FP7 VANAHEIM
IP project n°248907
February 2010 – September 2013

# Integrating innovative audio/video analysis in real-world surveillance platforms: VANAHEIM legacy publication

C. Carincotte[1*], A. Forchino[2], J.-M. Odobez[3], F. Bremond[4],
F. Sabourin[5], B. Ravera[6], A. Grifoni[7], K. Grammer[8]

*[1] Multitel asbl, Mons, Belgium, [2] GTT, Turin, Italy, [3] Idiap, Martigny, Switzerland, [4] INRIA, Sophia-Antiplois, France*
*[5] RATP, Paris France, [6] Thales, Colombes, France, [7] Thales Italy, Firenze, Italy, [8] University of Vienna, Vienna, Austria*

## Abstract

This publication presents the outcomes of the four-year VANAHEIM (Video/Audio Networked surveillance system enhAncement through Human-cEntered adaptIve Monitoring) project. We first review the main objectives of the project, and in a second step describe the main developments and individual achievements that were made along these planned objectives. Finally, we describe how they were integrated, deployed and evaluated in two real-world operational CCTV platforms (GTT Torino, RATP Paris).

*Keywords:* FP7 VANAHEIM, video surveillance, audio analysis, human behavior research, real-scale evaluation.

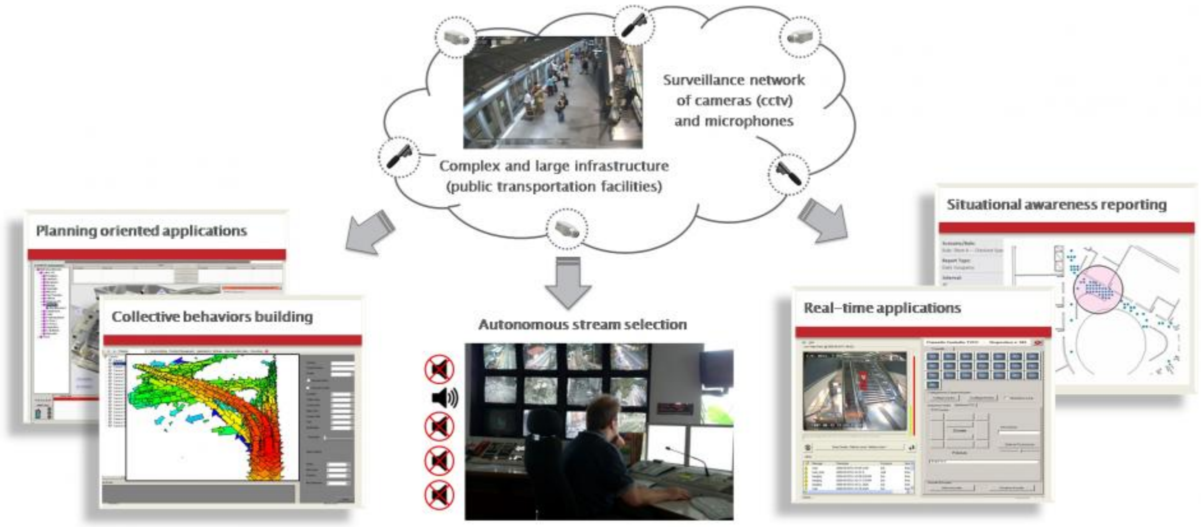
[*] Corresponding author information here. Tel.: +32 65 34 28 01; fax: +32 65 34 27 29.
***E-mail address*: carincotte@multitel.be**

Other authors information here. www.vanaheim-project.eu/project/consortium.php

# 1. Introduction

From single camera and monitor to complex surveillance networks with hundreds of cameras and network video recorders, more and more surveillance systems are deployed on a daily basis, such those already prevalent in shopping malls, parking lots or public transport facilities. However when dealing with large infrastructures like transportation ones, the primary role of automatic surveillance systems is usually to handle specific security issues, e.g. monitoring track rails between metro stations, or generate recorded material for *a posteriori* investigation (mainly for clarifying what happened after an incident and/or for forensic evidence). Less frequently are they exploited to detect events when they occur, or to collect statistics on passenger activities and behaviour trends.

In this context, the integrated project VANAHEIM studied and integrated innovative video analysis bricks in CCTV surveillance platform already in use in different European metros (Turin and Paris). Its multidisciplinary team comprises eight partners – four research institutes, two companies and two public operators from six different countries – with complementary competencies – which worked on the three specific application areas: 1. Automatic sensor selection for intelligent video wall management, 2. Human-centred monitoring for security and efficiency applications, and 3. Long-term statistics collection for infrastructure understanding and planning. The video analytics modules corresponding to these different areas have been integrated into a Video Management Solution featuring an innovative video wall module.


Surveillance network
of cameras (cctv)
and microphones
Complex and large infrastructure
(public transportation facilities)
Planning oriented applications
Collective behaviors building
Autonomous stream selection
Situational awareness reporting
Real–time applications

**Long-term statistics building for planning applications**

**Automatic sensor selection for videowall management**

**Human-centered monitoring using audio/video analysis**

Practically speaking, the resulting automatic surveillance prototype has been integrated in the Turin Automated Metro (publicly demonstrated to participants of the EXPO Ferroviaria fair held in Turin on March 28 2012), as well as in Paris metro (*Bibliothèque François-Mitterrand* station), with comparable public demonstration on September 27 2013. Next sections introduce in more details the innovative components developed within the project, as well as the integration and evaluation results constituting the VANAHEIM legacy, while the remaining of this section provides more information on the project objectives and results through these 4 years.

