Man Overboard Event Detection from RGB and Thermal Imagery: Possibilities and Limitations

Iason Katsamenis†
National Technical
University of Athens
Athens, Greece
iasonkatsamenis@mail.ntua.gr

Eftychios Protopapadakis
National Technical
University of Athens
Athens, Greece
eftprot@mail.ntua.gr

Athanasios Voulodimos
University of
West Attica
Athens, Greece
avoulod@uniwa.gr

Dimitris Dres
Telesto Technologies
Athens, Greece
jdres@telesto.gr

Dimitris Drakoulis
Telesto Technologies
Athens, Greece
dimitris@telesto.gr

ABSTRACT

A man overboard is an emergency incident, in which fast detection is the most crucial factor, for the quickest and most efficient recovery of the victim. As such, efficient monitoring methodologies should be employed. A variety of sensors is available today, supporting a continuous monitoring process, regardless of environmental conditions; RGB and thermal are two commonly used sensors. At the same time, several algorithms and techniques have been tested and proved to be efficient in human detection and situation recognition tasks. However, to this day a coherent methodology for fall detection over multiple sensors on a large-scale deployment, complying with related ISO standards on extremely low false positive alerts, has not been implemented. In this paper, we investigate the possibilities as well as the limitations of man overboard vision-based systems' development based on RGB and thermal imagery.

CCS CONCEPTS

• Computing methodologies → Artificial intelligence → Computer vision → Computer vision problems → Object detection • Computing methodologies → Artificial intelligence → Computer vision → Computer vision tasks → Scene anomaly detection

KEYWORDS

Man overboard, Human detection, Deep learning, Computer vision, Thermal imaging

†Corresponding author

PETRA '20, June 30-July 3, 2020, Corfu, Greece © 2020 Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-7773-7/20/06…$15.00 https://doi.org/10.1145/3389189.3397998

1 Introduction

A man overboard is an emergency situation where a member of the ship's crew or a passenger has fallen off the ship, in the sea. Each year, approximately 22 people fall off a cruise ship and 79% of them do not survive or are considered missing [1]. The fact that when an individual remains for 1 hour in water at 4.4°C and their body temperature drops at 30°C, is the main reason for such a small survival rate [2]. Thus, it is a critical event that demands immediate handling as time plays an important role and because the overboard casualty is exposed to various security risks, such as drowning at sea, hypothermia, injuries and rough sea. It is noted that the problem lies in the lack of timely and critical information, such as the accurate confirmation of the event as well as its exact time and position of the occurrence.

Traditionally, maritime surveillance systems consist of optical cameras, which are programmed to monitor statically predetermined positions of the security perimeter. However, since emergency response and rescue time play a crucial role in such incidents, a wide area multi-sensor approach (e.g. [3]) is more suitable, that is able to perform accurately under adverse conditions, such as extreme weather events (e.g. thunderstorms, dust storms, blizzards, tornadoes, rough sea) and various illumination situations (e.g. overexposure, underexposure, low light, foggy and cloudy environments). Moreover, it is noted that following the alarm, wide-area sensors are capable of scanning large areas at a 360° angle and therefore, are able to provide continuous monitoring of the man overboard event, enhancing the effectiveness and reducing the response time of the rescue team.

Over the last decades, interest, research and development of computer vision systems have grown rapidly. Optical cameras (e.g. cellphones, UAVs, DSLR and CCTV cameras), which capture visible light in grayscale or RGB images, are the most common visual sensors. However, these sensors conceal some serious disadvantages due to the fact that the object's visibility is inextricably linked to the light source (e.g. sun, artificial light sources) [4]. Hence, the main challenge that arises is to overcome illumination problems of RGB and grayscale cameras (e.g. overexposure, underexposure, color balance, total darkness, near-instantaneous switching of light intensity and direction).

Many of the aforementioned limitations can be addressed through the use of thermal cameras since human is a warm-blooded organism, a property that distinguishes him from the environment around him in thermal images. Thus, their use is an important aspect of computer vision systems related to human detection [5] and furthermore, if the observation is made at a close distance, we are able to extract information related to skin temperature distribution, which may be critical in processes such as face recognition [6], [7].

