VANAHEIM

# Video/Audio Networked surveillance system enhAncement through Human-cEntered adaptIve Monitoring

**Large-scale integrating project**
**Grant Agreement n°248907**
**01/02/2010 – 31/07/2013**

---

**Contractual delivery date: 30 April 2010**
**Actual delivery date: 30 April 2010**

## Deliverable D2.1
## End user requirement and system objectives (version 1)

---

## D2.1

**Version: 6.0**
**Author: GTT/RATP**
**Contributors: THALIT, MULT, INRIA, TCF, IDIAP**
**Reviewers: IDIAP, UNIVIE**
**Dissemination level: PU**
**Related document(s): –**
**Number of pages: 51**

SEVENTH FRAMEWORK PROGRAMME

## Document information

| Ver. | Date | Changes | Author (partic.) |
|---|---|---|---|
| 1.0 | 16/03/2010 | Edition of functional blocks requirements document | A. Grifoni (THALIT) |
| | 01/04/2010 | First edition of document | E. Jouneau (MULT) |
| | 02/04/2010 | Contributions to I3S & AVAS requirements | C. Carincotte (MULT) |
| 2.0 | 05/04/2010 | Contributions to audio and autonomous selection | B. Ravera (TCF) |
| | 07/04/2010 | Contributions to situational reporting, etc. | C. Carincotte (MULT) |
| 3.0 | 09/04/2010 | Contribution to mid-long term analysis | J.L. Patino (INRIA) |
| | | Review of HMI and AVAS requirements | A. Grifoni (THALIT) |
| 4.0 | 12/04/2010 | Contributions to functional and usage context requirements | A. Forchino (GTT) |
| | 13/04/2010 | Modifications to AVAS requirements | A. Grifoni (THALIT) |
| | | Contributions to event detection blocks requirements and scenarios | J.M. Odobez (IDIAP) |
| | 14/04/2010 | Last contributions and final formatting | C. Carincotte (MULT) |
| 5.0 | 26/04/2010 | General review of architecture & functional requirements and answers to comments | A. Grifoni (THALIT) |
| | | Contributions to safety, security and privacy requirements | A. Forchino (GTT) |
| 6.0 | 30/04/2010 | Final version | C. Carincotte (MULT) |

| Ver. | Date | Approval/Rejection decision/comments | Author (partic.) |
|---|---|---|---|
| 4.0 | 23/04/2010 | Comments on architecture & functional requirements<br>Approval of deliverable | J.M. Odobez (IDIAP) |
| 4.0 | 26/04/2010 | Comments on architecture & functional requirements<br>Approval of deliverable | K. Grammer (UNIVIE) |

## Copyright



Consisting of:

| | | |
|---|---|---|
| **Coordinator**: | Multitel asbl (MULT) | Belgium |
| **Participants**: | Gruppo Torinese Trasporti (GTT) | Italy |
| | Institut Dalle Molle d'Intelligence Artificielle Perceptive (IDIAP) | Switzerland |
| | Institut National de Recherche en Informatique et en Automatique (INRIA) | France |
| | Régie Autonome des Transports Parisiens (RATP) | France |
| | Thales Communications (TCF) | France |
| | Thales Italia (THALIT) | Italy |
| | University of Vienna (UNIVIE) | Austria |

# 1 Executive Summary

The objective of this deliverable is to collect and present the requirements needed for the development of the monitoring components foreseen in the project. Both technical requirements and end-users recommendations are gathered on respectively the technical aspect of the VANAHEIM architecture management, and on the end-user usage context scenarios envisaged in the project. This report therefore summarizes requirements on both technological needs and end-user monitoring usage expectations.

Reported requirements will be used for three purposes: to define the behavioural models required to perform the desired recognition tasks (T2.2), to specify the systems and components technical layout (T2.3), and to delineate the trials scenarios for the evaluation stage (T3.1).

# Table of contents

# List of Figures

## Glossary

| | |
|---|---|
| ACGTT | Audio-Codec GTT |
| ACRATP | Audio-Codec RATP |
| AVAS | Audio-Video Analysis System |
| AVAS-API | Audio-Video Analysis System - Application Program Interface |
| AVRS | Audio-Video Recording and Indexing System |
| Codec-GTT | Audio-Video Encoder in use for the demonstrator in GTT - Torino |
| Codec-RATP | Audio-Video Encoder in use for the demonstrator in RATP - Paris |
| DB | Data Base |
| HCI | Human Machine Interface |
| HD | High Definition |
| I3S | Integrated Security Supervision System |
| I3S Client | Integrated Security Supervision System – Client application |
| I3S Server | Integrated Security Supervision System – Server application |
| MMI | Man Machine Interface |
| SRV | Server |
| SSS | System/ Subsystem Specification |
| TBC | To Be Continued |
| TBD | To Be Defined |
| VCGTT | Video-Codec GTT |
| VCRATP | Video-Codec RATP |
| VS | VANAHEIM System |
| VWM | I3S Video Wall Manager application |
| WS | Workstation |

# 2 Introduction

This document provides a high level overview of the whole VANAHEIM System (VS) :

- summarizing the objectives and design of the VS, in Sec. 2.1;
- specifying the end users' functional and operational requirements, in Sec. 3.
- and describing the related architecture of the VS and functional blocks requirements, in Sec. 4;

In brief, the VS will be a flexible and fully comprehensive video-audio solution designed for the management of large and heterogeneous video and audio surveillance networks. The VS will give to end-users all the facilities to control the whole video-surveillance network (top-down management) and dynamically guide the operator to focus on the relevant event detected autonomously by the system (bottom-up automation), as more detailed in the following sections.

## 2.1 Operational system objectives

In more details, the aim of VS is to study **innovative surveillance components for autonomous monitoring of complex audio/video surveillance infrastructure**. To do so, the VS will address three main application-driven research questions:

- **Scene activity modeling algorithms for automatic sensor selection in control room.**

  In everyday practice , surveillance video wall monitors frequently show empty scenes, while there are obviously many cameras looking at scenes in which something (even normal) is happening. Performing a sensor selection at the control room level to autonomously select the streams to display therefore seems required. While this scenario is trivial when dealing with "*empty vs. occupied*" scenes, building models to characterize the streams content, in terms of usual and unusual activities, turns out to be necessary when dealing with almost all occupied scenes. Furthermore, the need for such selection is even more explicit when dealing with audio streams, for which mosaicing of data is not possible due to the transparent nature of sound.

- **Investigation of behavioral cues for human-centered monitoring and reporting.**

  One of the limiting factors in the development of CCTV is the slow technology advances in the field of video content analysis, that do not allow to efficiently perform the human modeling and spatio-temporal reasoning usually required for the user-desired applications. For example, when application such as group detection is targeted, the need for robust and reliable huma*n-centered modelin*g components is obvious. While significant academic research into this field is ongoing and corresponding literature already propose algorithms partially addressing the human and/or interactions modeling topics, most of the time studies have been conducted in restricted environments with little (acted) or even synthetic data, and usually do not rely on behavioral cues. As a matter of fact, video analysis software currently available are not able to efficiently detect, understand and recognize human behaviors in real-scale environment, and thus fail to fulfill the requirements of an operational monitoring system. Moreover, besides surveillance-like scenarios recognition capabilities, efficient human-centered monitoring components could also revolutionize the everyday operation of infrastructure monitored by CCTV, by providing real-time user-oriented reporting consisting in live environmental figures. For instance, such situational awareness reporting could provide the locations and numbers of people in the monitored areas through a real-time "occupation map" or also provide the user behavior category itself (e.g. people walking/waiting, people queuing at vending machine or getting on/off the train, etc).

- **Collective behavior building and online learning from long-term analysis of passenger activities.**

  While reliable monitoring system could answer the safety/security concerns mentioned above, it could also be of great help for infrastructure designers and managers. More precisely, transportation planners and public transport operators have become increasingly aware of the need to evaluate their transportation infrastructures under a variety of criterions (regularity, efficiency, capacity, accessibility, quality of service…). Nonetheless, the bottleneck of such analysis consists in the high variety and complexity of passenger behaviours in the infrastructure, which cannot be quantified and analyzed in a straightforward way. An example of this complexity lies in the fact that passengers sharing the same goal will have different routes depending on their entry points, the time of day, the passenger density in the area, their interactions with the infrastructure equipments or with other users, etc. Furthermore, while particular transportation facilities such as turnstiles or ticket vending machines can provide information about "localized behaviours", they are clearly not suited to provide finer understanding of the infrastructure usage, for example in terms of static and dynamic space occupancy (people walking, waiting...), passengers' in-station routing behaviour, etc.

Taking into account the VS targets the development of automatic audio/video components in these three specific areas, Sec. 3 details the end-user functional and operational requirements expressed by GTT/RATP in these three areas of interest, respectively in Sec. 3.2, 3.3 and 3.4. As for Sec.3.1 , it briefly introduces some more general end-user requirements about the system, together with the description template used to characterize the operational requirements.

## 2.2 System architecture design

At the submission time, a reference architecture for the project has been drafted, to illustrate the targeted architecture. It is an almost complete design integrated into a collaboration grid, in which the main functional and technical project requirements can be allocated to the represented blocks. Each block is autonomous and covers specific functionalities; the functionalities of the overall VS being provided through the collaboration of the various blocks that compose this design. As highlighted in Figure 1 and detailed below, each block is associated to a dedicated functional layer,

The considered internal system layers are:

- **Presentation Layer**: it implements the man-machine (or human-machine) interface (MMI or HMI) for the environment supervision.
- **Supervision Layer**: it implements the highest supervision level of the whole video-surveillance network, including data logging and reporting.
- **Analysis Layer**: it implements the smart analysis functionalities on the live and recorded audio/video streams.
- **Recording Layer**: it implements the permanent storage of the audio/video streams and of the data produced by the Analysis Layer.
- **Streaming Layer**: it represents the zone of publication of all the media streams produced or consumed by the various layers.
- The external actors of the VANAHEIM SYSTEM are:
- **Security User**: the operator of the system.
- **Field Device**: IP cameras, codec, audio streaming devices, etc.

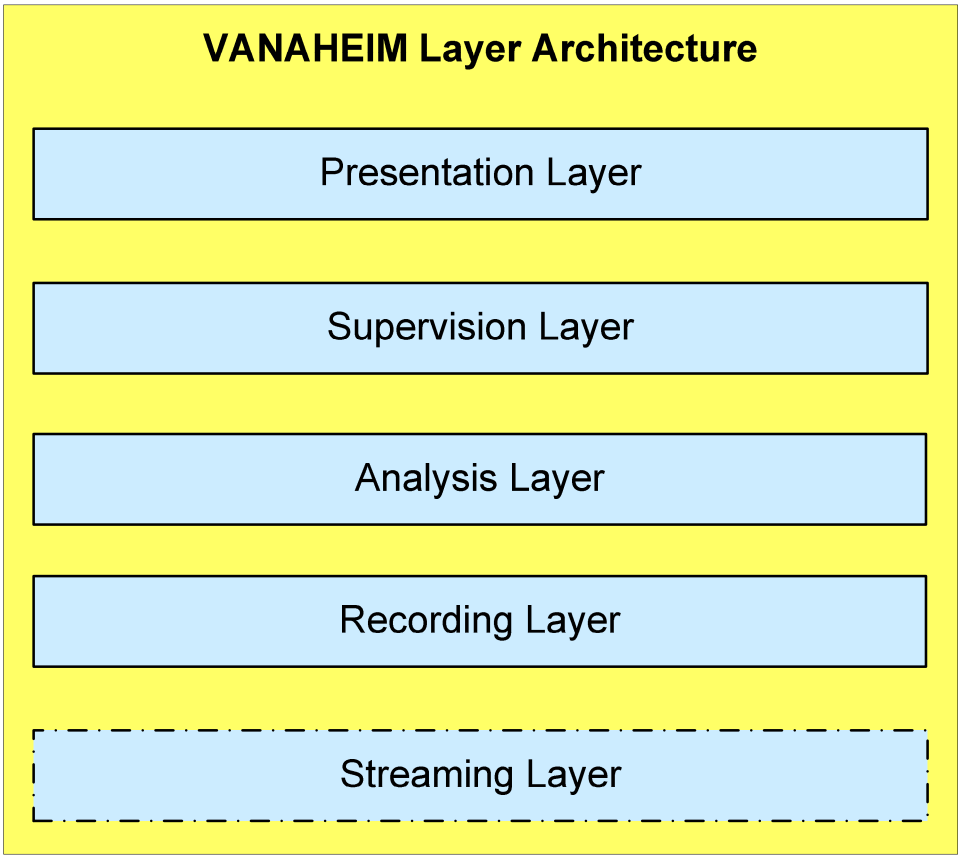


**Figure 1 VANAHEIM Layer architecture.**

Following this system architecture design, Sec. 4 then details the architecture and functional blocks requirements needed to achieve and fulfil the operational system objectives presented in Sec. 3.