## *1.1. Summary description of project context and objectives*

The aim of VANAHEIM was to study innovative surveillance components for autonomous monitoring of complex audio/video surveillance infrastructure, such as the ones prevalent in shopping malls or underground stations. To do so, VANAHEIM addresses three main application-driven research questions:

1. **Scene activity modelling 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 characterise 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. VANAHEIM thus targeted the development of such automatic audio/video components allowing to select the audio/video streams to present to operators in control room.

2. **Investigation of behavioural cues for human-centred monitoring and reporting**. VANAHEIM investigated the use of subtle human behavioural cues (head pose, body shape) and social models (e.g. about space occupancy) to perform the live detection of well-defined scenarios of interest. In addition, the project also targeted three specific levels of monitoring: 1. individuals, 2. groups of people and 3. crowd/people flow. Last, VANAHEIM targeted the development of a situational awareness reporting, which aims at translating the ongoing activities of people into meaningful user-oriented figures, through for example a map-based overlay of the approximate location/number/behaviour of people in the infrastructure.

3. **Collective behaviour building and online learning from long-term analysis of passenger activities**. By combining cognitive science and ethological analysis, VANAHEIM aimed at designing models for the identification and characterization of the structures inherent in collective human behaviour. In other words, by continuously analyzing, learning and clustering information about users' locations, routes, activities, interactions with others passengers and/or equipments, and contextual data (time of day, density of people...), VANAHEIM targeted a subsystem able to estimate of the long-term trends of large-scale human behaviour, thus allowing the discovery of collective comprehensive daily routines.

The evaluation/assessment of the sub-systems developed within the project has been covered thanks to the participation of Turin and Paris transport operators; as highlighted in next sections, VANAHEIM deployment in both sites allowed to demonstrate the scalability, as well as the performance of the developed system.

### *1.2. Main scientific & technical results*

***Feb. 2010 - Jan. 2011*** - During the 1st year, the audio/video scenarios of interest have firstly been defined, together with the project system architecture and sub-systems components. A first audio-video recording system has been deployed and configured at Turin trial site, and one first audio/video dataset has been acquired for algorithms development purposes. Several main research activities in audio/video analysis have then been initiated; different ways of modeling activities occurring in videos have been investigated, and audio features extraction and audio activity modeling studies have been initiated. Later on, complementary studies for the recognition of human actions from surveillance material have started, namely some studies on multi-person tracking, head localization, head pose and body orientation estimation, some work related to detection and tracking of groups of people, and a last study on people flow monitoring in escalator. Some work related to the development of the middleware components, API, etc. needed for the preliminary integration have also been initiated during this 1st period. Last, the project visibility has been guaranteed through the set-up of the project web-site, some dissemination materials (flyer, slideshows, etc.) and some regular news about the project. The first newsletter has been sent to a mailing list of contacts potentially interested by the project, inviting them to subscribe to the User-Board, leading to a doubling of User-Board members. Last, five publications resulting from the project have been accepted or published in this 1st period.

***Feb. 2011 - Jan. 2012*** - During the 2nd period, some work has first been achieved for the deployment of the acquisition system in Paris (BFM metro station), from both legal and technical point of view. The main research studies in audio/video stream activity modeling have been pursued, to propose different ways of modeling the (normal) activities occurring in the video streams, to perform video analysis in the compressed domain, and to perform audio features extraction and audio activity modeling. In addition, the task dedicated to developing the mechanisms for autonomously selecting the more salient audio/video streams has also been initiated. The studies for the recognition of human action from surveillance material have also been pursued, on multi-person tracking, head localization, pose estimation and body orientation estimation, detection and tracking of groups of people.

Some new work on crowd monitoring has also been achieved, mainly for estimating platform occupancy at rush hours. During this 2nd period, the studies on the applicative components required for integration have also been initiated. Briefly speaking, real-time applications like left-luggage detection and group monitoring have been implemented, as well as a situational reporting application to provide operators with the rough number and location of people in infrastructure. Some studies on adaptive scene understanding using online learning have also been conducted, to show how activity zones can be clustered meaningfully, and learnt and refined incrementally according to the station use. The work related to the first integration stage (mid-project integration) has also been pursued, mainly for finalizing the middleware components, API, etc. needed for this integration, and for proposing a simple but functional HMI to be used for this preliminary integration. As agreed with the EC, the deployment and assessment of the developed system, as well as the user-field trials and evaluations of the integrated system, have been postponed. Last, the project visibility has been guaranteed through numerous scientific publications, publication of regular news on the project website, organization of two workshops in the field, and management of several newsletters dedicated to a survey on video content analysis for transportation applications.

***Feb. 2012 - Jan. 2013*** - Regarding the 3rd period, the very first activities were dedicated to the finalization of the first integration stage (mid-project integration) at GTT pilot site, and the related user-field trials and evaluation tasks. Then, the project audio-video recording system has been deployed and configured at RATP trial site, and one first audio/video dataset has been acquired. Annotations from both scientific and sociologic point of view have been pursued on both GTT and RATP data. The research studies in audio/video stream activity modeling have been pursued and finalized; extension of previous models and new ones have been proposed and evaluated on the project dataset, both on previously available GTT dataset, as well as on the new dataset coming from RATP pilot site. The task dedicated to developing the software bricks for autonomously selecting the more salient audio/video streams has also been pursued, as well as some stream selection experiments on visual attention. The studies for the human-centered features extraction have also been pursued and almost finalized, on human detection and multi-person tracking, coupled head and body pose estimation, detection and tracking of groups of people, and crowd monitoring. Regarding the video surveillance applications themselves, the left-luggage detection module and the group monitoring one have been improved, as well as the situational reporting application which has been evaluated on RATP pilot site. The studies on adaptive scene understanding using online learning have also been pursued. As for the project integration steps, the project framework has been modified according to the RATP environment with respect to the GTT one. A first working version of the updated system has been deployed at RATP together with a first set of updated analytics modules. Last, the project visibility has been guaranteed through numerous scientific publications, a public demonstration at GTT pilot site, a sponsored summer school on human activity and vision at INRIA, a scientific demonstration at ECCV, and several articles in specialized press (Metro Report International, Mobility Magazine), news, newsletters, etc.

