

PETS 2017: Dataset and Challenge

Luis Patino*, Tahir Nawaz*, Tom Cane** and James Ferryman*

*University of Reading, Computational Vision Group, Reading RG6 6AY, United Kingdom,

{j.l.patinovilchis, t.h.nawaz, j.m.ferryman}@reading.ac.uk

**BMT Group Ltd., Teddington TW11 8LZ. United Kingdom, tcane@bmtmail.com

Abstract

This paper indicates the dataset and challenges evaluated under PETS2017. In this edition PETS continues the evaluation theme of on-board surveillance systems for protection of mobile critical assets as set in PETS 2016. The datasets include (1) the ARENA Dataset; an RGB camera dataset, as used for PETS2014 to PETS 2016, which addresses protection of trucks; and (2) the IPATCH Dataset; a multi sensor dataset, as used in PETS2016, addressing the application of multi sensor surveillance to protect a vessel at sea from piracy. The datasets allow for performance evaluation of tracking in low-density scenarios and detection of various surveillance events ranging from innocuous abnormalities to dangerous and criminal situations. Training data for tracking algorithms is released with the dataset; tracking data is also available for authors addressing only surveillance event detection challenges but not working on tracking.

1. Introduction

The 2017 International Workshop on Performance Evaluation of Tracking and Surveillance (PETS 2017) continues the series of highly successful PETS workshops held for almost twenty years (FG 00, CVPR '01, ECCV '02, ICVS '03, ICCV '03, ECCV'04, ..., WVM'13, AVSS'14, AVSS'15, CVPR'16). The goal of the PETS workshop has been to foster the emergence of methodologies for performance evaluation of tracking and surveillance. This includes development of performance evaluation metrics as well as a quantitative evaluation of tracking and surveillance results based on a common dataset.

This year, the BMTT Tracking Challenge (Benchmarking of Multi-Target Tracking) and PETS (Performance Evaluation of Tracking and Surveillance) have joined to organise the first BMTT-PETS workshop of tracking and

surveillance, in conjunction with CVPR 2017.

BMTT 2017 is focusing on object detection and its interaction with tracking in crowded scenarios. PETS 2017 continues the evaluation theme of on-board surveillance systems for protection of mobile critical assets as set in PETS 2016. 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 in PETS 2017: (1) a multi sensor dataset, as used for PETS2014 to PETS 2016, which addresses protection of trucks (the ARENA Dataset); and (2) a maritime dataset - the IPATCH Dataset - addressing the application of multi sensor surveillance to protect a vessel at sea from piracy. This dataset is unique in the sense it comprises a suite of heterogeneous sensors (GPS, visual and thermal cameras) and fills a previous void of publicly available annotated datasets on the maritime domain. This paper focuses on PETS datasets and its evaluation metrics. PETS2017 is releasing new ground truth for learning purposes and new more adapted tracking metrics.

2. Datasets

Two datasets are employed, which are multisensor sequences containing different activities around a mobile critical asset. The first dataset is the ARENA Dataset as used for the PETS2014 [2] challenge which addresses protection of trucks. The second dataset is the IPATCH Dataset as used for PETS2016 [3] - addressing the application of multi sensor surveillance to protect a vessel at sea from piracy. In this edition new ground truth is being released for algorithm learning purposes.

