Multisensor tracker

Tracking multiple people from standard cameras is challenging, specially when overlapping between several cameras is not covered (see, for instance, [2, 9]). In the proposed problem, when an alarm is generated in the System Core, the Multisensor Tracker manages all the information regarding the object or trajectory that caused the alarm. For instance, when a person is detected in a restricted area by the corresponding Activity, an alarm is generated in the System Core. Then, the System Core activates the Multisensor Tracker in order to collect all the data regarding that person from the different sensors. Notice that, to do that, this process configures the Suspect Detectors of the corresponding Activities.

3.3 Control manager

As presented in previous section, the security user configures the whole system through the Control process in the System Core module. The user sets the parameters of each Scene Manager Configuration process and the parameters to control the Global Alarms process in the System Core. Notice that each Scene Manager Configuration process controls the parameters of the Activities, Information Extractors and Sensors Managers involved in each scene. To facilitate the installation and configuration of the proposed system, three basic configurations are presented to the security user: Level 3 or low security, Level 2 or medium security, and Level 1 or high security. Table 1 shows a summary of these security configurations and the parameters established for each process.

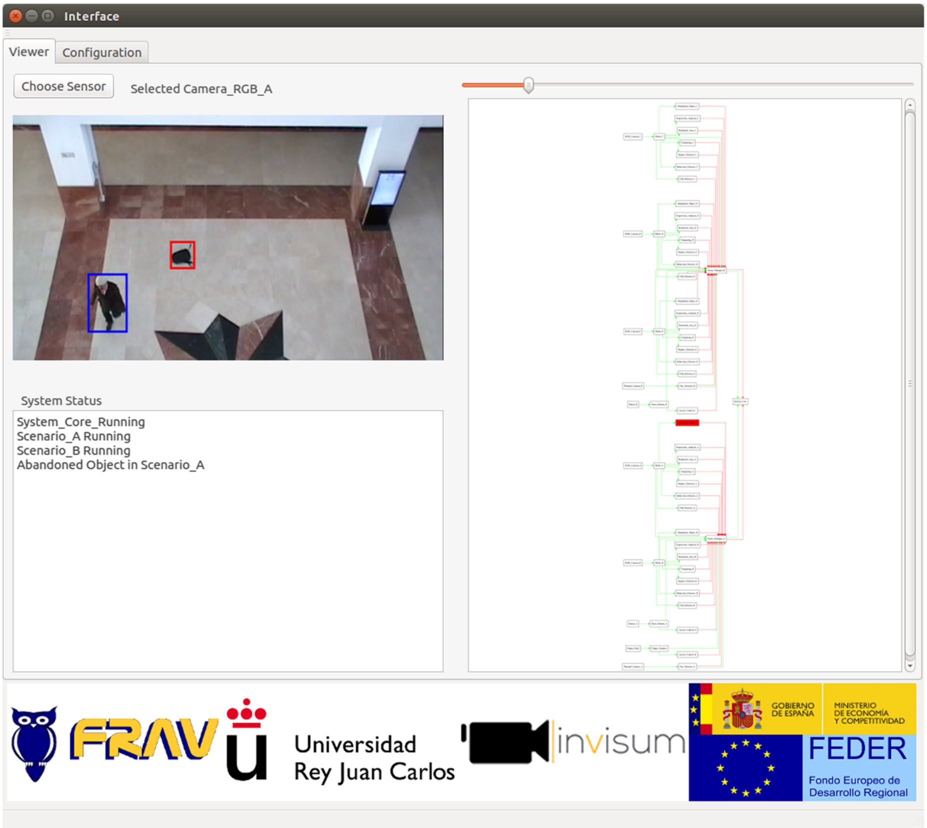
To manage the Control process in the System Core, a Control Manager Tool (CMT) was developed. The main aim of this tool is to facilitate the configuration of the system parameters to the security user. In addition, global information is presented as a result. The proposed CMT allows two visualizations (see Figs. 3 and 4). In the first visualization, the called Viewer, the CMT shows information about the global system, their architecture, sensors and connections. The user can choose a sensor to obtain the visual signal. In addition, the system status with all the generated alarms is presented. For instance, in the image in Fig. 3 an abnormal trajectory and an abandoned object have been detected in scenario A.

In the second visualization, the called Configurator, the CMT allows the user to configure the system. First, a control to run/stop the system is presented. To facilitate the configuration of the system, the user can easily chose the level of security associated with a color scale: green for low security, orange for medium security, and red for high security. The user can add/delete Scene Managers, Sensors Managers and Activities to the system. Furthermore, it is possible to change the level of security for Scene Managers and Activities. Notice that it is possible to select different levels of security for each scene and activity. For instance, the user could chose a low security level for the Trajectory Analysis activity, but a high level of security for the Abandoned Objects activity. In such a way, the system would respond almost immediately to an abandoned object situation, but slowly to an abnormal trajectory. Moreover, the security user could chose various security levels fot different scenes.

