ADVISOR

R7.3 Technical Validation of the ADVISOR Demonstrator

IST-1999-11287

Classification: Public

Author A. D. Barnett

Date 28 May 2003

The ADVISOR Consortium consists of:

Thales Research and Technology (UK)	Co-ordinator	United Kingdom
VIGITEC SA	Principal Contractor	Belgium
BULL SA	Principal Contractor	France
INRIA	Principal Contractor	France
University of Reading	Principal Contractor	United Kingdom
Kingston University	Principal Contractor	United Kingdom

Thales Research and Technology - UK,
Worton Drive,
Worton Grange Industrial Estate,
Reading,
Berkshire
RG2 0SB

Classification: Public

Documents that are classified as Internal may not be distributed outside the Project Team, without prior written consent of all members of the Steering Committee.

MODIFICATION RECORD

Version	Date	Author	Purpose/Description
Draft A	04-04-2003	A.D. Barnett	Initial draft of Demonstrator validation report
Draft B	11-04-2003	A.D. Barnett	Updated draft, incorporating review comments
Issue 1	23-05-2003	A.D. Barnett	Issue 1 for release, incorporating review comments of Advisor consortium partners

Contents

Tables

Executive Summary

This document describes the validation of the Annotated Digital Video for Intelligent Surveillance and Optimised Retrieval (ADVISOR) Demonstrator system [1]. The report includes the data used for the validation, the validation process employed, the results obtained and an analysis of the results.

The system was validated at the Sagrada Familia Metro station in Barcelona, where the Demonstrator was taken for evaluation, demonstration and validation purposes. The validation process involved testing the behaviour recognition, the archive search and the archive retrieval of the ADVISOR system. The validation of the behaviour recognition comprised the detection of fighting, blocking, overcrowding, jumping over the barrier and vandalism.

The evaluations, validation and demonstrations were conducted using both live and recorded data. For the validation task, the system was tested using four input channels in parallel, the four channels being composed of three recorded sequences and one live input stream from the main hall of the Sagrada Familia Metro station. The validation was conducted over four hours which, over the four channels, gave a combined total of 16 hours of validation. The three recorded data sequences were constructed using shorter sequences of the required behaviours, created using actors. The accuracy of the report was also measured as percentage of the ground-truth behaviour over which the correct report was generated by ADVISOR.

In total, out of 21 fighting incidents in all the Demonstrator sequences, 20 alarms were correctly generated, giving a very good detection rate of 95%. These twenty correctly identified alarms had an average report accuracy of 68%. Out of nine blocking incidents, seven alarms were generated, giving a detection rate of 78%. These seven alarms were found to be 60% accurate on average. Out of 42 instances of jumping over the barrier, including repeated incidents, the behaviour was detected 37 times, giving a success rate of 88%. The two sequences of vandalism were always detected with an overall accuracy of 71%, over six instances of vandalism. Finally, the two overcrowding alarms in camera C11 were consistently detected, with an overall accuracy of 80% over 7 separate instances of the alarms. The overcrowding alarms were also consistently detected in the live camera C10, with some 28 separate events being detected. In conclusion, the algorithms responded very successfully to the input data, with high detection rates and with all the reports being above approximately 70% accurate. No false alarms were generated during the playback of the recorded sequences although one false blocking alarm was generated in the live input.

An independent validation of the behaviour algorithms was performed by INRIA, outside of the ADVISOR system, and not in multi-channel real-time. The following results were obtained. Out of 17 blocking incidents, 16 incidents were correctly identified, giving a success rate of 94%. The fighting behaviour was detected in 23 out of 27 incidents, giving a detection rate of 85%. Overcrowding was detected in the two incidents of that behaviour and vandalism was detected all three times. However, one of the vandalism incidents also produced a false fighting alarm. Finally, all three jumping over the barrier sequences were detected.

The archive search and retrieval functionality was found to work as specified, through the controls provided on the HCI.

In conclusion, the ADVISOR system meets the requirements of the Demonstrator as laid out in the functional specification document [1]. The system was easily able to cope with live input, although the threshold for overcrowding may need to be adjusted. The system works very well on the constructed sequences, but the lack of real data or real fighting events during the validation makes it difficult to assess the ADVISOR system's ability with these events. Finally, the performance of the ADVISOR Demonstrator in Barcelona, showed that it has a lot of potential to improve the security and safety of passengers on a Metro system. Furthermore, the system could have a wide-range of applications to detect and respond to human behaviours in many different settings.

1 Introduction

This document describes the formal validation of the Annotated Digital Video for Intelligent Surveillance and Optimised Retrieval (ADVISOR) system at the Demonstrator stage [1]. The report includes the data used for the validation, the validation process employed, the results obtained and an analysis of the results.

The requirements of the Demonstrator validated in this report, as described in [1], are as follows;

- Demonstrator startup, including implementation of configuration files
- Demonstrator shut down
- Source material requirements
- Capture module functionality
- Machine vision processing functionality, comprising recognition of specified human behaviours
- Archive, Search and Retrieval functionality
- Human Computer Interface (HCI) functionality

The ADVISOR system detects the presence and motion of people by CCTV cameras and attempts to classify and report their behaviour. Identification of these behaviours in the system generates an alarm at the user interface.

The ADVISOR demonstrator consists of six different modules: Capture; Motion Detector; Crowd Monitor; Behaviour Recognition; Archive; Human Computer Interface (HCI). The modules reside on separate computers in a local area network (LAN) and are connected to each other using socket interfaces. Figure 1 illustrates the ADVISOR architecture.