Although thermal sensors hide some significant advantages over visual cameras that are related to illumination conditions during the critical incident, it is underlined that there are cases where the optical cameras are superior and therefore their presence in surveillance applications is considered necessary. For instance, in a thermal image, it is extremely difficult to separate two or more overlapping human silhouettes, since their pixels exhibit the same or similar intensities [4]. However, by utilizing the color data and contrast of the pixels, as well as the depth information derived from the optical sensors, we can overcome this obstacle. It is thus immediately apparent that optical and thermal cameras have both advantages and disadvantages in their use in computer vision applications. To get the best of both worlds and since the constraints of the two technologies are independent and mostly do not occur simultaneously, it is extremely beneficial to combine them. By utilizing the high resolution and low-cost optical camera's color characteristics and then enriching them with the thermal information, the system will be able to cope with any man overboard scenario with high precision.

Traditional surveillance systems, that have been developed for maritime safety purposes, require an operator who simultaneously monitors multiple real-time videos. This obviously results in an increased likelihood of error or inadequate response, due to reduced visibility. Hence a continuous monitoring approach and evaluation of the critical event should be adopted, in parallel with the precise detection of its position. That can be achieved through the development of signal processing and deep learning algorithms and frameworks for semantic information extraction (human detection and recognition, motion tracking [8], pose recognition [9], [10], anomaly detection) from heterogeneous information sources (RGB and thermal cameras). Consequently, through the implementation of multimodal information fusion techniques, it is possible to build an advanced decision support system for maritime surveillance applications and man overboard events. Given the heterogeneity of data flows originating from different optical sensors (RGB and thermal cameras), pattern recognition and deep learning algorithms appropriate for the different data modalities must be applied.

The remainder of this paper is organized as follows. Section 2 discusses related work in human detection frameworks, through RGB and thermal imaging. Section 3, briefly presents the explored methods for man overboard detection: Outlier detection based on Density-based spatial clustering of applications with noise (DBSCAN) algorithm and human detection based on Histograms of Oriented Gradients (HOG), Haar Cascade Classifier (HCC) and You Only Look Once–v3 (YOLOv3). Furthermore, we analyze the possibilities and highlight the limitations of the aforementioned methodologies. Finally, in Section 4, the conclusion is summarized and suggestions for future research are introduced.

2 Previous Work

In a universal maritime surveillance system, human detection is a key issue and must be completely independent of the environment as well as light and weather conditions. Hence a thermal camera is often superior to optical cameras. Several human detection methods through thermal imaging have been presented in the literate. Many of them exploit features extracted from Histograms of Oriented Gradients (HOG), in combination with various classification models (e.g. Support-vector machines (SVMs) [11], AdaBoost [12] and adaptive fuzzy C-means clustering [13]) in infrared images for pedestrian detection. The work of [14] implemented a person detection system in thermal video sequences by combining background-subtraction, gradient information, watershed algorithm and A* search, whereas in [15] the authors investigated segmentation of the head regions for human detection in far-infrared images. In the method of [16], Contour Saliency Maps and adaptive filters are utilized for person detection in thermal images. Several other approaches have also been proposed in the literature, including [17] and [18], yet these methodologies are based on the fact that body and surrounding temperatures differ significantly. This limitation, that arises especially during summer, can be managed by using Mahalanobis distance for each pixel in combination with edge orientation histogram [19].

Finally, it is noted that several studies have emphasized the importance of real-time home surveillance systems ([20], [21]) that focus on fall detection through visual sensors, deep learning and computer vision applications (e.g [22]–[24]), however, little work has been presented in the literature on the man overboard situation.

3 Explored Techniques for Man Overboard Detection

In the paper at hand, two types of approaches, complementary to each other, were considered: (i) outlier detection and (ii) human detection. The former case, which is responsible for the alert triggering, is easy to employ methodology and requires low resources. The latter case is any approach capable to detect a person (i.e. bounding box or pixel-level segmentation). If we have an alert and positive detection of a person, the operators are informed in order to recover the casualty.

3.1 Outlier detection-based approaches

Today, multiple video streams from RGB cameras are the common scenario; low-cost sensors, easy to install and accessibility are a few of the related advantages. However, processing the provided information is not an easy task; especially when we try to monitor something vague [25]. Towards that direction, a general detector, indicating unordinary behavior appears a viable approach. The core idea lies in detecting the change between successive frames [26]. Dense changes in the content indicate possible events worthy of detection. Such an approach offers two significant advantages: (i) limited requirement of hardware resources, i.e. easy on-field deployment and real-time operation.