To test the system, a set of experiments were designed to be executed on two real scenarios where our system was deployed. These experiments are presented in the next Section.

Table 1 Different security configurations for the security system

Information Extractors		Security Level		
Parameter	Description	Low	Medium	High
Indoorobject minimum size	minimum size in pixels of the object to be considered in an indoor scenario	20 × 20	10 × 10	5 × 5
Outdoor object minimum size	minimum size in pixels of the object to be considered in an outdoor scenario	30 × 30	20 × 20	10 × 10
Temperature	minimum temperature in degrees to be considered	110	90	70
Activities		Security Level		
Parameter	Description	Low	Medium	High
Abnormal Trajectory	maximum similarity to a normal trajectory	< 0.5	< 0.7	< 0.9
Restricted Area time	time in seconds to be considered as illegal	120	60	20
Behaviour crowd density	number of people	60	50	20
Behaviour crowd trajecto-ries	number of similar trajectories at the same time in the same place	30	20	10
Fire size	minimum size in pixels for the blob	20 × 20	10 × 10	5 × 5
Fire time	time in seconds to be considered a fire	20	10	5
Suspecto detector threshold	Normalized Euclidean distance between features to generate an alarm	< 0.1	< 0.2	< 0.3
Abandoned object time	time in seconds to be considered as abandoned	20	10	5
Tampering size	size in % of the image size, of the detected object	90	80	70
Fall detection time	time in seconds to be considered as fallen	7	5	3
Access Control Security	number of invalid access attempts before alarm	5	2	1



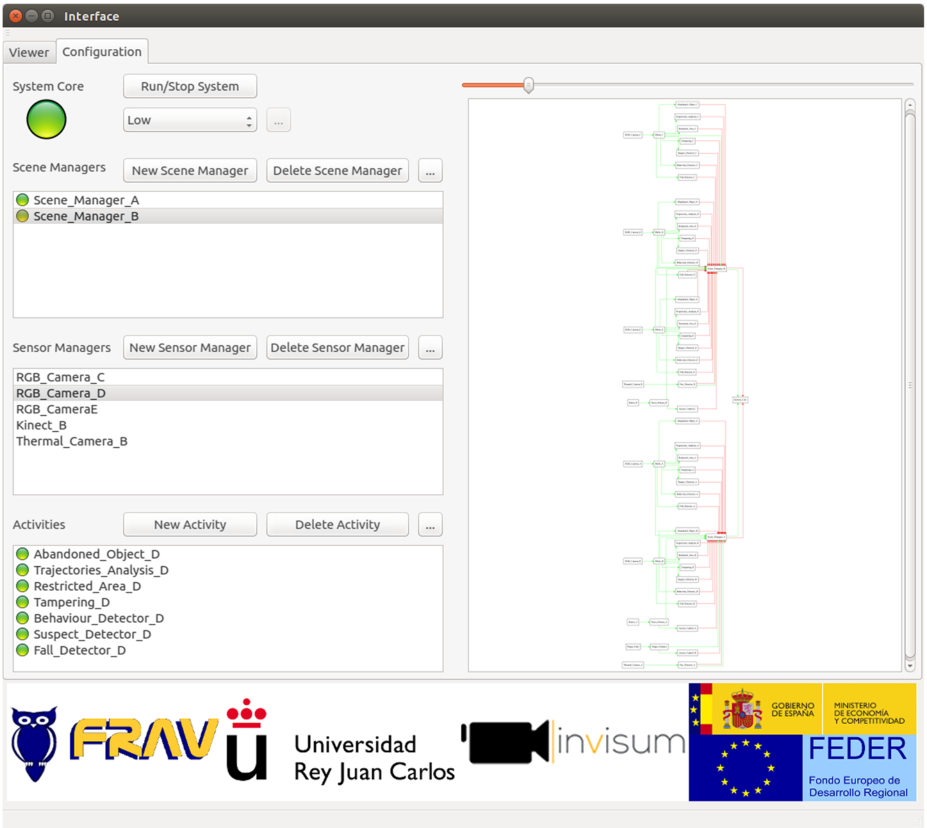
```
System_Core_Running
Scenario_A Running
Scenario_B Running
Abandoned Object in Scenario_A
```

Fig. 3 Viewer screen in the Control Manager Tool (CMT)

4 System tests (a case of study)

To validate the performance of our system, we have performed experiments using two academic scenes at the campus of the University Rey Juan Carlos in Mostoles, Spain.

