VEHICLE THREAT LOCATING VIA THE DETECTION OF ANOMALIES ON ROADS AND THEIR VERGES

Roel Heremans
AGT
R&D Group GmbH
Hilpertstrasse 35
64295 Darmstadt, Germany
email: RHeremans@agtinternational.com

Wim Mees
Royal Military Academy of Belgium
Communication, Information, Systems and Sensors (CISS)
Renaissancelaan 30
1000 Brussels, Belgium
email: wim.mees@rma.ac.be

ABSTRACT
Situational awareness originating from advanced sensor systems augments the ability of sound decision making. This paper considers the image analysis problem given a multi-sensor system mounted on an army patrol vehicle that serves as an early threat alert mechanism. Hence, it concerns forward looking sensors that move through a complex environment whilst detecting events or points of interest. The objective comprises identifying potential threats to the vehicle such as Improvised Explosive Devices (IED). This paper focuses on how an optical sensor and a forward-looking infrared (FLIR) are exploited for detecting and tracking stationary anomalies on the road’s surface and verges. The applied approach consists of a road scene analysis that extracts the road’s surface and verges followed by anomaly detection in each of the extracted regions separately. The detected anomalies from both sensors are tracked and combined in geographical coordinates where their threat-levels also increase due to the response of other sensors. The proposed methodology was assessed during a full system’s demonstration. The obtained results within a simplified real situation show considerable potential.

KEY WORDS
Scene segmentation, Road extraction, Anomaly detection in an Urban Environments using Mobile Sensors (SUM) project, a system is constructed with as primary objective providing the operator with decision support information concerning potentially dangerous objects that lurk in and on the road’s surface and verges. Well-known threats to armed forces on patrol are the IED. The one characteristic that most of these IEDs share is their explosive nature. The literature reports on several sensing technologies for detecting explosive materials, e.g. ground penetrating radar, and infra-red sensors. The SUM framework adopts a dedicated sensor array that consists of four dissimilar forward looking sensors. The idea is raising detection accuracy whilst lowering the false alarms by extracting evidence of potential threats from each sensor individual and fusing the results. Furthermore, using diverse sensors quells the problematic detectability under different lighting conditions.

The focus of this paper is on locating potentially dangerous objects using the output of the optical sensor and the FLIR separately. IEDs or other hazards do not really exhibit unique characteristics. Thus the detection of IEDs from optical data implies anomalies. These are of particular interest since they often correspond to significant information. Detecting outliers remains an important topic in numerous application domains. Chandola et al. [3] and Hodge et al. [8] present extensive surveys on this topic. For anomaly detection applied in the domain of forward looking sensors mounted on a moving platform, the literature is less rich. In [1], [5] and [12], a set of anomaly detectors aim at detecting explosive hazards using a FLIR. The detection of temporal changes is a related but dissimilar topic. It includes the detection of abandoned objects on road’s verges [10]. In the context of the SUM project, the anomaly detection follows an unsupervised approach. It consists of two consecutive steps. A first step segments the road scene in regions that are expected to exhibit similar characteristics. The second step identifies anomalous areas within these regions. Once anomalous areas are identified, their geographical location is approximated. The anomaly is considered as threat evidence and clusters of threat evidence are likely to correspond to physical objects.

1 Introduction
Situational awareness refers to the perception of environmental elements with respect to time and space, the comprehension of their meaning, and the projection of their status. It is concerned with perception of the environment critical to decision-makers in complex, dynamic areas. Situational awareness is an important facet of most application domains, e.g. surveillance, autonomous vehicles, advanced driver’s assistance systems and so forth. Consequently, research communities continuously endeavor augmenting situational awareness. The use of multi-sensor systems presents the opportunity of combining the advantages of different sensors whilst minimizing their drawbacks. Within the framework of the Surveillance

Proceedings of the IASTED International Conference
Computer Graphics and Imaging (CGIM 2013)
February 12 - 14, 2013 Innsbruck, Austria

DOI: 10.2316/P.2013.797-030
The paper is structured as follows: The first section concisely explains the complete threat detection system in its present shape. It focuses on explaining the image analysis task at hand and sketches the sensor registration. The following section introduces the road-scene analysis and details the extraction of the distinct areas. The final section tackles the detection of anomalous objects that consist of anomalous pixels detection, clustering, object extraction and filtering via approximated physical appearances. Each section discusses and illustrates the achieved outcomes.