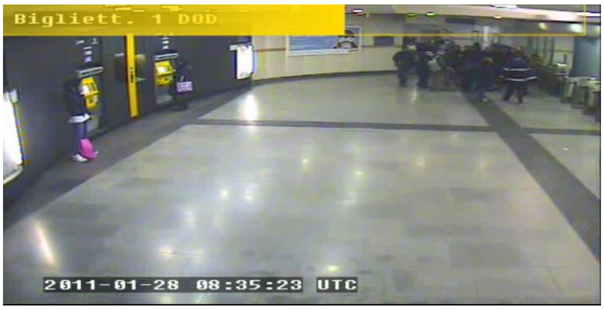
***Feb. 2013 - Sept. 2013*** - Regarding the 4th and last period of the project, the main activities were dedicated to the finalization of the second integration stage at both RATP & GTT pilot sites, and the related user-field trials and evaluation tasks. First, the last planned active research tasks on autonomous stream selection mechanisms, monitoring of people flow/crowd, contextual behaviors from long term analysis, and adaptive scene understanding using online learning were concluded with a focus on deployment. Then the consortium refined the user-field trials & demonstration scenarios so as to be able to prepare both the system evaluation and the final demonstration to be held at RATP. Following these refinements, the stream selection process has been improved to use both abnormality measurement on video modules, as well as events generated from real-time modules (static luggage detector, counter flow on escalator, etc…). Various improvements have been performed at the audio/video analytics level, e.g. for improving performance of left luggage, counter-flow or group detections, improving integration level of situational reporting (in the system or available for more partners), providing more audio analytics modules, integrating and demonstrating the long term (offline) analysis, or retraining models and updating modules for RATP and GTT deployments. Following the updates of the system done at each pilot sites, the user-field trials and evaluations were conducted. The very last efforts of the consortium were dedicated to the organisation of the final demonstration on Sept. 27 2013.

## 2. Automatic sensor selection for video wall management

Most of the time videos from large surveillance network are never watched. For instance, in the case of Turin metro, 28 monitors are used in the control room to supervise more than 1100 cameras; the probability of watching the right streams at the right time is therefore close to zero. Moreover, vigilance studies show that operators who spend hours 'screen gazing' at static scenes tend to become bored and less attentive. They are thus likely to miss low-frequency events such as a falling people, which reduces the overall effectiveness of CCTV.

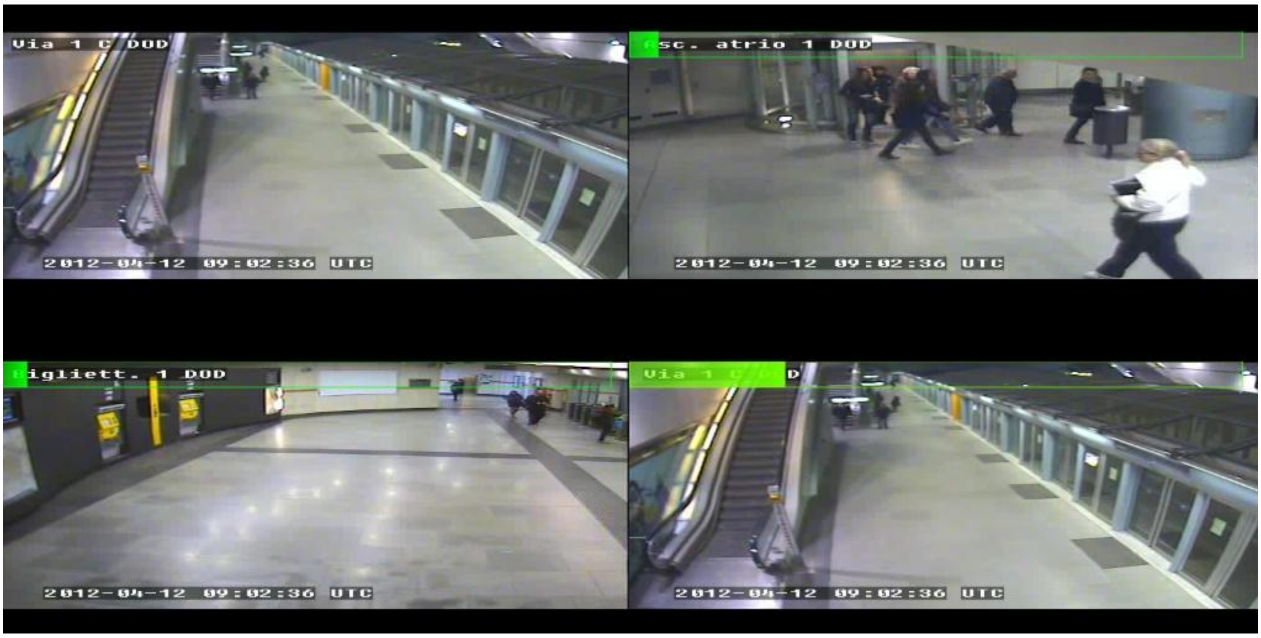
### *2.1. Video abnormality rating*

In this context, there is a need for automatic content-based selection systems that select the most informative / relevant data streams and provide them to the surveillance operators. To do so, different unsupervised activity modelling algorithms that capture recurrent activities from long recordings have been developed; *a contrario*, these algorithms are also able to provide unusualness measures that can be used to select the most interesting streams to be displayed in control rooms, as illustrated below.