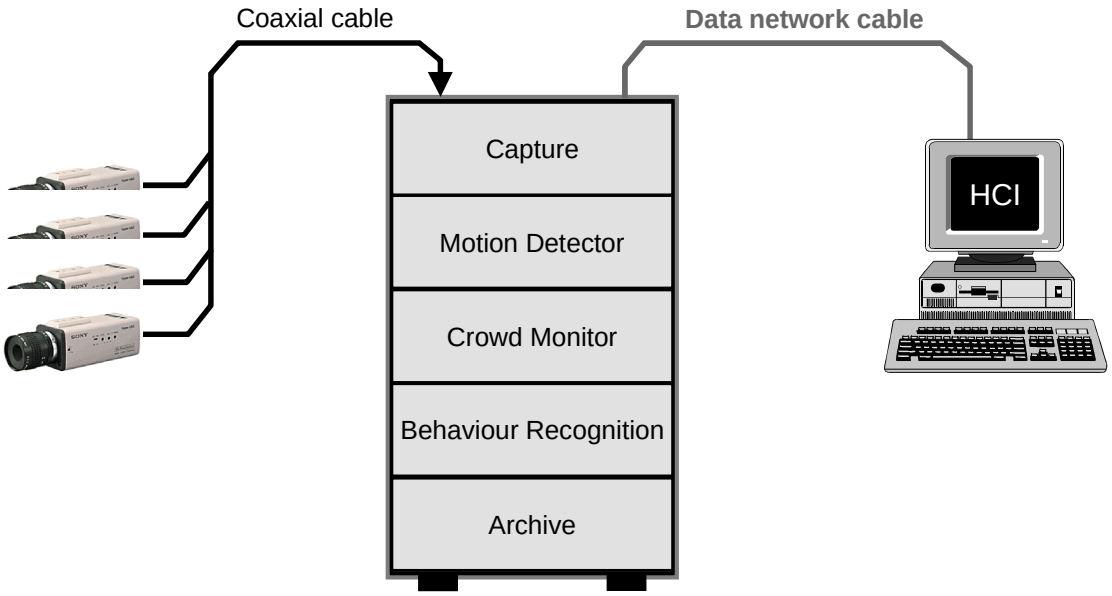

Figure 1: The ADVISOR system.

The behaviours tested in the Demonstrator are those of blocking, fighting, overcrowding, jumping over the barrier and vandalism against equipment, as described in [1]. The technical definitions of these behaviours are described in Section 3.1.

The ADVISOR system is intended for use on underground Metro systems. Therefore, the ADVISOR demonstrator was taken to the Sagrada Familia (SGFM) Metro station on the TMB (Transport Metropolitans de Barcelona) Metro to be used for three different tasks:

- The evaluation of the ADVISOR demonstrator by Metro CCTV operators
- A formal validation of the ADVISOR demonstrator in a live scenario
- Presentations and demonstrations of the ADVISOR project to the European Commission (EC), project reviewers and invited guests of the ADVISOR consortium

The ADVISOR system was presented to CCTV operators from the STIB (Brussels) and the TMB (Barcelona) Metro systems, followed by a tutorial on the use of ADVISOR. The operators were asked to use the system for a few hours and evaluate its performance and usefulness. The results of the evaluation were recorded by the completion of a comprehensive questionnaire, the findings of which can be found in [3].

The system was also presented and demonstrated to the European Commission project officer and the project reviewers during the final review of the ADVISOR project, at the TMB offices at Sagrera, Barcelona. Also present at the demonstrations at Sagrera were representatives from SRWT (Société Régionale Wallonne du Transport), Thales Security, and Alstom.

Lastly, whilst on-site at Sagrada Familia, a formal validation of the ADVISOR system was performed. The results of the validation are presented and analysed in this report.

The evaluations, validation and demonstrations were conducted using both live and recorded data. Four CCTV cameras at SGFM were connected to the ADVISOR system, providing live data from the Metro station. In addition, four recorded sequences, constructed from acted sequences containing the various human behaviours, were also fed into the system. These sequences were created with all three purposes in mind and were intended to demonstrate the capability of ADVISOR to recognise behaviours, such as fighting, that were unlikely to occur in the live cameras during the validation and demonstrations. For the validation task, the system was tested using four input channels in parallel, the four channels being composed of three recorded sequences and one live input stream.

The validation of the ADVISOR demonstrator was carried out by testing the three functions of ADVISOR seen by an end-user.

- live reporting of behaviour alarms
- searching the archive
- retrieving events and alarms from the archive

The validation of the behaviour recognition involved playing the sequences through the system and noting the resulting behaviour report generated by ADVISOR. The demonstrator sequences were ground-truthed by identifying the behaviour by manual inspection and recording the times at which those behaviours occurred. The results obtained when the sequence was played through the system were then compared with the ground-truth. If the system generated the correct behaviour report then an estimate of the accuracy of the report was obtained. This was achieved by measuring the percentage of the behaviour event during which the correct report was generated, allowing for a measurement delay of five seconds, the specified delay in the behaviour recognition algorithms.

In addition to the formal validation, the algorithms that detect the various behaviours were tested independently by INRIA, outside of the ADVISOR system. The results of these tests are also given in this report.

The validation of the archive search functionality consisted of searching the archive for the reported alarms using the HCI. The results were compared with the live alarms already generated. Similarly, the validation of the retrieval functionality consisted of playing back alarms stored in the archive and testing all the playback options, such as fast forward, jump to end of sequence, etc. The validation of the search and retrieval functionality implicitly tests the functionality of the HCI.

In summary, the validation process entails the explicit testing of the reporting, searching and retrieval of behaviour alarms and events generated by ADVISOR in response to three recorded cameras and one live camera being played through the system. Modules such as Capture, the HCI, the Motion Detector and the Crowd Monitor are tested implicitly in the overall validation: that is, the ADVISOR system as a whole would not give any results if these modules did not function properly.

1.1 Document Layout

In the following section, the validation process is described, followed by a description of the data used and the ground-truth of that data in Section 3. In Section 4, the results of the response of the ADVISOR system to the data is presented and analysed. Finally, in Section 5 a summary of the report and the conclusions drawn are given.

2 Validation Process

2.1 Behaviour Recognition Validation Process

The behaviour recognition of ADVISOR was validated by playing live and recorded data through the system. The data contained behaviours performed by actors, such as blocking and jumping over the barrier, and normal behaviours, such as people walking across platforms and halls.

The validation process was performed by firstly recording whether or not the ADVISOR system detected the observed events. This included any false reporting of events, such as the system reporting 'fighting' when no fighting occurred.

The time at which the event was reported, and the duration of the event, was compared with the ground-truth, giving, as the result of the validation process, a percentage of the event during which the correct behaviour was reported. This included true negative responses. So, for example, if ADVISOR reported a sequence as showing fighting for 45 seconds, when the ground-truth shows that 60 seconds of fighting occurred, then a score of 75% was awarded. The score also included true negative periods of the sequence, i.e. if nothing happens and no reports are generated then the report is counted as being correct. A delay of 5 seconds after the behaviour starts in the ground-truth is permitted in the measurement as this is the specified maximum delay in the behaviour detection algorithms. The algorithms must detect the behaviours occurring for the durations specified in [1] to generate an alarm, i.e. fighting must be detected for at least 2 seconds, blocking for at least 4 seconds, etc.

For the validation process the three long demonstrator sequences from cameras C01 (SGFM), C11 (SGFM) and C05 (YZER), together with the live input from camera C10 (SGFM) were used.