# 3 End users' operational scenario requirements

This section presents the functional and operational requirements of the system, described in terms of scenarios of interest from the end-user point of view.

The joint analysis of GTT and RATP missions, combined with three aspects/categories of the VS, namely the autonomous stream selection (WP4), the real-time applications for human monitoring (WP5/T6.1) and the collective behaviour building (T6.2/T6.3), led to the identification of several scenarios of interest which are detailed in this section.

In the next sub-section, we first introduce the general end-user requirements that are applicable to all the categories and the system as a whole. We then present the template used to describe the different scenarios, last review the detailed requirements of each scenarios by category.

## 3.1 General end-user requirements

### 3.1.1 General end-user requirements

- The HMI must be user-friendly, easy and intuitive to understand.
- Solutions for intellectual property rights and data ownership need to be found.
- End users shall receive reasonable training on the system before the demonstration in order to promote acceptance.
- HMI shall use pictograms and only a minimum of text explanations in order to support language-independent and quickly understandable information.
- The system shall use standard audio-video equipment, whenever possible.
- The system shall comply with already installed audio-video sensors, when available.
- Misuse of services shall be avoided by security features (e.g. passwords).
- The operator shall be able to select the event/function/query of interest from a pre-defined list.
- Different level of access to data must be defined, allowing different access to real-time analysis or to recorded data.
- The video surveillance recording system shall guarantee a sufficient storage capacity in line with the national privacy law requirements.
- The recorded images shall be kept in a secure place protected not only from the environment but also from unauthorised removal or viewing.
- The system shall automatically delete recorded data once they are no longer needed for analysis.

Furthermore it should also be mentioned that privacy requirements can be seen as horizontal requirements that shall be applicable to all aspects of the system, due to the high relevance of privacy issues in audio-video data management.

### 3.1.2 Scenario template

So as to describe the scenarios in a consistent way, a common structure to describe each scenario has been defined in agreement with all the partners. Hereafter the structure to describe each scenario is reported, with a short summary of the main content of each one.

---

**NAME OF THE SCENARIO**

**Description of the scenario**
*Brief description of the scenario from the end-user point of view.*

**Motivation from the user**
*Brief description of why the user is interested in the scenario.*

**Characterisation of the scenario**
*Defines the main characteristics of the scenario*

> **Definition of the scenario**
> *Presents a end-user definition of the scenario*
>
> **Usual location of the scenario**
> *Provides indication of most usual or useful location for the scenario.*
>
> **Human behaviour/actions**
> *Describes the human behaviour or actions related to the scenarios, in terms of expected or available audio and video data information.*

**What it implies in terms of action from the operator?**
*Describes the action requested by the operator before/after the scenario.*

**What is expected in terms of information from the system?**
*Describes the main output of the system from the end-user point of view.*

**How does it impact to other parts of the system?**
*Describes the possible connection with other part of the system, such as relation to other scenarios.*

---

## 3.2 User requirement for autonomous sensor selection

Regarding the autonomous sensor selection part, the developed components should perform an unsupervised and efficient sensor selection among the huge amount of data streams available in such CCTV systems. First, components for attracting operators' attention to relevant/salient audio streams, or spotting out corresponding microphones or areas, should be developed. Secondly, components allowing to perform a video sensor selection at the control room level to select the streams to be displayed on monitors, based on the video stream content normality/abnormality, should be developed. Last, audio-video joint-processing adaptation should be investigated in well-suited physical configurations, to exploit the complementary nature of video and sound in particular part of the infrastructure (e.g. at the platform where both audio and video data can be used to infer a metro arrival/departure, which could be of interest for monitoring operations).

---

**AUTONOMOUS SENSOR SELECTION**

**Description of the scenario**
The autonomous sensor selection system shall perform a sensor selection at the control room level allowing to autonomously detecting the most interesting/relevant/abnormal video streams to display on the videowall and audio streams to play on loudspeaker in the control room.

**Motivation from the user**
The motivation is that, in everyday practice, surveillance videowall monitors frequently show empty scenes, while there are many others cameras looking at scenes in which something (even normal) is happening. Furthermore, the need for selection mechanisms is even more explicit for audio streams, for which mosaicing of data is not possible due to the transparent nature of sound. Last, as vigilance studies confirm, operators who spend hours "screen gazing" at static scenes tend to become bored and less efficient, and are then likely to miss low-frequency events.

The main motivation for both RATP and GTT is thus to make the video/audio monitoring task easier for security operators in control room, and overall more efficient, by providing them with a fully or partially automatic monitoring system able to intelligently select the audio/video streams to play/display, therefore reducing the amount of manual operation requested by the monitoring task, and increasing the probability to watch/listen to the right streams at the right time.

**Characterisation of the scenario**

**Definition of the scenario**
The scenario can be defined as providing security operators with videowall / loudspeakers in the control room that are automatically managed by an intelligent system, allowing to autonomously detecting the most interesting/relevant/abnormal video streams to display on the videowall and audio streams to play on loudspeaker.

**Usual location of the scenario**
The location of cameras and microphones concerned by this scenario correspond to the usual requirements for surveillance coverage needs of areas in underground networks, i.e. entrances and (emergency) exits, automatic ticketing machines, lifts and escalators, passageways and platforms.

**Human behaviour/actions**

While the operational scenario corresponding to the autonomous sensor selection is clearly well-defined, there are no behavioural scenarios (in terms of human behaviour to detect) that can be drawn for this task. More precisely, what is envisaged here is to work on data stream in terms of normal/abnormal content, and not on semantic description of scene/stream content itself. As a matter of fact, the selected audio/video streams should be the less normal / more abnormal among the whole audio/video streams available at that time. In this context, the selected streams should not always contain abnormal events, but they should be the most relevant streams to play/display in comparison to the others ones, and the specification of human behaviour related to this scenario is therefore almost impossible. However, some indicative and obvious practical abnormal configurations that could be considered as abnormal in comparison to usual conditions can be delineated considering the same camera, for example people stopped in stairs or escalators, people standing still in areas where people are usually passing by (and vice versa), etc.

**What it implies in terms of action from the operator?**

The operator should be able to customize the display area to best suit the number of camera windows required; he should also be able to choose between a fully automatic mode and a partial mode (possibility of partial manual operation), to select the number and position of videowall screens and loudspeakers to be autonomously controlled, and last to select a set of main relevant cameras on which to apply the automatic selection.

**What is expected in terms of information from the system?**

The system should offer the ability to identify on the videowall / loudspeakers the ones which are automatically selected; it should also provide a way to identify, in the selected data streams, the ones which present a very high probability of abnormality. Last, the system should allow the combination and play/display of audio/video and metadata simultaneously, through for example text overlaid, or normality gauges. The sensor selection provided by the system should also be self evident: the audio and video streams should be presented to the operator in a way that the reason why the specific stream and sensors are relevant for the operations is fully clear. Relevant streams related to different instances of the same abnormal event should be presented in a human-centred logic and with regards to normal end user reactivity. This means that, as an example, if an aggression occurs and the system understands it from many different cameras and microphones it should present it displaying only the sensors that has the constant “average” maximum abnormality instead of presenting camera 1 and microphone 1 for 5 seconds and then camera 2 and microphone 2 for 5 seconds and then camera 3 and microphone 3 for 5 seconds and then another time camera 1 and microphone 1 for 5 seconds. This is because, even if changing camera frequently basing it on the results of the algorithms gives the operator the maximum amount of information, this information is not easily serviceable.

**How does it impact to other parts of the system?**

While there should not be any semantic description of the analysed audio/video data streams, the normality level corresponding the selected data streams should be stored in a database for further exploitation, e.g. through the mid-/long-term analysis.

---

## 3.3 User requirement for real-time applications for human monitoring

The goal and, thus, the user requirements for real-time applications for human monitoring mainly concern the following items:

- protecting real estate, infrastructures, technological systems and equipment as well as their assets, in order to avoid intentional deeds (theft or damage);
- protecting rolling stock both while moving and standing;
- deterring possible crimes and/or interference of people working in the stations or transiting;
- assisting and protecting company personnel working in direct contact with the public, in critical overcrowding, protest and turmoil situations;
- managing and preventing dangerous situations in order to reduce/remove the risk of robbery, theft, bag-snatching and harassment in stations;
- actually performing the task of protecting people and places, thus guaranteeing a secure environment.

Real-time applications should thus be defined as analysis of audio-video data providing immediate information and knowledge to operators on pre-defined scenarios detected by the system, which can be useful to manage problems without any delay. These applications should concern the detection of the critical safety/security situations that can slow down or interfere with the everyday use of public transportation (event detection applications), as well as the reporting of useful information extracted from the data streams (situational reporting).

Taking into account the project targets three specific levels of monitoring (individuals, groups of people and crowd/people flow), next sub-sections detail the end-user scenario for real-time applications which are related to these three levels, and which are summarized below:

- *Abandoned and stolen luggage* scenario (Sec. 3.3.1);
- *Group detection* scenario (Sec. 3.3.2);
- *People arguing, entering in conflict* scenario (Sec. 3.3.3);
- *Crowd/Flows of people* scenario (Sec. 3.3.4);
- *Monitoring equipment* scenario (Sec. 3.3.5);
- and *Situational reporting* scenario (Sec. 3.3.6).

### 3.3.1 Abandoned and stolen luggage scenario

---

**ABANDONED AND STOLEN LUGGAGE SCENARIO**

**Description of the scenario**
This scenario focuses on detecting lasting changes to the scene. In particular, it should allow to identify abandoned luggage. It should also explore finer activity analysis involving people to detect quiet stealing of luggage due to inattention and diversion.

**Motivation from the user**
Quickly detecting left luggage and theft have impact on overall security. The main motivation of this event is to help the operator to see abandoned objects, which could be a menace for the safety or security of the station. In addition, such information can be used to manually or automatically trigger vocal messages inviting passengers to check they haven't forgotten anything and to be more vigilant. Preventing people from forgetting their luggage saves interventions from the security guards.

**Characterisation of the scenario**

**Definition of the scenario**
A package is abandoned in an area monitored by the CCTV network. The operator is alerted if the abandoned package is not removed after few minutes.

**Usual location of the scenario**
While the event can occur in any areas of the station, Elevators and hall are best suited for this scenario. Another possible location is the platform.