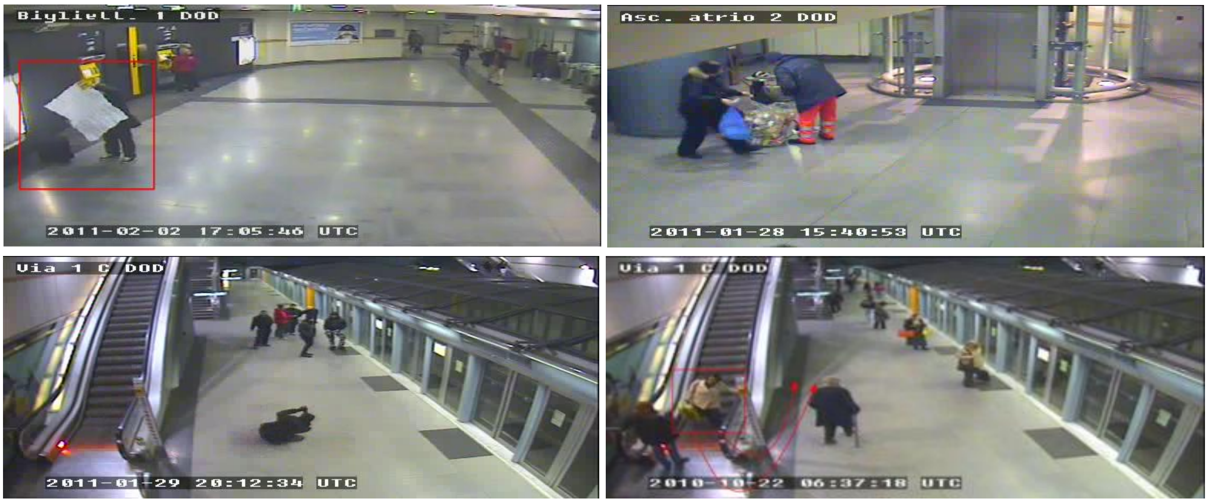
*Video stream abnormality rating – providing a continuous measure of abnormality (or unusualness). By convention, this measure is in the range [0:1] (represented by the color bar). On this example, one person is falling (top right corner) and people are gathering around him, which yields to a significant unusualness value.*

The developed system is then able to process unusualness measures coming from several video streams (e.g. the whole video streams from a station), and to compare them together in order to select the stream characterized by highest abnormality level**.** This video is then displayed as the video stream of interest for the operator, e.g. the top left video stream in the following figure.



*VideoWall module – able to display unusualness measures coming from different video streams (cf. color bars), and to select the stream characterized by highest abnormality levels (top-left screen).*

Finally, applied to more than 200 hours of video footage from 8 cameras of Turin metro, this automatic sensor selection enables the discovery of statistical anomalies such as of people using the escalators in counter flow, crowd of people linked to a fall, teenager skylarking, people lost or distributing leaflets, cleaning staff emptying a garbage, etc..



*Examples of video anomaly results identified by the stream selection system – people lost, cleaning staff emptying a garbage, teenagers skylarking, and people using the escalators in counter flow.*

It worth pointing out that, implicitly, some phenomena statistically rare but not necessarily abnormal or interesting are also identified. While this could be problematic for a detection system (typically an intrusion detection system or an abandoned object one), the presence of non-abnormal event is much less problematic in the case of a selection of video streams, where the objective is not to generate an alarm, but to prioritize one or more video streams from a large number of video streams (among which, all can be normal). In this context, displaying something even normal is not a problem, since it is what is done in a standard supervision task.

### *2.2. Audio-based selection mechanism*

One additional module has also been developed in the audio domain for this selection mechanism. This module consists in an abnormal event detection (AED) module which aims at detecting unusual audio signals that should be considered as abnormal. This module use models trained unsupervisedly on recorded signals; it then extracts online acoustic features from the live audio stream, and performs an automatic segmentation of the stream. This output can then be either displayed in a gauge as an overlay on the video wall, or used to select related video stream(s) based on the audio abnormality level of the area under surveillance.

### *2.3. Event-based selection mechanism*

With respect to first version of stream selection (mid-project integration), some important updates have been performed in the final version to allow the stream selection to use also events generated from real-time algorithms as an input of the stream selection module.

The stream selection mechanism was modified to, in case of the detection of a specific event (for example a static luggage detection), automatically present to the operator the camera that generated the event. In this way stream abnormality measure, both audio and video, and real-time events compete to select the best cameras to be selected and presented to the operators. The following modules are used for stream selection:

- Counter flow on the escalator detection
- Static luggage detection
- Abnormal Group detection

### *2.4. Stream selection from the social attention point of view*

In the project, we also developed a behaviour catalogue of annotated “interesting behaviours” and tested the occurrence of human social attention for video streams. When behaviour and social attention are compared to stream selection algorithms, we found that in some cases stream selection algorithms outperform humans in behaviour detection. This is not surprising because the stimulus presentation for humans was competitive. i.e. they had to decide between four videos. Generally we were able to demonstrate a correspondence, although it was not high in case of some behaviours.

These results have implications for future projects in stream selection tasks from surveillance cameras.