The validation process used for the Demonstrator is slightly different to that used for Test-Bed 2 [2]. Under Test-Bed 2, two stages of validation were carried out, firstly a simple binary comparison of detected events and then a measure of the accuracy of the report. In this validation of the Demonstrator, the two stages have effectively been combined since the detection rate was much higher than under Test-Bed 2 with the vast majority of all the behaviours being detected.

2.2 Search Validation Process

The search functionality of ADVISOR was tested by using the HCI to search for the alarms that had been generated during the validation of the behaviour recognition functionality and examining the response.

2.3 Retrieval Validation Process

The retrieval functionality of ADVISOR was tested by using the HCI to retrieve the alarms that had been generated during the validation of the behaviour recognition functionality. In addition, the following retrieve control commands were also tested:

- Start the forward play of an image sequence
- Start the reverse play of an image sequence
- Stop the playback of an image sequence
- Pause the playback of an image sequence
- Fast forward the playback of an image sequence
- Fast reverse the playback of an image sequence

- Jump to the end of an image sequence
- Jump to the beginning of an image sequence
- Play beyond the beginning/end of an image sequence

3 Ground-Truth of the Validation Data

The sequences used in the validation of the ADVISOR demonstrator are composed of the ADVISOR-SEQ-NNN short sequences containing acted behaviours and general sequences. The sequences contain the behaviours of fighting, blocking, overcrowding, jumping over the barrier and vandalism against equipment.

The behaviours recognised by the ADVISOR demonstrator are those of blocking, fighting, overcrowding, jumping over the barrier and vandalism against equipment. The technical definition of each behaviour is described in [1]. The behaviours correspond to the following situations.

- Blocking occurs when a group of at least 2 people is stopped in a predefined zone for at least 4 seconds and can potentially block the path of other people.
- Fighting occurs when a group of people (at least 2 persons) are pushing, kicking or grasping each other for at least 2 seconds.
- Overcrowding occurs when the density of the people in an image is greater than a specified threshold
- Jumping over the barrier occurs when a person jumps over a specified ticket validation zone
- Vandalism against equipment occurs when an individual is damaging a piece of equipment in the image

The ground-truth process of the input data is described in the next section.

3.1 Validation Ground-Truth Data

Consider an alarm or event occurring in one of the four input channel sequences. The ground-truth data for this sequence might be:

Fighting: yes

In other words, the behaviour event *should* generate a 'fighting' alarm and should *not* generate any other alarms. The ground-truth data is created by *inspection*. That is, a competent authority examines the sequence and decides on what behaviours are being exhibited.

Even given a suitable definition of what constitutes the behaviour that should produce an alarm, it is still a subjective assessment whether or not a particular sequence does in fact meet the criteria. The view before validation was that this should not be a problem, because alarms should be high-level ideas with human-understood meanings. As such, the user should easily understand them. This should actually help with producing useful ground-truth data.

For example, consider 'fighting'. The user (an operator) is not going to be concerned with the technical definition of 'fighting'. They are going to judge the system's performance quite simply: how often does a 'fighting' alarm actually show what *they* consider to be fighting, and how often is what *they* would call fighting actually ignored?

The long sequences used in the validation process are given in Table 1. The events in the long sequences are composed of acted sequences taken from the ADVISOR-SEQ-NNN sequences. The key for the alarm codes shown in Table 1 are given in Table 2. The light blue alarms indicate the fighting sequences, the purple indicates overcrowding, the grey indicates blocking, the red indicates jumping over the barrier and green indicates vandalism against equipment.

Alarm Code	Frame Number	Alarm Code	Frame Number
F1	3315	B14	41170
F2	4435	F9	41920
O1	6370	O2	44325
B18	7420	F24	45720
O2	9350	J3	47165
J1	11450	F10	48475
B15	12090	B15	49005
V1	13850	V1	49605
F25	15775	F11	50850
J3	16525	V3	52815
F4	19030	F25	54695
B16	19300	F12	55295
O1	21280	O1	56620
B17	23295	O2	58270
O2	24045	F13	61060
J2	27205	B16	61655
F26	27880	J1	62255
F6	28630	F26	64955
V3	29660	V1	65555
F27	31830	F14	66800
F7	32475	J3	67930
V1	33630	F15	70180
F8	36075	F1	71200
B18	36355	B17	71570
O1	37165	J2	72170
O1	40120		

Table 1: The combined alarms and events, in time order, in the three demonstrator sequences.

Alarm Code	Camera and Station	Behaviour Type	Sequence Label
F1	C01 TMB	Fighting	SEQ-042
F2	C01 TMB	Fighting	SEQ-043
F3	C01 TMB	Fighting	SEQ-044
F4	C01 TMB	Fighting	SEQ-045
F5	C01 TMB	Fighting	SEQ-046
F6	C01 TMB	Fighting	SEQ-047
F7	C01 TMB	Fighting	SEQ-048
F8	C01 TMB	Fighting	SEQ-049
F9	C01 TMB	Fighting	SEQ-050
F10	C01 TMB	Fighting	SEQ-051
F11	C01 TMB	Fighting	SEQ-052
F12	C01 TMB	Fighting	SEQ-053
F13	C01 TMB	Fighting	SEQ-054
F14	C01 TMB	Fighting	SEQ-055
F15	C01 TMB	Fighting	SEQ-056
F24	C05 STIB	Fighting	2.1.01
F25	C05 STIB	Fighting	2.1.02
F26	C05 STIB	Fighting	2.1.03
F27	C05 STIB	Fighting	2.1.04
J1	C11 TMB	Jumping over barrier	SEQ-043
J2	C11 TMB	Jumping over barrier	SEQ-044
J3	C11 TMB	Jumping over barrier	SEQ-045
V1	C11 TMB	Vandalism against equipment	SEQ-094
V2	C11 TMB	Vandalism against equipment	SEQ-095
V3	C11 TMB	Vandalism against equipment	SEQ-096
O1	C11 TMB	Overcrowding	SEQ-event2
O2	C11 TMB	Overcrowding	SEQ-event3
B14	C05 STIB	Blocking	1.1.01
B15	C05 STIB	Blocking	1.1.02
B16	C05 STIB	Blocking	1.1.03
B17	C05 STIB	Blocking	1.1.04
B18	C05 STIB	Blocking	1.1.05

Table 2: The original sequence label and type of the acted behaviours in the demonstrator sequences.

4 Validation Results and Analysis

The results of the formal validation are presented in this section. An analysis of the results is also given.