**Human behaviour/actions**
The scenario can be defined by a person (or group) entering the scene with some luggage then putting it on the floor. From this situation, one possibility is that the person (or group) leaves and forgets the luggage that is then detected as static for a long time and thus reported by the system. Another possible situation is that another person takes advantage of a moment of inattention (natural or caused by some external event) of the person (or group) to take the luggage without person A noticing it. No specific audio event is envisaged, and expected video information related to the event(s): A package stays for a few minutes in the same position within the monitored area of a CCTV camera.

**What it implies in terms of action from the operator?**
The action requested by the operator before starting the sub-system only consists in the selection of the cameras that should be processed by the sub-system. The operator should then be alerted if an abandoned package is not removed after few minutes. An acknowledgment step should also be included to allow the end-user to confirm or invalidate the alarm.

**What is expected in terms of information from the system?**
The system should alert the operator if the package is not removed within few minutes. The system should output alarms with one or more key locations and key frames (person entering, person dropping bag, another people stealing it, etc.). This information should help the operator in evaluating the alarm.

**How does it impact to other parts of the system?**
This scenario could exploit group information from the “Group Detection” scenario. Statistics about person's attention provided by this scenario could be used for the “collective behaviour building” scenario. Lasting scene change with only low confidence can also increase the abnormality level of the camera instead or before raising an alarm.

---

### 3.3.2 Group detection scenario

---

**<u>GROUP DETECTION</u>**

**Description of the scenario**
This scenario should focus on the identification of groups and analysis of people's interaction with each others.

**Motivation from the user**
The motivation for the group detection scenario is that the insecurity perception is often due to the presence of groups with loudly and disrespectful behaviours such as vandalism and fighting groups.

**Characterisation of the scenario**

**Definition of the scenario**
This scenario should focus on the recognition of the (social) interaction between two or more people, so as to be able to identify presence of groups in the monitored areas.

**Usual location of the scenario**
The Hall area could the most adequate monitoring space for this scenario. Specific attention could be paid to select a group of cameras to ensure the best views of the potential group.

**Human behaviour/actions**
The human behaviour or actions related to this scenarios, in terms of audio/video data available, is mainly related to people facing each other, people having a discussion, body shape and head pose,

**What it implies in terms of action from the operator?**
N/A

**What is expected in terms of information from the system?**
The outcome from this analysis should be the detection and notification of presence of groups in the monitored areas.

**How does it impact to other parts of the system?**
Other system modules shall employ this information, such as the general mid-long term analysis module.

---

### 3.3.3 People arguing, entering in conflict scenario

---

**PEOPLE ARGUING, ENTERING IN CONFLICT**

**Description of the scenario**
The video surveillance system shall provide a monitoring to alert operators in case of people arguing, entering in conflict or aggression happening in the infrastructure.

**Motivation from the user**
Safety and security in the station are naturally the main motivations for this scenario.

**Characterisation of the scenario**

**Definition of the scenario**
The video surveillance system shall provide a monitoring to alert operators in case of people arguing, entering in conflict or aggression happening in the infrastructure.

**Usual location of the scenario**
Due to the fact that in ticket vending machines area, passengers buy tickets and consequently money is exchanged, this area could be considered as a good location for the aggression scenario. In addition, passageways could be considered for this scenario, since they are considered very important for passenger security especially at night when there are not many passengers.

**Human behaviour/actions**
Characteristic audio and visual cues (people standing near each other and with abnormal levels of noise) should be extracted and analysed in order to perform this scenario.

**What it implies in terms of action from the operator?**
No previous action to the scenario recognition is needed; however an acknowledgment level should be included, to allow and ask the end-user to confirm the alarm.

**What is expected in terms of information from the system?**
This scenario should provide an autonomous notification to surveillance operators, alerting for a potential conflictual situation in the monitored area.

**How does it impact to other parts of the system?**
Depending of the configuration made, it could be envisaged that other modules can automatically be launched such as the ‘automatic sensor selection’ module.

---

### 3.3.4 Crowd/Flows of people scenario

---

**CROWD/FLOWS OF PEOPLE**

**Description of the scenario**
This scenario should provide an automatic system for obtaining detailed information about the movements and dynamics of collective people flow (crowded scene) observed by a camera.

**Motivation from the user**
GTT and RATP are interested in this scenario because crowd monitoring is obviously a key aspect to ensure public safety in transport networks, and to guarantee the efficient management of transport networks and public facilities.

**Characterisation of the scenario**

**Definition of the scenario**

This scenario should allow the analysis of crowd movement, and for example provide a continuous measurement of crowd activities, congestion status, identify some abnormal crowd behaviour, etc.

**Usual location of the scenario**
The most usual and useful location for this scenario is the platform area. Here it is very important that the video surveillance system guarantee surveillance to prevent any kind of accidents, not only in normal service, but also in crowded situations during the rush hours.

Additionally, considering the continuously increasing stream of passengers in RATP, some corridors or passageways in RATP pilot site could also be a relevant location to apply such analysis.

**Human behaviour/actions**
In this scenario, and in addition to the crowd movement/dynamic itself, the behaviours/actions that should be expected or available in audio and video data can be summarized as follows; crowd boarding/alighting metro, overcrowded situation, congestion or panic, rapid crowd dispersion, crowd agglomeration, etc.

**What it implies in terms of action from the operator?**
N/A.

**What is expected in terms of information from the system?**
The main output of the system from the end-user point of view for this scenario should be graphical user-friendly information about the movements within the crowded scene, highlighting for example passengers direction information, and/or pointing out the different flows of people within the crowd.

**How does it impact to other parts of the system?**
N/A.

---

### 3.3.5 Monitoring equipment scenario

---

**MONITORING EQUIPMENT**

**Description of the scenario**
This scenario should allow to monitor people behaviours evolution employing different equipment such as ticket vending machines, turnstiles...

**Motivation from the user**
The main motivation for this scenario is related to the management of the infrastructure.

**Characterisation of the scenario**

**Definition of the scenario**
The monitoring equipment scenario should provide useful statistics on the use of specific station equipments, e.g. number of users, percentage of use, duration of use by different time periods (e.g. morning or evening, weekdays or weekends) for the different equipments.

**Usual location of the scenario**
The set of cameras involved for this scenario will be those with a full view of the monitored equipments.

**Human behaviour/actions**
The human behaviour or actions related to the scenarios is related to the use of the concerned equipment

**What it implies in terms of action from the operator?**
The specific equipments of interest must be specified by the end-user.

**What is expected in terms of information from the system?**
The previously mentioned statistics should be computed and stored into the mid-long term specific databases.

**How does it impact to other parts of the system?**
N/A.

---

### 3.3.6 Situational reporting scenario

---

<u>**SITUATIONAL REPORTING**</u>

**Description of the scenario**

The situational reporting tool should able to translate the current presence of people in some part of the infrastructure into one meaningful graphical figure, through a map-based overlay of the approximate location and number of people in the infrastructure.

**Motivation from the user**

The main motivation for this scenario is the ability to roughly estimate the number and locations of people in the infrastructure, thus allowing to monitor an entire floor of the infrastructure through a unique screen, without the need to look at the whole corresponding cameras. A secondary motivation, in the safety context, is the ability to rapidly point out the presence of people at a floor level, for example in case of emergency exit procedure.

**Characterisation of the scenario**

**Definition of the scenario**

The situational reporting should provide the approximate locations and numbers of people in the monitored areas through a real-time "occupation map". It should detect the presence of people in the selected cameras and transpose these detections to real-world coordinate into an infrastructure map.

**Usual location of the scenario**

The most useful camera locations for this scenario could be considered as the set of cameras covering an entire floor of the infrastructure, typically the mezzanine level.

**Human behaviour/actions**

Practically speaking, there is no human behaviour or action related to this scenario, since it should focus on the presence of human itself. However, one should mention that areas uncovered by CCTV could raise problem, since the "detection of people" in these areas should be achieved by, if possible, extrapolating the presence of people in covered areas.

**What it implies in terms of action from the operator?**

The action requested by the operator before starting the situational reporting sub-system only consists in the selection of the cameras that should be processed by the sub-system. No action is required after due to the reporting nature of the application.

**What is expected in terms of information from the system?**

As already described, the main output of this sub-system should be the reporting into user-friendly MMI of the approximate position of the detected person in the selected cameras, and possibly, also the reporting of assumed people in uncovered area.

**How does it impact to other parts of the system?**

There should be no connection with other part of the system, except maybe the storage in a database of the number of person detected by camera, e.g. for further exploitation through the mid-/long-term analysis.

---

## 3.4 User requirement for collective behaviours building

Modern transport infrastructures and even older ones are built in a way that with a good knowledge and modelling of customer's behaviour, the operator can configure them in a way that the infrastructure itself can be used more safely and efficiently. A simple example of that is the escalators sense of working or the crowd exit suggested direction that can be changed depending of the situations and the surrounding environment.
To do so, it is necessary to get detailed and well defined trends and relationships among events both over short and long term periods.

Collective behaviour building should thus be defined as an offline analysis of audio-video data and metadata providing information and knowledge on past human behaviours and events to operators. This analysis should be used to compute information on trends, repeated events and relationships among events that can be useful for information retrieval and resources planning.

---

**HOW IS THE STATION USED OVER A MID-/LONG- TERM PERIOD?**

**Description of the scenario**
The objective of this scenario consists in learning and analysing passenger dynamics, activities and behaviors in the infrastructure over mid/long period of time, so as to provide a comprehensive long-term analysis of passenger activities' trends and evolutions within the infrastructure.

**Motivation from the user**
Due to the passenger demand growth and related capacity problems, transportation planners and public transport operators are now looking for ways to evaluate their transportation infrastructures under a variety of criterions (regularity, efficiency, capacity, accessibility, quality of service…). The main motivation for the mid-/long- term analysis foreseen in VANAHEIM is therefore to provide a fine and long-term understanding of the infrastructure usage by passengers, which can be useful for identifying and quantifying behaviour trends (e.g. by different time periods like morning or evening, weekdays or weekends), or for identifying relationships between the behaviours observed at different locations.

**Characterisation of the scenario**

**Definition of the scenario**

This subsystem should be able to estimate of the long-term trends of large-scale human behaviour that can be extracted using audio/video analysis, allowing the discovery of collective comprehensive daily routines.

**Usual location of the scenario**
The location of cameras and microphones concerned by this scenario correspond to the usual requirements for surveillance coverage needs of areas in underground networks, i.e. entrances and (emergency) exits, automatic ticketing machines, lifts and escalators, passageways and platforms.

**Human behaviour/actions**
The human behaviour or actions related to this scenario, in terms of available information, are quite numerous since all information that can be extracted using audio/video analysis, and even contextual information, could be used in this scenario. Briefly, people activities (walking, waiting, interactions with others passengers and/or equipments...), people trajectories (entry point, in-station routing behaviour, waiting/travelling time…), contextual information (such as location, time of day (morning or evening, weekdays or weekends), density of people, etc.), proxemic information (such as distance to walls) should be investigated.

**What it implies in terms of action from the operator?**
Regarding the action requested by the operator to manage this scenario, this should mainly consist in user-friendly compound query-based search using e.g. object type, camera location, etc. data mining applications dedicated to site activity summarization, site activity pattern discovery, etc. Transport operator should be enabled to create queries on the relationship among different events in order to get the correlation of them and the specific period trend.

