JDST, vol. 5, no. 2, p29-p43, 2022
https://doi.org/10.46713/jdst.05.002

www.jdst.eu, ISSN 2534-9805, e-ISSN 2534-9813

Research Article

Tracking of objects in a multi-sensor fusion system for border surveillance

Luis Patino [a], Michael Hubner [b], Martin Litzenberger [b] and James Ferryman [a]

[a] *University of Reading, Department of Computer Science, Polly Vacher Building, Reading RG6 6DH, United Kingdom*
https://www.reading.ac.uk/computer-science/

[b] *AIT, Giefinggasse 4, 1210 Vienna, Austria*
https://www.ait.ac.at/themen/new-sensor-technologies/situational-awareness-decision-support

ABSTRACT

In this work, we present a Fusion and Tracking system developed within the EU project FOLDOUT aimed to facilitate border guards work by fusing separate sensor information and presenting automatic tracking of objects detected in the surveillance area. The focus of FOLDOUT is on through-foliage detection in the inner and outermost regions of the EU. Fusing several sensor signals increases the effectiveness of detection, particularly in forested and other areas hidden by foliage. We use weighted maps (also called Heatmaps) to combine multi-sensor information; tracking is performed on the resulting fused objects; a track is created or updated based on cost calculation of associating fused detections temporally. We compare tracking results from individual sensors and from fused objects from data collected in a simulated border that is representative of actual EU borders in Bulgaria. The results show how tracking is enhanced if performed on fused data rather than from individual sensor information.

ARTICLE INFO

RECEIVED: 09 Oct 2021
REVISED: 10 Nov 2021
ACCEPTED: 30 Nov 2021
ONLINE: 12 DEC 2021

KEYWORDS

Border surveillance, Detection and Tracking, Illegal border crossing, Person detection, Sensor Fusion.

Corresponding Author: Luis Patino; Tel.: +44 118 378 7641; E-mail: j.l.patinovilchis@reading.ac.uk

I. OVERVIEW

Border surveillance has been a topic with increasing interest in the last few years [1-3], particularly given the fact that in the last years irregular migration has dramatically increased [4]. For border guards, the main concern remains to detect an illegal border crossing as soon as possible. Camera-based systems remain the most widespread kind of surveillance system because video streams allow operators to perform a direct situation assessment at the border. To increase the security of the border, multiple types of camera sensors may be deployed including, for instance, RGB, Infrared, Thermal or multispectral cameras. Beyond this kind of sensors, even acoustic, seismic or PIR (movement) sensors may be installed along the border. The downside for border guards is the overwhelming amount of data that may be generated from all different sensors. The task is even more complex if the operator must deal with false alarms, which are common on surveillance systems for natural spaces where weather, trees, and animal movements, among others, can cause false positives.

An obvious approach to reduce the false alarm rate is the data association of automatically validated detections by different sensor modalities that monitor the same perimeter. Furthermore, to facilitate border guards work automatic tracking of detected people can be performed.

In this work, we present a Fusion and Tracking system developed within the EU project FOLDOUT aimed to facilitate border guards work by fusing separate sensor information and presenting automatic tracking of objects detected in the surveillance area. FOLDOUT focus is on through foliage detection in the inner and outermost regions of the EU. Fusing several sensor signals increases the effectiveness of detection, particularly in forested and areas hidden by foliage where particularly traditional camera-based systems face the problem of severe occlusion given by the foliated area and must then potentially deal with fragmented detections.

In this work we use weighted maps to provide layers (also called *HeatMaps*) of sensor data and combine them in a logical and mathematically correct forumation. Our Fusion approach is derived from well-known probabilistic fusion techniques based on probabilistic occupancy grids [5,6] which are commonly used in mobile robot perception [7]. Additionally, a Linear Opinion Pool (LOP) [8] is used to fuse the different layers of sensor data. To solve for automatic tracking, we have developed a complete tracking approach based on cost calculation of associating fused detections temporally.