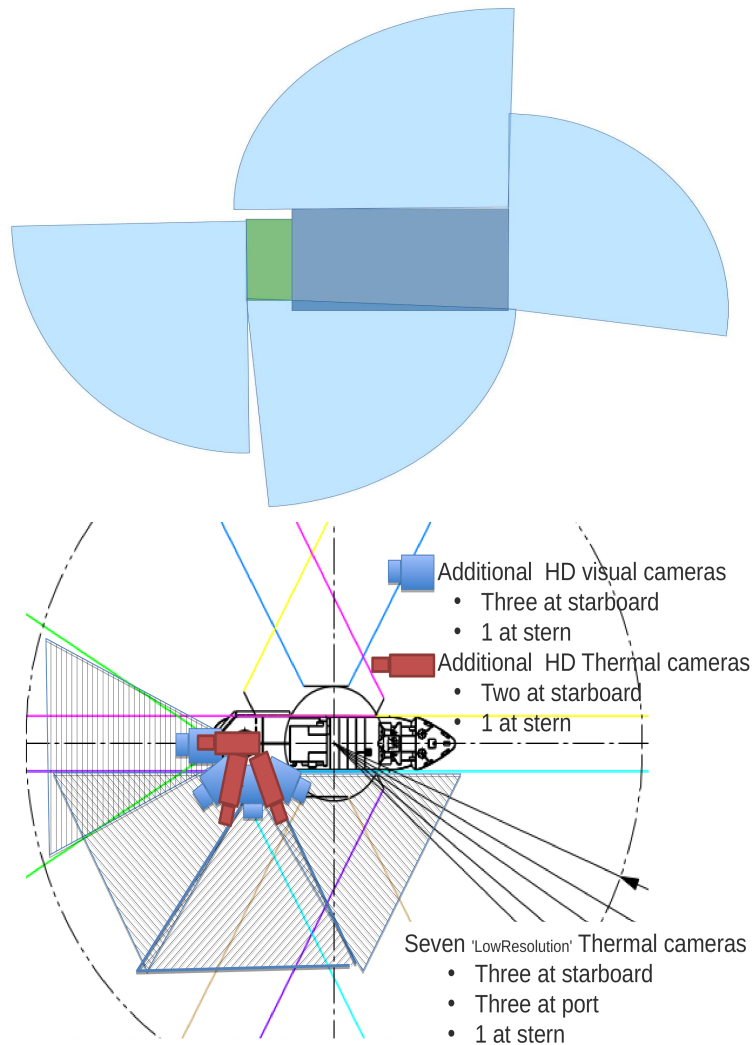


Figure 1. PETS2016-2017 Datasets, sensors and coverage. Top row: Sensor locations and coverage in the ARENA Dataset ; Bottom row: Sensor locations and coverage in the IPATCH Dataset.

2.1. 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 (see Figure 1). 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. The dataset has been available for download since PETS2014. Hence, PETS2017 provides

the opportunity for researchers and industry to submit methodological advances and results obtained using this data since the 2014 to 2016 workshops. More details on the dataset are available at: www.pets2014.net and [14, 10].

This year we have released a training set for tracking algorithms composed of three sequences with available ground truth. The data should be useful particularly for authors working on tracking methods with supervised learning stages such as deep learning.

2.2. IPATCH Dataset

The IPATCH dataset was first collected in April 2015. New recordings were carried out in September 2016. In this dataset real piracy inspired scenarios are recorded. The ship

Sequences for Detection and Tracking Analysis	
ARENA Dataset	IPATCH Dataset
person detection and tracking	boat detection and boat tracking
Sequences with GT for training	Sequences with GT for training
<ul style="list-style-type: none"> • 01_02_ENV_RGB_3 • 01_02_TRK_RGB_1 • 01_02_TRK_RGB_2 	<ul style="list-style-type: none"> • Sc2a_Tk1_UoR_Th_1 • Sc2a_Tk1_UoR_RGB_12 • Sc3_Tk2_TST_Th_1
Sequences for testing	Sequences for testing
<ul style="list-style-type: none"> • 11_03_ENV_RGB_3 • 11_03_TRK_RGB_1 • 15_06_ENV_RGB_3 • 15_06_TRK_RGB_2 	<ul style="list-style-type: none"> • Sc2a_Tk1_TST_Th_2 • Sc2a_Tk1_UoR_RGB_11 • Sc3_Tk2_TST_Th_2 • Sc3_Tk2_UoR_RGB_14

Table 1. Selected sequences on Detection and Tracking.

under attack is fitted with different types of sensors including visual and thermal cameras. The dataset contains a set of fourteen multi sensor recordings (see Figure 1) collected off the coast of Brest, France, in collaboration with AUTO-PROTECTION French national project [1]. The recordings present different challenges covering object detection and tracking (fusion of data from sensors with different modalities and sensor handover (tracking objects passing from one FOV to another with minimal overlapping FOV)), event detection and threat recognition. The video sequences are recorded from four high-resolution visual cameras and three high-resolution thermal cameras placed at starboard and at stern of the ship. Visual cameras are 5 megapixel PTZ with autofocus cameras and FoV 35° to 109°. Two of the Thermal cameras are 640 x 480 pixels in resolution, with motor focus and FoV 25°(H) x 19°(V). The remaining thermal camera is 640 x 512 pixels in resolution, with fixed focus and 45°(H) x 37°(V) with 13 mm lens.

