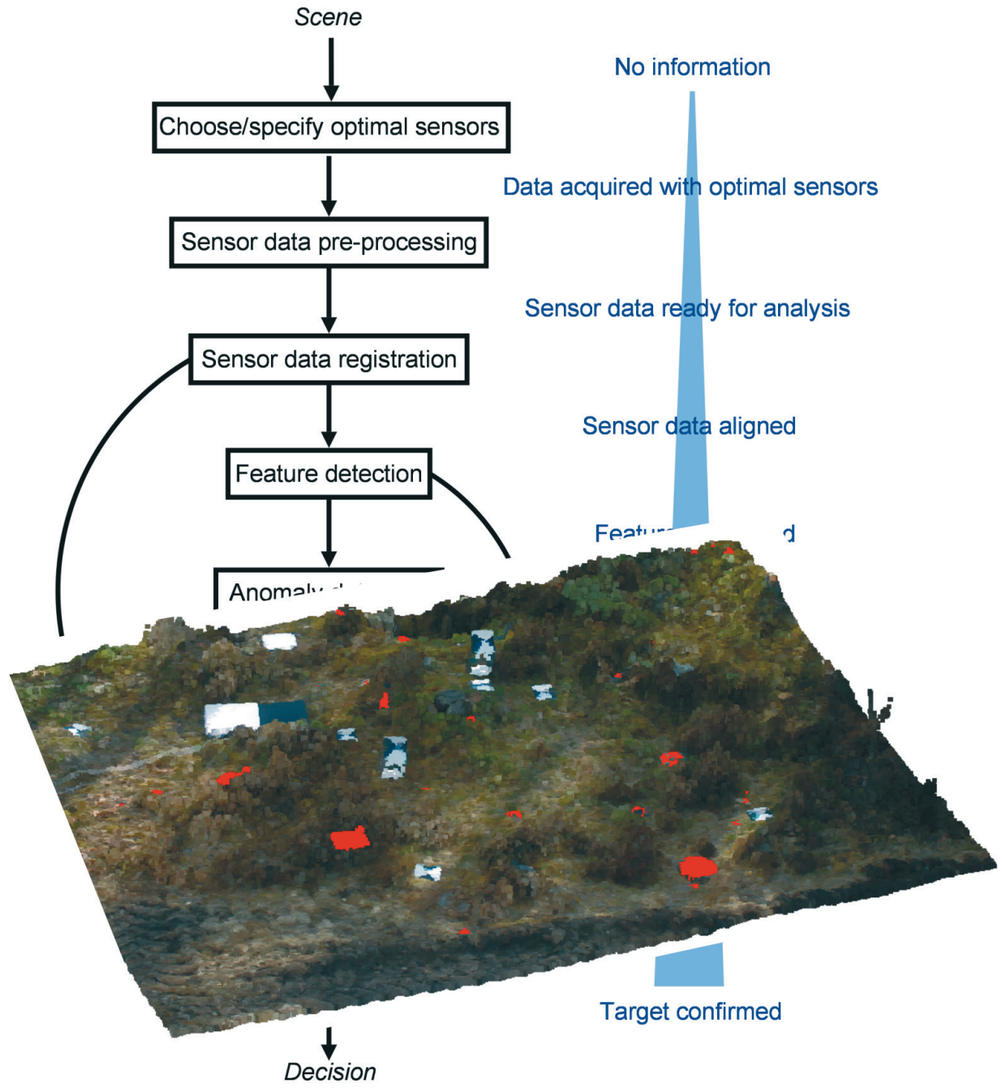
Detection and recognition of surface-laid mines

Summary of signal processing techniques

GUSTAV TOLT, HADI ESTEKI, ANDRIS LAUBERTS,
CHRISTINA GRÖNWALL, NICLAS WADSTRÖMER



FOI, Swedish Defence Research Agency, is a mainly assignment-funded agency under the Ministry of Defence. The core activities are research, method and technology development, as well as studies conducted in the interests of Swedish defence and the safety and security of society. The organisation employs approximately 1000 personnel of whom about 800 are scientists. This makes FOI Sweden's largest research institute. FOI gives its customers access to leading-edge expertise in a large number of fields such as security policy studies, defence and security related analyses, the assessment of various types of threat, systems for control and management of crises, protection against and management of hazardous substances, IT security and the potential offered by new sensors.

FOI
Defence Research Agency
Information Systems
P.O. Box 1165
SE-581 11 Linköping

Phone: +46 13 37 80 00
Fax: +46 13 37 81 00

www.foi.se

FOI-R--2777--SE
ISSN 1650-1942

Scientific report
June 2009

Information Systems

Gustav Tolt, Hadi Esteki, Andris Lauberts,
Christina Grönwall, Niclas Wadströmer

Detection and recognition of surface-laid mines

Summary of signal processing techniques

Titel	Detektion och igenkänning av ytlagda minor
Title	Detection and recognition of surface-laid mines

Rapportnr/Report no	FOI-R--2777--SE
Rapporttyp Report Type	Vetenskaplig rapport Scientific report
Sidor/Pages	77 p
Månad/Month	Juni/June
Utgivningsår/Year	2009
ISSN	ISSN 1650-1942
Kund/Customer	FM
Kompetenskloss	8 Samverkande sensorsystem
Extra kompetenskloss	9 sensorinformatik
Projektnr/Project no	E3084
Godkänd av/Approved by	Pär Carlshamre

FOI, Totalförsvarets Forskningsinstitut
Avdelningen för Informationssystem
Box 1165
581 11 Linköping

FOI, Swedish Defence Research Agency
Information Systems
Box 1165
SE-581 11 Linköping

Sammanfattning

I denna rapport sammanfattas resultaten av det signalbehandlingsarbete som bedrivits inom projektet Multi-optisk minspaning (MOMS). Ett antal metoder för och aspekter av bl a dataregistrering, anomalidetektion, egenskapsextraktion, datafusion och igenkänning av minor beskrivs och diskuteras. Ett antal särskilt intressanta metoder har testats och utvärderats på sensordata från olika scener, för att möjliggöra analys av respektive metods för- och nackdelar under olika förutsättningar. Slutsatser från experimenten presenteras och diskuteras, med särskild tonvikt på aspekter som rör signalbehandling i ett sensorsystemperspektiv.

Ett flertal olika elektrooptiska sensorer, såväl passiva som aktiva, har beaktats inom MOMS. I rapporten presenteras en metod för sensoroptimering som ger verktyg för att utforma en förhållandevis enkel elektrooptisk sensor som ändå är adekvat för uppgiften. Detta kan åstadkommas med hjälp av en informationsteoretisk dataanalys i vilken spektralband definieras utifrån mängden information de innehåller.

