Abnormal Behavior Detection in LWIR Surveillance of Railway Platforms

Kristof Van Beeck, Kristof Van Engeland, Joost Vennekens and Toon Goedemé
EAVISE, Technology Campus De Nayer, KU Leuven, Belgium
{firstname.lastname}@kuleuven.be

Abstract

In this paper we present a framework that is able to reliably and completely autonomously detect abnormal behavior in surveillance images. As input, we rely solely on a long-wave infrared (LWIR) image sensor. Our abnormal behavior detection pipeline consists of two consecutive stages. In a first stage, we perform efficient and fast pedestrian detection and tracking. In a second step, the detected paths are fed into a semi-supervised classifier that detects abnormal behavior. As test-case we recorded a unique real-life LWIR train station dataset – which will be made publicly available – containing natural occurrences of both normal and abnormal behavior. Our experiments indicate that our proposed framework achieves excellent accuracy results at real-time processing speeds.

1. Introduction

The detection of abnormal behavior in surveillance images has received a considerable amount of attention in recent literature. Indeed, several works are found concerning this topic [1, 8, 9, 10, 13, 21]. In this paper, however, we aim to detect abnormal behavior in long-wave infrared (LWIR) images. For this, we recorded and labeled an extensive LWIR pedestrian dataset focusing on the detection of abnormal behavior at a train station. Fig. 1 displays an example frame of our dataset (with detections – see further), which we made publicly available[1]. The goal of this work is to automatically detect abnormal behavior by fusing two modalities. We first perform pedestrian detection and tracking in the LWIR images. These detection tracks are then automatically evaluated in a second stage, using a Markov model of the normal behavior of passengers on the platform. If the likelihood of a given track according to the model falls below a certain threshold, the track is flagged as abnormal. This method has the benefit that it does not require an exhaustive enumeration of the different kinds of abnormal behavior that should be detected. Indeed, there is a wide range of possible abnormal behavior that should be flagged: the main focus is on suicide prevention (*i.e.*, detecting the typical behavioral pattern exhibited on the platform by people with the intent of committing suicide), but additionally, we should also detect such incidents as people fighting, crossing the train tracks, etc. The main advantage of using LWIR images over more traditional RGB images is that even in low-light and harsh weather conditions (*e.g.*, fog, heavy rain), pedestrians remain clearly visible. Furthermore, since pedestrians are not recognizable, LWIR images inherently avoid privacy issues which exist when utilizing traditional RGB images. However, the detection of pedestrian in LWIR images is more challenging. Indeed, color-sensitive features are not present, and we can only rely on a single gray scale image representation of the radiated heat. In summary, our main contributions are:

Figure 1. LWIR frame from our *Brugge* dataset with detections.

- We recorded and labeled an extensive LWIR pedestrian dataset (the *Brugge* dataset), resulting in 25000 frames in which 30000 pedestrians were labeled.
- We present an efficient pedestrian detection methodology for LWIR images. We apply scene constraints to increase the accuracy of our pedestrian detector. We track all detections and generate pedestrian paths.
- We present a semi-supervised classification methodology, which is able to efficiently detect abnormal behavior using only these generated pedestrian paths.

[1] http://eavise.be/viper/

978-1-5386-2939-0/17/$31.00 ©2017 IEEE

The remainder of this paper is structured as follows. We discuss related work in Section 2. Next, we present details on our complete detection pipeline, the abnormal behavior detection stage, and our LWIR recordings in Section 3. In Section 4, we experimentally validate our approach. Finally, we conclude and discuss future improvements in Section 5.