In addition to video data, GPS groundtruth data is also made available. Altogether, detection, tracking and scene understanding challenges in the maritime domain can now be addressed on this dataset. More details on the dataset are available at: www.pets2016.net and [13].

2.3. Recorded Scenarios

Both datasets, ARENA and IPATCH were introduced in previous PETS editions. We refer the reader to [PETS2015 PETS2016] for a detailed description on the recorded scenarios in these datasets. Generally speaking, the datasets

correspond to acted scenarios containing diverse abnormalities although not all of them represent a ‘dangerous situation’. The recorded scenarios are divided as follows:

- ‘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 the driver (physical aggression). In the maritime dataset, this corresponds to an attack to the vessel.

In the ARENA dataset the valuable asset is a truck that remains parked at the same site. In the IPATCH dataset the valuable asset is a ship that may be navigating or at anchorage.

3. Challenges

We aim to give opportunity to all users of previous editions to submit new or updated results. Tracking and surveillance challenges remain as set in PETS2016. We recall next sequences that can be processed for performance evaluation of detection/tracking and surveillance.

3.1. Detection and Tracking

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. Note that in this PETS edition we have released a training dataset to facilitate parameter learning on supervised learning tracking algorithms.

Sequences that can be processed in this category are stated in Table 1.

3.2. Surveillance

The task is to detect any of the events in this category and report the frame at the start of the event and the frame at the end of the event. There are two kind of events to detect:

- atomic events: corresponds to short temporal span events and can be one of the following list:
 - Falling (land-case only): Person losing balance and falling to ground. Can be caused by themselves or by a third person.
 - 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.
- complex events: can be composed of several atomic events and represent activities that may have a higher degree of difficulty for its representation and detection. Complex events in the dataset can be:
 - 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.

Sequences that can be processed, for each ‘atomic’ event, are given in Table 2. Sequences that can be processed in this category are those belonging to Table 3.

4. Tracking made available

As in previous editions, PETS offers tracking data for those authors working on event detection but not necessarily on tracking. In the land-case scenarios, the same real tracking results as given in PETS 2014 [2] are distributed. The employed tracker to generate these results has been described before in detail [14]. For the maritime-case all boats were equipped with GPS. Recorded tracks are available to authors.

5. Performance metrics

5.1. Detection and tracking

Performance evaluation of tracking algorithms is a non-trivial task. In fact, for a thorough tracking assessment, several aspects needs to be assessed for which numerous metrics have been proposed over the years [6, 16, 5, 15, 12]. The choice of appropriate metric(s) is indeed quite challenging and could depend mainly on the application under consideration. Here, for the PETS tracking challenge, we choose the metrics adopted by a recent evaluation campaign, the Multiple Object Tracking (MOT) challenge [9]. MOT challenge uses the CLEAR metrics [5] as well as a set of track quality measures [16]. Overall, the metrics include Multiple Object Tracking Accuracy (MOTA), Multiple Object Tracking Precision (MOTP), False Alarms per Frame (FAF), ratio of the Mostly Tracked trajectories (MT), ratio of the Mostly Lost trajectories (ML), number of False Positives (FP), number of False Negatives (FN), number of Identity Switches (ID Sw) and number of track Fragmentations (Frag). For more details about these metrics, we refer the reader to [9].