**What is expected in terms of information from the system?**
While by definition outcomes of such tools/analysis are not predictable, mains areas of analysis should be related to passenger counts and occupancy (density maps, station space utilization analysis), passenger speeds, paths and route choice, boarding/alighting analysis and waiting times, queuing time… Among the expected outcomes, the localization of common loitering areas and/or highly frequented aisles, the monitoring of the usage of stairs in comparison to escalators, as well as the identification of traffic patterns in the infrastructure could also be addressed. The identification of different routing behaviours and their influencing and limiting factors could also be targeted. Last, the results should be presented at different level of complexity depending if the analysis is dedicated to transport planners operations managers or security agents or infrastructure designers.

**How does it impact to other parts of the system?**
N/A

# 4 Architecture and functional blocks requirements

Following the operational system objectives and scenarios presented above, this section details the architecture and functional blocks requirements needed to technically achieve and fulfil these scenarios, as well as the standard requirements needed to perform the overall surveillance monitoring tasks foreseen in VANAHEIM.

In brief, this section is structured as follows:

- Sec. 4.1 first introduces the system components and the nomenclature used for specifying the functional blocks requirements;
- Sec. 4.2 then describes the main system capability requirements, in terms of:
    - MMI, Supervisor and video wall requirements (Sec. 4.2.1);
    - Audio-Video analysis requirements (Sec. 4.2.2);
    - Audio-video recording and metadata indexing requirements (Sec. 4.2.3);
    - Mid-long term analysis & reporting requirements (Sec. 4.2.4);
    - Streaming service requirements (Sec. 4.2.5);
- system external interfaces required to deploy VS in GTT/RATP sites are then presented in Sec. 4.3;
- computer resource and system environment requirements are presented in Sec. 4.4 and Sec. 4.5;
- safety, and security and privacy requirements are last presented in Sec. 4.6 and Sec. 4.7.

## 4.1 System components and functional block requirement nomenclature

The VANAHEIM system is designed as a multi-layer architecture composed, as shown in Figure 2, by several blocks:

- **Presentation Layer:**
  - Operator Client (MMI)
  - Video Wall Manager (VWM)
  - Video Wall (VW)
  - Audio Wall - Set of loudspeakers (AW)
- **Supervison Layer:**
  - Supervisor
  - Map Service
- **Analysis layer:**
  - Audio-Video Analysis Service (AVAS)
  - Mid-Long Term Analysis (Metadata Management) & Reporting Service (MTAS)
- **Recording Layer:**
  - Audio-Video Recording and Metadata Indexing Services (AVRS)
- **Streaming Layer:**
  - Streaming Service

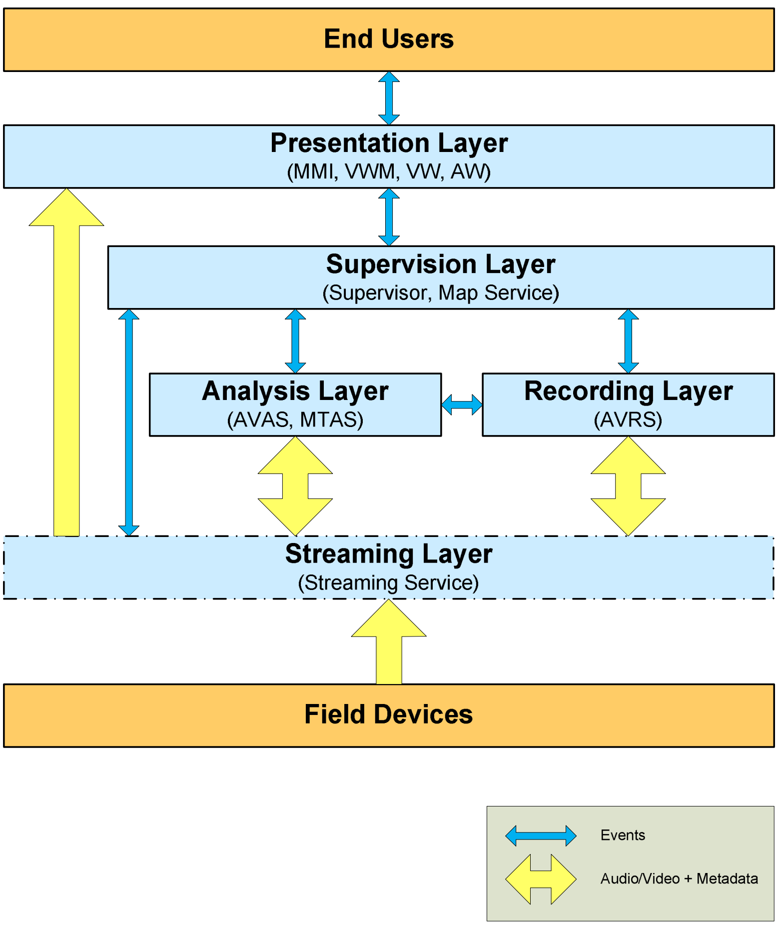


**Figure 2 VANAHEIM System Design.**

The **external actors** (field devices) considered are:

- the video surveillance sensors (IP cameras, audio-video codec);
- the users that are in charge of the supervision operations.

In more details, VANAHEIM architecture is represented in the figure below, together with the main data-flows. As illustrated, the *Switching/Streaming Layer* is responsible for the interfacing of the field sensors (cameras and microphone systems), and for providing the needed streams to the other project applications like recording, analysis and user visualization.

The *Audio-Video Recording and Metadata Indexing Server* (AVRS) is dedicated to the recording of the captured raw audio and video, and the storage of the metadata generated by the audio/video analysis components (AVAS and MTAS). It is also responsible for playing and streaming recorded audio-video streams towards the operators' *Video Wall Manager* (VWM).

The *Audio-Video Analysis Server* (AVAS) is used to embed the algorithms dedicated to the analysis of live (real-time algorithms) video streams. This component allow to acquire the live streams coming from the *Switching/Streaming Layer*, analyze them using one or several combined algorithms, generates the description metadata to be sent to the AVRS, and related real-time alarms that will be shown on user MMI by the *Supervisor (I3S)*.

The *Mid/Long Term Analysis and Reporting server* (MTAS) is dedicated to the mid-long term analysis (e.g. behaviour recognition/interpretation of collective behaviours) and to the consequent reporting. It directly interfaces the metadata DB in order to perform the requested data analysis and report.

The *Supervisor (I3S)* is the most important integration asset, because it interfaces all the components (e.g. streaming selection, recording/indexing start/stop, analysis modules activation/deactivation and parameterization/configuration, etc). The supervisor is in charge of providing the operators' interfaces (MMI and VWM) and other general supervision functions like alarm management.

Last, the User Interfaces (*MMI* and *VW/Speakers*) provide interfaces to the overall system, and the main functionalities to end-users (user access management, audio/video sensor selection and display, audio/video analysis algorithm selection and configuration, together with synchronized metadata visualization and alarm management, reporting, etc.).

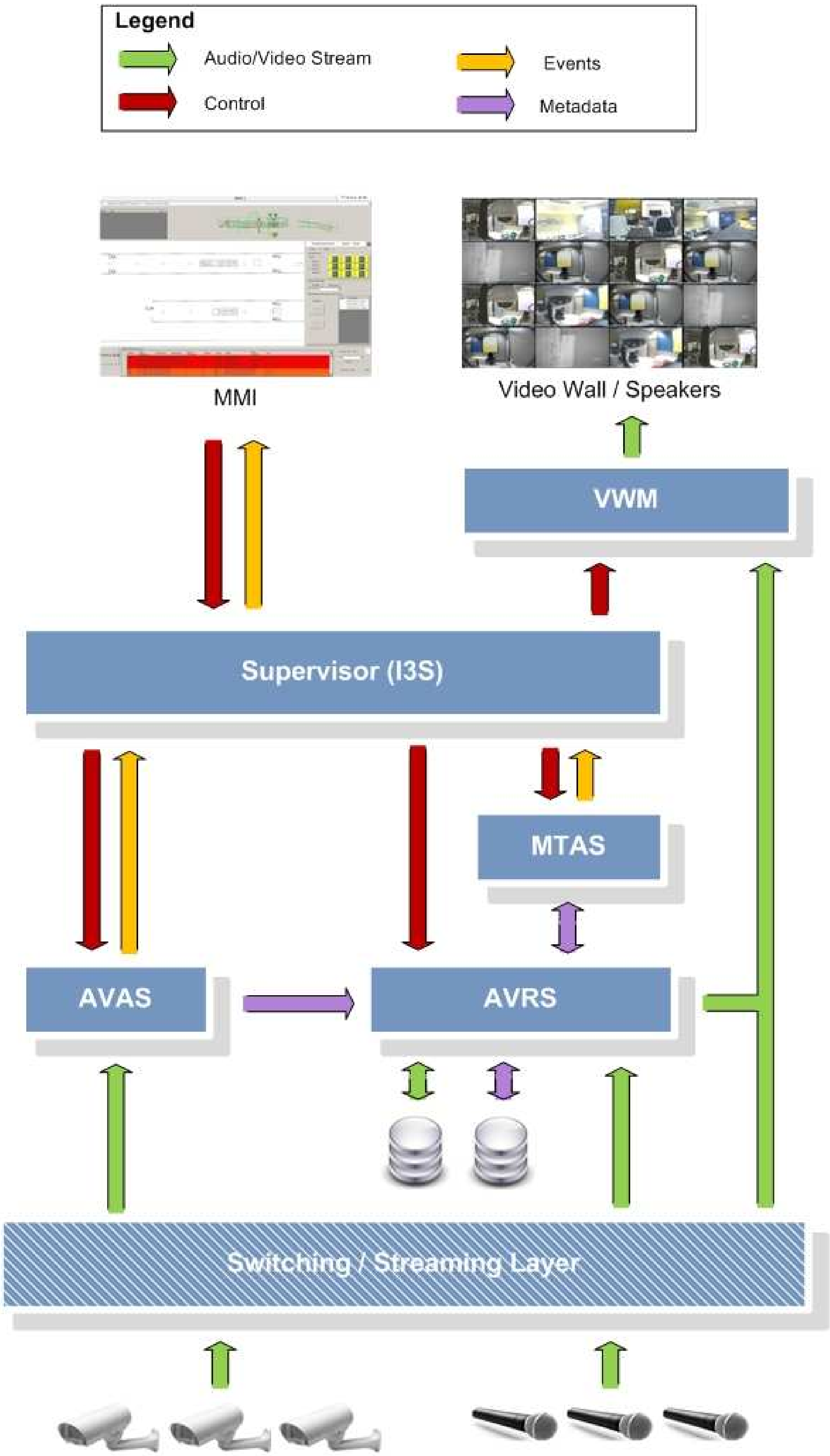


**Figure 3 VANAHEIM architecture and components.**

In the following sections, each requirement shall be identified by a code that follows the following rule:

**[SSS_AAA_BBB_CCC_DDD]**

Where:

“**SSS**” is a field used to identify the document type (System/Subsystem Specification).

“**AAA**” is a field used to identify the subsystem affected by the requirement:

“GEN”: for Generic requirements.

“VWM”: for Video Wall Manager requirements.

“I3S”: for I3S Client and/or Server requirements.

“AVAS”: for Audio-Video Analysis System requirements.

“AVRS”: for Audio-Video Recording System requirements.

“MTAS”: for Mid-Long Term Analysis System requirements.

“ACGTT”: for Audio-Codec-GTT requirements.

“ACRATP”: for Audio-Codec-RATP requirements.

“VCGTT”: for Video-Codec-GTT requirements.

“VCRATP”: for Video-Codec-RATP requirements.

“**BBB**” is a field used to identify, where applicable, the involvement of an external block or subsystem:

“GEN”: when no external block or subsystem is involved.

“VWM”: when the Video Wall Manager application is involved.

