i-LIDS User Guide

Imagery Library for Intelligent Detection Systems

Publication No. 29/09

v2.1

i-LIDS User Guide

Imagery Library for Intelligent Detection Systems

Publication No. 29/09

v2.1

FIRST PUBLISHED APRIL 2009

Home Office Scientific Development Branch
Langhurst House
Langhurstwood Road
Horsham
RH12 4WX
United Kingdom

Telephone: +44 (0)1403 213800
Fax: +44 (0)1403 213627
E-mail: hosdb@homeoffice.gsi.gov.uk
Website: http://science.homeoffice.gov.uk/hosdb/

Foreword

The United Kingdom continues to lead the world in the deployment of CCTV technology, with recent high profile cases demonstrating the valuable contribution it makes in the fight against crime and terrorism.

It is recognised that the workload of today's CCTV operators is increasing and that the full potential of the CCTV scheme in tackling crime may not be achieved. Video Analytics (VA) systems offer a potential solution to the problem of 'operator overload' by automatically alerting operators in real time to events of interest or identifying sequences of interest to speed up post event analysis.

To help realise this potential the Home Office has developed the Imagery Library for Intelligent Detection Systems (i-LIDS), which aims to stimulate the development of VA systems.

Through the i-LIDS initiative the Home Office assesses and promotes VA development for Event Detection scenarios (e.g. illegally parked vehicles) and Object Tracking scenarios (e.g. people in airports) that are key to UK Government requirements.

The i-LIDS datasets are widely regarded as the most comprehensive of their kind and have achieved substantial recognition since their launch in 2006. Demand from manufacturers has been extremely encouraging and the evaluation programme has subsequently shown that a number of systems can meet certain Government requirements.

This revision of the user guide marks the newest edition to the i-LIDS library – a standard for the development and testing of Multiple-Camera Tracking systems. Through this new scenario, and the release of further datasets, the Home Office will continue to challenge the research and development community to create robust solutions that will, with increasing effectiveness, support the efficient and effective delivery of justice and protect the public from crime and terrorism.

Alan Pratt. CSci CPhys FInstP

Director, Home Office Scientific Development Branch

May 2008

Contents

1 About i-LIDS

The i-LIDS video library is a Government initiative which provides a benchmark to facilitate the development and selection of Video Analytics (VA) systems which meet Government requirements.

i-LIDS is produced by the Home Office Scientific Development Branch (HOSDB) in partnership with the Centre for the Protection of National Infrastructure (CPNI) and consists of CCTV video based initially on five different scenarios:

Event Detection:

- Abandoned baggage detection
- Parked vehicle detection
- Doorway surveillance
- Sterile zone monitoring

Object Tracking:

- Multiple-camera tracking

Within each Event Detection scenario, certain 'alarm events' are defined – for example, the presence of a parked vehicle in a defined zone for more than 60 seconds. Video Based Detection Systems (VBDS) are required to report an alarm when any of these events occur in the footage, with minimal false alarm reports.

For i-LIDS Object Tracking scenarios, individuals or 'targets' identified in the CCTV imagery are presented to the tracking system. Object Tracking Systems are required to track the Target through a network of cameras until the Target is either no longer present or a new Target is specified.

The video from each scenario is split into three 'datasets', two of which are made available to VA manufacturers and academics to assist the development of suitable systems. The remaining dataset is retained by HOSDB and used to verify the performance of systems.

Systems which demonstrate a sufficient level of performance in HOSDB's trials will be listed in a catalogue of approved security equipment used by Departmental Security Officers for Government procurement.

2 Licensing and Distribution

Distribution of i-LIDS datasets is restricted to VA manufacturers and relevant academic research groups. Applications for datasets cannot be accepted from other organisations or individuals not connected with VA development or evaluation. An application form and End User Licence Agreement can be found on the HOSDB i-LIDS website:

http://scienceandresearch.homeoffice.gov.uk/hosdb/

i-LIDS data remains Crown copyright. The End User Licence Agreement permits that it may be disassembled and processed in any way such as to contribute to the development of Video Analytics algorithms. It must not be redistributed to any third party. i-LIDS imagery may be used in academic, but not commercial exhibition.

3 Video Format and System Requirements

i-LIDS Event Detection datasets are distributed on 500GB USB 2/Firewire external hard drives. The Multiple Camera Tracking (MCT) datasets are distributed on 1TB USB 2/Firewire/e-Sata external hard drives. Windows NTFS or Apple Mac format drives can be provided. NTFS is recommended for Linux users.

i-LIDS video is rendered in the cross-platform Quicktime MJPEG file format. The minimum system requirements to view the footage are:

- PC or Mac
- USB2 or Firewire port
- Apple Quicktime or an equivalent emulator

Users will require an up-to-date video card for Quicktime to render the video at full frame rate although this may not be necessary unless exporting a signal to external hardware. Users requiring a composite video output may also export footage to DVD using a video editing package and generate a composite signal using a stand-alone DVD player. It is recommended to use a high quality MPEG2 encoding when exporting MJPEG rendered footage to DVD so as to avoid noticeable degradation in image quality. For some windows-based PCs video files may need to be converted to .avi format before burning to DVD.

4 Principles of i-LIDS

4.1 Scenarios

i-LIDS is based around five 'scenarios' crucial to Government requirements:

- Abandoned baggage detection
  with alarm events consisting of unattended bags on the platform of an underground station
- Parked vehicle detection
  with alarm events consisting of suspiciously parked vehicles in an urban setting
- Doorway surveillance
  with alarm events consisting of people entering and exiting monitored doorways
- Sterile zone monitoring
  with alarm events consisting of the presence of people in a sterile zone between two security fences
- Multiple-camera tracking
  with Target events consisting of people ('Targets') travelling through a network of CCTV cameras

4.2 Training, Test and Evaluation datasets

In accordance with academic convention, footage from each i-LIDS scenario is divided into three equivalent datasets:

- A public 'training' dataset which can be used to develop effective recognition algorithms
- A public 'test' dataset which can be used to verify the performance of those algorithms
- A private 'evaluation' dataset held by HOSDB and used to certify the performance of systems submitted to their regular trials

4.3 Event Detection Sequences

Each i-LIDS Event Detection dataset comprises approximately 24 hours of 'sequences' recorded in different conditions; time of day, weather, background activity level etc.

Some sequences are augmented by alarm events 'acted out' for the footage. The remaining 'non-alarm' sequences contain only a background level of alarm events.

Normally, each sequence is rendered as a single Quicktime file in the i-LIDS library. In the public training datasets, however, 'alarm' sequences containing

many alarm events are spliced into 'clips' such that each alarm event acted out is rendered to a separate file. This makes it quicker for training dataset users to access footage of specific alarm events. The file naming convention is such that adjacent clips can easily be identified and concatenated should more pre- and post- event footage be needed e.g. for learning algorithms.

Each alarm sequence contains at least five minutes of footage prior to the first scheduled alarm event so as to assist learning systems in adapting to the conditions of the sequence.

4.4 Object Tracking Sequences

The i-LIDS Multiple-camera tracking public datasets comprise approximately ten hours of 'scenario' for both training and test datasets. This MCT scenario is formed from a network of five CCTV cameras, giving a total of approximately fifty hours video imagery per dataset.

Cameras may be selected from the scenario for overlapping and non-overlapping camera fields of view, or a mixture of the two.

Systems are required to accurately track a Target through the network of CCTV cameras. Targets are defined and structured as follows:

- **Target** – An operator nominated individual
- **Target Event** – An event in which a Target is present within the CCTV imagery
- **Target Event Set** – A set of video imagery collected from a group of CCTV cameras (five in this scenario) containing multiple Target Events

4.5 Ground Truth

Each public i-LIDS Event Detection dataset is supplied with an XML based index[1] describing at a high level the content and alarm events present in each video file.

Similarly, the Object Tracking datasets are supplied with an XML based index, or *ground truth*. This is more detailed than for the Event Detection scenarios and provides data for Target size and location within the video imagery.

In addition to the raw text of the ground truth index, a front-end user interface is provided to facilitate access to requisite footage.