För att data från flera sensorer ska kunna samutnyttjas måste data transformeras till ett gemensamt koordinatsystem. Kvaliteten på positionsbestämningen avgör på vilken nivå man kan fusionera data; ideal registrering möjliggör fusion ner på pixel- eller signalnivå. I ett distribuerat sensorsystem där, säg, data insamlade med en flygande plattform ska kombineras med data från en markbunden sensor, blir sannolikt pixelfusion av dessa data mycket svår att uppnå utan fusion måste då ske på en högre nivå, t ex beslutsnivå. Från ett signalbehandlingsperspektiv är det önskvärt att sensorerna sitter väldigt nära varandra, helst t o m med en gemensam optik eller detektor så att registreringen kan göras så noggrant som möjligt.

Bland de undersökta signalanalysmetoderna framstår anomalidetektion som en nyckelkomponent i ett systemkoncept. Denna metod syftar till att detektera sådant som avviker från det normala i scenen (bakgrunden) och ger därför en första indikation på var ev. minor kan finnas. Dessutom kan denna metod potentiellt användas för detektion av andra objekt än minor, t.ex. IEDer.

De detekterade anomalierna analyseras sedan genom att olika antaganden om målobjekten utnyttjas, bl a beträffande deras storlek. Detta leder till att minlika objekt kan detekteras. Om man dessutom har tillgång till detaljerad information om vissa måltyper, t ex i form av CAD-modeller eller bilder, erhållen innan eller insamlad under uppdraget, kan de detekterade objekten gå vidare till ett igenkänningssteg. Där undersöks de detekterade objektens likhet med ett antal måltyper.

Nyckelord: mindetektion, elektrooptiska sensorer, signalbehandling, datafusion

Summary

This report summarizes the signal processing work carried out within the project Multi-optical mine detection (MOMS). A number of methods for and aspects of data registration, anomaly detection, feature extraction, data fusion and mine recognition are described and discussed. A number of especially interesting methods have been tested and evaluated with sensor data from different scenes, in order to allow for analysis of pros and cons under certain conditions. Conclusions from the experiments are presented and discussed, with focus on aspects concerning signal processing in a sensor system perspective.

A number of electro-optical sensors, passive as well as active have been considered within MOMS. In this report, a method for optimized sensor design is presented, that provides a tool for designing a relatively simple sensor that still is adequate for the task. This can be achieved through analysis based on information theory, in which the spectral characteristics of the sensor are defined based on the information they contain.

In order for data from several sensors to be combined, the data has to been registered, i.e., transformed into a common coordinate system. The quality of the registration strongly influences the level at which data can be combined; ideal registration allows for fusion on the lowest level (pixel- or signal-level). In a distributed sensor system where, say, data from an airborne system shall be combined with data from a ground-based sensor, pixel-level fusion will probably be difficult to use. From a signal processing perspective, it is desirable that the sensors are mounted close to each other, preferably with common optics and/or detector array, so that the registration can be as accurate as possible.

Among the signal processing techniques considered, anomaly detection emerges as a key component in a system concept. This method detects things that are different from what is expected (the background) and thus gives a first indication of possible mines. In addition, this technique can potentially be used for detection of other objects, e.g. IED's.

The detected anomalies are then analyzed further, by using certain assumptions concerning the targets, e.g. their expected size. This leads to the detection of mine-like objects. If available detailed data about certain targets, e.g. CAD models or images, given before or collected during the mission, can be used for mine recognition, in which the similarity between detected objects and these targets is investigated.

Keywords: mine detection, electro-optical sensors, signal processing, data fusion

Contents

1 Executive summary

This report describes the signal processing work carried out within the MOMS project. The main purpose of this work is to provide knowledge about what results could be expected in terms of mine detection using certain sensors.

More specifically, the report contains the following:

- a proposal for a signal processing framework for mine detection and recognition
- a description of the relevant signal processing techniques involved
- performance assessments of signal processing techniques in terms of detection and false alarms based on real sensor data acquired within the project
- conclusions and discussion, with focus on aspects related to a realization of a system for mine detection

Figure 1 shows the main structure of the signal processing framework outlined in this report, illustrating the order in which different processing tasks are generally performed. Starting with a certain scene, the goal is to detect and, if possible, recognize any mines in the scene. Information about possible mines can then be handed over to an operator for further analysis and/or decision-making.

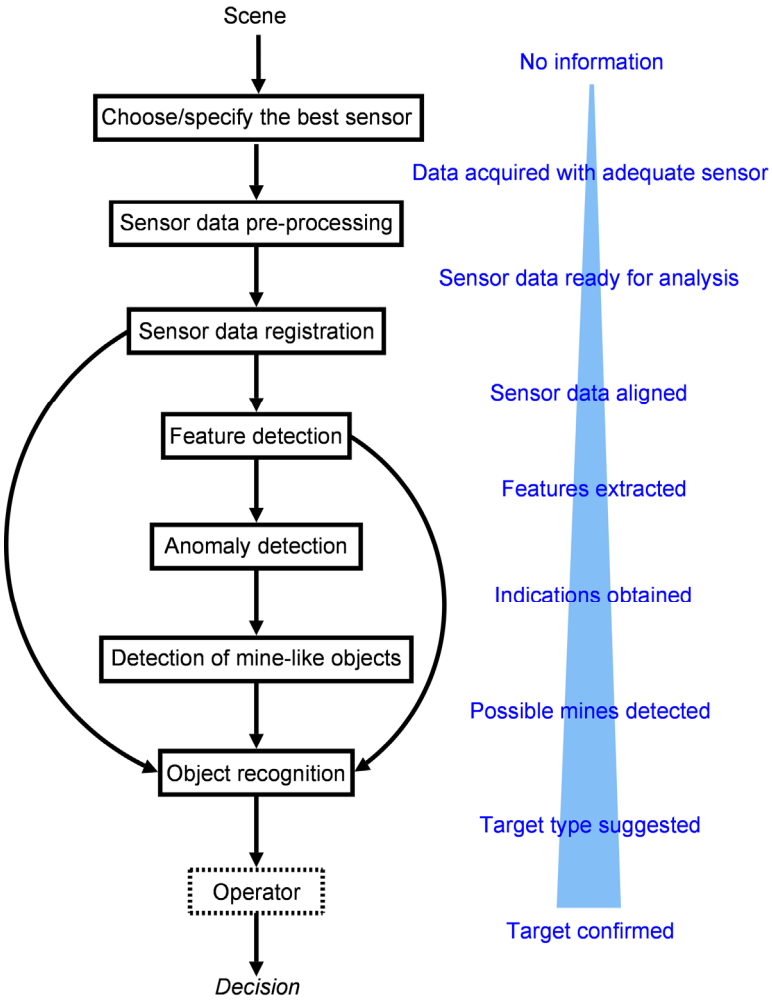


Figure 1. Description of the mine detection and recognition process developed in MOMS.