2 Threat detection system

The goal of the SUM system is to detect potential threats to a military patrol vehicle, both during the day and at night. The threats are most often IEDs, hidden from view in some way. For this reason a purely shape based detection in an optical or infrared image will not produce acceptable results. Indeed, any disturbance on the road surface or sufficiently large object on the side of the road might hide an IED. However, not all disturbances or road-side objects can be reported as possible threats since this would clutter the user interface with a huge amount of false alerts.

Therefore, the system is built around a combination of four sensors: an optical, a FLIR, a radiometer and a radar. It combines the sensor data with auxiliary data from outside sources in a central fusion engine. The optical and infrared sensors will detect anomalies on the road surface or on the side of the road, and the radiometer and radar will allow us to distinguish for instance an innocent garbage bag, cardboard box or pile of dirt from one that is hiding an unexploded ordnance or homemade explosive with metal shrapnel used to build an IED. The individual sensors acquire data in an asynchronous manner and at different sampling rates. In order to make the system robust against temporary sensor failures or occlusions due to the movement of the vehicle and the relative positions of the objects in the scene, there are no predefined decision scenarios for detecting different types of threats and there is no tracking. Every image that is made available to the fusion engine by the sensors is processed individually. The resulting candidate threats or targets are first evaluated based on their features and then spatially fused based on their absolute position. The spatial fusion requires an accurate registration of the sensors, in order to map image coordinates to local coordinates relative to the vehicle, as well as a precise tracking of the vehicle’s position and orientation in order to map the local coordinates to absolute coordinates, being geographic coordinates in the case of the SUM system. This also allows for an easy integration of auxiliary data received from external sources. Figure 1 outlines the detection process.

[Figure 1. Schematic view of the proposed multi-sensor anomalous object detection]

3 Sensor registration

The processing of the individual sensors yields anomalies or responses in their respective image spaces. Their combination and tracking require the use of geographical coordinates, i.e. longitude and latitude. A first transformation converts the image coordinates to world coordinates that are relative to the vehicle [11]. The second step uses the outcome of the Global Positioning System (GPS) and translates these world coordinates to longitude and latitude.

A sensor registration procedure allows estimating the image to world transformation for the optical sensor, the FLIR and the radiometer [7]. Applying these transformations on the respective pixel coordinates yields the world coordinate system relative to the vehicle. In the latter, the Y-axis is parallel to the vehicle’s bearing. Assuming a flat surface model, perspective geometry allows the construction of an invertible transformation based on the camera’s parameters [4]. The external parameters can be estimated via a registration procedure. The latter entails the acquisition of one or more images with several ground truth locations, i.e. locations for which the position with relative to vehicle is known. For this purpose, a grid was drawn on the road surface. Figure 2 shows the outcome of the registration procedure. For each sensor, the approximated grid is shown in red. Further verification of the estimated transformations is achieved by comparing the position of features that are visible in both sensors. Figure 2 highlights 3 such visual control points. The GPS provides the vehicle’s location. When the vehicle moves, it also provides a bearing. Both are essential for locating the detected features in the geographical space, i.e. the real world. Considering that the system concerns forward looking sensors and assuming that the vehicle travels forward, the global geographical position of each detected feature can be calculated as follows: Let $R_f$, $\theta_f$ and $\theta_v$, represent the feature’s distance to the vehicle, the feature’s angle with respect to the vehicle and,
the vehicle’s bearing respectively. Remark that $\theta_f = 0$ is the vehicle’s bearing and, $\theta_f > 0$ and $\theta_f < 0$ represent locations to the right and to left. Given the vehicle’s latitude $\phi_v$ and longitude $\lambda_v$, the feature is globally located at $(\phi_f, \lambda_f)$:

$$\begin{align*}
\phi_f &= \arcsin \left[ \sin \phi_v \cos \tilde{R}_f + \cos \phi_v \sin \tilde{R}_f \cos \theta \right] \\
\lambda_f &= \lambda_v + \arctan \left[ \frac{\sin \theta \sin \tilde{R}_f \cos \phi_v}{\cos \tilde{R}_f - \sin \phi_v \sin \phi_f} \right]
\end{align*}$$

where $\tilde{R}_f$ is the feature distance $R_f$ normalized by the Earth’s radius. $\theta$ represents the vehicle’s bearing if it would be traveling in the direction of the detected anomaly. It is expressed in radians clockwise from the North. Remark that the geographical positions of the detected features possess a degree of uncertainty. This depends on the accuracy of the GPS, the resolution of the sensor and, the precision of the image to world transformation. Moreover, the further away the feature from the vehicle, the less accurate its estimated location.

## 4 Potential threat detection

Usually, IEDs or other hazards do not exhibit specific characteristics. The IEDs generic composition consists of a trigger, a detonation mechanism and the explosive element. However, there are numerous varieties of these components and there exists uncountable assembly methods. Hence, locating this type of potentially dangerous object in a road scene advocates finding anomalies in the scene. Anomaly detection is applied in many application domains. There exist several general approaches and, it remains a widely studied topic [3][8]. Any type of anomaly detection commences with defining at least a notion of normal and aberrant data behavior. The latter categorizes approaches in three classes: (i) supervised, (ii) semi-supervised and, (iii) unsupervised. The first approach assumes that both types of behavior are well described. For the second approach, usually the normal demeanor is known, whereas for the final method both behaviors are undetermined.

In this work, an unsupervised approach is adopted in which prior knowledge of the scene structure defines the notion of normal behavior in certain areas. Figure 3 outlines the anomaly extraction in case of the optical sensor and the FLIR. A road-scene analysis identifies areas for which similar spectral characteristics can be assumed. This is followed by an anomaly detection that strives to find connected pixel clusters that do not correspond with these characteristics. The retained clusters are attributed with sets of spectral and shape features. Subsequently, an appropriate location in the vehicle’s world coordinate system is estimated and the cluster’s projected physical characteristics are approximated.

### 4.1 Road-scene analysis

#### 4.1.1 Road detection

The core of the adopted road-scene analysis is a road detector. Road detection from a mobile platform is a very challenging task. The algorithms should be able to deal with a continuously changing background, the presence of different objects with unknown movement and, they also should be capable of identifying a high variability of road types. These may differ in shape, size and visibility conditions [1]. Note that, road extraction differs from the related
lane detection. The former aims at extracting the whole road, whilst the latter solely requires the lane the vehicle is driving on. The extraction of the road or road-tracking can be achieved through several approaches [1, 6, 10]. One can discriminate two main methodologies. One focuses on the road color whilst the other assumes a road model. Within the context of the multi-sensor approach, it is valuable that the road-scene segmentation can be ported from one sensor to another. The latter allows detecting the road surface through the available sensor in which it is most opportune and using this result for all sensors. Furthermore, a scale-invariant approach is preferred. Therefore, the adopted approach approximates the road surface via two lines that represent the separation between the road surface and its verges. There exist methods that approximate this separation with piecewise linear lines. This enables the extraction of curved roads. For now, it suffices extracting the so-called road triangle whilst assuming a flat road surface model. The extraction of this triangle is achieved using a rudimentary form of the technique introduced in [10]. The adopted approach assumes that the vehicle is located on the road and that it is somewhat parallel to the road’s boundaries. Hence, the left road edge corresponds to the dominant left oriented edge in the left part of the image, and vice versa for the right road edge. In this way the edge detection can be achieved in parallel and on a smaller and possibly less cluttered image area. Moreover, it decouples the detection process which allows estimating the algorithm’s parameters for each road side separately. The latter can prove useful when the verges differ substantially.

