
# 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: June 30, 2012**
**Actual delivery date: October 11, 2012**

## Deliverable D2.4
## End user requirement and system objectives (version 2)

---

## D2.4

**Version: 1.0**
**Author: GTT/RATP**
**Contributors: THALIT, MULT, INRIA, TCF, IDIAP**
**Reviewers: MULT, THALIT**
**Dissemination level: PU**
**Related document(s): Deliverable D2.1 (End user requirement and system objectives (version 1))**
**Number of pages: 19**

# Document information

| Ver. | Date | Changes | Author (partic.) |
|---|---|---|---|
| | 30/04/2010 | Deliverable D2.1 (End user requirement and system objectives (version 1)) | |
| 1.0 | 25/09/2012 | Revision of version 1 deliverable | F. Sabourin (RATP) |

| Ver. | Date | Approval/Rejection decision/comments | Author (partic.) |
|---|---|---|---|
| 1.0 | 11/10/2012 | Approved | C. Carincotte (MULT) |
| 1.0 | 11/10/2012 | Approved | A. Grifoni (THALIT) |

# 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 end-user requirements needed for the development of the monitoring components foreseen in the project. In this document we build upon the end-user and technical specifications delineated in Deliverable D2.1 (End user requirement and system objectives (version 1)), and update end-users recommendations and end-user usage context scenarios with respect to the system deployed at GTT in March 2012. This report therefore mainly focuses on end-user monitoring usage expectations.

# Table of contents

# 2 Introduction

This document 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, and the feedbacks coming from the first system deployment at GTT in March 2012, led to the identification of several scenarios of interest which are detailed in this section.

In the next sub-section, is first introduced the general end-user requirements that are applicable to all the categories and the system as a whole. Template used are then presented to describe the different scenarios, then follows detailed requirements of each scenarios by category.

## 2.1 General end-user requirements

### 2.1.1 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 and cope with local regulations
- 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 levels of data access 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 regulation 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.
- The system should use a limited number of servers

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.

### 2.1.2 Scenario template

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

---

**NAME OF THE SCENARIO**

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

**User motivation**
*Brief description of why the user is interested in the scenario.*

**Scenario characterisation**
*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 impacts to other parts of the system?**
*Describes the possible connection with other part of the system, such as relation to other scenarios.*

---

## 2.2 User requirement for autonomous sensor selection

Components developed for autonomous sensors should perform an unsupervised and efficient selection among the huge amount of data streams available in such CCTV systems.

- First, sounds attracting operators' attention to relevant/salient audio streams, or spotting out corresponding microphones or areas, should be developed.
- Secondly, algorithms performing a video sensor selection at the control room level to select streams to display on monitors, based on the video stream content normality/abnormality, should be developed.
- Last, audio-video joint-processing adaptation should be searched in relevant physical configuration, to exploit the complementary nature of video and sound in specific part of the infrastructure (e.g. at the platform where both audio and video data can be used to link video action with sound during a metro arrival/departure).

---

**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 detect the most interesting/relevant/abnormal video streams to display on the videowall and audio streams to play on speakers in the control room.

**User motivation**
In everyday practice, surveillance videowalls frequently show empty scenes, while there are many others cameras looking at scenes in which something (even normal) is happening. 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. Then, as vigilance studies confirm, operators who spend hours "screen gazing" at static scenes tend to become bored and unefficient, and are then likely to miss sensible events.

The main motivation for both RATP and GTT is to make the video/audio monitoring task easier for security operators in control rooms, 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 operations requested by the monitoring task, and increasing the probability to watch/listen to the right streams at the right time.

**Scenario characterisation**

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

**Usual location of the scenario**
The location of cameras and microphones corresponds to the usual requirements for surveillance coverage in all space of underground, 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 done here is working on data stream in terms of normal/abnormal contents, and not on semantic description of scene/stream contents. 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. The system should also "turn off" the screen when the normality gauge result is lower than a certain level.

**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 and speakers 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. The system should also 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 appear to the operator in a way that the operational reason is fully clear. Relevant streams related to different instances of the same abnormal event should be presented in a human-centred logic and keeping in focus 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 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.

---

## 2.3 User requirement for real-time applications on human activity monitoring

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

- 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;
- performing the task of protecting people and places, thus guaranteeing a secure environment
- 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;
- managing and preventing dangerous situations in order to reduce/remove the risk of robbery, theft, bag-snatching and harassment in stations;

Real-time applications should 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. 2.3.1);
- *Group detection* scenario (Sec. 2.3.2);
- *People arguing, entering in conflict* scenario (Sec. 2.3.3);
- *Crowd/Flows of people* scenario (Sec. 2.3.4);
- *Monitoring equipment* scenario (Sec. 2.3.5);
- and *Situational reporting* scenario (Sec. 2.3.6).