2. Related work

The detection of abnormal behavior in surveillance images receives a considerable amount of attention in the current literature. Existing work often relies on standard RGB images as input [1, 8, 21]. However, LWIR cameras are increasingly installed in real surveillance applications because of their superior performance during low-light conditions (*i.e*. night), and their decreasing hardware costs. These thermal images are often fused with RGB images to increase the detection accuracy [9, 10]. The work in this paper however relies solely on LWIR images, which significantly increases the difficulty due to the lack of color and texture information. Traditionally, pedestrian detection in LWIR images therefore relies on simple cues. Existing techniques exploit the fact that in most cases a significant difference in temperature is found between a person and the background. Thus, pedestrians are simply detected as *hot spots* using *e.g*., shape measures [19] or Maximally Stable Extremal Regions (MSER) [16]. However, these methods fail when limited contrast is available (*e.g*., during warm weather or heavy rain). To overcome these limitations, in this work we employ an appearance-based LWIR pedestrian detection approach (see subsection 3.2). Furthermore, the detection of abnormal behavior is often analyzed based on the activity analysis of human motion, such as optical flow [13], gait or gestures [18]. Our detection pipeline however uses a Markov model to detect abnormal behavior. Such Markov models are used in several contexts, for example [20] uses a Markov-chain based approach for analyzing observed activities in a computer - network system to detect cyber attacks, and [12] for characterizing specific behaviors in trajectories. In [11], the authors represent routes of pedestrians as sequences of probability distributions around a central axis. Their approach works well to determine if the place where someone is walking is abnormal, but seems to offer little solution for errant behaviors that still follow normal routes. In [6] a Markov model technique is applied to analyze behavioral patterns in movement. This method forms the base of our approach to separate normal and abnormal behavior. However, the model they used is not very performant in situations where pedestrians tend to stand still or walk around in circular patterns, as is often the case in our railway platform environment. Furthermore, [5] attempts to detect abnormal behavior using a modified probabilistic neural network. The position of every person's head is tracked, and its speed magnitude is used as input feature. The results are accurate, but training time is long and increases dramatically with more samples. The fact that it uses the head's speed as the only feature also limits the usefulness for our case.

3. Technical details

As previously discussed, our detection pipeline consists of two consecutive steps. In this section we now discuss all technical details of both the pedestrian detection stage (subsection 3.2) and the abnormal behavior classification stage (subsection 3.3). However, we first present an overview of the technical details of our LWIR dataset.

3.1. Dataset statistics

To validate and train our abnormal classification system we recorded and labeled an extensive LWIR image dataset. Our dataset was recorded at the train station of Brugge (Belgium), in cooperation with the thermal imaging company FLIR Systems. As mentioned, the use of LWIR images facilitates the detection of pedestrians during difficult conditions (*e.g*., low-light, heavy fog or rain) since it relies on the radiated body heat. An example frame of our dataset, coined the *Brugge dataset*, is shown in Fig. 1. The images have a resolution of 640×512 and a frame rate of 7 frames per second. In total we labeled 24831 frames, divided over 27 videos. In the first 14 videos pedestrians cross the train tracks. The remaining 13 videos include mostly people strolling the railway platforms, waiting for their train to arrive. Occasionally a person crossing the train tracks is seen. In total 30128 pedestrians were labeled, of which 4523 pedestrians were occluded (an *occluded* flag is included in the labels). Aside from individual pedestrian labels, all pedestrian paths were also uniquely labeled, resulting in 79 track IDs. Our dataset is subdivided into a training and validation set. This is done in an interleaved manner to ensure that the training and validation set contain the same ratio of normal versus abnormal behavior. For this, we assigned all 27 videos alternately for training and testing. In the next subsections we now use this dataset to train, perform detection and evaluate both our pedestrian detector and the final complete abnormal behavior detection scheme.

3.2. Pedestrian detection in LWIR images

A commonly-used method for detecting pedestrians in fixed-camera surveillance images is background subtraction, often using relatively simple background estimation techniques, such as Gaussian Mixture Models (GMM) [22] or Fuzzy background subtraction [14]. These methods are not usable on our LWIR images. Indeed, while standard GMM techniques might sometimes work, they suffer from a multitude of problems. First, pedestrians walking close to each other are often merged. Second, the LWIR camera performs auto-contrast to achieve optimal temperature con-

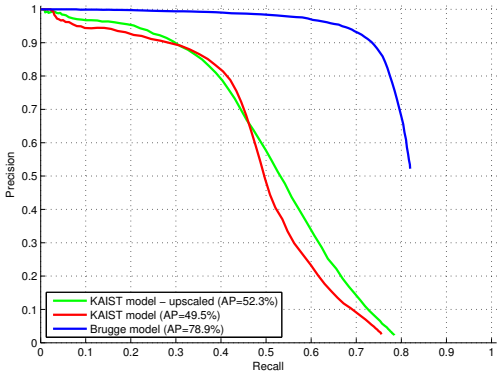

Figure 2. Accuracy of the KAIST model (red/green) and Brugge model (blue) on our LWIR *Brugge* validation set.