The above set of metrics primarily focus on evaluating the end performance that is important particularly for ranking trackers. To obtain a deeper insight as why a certain end performance is achieved, it would also be desirable to analyze the factors (false positives, false negatives, ID changes) that contribute to the attainment of a certain end performance [11]. Therefore, to complement the evaluation using above metrics and further aid the performance analysis, we also adopt a recently proposed method [11] that enables revealing a dissected picture of the performance of trackers based on the analysis of the computed probability density functions (PDFs) of different *fault types* (i.e. false positives, false negatives, ID changes) in a sequence, as well as the associated performance scores. PDFs for false positives, false negatives and ID changes in a sequence are computed as the corresponding normalized histograms, and are denoted as $Pr[0 \leq FP_k \leq N_{fp}]$, $Pr[0 \leq FN_k \leq N_{fn}]$

Surveillance Challenge (Sequences for ‘atomic’ abnormal event detection)	
ARENA Dataset	IPATCH Dataset
Person falling or pushed to ground <ul style="list-style-type: none"> • 11_04 • 11_03 • 08_02 	
Person speeding up (starting to run) <ul style="list-style-type: none"> • 11_03 • 14_01 • 08_03 	Boat speeding up <ul style="list-style-type: none"> • Sc3_Tk1 • Sc3_Tk3
Person loitering <ul style="list-style-type: none"> • 03_06 • 14_05 	Boat loitering <ul style="list-style-type: none"> • Sc3_Tk1 • Sc3_Tk3
Group formation/separation <ul style="list-style-type: none"> • 08_02 • 11_04 • 11_05 • 03_05 	Group formation/separation <ul style="list-style-type: none"> • Sc4_Tk3 • Sc3b_Tk1 • Sc4_Tk2
Person moving around vehicle <ul style="list-style-type: none"> • 06_01 • 10_03 	Boat moving around vessel <ul style="list-style-type: none"> • Sc1_Tk2 • Sc1_Tk3
Person suddenly changing direction <ul style="list-style-type: none"> • 08_02 • 08_03 • 03_05 	Skiff suddenly changing direction <ul style="list-style-type: none"> • Sc2a_Tk1 • Sc2_Tk2 • Sc3a_Tk2 • Sc2b_Tk3

Table 2. Abnormal events and selected sequences.

Surveillance Challenge (Sequences for ‘complex’ event detection)	
ARENA Dataset	IPATCH Dataset
Attack to person Real threat <ul style="list-style-type: none"> • 22_01 • 15_06 ‘Innocent’ abnormality <ul style="list-style-type: none"> • 08_02 • 11_04 	Attack to vessel Real threat <ul style="list-style-type: none"> • Sc1_Tk1 • Sc2_Tk2 ‘Innocent’ abnormality <ul style="list-style-type: none"> • Sc2a_Tk1 • Sc2b_Tk3
Stealing from vehicle Real threat <ul style="list-style-type: none"> • 14_01 • 14_07 ‘Innocent’ abnormality <ul style="list-style-type: none"> • 06_01 • 03_06 	

Table 3. Threat events and selected sequences.

and $Pr[0 \leq IDC_k \leq N_{idc}]$, respectively. FP_k , FN_k and IDC_k are the number of false positives, false negatives and ID changes, respectively, at frame k of a sequence that has a total of K frames, and $N_{fp} = \max\{FP_k\}_{k=1}^K$, $N_{fn} = \max\{FN_k\}_{k=1}^K$, and $N_{idc} = \max\{IDC_k\}_{k=1}^K$.

Fig. 2 shows example PDFs of false positives, false negatives and ID changes for a pair of trackers on PETS 2009 S2.L1 view1 dataset [8]. Note: $Pr[FP_k = 0]$ is read as a probability in terms of the percentage of frames in which the tracker produces zero false positive; similarly,