The first, and very important step, is to use an efficient sensor. Without a sensor that is able to capture the necessary information, it does not matter what kind of signal processing

is applied afterwards. In this report, we show that by using measures from information theory, techniques for finding the optimal sensor can be formulated (Section 4).

The data from the respective sensors typically have to be subject to some individual processing before the actual analysis can take place. This is referred to as sensor data *pre-processing*. Generally this includes very sensor-specific processing, such as radiometric calibration, geometric correction, laser radar pulse detection, etc., and thus falls outside the scope of this report.

After having acquired data from the different sensors, the different data sets have to be transformed into a common coordinate system, so that we can combine the information from the sensors. This is known as *data registration* (Section 5).

The next step is typically to *extract relevant features* from the data, features that discriminate expected targets from the background. Section 6 briefly discusses some approaches investigated within MOMS for such feature extraction, e.g. detection of 3D surfaces, convexity, temperature contrast, etc.

In many cases, the mines manifest themselves as being different from what is expected to be found in the scene. To discriminate objects that are different from the scene can be formulated as a task called *anomaly detection* (Section 7). In practice, one often makes use of the assumption that mines are relatively rare and implicitly that the vast majority of the data corresponds to the background. This technique is often quite appealing, as it does not involve any explicit information about particular characteristics of the targets.

Although the detected anomalies often do correspond to any mines in the scene, various clutter in the scene is also detected, as a result of the fact that the anomalies correspond to anything that is not background. The next step is then to refine the results further and to *detect mine-like objects* (Section 8). For example, the detected anomalies could be judged in terms of spatial and spectral characteristics. Or we may want to ignore anomalies that correspond to too small or too large objects that cannot possibly be mines. We can also analyze the similarities between them and remove those who appear too frequently in the scene. Several sources of information could also be combined by performing data fusion, in order to refine the results further.

The idea in mine recognition (addressed in Section 9) is to take a detected mine-like object and determine what type of mine it actually is, using a database with previously stored information about what characterizes the respective mine type. For example, such a database could consist of CAD models, high-resolution images, or certain spectral characteristics of the respective target. This step often requires high-resolution data about the sought-after objects.

The main conclusions can be summarized as follows:

- Most tested approaches can meet real-time or near real-time demands.
- Occlusion causes difficult problems when detecting small ground objects.
- Combinations of (broad band) spectral and spatial techniques have shown to be relatively robust.
- Fusion on the signal/pixel level requires very accurate data registration. Such accuracy is difficult to obtain with a distributed sensor system, and will probably require a common detector array or arrays situated very close to each other. Fusion on the decision-level copes considerably better with a less accurate registration, as the different sensor data streams are processed individually and only the final outputs are combined.
- Anomaly detection is a very useful tool for detecting possible mines. The real benefit is that the anomaly detector only has to be trained with background data, not with targets.

- Spatial feature extraction is attractive but somewhat difficult, as it requires good object/background separation. It is attractive from a computational viewpoint as the spatial feature extraction often is convolution-based and the amount of numerical operations per frame needed to compute the desired features is known beforehand. This makes spatial features suitable for hardware implementations close to the sensor.
- Mine detection based solely on supervised classification cannot be recommended; it is risky to rely on that our target database is kept up-to-date and contains information about all the possible threats the system may encounter. Nevertheless, such a technique can run in parallel with anomaly-based detection and report whenever the system encounters an object that is very similar to a target with which it was trained.
- The spatial resolution of the sensor must allow for having several pixels on the target. For detection the pixels should correspond to a resolution on the target of maybe about 2-3 cm, to enable the removal of small, irrelevant objects. For mine recognition based on spatial properties, the sensor resolution should be significantly better than 2 cm, probably around 5 mm or below. Even at that resolution, it may be difficult to determine object type.
- A system for detection of small ground objects, like land mines, would benefit from including both passive and active imaging sensors, preferably operating at a broader range of wavelengths. This will provide 24h capabilities and could reduce problems caused by uneven and unpredictable illumination of the scene (e.g. shadows). Stable illumination is favorable from a signal processing point of view.
- The training-based algorithms, i.e., the anomaly detection and the supervised approaches, have a benefit in that they can be updated under a mission, to adapt to the current conditions in the area of interest. Through an extra training phase, supervised by a skilled operator, the algorithms can be tuned to the new environment and the false alarm rate can be lowered while retaining the mine detection rate.

2 Introduction

The tactical land mine detection problem is very challenging, as illustrated by the lack of operational systems with rapid surface coverage in the international arena. In order to improve mine detection and classification performance, the project Multi-Optical Mine Detection (MOMS) was initiated to develop and investigate new ideas for optical sensing, signal processing, target and background characterization (Sjökvist, *et al.*, 2005). Within the project, both new enabling technologies and new system concepts are of interest.

The MOMS project was originally formed to build a deeper knowledge of the phenomena and potential sensor technology to use in a future system demonstrator. The MOMS mission has not been to build a real system but to deliver the specification and guidelines for such a system. With some changes concerning the task for MOMS, the focus of the project is now limited to an assessment of concepts; to analyze and describe the possibilities and shortcomings of various sensor combinations, signal processing techniques and system concepts.

Within MOMS, there have been several measurement campaigns aiming at collecting sensor data from realistic scenarios (e.g. see Letalick *et al.*, 2007 and Larsson *et al.*, 2008). The acquired sensor data, together with knowledge of physical properties and sensor characteristics, have then been used for investigating the relevance of various phenomena for the mine detection problem (e.g. see Letalick *et al.*, 2006, Renhorn *et al.*, 2008).

From the outcome of the sensor data analysis, initial results have emerged concerning what sensor types and objects properties that seem reasonable to exploit in a real system. Moreover, internal meetings and workshops with the customer have been held to discuss different total concepts for MOMS, including performance and potential ways of tactical operation. A number of system concepts have then been formulated for further evaluation during 2009 (Steinvall *et al.*, 2008).