trast, which may cause the background to suddenly reappear as foreground. Third, selecting an appropriate "memory size" (*i.e*. number of frames) is also problematic in our application. In order for passing trains to be classified as background, this memory should be relative short, but then waiting passengers—who tend to stand still for relatively long periods of time—will also blend into the background. Therefore, we do not consider background subtraction an appropriate method for our application. Instead, we employ a rigid pedestrian detection methodology, which relies on appearance-based features. In particular, we use the well-know *Aggregated Channel Features* (ACF) detector, introduced by Dollár *et al*. [3]. This detector works as follows. First, ten *feature channels* are calculated from the original image: three color channels (LUV), six gradient orientation histogram channels, and a gradient magnitude channel. Then, a sliding window approach is applied to the entire input image (and at multiple scales), in which multiple weak classifiers (decision trees) are evaluated using specific positions in the feature channels.

This evaluation is performed in a cascaded manner: the decision trees are evaluated sequentially and evaluation for a specific window stops when the score drops below a specific threshold. The specific features, sequence and threshold are learned during an off-line training phase. However, traditionally these detection models are trained on standard RGB images. Therefore, for our application we first need to train a detection model using only LWIR images, which we input as gray scale images. To obtain preliminary detection results, we first trained an LWIR detection model (coined the *KAIST model*) using training data from the publicly available KAIST dataset [4]. This KAIST dataset is a multispectral pedestrian dataset; it consists of about 100.000 image pairs containing both RGB and LWIR images. The images are captured using a color camera, thermal camera and beam splitter (and thus are perfectly aligned). For training, we learned a detection model of 50×20.5 pixels (proven to be ideal according to [2]), aiming for 4096 weak classifiers consisting of depth-5 decision trees. Every 20th frame

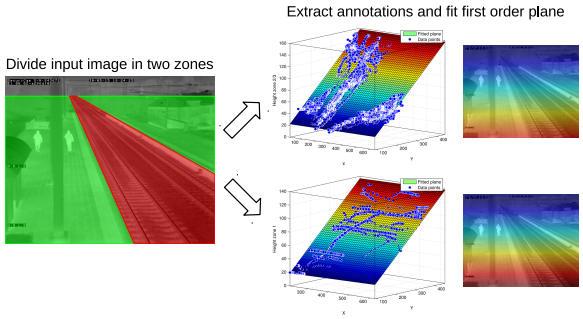

Figure 3. Employing ground plane constraints. We define two ground planes and fit a first order plane through the annotations.

was used for training, resulting in 2500 images containing 2120 annotations. A total of 3496 positive patches were retrieved (annotations were mirrored and pruned if too small), while about 50.000 negatives were randomly sampled in the dataset. The training automatically stopped after 2701 weak classifiers since the miss-rate showed no further significant decrease. As a qualitative result, Fig. 1 visualizes the detection output of the trained KAIST model on a single frame of our platform validation data. Fig. 2 displays the entire *Precision-Recall* curve (red) of this model on our data set.

As seen, we achieve an average precision (AP) of 49.5%. Since the model size is 50 pixels, upscaling is needed to detect smaller pedestrians. We opted to upscale the image such that pedestrians up to 40 pixels (representing 80% of all annotations) are detected, resulting in the green curve. Although slightly better, the accuracy is far from optimal. This is of course not surprising, because the training and evaluation dataset are significantly different [17]. Therefore, we trained a new ACF LWIR model using our own recorded *Brugge* training set. The model specifications and number of negatives remain identical, and every 5th image was used for training (resulting in 2921 annotations and 4212 positive patches). The accuracy of this detection model—coined the *Brugge model*—is shown as the blue curve (again for pedestrians up to 40 pixels). As can be seen, we now achieve excellent accuracy. For example, at a precision of 90%, the recall is 75% (AP = 78.9%).