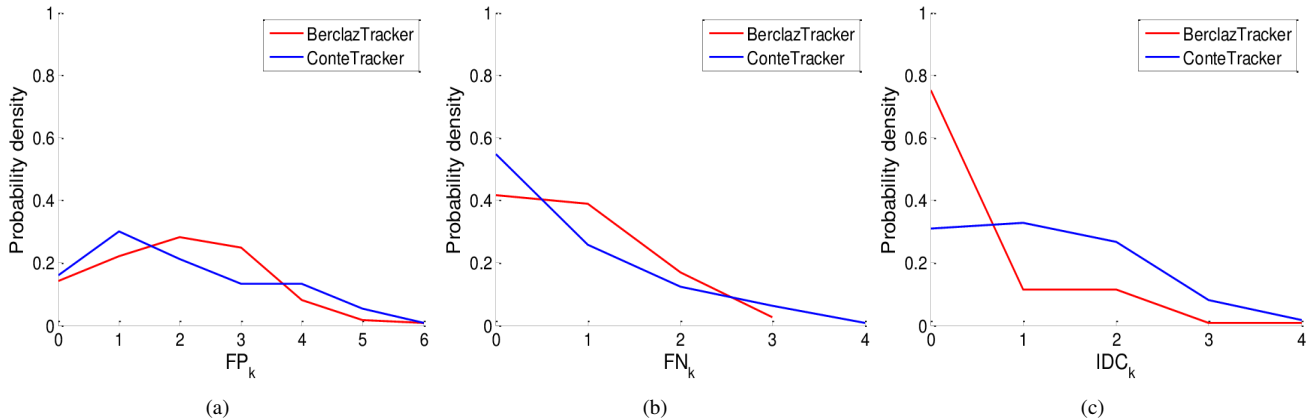


Figure 2. Probability density functions (PDFs) for a tracker by Berclaz *et al.* [4] and a tracker by Conte *et al.* [7] on PETS 2009 S2.L1 view1 [8] corresponding to false positives (first column), false negatives (second column), and ID changes (third column) in the sequence.

$Pr[FN_k = 2]$ refers to a probability in terms of the percentage of frames in which the tracker produces two false negatives; likewise, $Pr[IDC_k > 2]$ means a probability in terms of the percentage of frames in which the tracker produces more than two ID changes. Indeed, the analysis of a PDF could offer a more detailed picture of a tracker’s performance by revealing two aspects: tracker’s *robustness* and *per frame concentration* corresponding to each fault type, which are quantitatively accounted for by the following two performance scores. The first score tells the ability of a tracker to track without producing a fault across a sequence, and is called *robustness to a fault type (R)*:

$$R_{fp} = 1 - \frac{K_{fp}}{K}; R_{fn} = 1 - \frac{K_{fn}}{K}; R_{idc} = 1 - \frac{K_{idc}}{K}; \quad (1)$$

such that K_{fp} is the number of frames containing false positive(s), K_{fn} is the number of frames containing false negative(s), and K_{idc} is the number of frames containing ID change(s). $R_{fp} \in [0, 1]$, $R_{fn} \in [0, 1]$, $R_{idc} \in [0, 1]$: the higher the value ($R_{fp} / R_{fn} / R_{idc}$), the better the ability. The second score tells the tendency of a tracker to produce a fault type per frame, and is called *per frame concentration of a fault type (PFC)*:

$$PFC_{fp} = \frac{1}{K} \sum_{k=1}^K FP_k; PFC_{fn} = \frac{1}{K} \sum_{k=1}^K FN_k; \quad (2)$$

$$PFC_{idc} = \frac{1}{K} \sum_{k=1}^K IDC_k.$$

For more details, please refer to [11].

5.2. Surveillance

We are interested in behaviours that correspond to movements and displacements that may be indicative of abnormalities, dangerous or criminal situations as stated in table

2 and Table 3. The ground truth for such behaviours corresponds to temporal events defined by a start-time and an end-time.

For analysing event detection performance, a behaviour event is deemed as True Positive instance (TP) if it has a temporal overlap with the ground-truth (GT) event (see Figure 3). We call this type of binary event recognition ‘instance evaluation’ (the event is recognised inside the temporal boundaries given by the ground-truth or not). The extent of the agreement between the detected event and the ground-truth can be quantified more precisely as the number of seconds where event and ground-truth overlap. We refer to this as ‘duration evaluation’. Typical ROC measures can then be obtained:

$$TP = |Event \cap GT|, \quad FP = |Event \cap \overline{GT}|, \quad (3)$$

$$TN = |\overline{Event} \cap \overline{GT}|, \quad FN = |\overline{Event} \cap GT|,$$

where $Event$ and GT indicate the presence of a detected or ground-truth event and \overline{Event} and \overline{GT} is the absence of the event.