This scientific report is focused on the signal processing part – how to exploit the phenomena (e.g. spectral and spatial object properties) in order to obtain automatic mine detection based on sensor data. A user report will finalize the MOMS project.

2.1 Scope of this report

In this report we present a signal processing concept for detection of surface laid mines with an electro-optical system developed within the MOMS project. The report contains

- a proposal for a signal processing framework for mine detection and recognition
- a description of the relevant signal processing techniques
- performance assessments of signal processing techniques in terms of detection and false alarms based on real sensor data acquired within the project
- a discussion about system realization issues related to signal processing
- conclusions and recommendations

2.2 Limitations

For practical reasons, this report is subject to some limitations, of which the following are the most significant:

- Polarization effects, although considered within MOMS as an interesting phenomena candidate, have not been studied here, due to the lack of usable sensor data.

- The number of scenes and targets studied in this report are relatively limited. This is partly due to the difficulties associated with obtaining accurate data with the desired range of sensors for many different environment and many realistic targets, but also due to the fact that from a signal processing point of view, it was considered important to focus the work on certain scenes.

2.3 Assumptions

In order to provide a starting point for the signal processing, some assumptions have to be made about the targets and their relation to the scene. The main assumptions used in this work are:

- *The mines are surface-laid.*
  This follows from the directions given for the MOMS project (Sjökvist *et al.*, 2005, Sections 1.3–1.5)
- *There are a number (>1) of measurement samples, e.g. pixels and 3D points, on the target.*
  This assumption is actually a direct consequence of two other assumptions:
  - *The spatial resolution of the sensors is such that targets are spatially resolved at the viewing distance.*
    This implies that each target, when projected onto the sensor arrays, corresponds to several data samples (pixels, 3D points).
  - *A part of the target is visible for the sensors.*
    This basically means that the targets should not be occluded to a degree where only one or very few measurement samples are available. In the extreme case, a perfectly occluded object, e.g. hidden behind a tree, would not be detectable at all with the range of considered sensors. Still, however, this does not mean that there are no perfectly occluded objects in the data sets, but that such objects cannot be detected by any signal processing technique.
- *Mines are relatively rare, compared to the background.*
  We can thus detect possible mines by finding pixels and objects that are *unusual* compared to what is expected to be found in the scene.

2.4 Outline of the report

The remainder of the report is organized as follows. Section 3 gives an overview of the signal processing concept used within MOMS and a short summary of each of the signal processing stages involved. Sections 4 through 9 contain a deeper description of each data analysis step, from techniques for optimal sensor design to approaches for mine recognition. Section 10 is devoted to a description of data and tools for performance assessment, involving a short description of the sensors and the scenes considered as well as an introduction to the Receiver Operating Characteristics (ROC) analysis used for quantifying performance. In Section 11 the results of applying a set of selected algorithms to the sensor data are shown and discussed. Section 12 contains conclusions and a discussion about signal processing issues in a system context. In Section 13, references to related work are given and finally, in Appendix A some additional results are included.

The more system-oriented reader could focus primarily on Sections 1, 2, 3, 11 and 12, and if desired refer to Sections 4 through 10 for technical details.

3 Signal processing framework for mine detection and recognition

In the MOMS project, a signal processing framework has been developed for detection and recognition of surface-laid mines. In this section, a condensed description of the framework will be given, and the interested reader is referred to following sections for more detailed descriptions and discussions about each separate set of functions.

The goal of the signal processing work has been to design a framework that could help an operator detect and recognize potential threats (mines). In order to succeed with this, efficient signal processing techniques should be used that detect or high-light true targets while ignoring irrelevant objects and the background. Figure 2 shows a simplified schematic diagram of the main stages involved in the data processing. For simplicity, the graph emphasizes the conceptual signal processing layout and hence dependencies on other information than sensor data have been excluded from the diagram. Examples of such information are *a priori* information about expected target size and estimated target density, target model libraries, etc. that are still necessary for the signal processing.

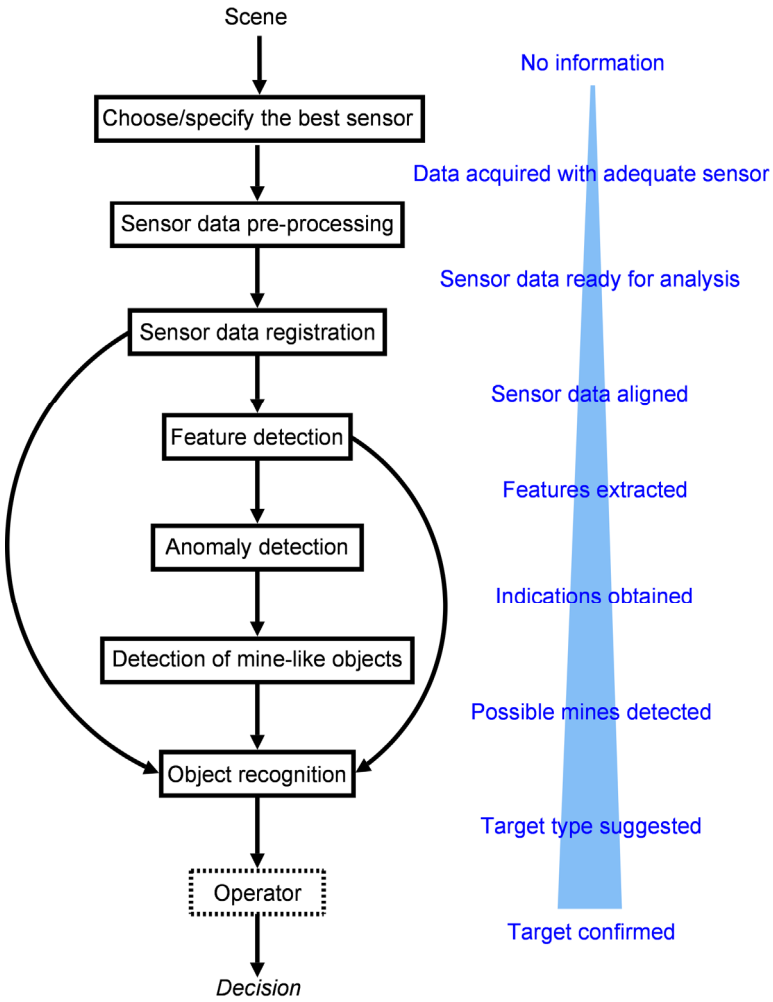


Figure 2. Schematic description of the mine detection and recognition process developed within MOMS.