“I3S”: when the I3S Client and/or Server application is involved.

“AVAS”: when the Audio-Video Analysis System application is involved.

“AVRS”: when the Audio-Video Recording System application is involved.

“MTAS”: when the Mid-Long Term Analysis System application is involved.

“ACGTT”: when the Audio-Codec-GTT is involved.

“ACRATP”: when the Audio-Codec-RATP is involved.

“VCGTT”: when the Video-Codec-GTT is involved.

“VCRATP”: when the Video-Codec-RATP is involved.

“**CCC**” is a field used to identify the type of requirement:

“GEN”: for Generic requirement type.

“FUN”: for Functional requirement type.

“INT”: for Interface requirement type.

“HW”: for Hardware requirement type.

“**DDD**” is a unique number associated to each requirement.

## 4.2 System capability requirements

### 4.2.1 MMI, Supervisor and video wall requirements

This section describes the functionalities that the MMI, the Supervisor (including the Map Service) and the Video Wall Manager shall provide.

These blocks shall be based on the following software components provided by THALIT:

- Supervisor + Map Server: ***I3S Server*** application.
- MMI: ***I3S Client*** application (see Figure 5).
- Video Wall Manager: ***VWM*** (I3S – Video Wall Manager) application (see Figure 4).
- Video Wall: ***HD Display*** dedicated to video visualization purposes.

The requirements specified in the following part of this chapter refer directly to the components above specified. In order to better clarify the possible ambiguity between the *Video Wall* and the *I3S Client*, here below two examples for clarification. As showed in Figure 4, the *Video Wall* is the unique application able to display video streams.

**Figure 4 Example of Video Wall application.**

On the contrary, the *I3S Client* (see Figure 5) is responsible for the configuration and control of the whole system. In the next figure is showed how the MMI manages the CCTV subsystem:

- Navigation Area (on the top). It is used to quickly navigate through all the areas of the configured stations.
- Alarm Banner Area (on the bottom). It collects the list of currently active events/alarms generated by the system.
- Map Area (on the centre). It is a schematic that represent the map of a station area and where several devices are located (CCTV cameras, microphones, etc.). Each device is an active object able to represent its state changing its layout.
- Control Panel Area (on the right). This zone can display different panels according to the selected device. From any panel it is possible to manage all the functionalities available for the selected device. In the figure below it is represented the CCTV Control Panel used to control, for instance, the Video Wall layout and displayed cameras.

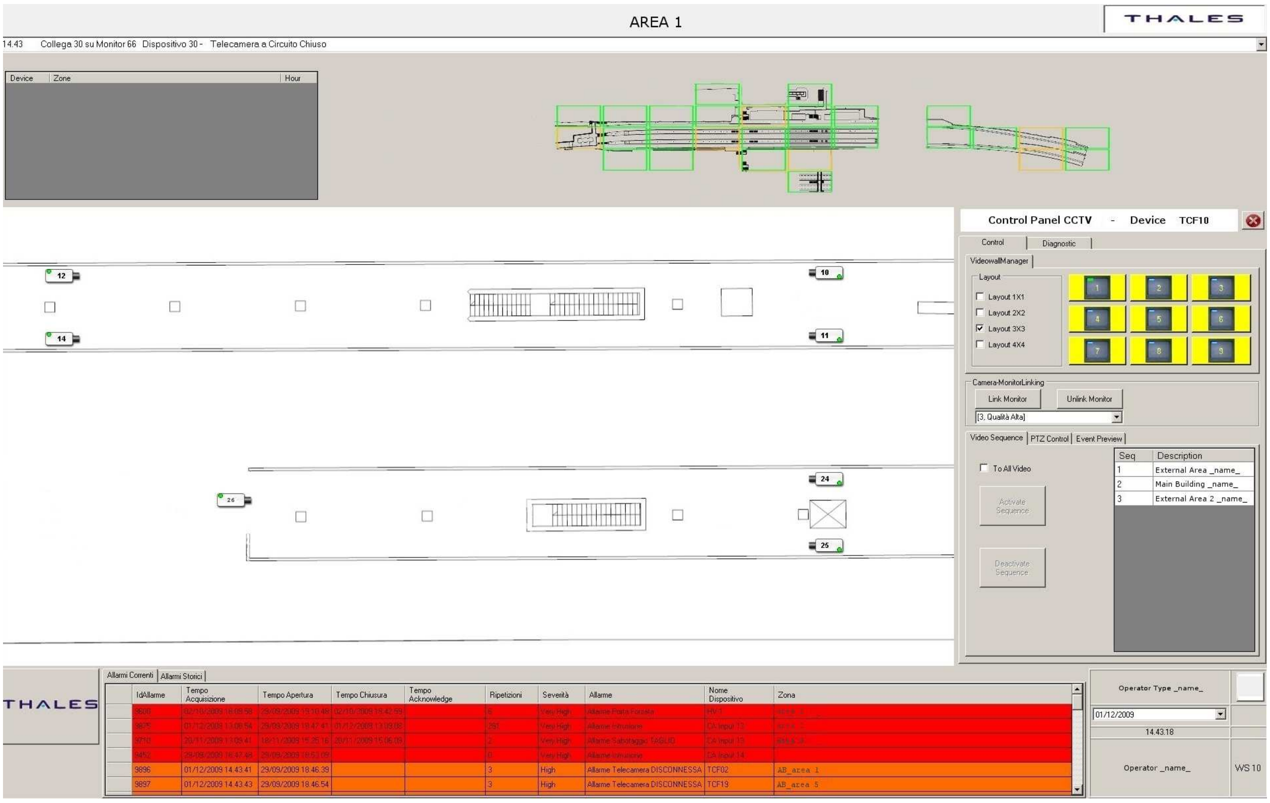


**Figure 5 Example of MMI (I3S Client) application.**

### 4.2.1.1 Installation requirements

The following requirements describe the I3S system constraints for installation on demonstrators.

**[SSS_I3S_GEN_HW_1]** The I3S Server application shall be installed as a stand-alone application. It shall not be redounded.

**[SSS_I3S_GEN_HW_2]** The I3S Server application shall be installed on a specific server (**SRV1**, see section 4.4.1.1).

**[SSS_I3S_GEN_HW_3]** The I3S Client application shall be installed on a specific workstation (**WS1**, see section 4.4.1.1).

**[SSS_VWM_GEN_HW_4]** The Video Wall Manager (VWM) application shall be installed on a specific server (**SRV5**, see section 4.4.1.1).

#### 4.2.1.2 I3S configuration requirements

Taking into account the I3S is the system in charge of data (mainly alarm) gathering in the architecture, the following requirements describe the functionality that should be configured in terms of alarms.

**[SSS_I3S_GEN_FUN_5]** The I3S Server application shall be configured in order to react with automatic actions to internal/external events.

**[SSS_I3S_GEN_FUN_6]** The I3S Server application shall be configured in order to raise alarms/events on critical situation.

**[SSS_I3S_GEN_FUN_7]** The I3S Server application shall provide the possibility to configure different alarm priorities.

**[SSS_I3S_GEN_FUN_8]** The I3S Server application shall be configured in order to associate a sound to a specific alarm priority.

#### 4.2.1.3 I3S alarm management requirements

The following requirement describes the I3S system behaviour required in case of alarm notification.

**[SSS_I3S_GEN_FUN_9]** The I3S Server and Client applications shall give the possibility to acknowledge an alarm and, in case, to associate a comment.

#### 4.2.1.4 I3S system requirements

Taking into account the I3S system is composed of a client (responsible for the displaying of the MMI) and of a server (that is in charge of the business services that support the MMI), the following requirements mainly concern the general functionalities that the I3S application (client and server) need to provide to the operator in order to monitor and control the whole system.

**[SSS_I3S_GEN_FUN_10]** The I3S Client application shall provide in the *Login Page* the facility to access to the MMI through *username* and *password*.

**[SSS_I3S_GEN_FUN_11]** The I3S Client application shall provide in a specific area of the MMI (*Navigation Area*) the facility to navigate through all the areas of the MMI.

**[SSS_I3S_GEN_FUN_12]** The I3S Client application shall display in a specific area of the MMI (*Map Area*) the graphic map related to the location selected in the *Navigation Area*.

**[SSS_I3S_GEN_FUN_13]** The I3S Client application shall display on each *Map Area* all the *CCTV Camera Objects* present in the location selected in the *Navigation Area*.

**[SSS_I3S_GEN_FUN_14]** The I3S Client application shall display in a specific area of the MMI (*Control Area*) the *CCTV Control Panel*.

**[SSS_I3S_GEN_FUN_15]** The I3S Client application shall display in a specific area of the MMI (*Alarm Area*) the list of the alarms occurred.

**[SSS_I3S_GEN_FUN_16]** The I3S Client application shall provide in a specific area of the MMI (*Alarm Area or Historical Alarm Area*) the facility to replay the audio or video streams corresponding to the alarms occurred.

**[SSS_I3S_GEN_FUN_17]** The I3S Client application shall display in a specific area of the MMI (*Historical Alarm Area*) the list of the historical alarms of the system.

**[SSS_I3S_GEN_FUN_18]** The I3S Client application shall provide in a specific area of the MMI (*Report Area*) the facility to view/print the reports produced by the system.

**[SSS_I3S_GEN_FUN_19]** The I3S Client application shall provide the facility to launch external applications (on demand or automatically).

**[SSS_I3S_GEN_FUN_20]** The I3S Client application shall provide in a specific area of the MMI (*AVAS Configuration Area*) the facility to configure each module present in the Audio-Video Analysis System.

**[SSS_I3S_GEN_FUN_21]** The I3S Client application shall provide in a specific area of the MMI (*AVAS Control Area*) the facility to control (start/stop) each module present in the Audio-Video Analysis System.

**[SSS_I3S_GEN_FUN_22]** The WWM application shall display in a specific virtual monitor of the VW the information generated by each module present in the Audio-Video Analysis System.

**[SSS_I3S_GEN_FUN_23]** The I3S Client application shall provide in a specific area of the MMI (*MTAS Configuration Area*) the facility to configure the Mid-Long Term Analysis System.

**[SSS_I3S_GEN_FUN_24]** The I3S Client application shall provide in a specific area of the MMI (*MTAS Control Area*) the facility to control (start/stop) the Mid-Long Term Analysis System.

**[SSS_I3S_GEN_FUN_25]** The I3S Client application shall display in a specific area of the MMI (*MTAS Result Area*) the information generated by the Mid-Long Term Analysis System.

**[SSS_I3S_GEN_FUN_26]** The I3S Client application shall provide in a specific area of the MMI (*VWM Configuration Area*) the facility to configure the modules present in AVAS, which are related to the VWM stream selection.

**[SSS_I3S_GEN_FUN_27]** The I3S Client application shall provide in a specific area of the MMI (*VWM Control Area*) the facility to control (start/stop) the modules present in AVAS, which are related to the VWM stream selection.

**[SSS_I3S_GEN_FUN_28]** The I3S Server and Client applications shall provide the facility to save the current AVAS, MTAS or VWM configuration in terms of modules configuration and selected audio/video streams.

**[SSS_I3S_VWM_FUN_29]** The I3S Server and Client applications shall provide the facility to retrieve a previously stored AVAS, MTAS or VWM configuration.

#### 4.2.1.5 CCTV Requirements

##### 4.2.1.5.1 Cameras/Virtual monitors status

These requirements describe the functionalities that should be provided by the system to remotely monitor the status of each CCTV device.