4.1 Behaviour Recognition Validation Results

The response of the ADVISOR demonstrator to the input sequences are presented in the following tables. The results are tabulated by camera. The validation was conducted over two days and the input sequence was started at 14:42. The system crashed at 15:56 and was restarted at frame 9300, without clearing the archive. The sequences then ran until 17:26 when it was paused. The system was then restarted, again without clearing the archive, at 09:42 the following morning and ran until it crashed at 10:56. The system was restarted at 11:45 and was terminated at 12:29.

Overall, the system was validated for over four hours using three recorded cameras and one live camera, giving a total of more than 16 hours of validation.

In the cases where multiple reports of the same event are generated, either because of overlapping reports or through sequential reports of the same incident, the combination of the reports is recorded. Overlapping reports occur because the algorithms are designed to associate similar behaviour reports together. For example, if fighting is detected, then ends and then starts again in the same location two seconds later, two separate reports are generated with the same identity number, but the second report has the start time of the first report.

4.1.1 Validation Results for Camera C01

The actual response of the ADVISOR system to the demonstrator sequence for camera SGFM C01 is given in Table 3 below.

Alarm Code	Ground-truth for C01	Actual response for C01	Percentage Accuracy
F1	Fighting 14:44:13 to 14:44:19	Fighting 14:44:15 to 14:44:19	100%
F2	Fighting 14:47:25 to 14:47:39	Fighting 14:47:33 to 14:47:39	79%
F4	Fighting 16:28:58 to 16:29:10	Fighting 16:29:04 to 16:29:12	92%
F6	Fighting 17:01:04 to 17:01:12	Fighting 17:01:11 to 17:01:15	75%
F7	Fighting 17:14:00 to 17:14:22	Fighting 17:14:21 to 17:14:25	27%
F8	Fighting 17:25:48 to 17:25:58	Fighting 17:25:57 to 17:26:02	60%
Paused			
F9	Fighting 10:01:08 to 10:01:13	Fighting 10:01:08 to 10:01:09	20%

Alarm Code	Ground-truth for C01	Actual response for C01	Percentage Accuracy
F10	Fighting 10:23:10 to 10:23:18	Fighting 10:23:17 to 10:23:19	75%
F11	Fighting 10:31:02 to 10:31:14	Not detected	0%
F12	Fighting 10:46:00 to 10:46:18	Fighting 10:46:14 to 10:46:23	50%
F13	Fighting 11:51:08 to 11:51:20	Fighting 11:51:20 to 11:51:22	42%
F14	Fighting 12:10:12 to 12:10:18	Fighting 12:10:19 to 12:10:23	67%
F15	Fighting 12:21:18 to 12:21:26	Fighting 12:21:27 to 12:21:28	50%
F1	Fighting 12:24:40 to 12:24:46	Fighting 12:24:45 to 12:24:50	100%

Table 3: Validation Actual Data for camera C01.

4.1.2 Analysis of Results for Camera C01

The results of the validation for camera SGFM C01 show that out of 15 different fighting events, 14 were correctly detected, a success rate of 93%. Of these 14 correctly identified fighting alarms, the reports were found to be 53% accurate in the timing and duration of the alarm report. Note that this value is subject to the human interpretation of when fighting begins, which is not always clear. For example two people might begin fighting by pushing each other, so it is unclear if fighting has begun at that point or when they actually start trading blows.

4.1.3 Validation Results for Camera C11

The actual response of the ADVISOR system to the demonstrator sequence for camera SGFM C11 is given in Table 4 below.

Alarm Code	Ground-truth for C11	Actual response for C11	Percentage Accuracy
O1	Overcrowding, 14:54:38 to 14:55:45	Overcrowding, 14:54:38 to 14:55:45	100%
O2	Overcrowding, 15:57:25 to 15:59:35	Overcrowding, 15:57:25 to 15:59:35	100%
J1	Jumping over barrier, 16:03:02	Not detected	0%
J1	Jumping over barrier, 16:03:25	Jumping over barrier 16:03:25	100%
J1	Jumping over barrier, 16:03:45	Jumping over barrier 16:03:45	100%

Alarm Code	Ground-truth for C11	Actual response for C11	Percentage Accuracy
J1	Jumping over barrier, 16:04:26	Jumping over barrier 16:04:26	100%
J1	Jumping over barrier, 16:04:47	Jumping over barrier 16:04:47	100%
J1	Jumping over barrier, 16:05:07	Jumping over barrier 16:05:07	100%
V1	Vandalism against equipment, 16:11:50 to 16:13:16	Vandalism against equipment, 16:12:13 to 16:13:16	73%
J3	Jumping over barrier, 16:20:23	Not detected	0%
J3	Jumping over barrier, 16:20:46	Jumping over barrier 16:20:46	100%
J3	Jumping over barrier, 16:21:12	Jumping over barrier 16:21:12	100%
J3	Jumping over barrier, 16:21:35	Jumping over barrier 16:21:35	100%
J3	Jumping over barrier, 16:21:56	Jumping over barrier 16:21:56	100%
J3	Jumping over barrier, 16:22:17	Jumping over barrier 16:22:17	100%
O1	Overcrowding, 16:37:10 to 16:38:17	Overcrowding, 16:37:10 to 16:37:25 16:37:47 to 16:38:17	67%
O2	Overcrowding, 16:46:30 to 16:48:40	Overcrowding, 16:46:32 to 16:46:49 16:48:13 to 16:48:34	32%
J2	Jumping over barrier, 16:55:54	Not detected	0%
J2	Jumping over barrier, 16:56:23	Jumping over barrier, 16:56:23	100%
J2	Jumping over barrier, 16:56:43	Jumping over barrier, 16:56:43	100%
J2	Jumping over barrier, 16:57:04	Jumping over barrier, 16:57:04	100%
J2	Jumping over barrier, 16:57:24	Jumping over barrier, 16:57:24	100%
J2	Jumping over barrier, 16:57:46	Jumping over barrier, 16:57:46	100%
V3	Vandalism against equipment, 17:04:22 to 17:07:16	Vandalism against equipment, 17:05:40 to 17:07:16	52%
V1	Vandalism against equipment,	Vandalism against equipment, 17:18:11	70%