It should be pointed out that the scheme aims at explaining the framework on a conceptual level, showing how the information about the scene gradually evolves as the processing goes on. In practice, the order in which the different tasks are carried out may vary. For example, it could be that the computational resources needed for feature detection are

preferably applied only to certain especially interesting regions, e.g. after the anomaly detection. Some tasks may even be left out completely. For example, an explicit feature detection step is not required if the spectral data associated with the image pixels (already spectral features in themselves!) is enough to solve the task at hand. And obviously data registration is needed only when several sources of information are to be combined.

3.1 The framework

An initial, and critical step, is to use an efficient – preferably the optimal – sensor. Ideally, one would like to have a sensor that could measure "everything" very accurately and very quickly. In practice the problem is to find the best balance between obtaining good results in terms of detection and false alarm rates, and having sensors that meet operational requirements. In Section 4, a framework for finding the optimal sensor characteristics is described, based on information theory. More specifically, an approach for finding the near-optimal sensor configuration in terms of spectral bands and dynamic range is outlined.

The data from the each sensor typically have to be subject to some individual processing before the actual analysis can take place. This is referred to as sensor data *pre-processing*. It is beyond the scope of this report to describe techniques for such low-level data processing. Instead, the reader is referred to other work for reading more about typical pre-processing techniques, such as geometric corrections to the images due to known system imperfections, performing radiometric calibration (Nelsson and Nilsson, 1999), extracting 3D data from laser waveform signals (Tolt and Larsson, 2007), enhancing image contrasts (Rahman *et al*., 2005), etc.

Then, in order to be able to compare or combine data obtained from different measurements, the different data sets have to be transformed into a common coordinate system, so that we know the relationship between the data elements (pixels, 3D points) for all the sensors. This is known as *data registration* (Section 5).

An important part of the signal processing concept for land-mine detection is the ability to *extract relevant features* that are (more or less) discriminative to the targets. These features are often spatial in nature, meaning that information from several neighbouring pixels is combined to form more high-level information related to shape, structure, local similarities, etc. Section 6 contains a brief discussion about some approaches investigated within MOMS for such feature extraction, e.g. detection of 3D surfaces, convexity, temperature contrast, etc.

In many cases, it is quite difficult to extract all the relevant mine features that would help discriminate between mines and the background. Among other things, this is due to the great variations in object appearance and background characteristics, as well as limitations due to sensor noise. However, even if we are not able to extract features that are similar to those of the targets, we can use the fact that targets often manifest themselves in the sensor data by being *dissimilar* to the background, i.e., different from what is expected to be found in the scene. This can be formulated as a task called *anomaly detection* (Section 7). Its goal is to detect regions that are considered to be different compared to the background. A key here is hence to obtain, and maintain, an accurate model of the background as well as defining metrics for measuring quantifying the dissimilarity. In practice, one often makes use of the assumption that mines are relatively rare and implicitly that the vast majority of the data corresponds to background.

In the anomaly detection step we typically do not use explicit information about particular characteristics of the targets; it is only the dissimilarity to the background that is measured. Thus, an anomaly can correspond to any type of object or structure that is different from the background model, be it a true target, clutter, natural objects or shadows, etc. And the features detected (surfaces, convex regions, etc) is typically not enough to actually make a decision about the existence of a mine, as it primarily gives indications. However, by

combining results, using more sophisticated algorithms and using more information about the expected nature of the target, we move on to *detection of mine-like objects* (Section 8). For example, the detected anomalies could be judged in terms of spatial and spectral characteristics (Section 8.1 and 8.2, respectively). For example, following the assumptions outlined in Section 2.3, we could ignore anomalies that correspond to single pixels. We could also remove anomalies that appear too frequently in the scene and thus are likely to be natural objects not captured by the current background model. Several sources of information could also be combined by performing data fusion (Section 3.2), for example using the EM-MML-based unsupervised clustering approach described in Section 8.3 or via supervised classification (Section 8.4).

In Section 9, we describe some approaches for mine recognition. The idea here is to take the mine-like object (Section 8), and determine what type of mine it actually is, using a database with previously stored information about what characterizes the respective mine type. For example, such a database could consist of CAD models, high-resolution images, and detailed spectral characteristics of the respective target. Now, a universal and complete database covering the whole range of existing mine types will be very challenging to obtain and maintain, not to mention the difficulties of processing a huge object database for a real-time system. However, if there is *a priori* information available about certain expected targets, the mine recognition module could be trained to detect those targets, operating in parallel with the anomaly detection-based detection. Stated differently, we could have a *top-down*, or *hypothesis testing*, approach ("Is there any TMM-1 mine anywhere in this scene?) working side by side with a bottom-up approach ("Is there any mines in this scene, and in case there are, which kind?").

3.2 Data fusion

It could well be that no single sensor alone carries enough information about the scene to provide reliable results, but that data obtained from several sensors correlates in a way that the sought-after objects can be found with greater success. This is the underlying idea in *data fusion*. Broadly speaking, the term fusion only refers to that different sources of information have been combined; in itself it does not specify how the information has been treated. Often one distinguishes between fusion performed on different levels, from signal-level, via feature-level to decision-level. Different aspects and interpretations of these terms may exist (and indeed be perfectly natural) depending on the purpose of the fusion. In this section, we describe the data fusion perspective used within MOMS and give a couple of examples of it in practice.

Signal-level fusion can be considered analogous to the hyper-spectral analysis described in Section 7, where the spectral bands can actually be seen as individual "sensors" whose data are stacked to form a multi-dimension data cube. The principle is basically the same for stacking data from different sensors although many other problems may then occur (registration, occlusion, etc).

In signal-level fusion, no feature extraction for the individual sensors has taken place prior to the stage where the data from the different sensors are actually combined. Often, however, data from all or some of the participating sensors are pre-processed to yield more specific pieces of information, *features*. The underlying idea is simply to make the best possible use of each sensor. For example, "raw" laser radar data (3D points) data may be used to find surfaces (Section 6.1), thermal images may be used to find interesting spatial characteristics (Section 6.2), etc, and the results for each sensor can then be combined in a feature-level framework. Theoretically, feature fusion should give the best result as it allows for combining the features in a way that best solves the problem at hand.

Data fusion can also be applied to the decisions made from processing data from each sensor individually. This is known as *decision-level fusion* and it leaves the option for making decisions based on different (possibly stand-alone) classifiers. This makes it very easy to include more, or other, classifiers and combine in the existing framework.

Decision-level fusion is also generally more attractive from a computational viewpoint since it does not require access to, or processing of, features from all the sensors at the same time. The fact that each sensor data stream is processed separately and only the final results for each sensor need to be communicated is advantageous for a distributed system.