---

[1] Whilst considerable care is taken to ensure that every index is as accurate as possible, it should be considered that HOSDB cannot guarantee the integrity of index data.

5 i-LIDS Data

5.1 Event Detection Scenarios

For Object Tracking Scenarios (such as Multiple-Camera Tracking) see section 5.2.

5.1.1 File and Folder Structure on Hard Drives

Each i-LIDS dataset is supplied on an individual hard-drive containing the following:

- i-LIDS User Guide (User_Guide_v2.0.pdf) – this document
- i-LIDS Flyer (i-LIDS Flyer.pdf) – a one page flyer describing the i-LIDS library
- Scenario definition file (eg. Sterile Zone.pdf) – defining alarms and other attributes specific to the scenario; see section 5.1.2
- Text index (index.xml) – XML description of each video file in the dataset using the schema defined in the scenario definition
- User interface gateway (index.html) – see section 5.1.4
- User interface support files (‘index-files’ folder)
- Video (‘video’ folder) – rendered in Quicktime MJPEG format; see section 5.1.3
- Calibration stills (‘calibration’ folder) - .tif stills from each camera view used in the scenario. The HOSDB Rotakin® calibration test target is placed within each scene along with a white, right-angled triangle of sides 600 x 800 x 1000mm laid flat on the ground.
- Frame based annotation (‘annotation’ folder) – available on selected datasets only. Provided by our colleagues in the US National Institute of Standards and Technology

5.1.2 Scenario Definition File

The .pdf scenario definition contains the following information specific to the scenario:

5.1.2.1 Alarm Definitions

Describes the circumstances which constitute an ‘alarm event’ in that scenario. Several different types of alarm event may be defined, all of which should be recognised by VBDS and cause an alarm.

Each scenario typically contains footage from several fixed camera views. The alarm definitions will contain an image from each of these ‘stages’ with areas relevant to the definition of alarm events or XML markup highlighted.

5.1.2.2 XML Indexing Schema

Describes each XML element used in the index for the scenario, from 'clip' level down. The descriptive syntax uses several means to define possible element values:

- One of several discrete values eg. Time of Day - <Dawn|Day|Dusk|Night>
- One of a range of discrete values eg. Grade - <a…z>
- By format eg. Duration - <hh:mm:ss>

In all scenarios, the text index file contains, as a header, a number of high level elements describing the name of the scenario, dataset and version number.

More detail on this propriety and self contained HOSDB schema can be found in the appendices at the end of this document.

5.1.2.3 Recall Bias

VBDS may be evaluated by HOSDB for either an 'Operational Alert', or 'Event Recording' role. In the former, the system provides real-time detection of suspicious events which must be dealt with by a human controller. In the latter, the system acts as a trigger for recording of suspicious events, where all the recordings obtained are to be analysed at a later time.

HOSDB assess the performance of systems based on a criterion called the F1 measure, defined in section 6.4.1. This criterion is dependent on a parameter called the recall bias (α) which determines the influence of detection rate (recall) with respect to that of false alarm rate on the value of F1.

A higher value of recall bias is used to assess systems for the 'Event Recording' role since in this role false alarms are a less significant problem. Knowledge of the recall bias value enables manufacturers to optimise their systems for either role under HOSDB evaluation.

5.1.3 Video Folder

Contains all the video in the dataset in Quicktime MJPEG (.mov) format. For each .mov file, a matching .qtl 'reference file' is present. This is a small file used by the user interface to access the stand-alone Quicktime player and play the .mov video. Files are named according to the following nomenclature (eg. 'PVTRA301b05.mov'):

- Scenario

  AB=Abandoned Baggage

  PV=Parked Vehicle

  SZ=Sterile Zone

  DS=Doorway Surveillance.

- Dataset

  TR=Training

TE=Test

- Alarm or Non-Alarm sequence

  A=Alarm

  N=Non-alarm

- Stage ie. camera view

  1

  2

  3 etc.

- Archive Tape (of relevance to HOSDB only)

  01

  02

  03 etc.

- Sequence

  a

  b

  c etc.

- Clip (training dataset alarm sequences only)

  01

  02

  03 etc.

NB. Adjacently numbered clips provide continuous footage when concatenated.

5.1.4 User Interface

i-LIDS is provided with a web browser based user interface facilitating cross-platform search and access to requisite footage. Full user-interface functionality is assured by using a DOM Level 2 compliant web browser. The following browsers have been tested and are recommended:

- Windows: Internet Explorer 7

  Firefox 1.0.4

- MacOS X: Netscape 7.2

  Firefox 1.0.4

Internet Explorer is preferable as the clip launch process does not generate an additional browser window. The viewer (Quicktime or emulator) should be registered to handle the .qtl MIME type. This is done automatically on setup with Quicktime version 5 and later.

Although similar, the interface for the i-LIDS Event Detection scenarios has a slightly different layout and software requirement to the Multiple Camera Tracking Scenario (see section 5.2.4 for further details).

To start the user interface, launch 'index.html' from the root folder of the i-LIDS hard-drive. After a few moments, this should bring up the main user interface, similar to that shown in figure 1, below:

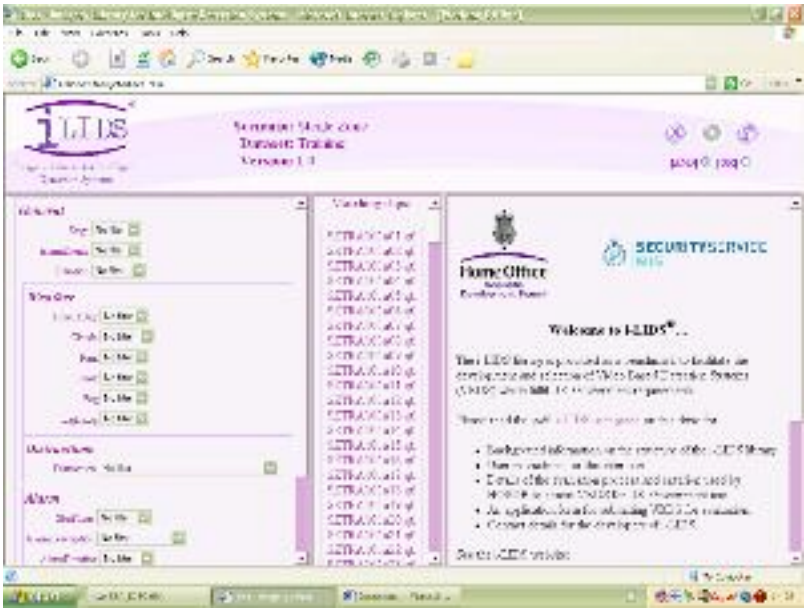

Figure 1

In the left hand pane are presented a number of combo boxes used for filtering the available footage based on the XML schema pertinent to the scenario. Each box offers the full range of field values present within the XML index for the scenario.

In the middle pane are presented a list of the video files (clips or sequences) matching any search terms selected in the left hand pane. Initially this list will contain all the video files in the dataset.

A welcome page containing a copyright notice summarising the i-LIDS licensing conditions is initially presented in the right hand pane. When a video file is selected in the middle pane, this is replaced by a formatted view of the complete index data pertaining to that file as shown in figure 2, overleaf:

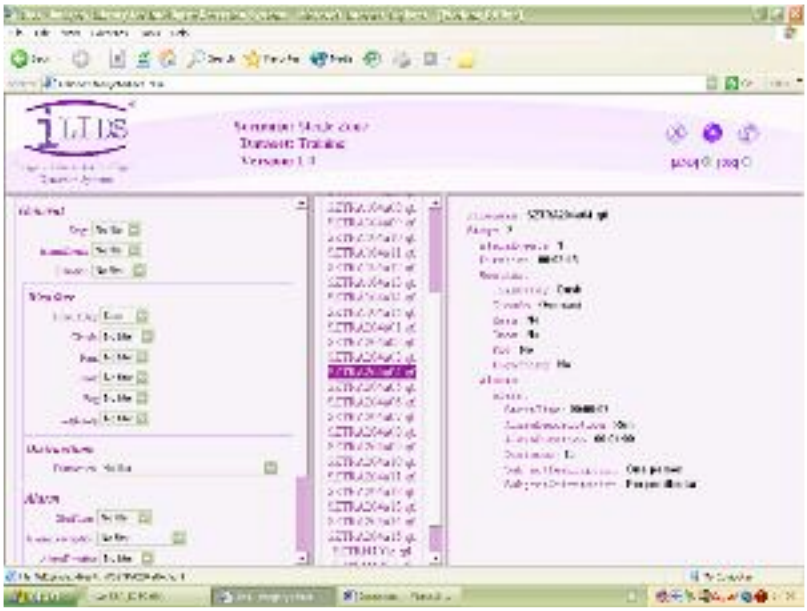

Figure 2

On the right of the banner at the top of the user interface are three icons and two radio buttons providing various controls:

Table 1: User interface controls

Control	Function
	Clears all search filters
	Launches video in Quicktime, or whichever other application is registered to handle .qtl files.
	Re-displays the welcome and copyright information in the right hand pane.
[AND] / [OR]	Radio buttons to determine whether multiple search filters should be applied with AND or OR logic. Default is AND[2].