Let us assume a video sequence of n frames. Every frame i is divided into m non-overlapping image patches. Each patch is then described by m features, which refer to contrast, energy, homogeneity and dissimilarity. Then, these patches are clustered using DBSCAN [27], a density-based approach. Finally, by comparing the number of clusters or the number of outliers between consecutive frames, we get an indication of something unordinary that appeared, i.e. a significant change in the number of clusters or outliers. Once the alert is activated, deep learning approaches will scan the latest frames of the sequence and inform the operators about the current situation.

(a) No outliers detected. (b) Two outliers detected.

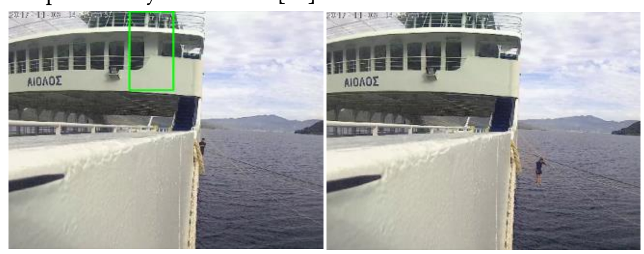
Figure 1: Outlier detection during a man overboard event.

Figure 1 compares two representative frames of a test RGB video before (Fig. 1a) and after (Fig. 1b) the man overboard situation. In the former, there are no outliers detected, while on the latter, due to the significant differences between the consecutive frames (i.e. human fall off and turbulence at sea) the system indicates that something unusual has happened. When a certain amount of outliers, beyond a predetermined threshold, is detected, the surveillance framework must trigger deep learning algorithms in order to decide if there is indeed a man overboard incident (and not i.e. a passing-by bird from the camera) and if so, its exact location for the immediate recovery of the victim.

3.2 Object detection-based approaches

3.2.1 Histograms of Oriented Gradients (HOG). HOG was a popular method used in human detection applications [28]. A detection window slides across an image frame wherein a grid of cells is created. A histogram of edge orientations, for each cell, is extracted, allowing the identification of objects, described through the distribution of local intensity gradients. The cells are inside larger blocks which are used to overcome illumination variations.

HOG-based detectors using the multi-scale sliding window mechanism have long been the dominant approaches for pedestrian detection. While no single hand-craft feature has been shown to outperform HOG, the combinations of HOG with other features have made improvements by making use of complementary visual cues [29].



(a) False positive detection. (b) False negative detection.

Figure 2: Unsuccessful detection of human fall off from a ship, using HOG.

As highlighted in Figure 2, it is impossible to predict a man overboard situation by using solely HOG features since multiple false or non-call alarms occurred (see Fig. 2a and 2b respectively). The specific fact contradicts with the ISO/PAS 21195 standard, which specifies all the technical requirements for systems designed to detect a person while going overboard from a passenger or cruise ship.

3.2.2 Haar Cascade Classifier (HCC). HCC is a machine learning object detection algorithm where a cascade function is trained from a lot of positive and negative images. It is then used to detect objects in other images [30]. The cascade classifier consists of a collection of stages, where each stage is an ensemble of weak learners. The weak learners are simple classifiers called decision stumps. HCC is trained using boosting, which provides the ability to train a highly accurate classifier by taking a weighted average of the decisions made by the weak learners.

Each stage of the classifier labels the region defined by the current location of the sliding window as either positive or negative. Positive indicates that an object was found and negative indicates no objects were found. If the label is negative, the classification of this region is complete, and the detector slides the window to the next location.

Figures 3a and 3b pinpoint that HCC is able to locate a human fall off with high accuracy. However, similar to the HOG-based technique, it triggers numerous false alarms (see Fig. 3c), in contrary to preconditions of the ISO/PAS 21195 standard. Moreover, in spite of the fact that HCC can detect a person before and during their fall off, it is not capable of recognizing them while they are in the water (see Fig. 3d). This happens because the algorithm does not locate many of the distinct features that characterize a human (e.g. arms and legs). This drastically reduces the reliability of the video maritime surveillance system since (i) the possibility of detecting a person gone overboard is lowered and (ii) it is not able to locate the exact position of the victim making their rescue more difficult.