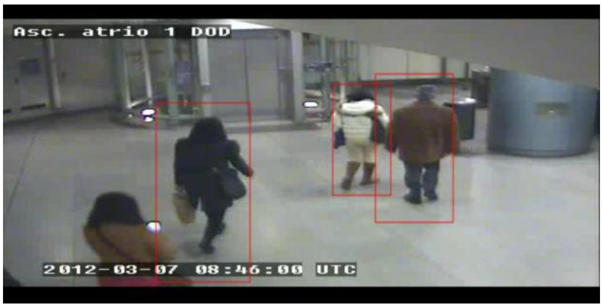
- In any case there is no absolute “ground truth” - human behaviour codings have a limited reliability, as it is the case for social attention. Given this, stream selection algorithms performed well and we developed a new procedure for the calculation of reliability.
- The behaviour catalogue might be not extensive enough – this suggests that we should reanalyse the false positives generated by the algorithms. This was done in a first attempt and we have not found a really clear picture.
- The error analysis was done only for behaviours and not human social attention – this seems necessary in order to get a complete picture.
- The results which are interesting for us are that human brains seem to use certain cues like motion and colour change for attention giving in a competitive situation. This is new for our field.

In our view this should be the starting point for a new project where iterative procedures are used for the development of stream selection algorithms. The caveat in this project was that the research was running parallel.

## 3. Human-centered monitoring using video analysis

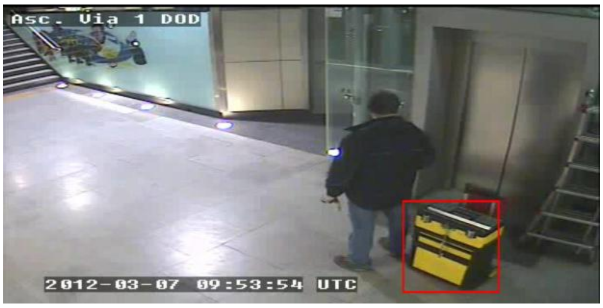
In a second research thread, different real-time scene analysis algorithms have been investigated and integrated in the pilot sites. As detailed below, these modules range from human and group detection to left luggage detection, through people counting and flows estimation at different locations in the Turin station.

**Human detection.** This module focuses on the capability to efficiently and reliably detect persons in the station. The processing is done using appearance and spatio-temporal features. A human/non-human classifier is then used to detect people presence.



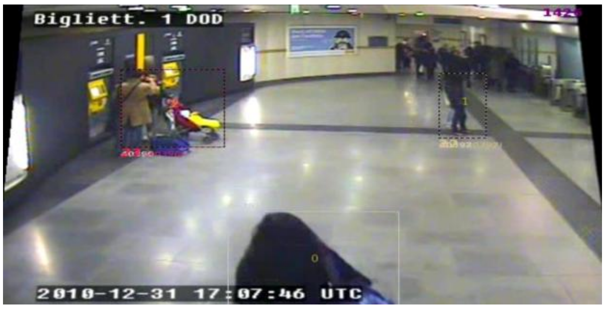
*Human detector – allowing to detect passenger's presence in the data stream.*

**Static-luggage detection**. The algorithm relies on a multi-layer background subtraction method by distinguishing recent background layers from old and long term background layers and foreground regions. It is enhanced by a human detection algorithm to remove false alarms due to people waiting and hence remaining static for quite some time.



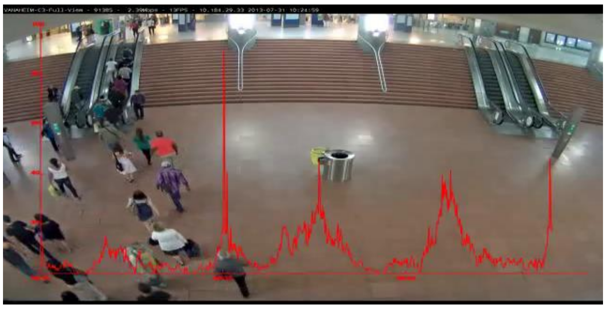
*Left luggage detector – monitoring a video stream and detecting abandoned objects like an unattended luggage.*

**Group detection and behavior analysis**. The algorithm performs real-time group detection and tracking and detect events corresponding to predefined behaviors of interest. The group detection and tracking method works by first tracking mobile objects and then grouping them recursively over time maintaining a spatial and temporal group coherence, using proximity as well as walking similarity (based on speed and direction).



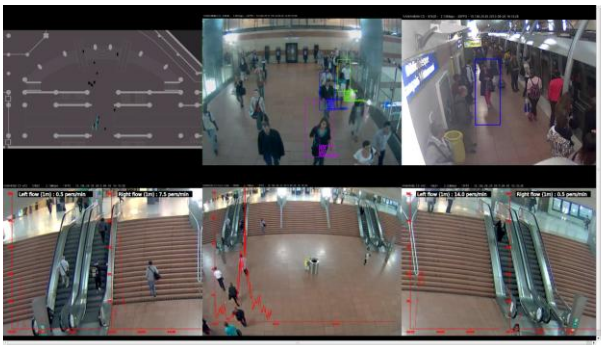
*Group module – detecting and tracking groups of people in video-surveillance videos, leading to the detection of group behavior or events relative to groups.*

**Flow monitoring**. Several crowd characterization approaches have been developed to monitor crowd/flow of people.



*People flow monitoring in escalator – counting people that cross a virtual line on the escalator.*

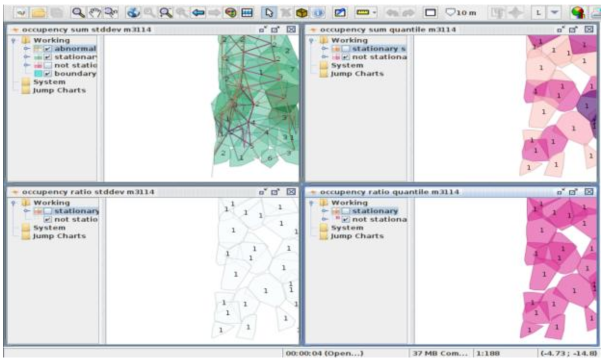
**Situational reporting.** Finally, the full set of methods that provide estimates of people locations and numbers (i.e. people flow in escalators, people density at platforms and human presence at lift, and, whenever possible, multi-camera/-object tracking) are used to feed a situational reporting tool. The location information is back-projected in real time on the infrastructure map of Turin metro, providing the operators with synthetic views of the metro occupancy and activity.