To further increase the detection accuracy, we exploit *scene constraints*. For this, we assume a fixed and flat ground plane. This implies that—for the specific camera viewpoint in our application—a linear relationship exists between the height of the pedestrians and their location in the image [15]. Thus, after a one-time calibration (based on the annotations), we know the expected height at each position in the image, and can reject detections which significantly diverge from this expected height. This concept is illustrated in Fig. 3. We define two ground planes (one for the platform and one for the train tracks, due to their difference in height), extract the height for all annotations in each zone and fit a first order plane through these data points. Both planes are then used as lookup function (LUF): for

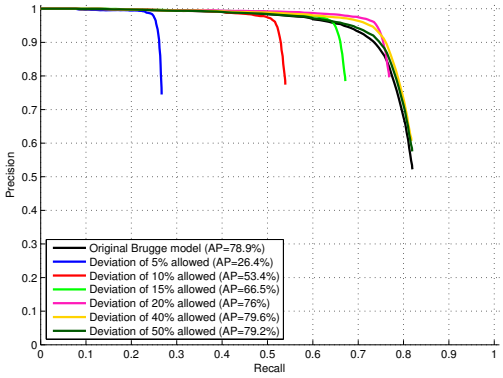

Figure 4. The detection accuracy with ground plane estimation.

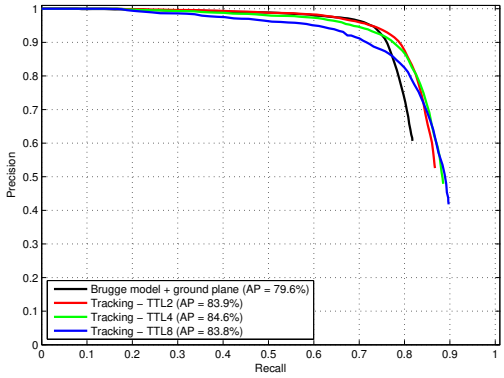

Figure 5. The detection accuracy with tracking.

each new detection we validate whether the divergence is tolerable or not. Fig. 4 displays the accuracy results (compared to the original Brugge model, drawn in black) versus different levels of allowed deviation.

If we only allow a small deviation, the recall drops significantly (since correct detections are also discarded – blue curve). On the other hand, if much deviation is allowed, almost no detections are discarded and we converge to the original detection accuracy (dark green curve). An optimal point is found for an allowed deviation of about 40% (yellow curve). Here, the accuracy significantly increases: *e.g.*, for a recall of 75%, the precision increases by about 5%.

Finally, we integrate temporal information in our detection pipeline using a *tracking* approach: we aim to assign a single tracking ID to each unique pedestrian, which is then fed into our abnormal behavior classification stage. Furthermore this approach enables us to cope with missing detections (due to *e.g.*, a low gradient). For this, we employ the well-know Kalman tracker [7]. A constant velocity motion model is used (with state vector $x_k = \begin{bmatrix} x & y & v_x & v_y \end{bmatrix}^T$, using the center of mass of the detections). We perform the aforementioned pedestrian detection scheme on each new input image, and evaluate for each detection if it matches a predicted pedestrian track using the Euclidean distance. When no match is found, a new track is started. When a track is not matched for a number of frames in a row—called the *Time To Live* (TTL)—the track is discarded. Fig. 5 displays the accuracy when tracking is included, for multiple values of the TTL (compared with our best ground plane constraint model in black). Again, an optimal value for the TTL needs to be determined. For low values of the TTL, a smaller increase in recall is observed, whereas for higher values the precision drops (since false detections are tracked as well). Our best implementation achieves an average precision of 84.6%. Concerning the detection speed, we achieve on average 8 frames per second on the upscaled images, and 10.98 frames per second if no upscaling is performed (evaluated on a CPU only implementation, running on an Intel Xeon E5-2687 at 3.1GHz).

3.3. Abnormal behavior detection

We are mainly interested in detecting three different behavioral patterns, namely suicidal demeanor, reckless and/or aggressive movements, and crossing the railroad tracks. The latter could be detected rather easily by adding a regional exclusion zone to the pixel area corresponding to the railroad tracks and flagging every detection in this zone as an abnormality. However, this is much harder to do for the former two patterns. Furthermore, the definition of such a zone is an extra manual tuning step which we would have to redo for every different track or camera viewpoint. Rather than defining precisely the kind of abnormal behavior that is to be detected, we therefore adopt a different approach: we construct a model of the *normal* behavior and flag each behavior that deviates significantly from this norm. The details of this approach are as follows.

