PETS 2016: Dataset and Challenge

Luis Patino[*], Tom Cane[**], Alain Vallee[***] and James Ferryman[*]

[*]University of Reading, Computational Vision Group, Reading RG6 6AY, United Kingdom,
{j.l.patinovilchis, j.m.ferryman}@reading.ac.uk
[**]BMT Group Ltd., Teddington TW11 8LZ. United Kingdom, tcane@bmtmail.com
[***]SAGEM, 92659 Boulogne-Billancourt, France, alain.vallee@sagem.com

Abstract

This paper describes the datasets and computer vision challenges that form part of the PETS 2016 workshop. PETS 2016 addresses the application of on-board multi sensor surveillance for protection of mobile critical assets. The sensors (visible and thermal cameras) are mounted on the asset itself and surveillance is performed around the asset. Two datasets are provided: (1) a multi sensor dataset as used for the PETS2014 challenge which addresses protection of trucks (the ARENA Dataset); and (2) a new dataset - the IPATCH Dataset - addressing the application of multi sensor surveillance to protect a vessel at sea from piracy. The dataset specifically addresses several vision challenges set in the PETS 2016 workshop, and corresponding to different steps in a video understanding system: Low-Level Video Analysis (object detection and tracking), Mid-Level Video Analysis ('simple' event detection: the behaviour recognition of a single actor) and High-Level Video Analysis ('complex' event detection: the behaviour and interaction recognition of several actors).

1. Introduction

There is nowadays a significant amount of research achieved in the field of video surveillance. A large number of algorithms have been designed and tested for the tasks of object detection and tracking as well as for detection of events of interest, abnormalities or criminal behaviours. However, it is still difficult to compare or evaluate such algorithms because of the lack of standard metrics and benchmarks that indicate how detection, tracking and threat analysis perform against a common database. Furthermore, there is an imbalance on the progress and development of such algorithms depending on the application domain. This occurs mainly because of a lack of public data available for research and evaluation. The goal of the PETS workshop has been to foster the emergence of computer vision technologies for detection, tracking and surveillance by providing datasets and metrics that allow an accurate assessment and comparison of such methodologies. PETS 2016 is sponsored by the EU project IPATCH[1] (Intelligent Piracy Avoidance using Threat detection and Countermeasure Heuristics) and continues the series of highly successful PETS workshops held for over fifteen years (FG 00, CVPR '01, ECCV '02, ICVS '03, ICCV '03, ECCV'04, ..., AVSS'12, WVM'13, AVSS'14, AVSS'15).

PETS 2016 addresses the application of on-board multi sensor surveillance for protection of mobile critical assets. Such assets (including trucks, trains, and shipping vessels) could be considered as targets for criminals, activists or even terrorists. The sensors (visible and thermal cameras) are mounted on the asset itself and surveillance is performed around the asset. Two datasets are provided: (1) a multi sensor dataset as used for the PETS2014 challenge which addresses protection of trucks (the ARENA Dataset); and (2) a new dataset - the IPATCH Dataset - addressing the application of multi sensor surveillance to protect a vessel at sea from piracy. This new dataset is unique in the sense it comprises a suite of heterogeneous sensors (GPS, visual and thermal cameras) and will fill the current void of publicly available annotated datasets in the maritime domain. The ARENA dataset has been available for download since PETS2014. Hence, PETS2016 provides the opportunity for researchers and industry to submit methodological advances and results obtained using this data since the 2014 and 2015 workshops.

Ultimately, the visual challenge in PETS 2016 comes down to deciding if the detected activities in the video are non-dangerous abnormalities or if they constitute a real

[1] www.ipatchproject.eu

threat, in the land case to the motor vehicle, its driver or any asset contained inside; in the maritime case to the main vessel. For this, visual cues and the temporal history of the scenario must be analysed to differentiate each behaviour.

The remainder of the paper is organised as follows. Section 2 presents the PETS 2016 datasets in detail. The challenges addressed with this datasets are given in Section 3. For authors willing to investigate only abnormal event detection and behaviour understanding topics, tracking data has been made available. A brief explanation of such trajectories is given in Section 4. Some conclusions on the current datasets as well as some potential new challenges in forecoming PETS workshop are presented in Section 5.