[2] In AND search mode, the application is designed such that 'matching filter terms may not be descended from different incidences of the same element type.'
An example of this is where more than one 'alarm' field is filtered, for instance <Distance> and <Subject Description>. In this case, a video file with several alarm events will only pass the AND filter if the required <Distance> and <Subject Description> both occur within the same alarm.

5.2 Object Tracking Scenarios

5.2.1 File and Folder Structure on Hard Drives

Each i-LIDS MCT dataset is supplied on an individual hard-drive containing the following:

- i-LIDS User Guide (User_Guide_v2.0.pdf) – this document
- i-LIDS flyer (i-LIDS_Leaflet_v1.pdf) – a two page flyer describing the i-LIDS library
- MCT scenario definition (MCT_Scenario_Definition_Mar08_v1.0.pdf) – defining the tracking requirements and providing an example of the XML schema used
- User interface gateway (index.html) – see section 5.2.4
- User interface support files ('http' folder)
    - Video files ('video' folder within 'http' folder) – rendered in QuickTime MJPEG format; see section 5.2.3
    - Text index ('xml' folder within 'http' folder) – XML description of all Target Events on the dataset using the schema provided in the scenario definition
- Calibration stills ('Calibration' folder) - .jpeg stills from each camera view used in the scenario. The HOSDB Rotakin® calibration test target is placed within each scene.

5.2.2 Scenario Definition File

The .pdf scenario definition should be read in conjunction with the following information specific to the scenario:

5.2.2.1 Tracking Requirements

Describes the circumstances which constitute when a Target is required to be tracked within each camera view and thus when the tracking systems should provide an output as described in section 6.3.3.

Each Target Event Set contains footage from five fixed camera views. The Target Acquisition section contains an image from each of these cameras along with a short description of when the person counts as a valid Target.

At the end of this section, there is a schematic of the camera layout used to collect the imagery. This map is intentionally not to scale and does not include all of the scenery furniture as many sites are unlikely to hold such detailed CCTV maps.

5.2.2.2 XML Indexing Schema

The HOSDB *SABRE* annotation tool was used to create the ground truth documentation in the form of a VIPER compliant [1] XML document.

At its highest level of abstraction the Object Tracking datasets are organised hierarchically as shown below:

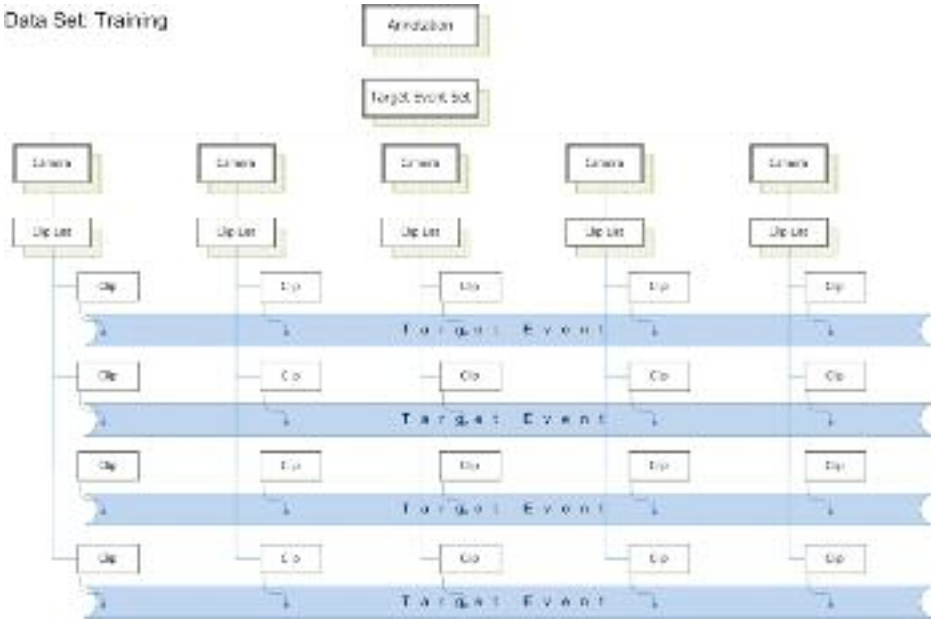

Figure 3: Hierarchical structure of training dataset

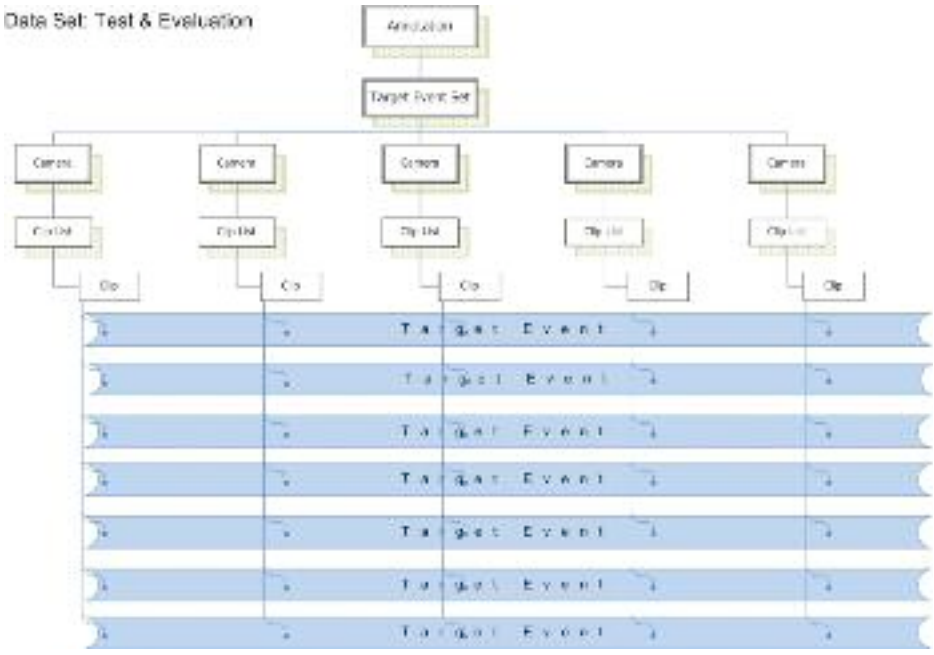

Figure 4: Hierarchical structure of test and evaluation datasets

The following terminology applies:

- A Target Event pertains to a filmed event featuring one human target on one camera.
- A Clip may contain many Target Events filmed as a single continuous piece of video for the same single camera.
- A ClipList is a simple un-ordered set of Clips for the same single camera.
- Camera pertains to a single ClipList.
- A Target Event Set contains multiple Cameras, containing ClipLists for different camera views.