First, we automatically divide the camera image into separate regions. By doing this, we reduce the spacial state space from $w \times h$ to n, with w and h the width and height of the camera view (in pixels) and n as the number of regions. In order to make our system fully automated, this step is preformed by a clustering algorithm, which clusters the different location at which pedestrians have been detected into larger regions. In particular, we use structured agglomerative clustering. This is a bottom-up hierarchical clustering approach in which a dendrogram is constructed by repeatedly merging the clusters with the smallest spacing between them. The structure, provided by a k-nearest neighbor graph computed on a representative sample of detection coordinates, adds connectivity constraints to the clusters and also decreases the calculation time. We measure the distance between individual detection coordinates a, b using the Euclidean distance $d(a, b)$, and the distance between different clusters A, B using the average linkage criterion $D(A, B)$:

$$d(a,b) = \|a-b\|_2 = \sqrt{\sum_{i=1}^{2}(a_i - b_i)^2} \qquad (1)$$

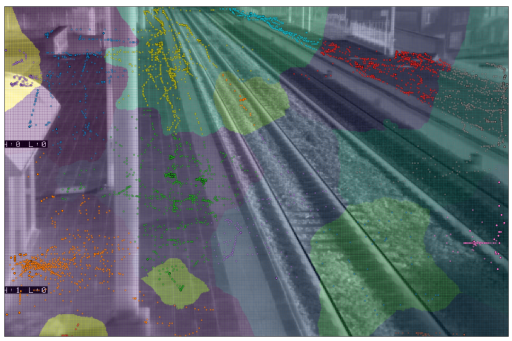

Figure 6. Clustering ($n = 15$) applied to the *Brugge* dataset.

$$D(A,B) = \frac{1}{|A||B|} \sum_{a \in A} \sum_{b \in B} d(a,b) \qquad (2)$$

The corresponding clusters for the other pixels in the image are generated by a k-nearest neighbors classifier, which uses the detections sampled and clustered earlier as training input. This approach generally leads to clusters of different sizes, where large clusters contain a large number of detection points and small clusters a small number. An example of this clustering can be seen in Fig. 6. Once this clustering has been computed, we use it to transform each pedestrian track from sequence of pixel coördinates $((x_0, y_0), \ldots, (x_n, y_n))$ into a sequence of regions $(r_0, \ldots, r_n)$. Note that it may be the case that $r_i = r_j$ for subsequent steps $i > j$. In particular, for slow-moving pedestrians, this will often be the case.

Second, we then use these sequences to estimate an initial position probability matrix and a transition probability matrix, forming the base of a Markov chain. The dimensions of the matrices are $m \times 1$ for the initial position probabilities and $m \times m$ for the transition probabilities, with m the number of clusters. To learn these matrices, we use a maximum likelihood estimation method[2]. Once the matrices have been learned, we can then use the Markov model to calculate the probability of a particular sequence. Three observations about the probability of a sequence can be made when examining the output of the model evaluation: (i) it decreases slightly with increasing length, (ii) it decreases strongly when it incorporates a region with a low transitional probability, (iii) it decreases strongly when it passes through many different regions. These observations are in accordance with the kind of abnormal behavior we want to detect: people with suicidal intent tend to walk the entire length of the platform to its very end (iii), where they pace back and forth during a long time (i); and people who cross the railroad tracks are relatively rare, so the "forbidden" zones have a very low transitional probability (ii).

Finally, we determine the threshold value P_{th} that defines the lowest sequence probability that is still considered as normal behavior. This value depends heavily on the nature of the dataset (*i.e.* number of clusters and the order of the Markov model. In order to choose a P_{th} that provides a good working point for our algorithm, we make use of the following method. First, we estimate the density of the calculated path probabilities using a Gaussian kernel. Second, we use this density function to calculate the cumulative distribution function $F(x)$ and its inverse, the quantile function $Q(p) = \inf\{x \in \mathbb{R} : p \leq F(x)\}$. This function $Q(p)$ maps a selected fraction p of the most improbable paths to the corresponding maximum probability of this subset of paths. By choosing a specific fraction of these ordered estimated path probabilities, we can thus easily calculate the corresponding P_{th}, to which all future path probabilities will be compared. All paths with a lower probability than P_{th} will be considered abnormal.