Others scenarios can also be defined

- collapse of people (escalators, tracks, terrorism, etc.)
- unauthorized trespassing on tracks (not easy to test on lines with platform screen doors, but realistic for conventional lines)

### 2.3.1 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 more detailed activity analysis involving people to detect discret stealing of luggage due to inattention and diversion.

**User motivation**
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 threat for the station security. In addition, such information can be used to manually or automatically trigger vocal messages inviting passengers to make sure they haven't forgotten anything and to be more vigilant. Preventing people from forgetting their luggage saves interventions from the security staff.

**Scenario characterisation**

**Definition of the scenario**
An item is abandoned in an area monitored by the CCTV network. The operator is alerted if the abandoned item is not removed within few minutes.

**Usual location of the scenario**
While the event can occur in any areas of the station, platforms are the most sensible area where abandoned items may suspend traffic. Elevators and hall are also concerned by this scenario.

**Human behaviour/actions**
The scenario can be defined by a individual (or group) entering the scene with some luggage then leaving it on the floor. From this situation, one possibility is that the person (or group) leaves and forgets the luggage which 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 with nobody noticing it. No specific audio event is concerned, and expected video information related to the event(s): An item 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 consists only in selection of the cameras that should be processed by the sub-system. The operator should then be alerted if an abandoned item is not removed after a 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 a few minutes. The system should output alarms with one or more key locations and key frames (person entering, person dropping bag, other people stealing it, etc.). This information should help the operator in evaluating the alarm.

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

### 2.3.2 Group detection scenario

---

**GROUP DETECTION**

**Scenario description**
This scenario should focus on the identification of groups and analysis of people interaction with each others.

**User motivation**
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 up to vandalism and fighting groups.

**Scenario characterisation**

**Definition of the scenario**
This scenario should focus on the recognition of the (social) interaction between two or more people, 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 scenario, 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?**
Confirm alarm and ask for security staff.

**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 other parts of the system?**
Other system modules shall use this information, such as the general mid-long term analysis module.

---

### 2.3.3 People arguing, entering in conflict scenario

---

**PEOPLE ARGUING, ENTERING IN CONFLICT**

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

**User motivation**
Safety and security in the station are naturally the main motivations for this scenario.

**Scenario characterisation**

**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 premises.

**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 situations (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 other parts of the system?**
Depending of the configuration, it could be envisaged that other modules would automatically be launched such as ‘automatic sensor selection’ module.

---

### 2.3.4 Crowd/Flows of people scenario

---

**CROWD/FLOWS OF PEOPLE**

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

**User motivation**
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.

**Scenario characterisation**

**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 entrances, halls 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 other parts of the system?**
N/A.

---

### 2.3.5 Monitoring equipment scenario

---

**MONITORING EQUIPMENT**

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

**User motivation**
The main motivation for this scenario is related to transit statistics (including fraud) management of the infrastructure and the availability of service according to traffic.

**Scenario characterisation**

**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 compulsory use of the concerned equipment.

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

---

### 2.3.6 Situational reporting scenario

---

**SITUATIONAL REPORTING**

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

**User motivation**
The main motivation for this scenario is the ability to roughly estimate the number and locations of people in the premises, thus allowing to monitor an entire floor of the station 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.

**Scenario characterisation**

**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 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.

---

## 2.4 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?**

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

**User motivation**
Due to the passenger demand growth and related capacity problems, transportation planners and public transport operators are now looking for ways to evaluate their transport 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 use 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 and times.

**Scenario characterisation**

**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 route behaviour, waiting/traveling 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 outcomes of such tools/analysis are not predictable, main areas of analysis should be related to passenger counts and occupancy (density maps, analysis of the use of station space), passenger's speed, paths and route choices, boarding/alighting analysis and waiting times, queuing time… Among the expected outcomes, localization of common loitering areas and/or highly frequented areas, monitoring of the use of stairs in comparison to escalators, 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. Finally, the results should be presented at different levels of complexity depending if the analysis is dedicated to transport planners operations managers, or security agents, or infrastructure designers.

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

---

# 3 Conclusion

In this deliverable, the requirements needed for the development of the monitoring components were reported. End-users recommendations have been gathered on the end-user usage context scenarios envisaged in the project, and refined following the mid-term project integration at GTT.