We analyse in this work tracking results from individual sensors and from fused objects from data collected in a simulated border that is representative of actual EU borders in Bulgaria. The proposed approach and deployment were discussed with actual EU border guards to guarantee the usefulness and validation of the system. We simulated realistic illegal border crossings in an area. The surveillance system deployed includes two camera sensors and PIR sensors to monitor the simulated border.

We observe how tracking results change from processing single sensor information to processing fused data with different sensor combinations. The results show how tracking results are enhanced when different sensor information is fused.

II. MATERIALS AND DATA

The work in this paper derives from direct collaboration with EU border guards. In the following paragraphs we detail how actual EU borders are characterised and how we simulated an actual border and the data we have collected for analysis.

A. Current Situation at the borders

The situations as-is in current border surveillance installation is characterized by the following:

- A cleaned strip (typically 10m wide) along the border line that is kept clear of vegetation, often a small road (as seen, for instance, in Bulgaria, Finland, Greece).
- A line of passive motion detection sensors along the border line typically in a distance 10-20m from each other with automatic detection (these can be PIR sensors, acoustic sensors, seismic sensors, or a combination of all of them).
- A line of fixed field-of-view cameras visual and/or thermal with view direction along the border line covering an area of about 5-20 of the above-mentioned motion detection sensors.
- A steerable (pan-tilt-zoom) high resolution RGB camera and/or thermal camera on a high mast or tower at some distance from the border. the camera can be steered by an operator (or sometimes is automatically steered by the triggered motion detection sensors). That camera covers a wide area, typically 1 – several kilometres, of border line.

B. Simulation of the border installations

This study has been carried out in a simulated border implemented in a realistic way. The deployment took place in a forested area in Bulgaria simulating a border line of approx. 100m length:

- A cleaned strip (typ. 10m wide) along border: the road
- A line of passive motion detection sensors (PIR) along the border line.
- A line of fixed view cameras with view direction along the border: Implemented with a FLIR thermal camera
- A steerable (pan-tilt-zoom) high resolution RGB camera: Implemented with a PTZ camera mounted on a tower.

In Figure 1, it is depicted the simulated border with the 100m length cleaned strip (the road) separating two areas of vegetation (corresponding to two different countries in the simulation).

C. Deployed sensors

RGB camera

In this study we employed the following camera, DH-SD6AL830V-HNI 4K 30x Laser PTZ Network Camera featuring powerful optical zoom (30x optical zoom) and accurate pan/tilt/zoom performance. Together with infrared illumination, the camera represents a good solution for dark, lowlight applications. The series combines a day/night mechanical IR cut filter for the highest image quality in variable lighting conditions during the day.

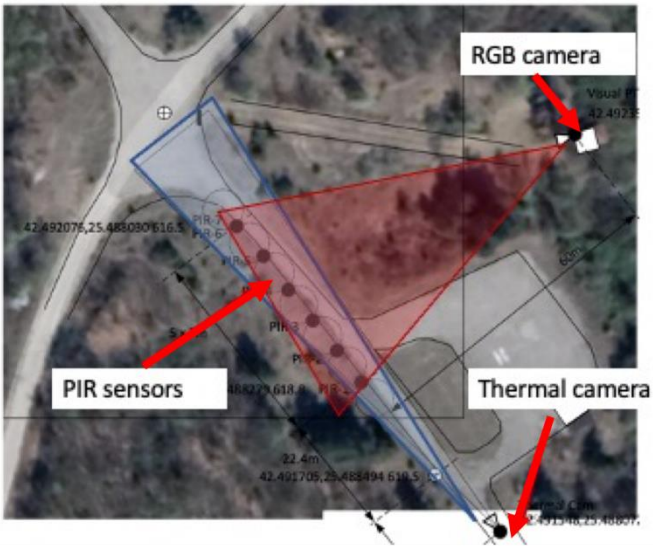

Figure 1: Simulated border for this study. A 100m length cleaned strip (the road) separates two hypothetical countries. The border is surveilled with three different sensor modalities: PIR sensors; RGB and Thermal cameras.