Figure 4 illustrates the adopted approach. First the image is split into a left and a right part. In each image part, the gradient magnitude and orientation (Figure 4.b-c) using the vector-valued gradient [13] are estimated. Hereafter the Canny edge map [2] is calculated that solely contains pixels that have a gradient orientation corresponding to the sought after road edge (Figure 4.f-g). Next the Hough transform extracts about 5 lines that correspond to the sought after road edge (Figure 4.h). To reduce the amount of lines, the result of a previous frame is consulted. Given the assumption that the road-scene evolves slowly; lines for which the orientation differs too much from the result of the previous frame are discarded. In the example given in Figure 4.h, this eliminates all lines corresponding to the shadow. Hence, there remain 10 road triangles, from which the one with the most dominant edges is selected. In the case that the first filtering operation rejects all lines, the algorithm returns the previously detected road triangle. Successive failures to detect a new road triangle reset the algorithm. Note that a scaling operation reduces the optical image’ size so that the smallest size is around but not below 128 pixels. The latter provides sufficient accuracy whilst decreasing the computational burden during the conducted experiments.

Figure 4. Road triangle extraction: (a) the input image, (b,c) gradient orientation, (d,e) Canny edge map, (f,g) modified Canny edge map, (h) all extracted Hough-lines, (i) extracted road triangle
4.1.2 **Scene segmentation**

The scene segmentation determines the different areas of interest, i.e. the target areas. An individual target area is a set of connected pixels which are expected to exhibit similar spectral characteristics. Furthermore, the target area corresponds to a relevant physical area. The latter implies that the found threats could have an impact on the operator’s decision making process.

A first partitioning stems from excluding portions of the image in which threats either are too close to be handled in timely fashion or they are too far and cannot be detected with sufficient accuracy. Both distances depend on the vehicle’s traveling speed. The portion of the image that falls between these two distances is determined by calculating two point couples that describe a line perpendicular to the bearing at the minimum and maximum distance respectively. In Figure 5, these lines are depicted in green. The former yields the distance limitation in the vehicle’s traveling direction. The other distance limitation considers those locations that are in the road’s verges. There is a distance at which a potential threat ceases to be harmful. Therefore, the image analysis includes that part of the verge that is within a certain distance to the road’s edge. Hence, this corresponds to the portion of the image that falls within the extended road triangle. Figure 5 depicts it in blue. It can easily be calculated in world coordinate system as it concerns finding lines parallel to the road’s edges.

A further scene partitioning aims at identifying areas in which the majority of the pixels exhibit some type of similar behavior whilst pixels that correspond to anomalous object do not. It is assumed that the road surface pixels have similar color and/or texture and, that the same holds for each verge separately. The transition between road and verge is not always abrupt. Often there is a littered gutter for which it cannot be expected that the pixels exhibit similar behavior. Although one may expect potential threats here, in this work gutters are excluded from processing. The gutter is set via road triangles that have boundaries parallel to those of the detected road. In Figure 5, the demarcations of the internal gutters are the yellow lines and, that of the external gutters are the white lines. In the sequel, anomaly detection is performed on following target areas:

- Road surface \( \equiv \) the detected road triangle without the internal gutter
- Road’s verges \( \equiv \) the road sides without the external gutter

Figure 6 shows the different target areas for both the optical sensor and the FLIR given the parameters displayed in Figure 5. During the experiments, the left road verge is not considered. Hence the left road verge is not processed.