To summarize, the different types of fusion differ from each other by the level of processing applied prior to the stage where the data from the sensors are used together; if all data are considered directly it is signal-level fusion, if the data from all sensors are treated separately and only the results are combined it is decision-level fusion, with feature-level fusion somewhere in-between. In practice, the applied fusion procedure is often a combination of the different levels.

4 Optimal sensor configuration

The goal is to design a system that conveys as much information as possible about the number of targets in the scene and their positions and appearances. The number of targets and their specific appearances are unknown. Also the terrain is previously unseen. It is necessary to make assumptions about the size and density of the targets. In our examples we assume that the targets are from centimetres to a few decimetres across the top. They are usually square or round but other shapes are possible. The densities of targets that are expected can be divided into three groups: there are no targets at all in the scene, there are a few targets in the scene and there are many equal or similar targets in the scene with distance between them ranging from metres to tenth of meters. The targets are the only man made objects in the scene except for a few objects which may be mistaken for targets.

The goal of the system is to for each measurement (pixel) determine if there is a target or not. The goal can also be formulated as to determine the number of targets in the scene and their positions. If the probability of a target in each pixel is known then it is possible to make a probabilistic decision about whether there is target or not. The optimal decisions can be made if the conditional probability $p(x \mid y)$ of the presence of a target given the sensor data is known.

In this section we will describe how some of the parameters of the sensor can be optimized using the mutual information and the ROC curve.

4.1 An information theoretic approach

Target discrimination has similarities with communication though in the target discrimination case the deployer of the targets does not wish that the positions of the targets be known to the discriminator. However the positions of the targets are information that is disclosed by emitted and reflected radiation.

Information theory is about theoretical limits to communication where the system is described by stochastic models. Entropy is a measure of the amount of information coming from a stochastic source and the entropy has direct implications on the number of bits needed to encode the information. Mutual information is a measure of the mutual relation between two stochastic variables. The mutual information is interesting when there is some dependence between the two stochastic variables. The mutual information between two variables can be interpreted as the amount of information that one of them convey about the other. In this context it may be appropriate to remind that the measure is statistical and the information in a certain context may have a different value and which will affect the decision. For example, in the mine discrimination case it usually has much more severe consequences to miss a mine than to make false detection of a mine. The consequences are often called costs. Hence to make the decision that will minimize the costs these has to be known together with the probabilities of making erroneous and correct decision. In this work will only consider the probabilities of correct and erroneous decision because when these probabilities are known it is possible to weigh them with the costs to get the decision that minimize the average expected cost.

Figure 3 shows a model of target discrimination from an information theoretic point of view. The targets are deployed at specific positions in the scene which may have been chosen deterministically or stochastically. In the model (Figure 3) the positions are represented by the stochastic variable X . The targets are deployed in an environment which emits electro-optical signals represented in the model by the stochastic variable Z . The sensor measures the radiation and gives some values as output. In the model the measurements are represented by the stochastic variable Y . The detector is a function that estimates the presence or absence of a target at a given position based on the sensor output and possibly previous sensor data. The estimated presence or absence of a target or the positions of the targets are represented by the stochastic variable $\widetilde{X}$. The challenge is to

construct a system where the estimated positions $\widetilde{X}$ are as close as possible to the true positions X. It is also interesting to find out any theoretical limits to how much information $\widetilde{X}$ may give about the true positions X.

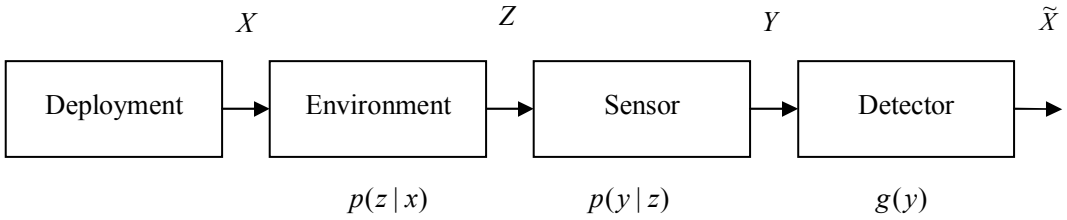


Figure 3. An information theoretic model of target discrimination.

The occurrence of a target and the sensor data in a specific pixel is unknown before the measurement. However it is likely that some statistical knowledge is available. Thus the parts in the system will be modelled with probability distributions. Let X be a stochastic variable with two outcomes 0 for background and 1 for target and with the probability distribution $p(x)$. The sensor and the environment can be described by a conditional probability distribution $p(y \mid x)$. The detector is a deterministic system that is best described by a function. Y is a stochastic variable describing the sensor output which in the case of hyperspectral images has many dimensions. In our case the instrument presents integer valued measurements. The outcome is described by the probability distribution $p(y)$. However it is difficult to estimate $p(y)$ because the space is too large compared to the number of possible values. We have used Gaussian mixture models which have fewer variables and are possible to estimate with the available number of data points.

A similar model which we have not used is to let X describe the position of the targets instead of binary variable describing the presence or absence of a target for each pixel.

The entropy $H(X)$ is the a priori uncertainty if the point is a target or background. The mutual information how much information the sensor data Y gives about the presence of a target and $H(X \mid Y) = H(X) - I(X,Y)$ which is the remaining uncertainty after the sensor data has been considered. The remaining uncertainty $H(X \mid Y)$ tells how many bits will be needed to encode the presence of a target given that the sensor data is known. The conditional entropy $H(X \mid Y)$ can also be interpreted as the number of possible decisions that cannot excluded.

Mutual information is defined as

$$I(X,Y) = \sum_{(x,y)} p(x,y) \log_2 \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

and can be written as $I(X,Y) = H(X) - H(X \mid Y)$ which should be interpreted as the difference between the amount of information in X and the information in X given that the sensor data Y is known. The amount of information is the least number of bits needed to encode the data so that it can be reconstructed with an arbitrarily small probability of error. The amount of information is the average limit when long sequences of data are coded jointly. If each data is assigned a code word with the length $-\log_2(p(x))$ then these code words will give a mean code word length. The chosen sequences will roughly have equal probability. In the case of mutual information for target detection the mutual information show how many bits of information that the sensor data convey about the presence of a target. Assume that there is a target in a sensor image with 1000×500 pixels corresponding to 10x5 meters on the ground. Also assume that the location of the target needs to be determined by decimetre precision. Then there are 100x50 possible positions and they all have the same probability then around 12 bits of data is needed to describe the position. The information of primary interest is very limited. However, probably sensor data need to be used by the operator to determine if an indication is a target or not. Thus it is not possible to keep the amount of data down to this lower limit.