Alarm Code	Ground-truth for C11	Actual response for C11	Percentage Accuracy
	17:17:45 to 17:19:12	to 17:19:12	
		Paused	
O1	Overcrowding, 09:45:32 to 09:45:55	Overcrowding, 09:45:32 to 09:45:55	100%
O1	Overcrowding, 09:55:14 to 09:55:38	Overcrowding, 09:55:14 to 09:55:18 09:55:23 to 09:55:38	79%
O2	Overcrowding, 10:09:26 to 10:11:35	Overcrowding, 10:09:26 to 10:09:30 10:09:33 to 10:09:50 10:11:13 to 10:11:35	80%
J3	Jumping over barrier, 10:17:52	Not detected	0%
J3	Jumping over barrier, 10:18:20	Jumping over barrier, 10:18:20	100%
J3	Jumping over barrier, 10:18:45	Jumping over barrier, 10:18:45	100%
J3	Jumping over barrier, 10:19:08	Jumping over barrier, 10:19:08	100%
J3	Jumping over barrier, 10:19:29	Jumping over barrier, 10:19:29	100%
J3	Jumping over barrier, 10:19:50	Jumping over barrier, 10:19:50	100%
V1	Vandalism against equipment, 10:26:25 to 10:27:52	Vandalism against equipment, 10:26:49 to 10:27:52	72%
V3	Vandalism against equipment, 10:36:58 to 10:39:42	Vandalism against equipment, 10:37:25 to 10:39:31	84%
J1	Jumping over barrier, 11:54:00	Jumping over barrier, 11:54:00	100%
J1	Jumping over barrier, 11:54:18	Jumping over barrier, 11:54:18	100%
J1	Jumping over barrier, 11:54:35	Not detected	0%
J1	Jumping over barrier, 11:54:56	Jumping over barrier, 11:54:56	100%
J1	Jumping over barrier, 11:55:19	Jumping over barrier, 11:55:19	100%
J1	Jumping over barrier, 11:55:37	Jumping over barrier, 11:55:37	100%

Alarm Code	Ground-truth for C11	Actual response for C11	Percentage Accuracy
V1	Vandalism against equipment, 12:05:19 to 12:06:45	Vandalism against equipment, 12:05:43 to 12:06:45	72%
J3	Jumping over barrier, 12:12:46	Jumping over barrier, 12:12:46	100%
J3	Jumping over barrier, 12:13:16	Jumping over barrier, 12:13:16	100%
J3	Jumping over barrier, 12:13:43	Jumping over barrier, 12:13:43	100%
J3	Jumping over barrier, 12:14:05	Jumping over barrier, 12:14:05	100%
J3	Jumping over barrier, 12:14:26	Jumping over barrier, 12:14:26	100%
J3	Jumping over barrier, 12:14:46	Jumping over barrier, 12:14:46	100%
J2	Jumping over barrier, 12:26:55	Jumping over barrier, 12:26:55	100%
J2	Jumping over barrier, 12:27:25	Jumping over barrier, 12:27:25	100%
J2	Jumping over barrier, 12:27:45	Jumping over barrier, 12:27:45	100%
J2	Jumping over barrier, 12:28:05	Jumping over barrier, 12:28:05	100%
J2	Jumping over barrier, 12:28:25	Jumping over barrier, 12:28:25	100%
J2	Jumping over barrier, 12:28:47	Jumping over barrier, 12:28:47	100%

Table 4: Validation Actual Data for camera C11.

4.1.4 Analysis of Results for Camera C11

The two overcrowding alarms were consistently detected, with an overall accuracy of 80% over seven separate instances of the behaviour. The ground-truthing of an overcrowding alarm is somewhat subjective since it is not exactly obvious at which point the scene becomes overcrowded. In addition, if there is a stationary group of people triggering the alarm and they spread out slightly and move back together then, again, it is unclear exactly when the overcrowding stops and starts again. For this reason, the accuracy of the overcrowding alarms should be regarded as approximate.

There are three separate sequences of jumping over the barrier, with each sequence containing six individual instances of the event. In 42 separate instances of jumping over the barrier, the behaviour is detected 37 times, giving a success rate of 88%. This figure includes repeated instances of the same as events since sometimes the same event was not detected and sometimes it was. It can also be seen from the results that, at the beginning of the validation, 5 out of the 6 individual instance were detected, but this increased to 6 out of 6 later in the scenario. This is due to the behaviour algorithms being able to change their thresholds over time.

The vandalism against equipment scenarios contained an actor repeatedly going to a piece of

equipment and attempting to break it open. As people approach he moves away from the equipment and returns to it later. The system recognises this as one prolonged act of vandalism rather than several individual acts and, therefore, has been scored as such. The two sequences of vandalism are consistently detected with an overall accuracy of 71%. This figure is lowered by the fact that the first instance of vandalism is not detected but the second instance is. If there is a relatively large gap between instances of vandalism then the report is scored as being wrong, when in fact nothing is happening. However, since the acts of vandalism are reported as one act, this is still the best way of scoring the accuracy. Hence the algorithm should be regarded as more accurate than the score suggests.

4.1.5 Validation Results for Camera C10

The actual response of the ADVISOR system to the live input for camera SGFM C10 is given in Table 5 below.

Alarm Code	Ground-truth for C10	Actual response for C10	Percentage Accuracy
N/A	Overcrowding, 14:32:45 to 14:33:12	Overcrowding, 14:32:45 to 14:33:12	100%
N/A	Overcrowding, 14:33:17 to 14:33:32	Overcrowding, 14:33:17 to 14:33:32	100%
N/A	Overcrowding, 14:33:33 to 14:33:57	Overcrowding, 14:33:33 to 14:33:57	100%
N/A	Overcrowding, 14:39:28 to 14:39:33	Overcrowding, 14:39:28 to 14:39:33	100%
N/A	Overcrowding, 14:43:19 to 14:43:28	Overcrowding, 14:43:19 to 14:43:28	100%
N/A	Overcrowding, 14:46:47 to 14:47:05	Overcrowding, 14:46:47 to 14:47:05	100%
N/A	Overcrowding, 14:47:22 to 14:47:27	Overcrowding, 14:47:22 to 14:47:27	100%
N/A	Overcrowding, 14:54:39 to 14:54:47	Overcrowding, 14:54:39 to 14:54:47	100%
N/A	Overcrowding, 14:54:49 to 14:55:00	Overcrowding, 14:54:49 to 14:55:00	100%
N/A	Overcrowding, 15:55:30 to 15:55:49	Overcrowding, 15:55:30 to 15:55:49	100%
N/A	Overcrowding, 15:55:11 to 15:55:14	Overcrowding, 15:55:11 to 15:55:14	100%
N/A	Overcrowding, 16:08:12 to 16:08:21	Overcrowding, 16:08:12 to 16:08:21	100%
N/A	Overcrowding, 16:11:51 to 16:11:57	Overcrowding, 16:11:51 to 16:11:57	100%
N/A	Overcrowding, 16:12:10 to 16:12:23	Overcrowding, 16:12:10 to 16:12:23	100%