2. Datasets

The datasets are multisensor sequences containing different activities around a mobile critical asset. The first dataset is the ARENA Dataset as used for the PETS2014 challenge which addresses protection of trucks. The second dataset is the IPATCH Dataset - addressing the application of multi sensor surveillance to protect a vessel at sea from piracy. Only a brief summary of the former dataset is given here as it is originally described in PETS 2014 [5]. The latter dataset is presented in detail in this paper.

2.1. ARENA Dataset

The ARENA dataset comprises of a series of multi-camera video recordings (22 acted scenarios) where the main subject is the detection and understanding of human behaviour around a parked vehicle, with the main focus on discriminating behaviour between normal, abnormal/rare behaviour and real threats. The main objective is to detect and understand the different behaviours from four visual (RGB) cameras mounted on the four corners of the vehicle itself.

2.1.1 Camera setup and characteristics

The recordings were carried out at the University of Reading, more precisely at the road intersection and car park in front of the School of Systems Engineering.

The on-board camera configuration during the recordings is shown in Figure 2. Four visual cameras are employed. Their characteristics are as follows: Model: BIP2-1300c-dn (http://www.baslerweb.com/products/Fixed-Box.html?model=178); Resolution: 1280 x 960 pixels; frame rate: 30 fps.

2.2. IPATCH Dataset

A new IPATCH dataset, collected in April 2015, contains a set of fourteen multi sensor recordings (visible, themal) collected off the coast of Brest, France, as a collaboration between the EU IPATCH project [2] and the AUTOPROTECTION project [1]. The IPATCH project addresses non-military protection measures for merchant shipping against piracy. The IPATCH system uses advanced sensors and data fusion to provide the Master of the ship with the information needed to decide how best to mitigate a threat. AUTOPROTECTION is French national project started in 2012. The project aim is to build a protection system against maritime piracy avoiding lethal weapons and thus reducing violence at sea by incorporating, among others: Automatic Threat Detection, Graduated Warning and Dissuasion Responses (for example, search lights and sound cannons); Anti-boarding Actions (for example, water cannons and smoke generators).

The recordings, which represent a series of realistic maritime piracy (abnormal / attack to vessel) scenarios, present different challenges covering object detection and tracking (fusion of data from sensors with different modalities and sensor handover (tracking objects passing from one field of view (FOV) to another with minimal overlapping FOV)), event detection and threat recognition. In addition to video data, GPS data is also made available. Altogether, detection, tracking and scene understanding challenges in the maritime domain can now be addressed on this dataset.

Figure 1. Partisan Vessel.

The main ship employed for these recordings is the VN Partisan; depicted in Figure 1. The VN Partisan is a multi-purpose offshore vessel frequently used for training by the French Navy and owned by SeaOwl [3]. The VN Partisan has the following characteristics:

- Ship Type: Tug
- Built: 1978
- GT (Gross Tonnage): 2342 t
- DWT (Deadweight): 2250 t
- LOA (Length Overall): 79 m
- Beam: 15 m
- Draft (avg): 5.7 m