The RGB camera Features are:

- 1/1.7" 12Megapixel STARVIS™ CMOS
- Powerful 30x optical zoom
- H.265 Encoding
- Max. 25/30fps@4K
- Auto tracking and IVS
- Support Hi-PoE
- IR distance up to 500m
- IP67

Thermal camera

The thermal imaging camera deployed is the FLIR camera FLIR F-606E. FLIR F-Series cameras are high-resolution thermal security cameras that provide video and control over both IP and analogue networks. Because thermal cameras detect heat, they can reveal persons in all lighting conditions, including complete darkness, rain, light fog, and smoke. Therefore, they are a good complement to visual cameras for border surveillance.

PIR sensors

We developed custom sensors for the purpose of this study. The device is composed of a Passive Infrared sensor, a microcontroller and a wireless communication module. The sensor is powered by internal batteries and connected wirelessly via 2.4 GHz XBee module.

D. Collected Data

We carried out a series of recordings in November 2019 in a forested area in Bulgaria. Actors were instructed to perform realistic activities at the border (set out as described in Section II.B). These include: i) irregular border crossings (illegal person + vehicles); ii) illegal transport and entry of goods (trafficking); iii) Detection of persons & vehicles in a search & rescue operation in forest terrain. For each of these scenarios, and different variations of them, scripts were created giving specific instructions to the actors on how they should move to mimic realistic situations. Recordings were performed over two days including day and night conditions. Fifteen script sequences were recorded with all sensors operational. In Figure 2 it is presented, as an example, the movement instructed to actors for one of the illegal border crossing scenarios played during the data collection.

III. Detection & Tracking methodology

Border guards' main interest is the localisation in a global map of detected people on the surveilled area as well as its tracking. To achieve this, detections from a single person, observed on separate sensor systems, are to be fused first. When detections have been correlated and made consistent, the tracking of separate targets on a common map can be performed.

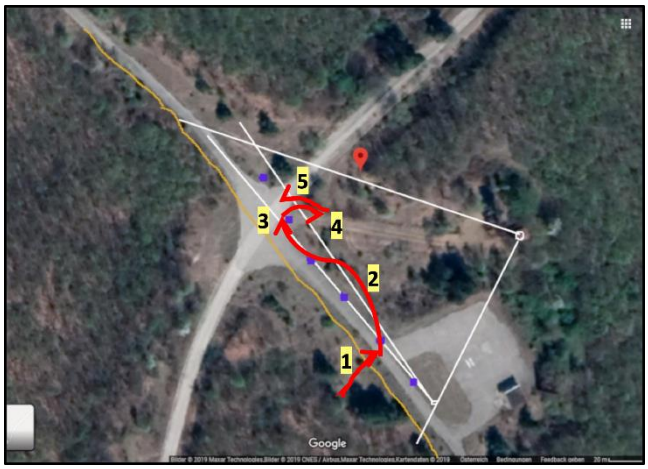

Figure 2: Instructed movement (red lines) to actors to play an illegal crossing scenario: 1. one person has crossed the border by walking. 2. The person walks along the border path towards the main road. 3. The person stops and stays long time on the road (potentially waiting for a smuggler in a car). 4. At some point leaves the road to go hiding into the foliage. 5. Being amongst the foliage, the person comes back to the road again (probably looking again for the car)

A. Single sensor detection

Person detection in RGB and Thermal cameras

A comprehensive Deep Learning-based object detection is applied on the camera image. Deep Learning methods have been shown to outperform previous state-of-the-art machine learning techniques. Deep neural networks (DNNs) mimic how the brain perceives and processes information. In contrast to previous approaches, DNNs learn the features required for tasks such as person detection. In recent years, DNNs have shown outstanding performance on object detection and classification tasks [9, 10]. For this work, the object detection is based on a well-known DNN implementation, the YOLO detector [11].

Person detection in PIR sensors

The detector is tuned so that a Passive Infrared sensor will trigger the presence of a person within a radius of 7.5 metres around the PIR.