Alarm Code	Ground-truth for C10	Actual response for C10	Percentage Accuracy
N/A	Overcrowding, 16:12:33 to 16:12:49	Overcrowding, 16:12:33 to 16:12:49	100%
N/A	Overcrowding, 16:27:38 to 16:27:49	Overcrowding, 16:27:38 to 16:27:49	100%
N/A	Overcrowding, 16:44:06 to 16:44:51	Overcrowding, 16:44:06 to 16:44:51	100%
N/A	Overcrowding, 16:58:38 to 16:59:02	Overcrowding, 16:58:38 to 16:59:02	100%
N/A	Overcrowding, 17:06:15 to 17:06:28	Overcrowding, 17:06:15 to 17:06:28	100%
N/A	Overcrowding, 17:14:09 to 17:14:45	Overcrowding, 17:14:09 to 17:14:45	100%
N/A	Overcrowding, 17:14:55 to 17:15:05	Overcrowding, 17:14:55 to 17:15:05	100%
N/A	Overcrowding, 17:18:49 to 17:18:57	Overcrowding, 17:18:49 to 17:18:57	100%
N/A	Overcrowding, 17:19:38 to 17:19:53	Overcrowding, 17:19:38 to 17:19:53	100%
N/A	Overcrowding, 17:23:32 to 17:23:37	Overcrowding, 17:23:32 to 17:23:37	100%
N/A	Overcrowding, 17:23:39 to 17:23:53	Overcrowding, 17:23:39 to 17:23:53	100%
N/A	Overcrowding, 17:26:37 to 17:26:52	Overcrowding, 17:26:37 to 17:26:52	100%
Paused			
N/A	Overcrowding, 09:58:18 to 09:58:29	Overcrowding,09:58:18 to 09:58:29	100%
N/A	No events	Blocking, 11:56:56 to 11:56:56	0%
N/A	Overcrowding, 12:01:32 to 12:01:47	Overcrowding, 12:01:32 to 12:01:47	100%

Table 5: Validation Actual Data for camera C10.

4.1.6 Analysis of Results for Camera C10

The camera SGFM C10 provided the live feed from the Sagrada Familia station into the ADVISOR demonstrator. The camera was situated in the main hall and overlooked the escalator from one of the platforms. Therefore, during busy periods, a large number of people disembarked from the train, up the escalator and into the field of view of the camera. The relatively high density of people caused the ADVISOR system to trigger an 'overcrowding' alarm. This is demonstrated by the fact that many such alarms were triggered on the busy Friday afternoon, whereas only two were generated on the much quieter Saturday morning.

In truth, the station was not really overcrowded, but that does not mean that the alarms were false alarms, i.e. reporting overcrowding when there was none. Instead, the high number of overcrowding alarms suggests that the thresholds for overcrowding need to be adjusted to make the system less sensitive to this scenario. Therefore, the overcrowding alarms have been scored as being correct because they were generated by a relatively high density of people emerging from the escalator after getting off a train.

One false blocking report was generated during the validation, when there was only one person standing by the exit barriers. At least two people are required to be blocking a designated area to constitute a blocking event. No other behaviours, such as fighting or vandalism against equipment were observed during the validation.

4.1.7 Validation Results for Camera C05

The actual response of the ADVISOR system to the live input for camera YZER C05 is given in Table 6 below.

Alarm Code	Ground-truth for C05	Actual response for C05	Percentage Accuracy
B18	Blocking, 14:57:20 to 14:59:15	Blocking, 14:57:21 to 14:58:36 14:58:48 to 14:58:53 14:59:11 to 14:59:16	74%
B15	Blocking, 16:05:45 to 16:07:12	Not detected	0%
F25	Fighting, 16:18:20 to 16:18:26	Fighting, 16:18:22 to 16:18:27	100%
B16	Blocking, 16:29:52 to 16:31:50	Blocking, 16:29:52 to 16:31:26	80%
B17	Blocking, 16:43:09 to 16:44:53	Not detected	0%
F26	Fighting, 16:58:32 to 16:58:42	Fighting, 16:58:37 to 16:58:38	60%
F27	Fighting, 17:11:50 to 17:11:58	Fighting, 17:11:58 to 17:11:59	63%
Paused			
B18	Blocking, 09:42:30 to 09:45:54	Blocking, 09:42:34 to 09:42:50 09:43:00 to 09:43:18 09:43:30 to 09:43:38 09:43:48 to 09:43:58	28%
B14	Blocking, 09:58:00 to 10:00:22	Blocking, 09:58:05 to 09:59:08 09:59:46 to 10:00:11	65%

Alarm Code	Ground-truth for C05	Actual response for C05	Percentage Accuracy
F24	Fighting, 10:13:35 to 10:13:48	Fighting, 10:13:40 to 10:13:46	85%
B15	Blocking, 10:24:07 to 10:25:40	Blocking, 10:24:15 to 10:25:27	52%
F25	Fighting 10:43:28 to 10:43:35	Fighting 10:43:28 to 10:43:36	100%
B16	Blocking, 11:52:00 to 11:54:10	Blocking, 11:52:09 to 11:53:45	78%
F26	Fighting, 12:03:14 to 12:03:22	Fighting, 12:03:20 to 12:03:21	75%
B17	Blocking 12:25:02 to 12:26:57	Blocking 12:25:07 to 12:25:27 12:25:39 to 12:25:55 12:26:51 to 12:26:57	41%

Table 6: Validation Actual Data for camera C05.

4.1.8 Analysis of Results for Camera C05

For camera C05, the six different fighting alarms were all detected. Of these six alarms the accuracy of the report, regarding timing and duration was found to be 81%. Of the five different blocking alarms, events B14, B16 and B18 were detected each time. Events B15 and B17 were both detected only once out of two times. This is because the algorithms are dynamic and change their thresholds during the sequences. In addition, any missing frames may affect the performance of the algorithm detection. Overall, out of a total of nine blocking events, seven events were detected giving detection rate of 78%. Of these seven events, the reports were found to be 60% accurate on average.