**[SSS_I3S_GEN_FUN_30]** The I3S Server application shall monitor the availability of each configured camera.

**[SSS_I3S_GEN_FUN_31]** The I3S Server and Client applications shall represent the connection status of each configured camera on the related *CCTV Camera Object.*

**[SSS_I3S_GEN_FUN_32]** The I3S Server application shall monitor the availability of each configured virtual monitor of the Video Wall Manager application.

**[SSS_I3S_GEN_FUN_33]** The I3S Server and Client applications shall represent the status of each configured virtual monitor on the related *Virtual Monitor Object* present in the *CCTV Control Panel*.

#### 4.2.1.5.2 Video wall configuration

These requirements describe the functionalities that should be provided by the system to manage the Video Wall layout and configuration.

**[SSS_VWM_GEN_FUN_34]** The Video Wall and the Video Wall Manager application shall be configurable as a square matrix of video windows that shall act as virtual monitors.

**[SSS_VWM_GEN_FUN_35]** The Video Wall and the Video Wall Manager application shall be able to display one or more streams at the same time and up to a maximum of 16.

**[SSS_I3S_VWM_FUN_36]** The I3S Server and Client applications shall provide the facility to configure the Video Wall Manager application in order to change the layout of the Video Wall application (1x1, 2x2, 3x3, 4x4 virtual monitors).

**[SSS_I3S_VWM_FUN_37]** The I3S Server and Client applications shall provide the facility to request the current layout configuration of the Video Wall Manager application.

**[SSS_I3S_VWM_FUN_38]** The I3S Server and Client applications shall provide the facility to save the current Video Wall Manager application configuration in terms of layout and cameras/video sequences displayed.

**[SSS_I3S_VWM_FUN_39]** The I3S Server and Client applications shall provide the facility to retrieve a previously stored Video Wall Manager application configuration.

#### 4.2.1.5.3 Live streaming on video wall

These requirements describe the functionalities that should be provided by the system to manage the visualization of live video streams coming from the CCTV devices on the Video Wall.

**[SSS_I3S_VWM_FUN_40]** The I3S Server and Client applications shall provide the facility to display a live stream coming from a camera to a virtual monitor of the Video Wall Manager application.

**[SSS_I3S_VWM_FUN_41]** The I3S Server and Client applications shall provide the facility to display a cycle of live images (video sequence) coming from various cameras to a virtual monitor of the Video Wall Manager application.

**[SSS_I3S_VWM_FUN_42]** The I3S Server and Client applications shall provide the facility to pause a cycle of live images (video sequence) coming from various cameras to a virtual monitor of the Video Wall Manager application. Hence, the live images of the currently active camera shall be displayed from that moment on.

**[SSS_I3S_VWM_FUN_43]** The I3S a Server and Client applications shall provide the facility to resume from pause the cycle of live images (video sequence) coming from various cameras to a virtual monitor of the Video Wall Manager application.

**[SSS_I3S_VWM_FUN_44]** The I3S Server and Client applications shall provide the facility to stop the visualization of a live stream coming from a camera to a virtual monitor of the Video Wall Manager application.

**[SSS_I3S_VWM_FUN_45]** The I3S Server and Client applications shall provide the facility to stop the visualization of a cycle of live images (video sequence) coming from various cameras to a virtual monitor of the Video Wall Manager application.

**[SSS_I3S_GEN_FUN_46]** The I3S Server and Client applications shall provide the facility to store a video sequence configuration.

##### 4.2.1.5.4 Recorded video on video wall

These requirements describe the functionalities that should be provided by the system to manage the visualization of recorded video streams coming from the Audio-Video Recording and Metadata Indexing Services on the Video Wall.

**[SSS_I3S_VWM_FUN_47]** The I3S Server and Client applications shall provide the facility to reproduce a recorded video coming from the **AVRS** application to a virtual monitor of the Video Wall Manager application.

**[SSS_I3S_VWM_FUN_48]** The I3S Server and Client applications shall provide the facility to set the start and stop reproduction parameters (date/time) of the recorded video to be displayed on a virtual monitor of the Video Wall Manager application.

##### 4.2.1.5.5 Video analysis alarms

These requirements describe how the system should react to alarms coming from the Audio-Video Analysis Service in terms of automatic actions and functionalities provided to the operators.

**[SSS_I3S_AVAS_FUN_49]** The I3S Server application shall be able to receive events from the **AVAS** application. This kind of events is called *Video Analysis Alarms*.

**[SSS_I3S_VWM_FUN_50]** The I3S Server application could be configured in order to react to any *Video Analysis Alarm*. In particular, it should be able to (i) display the related live stream(s) associated to an alarm on a virtual monitor of the Video Wall application (ii) display and replay the recorded streams associated to this event (e.g. to validate the event or better understand what happen) (iii) overlay information related to the event on top of the video stream (e.g. the image location where an event has been identified).

**[SSS_I3S_AVAS_FUN_51]** The I3S Server application shall be able to manage received events that refers to previously sent events, e.g. providing an update of the status of an event: for instance, if a left luggage has been identified, then if later on the same luggage is handled again by a person, the status of the luggage in the I3S system should be updated.

**[SSS_I3S_GEN_FUN_52]** The I3S Server and Client applications shall provide the facility to access the list of *Video Analysis Alarms* related to any configured camera, and to access the list of alarms of one specific type only.

**[SSS_I3S_VWM_FUN_53]** The I3S Server application shall provide the facility to reproduce the video corresponding to any *Video Analysis Alarm* using as input the related stored *Video Analysis Alarm*.

#### 4.2.1.6 Audio Requirements

##### 4.2.1.6.1 Microphones Status

These requirements describe the functionalities that should be provided by the system to remotely monitor the status of each microphone.

**[SSS_I3S_GEN_FUN_54]** The I3S Server application shall monitor the availability of each configured microphones.

**[SSS_I3S_GEN_FUN_55]** The I3S Server and Client applications shall represent the connection status of each configured microphones on the related *Audio-CCTV Microphone Object.*

**[SSS_I3S_GEN_FUN_56]** The I3S Server application shall monitor the availability of each configured microphones.

**[SSS_I3S_GEN_FUN_57]** The I3S Server and Client applications shall represent the status of each configured microphones on Video-CCTV Control Panel.

##### 4.2.1.6.2 Audio sensors Configuration management

These requirements describe the functionalities that should be provided by the system to manage the audio and microphones configuration.

**[SSS_VWM_GEN_FUN_58]** The surveillance system shall include loudspeakers (at least one) that shall act as audio player.

**[SSS_VWM_GEN_FUN_59]** The surveillance system shall be able to play one or more audio streams (signal summation) at the same time.

**[SSS_I3S_VWM_FUN_60]** The I3S Server and Client applications shall provide the facility to configure the surveillance system in order to change the selected microphones.

**[SSS_I3S_VWM_FUN_61]** The I3S Server and Client applications shall provide the facility to request the current selected microphones configuration.

**[SSS_I3S_VWM_FUN_62]** The I3S Server and Client applications shall provide the facility to save the current selected microphones configuration.

**[SSS_I3S_VWM_FUN_63]** The I3S Server and Client applications shall provide the facility to retrieve a previously stored selected microphones configuration.

##### 4.2.1.6.3 Audio Live Streaming

These requirements describe the functionalities that should be provided by the system to manage the listening of live audio streams coming from each microphone.

**[SSS_I3S_VWM_FUN_64]** The I3S Server and Client applications shall provide the facility to play a live stream coming from a microphone to a loudspeaker.

**[SSS_I3S_VWM_FUN_65]** The I3S Server and Client applications shall provide the facility to play a live streams coming from microphones to loudspeakers (at least one).

**[SSS_I3S_VWM_FUN_66]** The I3S Server and Client applications shall provide the facility to pause a live stream coming from a microphone to a loudspeaker.

**[SSS_I3S_VWM_FUN_67]** The I3S Server and Client applications shall provide the facility to pause a live streams coming from microphones to loudspeakers (at least one).

**[SSS_I3S_VWM_FUN_68]** The I3S Server and Client applications shall provide the facility to resume from pause a live stream coming from a microphone to a loudspeaker.

**[SSS_I3S_VWM_FUN_69]** The I3S Server and Client applications shall provide the facility to resume from pause a live streams coming from microphones to loudspeakers (at least one).

**[SSS_I3S_GEN_FUN_70]** The I3S Server and Client applications shall provide the facility to store audio live stream(s) sequence configuration (number of streams and related microphones).

##### 4.2.1.6.4 Recorded audio streams

These requirements describe the functionalities that should be provided by the system to manage the listening of recorded audio streams coming from the Audio-Video Recording and Metadata Indexing Services.

**[SSS_I3S_VWM_FUN_71]** The I3S Server and Client applications shall provide the facility to reproduce a recorded audio coming from the **AVRS** application to loudspeakers (at least one).

**[SSS_I3S_VWM_FUN_72]** The I3S Server and Client applications shall provide the facility to set the start and stop reproduction parameters (date/time) of the recorded audio to be played on loudspeakers (at least one).

##### 4.2.1.6.5 Audio Analysis Alarms

These requirements describe how the system should react to alarms coming from the Audio-Video Analysis Service in terms of automatic actions and functionalities provided to the operators.

**[SSS_I3S_AVAS_FUN_73]** The I3S Server application shall be able to receive events from the **AVAS** application. This kind of events is called *Audio Analysis Alarms*.

**[SSS_I3S_VWM_FUN_74]** The I3S Server application could be configured in order to react to any *Audio Analysis Alarm* playing the related live stream on a loudspeaker.

**[SSS_I3S_GEN_FUN_75]** The I3S Server and Client applications shall provide the facility to access the list of *Audio Analysis Alarms* related to any configured microphones.

**[SSS_I3S_VWM_FUN_76]** The I3S Server application shall provide the facility to reproduce the audio stream corresponding to any *Audio Analysis Alarm* using as input the related stored *Audio Analysis Alarm*.

### 4.2.2 Audio-Video analysis requirements

This section describes the functionalities that the Audio-Video Analysis System application (AVAS) shall provide: The AVAS application is based on the following components:

- ***AVAS-API***: Framework that shall be used to develop the AVAS application (THALIT). The AVAS-API framework shall provide functionalities that can be shared among all the AVAS Modules.
- ***AVAS Modules***: Processing modules and audio-video algorithms (INRIA, IDIAP, MULT, TCF).

  In particular the main sets of AVAS Modules currently foreseen are:

  - AVAS Modules set 1 – Autonomous sensor selection modules
  - AVAS Modules set 2 – Event detection and recognition modules
  - AVAS Modules set 3 – Situational reporting modules

#### 4.2.2.1 Installation requirements

The following requirements describe the Audio-Video Analysis Service system constraints for installation on demonstrators.

**[SSS_AVAS_GEN_HW_77]** The AVAS application shall be installed on specific server(s) (**SRV2,** see section 4.4.1.1).

**[SSS_AVAS_GEN_FUN_78]** The AVAS application could require a particular calibration procedure, to be done on each configured camera, in order to properly work.

#### 4.2.2.2 System requirements

##### 4.2.2.2.1 AVAS Modules set 1 requirements – Autonomous selection

The following requirements describe the system capabilities required in order to manage AVAS modules for autonomous sensor selection, to display its output information (e.g. to automatically switch/playback audio/video streams depending on analysis results).

**[SSS_AVAS_I3S_FUN_79]** The I3S Client application shall provide in a specific area of the MMI (*AVAS Configuration Area*) the facility to select a set of cameras and microphones to be processed by the by "slaves module(s)" present in the Autonomous Sensor Selection Module (ASSM) in AVAS, and listed in the *AVAS Configuration Area*.

**[SSS_AVAS_I3S_FUN_80]** The I3S Client shall be able to associate multiple cameras and microphones to a single "slave module".