The first scenario is an indoor scene shown in Fig. 5. The image covers most of the main hall of a classroom building. The hall is mainly used to move from one classroom to other and to get into and out of the building. In this case, five inexpensive sensors were used: two RGB cameras, a 2D-3D camera, a thermal camera and a fingerprint sensor. The RGB cameras have a CMOS sensor with 1920 × 1080 pixels of resolution and a frame rate of 30 fps. These cameras have WI-FI connection. In these examples the standard protocol 802.11/b/g was used. They were mounted overlooking the ground floor, on the first and second floors, respectively. The 2D-3D camera is a Kinect V1.0 which has a CCD sensor with 640 × 480 pixels of resolution for RGB images and a CCD sensor with 320 × 240 resolution for infra-red and 3D images. Besides, the Kinect sensor has a infra-red proyector that emits a patron used for the 3D images calculation. The Kinect is connected to a Raspberry Pi B+ through USB 2.0. It has a ARM1176JZF-S CPU at 700 MHz and 512 Mb of RAM. The Sensor Manager process runs on the Raspberry Pi that sends the collected data through WI-FI to the corresponding Information Extractor. The range of temperatures for the thermal camera is from −20 to 120 degrees. It gives images of 640 × 480 pixels in a range of 30 m. The thermal camera is



Scene_Manager_A
Scene_Manager_B

RGB_Camera_C
RGB_Camera_D
RGB_CameraE
Kinect_B
Thermal_Camera_B

Abandoned_Object_D
Trajectories_Analysis_D
Restricted_Area_D
Tampering_D
Behaviour_Detector_D
Suspect_Detector_D
Fall_Detector_D

Fig. 4 Configuration screen in the Control Manager Tool (CMT)

connected to a smartphone through micro USB. The smartphone is able to store, process in real-time and send, through WI-FI, the information to the corresponding Activity process. The fingerprint sensor is used as an employee time clock attendance machine. It is supposed to be used by the teachers every hour from ten minutes before to ten minutes after o'clock. However, it is supposed not to be used out of that period of time. That is, there is a time window of twenty minutes per hour to use the sensor. This context information is fixed by the configuration control.

The second area of insterest is an outdoor scene shown in Fig. 6. The image covers a parking area close to the classroom building considered in the previous scene. In this case, five inexpensive sensors were used: three RGB cameras, a 2D-3D camera, and a thermal camera. To monitor the entire scene two RGB cameras, similar to those in the indoor are, are used. To extract vehicles information, one RGB camera is used. It has a CMOS sensor with 2048 × 1536 pixels of resolution and a frame rate of 12.5 fps. The thermal camera has the same characteristics than the camera used in the indoor scenario. The 2D-3D camera is a Kinect V1.0. In this case, this camera is set in the entry of the nearest building to the parking, to people access control.

For our purposes, we have used a distributed achitecture of computers. In this case, two Raspberry Pi computers, two smartphones and four Intel Core i5 (4GB of RAM) computers,

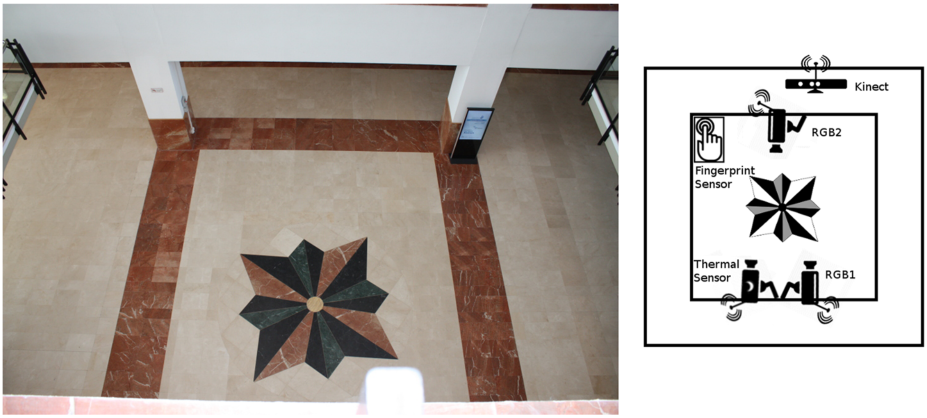


Fig. 5 Indoor Scenario and sensors positions

are considered. The connections between them have been made through WI-FI 802.11/b/g. One computer (server) is used for the System Core and the two Scene Managers. A second computer is used for the High Level Components module. A third computer is used for the Sensor Managers, Activities and Information Extractors of the indoor scene, and a fourth computer is used for the Sensor Managers, Activities and Information Extractors of the outdoor scene.