Mutual information can be used when it is possible to estimate the probability distributions and when it is possible to compute the value. We have used the measure to evaluate different set ups and different parameter values. We have not found any ways to estimate parameters that maximizes the mutual information.

The ROC curve can also be used to compare different configurations. The curve shows the probability of detection and the probability of false alarm as a function of a decision parameter. If the decision variable has a natural order then the probabilities can be obtained by choosing the decision level.

If the sensor variable does not have a natural order than the decision function may be too complex to be realised. If the decision for each sensor value is completely different from other values then the description of the decision function may be too complicated to be realised.

We show two examples to illustrate the relation between the mutual information between two variables and the optimal decision based on each sample independently from other samples. In the first example it is optimal to guess that there is no target present no matter what the sensor data Y is and in the second example it is optimal to guess different for the two Y values.

Table 1. Example 1: A joint probability distribution $p(x, y)$

$p(x, y)$		Y		
		0	1	
X	0	0.75	0.05	0.8
	1	0.16	0.04	0.2
		0.91	0.09	

In example 1 the X variable describes the occurrence of a target or not and the Y variable describe the sensor output which in this case only has two values. Table 1 shows the joint probability distribution between the two variables. In Example 1 the mutual information $I(X, Y) = 0.0223$. The grey cells show the decision for each sensor output with the largest probability of being correct. The probability of error $p_e = 0.25$. Taking decisions pixel by pixel will not give any information since the decision will always be 0.

Table 2. Example 2: A joint probability distribution $p(x, y)$.

p(x,y)		Y		
		0	1	
X	0	0.76	0.04	0.8
	1	0.15	0.05	0.2
		0.91	0.09	

In Example 2 the mutual information $I(X, Y) = 0.0451$. The grey cells show the decision for each sensor output with the largest probability of being correct. The probability of error $p_e = 0.19$. In this case the decision will give some information since the decisions are different.

Even if there is mutual information between sensor data and data source it is not always the case that this information can be retrieved. However it is possible to get information out of the situation in Example 1 by making simultaneous decisions. Consider a block of n pixels. For each block of sensor pixels the most likely decision will have a small probability. However there will be a relatively small set of possibilities that will have a probability that is close to one. There are many possible decisions that can be ruled out.

The situation can be illustrated by a binary variable with probabilities $p(x = 0) = 0.9$ and $p(X = 1) = 0.1$. Each variable is drawn independently from the others. Consider the probability of blocks of $n = 10$ values. The probability of the most probable block is 0.9^{10}=0.35 and decrease as n increases. However the eleven most likely blocks will have

the probability $0.9^{10} + 10 \times 0.9^{9} \times 0.1 = 0.74$ and yet only $11/1024 = 0.011$ of the total number of blocks.

The mutual information may not be enough to point out a certain position as a possible target with high probability of being correct but many positions may be ruled out as having a target with high probability of being a correct decision. It may be relevant to rule out positions if there are only a few remaining positions that possible could have a target.

5 Sensor data registration

In order to be able to benefit from using data acquired with multiple sensors and/or from different positions, the data have to be brought from each sensor's own coordinate system into a common coordinate system. This is called *data registration*. Ideally, we would like to arrive at a situation where we know exactly, for each sensor, on what pixel a certain position in the real world is mapped. That would allow us to basically overlay the data and have all individual sensors effectively function as one. Then we would know that data from all the different sensors would coincide even for the smallest mine. In reality, however, such a situation rarely occurs. The sensors are placed at a certain, albeit small, distance from each other and unless the scene under consideration is planar, the sensors will not observe exactly the same thing; what is perfectly visible from the position of one sensor, may be occluded in another sensor's view. Now, even if the sensors would view exactly the same part of the scene, irregularities and distortion in the image forming elements (lenses, detector array etc) add to the complexity of the problem. Obviously, the ideal sensor in this respect would be one where the same optical aperture is used for all sensor modes and where there is no, or at least known, distortion.

In MOMS, a couple of registration techniques have been considered and tested. One option is to use the laser data as reference data and project it onto the image planes of the respective sensors (Chan, 2006). In that way, since the laser gives us 3D information (x,y,z in meters), the problem with different perspectives due to the scene deviating from a plane is reduced. However, it requires a complete set of high-resolution, accurate 3D scene data, which is difficult to obtain in practice. This is due to a number of factors. First, laser points may be missing (*dropouts*) in regions where the laser beam hits mirror-like (e.g. wet) surfaces or surfaces with very low reflectance. Second, the range resolution of most laser radar systems is very uncertain in cluttered regions (e.g. bushes, grass, and sprigs) and around object edges. Another aspect is that registration requiring high-density and accurate 3D laser data for large areas may not be feasible in a practice for a reconnaissance system, due to the amount of data and calculations.

Another, more straightforward and realistic, approach to the registration problem is to apply standard 2D image transformations between images. This approach is quite suitable for planar scenes, an obvious shortcoming being that it cannot handle occlusion phenomena correctly. For most of the scenes studied within MOMS the planar approximation often made sense, as most of the upright objects, which causes significant occlusion problems, were reference targets, trees and some laboratory equipment. Whether those objects are perfectly overlaid in the different sensor images and hence whether or not they can be analyzed on a signal-level was not critical, as they are no interesting targets *per se*. The important thing to notice is that a *projective transform* was often good enough to produce significant overlap of the target objects (mines) between the sensors. The projective transform is defined as:

$$\begin{aligned} \hat{x} &= \frac{xh_1 + yh_2 + h_3}{xh_7 + yh_8 + h_9} \\ \hat{y} &= \frac{xh_4 + yh_5 + h_6}{xh_7 + yh_8 + h_9} \end{aligned},$$

where $[x, y]$ and $[\hat{x}, \hat{y}]$ are the original and new image coordinates, respectively, and the $h_i's$ are the transformation parameters, estimated in Matlab through optimization given a number of control point pairs. In this work, the control point pairs were selected manually to ensure that the points were placed in the right locations. See Figure 4 for an example of applying a projective transformation.

For most objects, the registration was adequate in the sense that after transformation, data from the different sensors overlapped significantly on most targets. A brief inspection of the registration result was often satisfactory to determine for which, if any, objects in the scene data fusion attempts would not be possible.