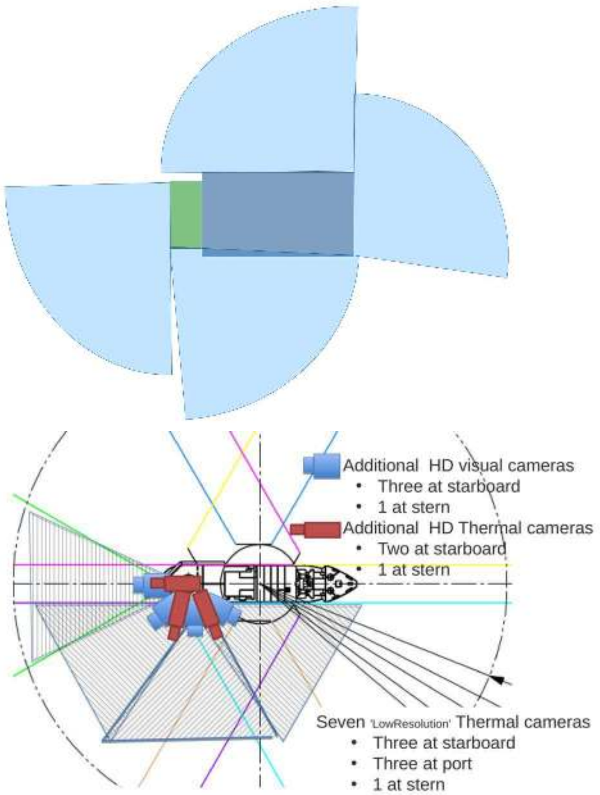

Figure 2. PETS2016 Datasets, sensors and coverage. Top row: Sensor locations and coverage in the ARENA Dataset ; Bottom row: Sensor locations and coverage in the IPATCH Dataset.

2.2.1 Camera setup and characteristics

The Partisan vessel is already equipped with 7 fixed 'low resolution' IR video cameras all around the ship. For the close range detection of piracy threats, supplementary visual and thermal cameras were added to the vessel as shown in Figure 2 and Figure 3. The images corresponding to these supplementary cameras are those released in the IPATCH dataset.

Visual sensors. Four AXIS P1427-E Network cameras were added to the ship; three of them at starboard and one at stern.

Thermal sensors. Three thermal cameras were added to the ship; two of them at starboard (FLIR SC655 Network Cameras). One thermal camera was placed at stern (FLIR A65 Network Camera).

The visual and thermal camera technical characteristics are given in Table 1.

2.3. Recorded Scenarios

In both, the ARENA and IPATCH datasets, scenarios include three different types of activity:

Figure 3. IPATCH cameras were positioned on a tower located at the stern of the Partisan Vessel.

- 'Normal activity': Made up of behaviours that are frequently observed within the context of the given dataset.
- 'Abnormal activity': Abnormal behaviour that however cannot be considered as a real threat as the sensitive mobile asset or its crew is not attacked.
- 'Criminal activity': The security/safety of the sensitive mobile asset or its crew has been breached. In the land-case dataset this corresponds to people succeeding to access the truck and steal an object from it. Other scenarios include an attack to to the driver (physical agression). In the maritime dataset, this corresponds to an attack to the vessel.

While in the land-case dataset the truck remains parked at the same site, two different recordings are made in the maritime dataset that assume the ship is at open sea, or at anchorage. The two type of scenarios, reagarding the vessel kinematics, are as follows:
Navigating. The ship travels steadily at normal speed. Piracy incidents occur while the ship is moving on the open sea.
At anchorage. Although the activity could be near to a port, it is assumed that anchorage is happening at open sea itself, or at an isolated dock. First, this prevents potential privacy issues during data acquisition, secondly this bounds scenario complexity by limiting the number of objects and actors to be detected, tracked and whose activity is to be analysed.
At port. The ship is anchored at port or travelling at low speed inside the port.

The actual location for the recordings is the Brest harbour, the Crozon area and neighbouring maritime seas as shown in Figure 4. Scenarios at anchorage and at open sea were recorded on the 21st April 2015. Recording of the scenarios at anchorage occurred inside the Crozon area. Recording of scenarios at open sea while the vessel is navigating occurred in the transit between the Brest harbour and

	AXIS P1427 E	FLIR SC655 Network Camera	FLIR A65 Network Camera
Resolution	5 megapixel	640 x 480 pixels	640 x 512 pixels
Day & Night functionality	yes	yes	yes
Object temperature		-20°C to +150°C	-25°C to +135°C
Environment temperature	wide range and weatherproof	15°C to +50°C	15°C to +50°C
Frame rate	30 fps	25 fps	30 fps
autofocus	PTZ with autofocus	motor focus	fixed focus
FoV	35 ° to 109 °	25°(H) x 19° (V)	45°(H) x 37° (V) with 13 mm lens

Table 1. Technical characteristics of additional high resolution visual and thermal cameras placed on the Partisan vessel.