4.1 Configuration for the tests

Given the sensors and the communications between the system modules previously presented, it is possible to define the set of rules to manage the system. In the indoor scenario, when one of the RBG cameras or the fingerprint sensor generate an alarm, the Kinect sensor switches on to collect as much information as possible about the person that generates that alarm. Similarly, in the outdoor scenario, when one of the RBG cameras generates an alarm, the Kinect sensor (in the entry of the building) switches on in order to collect as much information as possible about the person that generates that alarm. When an alarm is generated in one scenario, the Multisensor Tracker process looks for the person that generates the alarm in the RGB cameras

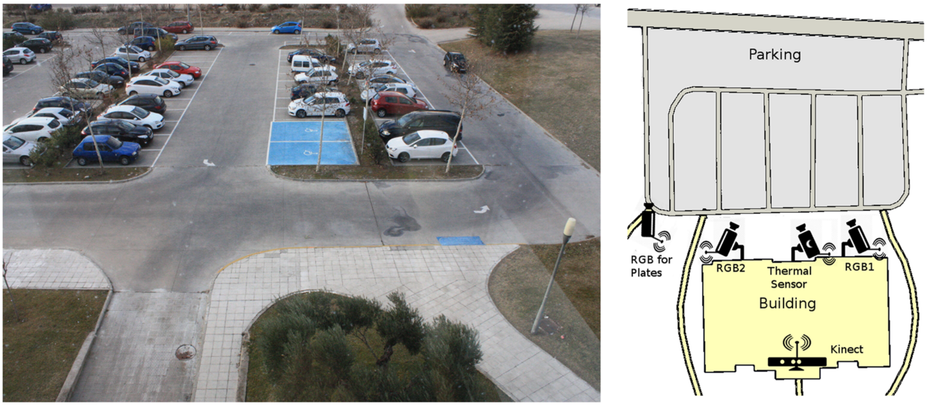


Fig. 6 Outdoor Scenario and sensors positions

of the two scenarios. In addition, the information collected by the plates reader RGB is retrieved in order to merge a car with the suspect. All the acquired data are stored in a 2GB memory.

4.2 Experiments

To test the perfomance of the proposed system, several security incidents were preprogrammed in each scenario to induced alarm situations.

The first security incident is performed to test context information. In the indoor scenario, the fingerprint sensor area is not physically protected. It is mandatory to generate an alarm when someone (an actor) acts over the fingerprint sensor or stands close to the sensor for a long period of time, out of the allowed time interval. If an alarm is generated, then a signal is sending to the Kinect sensor to switch on. The face information of the actor is collected and their identity is verified by the Access Control Activity. That is, an actor is labelled as suspect given the information of one sensor, and new information about the actor is collected by other sensors.

The second security incident is performed in the two scenes. First, an actor leaves an object (a backpack) on the floor in the indoor scenario. This should be detected as an abandoned object. Then, an alarm should be generated and the features of the actor collected from the RGB cameras. Next, the actor moves to the outdoor scenario. The system tries to locate the actor in the outdoor scenario using the information from the two RGB cameras.

5 Data analysis

A data analysis of the collected information was performed using the pre-configured security configurations presented in Table 1. The description of each alarm situation detected by our system are summarized in Table 2. Notice that, in the High security level, the system is able to detect six situations as alarms. In the case of performing under a Low security level, only one situation was reported as alarm. Finally, when the system performs under a Medium security level, three situations were considered as alarms.

As expected, the system is able to detect all the preprogrammed security incidents. First, the Restricted Area activity sent an alarm to the Scene Manager in the scenario indoor, when a person stayed in the predefined area for a time higher than the preconfigured time threshold

Table 2 Different security situations detected for the security system. Given the security level configurations, the secutiry levels in which the situation would be detected as an alarm is marked (X)

N.	Description	Detected by	Security Level		
			Low	Medium	High
1	A person in a restricted area during 90s	Restricted Area	–	X	X
2	A hand in front of the camera	Tampering	X	X	X
3	A backpack in the floor during 7 s	Abandoned objects	–	–	X
4	A car in a restricted area during 68 s	Restricted Area	–	X	X
5	A group of 13 people in the scenario walking on similar trajectories	Behaviour crowd trajectories	–	–	X
6	Suspect labeled in scenario A, detected in scenario B	Suspect Detector	–	–	X

(see Table 1). Notice that the alarm will be generated (or not) depending on the level of security that the system is working with. On the one hand, if the secutiry level is low, the person detected in the restricted are during 90s will not generate an alarm. On the other hand, if the secutiry level is medium or high, the same situation will generate an alarm. When an alarm was generated, a signal to switch on the Kinect sensor was sent. In such a case, the Face information was recorded in order to verify (if possible) their identity. Thus, several sensors are used to detect an event. In this case, the event is the presence of a suspect behaviour of a previous signed up suspect. A detailed schedule listing the events that take place during this situation is presented in Fig. 7. The RGB camera sensor sends images to the Sensor Manager located in a Computer A. The Sensor Manager sends the images in the proper format to the Objects Information process.