B. Fusion of heterogenous sensors

In this work we use weighted maps to provide layers (also called HeatMaps) of sensor data and combine them in a logical and mathematical way. Its dynamics is completely event driven using the sensor detection hypothesis of different sensor modalities. These sensor hypotheses comprise the location (WGS84 datum), a timestamp (Unix Timestamp) and a weight (e.g. confidence taken from a sensor detection). To achieve this, two components are essential: Weighted Maps (HeatMaps); Linear Opinion Pool. Figure 3 shows the basic concept of this approach.

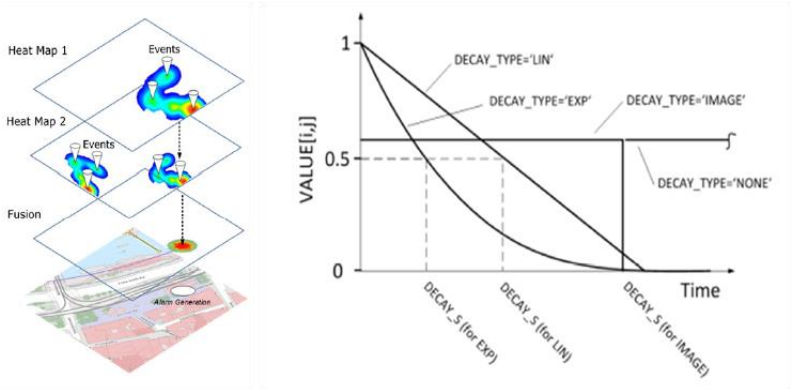

Figure 3: Basic concept of Fusion approach (left) as an example using two Weighted Maps (Heat Maps). Different decay functions (right) applied to establish time dynamic behaviour of the Weighted Maps.

Weighted Map (HeatMap)

The weighted map is the first of the two essential components of our data fusion approach. The basic idea of the weighted map is, to hold and update spatial-temporal information regarding different sensor detection hypothesis. The weighted map is derived from a probability occupancy grid but interprets incoming data in form of weights. Also, a decay in time is employed to model the timely behaviour of the sensor data. The weights are stored in an array of selectable resolutions, representing a rectangular area of interest in WGS84 coordinates. Figure 3, demonstrates possible decay functions for modelling the dynamic behaviour of the Weighted Maps.

Usually, a Weighted Map corresponds to any kind of sensor data or sensor modality (e.g., bounding boxes of person detection from camera images) in space-time. The

sensor data are ingested into a dedicated Weighted Map, which leads to an increase (replacement) of the Weighted Map's values according to the weight of the incoming senor hypothesis. Respectively, the decay will be applied in time to the Weighted Map's value matrix. Every time a sensor hypothesis is ingested into the map, it will be updated by recalculating the weights of the Weighted Map and decaying the values of previous states.
Finally, a Linear Opinion Pool allows us to combine multiple Weighted Maps and therefore multi sensor modalities with the goal to decrease the overall False Discovery Rate of a sensor system.

Linear Opinion Pool (LOP)

The second essential component of our fusion approach is the Linear Opinion Pool [8]. We use it to combine multiple Weighted Maps according to the following formula.

$$F_{j,k}(t) = \alpha \sum_{i \in I} \omega_i \cdot m^i_{j,k}(t), \quad \alpha = [\sum_{i \in I} \omega_i]^{-1} \qquad (3.1)$$

Where $F_{j,k}$ denotes the Fused Map at each cell (i,j) and ω_i denotes the contribution of the individual weighted maps $m^i_{j,k}(t)$. In our work we chose the special case of the LOP where $\alpha = 1$.

The LOP is applied every time a state of a Weighted Map is updated due to new sensor detection hypothesis. After the LOP has been evaluated, thresholding allows us to generate an alert. To determine the position of the alarm, a segmentation algorithm (blob detection) is used on threshold exceeded areas of the combined value matrix.
These alerts result from multiple sensor hypothesis and are used to provide the necessary input data for tracking, which will be described in the next section.