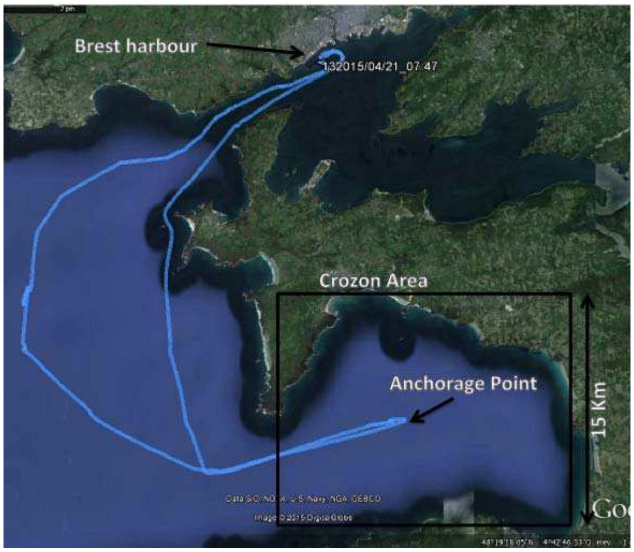

Figure 4. Ship journey for the scenario recordings at open sea and at anchorage. Recordings at port were held in Brest harbour.

the anchorage point. Scenarios at port were recorded inside the Brest harbour on 22nd April 2015.

Because of the specifity of piracy attacks at sea, some further hypotheses were made while preparing the scenarios.

1. The VN Partisan has a maximum speed of 12 knots.
2. Ships in general have a normal speed of about 20 knots, as assessed from an end-user requirements analysis [4].
3. Skiffs may accelerate up to 40 knots [4].
4. Scenario activity is not expected to spread out more than 2 km beyond the ship.
5. The number of boats 'acting' in the scenario, excluding the main vessel, is not more than three and correspond either to fishing (maximum of one) or pirate boats (maximum of two); see Section 2.3.1 for further details.

2.3.1 IPATCH Supplementary Target boats

As targets for the scenarios, two RHIBs and two Fishing boats were used (Target boats are depicted in Figure 5):

Bourre-pif (called BP in recordings) is a 7.4 metres long RHIB (Zodiac SRR750) used by SeaOwl for technical and practical needs. Maximum speed is around 25 knots. It is equipped with 2 motors, radar antenna, GPS and UHF radio transceiver.

Black Bull (called BB in recordings) is a 7.3 metres long RHIB (Hydrosport 737 GTR) that was hired for the data collection trials needs. Maximum speed is around 50 knots.

Fishing boats (called FB in recordings). For the first part of the data collection (morning of 1st day Recordings), a first fishing boat was used. For the second part of the data collection (afternoon), a second fishing boat was used. In both cases the maximum speed is around 8 knots.

2.4. Contained Behaviours

Each scenario contains acted behaviours, which may be taken as indications or pieces of information allowing to infer when the scenario activity corresponds to innocuous abnormal activity ('something is wrong') or if the activity can be considered as 'potential criminal behaviour' or 'criminal behaviour'. Acted behaviours are divided as 'normal', 'abnormal' or 'threats'. They are defined as follows:

Normal behaviours. This corresponds, in the land-case, to people simply walking along the pathways around the parked truck (see Figure 6). In the maritime case this corresponds to boats fishing in the area, or passing by the main vessel at normal speed (see Section 2.3).

Abnormal behaviours. There are a large number of abnormalities recorded in the dataset. However, only some of them are annotated. Behaviours that can be clear indications of abnormalities are:

- Falling (land-case only): Person losing balance and falling to ground. Can be caused by themselves or by a third person.