This process obtains the blobs and sends them to the Restricted Area process. The Restriced Area process generates an alarm and sends that signal to the Scene Manager A, located in a server. The System Core (located in the same server) recieves such alarm and send a control signal, in order to switch on a new sensor, to the Scene Manager A. This signal arrives to the Sensor Manager B (located in a Raspberry Pi), that manages the Kinect sensor. Figures 8a and b presents images of this security incident.

In addition to the pregrogrammed secutiry incidents, other ones were detected. For instance, during the record of the experiment, a hand appears in front of the camera (see Fig. 8c). This was detected by the Tampering process and an alarm was generated. Since the hand was very close to the camera, the alarm will be generated in any of the predefined security levels.

The abandoned objects module detected a backpack in the floor, and the person was labeled as suspect (see Figs. 8d and e, for the RGB and Thermal Images, respectively). By using the features extracted in the Objects Information process, the suspect was detected in the outdoor scenario (using the Suspect Detector activity) and the person and vehicle information was recorded in the alarm database (see Fig. 8g).

The Behaviour Detector process generated an alarm related to the number of similar trajectories of people in the scene. In this particular case, a group of 13 people were walking on similar trajectories at the same time in the image (see Fig. 8f). Given the parameters in the Security Configurations (see Table 1), this was detected as an alarm when the system performed in a High security level. The detected situation corresponds to a movement of students from their classroom to the exit of the classrooms building, when the class has finished. That is, it could be considered as a false alarm, since it is not a risk situation. Thus, in the indoor scenario, it was decided to increase this particular Security Configuration parameter from 10 to 20. This case indicates that a new rule adding context knowledge in the Behaviour Detector activity should be considered.

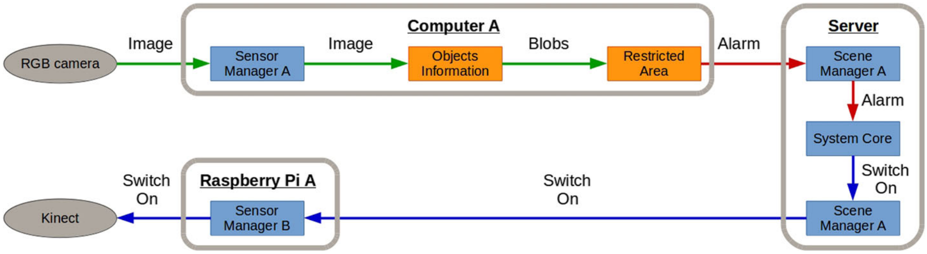
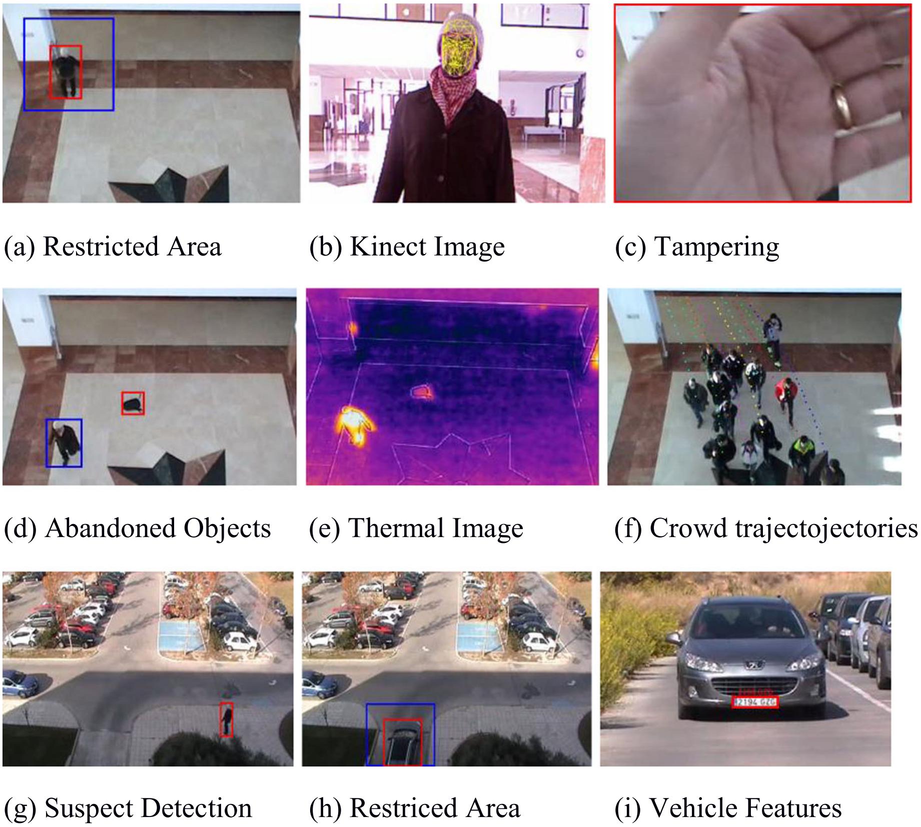