C. Multi-target tracking

In order to follow the movement of an intruder crossing a border into a forbidden or sensitive area, we have developed a custom algorithm based on cost calculation of associating object detections spatially and temporally.

The tracking system works by building a model of the object exclusively based on its position and time stamp.

At the first object detection, the model is initialised with the position and timestamp of that detection. A track model is defined with the following tuple:

$$T_i = (x_i, y_i, t_i) \qquad (3.2)$$

where *x, y* and *t* correspond respectively to the latitude, longitude and timestamp of the point.

If several object detections occur at the same time, there are as many model templates created as there are detections simultaneously received. Subsequent detections are added to a given Track model depending on the cost involved on appending the detection to the track. The cost is defined as the distance between the incoming detection and the Track candidate.

Let $d_s(T_i, o)$ be the spatial distance between the most recent point in the track T_i and the incoming detection *o*. The spatial distance is calculated as the Euclidean distance between the latitude and longitude of the two points.

$$d_s(T_i, o) = \sqrt{(x_i^T - x^o)^2 + (y_i^T - y^o)^2} \qquad (3.3)$$

Let $d_t(T_i, o)$ be the temporal distance between the most recent point in the track T_i and the incoming detection *o* given by the subtraction of the two point timestamps.

$$d_t(T_i, o) = \left| t_i^T - t^o \right| \qquad (3.4)$$

The cost of appending object *o* to track T_i will be then calculated as:

$$C = 2 - e^{-(d_s * \tau_s)^2} - e^{-(d_t * \tau_t)^2} \qquad (3.5)$$

Where τ_s and τ_t are respectively spatial and temporal similarity parameters tuned empirically for our current implementation.

The cost *C* is bounded between *0* and *2*; with *C=0* for a perfect match when appending object *o* to track T_i. The object is appended to the track if the cost is less or equal than a given threshold set to 0.5 in our implementation; otherwise, that object will initialise another Track.

In case of multiple incoming detections and multiple Track candidates, a Hungarian algorithm [12] has been implemented so that the associations between detections and Tracks incurs the minimum cost.

IV. Evaluation Methodology

We have focused the evaluation in a zone of interest (ZoI) around the road defining the border in the simulated scenario. The aim is to obtain measures for detection and tracking informing the performance of the system regarding detecting illegal crossings through the border. The ZoI analysed is depicted in Figure 4.

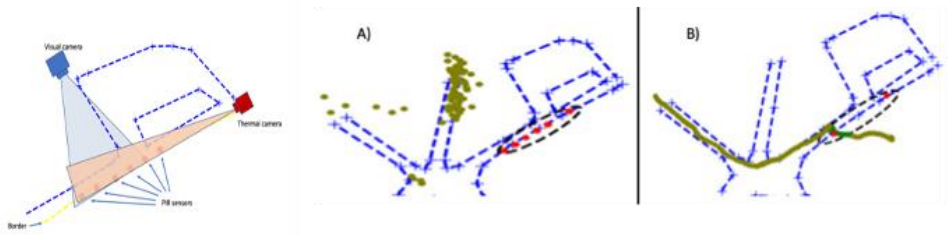

Figure 4: Left panel: sensors deployed in the simulated border; the blue dashed lines delineate the road defining the border and an open area by the road. Middle panel: ZoI depicted with a black dashed line; for reference, the PIR sensors are also plotted with red dots. Right panel: one person crossing the ZoI. PIR detections are shown in red; GT data not included in the analysis in light green; GT data included in the analysis in dark green.

Actors were instructed to move in the area and cross the simulated border with a variety of different movements from quick moves (running) to slow and calm movements. Actors were carrying a phone and their position was collected via the 'GPS logger' phone application. The Ground-Truth (GT) for evaluation corresponds to GPS data collected through actors' phones.

To solve for different sampling frequencies all data was analysed in temporal windows of 1 sec length. Detection data and GT data were compared inside these temporal windows and within the ZoI. In this frame, typical Receiver Operator Characteristic performance measures of True Positives, False Positives, True Negatives, False Negatives defined as follows:

- True Positive: a system detection and a GT object exist inside the ZoI.
- False Positive: a system detection exists inside the ZoI but no GT object is found.
- True Negative: no system detection exists inside the ZoI and no GT object is found.
- False Negative: no system detection exists inside the ZoI, however a GT object is found.