XML database structure

This notional hierarchical structure is documented by a flat XML structure. Each of the high level entities is characterised by the following attributes:

Table 2:

	Attribute name	Description	Possible values
Clip object	Sequence	For internal use by SABRE	Integer generated by SABRE
	DATA-SOURCE	File name of source media	Any permissible OS value e.g. "MCTTR01a.mov"
	CAMERA	ID of camera	Text field e.g. "Customs Hall"
	Annotation	Identifies which Annotation this Clip belongs to	Integer generated by SABRE
	Target-Event-Set	Identifies which Target Event Set this Clip belongs to	Integer generated by SABRE
	Target-Event	Identifies which Target Event is documented by this Clip	Integer generated by SABRE
Annotation	NAME	Text string identifying annotation	Text field e.g. "MCTTR1"
	DATA-SET	Text string identifying whether Annotation relates to training, test or evaluation dataset	Training | Test | Evaluation
Target Event Set object	NAME	Identifies which Annotation this set belongs to	Text field e.g. "MCTTR1"
	TIME-OF-DAY	Characterises time of day for this set	Day | Dawn | Dusk | Night
	DURATION	Text string recording duration in format `hh:mm:ss`	e.g. 00:45:00 for 45 minutes
	DISRACTION	Characterises whether distracting behaviours occur during this set as a whole	(none)
Target Event object	CROWD-DENSITY	Characterises background scene density for this set as a whole	High | Medium | Low
	Target	ID of human target described by this Target Event Object	Integer generated by SABRE
Target object	NAME	Text string describing human target object	Text field e.g. "John"
	DRESS	Text description of dress code of human target	Casual | Smart
	SEX	Sex of human target	Male | Female
	COLOUR	Any obvious single colour associated with human target (e.g. jacket colour)	Text field e.g. "Red"
	BAG	Boolean denoting whether human target is	true | false

		carrying a bag	
	BOUNDING-BOX	Exterior bounding box around an unoccluded human target	VIPER `bbox` type
	OCCLUDED-BOUNDING-BOX	Exterior bounding box around the observable part of a partly obscured human target	VIPER `bbox` type
	INITIAL-BOUNDING-BOX	Exterior bounding box around a human target in a Target Event where that target first meets the minimum screen height criteria and is unoccluded	VIPER `bbox` type
	INITIAL-OCCLUDED-BOUNDING-BOX	Exterior bounding box around a human target in a Target Event where that target first meets the minimum screen height criteria but that object is part occluded	VIPER `bbox` type

Note that with the exception of the Bag attribute of Target Events and the attributes defined by VIPER `bbox` types, all data in the ground truth is described using text representations in the form of VIPER `lvalue` attributes. This is due to restrictions on available data types in the VIPER schema.

The target data is specified in the data chunk of the XML document. The most core data consists of bounding box and occlusion bounding box data for targets.

5.2.2.3 Annotation Guidelines

Tracking systems may be evaluated by HOSDB for either an 'overlapping camera' or 'mixed camera' role. The overlapping role comprises cameras 2, 3 and 4, with the mixed role including all five cameras. In both roles the systems should provide real-time XY coordinates for the Target of interest associated with the correct camera.

When the Target meets the following requirements it is annotated and should be tracked:

Camera 1:

- 100% of Target height is visible within the scene, or would be visible if not occluded by objects that are not considered to be scene exit points[3].
- Target is equal to or greater than 10% screen height (58 pixels).
- Both shoulders of the Target are visible within the scene, or would be visible if not occluded by objects that are not considered to be scene exit points.

Camera 2:

- 100% of Target height is visible within the scene, or would be visible if not occluded by objects that are not considered to be scene exit points.
- Target is equal to or greater than 10% screen height (58 pixels).
- Both shoulders of the Target are visible within the scene, or would be visible if not occluded by objects that are not considered to be scene exit points.

[3] A scene exit point is either the edge of the camera field of view or scene furniture which will occlude the target until they would otherwise reach a camera field of view extremity.

Camera 3:

- 100% of Target height is visible within the scene, or would be visible if not occluded by objects that are not considered to be scene exit points.
- Target is equal to or greater than 10% screen height (58 pixels).
- Both shoulders of the Target are visible within the scene, or would be visible if not occluded by objects that are not considered to be scene exit points.

Camera 4:

- 75% of Target height is visible within the scene, or would be visible if not occluded by objects that are not considered to be scene exit points.
- Annotatable portion of Target within scene is equal to or greater than 10% screen height (58 pixels).
- Both shoulders of the Target are visible within the scene, or would be visible if not occluded by objects that are not considered to be scene exit points.

Camera 5:

- 100% of Target height is visible within the scene, or would be visible if not occluded by objects that are not considered to be scene exit points.
- Target is equal to or greater than 10% screen height (58 pixels).
- Both shoulders of the Target are visible within the scene, or would be visible if not occluded by objects that are not considered to be scene exit points.

The Target is the only object within scene that has been annotated. Annotation does not include luggage carried or pushed by the Target, but does include anything being worn by the Target, including hats, scarves and coats that may add size to the Target.

The Target is annotated from the first frame that the Target meets the above requirements and for every fifth frame until the last frame the requirements continue to be met.

Example: The Target enters camera 2 from the bottom of the screen. At first only the Target's head is visible, thus the Target is not annotated (Frame 2147). Then both shoulders become visible, but the complete Target height is still not within scene (frame 2150). Finally, the entire Target height is visible within the scene (frame 2153) and is annotated. Thereafter every fifth frame is annotated. The Target then leaves the scene on frame 2296, making the last annotated frame in our five frame sequence for this Target event 2293.

Occluded annotation is used when 100% of the Target area is within the scene, but 50%+ of the Target is occluded from the camera view.

Initial annotation is used for the first five annotated frames of a Target Event. These annotations are intended to represent an operator selecting the Target for the first time and should be processed as such. These frames are supplied each time a new Target is designated. Initial annotation is in the same format

as any other annotation within the datasets and systems will need to interpret this information in real-time to initiate tracking of each Target.

Initial – Occluded annotation is used when the annotation meets both of the previous rules.

5.2.3 Video Folder

The video folder contains sub folders for each Target Event Set. Each of these sub folders contain further sub folders for each Target Event. These contain five video files (camera 1-5) for each Target Event in QuickTime MJPEG (.mov) format. Files are named according to the following nomenclature (e.g. MCTTR0101a.mov). The xml folder contains the same file structure as the video folder and uses the same naming convention for each file (but ends in .xml).

- Scenario

  MCT=Multiple Camera Tracking

- Dataset

  TR=Training
  TE=Test

- Target Event Set

  01
  02
  03 etc…

- Camera

  01=Duty free
  02=Left baggage
  03=Café
  04=Lift
  05=Information desk

- Target Event

  a
  b
  c etc…

NB. Linking Target Events alphabetically provides continuous footage when concatenated.

Example:

- Video (folder)
  - MCT TR 01 (folder)
    - MCTTR01a (folder)
      - MCTTR0101a.mov
      - MCTTR0102a.mov
      - MCTTR0103a.mov
      - MCTTR0104a.mov
      - MCTTR0105a.mov

- MCTTR01b (folder) etc…

- MCT TR 02 (folder) etc…

5.2.4 User Interface

For the MCT datasets there is an additional requirement for the browser to be Java compliant because the user interface uses Java embedded in an HTML page. To start the user interface, launch 'index.html' from the root folder of the i-LIDS hard-drive. The following browsers have been tested and are recommended:

- Windows: Internet Explorer 7

When loaded the user interface will look similar to that shown in figure 5, below:

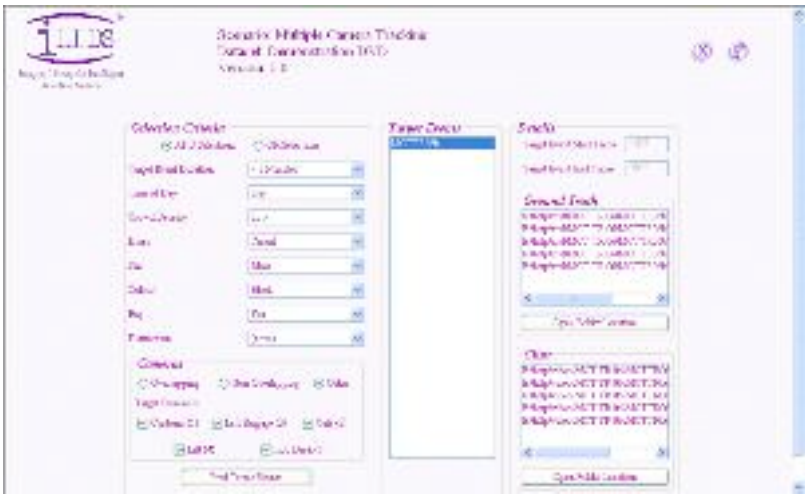

Figure 5

In the left hand pane are a number of combo boxes used for filtering the available footage based on the XML schema pertinent to the scenario. Each box offers the full range of field values present within the XML index for the scenario.