Fig. 7 Flow diagram of the Restriced Area experiment



(a) Restricted Area (b) Kinect Image (c) Tampering

(d) Abandoned Objects (e) Thermal Image (f) Crowd trajectojectories

(g) Suspect Detection (h) Restriced Area (i) Vehicle Features

Fig. 8 Secutiry situations detected as alarms by the system

Furthermore, in the outdoor scenario, one no-pregrogrammed security incident was detected (see Fig. 8h). A vehicle stayed in a restricted area during 68 s. An alarm will be generated when the system was performed under medium or high security levels. Figure 8i shows the information extracted by the Vehicle Features process (see Section 2.2.2) regarding the vehicle that caused the alarm.

Thus, the results of the tests show that the proposed system is reliable for security incidents identification. In addition, some of the alarms will be useful to redefine some of the default parameters in the system.

Further qualitative evaluation results, objective evaluation of the system has been performed. Notice that this goes beyond the purpose of the present paper. However, to validate the performance of the system, several studies have been developed. Using the architecture first presented in this paper, a context-aware distance for anomalous human trajectories detection has been recently presented [27]. The presented method outperforms the alternative distances: 75.8% of outliers detections in a database of 182 real trajectories. Another paper that presents a novel methodology for human reidentification in multi-camera VideoSurveillance environment scenarios using the proposed distributed architecture has been accepted for its publication [14]. A number of discriminative features extracted in a scene are used in a new scene in order to detect suspicious persons using the images from a non-overlapping camera. In this research, the 100% of the suspicious persons were detected, and a 5.88% of false alarms were generated.

6 Conclusions

In this paper a wireless and distributed architecture intelligent surveillance system has been presented and tested. The proposed architecture has been designed to manage several scenenarios with several wireless interconnected sensors, as part of a video surveillance system. Several state-of-the-art methods to process information and to detect abnormal activites have been used. The proposed architecture has been built using the Robot Operating System that manages wireless communications in a multimedia sensor network. The wireless multimedia sensor architecture presented in this paper has been tested in real situations. It has been shown, by employing multiple and heterogenous sources of information, that the system is capable to detect risk situations and generate appropriate alarm signals. The flexibility and adaptability of the poposed architecture has been tested. In a real experiment it has been shown that the detected alarms updated configuration parameters and rules of the system to improve its reliability. Overall, the results have been promising and the proposed architecture can serve as the foundation for further enhancements.

An objective evaluation of the distributed architecture intelligent surveillance system has been performed via experts' analysis. Two experienced experts working (>25 years) for the company responsible of the security at the University [6] were questioned regarding the usabilty, features, and capabilities of the system. They provided an average score of 9.5 over 10 about the global performance and the security support provided for the system.

Although this goes beyond the purposes of the present paper, preliminary statistical analysis and evaluation based on Data Analysis section have been promising and show that the proposed architecture can serve as the foundation for further research.

Although in this paper only academic scenarios were considered, the proposed architecture can be fully extended to smart world. As smart cities, our system relies on multimedia sensory data acquired from multiple sensors in distributed locations.

In the future, the proposed multimedia sensor network can be enriched with new sensors, such as road sensors and portable air sensors, with new multimedia devices capabilities.

To extend the system, a complete alarm module that combines information from several (and probably asynchronous) alarms is under development. This module will be part of the INVISUM project. It will consider relevant variables to increase the alarm level or decrease the alarm level depending on the security incidents. In the past, the same idea has been successfully used to build a risk buffer in the area of driving risk detection [16]. The fusion of several security experts' knowledge is considered to improve the future alarm module.

Acknowledgements This work is supported by the Ministerio de Economíıa y Competitividad from Spain INVISUM (RTC-2014-2346-8). The authors would like to thank Cesar Benavente Peces, professor at 'Universidad Politecnica de Madrid, for his time and effort in the form of review and feedback'.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Ahad Md AR (2013) Motion History Images for Action Recognition and Understanding. Springer