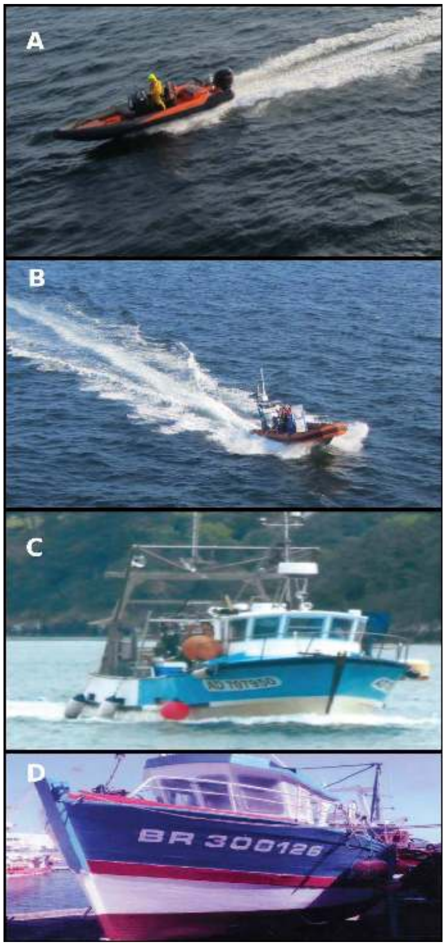

Figure 5. Supplementary boats employed in IPATCH recordings. A) skiff boat 'Black Bull' (called BB in recordings) B) skiff boat 'Bourre-pif' (called BP in recordings). C) fishing boat employed for the first part of the data collection. D) fishing boat employed for the second part.

- Person or skiff speeding up: Sudden acceleration of the mobile object.
- Person or skiff (boat) loitering: Detected object stands/moves slowly in the same area.
- Person or skiff group formation: A mobile comes close to another and holds an interaction.
- Person or skiff group separation: A mobile departs from a group.
- Person or skiff moving around the mobile asset: Mobile object moving and covering at least two sides of the vehicle or vessel.
- Person or skiff suddenly changing direction: Mobile object has sudden change of trajectory.

Figure 7. Examples of diverse froms of attack to the driver in ARENA dataset. A) Driver hit by a group of three people B) Driver agressed by another person.

Threats. Behaviours that are clearly criminal behaviour:

- Attack to driver (land-case only): Physical and intentional aggression to driver where they are hit or menaced with an arm, and possibly brought to the ground.
- Stealing from vehicle (land-case only): Someone penetrates the vehicle completely or partially and departs with an object removed from the vehicle.
- Attack to vessel (maritime-case only): Abnormal approach to the vessel of a boat (possibly from a sudden change of direction and/or the boat speeding up) that concludes with the skiff staying at the vessel starboard.

In both, ARENA and IPATCH datasets, attacks to the truck driver and to the vessel occur in varied forms. In ARENA, an attack can come from an organised group of people or from a single individual. They can also be direct attacks or in 'two-stages' after the driver is trapped by the attackers desguising their intentions, for instance, by first asking some directions then attacking him. Figure 7 shows two different attacks to the truck driver.

In IPATCH dataset the attacks are varied as well and these correspond to general facts asserted by end-users interviewed on piracy instances. The attacks to the vessel are based on five types of scenario. Their description is as follows:

Scenario 1 (Sc1) Ship pursuit: The pirate boat appearing as fishermen by behaving as if anchored or remaining in a particular area (loitering) in the vicinity of other fishing vessels. The main ship passes by following its normal journey. The pirate boat suddenly picks up speed and approaches the target astern.

Scenario 2 (Sc2): Attack from 'Tuna fishers': The main ship travels following its normal journey. One pirate is disguised as a fisherman speeding a fishing boat to bring tuna fish to the surface. A second pirate is disguised as fisherman by following the wake of the first ocean-going fishing boat in pursuit of the fish (tuna) brought to the surface. Suddenly both pirate boats take on blocking the front and rear of the attacked main ship.

Figure 6. A group of people detected and tracked walking by the Truck.

Scenario 3 (Sc3): Attack with ship at anchorage; slow approach from skiffs: Pirates disguise themselves as fishermen by behaving as one or more (near-) stationary fishing boats. The pirates slowly approach the anchored target ship. The pirates speed up at the final stage. Before boarding, the pirates may circle the vessel to look for activity and assess value of vessel.
Scenario 4 (Sc4): Attack during ship meeting: One vessel is at anchorage. A second small vessel approaches and stops, i.e. in order to bring a supplementary person on board to the main vessel. Suddenly two stationary skiffs, simulating being fishermen, speed up to attack the main cargo.
Scenario 5 (Sc5): Attack during queuing-up: One vessel arrives to the port for anchorage (arrives with low speed then almost not moving). A second vessel arrives and stops; moving slowly (queuing up). Suddenly two stationary skiffs, simulating being fishermen, speed up to attack the main cargo ship.