**[SSS_AVAS_I3S_FUN_81]** The I3S Client application shall provide in a specific area of the MMI (*AVAS Control Area*) to control (start/stop) the selected sensors (cameras and microphones) and ASSM.

**[SSS_AVAS_I3S_FUN_82]** The I3S Server application shall address and acquire the selected audio-video streams, and provide them to the slaves ASSM in AVAS. In parallel, the I3S Server application shall activate the "master ASSM" in AVAS.

**[SSS_AVAS_I3S_FUN_83]** The I3S Server application shall monitor the availability of meta-data coming from the selected slaves ASSM, and send them to the master ASSM. The I3S Server application shall monitor the availability of meta-data coming from the master ASSM, and send them to the *VWM* for controlling the *VWM* itself.
**[SSS_AVAS_I3S_FUN_84]** The I3S Client application shall display in a specific area of the MMI (*AVAS Result Area*) the information generated by the master ASSM in AVAS. This could be abnormality level for

each selected stream to be displayed as gauges. The I3S Client should group streams when they are redundant and have similar abnormality measures, e.g. two cameras providing two viewpoints of the same area.

#### 4.2.2.2.2 AVAS Modules set 2 requirements – Event detection

The following requirements describe the system capabilities required in order to manage the AVAS module for event detection, to display its output and the generated alarms.

**[SSS_AVAS_I3S_FUN_85]** The I3S Client application shall provide in a specific area of the MMI (*AVAS Configuration Area*) the facility to select one or multiple camera(s) to be processed by a specific module present in the Audio-Video Analysis System, and listed in the *AVAS Configuration Area*.

**[SSS_AVAS_I3S_FUN_86]** The I3S Client application shall provide in a specific area of the MMI (*AVAS Control Area*) the facility to control (start/stop) the selected cameras/modules.

**[SSS_AVAS_I3S_FUN_87]** The I3S Server application shall address and acquire the selected video stream(s), and provide it(them) to the corresponding module in the Audio-Video Analysis System.

**[SSS_AVAS_GEN_FUN_88]** The Event detection modules in AVAS shall analyse provided video stream(s) and provide the metadata related to any recognised scenario (see section 5.3).

**[SSS_AVAS_I3S_FUN_89]** The VWM application shall display in a specific virtual monitor of the VW the information generated by the selected camera/module in the Audio-Video Analysis System. This could be graphics (e.g. bounding boxes) or text information, overlaid on the video stream displayed in a synchronised way. At the same time the related alarm shall be listed in the Alarm Banner Area of the I3S Client.

#### 4.2.2.2.3 AVAS Modules set 3 requirements – Situational reporting

This section describes the functionalities that shall be provided to the situational reporting modules. Roughly speaking, it is a group of individual "event detection" modules running in parallel, and which outcomes are used as single output to a dedicated map-based MMI. Its functionality shall be allocated to system blocks as follows[1].

**[SSS_AVAS_I3S_FUN_90]** The I3S Client application shall provide in a specific area of the MMI (AVAS *Configuration Area*) the facility to select a set of cameras to be processed (from one camera to e.g. all cameras of a level) by module(s) present in the Audio-Video Analysis System.

**[SSS_AVAS_I3S_FUN_91]** The I3S Client application shall provide in a specific area of the MMI (AVAS *Control Area*) the facility to control (start/stop) the selected cameras/modules.

**[SSS_AVAS_I3S_FUN_92]** The I3S Server application shall address and acquire the selected video streams, and provide them to the corresponding modules in the Audio-Video Analysis System.

**[SSS_AVAS_I3S_FUN_93]** The I3S Client application shall display in the Map Area of the MMI the information generated by the selected cameras/modules in the Audio-Video Analysis System. This area shall contain graphic maps where metadata information could be displayed dynamically: preconfigured zones of each graphic map could change their appearance (background colour) according to incoming metadata information.

[1] Note that some requirements are already partially covered by the I3S

### 4.2.3 Audio-video recording and metadata indexing requirements

This section describes the functionalities that the Audio-Video Recording And Metadata Indexing Service application shall provide.

#### 4.2.3.1 Installation requirements

The following requirements describe the Audio-Video Recording And Metadata Indexing Service system constraints for installation on demonstrators.

**[SSS_AVRS_GEN_HW_94]** The AVRS application shall be installed on a specific server (**SRV4**, see section 4.4.1.1).

#### 4.2.3.2 System requirements

The following requirements describe the Audio-Video Recording And Metadata Indexing Service capabilities required to record video/audio/metadata streams and to provide recorded data.

**[SSS_AVRS_GEN_FUN_95]** The AVRS application shall continuously record live streams coming from all the configured cameras.

**[SSS_AVRS_AVAS_FUN_96]** The AVRS application shall be able to receive metadata information from the **AVAS** application.

**[SSS_AVRS_AVAS_FUN_97]** The AVRS application shall store the metadata information received from the **AVAS** application.

**[SSS_AVRS_GEN_FUN_98]** The AVRS application shall be able to provide the recorded audio-video and metadata information when requested.

**[SSS_AVRS_GEN_FUN_99]** The AVRS application shall automatically delete recorded data once they are no longer needed for analysis. This action shall be performed after a configurable predefined time.

### 4.2.4 Mid-long term analysis & reporting requirements

This section describes the functionalities that the Metadata Management & Reporting shall provide. Basically, it is a collection of modules that will analyse data on the mid and long-term basis to allow to the end-user to 1) find relationships between the behaviours observed at different time periods, and 2) find relationships between the behaviours observed at different locations. For these tasks, two dedicated databases will be set up: the mid-term activity DB and the long-term activity DB.

The dedicated modules for mid-long term analysis are the following:

**[SSS_MTAS_GEN_FUN_100]** **Mid-term indexing sub-module**. The Mid-long term analysis & Reporting application shall monitor the availability of new audio-video meta-data (only for selected monitored spaces) and after analysis update the mid-term DB. The results written to this DB consists of a series of statistics on the number of persons and events observed within the analysis period plus the data characterising the persons and events themselves.

**[SSS_MTAS_GEN_FUN_101]** **Context building sub-module**. The Mid-long term analysis & Reporting application will analyse the mid-term statistics on person position and displacement (trajectory analysis) to update a model of mid-term long-term space occupancy. The result will be written into the long-term DB.

**[SSS_MTAS_GEN_FUN_102]** **Activity clustering sub-module**. The Mid-long term analysis & Reporting application will analyse trajectories, events and other features stored in the mid-term DB to find activity patterns not necessarily previously defined. The result will be written into the long-term DB.

#### 4.2.4.1 Installation requirements

The following requirements describe the Mid-Long Term Analysis (Metadata Management) & Reporting Service system constraints for installation on demonstrators.

**[SSS_MTAS_GEN_FUN_103]** The Metadata Management & Reporting application shall be installed on a specific server (**SRV3**, see section 4.4.1.1).

**[SSS_MTAS_GEN_FUN_104]** The Mid-long term analysis & Reporting client application shall be installed on a specific server (**SRV3**, see section 4.4.1.1).

**[SSS_MTAS_GEN_FUN_105]** The Mid-long term analysis & Reporting dedicated databases shall be installed on a specific server (**SRV6**, see section 4.4.1.1).

#### 4.2.4.2 System requirements

The following requirement describes the Mid-Long Term Analysis (Metadata Management) & Reporting Service system Database constraints.

**[SSS_MTAS_GEN_FUN_106]** The dedicated Mid-long term databases shall have PostgreSQL as standard configuration (see section 4.10.2).

### 4.2.5 Streaming service requirements

This section describes the audio and video streaming functionalities that shall be provided.

The Streaming Service is a functionality that shall be allocated to system blocks as follows:

- Streaming Service for the audio and video streams from **Field Sensors** to **AVAS**, **AVRS** and **VWM** applications: Codecs provided by GTT and RATP for the respective demonstrators:
    - **VCGTT** (Video-Codec-GTT, see section 4.3.1.1).
    - **VCRATP** (Video-Codec-RATP, see section 4.3.1.2).
    - **ACGTT** (Audio-Codec-GTT, see section 4.3.1.3).
    - **ACRATP** (Audio-Codec-RATP, see section 4.3.1.4).
- Streaming Service for the audio and video streams from **AVAS** application to **AVRS** and **VWM** applications: **AVAS** application.
- Streaming Service for the audio and video streams from **AVRS** application to **VWM** and **AVAS** applications (depending on functionalities): **AVRS** application.

#### 4.2.5.1 System requirements

##### 4.2.5.1.1 Video-codec-GTT streaming requirements

The following requirements describe the general video field devices constraints for the deployment of the demonstrator in Turin (GTT).

**[SSS_VCGTT_GEN_FUN_107]** The Video-Codec-GTT shall manage the live streams from the configured cameras in use in Turin demonstrator site and shall be capable to provide them to the AVAS, the AVRS and the VWM applications.

**[SSS_VCGTT_GEN_FUN_108]** The Video-Codec-GTT shall be capable to synchronize with a centralized time-clock reference and to embed time-stamps into streams, following the used standard, originated and sent to the AVAS, the AVRS and the VWM applications.

**[SSS_VCGTT_GEN_FUN_109]** The Video-Codec-GTT shall be configured with a static IP address.

##### 4.2.5.1.2 Video-codec-RATP streaming requirements

The following requirements describe the general video field devices constraints for the deployment of the demonstrator in Paris (RATP).

**[SSS_VCRATP_GEN_FUN_110]** The Video-Codec-RATP shall manage the live streams from the configured cameras in use in Paris demonstrator site and shall be capable to provide them to the AVAS, the AVRS and the VWM applications.

**[SSS_VCRATP_GEN_FUN_111]** The Video-Codec-RATP shall be capable to synchronize with a centralized time-clock reference and to embed time-stamps into streams, following the used standard, originated and sent to the AVAS, the AVRS and the VWM applications.

**[SSS_VCRATP_GEN_FUN_112]** The Video-Codec-RATP shall be configured with a static IP address.

#### 4.2.5.1.3 Audio-codec-GTT streaming requirements

The following requirements describe the general audio field devices constraints for the deployment of the demonstrator in Turin (GTT).

**[SSS_ACGTT_GEN_FUN_113]** The Audio-Codec-GTT shall manage the live streams from the microphones in use in Turin demonstrator site and shall be capable to provide them to the AVAS and the AVRS applications.

**[SSS_ACGTT_GEN_FUN_114]** The Audio-Codec-GTT shall be capable to synchronize with a centralized time-clock reference and to embed time-stamps into streams, following the used standard, originated and sent to the AVAS, the AVRS and the VWM applications.

**[SSS_ACGTT_GEN_FUN_115]** The Audio-Codec-GTT shall be configured with a static IP address.

#### 4.2.5.1.4 Audio-codec-RATP streaming requirements

The following requirements describe the general audio field devices constraints for the deployment of the demonstrator in Paris (RATP).

**[SSS_ACRATP_GEN_FUN_116]** The Audio-Codec-RATP shall manage the live streams from the microphones in use in Paris demonstrator site and shall be capable to provide them to the AVAS and the AVRS using applications.

#### 4.2.5.1.5 AVRS streaming requirements

The following requirements describe the interface constraints between the audio/video field devices and the Audio-Video Recording and Metadata Indexing Services module.