*Situational reporting tool – aiming at reporting passengers location (cf. dots at escalators and in the hall, tracks on footbridge) on the infrastructure ma.*

## 4. Long-term statistics building for planning applications

Finally, an offline line tool that aims at analyzing long term recording has been developed. It aims at three different tasks summarized below: i) learning of (floor) activity zones within the scene; ii) learning of activity classes, and iii) calculation of activity statistics.

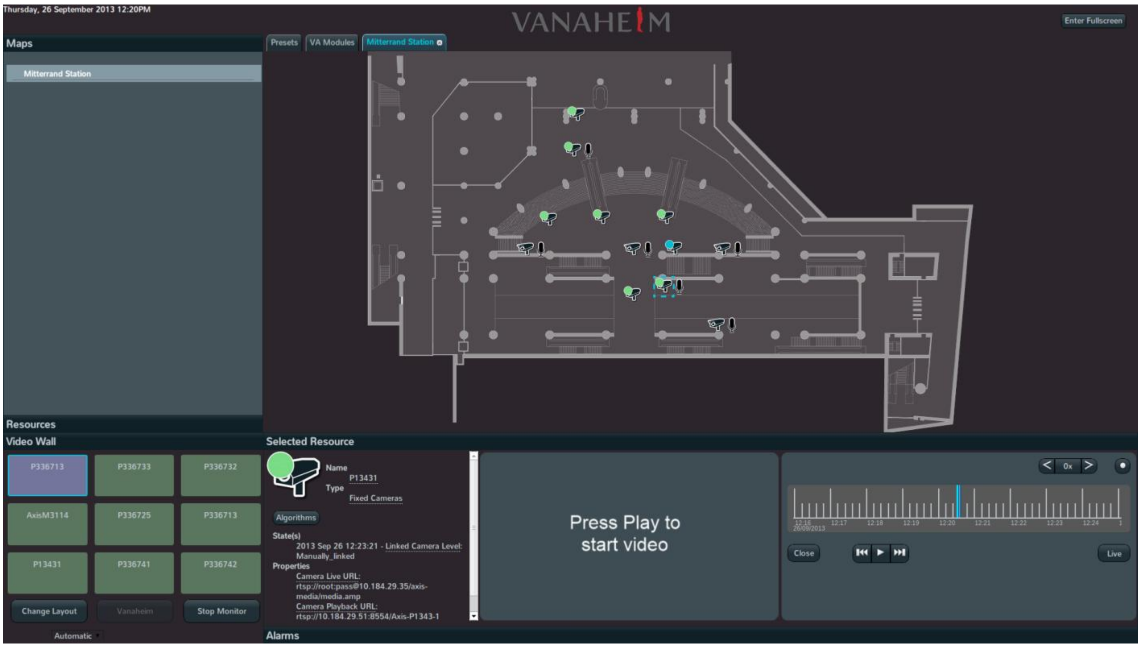


*Offline analysis tool – aiming at three different tasks: learning of activity zones, learning of activity classes, activity statistics calculations*

## 5. Integrated systems in Turin/Paris metros

The final integrated system consists of a professional Video Management Solution, where all the analytics modules developed for the VANAHEIM project have been integrated. The user of the system can easily launch any available algorithm, select the audio or video stream to use, and configure the graphical overlay (video wall) where to show the analytics' results. The system is currently functional (with online processing) on 26 real data streams from DOD station in Turin, and 14 data streams from BFM station in Paris.

The assessment of the VANAHEIM prototype has been covered by Turin (GTT) and Paris (RATP) metro operators (cf. conclusion for the main outcomes of the evaluation). Following these assessment/validation stages, several live demonstrations have been organized at both pilot sites – around project mid-term at GTT site (March 28 2012 during EXPO-Ferroviaria) and at the end of the project at TRATP site (September 27 2013).