To evaluate detection and tracking, we have then calculated the well-known MOT measures [13]: Multiple Object Tracking Precision (MOTP) and Multiple Object Tracking Accuracy (MOTA)

Multiple Object Tracking Precision

The Multiple Object Tracking Precision is the average dissimilarity between all true positives and their corresponding ground truth targets. MOTP is computed as:

$$MOTP = \frac{\sum_{i,t} d_t^i}{\sum_t c_t} \qquad (4.1)$$

Where c_t denotes the number of matches in frame (or at time) t and d_t is the distance between the reported detection i and the matched object Ground Truth.

MOTP is one of the CLEAR MOT metrics serving to measure localization precision. For evaluation of tracking performance itself, we employ the multiple object tracking accuracy *MOTA*.

Multiple Object Tracking Accuracy

MOTA has been defined as derived from 3 error ratios [13]: The ratio of Misses per Ground Truth objects (calculated from False Negative detections); the ratio of false positives per Ground Truth objects and the ratio of mismatches (or ID switches) per Ground Truth objects.

Overall, *MOTA* is calculated with the following formulae:

$$MOTA = 1 - \frac{\sum_t (FN_t + FP_t + IDswitch_t)}{\sum_t g_t} \qquad (4.2)$$

where g_t are the number of Ground Truth objects at time (or frame) t. FN_t, FP_t and $IDswitch_t$ are respectively, the number of False Negatives, False Positives and ID switches at time (or frame) t.

IV. Results and Discussion

The results from evaluating the collected data (described in Section II) are shown in Table 1. The evaluation corresponds to seven different tracking results (taking into account different input data). Three tracking results are from taking as input single sensor detections and four tracking results are from taking as input the fused detections from all different combinations possible with the single sensors. Generally, it can be said that tracking performs better having fused data as input to the tracker rather than detection data from single sensors. In most cases the MOTA values with fused data are higher than with data from single sensors (remark that in MOTA measures, the sign counts so that larger negative values indicate actually worst performance).

Table 1. Tracking evaluation results when single sensor detections or detections fused from different sensor combinations are employed. Best results are highlighted in green; particularly low (bad) values are highlighted in red and yellow indicate some performance values differentiating between the two best tracking results.

sensor	sequences evaluated	Precision	Recall_or_TPR	Specificity_or_TNR	MOTP (in m)	misses_per_objs	FP_per_objs	switches_avrg	MOTA
PIR	15	0.23	0.13	0.97	7.10	0.87	0.32	1.80	-0.23
RGB	15	0.11	0.48	0.68	5.91	0.52	8.09	10.40	-7.88
Thermal	15	0.21	0.70	0.56	17.00	0.30	14.30	0.13	-13.61
Fusion-RGB-PIR	15	0.18	0.06	0.99	7.18	0.94	0.16	0.93	-0.12
Fusion-Thermal-PIR	15	0.28	0.37	0.87	6.82	0.63	3.04	0.27	-2.67
Fusion-Thermal-RGB	15	0.37	0.37	0.92	7.23	0.63	1.39	0.47	-1.03
Fusion-Thermal-RGB-PIR	15	0.36	0.27	0.95	7.36	0.73	1.04	0.20	-0.77

In principle, it could be thought that the best tracking performance is achieved with fusion of RGB-PIR data as this combination provides the best MOTA value; however, the recall is particularly low. This is highlighted in red for the corresponding cell in Table 1. A general rule for trackers is that they tend to perform better in less crowded environments so that in this case the lower recall on one side helps to facilitate tracking but for real operational purposes a high recall is preferred as this ensures the correct detection of the object in the first place. It has to be noted that actually fusing of RGB and PIR data helps to increase the detection precision value of RGB, which is the lowest from all different sensor input combinations (see Table 1).