**[SSS_AVRS_VCGTT_FUN_117]** The AVRS application shall be capable to receive video streams from the Video-Codec-GTT.

**[SSS_AVRS_VCRATP_FUN_118]** The AVRS application shall be capable to receive video streams from the Video-Codec-RATP.

**[SSS_AVRS_ACGTT_FUN_119]** The AVRS application shall be capable to receive audio streams from the Audio-Codec-GTT.

**[SSS_AVRS_ACRATP_FUN_120]** The AVRS application shall be capable to receive audio streams from the Audio-Codec-RATP.

**[SSS_AVRS_GEN_FUN_121]** The AVRS shall be capable to stream, when requested, the recorded audio-video + metadata information to the VWM and the AVAS applications.

#### 4.2.5.1.6 VWM streaming requirements

The following requirements describe the interface constraints between the audio/video field devices and the Video Wall Manager module.

**[SSS_VWM_VCGTT_FUN_122]** The VWM application shall be capable to receive the video streams from the Video-Codec-GTT.

**[SSS_VWM_VCRATP_FUN_123]** The VWM application shall be capable to receive the video streams from the Video-Codec-RATP.

**[SSS_VWM_AVRS_FUN_124]** The VWM application shall be capable to receive the video streams from the AVRS application.

**[SSS_VWM_AVAS_FUN_125]** The VWM application shall be capable to receive the video streams from the AVAS application.

#### 4.2.5.1.7 AVAS streaming requirements

The following requirements describe the interface constraints between the audio/video field devices and the Audio-Video Analysis Service module.

**[SSS_AVAS_VCGTT_FUN_126]** The AVAS application shall be capable to receive video streams from the Video-Codec-GTT.

**[SSS_AVAS_VCRATP_FUN_127]** The AVAS application shall be capable to receive video streams from the Video-Codec-RATP.

**[SSS_AVAS_ACGTT_FUN_128]** The AVAS application shall be capable to receive audio streams from the Audio-Codec-GTT.

**[SSS_AVAS_ACRATP_FUN_129]** The AVAS application shall be capable to receive audio streams from the Audio-Codec-RATP.

**[SSS_AVAS_AVRS_FUN_130]** The AVAS application shall be capable to receive the audio and video streams from the AVRS application.

## 4.3 System external interface requirements

This section describes the external systems that the VANAHEIM SYSTEM shall interface. In VANAHEIM context, it mainly concerns the interfacing of the VS with GTT and RATP current installation. Each equipment type is thus identified by vendor, protocol version, hardware version (optional) and, in case, the relative referenced projects.

### 4.3.1 CCTV external interface requirements

#### 4.3.1.1 Funkwerk encoder VNS 102 Interface (Video-codec-GTT)

The following requirements describe the constraints of the specific video encoder that will be used in the demonstrator in Turin.

**[SSS_VCGTT_GEN_INT_131]** The encoder shall compress the video coming from the configured cameras in **MPEG-4 ASP** format.

**[SSS_VCGTT_GEN_INT_132]** The encoder shall stream data coming from the configured cameras using the standard **RTP** (Real-time Transport Protocol) on **UDP** (User Datagram Protocol).

**[SSS_VCGTT_GEN_INT_133]** The encoder is able to transmit video streams at two frame rates: 6 fps and 25 fps.

**[SSS_VCGTT_GEN_INT_134]** The encoder shall be able to stream the video to multiple clients (up to 10) at the same time with no drawbacks on video quality and frame-rate.

**[SSS_VCGTT_GEN_INT_135]** The encoder shall provide an interface to read current encoding, streaming and addressing parameters.

#### 4.3.1.2 RATP video interface (Video-codec-RATP)

The following requirements describe the constraints of the specific video encoder that will be used in the demonstrator in Paris.

**[SSS_VCRATP_GEN_INT_136]** The encoder shall compress the video coming from the configured cameras in **MPEG-4 Part 10 (H.264)** format.

**[SSS_VCRATP_GEN_INT_137]** The encoder shall stream data coming from the configured cameras using the standard **RTP** (Real-time Transport Protocol) on **UDP** (User Datagram Protocol).

**[SSS_VCRATP_GEN_INT_138]** The encoder shall be able to stream the video to multiple clients (up to 10) at the same time with no drawbacks on video quality and frame-rate.

**[SSS_VCRATP_GEN_INT_139]** The encoder shall provide an interface to read current encoding, streaming and addressing parameters.

#### 4.3.1.3 MX9000 interface (Audio-codec-GTT)

The following requirements describe the constraints of the specific audio encoder that will be used in the demonstrator in Turin.

**[SSS_ACGTT_GEN_INT_140]** The encoder shall compress and stream the audio using standard protocols.

**[SSS_ACGTT_GEN_INT_141]** The encoder shall stream data coming from the configured cameras using the standard **RTP** (Real-time Transport Protocol) on **UDP** (User Datagram Protocol).

#### 4.3.1.4 RATP audio interface (Audio-codec-RATP)

The following requirements describe the constraints of the specific audio encoder that will be used in the demonstrator in Paris.

**[SSS_ACRATP_GEN_INT_142]** The encoder shall compress and stream the audio using standard protocols.

## 4.4 Computer resource requirements

### 4.4.1 Computer hardware requirements

#### 4.4.1.1 Installation requirements

The following requirements describe the global system constraints, in term of modules allocation on different servers, for the installation on demonstrators.

**[SSS_I3S_GEN_HW_143]** On **SRV1** machine shall be installed the **I3S Server** application.

**[SSS_GEN_GEN_HW_144]** On several **SRV2** machines shall be installed the **Audio-Video Analysis** applications.

**[SSS_GEN_GEN_HW_145]** On **SRV3** machine shall be installed the **Mid-Long Term Analysis & Reporting** applications.

**[SSS_AVRS_GEN_HW_146]** On **SRV4** machine shall be installed the **Audio-Video Recording And Metadata Indexing**.

**[SSS_VWM_GEN_HW_147]** On **SRV5** machine shall be installed the **Video Wall** application.

**[SSS_MTAS_GEN_HW_148]** On **SRV6** machine shall be installed the **dedicated mid-long term Databases**.

**[SSS_I3S_GEN_HW_149]** On **WS1** machine shall be installed the **I3S Client** application.

#### 4.4.1.2 Hardware requirements

The following requirements describe the reference hardware to be taken into account for the installation on demonstrators.

**[SSS_I3S_GEN_HW_150]** A typical configuration for the hardware that shall host the **I3S Server** application (**SRV1**) is: "Intel Quad Core 2.66 GHz CPU, 4 GB of RAM, 4 GB of free disk space and 100/1000 Ethernet interface".

**[SSS_I3S_GEN_HW_151]** A typical configuration for the hardware that shall host the **I3S Client** application (**WS1**) is: "Intel Dual Core 2.33 GHz CPU, 4 GB of RAM, 4 GB of free disk space and 100/1000 Ethernet interface".

#### 4.4.1.3 Computer hardware resource utilization requirements

The following requirement describes the global system performance constraints.

**[SSS_I3S_GEN_HW_152]** The resource utilization of the I3S application shall be strictly dependent on the device to integrate and the overall system architecture.

### 4.4.2 Computer software requirements

The following requirement describes the software constraints to be taken into account during the servers/workstations installation on demonstrators.

**[SSS_I3S_GEN_GEN_153]** The I3S application shall use a Microsoft SQL Server 2003 as configuration and reporting database.

**[SSS_I3S_GEN_GEN_154]** The dedicated Mid-long term Databases shall use PostgreSQL 8.4 as standard configuration.

### 4.5 System environment requirements

The following requirements describe the Operating System constraints to be taken into account during the servers/workstations installation on demonstrators.

**[SSS_I3S_GEN_GEN_155]** The I3S Server application shall be able to run on Windows Server 2003 R2 operating systems.

**[SSS_I3S_GEN_GEN_156]** The I3S Client application shall be able to run on Windows XP Sp2 operating systems.

**[SSS_AVAS_GEN_GEN_157]** The AVAS application and AVAS-API shall be able to run on Windows and/or Linux (Ubuntu, Fedora, etc.) operating systems.

**[SSS_VWM_GEN_GEN_158]** The VWM application shall be able to run on Windows XP Sp2 operating systems.

**[SSS_AVRS_GEN_GEN_159]** The AVRS application shall be able to run on Linux operating systems.

**[SSS_MTAS_GEN_GEN_160]** The Mid-long term analysis & Reporting applications and dedicated databases shall run on Linux (Fedora) operating system.

## 4.6 Safety Requirements

The following requirements describe the safety requirements to be taken into account for the system installation and operation/demonstration.

**[SSS_I3S_GEN_GEN_161]** The I3S Server application must not affect the performance with already installed systems that it interfaces, in particular:

**[SSS_I3S_VCGTT_GEN_162]** The I3S Server application must not affect the performance with already installed systems that it interfaces GTT Video Codec.

**[SSS_I3S_ACGTT_GEN_163]** The I3S Server application must not affect the performance with already installed systems that it interfaces GTT Audio Codec.

**[SSS_I3S_VCRATP_GEN_164]** The I3S Server application must not affect the performance with already installed systems that it interfaces RATP Video Codec.

**[SSS_I3S_ACRATP_GEN_165]** The I3S Server application must not affect the performance with already installed systems that it interfaces RATP Audio Codec.

## 4.7 Security and privacy requirements

The following requirements describe the security and privacy requirements to be taken into account for the system operation/demonstration.

**[SSS_AVAS_GEN_GEN_166]** The Audio-Video Analysis System (AVAS) shall be configured by minimising the use of personal data and identification.

**[SSS_MTAS_GEN_GEN_167]** The Mid-Long Term Analysis System (MTAS) shall be configured by minimising the use of personal data and identification.

**[SSS_AVRS_GEN_GEN_168]** Data recording in AVRS must be relevant, complete and not excessive in relation to the purposes for which they are collected or subsequently processed.

**[SSS_MTAS_GEN_GEN_169]** Analysis of recorded data through the Mid-Long Term Analysis System must be relevant, complete and not excessive in relation to the purposes for which they are processed.

# 5 Conclusion

In this deliverable, the requirements needed for the development of the monitoring components foreseen in the VANAHEIM project were reported. Both technical requirements and end-users recommendations have been gathered on respectively the technical aspect of the VANAHEIM architecture management, and on the end-user usage context scenarios envisaged in the project.

In more details, a common format for video and audio scenario specification has been proposed, and end-users scenarios have been presented and specified with respect to this template. Various audio/video scenarios, which should offer the ability to monitor and exploit audio/video surveillance data, have been specified for the three specific VANAHEIM application areas (autonomous sensor selection, real-time applications, and long-term analysis). While video and audio analytics could be used to alert operators or investigators of an event occurrence, they should also be exploited in a wide range of applications. For example, the identification of common patterns such as traffic flow of people (long-term analysis) should be an interesting way to exploit such CCTV infrastructure. Video analytics should also create new opportunities for the use of video surveillance by non-traditional users, who are not generally focused on safety or security; for example, these users may be in other parts of an organization, such as in marketing in the retail industry or manufacturing and production control; for example, administrators should be able to use these analytics to trigger automated alarms and other responses.

The surveillance architecture, that will be deployed in VANAHEIM and will allow the audio/video analytics to be demonstrated, has then been specified in order to allow the different components of the system (standard surveillance platform components and innovative audio/video analysis blocks) to interact with each other, and also with third party sensors and transport operators available on-site (field sensors already deployed). Main system capability requirements were formulated in terms of - MMI, supervisor and video wall requirements, - audio-video analysis requirements, - audio-video recording and metadata indexing requirements, - mid-long term analysis & reporting requirements, and - streaming service requirements .

To conclude, reported requirements will be used for three purposes: to define the behavioural models required to perform the desired recognition tasks (T2.2), to specify the systems and components technical layout (T2.3), and to delineate the trials scenarios for the evaluation stage (T3.1).