Users also have the option to select only overlapping, non-overlapping or a user defined selection of cameras. Once "Find Target Events" is selected, the filtered Target Events will be listed in the middle pane. Initially this list will be empty.

Once a Target Event is selected from the middle pane, the Details pane will update with Start and End frames and the location of the video files and XML schema for the selected Target Event. After selecting a file path, "Open Folder Location" will open Windows Explorer (or equivalent) to the relevant file location.

The two buttons in the top right hand corner have the same function as those in the Event Detection user interface (see section 5.1.4).

6 System Evaluation

6.1 UK Government VA trials

The Home Office Scientific Development Branch advises the UK Government on the effectiveness of different VA Systems based on the results of regular, scenario based i-LIDS trials.

Departmental Security Officers involved in Government procurement are notified of any systems whose performance in these trials merits recommendation for operational use in the relevant scenario. This can lead to increased revenue for the manufacturers concerned and is seen as a strong incentive to submit systems for evaluation.

Manufacturers whose systems meet the highest level of performance classification during evaluations will be entitled to use the trademarked i-LIDS logo in their trade literature, as in figure 6.

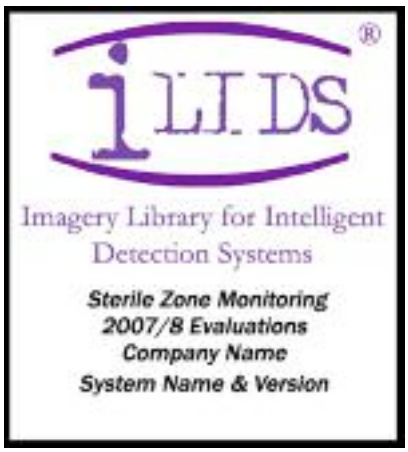

Figure 6

6.2 Applying for System Evaluation by HOSDB

Manufacturers wishing to submit a system for HOSDB evaluation should print off and fill out the 'Application for i-LIDS Evaluation' form found in the appendices to this user guide and send it to the address indicated. The application deadlines for forthcoming trials are posted on the i-LIDS web site:

http://scienceandresearch.homeoffice.gov.uk/hosdb/

The application form requires manufacturers to declare the measured performance of their system based on the F1 criterion as described in section 6.4. The reported performance should be based upon the entire test dataset for the relevant scenario, no part of which should have been used to configure the system.

The flowchart overleaf, figure 7, illustrates the end-to-end process of i-LIDS dataset procurement, system development and evaluation.

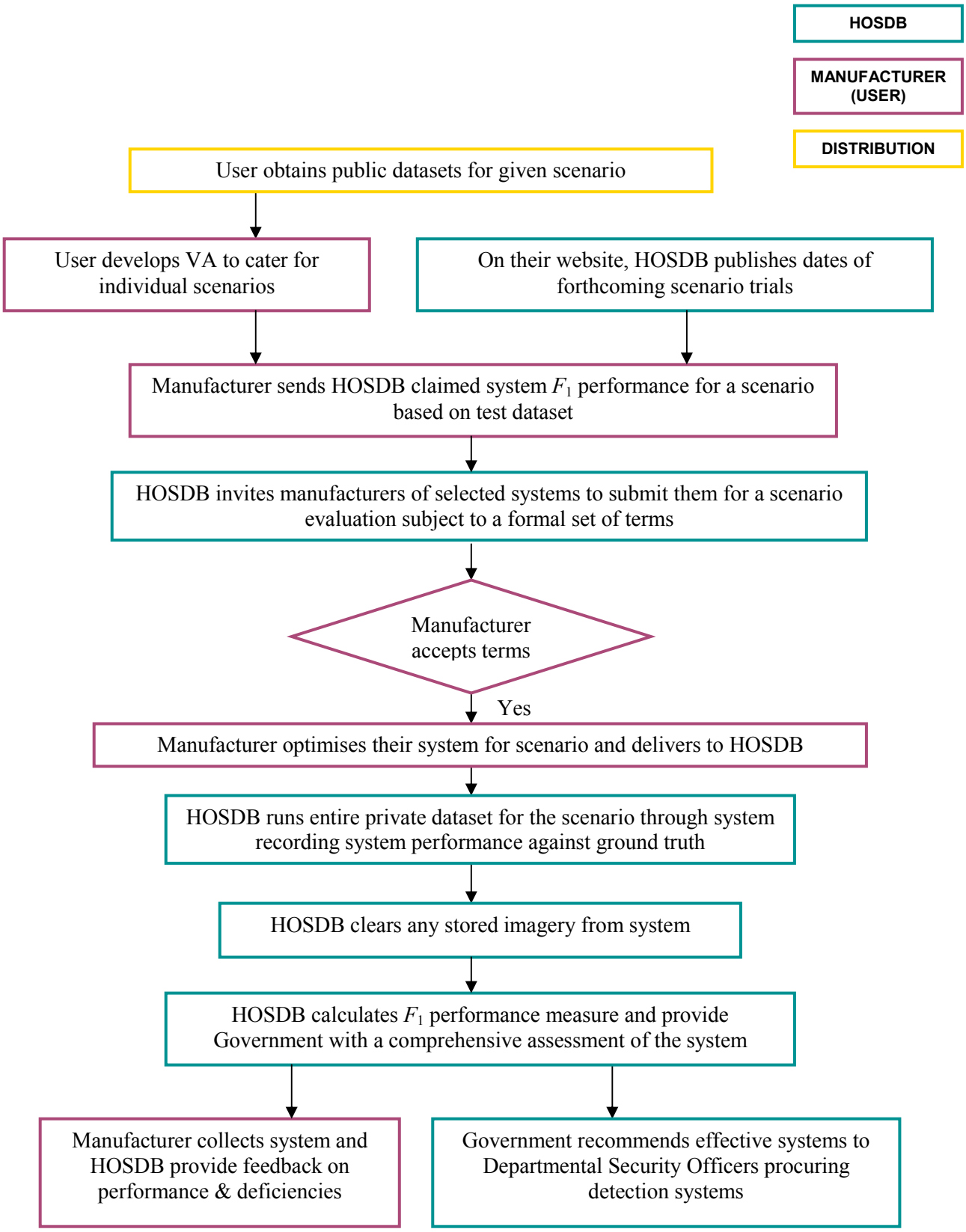
HOSDB
MANUFACTURER (USER)
DISTRIBUTION
User obtains public datasets for given scenario
User develops VA to cater for individual scenarios
On their website, HOSDB publishes dates of forthcoming scenario trials
Manufacturer sends HOSDB claimed system F_1 performance for a scenario based on test dataset
HOSDB invites manufacturers of selected systems to submit them for a scenario evaluation subject to a formal set of terms
Manufacturer accepts terms
Yes
Manufacturer optimises their system for scenario and delivers to HOSDB
HOSDB runs entire private dataset for the scenario through system recording system performance against ground truth
HOSDB clears any stored imagery from system
HOSDB calculates F_1 performance measure and provide Government with a comprehensive assessment of the system
Manufacturer collects system and HOSDB provide feedback on performance & deficiencies
Government recommends effective systems to Departmental Security Officers procuring detection systems

Figure 7

A full set of terms applicable to the HOSDB evaluation procedure can be found on the i-LIDS web site. Manufacturers must signify their consent to these in order for their systems to be accepted for evaluation.

As part of the terms of evaluation, manufacturers are required to optimise and submit their systems to HOSDB on loan at their own expense. Provision must be made for HOSDB to permanently erase any recorded footage from trialled systems. Once evaluation is complete, manufacturers will be asked to collect their systems and be given feedback on their performance.

6.3 Evaluation Procedure

6.3.1 General

To ensure they are familiar with the operation of each system loaned to them for evaluation, HOSDB staff will liaise with participating manufacturers. Manufacturers must sign an Evaluation Agreement which will contain detailed instructions for each evaluation.