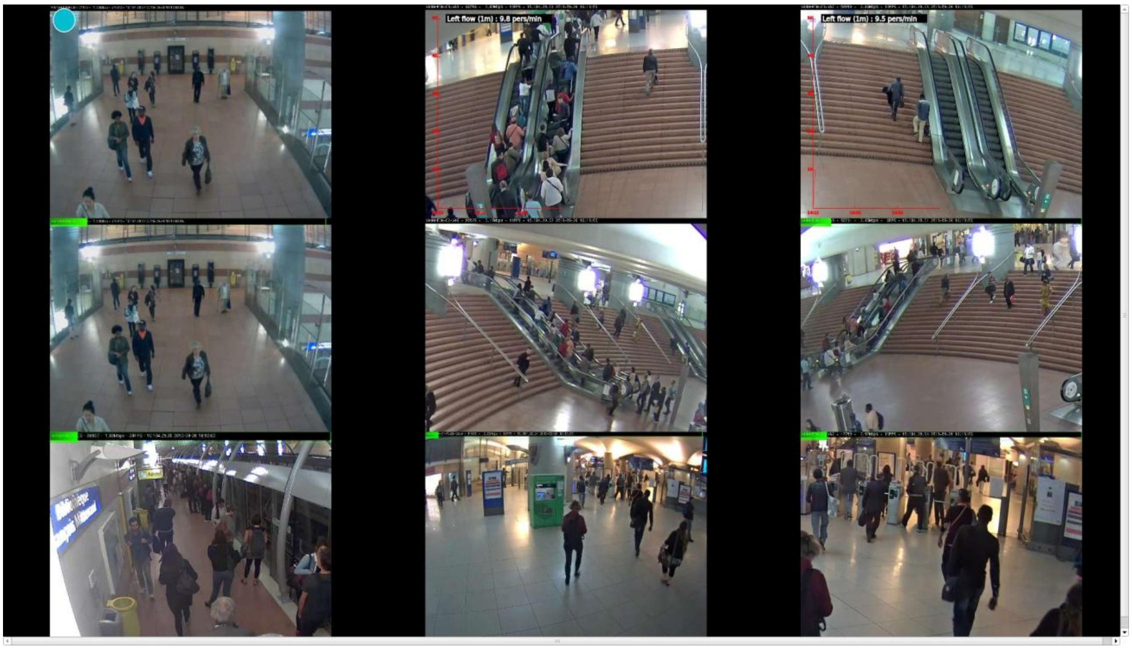
Name: P13431

Type: Fixed Cameras

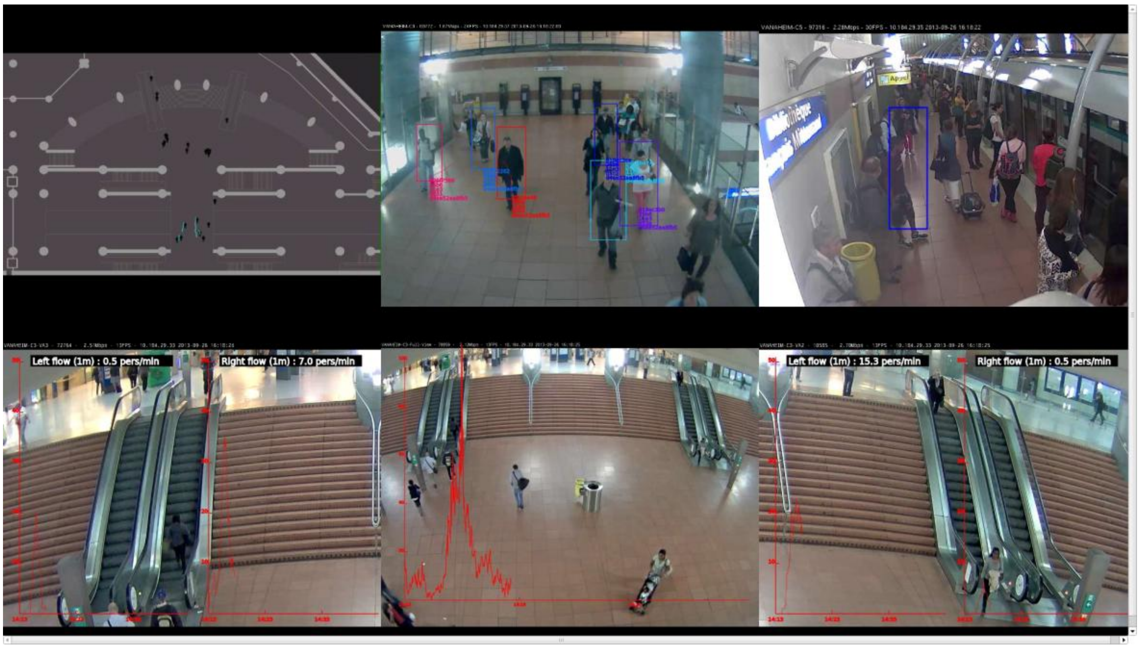
State(s)
2013 Sep 26 12:23:21 - Linked Camera Level: Manually_linked

Properties
Camera Live URL: rtsp://root:pass@10.184.29.35/axis-media/media.amp
Camera Playback URL: rtsp://10.184.29.51:8554/Axis-P1343-1

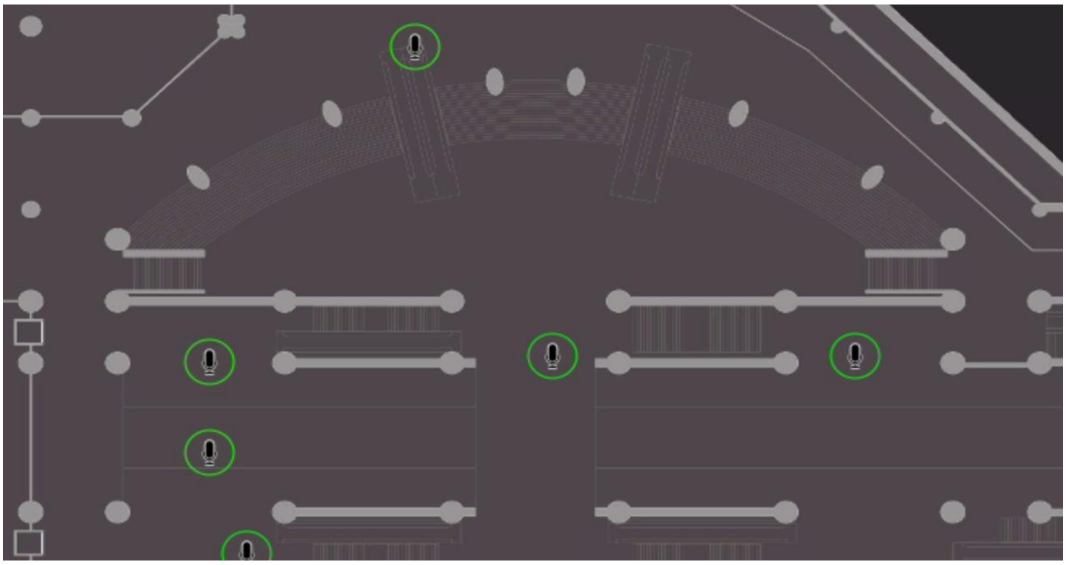
*VANAHEIM Web Interface displaying RATP station map.*