4.1.9 Overall Analysis of Behaviour Recognition Functionality

In total, out of 21 fighting incidents in all the demonstrator sequences, 20 alarms were correctly generated, giving a detection rate of 95%. These twenty correctly identified alarms had an average accuracy of 68%. Out of nine blocking incidents, seven alarms were generated, giving a detection rate of 78%. These seven alarms were found to be 60% accurate on average. Out of 42 instances of jumping over the barrier, including repeated incidents, the behaviour was detected 37 times, giving a success rate of 88%. The two sequences of vandalism were always detected with an overall accuracy of 71%, over six instances of vandalism. Finally, the two overcrowding alarms in C11 were consistently detected, with an overall accuracy of 80% over 7 separate instances of the alarms. The overcrowding alarms were also consistently detected in C10, with some 28 separate events being detected. The thresholds for this particular detection may need to be adjusted, as explained previously.

No false alarms were generated during the playback of the recorded sequences although one false blocking alarm was generated in the live input.

4.2 Archive Search Validation Results and Analysis

The archive was tested when the system was paused at 17:26. Therefore the search results

contain only those alarms and events generated before this time. The results of the archive search validation are presented in Table 7 as follows.

Behaviour Searched	Expected number of results	Actual number of results
Fighting	6 alarms from SGFM C01 3 alarms from YZER C05	8 alarms (2 alarms repeated from the same events) from SGFM C01 3 alarms from YZER C05
Blocking	3 alarms from YZER C05	17 alarms from YZER C05 (1st alarm occurred 3 times, 2nd alarm occurred 7 times, 3rd alarm occurred 7 times)
Vandalism Against Equipment	3 alarms from SGFM C11	7 alarms from SGFM C11 (1st alarm occurred 2 times, 2nd alarm occurred 2 times, 3rd alarm occurred 3 times)
Jumping Over Barrier	15 alarms from SGFM C11	15 alarms from SGFM

Table 7: The validation results of the archive search functionality.

The results in Table 7 show that the archive stores multiple reports of what are observed to be single events, especially in the cases where there are gaps in the behaviour or in the reports of the behaviour. As stated in the previous section, the results from the behaviour reporting are combined to give the overall ADVISOR report. However, the reports are stored individually in the archive, since, when a report is stored it is not known if there will be a further report which supersedes the previous one.

In summary, all the alarms were correctly searched and found in the archive.

4.3 Archive Retrieval Validation Results and Analysis

A selection of behaviour alarms stored in the archive was retrieved and the functionality of the playback options tested via the HCI controls. The results of the validation are presented in Table 8.

Functionality	Functionality Confirmed
Start the forward play of an image sequence	Yes
Start the reverse play of an image sequence	Yes
Stop the playback of an image sequence	Yes
Pause the playback of an image sequence	Yes
Fast forward the playback of an image sequence	Yes
Fast reverse the playback of an image sequence	Yes
Jump to the end of an image sequence	Yes
Jump to the beginning of an image sequence	Yes
Play beyond the beginning/end of an image sequence	Yes

Table 8: The validation results of the archive retrieval functionality.

In summary, the retrieval process, including the various playback options, complied with the specification [1].

4.4 Independent Validation of the Behaviour Algorithms

The behaviour recognition algorithms were also independently validated by INRIA. The algorithms were tested outside the ADVISOR system and were not tested in real-time. This means that the algorithms would not have to cope with any missing frames or processing lags. The algorithms were scored using a simple binary comparison of whether or not the behaviour was detected. There is no information on the accuracy of the detection. The results of the validation are given in the following table.

The algorithms were also tested for camera C12, which was not included in the formal validation, since only four cameras could be processed at one time.

SEQUENCES			ALARMS				
	C12 TMB	Time sec	Blocking	Fighting	Jumping over barrier	Vandalism	Overcrowding
B1	SEQ-046	99	Detected	None	None	None	None
B2	SEQ-047	114	Detected	None	None	None	None
B3	SEQ-048	86	Detected	None	None	None	None
B4	SEQ-049	106	Detected	None	None	None	None
B5	SEQ-050	59	Detected	None	None	None	None
B6	SEQ-051	96	Detected	None	None	None	None
B7	SEQ-052	62	Detected	None	None	None	None
B8	SEQ-053	70	Detected	None	None	None	None
B9	SEQ-054	70	Detected	None	None	None	None
B10	SEQ-055	58	Detected	None	None	None	None
B11	SEQ-056	59	Detected	None	None	None	None
B12	SEQ-057	68	Detected	None	None	None	None
B13	SEQ-058	73	None	None	None	None	None
	C01 TMB						
F1	SEQ-042	74	None	Detected	None	None	None
F2	SEQ-043	53	None	Detected	None	None	None
F3	SEQ-044	53	None	None	None	None	None
F4	SEQ-045	54	None	Detected	None	None	None
F5	SEQ-046	49	None	None	None	None	None
F6	SEQ-047	56	None	Detected	None	None	None
F7	SEQ-048	81	None	Detected	None	None	None
F8	SEQ-049	56	None	Detected	None	None	None
F9	SEQ-050	111	None	Detected	None	None	None
F10	SEQ-051	106	None	Detected	None	None	None
F11	SEQ-052	101	None	Detected	None	None	None
F12	SEQ-053	105	None	Detected	None	None	None
F13	SEQ-054	119	None	Detected	None	None	None
F14	SEQ-055	106	None	Detected	None	None	None
F15	SEQ-056	84	None	Detected	None	None	None
	C12 TMB						
F16	SEQ-061	34	None	Detected	None	None	None
F17	SEQ-062	43	None	Detected	None	None	None
F18	SEQ-063	52	None	Detected	None	None	None
F19	SEQ-064	43	None	None	None	None	None
F20	SEQ-065	50	None	None	None	None	None
F21	SEQ-066	31	None	Detected	None	None	None
F22	SEQ-067	52	None	Detected	None	None	None
F23	SEQ-068	36	None	Detected	None	None	None
	C11 TMB						
J1	SEQ-043	128.8	None	None	Detected	None	None
J2	SEQ-044	135	None	None	Detected	None	None
J3	SEQ-045	142	None	None	Detected	None	None
	C11 TMB						
V1	SEQ-094	129	None	None	None	Detected	None
V2	SEQ-095	158	None	Detected	None	Detected	None
V3	SEQ-096	197	None	None	None	Detected	None

	C11 TMB						
O1	**SEQ-event2**	210	None	None	None	None	**Detected**
O2	**SEQ-event3**	270	None	None	None	None	**Detected**
	C12 TMB						
G1	**SEQ-023**	240	None	None	None	None	None
	C11 TMB						
G2	**SEQ-023**	240	None	None	None	None	None
	C01 TMB						
G3	**SEQ-115**	561	None	None	None	None	None
	C05 STIB						
F24	**2.1.01**		None	**Detected**	None	None	None
F25	**2.1.02**		None	**Detected**	None	None	None
F26	**2.1.03**		None	**Detected**	None	None	None
F27	**2.1.04**		None	**Detected**	None	None	None
	C05 STIB						
B14	**1.1.01**		**Detected**	None	None	None	None
B15	**1.1.02**		**Detected**	None	None	None	None
B16	**1.1.03**		**Detected**	None	None	None	None
B17	**1.1.04**		**Detected**	None	None	None	None
B18	**1.1.05**		**Detected**	None	None	None	None

Table 9: Results of the Validation of the Behaviour Algorithms.