2. Alaya Cheikh F, Saha SK, Rudakova V, Wang P (2012) Multipeople tracking across multiple cameras. International Journal on New Computer Architectures and Their Applications (IJNCAA) 2(1):23–33 The Society of Digital Information and Wireless Communications
3. Babu RV, Ramakrishnan K (2004) Recognition of human actions using motion history information extracted from the compressed video. Image Vis Comput 22(8):597–607
4. Chua JL, Chang YC, Lim WK (2013) A simple vision-based fall detection technique for indoor video surveillance. SIViP 9(3):623–633
5. Du S, Ibrahim M, Shehata M, Badawy W (2013) Automatic license plate recognition (alpr): A stateof-the-art review. Circuits and Systems for Video Technology, IEEE Transactions 23(2):311–325
6. GOMSEGUR (2014). www.gomsegur.com. Accessed Dec 2017
7. Huynh T (2015) Fundamentals of thermal sensors. In: Jha CM (ed) Thermal Sensors. Springer, New York
8. INVISUM (2014). http://www.invisum.es. Accessed Dec 2017
9. Ju J, Ku B, Kim D, Song T, Han DK, Ko H (2015) Online multi-person tracking for intelligent video surveillance systems. In: Consumer Electronics (ICCE), 2015 IEEE International Conference on, pp. 345–346. IEEE
10. Kadhm MS, Yun LS (2015) Propose a simple and practical vehicle logo detection and extraction framework. International Journal of Emerging Tendrs and Technology in Computer Sciences 4(3)
11. Kalantar-Zadeh K (2013) Sensors: an introductory course. Springer Science & Business Media
12. Maltoni D, Maio D, Jain AK, Prabhakar S (2009) Handbook of Fingerprint Recognition, 2nd edn. Springer Publishing Company, Incorporated
13. Martín de Diego I, Muñoz A, Moguerza JM (2010) Methods for the combination of kernel matrices within a support vector framework. Mach Learn 78(1–2):137–174
14. Martín de Diego I, San Román I, Conde C, Cabello E (2017) WYA 2: Optimal Individual Features Extraction for VideoSurveillance Re-identification. SIGMAP: 14th International Conference on Signal Processing and Multimedia Applications
15. Martín de Diego I, Serrano A, Conde C, Cabello E (2010) Face verification with a kernel fusion method. Pattern Recogn Lett 31(9):837–844
16. Martín de Diego I, Siordia OS, Crespo R, Conde C, Cabello E (2013) Analysis of hands activity for automatic driving risk detection. Transportation Research Part C: Emerging Technologies 26:380–395
17. Moctezuma D, Conde C, De Diego IM, Cabello E (2015) Soft-biometrics evaluation for people reidentification in uncontrolled multi-camera environments. EURASIP Journal on Image and Video Processing 2015(1):1–20
18. Moctezuma D, Conde C, Martin de Diego I, Cabello E (2013) Incremental learning with soft-biometric features for people re-identification in multi-camera environments. In: Digital Image Computing: Techniques and Applications (DICTA), 2013 International Conference on, pp. 1–7. IEEE
19. Morris BT, Trivedi MM (2008) Learning and classification of trajectories in dynamic scenes: A general framework for live video analysis. In: Advanced Video and Signal Based Surveillance, 2008. AVSS'08. IEEE Fifth International Conference on, pp. 154–161. IEEE
20. Nam Y, Rho S, Park JH (2012) Intelligent video surveillance system: 3-tier context-aware surveillance system with metadata. Multimedia Tools Appl 57(2):315–334. https://doi.org/10.1007/s11042-010-0677-x. URL10.1007/s11042-010-0677-x
21. Nuppeney M (2014) Automated border control–state of play and latest developments. NIST International Biometric Performance Conference, In, pp 1–3
22. Oh JM, Kim KH, Hong S, Moon N (2014) Execution behavior modeling methodology for large scale surveillance system design and evaluation. Advances in Multimedia 2014:15
23. Quigley M, Conley K, Gerkey B, Faust J, Foote T, Leibs J, Wheeler R, Ng AY (2009) Ros: an open-source robot operating system. ICRA Workshop on Open Source Software 3:5
24. Rho S, Rahayu W, Nguyen UT (2015) Intelligent video surveillance in crowded scenes. Information Fusion 24:1–2
25. Ribnick E, Atev S, Masoud O, Papanikolopoulos N, Voyles R (2006) Real-time detection of camera tampering. In: Video and Signal Based Surveillance, 2006. AVSS'06. IEEE International Conference on, pp. 10–10. IEEE
26. Rougier C, Meunier J, St-Arnaud A, Rousseau J (2007) Fall detection from human shape and motion history using video surveillance. In: Advanced Information Networking and Applications Workshops, 2007, AINAW'07. 21st International Conference on, vol. 2, pp. 875–880. IEEE
27. San Román I, Martín de Diego I, Conde C, Cabello E (2017) Context-aware distance for anomalous human trajectories detection. IbPRIA 2017: 8th Iberian Conference on Pattern Recognition and Image Analysis, Faro