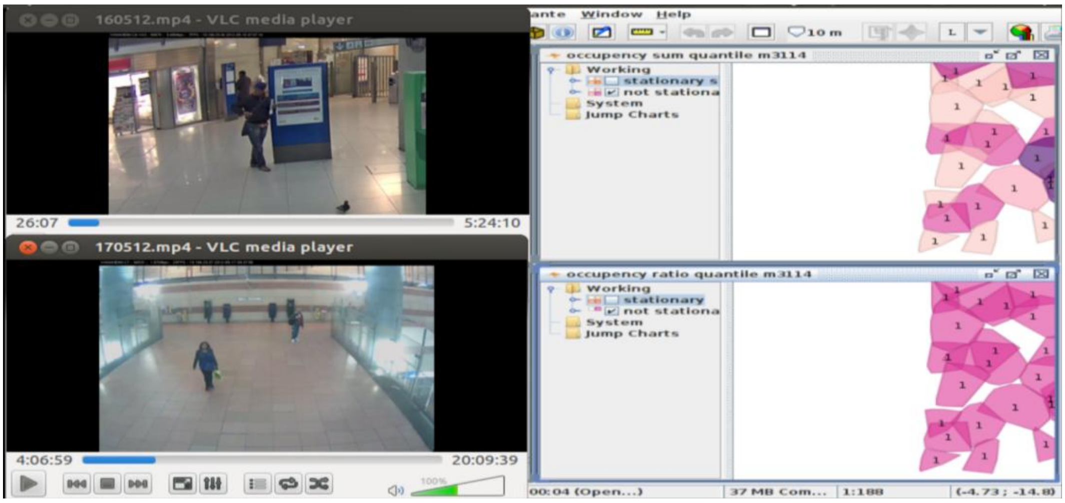
*VANAHEIM VideoWall displyaing abnormality values (color bar) and counting results (top mid/right streams).*


Left flow (1m) : 0.5 pers/min
Right flow (1m) : 7.0 pers/min
Left flow (1m) : 15.3 pers/min
Right flow (1m) : 0.5 pers/min

*Situational reporting module displaying information coming from several modules.*



*Audio Situational Awareness providing the abnormality score of the microphones.*



*Offline tool showing the possibility to replay the recorded video related to a selected activity*

## 6. Evaluation and Perspectives

Regarding the evaluation, the chosen methodology was intended to have both a qualitative and a quantitative assessment of VANAHEIM results. In order to do so, the evaluation has been split in 2 different steps:

- the first step consisted in interviews during which the VANAHEIM staff demonstrated in detail the whole system, asking specific questions to GTT/RATP staff with different roles to get their opinion
- the second step consisted in letting operators play with the prototype autonomously with the only help of a user manual, and asking them to fill a data sheet that has been designed on purpose.

The most interesting consideration raised from the questionnaire at both sites were the following:

- the highest score was given to the ease of use that was one of the goals
- it was given an high score to the ease to understand information given that is, together with the previous statement, an interesting goal. In fact, the consortium knew, from the scientific point of view, the quality the results of different algorithms (in terms of accuracy, etc..); what was needed and therefore tested was the usability and understandability of those results from the end user point of view
- the lowest score was given to the acceptability of an high percentage of false alarms, that means that is better to miss some alarms than to have a system unusable due to elevate number of false alarms
- the credibility of the information provided has a quite high score that means that the algorithms are working quite well and there was a good trade-off between false alarms and accuracy of the results

Regarding the interviews, the first thing to say is that people working on train operation and movement feels as hazardous and therefore abnormal, all situations in which a huge amount of people overcrowd in a specific part of the station. Therefore a combination of abnormality detection, stream selection, people counting and space occupancy can be used as an early alarm to help on operation. Security related people are, on the contrary, more interested in the behaviour of the single person.

The second consideration is that operator and managers have a different feeling on the usefulness of online and offline tools: operators don’t care at all about offline tools while managers are really interested on offline tools and statistic. They even would like to be able to extract statistical data from all the different algorithms.

The third consideration, that was clear talking about stream selection, even if not limited to it, is that all people interviewed considered VANAHEIM idea very interesting and very useful and they would push more and more the correlation between the results of different algorithms: in fact they don’t care which algorithm is giving the information, they want to have a correlation of the maximum amount of information already packed in order to give clearly the idea of “what’s going on”.

As perspectives, the VANAHEIM system is now operational with all the algorithms and should be tested over a longer period. At RATP, the system could come to interface the system already deployed in multimodal video center station of Châtelet les Halles. This would assess supervisors video presents in this new center the overall level of system performance. At GTT side, GTT would like to select some algorithms and ask the different partners to extend them to the whole metro.

## Acknowledgements

The research leading to these results has received funding from the European Community’s Seventh Framework Programme FP7/2007-2013 - Challenge 2- Cognitive Systems, Interaction, Robotics - under grant agreement n°248907-VANAHEIM.

**Project type:** Large-scale integrated (IP)
**Duration**: 44 months (February 2010 – September 2013)
**Budget**: €5,471,851  (EU contribution €3,717,998)

For more information, please visit www.vanaheim-project.eu