Figure 8 depicts graphically some of these scenarios. Figure 2.4 shows recorded skiffs approaching at fast speed the main vessel with IPATCH cameras. Variants of these scenarios were recorded as 'innocuous' abnormalities where the suspicious boat never perpetrates an attack.

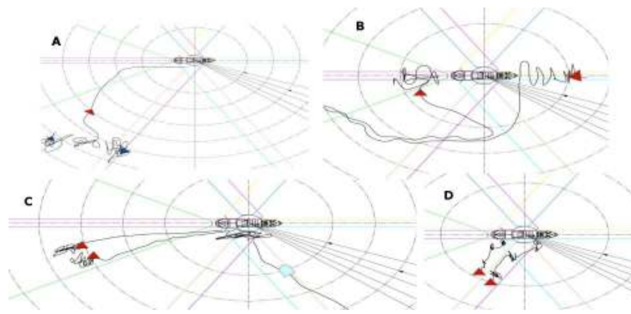

Figure 8. Examples of diverse froms of attack in IPATCH dataset. A) Ship pursuit B) Attack from 'Tuna fishers' C) Attack during ship meeting D) Attack with ship at anchorage; slow approach from skiffs. See Section 2.4 for detailed description.

3. Challenges

3.1. Low-Level Video Analysis

The task is to detect and track objects in all frames from video sequences and report detected/tracked object bounding boxes for each object at each frame. The datasets made available present low-level challenges on object detection and tracking, such as :
Pass: When mobiles are passing by the Truck or Vessel, tracking algorithms must perform basic tracking of objects of different size at different distances.
Approach: When mobiles are approaching the Truck or Vessel, tracking algorithms must perform basic tracking of changing-distance objects.
Leave: When mobiles are travelling away from the Truck or Vessel, tracking algorithms must perform basic tracking of changing-distance objects.
Changing appearance: When mobiles are turning to respect to the Truck or Vessel, their appearance may change. Tracking algorithms must perform on objects that change appearance (due to viewing angle).
Occlude: When there are several mobiles around the Truck or Vessel, occlusions may appear. Tracking algorithms must perform on multiple objects while maintaining identity.
Merge and split: When mobiles meet, and/or travel together and eventually split, tracking algorithms must perform on these multiple objects while maintaining identity.

Sequences that can be processed in this category are stated in Table 2.

3.2. Mid-Level Video Analysis

The task is to detect any of the events in this category and report the frame of the individual at the start of the event and the frame of the individual at the end of the event.

Figure 9. Example of synchronised images from IPATCH Dataset. Top row: Thee visible cameras; Bottom row, left: visible camera. Bottom row, right: thermal camera.

- Sequences that can be processed, for each 'atomic' event, are given in Table 3.
- The learnt/modelled events should comply with the definition of the behaviour given in Section 2.4.

3.3. High-Level Video Analysis

The task is to detect any of the events in this category and report the frame of the individual or asset under attack or individual stealing from vehicle, at the start of the event, and the frame of the individual or asset under attack or individual stealing from vehicle, at the end of the event.

- Sequences that can be processed in this category are those belonging to Table 4.
- The learnt/modelled events should comply with the definition of the behaviour given in Section 2.4.

4. Tracking made available

For authors addressing only the Mid-level or High-level video analysis challenge, tracking data for the corresponding sequences is provided, so that it is not compulsory to work on the Low-level challenges to address those in Mid-level or High-level categories. In the land-case scenarios, the same real tracking results as given in PETS 2014 [5] are distributed. The employed tracker to generate these results has been described before in detail [6]. For the maritime-case all boats were equipped with GPS. Recorded tracks are available to authors. Each boat is identified as mentioned in Section 2.3.1.

Low-Level Video Analysis	
ARENA Dataset	**IPATCH Dataset**
Person detection and person tracking	Boat detection and boat tracking
• 01_02_ENV_RGB_3	• Sc2a_Tk1_TST_Th_2
• 01_02_TRK_RGB_1	• Sc2a_Tk1_UoR_Th_1
• 01_02_TRK_RGB_2	• Sc2a_Tk1_UoR_RGB_11
• 11_03_ENV_RGB_3	• Sc2a_Tk1_UoR_RGB_12
• 11_03_TRK_RGB_1	• Sc3_Tk2_TST_Th_1
• 15_06_ENV_RGB_3	• Sc3_Tk2_TST_Th_2
• 15_06_TRK_RGB_2	• Sc3_Tk2_UoR_RGB_14

Table 2. Selected sequences on Detection and Tracking.