Each system on trial will be presented with an interlaced PAL composite video signal (via a BNC type connector) of the entire private evaluation dataset for the relevant scenario. The video will contain short title blocks between each sequence, and there will be a break in the signal each time i-LIDS archive tapes need to be changed. Archive tapes will be presented in a random order.

6.3.2 Event Detection

A system should meet an evaluation commissioning acceptance criteria of an overall F1 score of 0.7 at the application phase. Applicants should be advised that systems must meet considerably more stringent performance levels to meet the i-LIDS performance standards for Government use

For Event Detection evaluations systems are required to indicate alarms through a relay output. Manufacturers should declare to HOSDB staff whether an open or closed circuit denotes an alarm state.

Multi-channel VBDS should be supplied to HOSDB with one channel optimised to handle each stage (camera view) used in the scenario. For single-channel systems, one system will need to be submitted for each stage.

During the title blocks and for the first five minutes of each sequence any system alarms reported will be ignored. Likewise, any alarm events present in the first five minutes of each sequence will not contribute towards the calculation of system performance.

For the remainder of each sequence the start time of any system alarms will be logged and compared to ground truth data to evaluate the number of ‘true positive’, ‘false positive’ and ‘false negative’ alarms. This comparison process is illustrated in figure 8, overleaf.

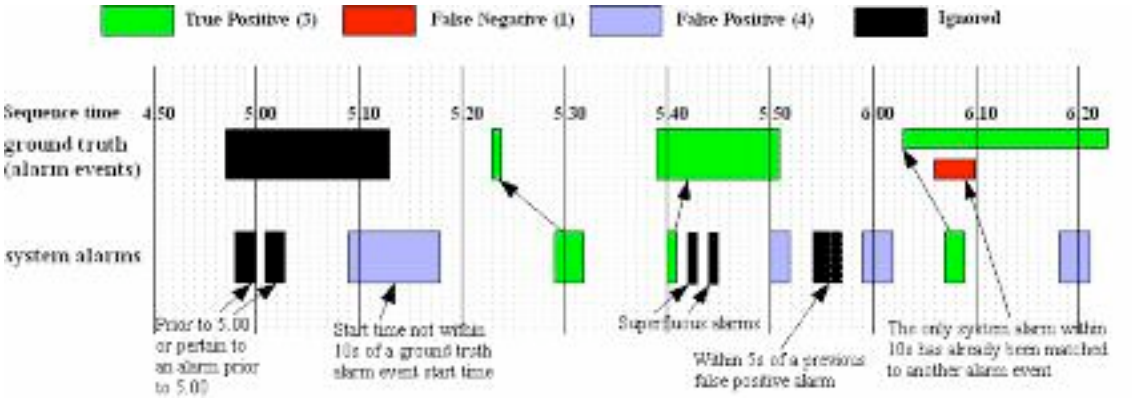

Figure 8

Systems have ten seconds to report an alarm state after an alarm event begins in the evaluation footage. During this time multiple alarm reports will be disregarded; an alarm event is either detected or not. After this ten second window, any further alarms reported will be deemed 'false positives'.
NB. Systems should NOT continue to alarm over the duration of alarm events.

Where a system false alarms several times in quick succession, only one false positive alarm will be logged every five seconds.

Where two or more alarm events occur together in the footage, systems must generate a separate alarm for each. For evaluation purposes, it is not necessary for a system to indicate the type of alarm event detected on reporting an alarm although this could be favourable for operational deployment.

6.3.3 Object Tracking

Each system will be presented with five separate frame-synchronised interlaced PAL composite video inputs (via a BNC type connector). Each video input will be from the private evaluation dataset and will contain exactly the same camera fields of view as the public test and training datasets.

The private evaluation dataset will be played out on Digital Betacam tapes. Each tape will contain two (approximately) 45 minute sequences (Target Event Sets). The footage will contain short title blocks at the start of each

Target Event Set, lasting no longer than 30 seconds. There will be a break in signal each time a tape is changed and tapes will be presented in a random order.

During the title blocks and for the first five minutes of any Target Event Set, systems are not expected to track any targets. Any targets that are tracked will be ignored and will therefore not contribute towards the final calculation of system performance.

Systems will be evaluated using an automated test system called CLAYMORE. Systems will need to be able to reliably and accurately report SMPTE [2] standard timecode information for the frames they are reporting on. The most precise method for doing this is to read the SMPTE timecode directly from the video source. This ensures that any latency in the frame

being read into a system and the processed date being written out is minimised. There are a number of commercially available VITC reader cards available. Systems are expected to interpret this timecode and use it as a timestamp for any tracked results.

CLAYMORE is designed to provide an effective and repeatable infrastructure for the testing of systems. The system architecture is shown below in diagram

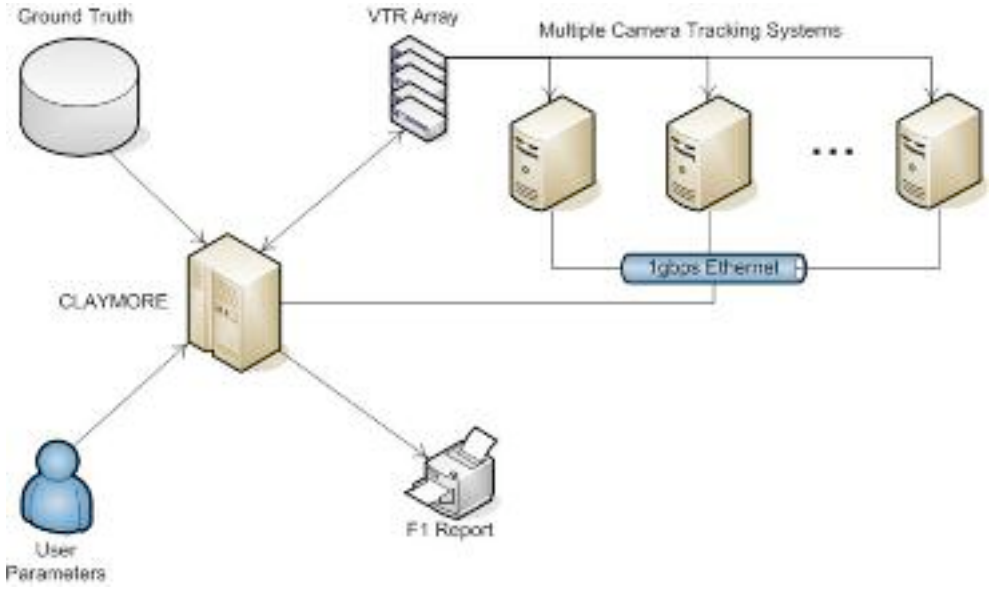

Figure 9

CLAYMORE will send messages to systems under test. These messages include:

- System check messages
- Target initialisation data

Systems will send messages to CLAYMORE. These message include:

- System check responses
- Bounding box and target ID information

The CLAYMORE test environment is design to simulate a system working in a real world type environment. The target initialisation data (which is documented in the ground truth files) is equivalent to a human operator selecting a target for tracking. During the evaluation process, CLAYMORE will send target initialisation data to systems and expect to receive in turn bounding box data and target ID data that it will then compare with the privately held ground truth data. Systems are expected to output target tracking information in a simple format that CLAYMORE can understand. The form and structure of messaging for CLAYMORE is described in a separate document available on request. These message are implemented as

simple character based messages sent over an IP link between CLAYMORE and the system under evaluation.

It is HOSDB's preference that all systems will be evaluated using the CLAYMORE infrastructure. In the event that participants are unable to modify their system to interface directly with CLAYMORE, HOSDB may, at its discretion, modify this evaluation procedure.

Systems should only provide output when the relevant Target is within a camera field of view, as described in the scenario definition and annotation guidelines in section 5.2.2.3.

This data will then be compared to the private evaluation dataset ground truth, and used to calculate the overall performance of the system using the i-LIDS Multiple Camera Tracking metric (see section 6.4.2).