(a) True positive detection. (b) True positive detection.

(c) False positive detection. (d) False negative detection.

Figure 3: Partially successful detection of a man overboard situation, but with possible false alarms, based on HCC.

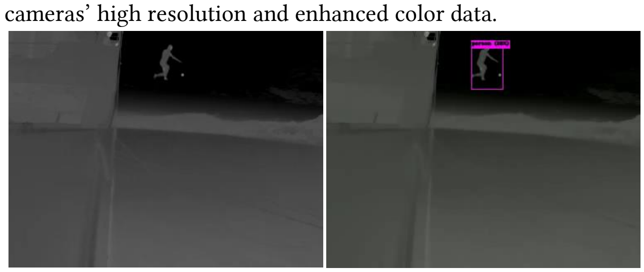
3.2.3 Single Shot Detectors (SSD). SSDs (i.e. YOLO), in contrast to other algorithms (i.e. RCNN family) that focus on certain parts of the image with a higher probability of containing an object, use the whole image in one go in order to detect multiple objects in it. In particular, an SSD predicts several bounding boxes with their corresponding class probabilities and coordinates, a fact that enhances the framework's performance in terms of accuracy and speed.

The YOLO framework was proposed in 2016 and unlike other region-based methods, passes the input image only once to a Fully Convolutional Network (FCN) for prediction purposes [31]. It is underlined that it is able to process images at more than 45 frames per second, which is a significant advantage in real-time applications, such as the man overboard event. Afterwards, YOLOv2 managed to overcome the obstacles of localization errors and relatively low recall, comparing to other region-based algorithms [32]. More recently, in YOLOv3 the softmax function is replaced by independent logistic classifiers for every class and consequently, the model maintains outstanding speed but is also more accurate, comparing to its predecessor [33]. Additionally, it is noted that in YOLOv3, we observe significant improvement in small object detection, a fact that may be crucial in an overboard event that takes place far away from the predetermined location of the surveillance camera. In this paper, we discuss a performance evaluation analysis of the YOLOv3 algorithm for a challenging but also critical human detection and tracking problem, such as the man overboard situation.

(a) Original RGB frame extracted from a test video. (b) Falling off human detected by YOLOv3.

Figure 4: Fall from ship detection, during day time, using an RGB camera and YOLOv3.

Figure 4 illustrates a human during his fall off the ship under good lighting conditions (e.g. daylight). Figure 4a is an extracted frame from an RGB video footage, whereas Figure 4b shows the corresponding YOLOv3 output, that has detected the falling person with high confidence. In such situation, the YOLOv3 algorithm in conjunction with optical sensors, constitutes the optimal man overboard detection system, due to the RGB cameras' high resolution and enhanced color data.



(a) Original thermal frame extracted from a test video. (b) Falling off human detected by YOLOv3.

Figure 5: Fall from ship detection, during night time, using a thermal camera and YOLOv3.

Nevertheless, optical sensors are almost impossible to cope with low light conditions and therefore, thermal imaging is considered necessary for an efficient video maritime surveillance system. In Figure 5a we can see a man overboard situation during night hours and Figure 5b clearly depicts that the algorithm has successfully detected the falling individual.

(a) Original RGB frame extracted from a test video. (b) Passenger detected by YOLOv3.

Figure 6: Passenger detection, during day time, using an RGB camera and YOLOv3.

(a) Original thermal frame extracted from a test video. (b) Passenger detected by YOLOv3.

Figure 7: Passenger detection, during night time, using a thermal camera and YOLOv3.

It is noted that human detection solely must not activate any kind of alarm. The alerts should be triggered if the position of the human, e.g. centroid coordinates of the bounding box, is located outside the safety region. That way, cases as in Fig. 6 and 7, where just an onboard passenger is detected, should not trigger an alert.

(a) Original RGB frame extracted from a test video. (b) Man overboard detected by YOLOv3.

Figure 8: Man overboard detection for ship safety, using an RGB camera and YOLOv3.

Given the orientation of the cameras and the geometry of the ship, fall detection is a trivial matter; if a person is detected outside a predefined area of the image (see Fig. 8) then operators have to deal with a dangerous situation. However, human detection, with minimum or no false-negative rates, requires sophisticated deep learning source-consuming solutions.