Out of 17 blocking incidents, 16 incidents were correctly identified, giving a success rate of 94%. The fighting behaviour was detected in 23 out of 27 incidents, giving a detection rate of 85%. Overcrowding was detected in the in the two incidents and vandalism was detected all three times. However, one of the vandalism incidents also produced a false fighting alarm. All three jumping over the barrier sequences were detected. Finally, the general sequences of people moving around normally, with no behaviours, did not generate any false alarms.

5 Summary and Conclusions

In this report, the ADVISOR Demonstrator system has been validated against the functional specification [1]. The detection of the behaviours of fighting, blocking, overcrowding, jumping over the barrier and vandalism was investigated. The system was validated at the Sagrada Familia Metro station in Barcelona, where the Demonstrator was taken for evaluation, demonstration and validation purposes.

In response to the functional specification [1] the following requirements were met during the validation.

- Demonstrator startup, including implementation of configuration files
- Demonstrator close down
- Source material requirements
- Capture module functionality
- Machine vision processing functionality, comprising recognition of specified behaviours
- Archive, Search and Retrieval functionality
- Human Computer Interface (HCI) functionality

The functions of behaviour detection, archive search and archive retrieval were validated explicitly. Implicit in the validation of the system was the testing of the HCI, Capture and Motion Detector functionality.

The first comment to be made is on the robustness of the ADVISOR system as a whole. The system was on-site at the Sagrada Familia metro station for two weeks and was used for a number of presentations to invited guests, evaluations by operators and validation. The system processed four input channels in parallel, including one live channel. Overall, the system was very robust. The one exception to this was a software bug in the behaviour recognition module which sometimes caused the system to crash after about one and a half hours.

In total, out of 21 fighting incidents in all the demonstrator sequences, 20 alarms were correctly generated, giving a very good detection rate of 95%. These twenty correctly identified alarms had an average accuracy of 68%. Out of nine blocking incidents, seven alarms were generated, giving a detection rate of 78%. These seven alarms were found to be 60% accurate on average. Out of 42 instances of jumping over the barrier, including repeated incidents, the behaviour was detected 37 times, giving a success rate of 88%. The two sequences of vandalism were always detected with an overall accuracy of 71%, over six instances of vandalism. Finally, the two overcrowding alarms in camera C11 were consistently detected, with an overall accuracy of 80% over 7 separate instances of the alarms. The overcrowding alarms were also consistently detected in camera C10, with some 28 separate events being detected.

No false alarms were generated during the playback of the recorded sequences although one false blocking alarm was generated in the live input. In conclusion, the algorithms responded very successfully to the input data, with high detection rates and with all the reports being above roughly 70% accurate.

A separate validation of the behaviour algorithms was performed by INRIA, outside of the ADVISOR system, but not in real-time with multiple channels. The following results were obtained. Out of 17 blocking incidents, 16 incidents were correctly identified, giving a success rate of 94%. The fighting behaviour was detected in 23 out of 27 incidents, giving a detection rate of 85%. Overcrowding was detected in the two incidents and vandalism was detected all three times. However, one of the vandalism incidents also produced a false fighting alarm. Finally, all three jumping over the barrier sequences were detected.

The archive search and retrieval functionality was found to work as specified, through the controls provided on the HCI.

It should be noted that the behaviour recognition algorithms worked very well against the behaviours in the sequences. It is very difficult to know how well the system would react to real behaviours, especially fighting and vandalism, without the availability of real data. This is obviously very difficult data to acquire.

Finally, it is worth saying that the invited guests and evaluators were all impressed with the ADVISOR demonstrator and with its potential as a future product to improve security and safety in Metro stations. Furthermore, the system could have a wide-range of applications to detect and respond to human behaviours in many different settings.

5.1 Other comments on the operation of ADVISOR

Whilst validating the ADVISOR system, a few problems and suggested improvements came to light. Firstly, it was found that after prolonged use, that is more than an hour and a half, the archive began to overflow. It was discovered that this was due to the deficiency of the MS Windows file system. When a directory has more than 26,000 files in it, the reading and writing of files takes a long time, so that the computer cannot write 10 files a second to its disk. It should be a relatively straightforward task to redesign the archive to write its files into smaller directories, for example ordered by hour or with a maximum of 10,000 files.

Secondly, there is a software bug in the behaviour algorithms that needs to be found and fixed. Again, this should be a straightforward but time consuming task.

Thirdly, when repeated alarms of the same incident are generated, a long list of alarms can be formed. In the cases where the alarms have the same identity number, it may be preferable to replace the previous alarm with the updated one. This would make the screen less cluttered with the repeated reports of the same incident. This suggestion could also be applied to the archive database, which would remove unnecessary or repeated alarms.

Finally, it was found that the system was probably physically too large. It certainly is not a very portable system at the moment. This is because the modules were distributed across six computers, five of which were stored in a rack, to maximise processor capability. Whilst on-site the performance of the processors was monitored with four cameras. It was found that the processing was not especially intensive and, so, the whole system could be transferred to three machines. One would be used for the capture and playback of the CCTV feeds, one for the processing of the algorithms, and one for the archive and HCI (which has to be separate anyway).

6 References

[1] ADVISOR-DOC-006, ADVISOR Functional Specification.

[2] ADVISOR-DOC-035, R7.2 Technical Validation of ADVISOR Test-Bed 2.

[3] ADVISOR-DOC-039, Final Evaluation Report.

7 Glossary, Terms and Abbreviations

ADVISOR	Annotated Digital Video for Intelligent Surveillance and Optimised Retrieval
CCTV	Closed-Circuit Television
HCI	Human Computer Interface
INRIA	Institut National De Recherche En Informatique Et En Automatique (The French National Institute For Research In Computer Science And Control)
SGFM	Sagrada Familia Metro station on the TMB
STIB	Société des Transports Intercommunaux de Bruxelles (Brussels Metro)
TMB	Transports Metropolitans de Barcelona (Barcelona Metro)
TRT(UK)	Thales Research & Technology (UK)
YZER	Metro station on the STIB