6.4 Performance Metrics

6.4.1 Event Detection

VBDS performance on a scenario is rated using a weighted harmonic mean of a system's 'recall' and 'precision' known as the F1 measure; see reference [3].

Presented with a full dataset of evaluation footage under the conditions described in section 6.3, each VBDS yields a number of

(a) True positive alarms
(system alarms in response to a genuine alarm event)

(b) False positive alarms
(system alarms without the presence a genuine alarm event)

(c) False negative alarms
(genuine alarm events not resulting in a system alarm)

The recall (detection rate), $r = a / (a+c)$

The precision (probability of an alarm being genuine), $p = a / (a+b)$

$$F_1 = \frac{(\alpha + 1)rp}{r + \alpha p}$$

where α is the 'recall bias'; a weighting of recall relative to precision declared in each scenario definition (cf. section 5.1.2.3)

Subject to the agreed terms and conditions governing the evaluation process, systems demonstrating an F1 performance measure in excess of set boundaries will be recommended for practical use in the relevant scenario and role. These systems will be listed in a catalogue of approved security equipment used by purchasers in Government and other parts of the UK's critical national infrastructure.

The F1 values which must be obtained in order to qualify for practical recommendation are not made public.

6.4.2 Object Tracking

The output bounding box track from a system will be compared against the annotated ground truth for each frame. The precision and recall of the tracked bounding box when compared to the ground truth bounding box will determine if the track for that frame is a True Positive (TP), False Positive (FP) or False Negative (FN).

Precision and Recall are defined as:

$$\mathrm{Re}\,call = \frac{OP}{GTP}$$

$$\mathrm{Pr}\,ecision = \frac{OP}{TTP}$$

Where:

GTP = Total number of Ground Truth Pixels.

TTP = Total number of Tracker Pixels.

OP = Total number of overlapping pixels.

Precision and Recall will then be used in an F1 harmonic mean calculation to determine the quality of the track:

$$F1 = \frac{2(\mathrm{Re}\,call \times \mathrm{Pr}\,ecision)}{\mathrm{Re}\,call + \mathrm{Pr}\,ecision}$$

From this unweighted F1 score, a boundary cut off point will be used to determine a TP, FP or FN.

Criteria for each of these categories are:

- True Positive = F1 $\geq$ 0.25 **AND** TTP < 3 x GTP
- False Negative = F1 < 0.25 **AND/OR** TTP > 3 x GTP
- False Positive = F1 < 0.25 **AND** Precision < 1

OR TTP > 3 x GTP

The boundaries penalise systems that produce a Target track more than three times greater than the number of pixels of the ground truth bounding box.

Also, systems that produce a very small bounding box (e.g. less than 10% of the ground truth) will produce a False Negative, but, if the track is completely within the ground truth bounding box (i.e. Precision = 1), the system will not incur the additional penalty of a False Positive.

Figure 10

Figure 10 shows three separate examples of how the metric applies. In the first example, the tracker and ground truth are the same size, with a 25% overlap, therefore producing an F1 of 0.25 (True Positive).

Example 2 shows a track that is 10% of the ground truth. This will produce an F1 of 0.18. Therefore this is a False Negative. However, as the precision is still 1, the track will not produce a False Positive.

Example 3 shows a tracker bounding box with an area greater than three times the ground truth. In this instance, the track will produce a False Negative and a False Positive, despite the fact that the F1 score is 0.33.

Each True Positive, False Positive and False Negative for a frame will be counted and added to a final F1 metric.

$$FinalF1 = \frac{2(\mathrm{Re}call \times \mathrm{Pr}ecision)}{\mathrm{Re}call + \mathrm{Pr}ecision}$$

Where, in the final F1 score:

$$\mathrm{Re}call = \frac{TotalTP}{TotalTP + TotalFN}$$

$$\mathrm{Pr}ecision = \frac{TotalTP}{TotalTP + TotalFP}$$

This method will produce an overall metric similar to the i-LIDS Event Detection scenarios that can be used to determine the quality of an algorithm over the entire dataset.

Additionally, a Track Percentage score will be provided for additional information, as this provides a more intuitive way of interpreting the results. Percentage Track is calculated as shown:

$$\%Track = \mathrm{Re}call \times 100$$

Subject to the agreed terms and conditions governing the evaluation process, systems demonstrating a Final F1 performance measure in excess of set boundaries will be recommended for practical use. These systems will be listed in a catalogue of approved security equipment and used by purchasers in Government and other parts of the UK's critical national infrastructure. The F1 values which must be obtained in order to qualify for practical recommendation are not made public.

7 Appendix A: References

[1] http://viper-toolkit.sourceforge.net/

[2] http://www.smpte.org/home

[3] C.J. van Rijsbergen. Information Retrieval. Butterworths, London, 1979.

[4] K. Smith et al. “Evaluating Multi-Object Tracking”, CVPR 2005.

[5] F. Yin et al. “Performance Evaluation of Object Tracking Algorithms”, 10th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS2007), October 2007.