5. Conclusions

We have presented in this paper the dataset and vision challenges that form part of the PETS 2016 workshop. The recorded videos contained in the dataset, address the application of on-board multi sensor surveillance for protection of mobile critical assets. Two datasets are provided: (1) a multi sensor dataset as used for the PETS2014 chal-

Mid-Level Video Analysis ('simple' abnormal event detection)	
ARENA Dataset	**IPATCH Dataset**
Person falling or pushed to ground • 11_04 • 11_03 • 08_02	
Person speeding up (starting to run) • 11_03 • 14_01 • 08_03	Boat speeding up • Sc3_Tk1 • Sc3_Tk3
Person loitering • 03_06 • 14_05	Boat loitering • Sc3_Tk1 • Sc3_Tk3
Group formation/separation • 08_02 • 11_04 • 11_05 • 03_05	Group formation/separation • Sc4_Tk3 • Sc3b_Tk1 • Sc4_Tk2
Person moving around vehicle • 06_01 • 10_03	Boat moving around vessel • Sc1_Tk2 • Sc1_Tk3
Person suddenly changing direction • 08_02 • 08_03 • 03_05	Skiff suddenly changing direction • Sc2a_Tk1 • Sc2_Tk2 • Sc3a_Tk2 • Sc2b_Tk3

Table 3. Abnormal events and selected sequences.

High-Level Video Analysis ('complex' Threat event detection)	
ARENA Dataset	**IPATCH Dataset**
Attack to person Real threat • 22_01 • 15_06 'Innocent' abnormality • 08_02 • 11_04	Attack to vessel Real threat • Sc1_Tk1 • Sc2_Tk2 'Innocent' abnormality • Sc2a_Tk1 • Sc2b_Tk3
Stealing from vehicle Real threat • 14_01 • 14_07 'Innocent' abnormality • 06_01 • 03_06	

Table 4. Threat events and selected sequences.

lenge which addresses protection of trucks (the ARENA Dataset); and (2) a new dataset - the IPATCH Dataset - addressing the application of multi sensor surveillance to protect a vessel at sea from piracy. In both cases, recorded scenarios include three different types of activity: 'normal activity', 'abnormal activity' and 'criminal activity'. The workshop vision challenge comes down to deciding if the detected activities in the video are part of innocent abnormalities or if they constitute a real threat. For this, behavioural cues and the temporal history of the scenario must be analysed. PETS2016 has enlarged the set of addressed behaviours in comparison to last two previous editions. Behaviours of interest included in the dataset are: Fall on floor (land-case only), Loitering; Moving around the sensitive asset; Group formation; Group separation; Speeding up; Suddenly changing direction; Fight and stealing (land-case only); Attack to vessel (maritime-case only). The dataset gives the opportunity to evaluate different steps in a video understanding system: Low-Level Video Analysis (object detection and tracking), Mid-Level Video Analysis ('simple' event detection: the behaviour recognition of a single actor) and High-Level Video Analysis ('complex' event detection: the behaviour and interaction recognition of several actors). In particular, the new dataset (IPATCH dataset) provides a maritime benchmark dataset, which is filling the current void for such datasets in the maritime domain. The datasets are publically available for the purposes of the PETS workshops and academic and industrial research (see download instructions at www.pets2016.net). Where the data is disseminated (e.g. publications, presentations) this paper should be acknowledged.

Acknowledgement

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 607567.

References

[1] AUTOPROTECTION http://en.sofresud.com/Maritime-Surveillance/AUTOPROTECTION/. 2

[2] The IPATCH project http://www.ipatchproject.eu/. 2

[3] The SeaOwl group www.seaowlgroup.com. 2

[4] IPATCH project deliverable D2.1: User Requirements Analysis, 2015. 4

[5] PETS 2014 the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, www.pets2014.net. 2, 7

[6] L. Patino and J. Ferryman. Pets 2014: dataset and challenge. In *Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference on*, pages 355–360. IEEE, 2014. 7