6. Conclusions

This paper has presented the dataset and vision challenges evaluated under PETS2017. PETS, in this edition, continues the evaluation theme of on-board surveillance systems with two datasets as employed in PETS2016: (1) the ARENA Dataset; an RGB camera dataset, as used for PETS2014 to PETS 2016, which addresses protection of trucks; and (2) the IPATCH Dataset; a multi sensor dataset, as used in PETS2016, 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 datasets should prove useful for authors working on Object Detection and Tracking,

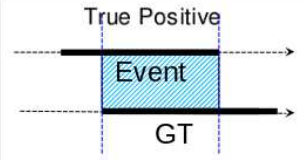
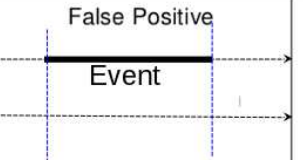
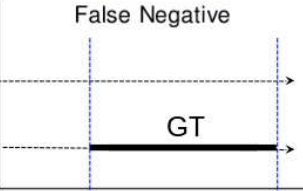
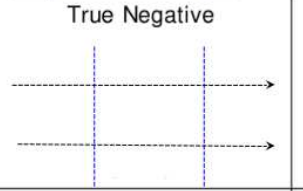
	Reference Data (RD)	Noise (N)	
Event Recognition (ER)	True Positive 	False Positive 	Precision
No Recognition (NR)	False Negative 	True Negative 	
	Sensitivity	Specificity	

Figure 3. Detected surveillance events are evaluated according to their overlap with the ground-truth; also stated in equation 3.

Multi-camera analysis, Visual surveillance, Abnormal activity recognition, Context analysis, Activity analysis and monitoring, and Maritime abnormal event detection among others. Training data for tracking algorithms is released with the dataset; tracking data is also available for authors addressing only surveillance event detection challenges but not working on tracking.

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/>. 3
- [2] PETS 2014 the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, www.pets2014.net. 1, 4
- [3] PETS 2016 the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, www.pets2016.net. 1
- [4] J. Berclaz, F. Fleuret, and P. Fua. Multiple object tracking using flow linear programming. In *IEEE Int. Work. on PETS, Snowbird, UT*, 2009. 6
- [5] K. Bernardin and R. Stiefelhagen. Evaluating multiple object tracking performance: The clear mot metrics. *EURASIP JIVP*, 2008:1–10, 2008. 4
- [6] J. Black, T. Ellis, and P. Rosin. A novel method for video tracking performance evaluation. In *Proc. of IEEE Int. Work. on PETS*, 2003. 4
- [7] D. Conte, P. Foggia, G. Percannella, and M. Vento. Performance evaluation of a people tracking system on the pets2009 database. In *Proc. of IEEE Int. Conf. on AVSS*, Boston, MA, 2010. 6
- [8] J. Ferryman and A.-L. Ellis. Performance evaluation of crowd image analysis using the pets2009 dataset. *Patt. Rec. Lett.*, 44:3–15, July 2014. 5, 6
- [9] L. Leal-Taixe, A. Milan, I. Reid, S. Roth, and K. Schindler. Motchallenge 2015: Towards a benchmark for multi-target tracking. *arXiv:1504.01942*, 2015. 4
- [10] L. Li, T. Nawaz, and J. Ferryman. Pets 2015: Datasets and challenge. In *2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6, Aug 2015. 2
- [11] T. Nawaz, A. Ellis, and J. Ferryman. A method for performance diagnosis and evaluation of video trackers. *Signal, Image and Video Processing*, DOI: 10.1007/s11760-017-1086-7, 2017. 4, 6
- [12] T. Nawaz, F. Poiesi, and A. Cavallaro. Measures of effective video tracking. *IEEE Trans. on Image Processing*, 23(1):376–388, 2014. 4
- [13] L. Patino, T. Cane, A. Vallee, and J. Ferryman. Pets 2016: Dataset and challenge. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2016. 3
- [14] 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. 2, 4
- [15] B. Ristic, B.-N. Vo, D. Clark, and B.-T. Vo. A metric for performance evaluation of multi-target tracking algorithms. *IEEE Trans. on Signal Processing*, 59(7):3452–3457, 2011. 4
- [16] B. Wu and R. Nevatia. Tracking of multiple, partially occluded humans based on static body part detection. In *Proc. of CVPR*, 2006. 4