8 Appendix B: HOSDB Event Detection XML Schema

The XML database structure is summarised by the following XML schema. Optional elements are tagged as `minOccurs="0"`, and most elements and attributes are unordered indicated by the use of the `<xs:all></xs:all>` tags:

```
<?xml version="1.0" encoding="utf-8"?>
<xs:schema id="NewSchema" xmlns:xs="http://www.w3.org/2001/XMLSchema" />
  <xs:element name="IlidsLibraryIndex">
    <xs:complexType>
      <xs:element name="Library">
        <xs:complexType>
          <xs:sequence>
            <xs:element name="scenario" type="scenarioType"/>
            <xs:element name="dataset" type="datasetType"/>
            <xs:element name="libversion" type="xs:string"/>

            <xs:element name="clip">
              <xs:complexType>
                <xs:all>
                  <xs:element name="Filename" type="filenameType"/>
                  <xs:element name="StartOffset" type="timeType"/>
                  <xs:element name=”Stage" type="xs:integer"/>
                  <xs:element name="Duration" type="timeType"/>
                  <xs:element name="PeriodOfDay" minOccurs="0"
                          type="periodOfDayType"/>
                  <xs:group ref="weatherGroup"/>
                  <xs:element name="Distraction" type="xs:distractionType"
                              minOccurs="0" maxOccurs="unbounded"/>
                  <xs:element name="AlarmEvents" type="xs:integer"/>
                  <xs:element name="Alarms" type="xs:string">
                    <xs:complexType>
                      <xs:element name="Alarm" type="xs:string" minOccurs="0"
                              maxOccurs="unbounded">
                        <xs:group ref="alarmGroup"/>
                      </xs:element>
                    </xs:complexType>
                  </xs:element>
                 </xs:all>
              </xs:complexType>
            </xs:element>
          </xs:sequence>
        </xs:complexType>
      </xs:element>
    </xs:complexType>
  </xs:element>

  <xs:simpleType name="booleanType">
    <xs:restriction base="xs:string">
      <xs:pattern value="(Yes)|(No)"/>
    </xs:restriction>
  </xs:simpleType>

  <xs:simpleType name="timeType">
    <xs:restriction base="xs:string">
      <xs:pattern value="[0-9]{2}:[0-9]{2}:[0-9]{2}"/>
    </xs:restriction>
  </xs:simpleType>

  <xs:simpleType name="timeOfDayType">
    <xs:restriction base="xs:string">
      <xs:pattern value="(Dawn)|(Day)|(Dusk)|(Night)"/>
    </xs:restriction>
  </xs:simpleType>

  <xs:simpleType name="periodOfDayType">
    <xs:restriction base="xs:string">
```

```
    <xs:pattern value="(Low) | (Medium) | (High)"/>
  </xs:restriction>
</xs:simpleType>

<xs:simpleType name"scenarioType">
  <xs:restriction base="xs:string">
     <xs:pattern value="(Parked Vehicle) | (Abandoned Baggage) |
                        (Sterile Zone) | (Doorway Surveillance)"/>
  </xs:restriction>
</xs:simpleType>

<xs:simpleType name"distractionType">
  <xs:restriction base="xs:string">
     <xs:pattern value="(Bag) | (Bats) | (Bird) |
                        (Camera switch from colour to monochrome) |
                        (Camera switch from monochrome to colour) |
                        (Flickering light) | (Foxes) | (Insect on camera) |
                        (Insects) | (Rabbits) | (Shadow through fence) |
                        (Squirrel)|
                        (Cyclist) | (Moving vehicle) | (Parked vehicle)”/>

  </xs:restriction>
</xs:simpleType>

<xs:simpleType name="filenameType">
  <xs:restriction base="xs:string">
    <xs:pattern value="(PV|AB|DS|SZ)(TE|EV|TR)(A|N)[1-9][0-99]{2}[a-z].qtl"/>
  </xs:restriction>
</xs:simpleType>

<xs:simpleType name"datasetType">
  <xs:restriction base="xs:string">
     <xs:pattern value="(Evaluation) | (Training) | (Test) /">
  </xs:restriction>
</xs:simpleType>

<xs:simpleType name="cloudType">
  <xs:restriction base="xs:string">
    <xs:pattern value="(None)|(Some)|(Overcast)"/>
  </xs:restriction>
</xs:simpleType>

<xs:simpleType name="subjectDescType">
  <xs:restriction base="xs:string">
    <xs:pattern value="(One person) | (Two people)" |
                      (Ambulance) | (Car) | (Minibus) | (MPV) |
                      (Pedestrian) | (Truck) | (Van)/>
  </xs:restriction>
</xs:simpleType>

<xs:simpleType name="subjectNumType">
  <xs:restriction base="xs:integer">
    <xs:pattern value="[1-12]"/>
  </xs:restriction>
</xs:simpleType>

<xs:simpleType name="approachDescType">
  <xs:restriction base="xs:string">
    <xs:pattern value="(Body drag)|(Crawl)|(Creep walk)|(Crouch run)|
                    (Crouch walk)|(Log roll)|(Run)|(Walk)|(Walk with ladder)"/>
  </xs:restriction>
</xs:simpleType>

<xs:simpleType name="approachOrientType">
  <xs:restriction base="xs:string">
    <xs:pattern value="(Perpendicular) | (Diagonal) |
                      (Facing away from camera) | (Facing towards camera)"/>
  </xs:restriction>
</xs:simpleType>

<xs:simpleType name="objectZoneType">
  <xs:restriction base="xs:string">
    <xs:pattern value="(Near) | (Mid) | (Far)"/>
  </xs:restriction>
</xs:simpleType>

<xs:simpleType name="objectDescType">
```

```
    <xs:restriction base="xs:string">
      <xs:pattern value="(Bottle) | (Drinks can) | (Family suitcase) |
                         (Newspaper) | (Paper) | (Rucksack) | (Sports Bag)"/>
    </xs:restriction>
  </xs:simpleType>

  <xs:simpleType name="dressCodeType">
    <xs:restriction base="xs:string">
      <xs:pattern value="(Smart) | (Casual)"/>
    </xs:restriction>
  </xs:simpleType>

  <xs:group name="weatherGroup">
    <xs:all>
      <xs:element name="TimeOfDay" type="timeOfDayType"/>
      <xs:element name="Clouds" type="cloudType"/>
      <xs:element name="Rain" type="booleanType"/>
      <xs:element name="Snow" type="booleanType"/>
      <xs:element name="Fog" type="booleanType"/>
    </xs:all>
  </xs:group>

  <xs:group name="alarmGroup">
    <xs:all>
      <xs:element name="StartTime" type="timeType"/>
      <xs:element name="AlarmDescription" type="xs:string"/>
      <xs:element name="AlarmDuration" type="timeType"/>
      <xs:element name="Distance" minOccurs="0" type="xs:string"/>
      <xs:element name="SubjectDescription" minOccurs="0" type="subjectDescType"/>
      <xs:element name="NumberOfSubjects" minOccurs="0" type="subjectNumType"/>
      <xs:element name="SubjectApproachType" minOccurs="0" type="approachDescType"/>
      <xs:element name="SubjectOrientation" minOccurs="0" type="approachOrientType"/>
      <xs:element name="ObjectZone" minOccurs="0" type="objectZoneType"/>
      <xs:element name="ObjectDescription" minOccurs="0" type="objectDescType"/>
      <xs:element name="SuspectDressCode" minOccurs="0" type="dressCodeType"/>
    </xs:all>
  </xs:group>
</xs:schema>
```

9 Appendix C: Contact Information

The i-LIDS team are part of the Home Office Scientific Development Branch (HOSDB) and can be contacted by mail at:

i-LIDS team
Home Office Scientific Development Branch
Langhurst House
Langhurstwood Road
Horsham
West Sussex RH12 4WX

or, by voicemail on:

(+44) (0)1403 213823

or, by fax marked 'FAO i-LIDS team' on:

(+44) (0)1403 213827

or, email at:

i-LIDS@homeoffice.gsi.gov.uk

The i-LIDS website can be found at:

http://scienceandresearch.homeoffice.gov.uk/hosdb/

10 Appendix D: Event Detection System Evaluation Application Form

Please send to: i-LIDS team, HOSDB, Langhurst House, Langhurstwood Road, Horsham, West Sussex. RH12 4WX.

Alternatively, fax to 01403 213827, marked FAO: i-LIDS team.

Contact name:

Organisation:

Address:

Telephone:

Email:

System Name:

Version:

Date:

Test dataset performance in proposed scenario and role

Scenario	Recall	Precision	Role	Recall Bias	F_1
eg. Parked Vehicle	0.40	0.80	Operational Alert	0.55	0.59

11 Appendix E: Object Tracking System Evaluation Application Form

Please send to: i-LIDS team, HOSDB, Langhurst House, Langhurstwood Road, Horsham, West Sussex. RH12 4WX.

Alternatively, fax to 01403 213827, marked FAO: i-LIDS team.

Contact name:

Organisation:

Address:

Telephone:

Email:

System Name:

Version:

Date:

Test dataset performance in proposed scenario and role

Scenario	Recall	Precision	F_1
e.g. Multiple Camera Tracking	0.81	0.54	0.65

12 Appendix F: FAQs

Q1. How can we play the footage back via a composite output?

A1. We recommend converting the footage to the MPEG2 format and creating DVDs. This will allow the footage to be played back via any DVD player.

Q2. We have already signed the End User License Agreement, but wish to purchase further datasets. Do we need to sign a new End User License Agreement?

A2. No. The End User License Agreement is indiscriminate of scenario and dataset.

Q3. What is the difference between Training and Test datasets?

A3. The Training Datasets splits the footage into small sequences that Video Analytics systems should be trained with to optimise their performance. VA systems can then be tested with the Test Dataset. Performance results from the Test Dataset can then be submitted to HOSDB.

Q4. Do I need to buy both Training and Test datasets?

A4. No. VA developers are not required to obtain both datasets, Training and Test Datasets can be ordered individually. For a system to be considered for evaluation, testing on the relevant Test Dataset must have been conducted by the manufacturer.

Q5. What format are the image sequences in the datasets?

A5. The image sequences are captured in MJPEG format CIF-4 resolution of 576*704 (4:3 aspect ratio) with 25 interlaced frames per second and 8 bit colour quantisation. Individual frames are compressed to approximately 90% of their original size. As this is MJPEG, there is no interframe dependency.

Q6. How do we have our systems evaluated by HOSDB?

A6. Submit your F1 Result from the Test Dataset of the relevant scenario along with the Evaluation Application Form, which can be found in these appendicies.

Q7. What is Event Recording?

A7. VBDS act as a trigger for recording of suspicious events, where all the recordings obtained are to be analysed at a later time.

Q8. What is Operational Alert?

A8. VBDS provide real-time detection of suspicious events which must be dealt with by a human controller.

Q9. Can I get a sample of i-LIDS before I decide to buy it?

A9. Yes. Email i-LIDS enquiries with your name, address and company details and we will send you a demonstration DVD.

Home Office Scientific Development Branch
Langhurst House
Langhurstwood Road
Horsham
RH12 4WX
United Kingdom

Telephone: +44 (0)1403 213800
Fax: +44 (0)1403 213627
E-mail: hosdb@homeoffice.gsi.gov.uk
Website: http://science.homeoffice.gov.uk/hosdb/

ISBN 978-1-84726-897-6