The PIR sensor appears to have the second best MOTA value. Again, the recall is very low, and it is the second lowest recall value that can be observed in Table 1. Tracking is done with a better performance, but it is much less informative, particularly holding the fact that the tracking positions are limited to the point fixed positions of PIR sensors along the "border" line (as shown in Figure 1).

The best combinations for better tracking results, are fusion from Thermal-RGB or Thermal-RGB-PIR data. In both cases, the Precision and Recall values are the best as observed in Table 1. With higher detection recall the problem of ID switches is more susceptible to appear as seen from the data. In our case, the tracker may exchange the ID among detected objects close from each other. Maintaining the correct ID is a topic of our future work where we may include more tracking features to achieve this.

V. Conclusions

In this paper we have addressed the problem of tracking persons in forested border areas where object detection can be particularly challenging given the difficulties of through foliage detection.

We analyse in this work tracking results from individual sensors (PIR, thermal and RGB cameras) and from fused objects from data collected in a simulated border that is representative of actual EU borders in Bulgaria. We observe how tracking results change from processing single sensor information to processing fused data with different sensor combinations. The results show that tracking results are enhanced when different sensor information is fused.

Acknowledgements

This research was funded by EU H2020 Research and Innovation Programme under grant agreement no. 787021 (FOLDOUT).

References

[1] A. Beduschi, “International migration management in the age of artificial intelligence,” *Migration Studies,*, 2020. https://doi.org/10.1093/migration/mnaa003.

[2] S. Ghaffary, “The “smarter” wall: How drones, sensors, and AI are patrolling the border,” *Recode*, 2020. https://www.vox.com/recode/2019/5/16/18511583/smart-border-wall-drones-sensors-ai.

[3] Kowalski, M.Ł.; Pałka, N.; Młyńczak, J.; Karol, M.; Czerwińska, E.; Życzkowski, M.; Ciurapiński, W.; Zawadzki, Z.; Brawata, S. “Detection of Inflatable Boats and People in Thermal Infrared with Deep Learning Methods” *Sensors* 2021, 21, 5330. https://doi.org/10.3390/s21165330

[4] International Organisation of Migration (IOM), “IOM-UN Migration.” https://migration.iom.int/europe?type=arrivals (accessed Feb. 05, 2021).

[5] O. Erdinc, P. Willett, and Y. Bar-Shalom, “The Bin-Occupancy Filterand Its Connection to the PHD Filters,” IEEE Transactions on Signal Processing, vol. 57, no. 11, pp. 4232–4246, Nov. 2009.

[6] J. Tong, L. Chen, and Y. Cao, "Human positioning based on probabilistic occupancy map," in 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). Huangshan, China: IEEE, Jul. 2018, pp. 271–275.

[7] A. Elfes, "Using occupancy grids for mobile robot perception and navigation,"Computer, vol. 22, no. 6, pp. 46–57, Jun. 1989.

[8] J. D. Adarve, M. Perrollaz, A. Makris, and C. Laugier, "Computing occupancy grids from multiple sensors using linear opinion pools," in 2012 IEEE International Conference on Robotics and Automation.

[9] M. Anusha, and P. Kiruthika, "Comparative Study on Augmented Analytics Using Deep Learning Techniques," Smart Innovation, Systems and Technologies, 243, pp. 135-142, 2022.

[10] L. Liu, W. Ouyang, X. Wang et al., "Deep Learning for Generic Object Detection: A Survey," Int J Comput Vis 128, 261–318, 2020. https://doi.org/10.1007/s11263-019-01247-4.

[11] Ultralytics, "Yolov5 in pytorch," Sep. 2021. [Online]. https://github.com/ultralytics/yolov5

[12] A. Yilmaz, O. Javed and M. Shah, "Object Tracking: A Survey," ACM Computing Surveys, Vol. 38, No. 13, 2006.

[13] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: The clear mot metrics," EURASIP J. Image Video Process., vol. 2008, 2008. http://dblp.uni-trier.de/db/journals/ejivp/ejivp2008.html #BernardinS08