4 Conclusions

In this paper, we evaluated different approaches for man overboard detection. The employed techniques ranged from traditional approaches to deep learning ones. Early stage results indicate that none of these approaches alone can be used for an efficient man overboard detection system. However, the trade-offs for each approach can be mitigated by combinations of more than one. The type of combination and the expected outcomes were introduced. Future work should concentrate on exploring the possibility of using additional information modalities, such as radar signals, to improve maritime surveillance system's performance.

ACKNOWLEDGMENTS

This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T1EDK-01169)

REFERENCES

[1] E. Örtlund and M. Larsson, "Man Overboard detecting systems based on wireless technology," 2018.

[2] A. Sevin, C. BAYILMIŞ, İ. ERTÜRK, H. EKİZ, and A. Karaca, "Design and implementation of a man-overboard emergency discovery system based on wireless sensor networks," *Turk. J. Electr. Eng. Comput. Sci.*, vol. 24, no. 3, pp. 762–773, 2016.

[3] S. Hennin, G. Germana, and L. Garcia, "Integrated perimeter security system," in *2007 IEEE Conference on Technologies for Homeland Security*, 2007, pp. 70–75.

[4] R. Gade and T. B. Moeslund, "Thermal cameras and applications: a survey," *Mach. Vis. Appl.*, vol. 25, no. 1, pp. 245–262, Jan. 2014, doi: 10.1007/s00138-013-0570-5.

[5] N. Bakalos *et al.*, "Protecting Water Infrastructure From Cyber and Physical Threats: Using Multimodal Data Fusion and Adaptive Deep Learning to Monitor Critical Systems," *IEEE Signal Process. Mag.*, vol. 36, no. 2, pp. 36–48, Mar. 2019, doi: 10.1109/MSP.2018.2885359.

[6] S. J. Krotosky, S. Y. Cheng, and M. M. Trivedi, "Face detection and head tracking using stereo and thermal infrared cameras for" smart" airbags: a comparative analysis," in *Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No. 04TH8749)*, 2004, pp. 17–22.

[7] J. Mekyska, V. Espinosa-Duró, and M. Faundez-Zanuy, "Face segmentation: A comparison between visible and thermal images," in *44th annual 2010 ieee international carnahan conference on security technology*, 2010, pp. 185–189.

[8] C. Lalos, A. Voulodimos, A. Doulamis, and T. Varvarigou, "Efficient tracking using a robust motion estimation technique," *Multimed. Tools Appl.*, vol. 69, no. 2, pp. 277–292, Mar. 2014, doi: 10.1007/s11042-012-0994-3.

[9] I. Rallis, I. Georgoulas, N. Doulamis, A. Voulodimos, and P. Terzopoulos, "Extraction of key postures from 3D human motion data for choreography summarization," in *2017 9th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, Sep. 2017, pp. 94–101, doi: 10.1109/VS-GAMES.2017.8056576.

[10] I. Rallis, N. Doulamis, A. Doulamis, A. Voulodimos, and V. Vescoukis, "Spatio-temporal summarization of dance choreographies," *Comput. Graph.*, vol. 73, pp. 88–101, Jun. 2018, doi: 10.1016/j.cag.2018.04.003.

[11] R. O'Malley, E. Jones, and M. Glavin, "Detection of pedestrians in far-infrared automotive night vision using region-growing and clothing distortion compensation," *Infrared Phys. Technol.*, vol. 53, no. 6, pp. 439–449, Nov. 2010, doi: 10.1016/j.infrared.2010.09.006.

[12] Li Zhang, B. Wu, and R. Nevatia, "Pedestrian Detection in Infrared Images based on Local Shape Features," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2007, pp. 1–8, doi: 10.1109/CVPR.2007.383452.

[13] V. John, S. Mita, Z. Liu, and B. Qi, "Pedestrian detection in thermal images using adaptive fuzzy C-means clustering and convolutional neural networks," in *2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, May 2015, pp. 246–249, doi: 10.1109/MVA.2015.7153177.

[14] J. W. Davis and V. Sharma, "Robust detection of people in thermal imagery," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, 2004, vol. 4, pp. 713–716.

[15] T. T. Zin, H. Takahashi, and H. Hama, "Robust person detection using far infrared camera for image fusion," in *Second International Conference on Innovative Computing, Informatio and Control (ICICIC 2007)*, 2007, pp. 310–310.