![Figure 5. Scene segmentation](image)

![Figure 6. Target areas. (a)-(b) road surface, (c)-(d) right road verge.](image)

<table>
<thead>
<tr>
<th>Road scene segmentation [m]</th>
<th>Optical sensor</th>
<th>FLIR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road target distance</td>
<td>[10.0,80.0]</td>
<td>[10.0,80.0]</td>
</tr>
<tr>
<td>Road side maximum distance</td>
<td>7.0</td>
<td>7.0</td>
</tr>
<tr>
<td>Internal gutter</td>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>External gutter</td>
<td>0.6</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Table 1. Scene segmentation parameters in meter
4.2 Anomaly detection

The identification of anomalous objects involves identifying spatial clusters of anomalous pixels, determining their spatial location and characterizing them. The first step estimates an anomaly response for each pixel. The latter solely operates in spectral domain and does not include any spatial relationship. This is handled in the subsequent step which isolates the spatial clusters. In the final step, each cluster is attributed with spatial and spectral features. Remark that the spatial features include traits calculated using the projected world coordinates. The latter determine also the geographical location of the anomaly. The cluster’s features form the basis of the subsequent filtering that retains or rejects the clusters.

4.2.1 Anomaly response

There exist various anomaly detectors ranging from statistical, parametric, non-parametric, neural network, hybrid methods [4][10]. Choosing an optimal approach depends on the task at hand. The scene segmentation simplifies the matter but does not yield precise prior knowledge concerning normal or abnormal pixels. Hence, the detection adopts an unsupervised approach. Furthermore, the fact that the vehicle travels through changing surroundings, may hamper a learning approach. What is normal at one location may turn out to be abnormal elsewhere. Other limitations stem from the near real-time demand and the fact that the method should operate under a wide range of circumstances. The latter imposes that the approach’s parameters have to be robust, physically meaningful and/or preferably extractable from the data or other situational conditions.

The adopted approach, a statistical method, relies solely on the spectral information present in the target area. It resembles the K-nearest-neighbors for which many alterations exist that deal with its computational burden, and clustering approaches. The proposed method commences by extracting representative spectral features from the target areas’ training set. The latter can be the target area or another part of the scene that was deemed to be more suited than the complete target area. The representative spectral features are defined as the colors or gray-levels that exhibit high frequency given a multi-dimensional histogram. In the case of the optical sensor, each RGB-channel is partitioned into 16 bins over the values present in the training area. For each tri-tuple, the frequency of occurrence over the target area is counted. Assuming that the road-scene analysis yields target areas in which a certain percentage \( f_p \in [0, 1] \) surely corresponds to the normal pixels, allows for the determination of the representative tri-tuples. To this end, the bins are sorted by their frequency and the minimum frequency corresponding to this percentage is calculated. A representative tri-tuple is defined a tri-tuple that occurs at least as much as the minimum frequency. For the FLIR, the method uses a similar one-dimensional method.

The anomaly response of each pixel in the target area is given by the minimum distance with the set of representative spectral features. In this work, the adopted metric is the Euclidean distance. The proposed approach has parameters pertaining to the calculation of the histogram, i.e. the adopted dynamic range and the amount of bins for each spectral channel. As mentioned earlier, it was chosen to use 16 bins and set the dynamic range of the spectral channels to correspond with the training data. The only remaining parameter is the percentage \( f_p \). Since the road verge is likely to be less homogenous and aberrations are expected, the threshold is set to \( f_p^{(rv)} = 0.6 \). For the road surface a higher threshold, i.e. \( f_p^{(rs)} = 0.9 \), reflects the expectation of a uniform surface with little to no disturbance. Figure 7 depicts the anomaly responses for the target areas shown in Figure 6.

4.2.2 Anomalous clusters

Applying the predefined threshold yields the binary image. A spatial cluster of connected anomalous pixels defines one unique anomalous object. To increase spatial coherence a set of mathematical morphological filters enhances the binary image. The goal also comprises reducing false positives due to noise. The adopted set of filters depends on the target area because the expectation of anomalous pixels and the spatial scope differs. In the case of the road surface, the spatial area corresponds to a relatively large image part in the middle of the image. Most pixels exhibit a low anomaly response. In general, there will be very few anomalous clusters and unless they correspond to road markings their response shall be limited to a small set of pixels. Therefore, an image dilation with a rectan-
4.3 Attributed anomalous objects