[2] https://github.com/jmschrei/pomegranate

4. Experimental results

In this section, we evaluate the effectiveness of our complete abnormal behavior detection pipeline. Based on discussions with domain experts, we identified different kinds of abnormal behavior (*e.g.*, the kind of behavior typical of people contemplating suicide, people crossing the railroad tracks) to be detected in the footage. We then manually labeled all pedestrian paths which were generated during the detection stage (using the specific detection threshold for P=91.5%, R=77.8%) as normal or abnormal accordingly. Of the 154 pedestrian tracks that were generated, we labeled 36 paths as abnormal. Fig. 7 displays an ROC curve of our final classification algorithm, where we varied the threshold P_{th} set to distinguish abnormal behavior (green curve). As can be seen, our method has reasonable accuracy, but there is still room for improvement.

The prediction errors made by our system may be due to either the classification method or the pedestrian tracking component. While Section 3.2 has shown that our pedestrian detector in itself has excellent detection accuracy, it may still be the case that track IDs get lost (*e.g.*, when a passenger temporarily leaves the frame) or switched (*e.g.*, due to occlusions). To further analyze our results, we have therefore also conducted a second experiment, in which we ran the classification component on the annotated pedestrian paths in our dataset (*i.e.*, the ground truth used to evaluate our pedestrian tracker), rather than on the detected paths. The classification accuracy on these annotated paths is shown as the red curve in Fig. 7. We find that this indeed improves the results, which demonstrates that part of the limited prediction accuracy of our entire pipeline was indeed due to problems with the pedestrian tracking.

In general, the prediction task for this dataset is quite challenging. To illustrate, Fig. 8 displays all paths in the dataset, color-coded according to their status w.r.t. our classification algorithm (for a specific value of P_{th}, such that TPR=83.3% and FPR=20.4%, see caption for color cod-

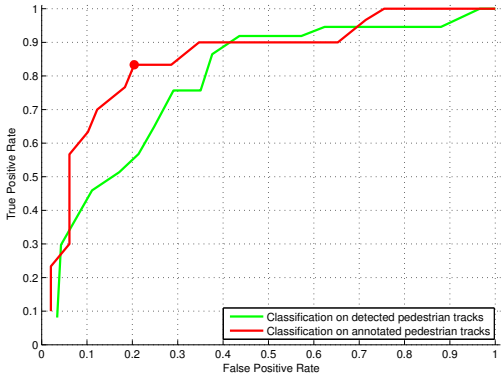

Figure 7. Accuracy of our behavior classification system.

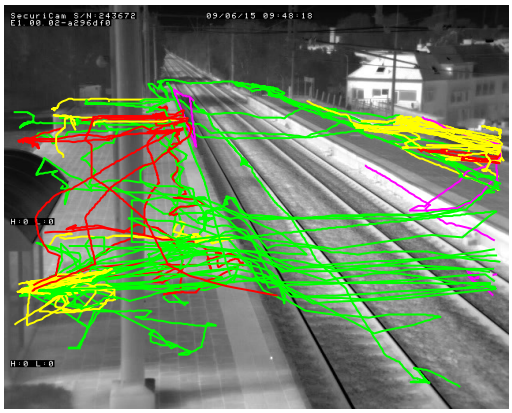

Figure 8. Classification result for all paths. Green: true positive, red: false positive, yellow: true negative, magenta: false negative.

ing). As can be seen, our classification approach is able to correctly classify most paths, including most of the "easy" examples of abnormal behavior that involve crossing the railroad tracks. In addition, our method is also able to correctly detect some instances of the abnormal behavior typical of people contemplating suicide. However, several other paths—typically those that are longer than usual—erroneously also get labeled as such, mostly on the left platform.