[16] J. W. Davis and M. A. Keck, "A two-stage template approach to person detection in thermal imagery," in *2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05)-Volume 1*, 2005, vol. 1, pp. 364–369.

[17] H. Nanda and L. Davis, "Probabilistic template based pedestrian detection in infrared videos," in *Intelligent Vehicle Symposium, 2002. IEEE*, 2002, vol. 1, pp. 15–20.

[18] F. Xu, X. Liu, and K. Fujimura, "Pedestrian detection and tracking with night vision," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 1, pp. 63–71, 2005.

[19] A. Jo, G.-J. Jang, Y. Seo, and J.-S. Park, "Performance Improvement of Human Detection Using Thermal Imaging Cameras Based on Mahalanobis Distance and Edge Orientation Histogram," in *Information Technology Convergence*, Dordrecht, 2013, pp. 817–825, doi: 10.1007/978-94-007-6996-0_85.

[20] A. S. Voulodimos, D. I. Kosmopoulos, N. D. Doulamis, and T. A. Varvarigou, "A top-down event-driven approach for concurrent activity recognition," *Multimed. Tools Appl.*, vol. 69, no. 2, pp. 293–311, Mar. 2014, doi: 10.1007/s11042-012-0993-4.

[21] N. D. Doulamis, A. S. Voulodimos, D. I. Kosmopoulos, and T. A. Varvarigou, "Enhanced Human Behavior Recognition Using HMM and Evaluative Rectification," in *Proceedings of the First ACM International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams*, New York, NY, USA, 2010, pp. 39–44, doi: 10.1145/1877868.1877880.

[22] K. Makantasis, E. Protopapadakis, A. Doulamis, N. Doulamis, and N. Matsatsinis, "3D measures exploitation for a monocular semi-supervised fall detection system," *Multimed. Tools Appl.*, vol. 75, no. 22, pp. 15017–15049, Nov. 2016, doi: 10.1007/s11042-015-2513-9.

[23] C. Rougier, J. Meunier, A. St-Arnaud, and J. Rousseau, "Robust video surveillance for fall detection based on human shape deformation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 5, pp. 611–622, 2011.

[24] M. Yu, A. Rhuma, S. M. Naqvi, L. Wang, and J. Chambers, "A posture recognition-based fall detection system for monitoring an elderly person in a smart home environment," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 6, pp. 1274–1286, 2012.

[25] E. Protopapadakis, A. Voulodimos, A. Doulamis, N. Doulamis, D. Dres, and M. Bimpas, "Stacked autoencoders for outlier detection in over-the-horizon radar signals," *Comput. Intell. Neurosci.*, vol. 2017, 2017.

[26] N. Papadakis, A. Litke, A. Doulamis, E. Protopapadakis, and N. Doulamis, "Multimedia Analysis on User-Generated Content for Safety-Oriented Applications," in *Social Media Strategy in Policing: From Cultural Intelligence to Community Policing*, B. Akhgar, P. S. Bayerl, and G. Leventakis, Eds. Cham: Springer International Publishing, 2019, pp. 161–175.

[27] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "DBSCAN revisited, revisited: why and how you should (still) use DBSCAN," *ACM Trans. Database Syst. TODS*, vol. 42, no. 3, pp. 1–21, 2017.

[28] T.-R. Liu, V. Copin, and T. Stathaki, "Human Detection from Ground Truth Cameras through Combined Use of Histogram of Oriented Gradients and Body Part Models.," in *VISIGRAPP (4: VISAPP)*, 2016, pp. 735–740.

[29] T. Liu and T. Stathaki, "Enhanced pedestrian detection using deep learning based semantic image segmentation," in *2017 22nd International Conference on Digital Signal Processing (DSP)*, Aug. 2017, pp. 1–5, doi: 10.1109/ICDSP.2017.8096045.

[30] L. Cuimei, Q. Zhiliang, J. Nan, and W. Jianhua, "Human face detection algorithm via Haar cascade classifier combined with three additional classifiers," in *2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*, 2017, pp. 483–487.

[31] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *ArXiv150602640 Cs*, May 2016, Accessed: Mar. 23, 2020. [Online]. Available: http://arxiv.org/abs/1506.02640.

[32] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.

[33] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *ArXiv Prepr. ArXiv180402767*, 2018.