4.3.1 Localization

Once the clusters of connected anomalous pixels are extracted, they are localized in image coordinates. The latter structuring element is applied to the response image. This strengthens the anomalous clusters prior to the thresholding. The binary image is further improved by filling the cluster’s holes and by eliminating small sized clusters. The complete filtering procedure can be expressed by:

\[ I_b^{rs} = BWO \left[ IF \left( ID(I^{rs}, R_{2x8}), T_a^{rs}, \text{holes}' \right), T_b^{rs} \right] \]  \hspace{1cm} (2)

where \( I_b^{rs} \) represent the enhanced binary anomaly image, \( I^{rs} \) is the anomaly response image, \( BWO, IF \) and \( ID \) are the Matlab commands for removing small clusters \textit{bwareaopen}, filling holes \textit{imfill} and dilation operator \textit{imdilate} respectively. \( R_{2x8} \) represents the rectangular structuring element of 2x8 pixels. The parameters \( T_a^{rs} \) and \( T_b^{rs} \) are thresholds. The former is the minimum response for an anomalous pixel. It is set to the distance between two adjacent bin-centers given the full spatial resolution. In other words, for the optical sensor it is \( T_a^{rs} = 16\sqrt{3} \) whereas for the FLIR it is \( T_a^{rs} = 16 \). The latter threshold sets the minimum amount of pixels in a cluster. It depends on the spatial size of the image and the sensor. Remark that this parameter is given for the full resolution of the image. In case of a prior uniform down-sampling, the parameter is modified as follows:

\[ T_b^{(rs,s_i)} = \max \left[ \text{ceil} \left( \frac{T_b^{(rs,s_0)}}{i^2} \right), 1 \right] \]  \hspace{1cm} (3)

where \( i \) represents the scaling factor. The adopted optical sensor has a full spatial resolution of 1080x1920, whilst the FLIR has a usable spatial resolution of 395x563. For this configuration, \( T_b^{rs} = 160 \) and \( T_b^{rs} = 16 \).

In the road’s verge, the expectation of noisy responses is higher. Also, this part of the scene may contain objects of a certain height. Therefore an alternative set of filters is applied:

\[ I_b^{rv} = BWO [ IO \left( IF \left( I^{rv} \geq T_a^{rv}, \text{holes}' \right), S_4 \right), T_b^{rv} ] \]  \hspace{1cm} (4)

Figure 8 depicts the detected clusters prior to the physical size filtering. The example shown does not exhibit any true anomaly in the road surface. This corresponds well with the detection result of the FLIR. In the case of the optical sensor, the road markings and the shadow of the lighting pole are deemed anomalous. In road’s verge, the optical sensor provides a better detection result. The FLIR creates several anomalous clusters in the verge that correspond to weeds in the grass. Increasing the threshold rejects these clusters whilst retaining the true anomaly in this particular example. However, in other circumstances the anomaly will be missed. Therefore, another filtering based on shape and other features is applied in the next steps.

4.3 Attributed anomalous objects

4.3.2 Characteristics

Each cluster corresponds to a set of spectral values and anomaly responses. For each set, the mean and median are

Once an adequate location for a cluster is calculated, the geographical coordinates of the cluster are calculated using EQ.2.
used as representatives. Hence in case of the optical sensor, there are 6 spectral features and 2 anomaly responses. In case of the FLIR, there are only 2 spectral features. The Hu set of geometric invariant moments [9] describes the cluster’s shape. The latter is done both in pixel and in world coordinates. For each cluster an area estimate is calculated by assuming that the cluster represents a flat object. Based on this estimate, too small and too large clusters are rejected from further processing. Additionally, all clusters detected by the optical sensor that correspond to white blobs in the center of the road are recognized as road markings and withheld from further processing. Consequently, all clusters that were detected on the road' surface, shown in Figure 8, are filtered out. Figure 9 shows the retained threat evidence present in the right verge.

![Figure 9. Threat evidence in the right road verge.](image)

**Acknowledgements**

The reported work is carried out in context of E.D.A. project A – 0444 – RT – GC, entitled Surveillance in an Urban Environments using Mobile Sensors.

**References**