5. Conclusion and future work

In this paper, we presented a framework to autonomously detect abnormal behavior in LWIR surveillance images of railway platforms. Our framework consists of two consecutive stages: we first perform reliable pedestrian detection, and then classify the detection paths using a Markov model. We are able to correctly classify most detected pedestrian tracks. Our approach is easily generalizable to other applications where abnormal behavior needs to be detected (*e.g.*, shoplifting). However, several future optimizations exist. In some cases the track assignment fails (*e.g.*, due to occlusion). Furthermore a different spatial segmentation approach might be evaluated.

References

[1] B. Benfold and I. Reid. Stable multi-target tracking in real-time surveillance video. In *Proc. of CVPR*, 2011. 1, 2

[2] F. De Smedt, S. Puttemans, and T. Goedemé. How to reach top accuracy for a visual pedestrian detection warning system from a car? In *Proc. of IPTA*, 2016. 3

[3] P. Dollár, R. Appel, S. Belongie, and P. Perona. Fast feature pyramids for object detection. *In PAMI*, 2014. 3

[4] S. Hwang, J. Park, N. Kim, Y. Choi, and I. So Kweon. Multispectral pedestrian detection: Benchmark dataset and baseline. In *Proc. of CVPR*, 2015. 3

[5] T. Jan. Neural network based threat assessment for automated visual surveillance. In *Proc. of IJCNN*, 2004. 2

[6] N. Johnson. *Learning object behaviour models*. PhD thesis, University of Leeds, 1998. 2

[7] R. E. Kalman. A new approach to linear filtering and prediction problems. *In JBE*, 1960. 4

[8] N. Kiryati, T. R. Raviv, Y. Ivanchenko, and S. Rochel. Real-time abnormal motion detection in surveillance video. In *Proc. of ICPR*, 2008. 1, 2

[9] P. Kumar, A. Mittal, and P. Kumar. Fusion of thermal infrared and visible spectrum video for robust surveillance. In *Proc. of CVGIP*. 2006. 1, 2

[10] A. Leykin and R. Hammoud. Pedestrian tracking by fusion of thermal-visible surveillance videos. *In MVA*, 2010. 1, 2

[11] D. Makris and T. Ellis. Spatial and probabilistic modelling of pedestrian behaviour. In *Proc. of BMVC*, 2002. 2

[12] J. C. Nascimento, M. A. Figueiredo, and J. S. Marques. Trajectory classification using switched dynamical HMMs. *In IEEE Transactions on Image Processing*, 2010. 2

[13] N. Rasheed, S. A. Khan, and A. Khalid. Tracking and abnormal behavior detection in video surveillance using optical flow and neural networks. In *Proc. of WAINA*, 2014. 1, 2

[14] M. H. Sigari, N. Mozayani, and H. R. Pourreza. Fuzzy running average and fuzzy background subtraction: Concepts and application. In *Proc. of IJCSNS*, 2008. 2

[15] P. Sudowe and B. Leibe. Efficient use of geometric constraints for sliding-window object detection in video. In *Proc. of ICVS*, 2011. 3

[16] M. Teutsch, T. Muller, M. Huber, and J. Beyerer. Low resolution person detection with a moving thermal infrared camera by hot spot classification. In *Proc. of CVPRW*, 2014. 2

[17] A. Torralba and A. A. Efros. Unbiased look at dataset bias. In *Proc. of CVPR*, 2011. 3

[18] C. Wang, X. Wu, N. Li, and Y.-L. Chen. Abnormal detection based on gait analysis. In *Proc. of WCICA*, 2012. 2

[19] W. Wang, J. Zhang, and C. Shen. Improved human detection and classification in thermal images. In *Proc. of ICIP*, 2010. 2

[20] N. Ye, Y. Zhang, and C. M. Borror. Robustness of the Markov-chain model for cyber-attack detection. *In IEEE Transactions on Reliability*, 2004. 2

[21] Y.-Y. Zhu, Y.-Y. Zhu, W. Zhen-Kun, W.-S. Chen, and Q. Huang. Detection and recognition of abnormal running behavior in surveillance video. *In MPE*, 2012. 1, 2

[22] Z. Zivkovic. Improved adaptive Gaussian mixture model for background subtraction. In *Proc. of ICPR*, 2004. 2