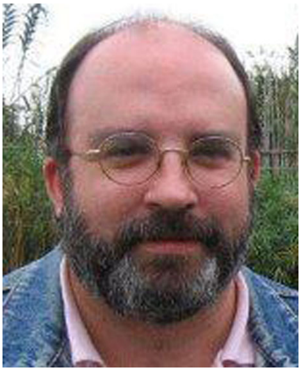
28. Sanchez del Rio J, Conde C, Tsitiridis A, Raul Gomez J, Martin de Diego I, Cabello E (2015) Face-based recognition systems in the abc e-gates. In: Systems Conference (SysCon), 2015 9th Annual IEEE International, pp. 340–346. IEEE
29. Serrano A, de Diego IM, Conde C, Cabello E, Shen L, Bai L (2007) Influence of wavelet frequency and orientation in an SVM-based parallel Gabor PCA face verification system. In: Intelligent Data Engineering and Automated Learning-IDEAL 2007, pp. 219–228. Springer
30. Siordia OS, Martin de Diego I, Conde C, Cabello E (2014) Subjective traffic safety experts' knowledge for driving-risk definition. Intelligent Transportation Systems, IEEE Transactions on 15(4):1823–1834
31. Tian Y, Feris RS, Liu H, Hampapur A, Sun MT (2011) Robust detection of abandoned and removed objects in complex surveillance videos. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions 41(5):565–576
32. Tung F, Zelek JS, Clausi DA (2011) Goal-based trajectory analysis for unusual behaviour detection in intelligent surveillance. Image Vis Comput 29(4):230–240
33. Valera M, Velastin SA (2005) Intelligent distributed surveillance systems: a review. In: Vision, Image and Signal Processing, IEE Proceedings, vol. 152, pp. 192–204. IET
34. Video surveillance system (camera, storage, monitor, software and services) market size: Global analysis and forecast - 2014 to 2020. (2014). http://www.researchandmarkets.com/reports/2876216/
35. Wang X (2013) Intelligent multi-camera video surveillance: A review. Pattern Recogn Lett 34(1):3–19
36. Zhan B, Monekosso D, Remagnino P, Velastin S, Xu LQ (2008) Crowd analysis: a survey. Mach Vis Appl 19(5–6):345–357. https://doi.org/10.1007/s00138-008-0132-4
37. Zhang T, Chowdhery A, Bahl PV, Jamieson K, Banerjee S (2015) The design and implementation of a wireless video surveillance system. In: Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, pp. 426–438. ACM
38. Zhang Z, Conly C, Athitsos V (2015) A survey on vision-based fall detection. In: Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments. ACM

Isaac Martín de Diego received the B.Sc. Degree in statistics and the Ph.D. degree in mathematical engineering from Universidad Carlos III de Madrid, Madrid Spain. In 1994–1999 he was a Researching Assistant with University of Valladolid, Valladolid, Spain, and with Universitat Autónoma de Barcelona, Barcelona, Spain. In 1999–2005 he was a Teaching Assistant with Universidad Carlos III de Madrid. Since 2006 he has been an Associate Professor with the Face Recognition and Artificial Vision Group, Universidad Rey Juan Carlos, Madrid, Spain. His research interests include computer science, computer architecture, pattern recognition, traffic safety, the representation and combination of information, and data science.

Ignacio San Román received his B.Eng in Telecomunications (Image and Sound) from de Public University of Navarra (Spain) in 2012. In 2015, he received the Master's degree in Computer Vision from King Juan Carlos University (Madrid, Spain), and the Master's degree in Automatization and Robotics from Charles III University of Madrid (Spain). In the same year, he joined FRAV research group as a Ph.D. student. He also is one of the main researchers in the INVISUM project. His research interests include Computer Vision, Image Processing, Pattern Recognition, Artificial Intelligence, Robotics, Bio-inspired Computer Systems, Distribured Architectures, Internet of Things, Smart Cities and Intelligent video Surveillance.

Javier Cano Montero received his B.E. in Computer Science in 2008 and his Ph.D in 2015. His main research interests are pattern recognition, parallel and distributed computing, real time computing, computer vision, Internet of things, smart cities and hardware-software design.

Cristina Conde received the M.E. Physics degree from University Complutense, Madrid, Spain, in1999. And the Ph.D. degree in Computer Science from the University Rey Juan Carlos, Madrid, Spain, in2006. She has been main research in projects funded by national and international institutions and companies. She has been Vice dean of Studies from 2008 to 2015 at the Computer Science Faculty. Her primary research interests include Computer Vision, Image Processing, Biometrics, Pattern Recognition, Intelligent Traffic Systems and Bio-inspired computer systems.



Enrique Cabello received the B.S.in Physics(Electronics) from the University ofS alamanca and the Ph.D. degree from the Polytechnical University of Madrid (Spain). Was awarded with the Extraordinary Prize as one of the best Ph.D.thesis. In 1990, he joined the University of Salamanca, where he was an Assistant Professor. From 1998 he is at the University Rey Juan Carlos (Spain). Since 2001 he is the Coordinator of the FRAV research group. He has been main research in projects funded by national and international institutions and companies. His research interest includes image and video analysis for smart videosurveillance, pattern recognition and risk estimation.