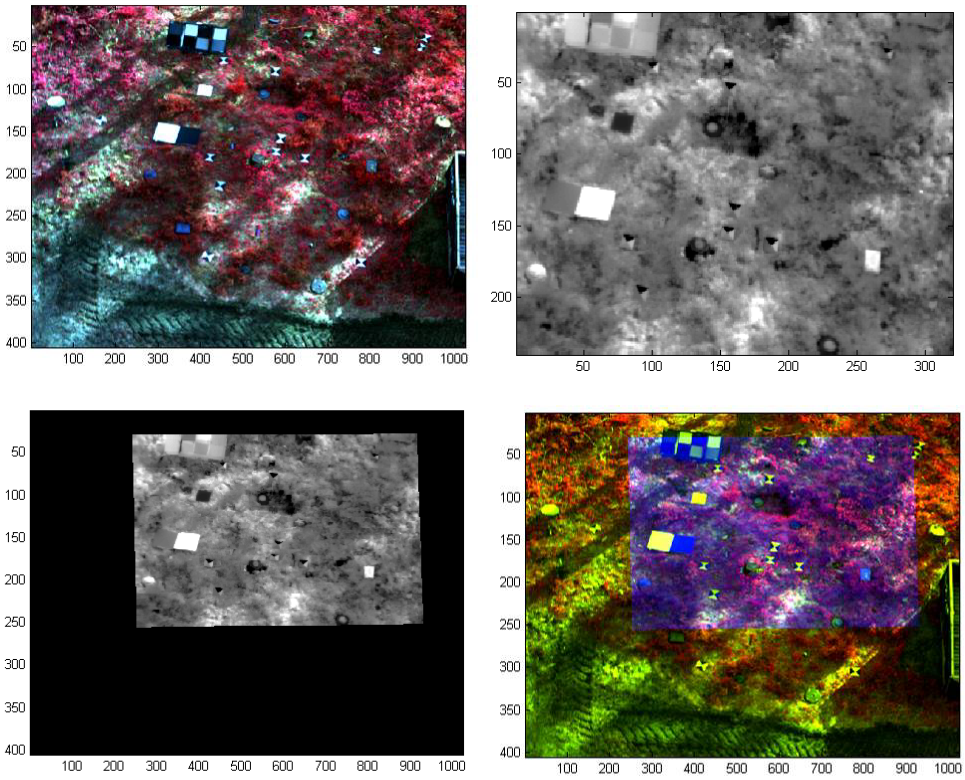


Figure 4. Top row: Original images, multi-spectral (left) and IR (right). Bottom row: Transformed IR image (left) and a pseudo-colored image showing the original and transformed images overlaid (right).

In MOMS, the sensors used for data collection were often kept fixed relative to each other and the range to the scene is measured with a laser, in order to facilitate the registration. In a system, however, with distributed sensors and/or asynchronous data collection, it would obviously be a bit problematic to rely on human involvement in the registration step. Methods for automatic control point detection and registration quality assessment will probably be needed. See (Svensson *et al.*, 2008) for a discussion.

6 Feature extraction

In this section we describe some techniques investigated within MOMS to extract certain, primarily spatial, features from the data. The idea is to process the raw sensor data to more robust features that are easier to use in detection and classification algorithms.

6.1 3D surfaces in laser radar data

In environments with lots of irregular structures (e.g. bushes, grass and sprigs) human-made objects, such as mines, are often among the most regular and smooth structures in the scene. Typically these objects tend to be more "surface-like" than the background. 3D data, e.g. acquired with ladar sensors, can be used to enhance these surfaces. Surface detection can be performed in many different ways. The technique presented in (Andersson and Tolt, 2007) was originally designed for detecting vehicles in forest environments but could in principle be used for detecting land-mines. The method is based on partitioning the point data set into a set of voxels in which the degree of surfaceness is found through Principal Components Analysis (PCA). This approach is quite fast, which is clearly advantageous for object detection in large areas. Another surface detection option is local surface-fitting (Westberg *et al.*, 2008), which is more accurate but considerably more computationally demanding. Therefore such a technique is likely to be more suited for detailed analysis of already identified suspicious regions (e.g. anomalies). See an example in Figure 5.

Common for all 3D surface detection techniques is that the performance is governed by the noise level relative to the size of the sought-after objects. If the noise is comparable to the size of the object (as is the case with some of the mines measured with the ILRIS sensor), 3D surface matching and shape detection will be more difficult.

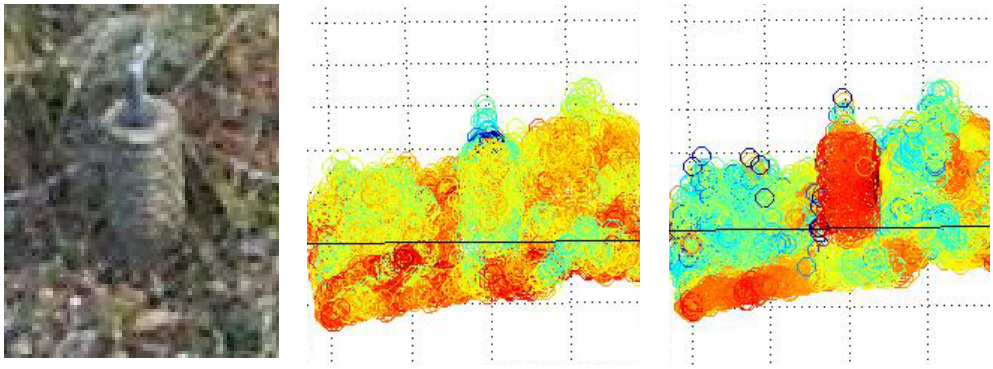


Figure 5. This mine is difficult to detect directly in the laser intensity domain (left), but by computing a surface score in 3D, it suddenly stands out clearly from the background (right).

6.2 Spatial features in IR images

Within MOMS a number of spatial operators aiming at capturing certain object characteristics have been tested on thermal IR images.

- The *curvature* operator responds strongly to circular and elliptical objects.
- The *convexity* operator is designed to find convex shaped objects. If the grayscale values in the IR image are viewed as range estimates, objects that are cooler around the edge appear convex.
- The *blob* detector finds segments of roughly uniform temperature.
- The *Laplace* operator detects regions of a certain size corresponding to local temperature extrema.

Some examples are shown in Figure 6. The principal advantage of using this kind of operators is that they allow for finding objects based on shape, rather than the magnitude of the difference in (apparent) temperature. On the other hand, a common problem with such operators is that they often fail to produce satisfactory results in complex environments. Many objects, typically smaller ones, are often very difficult to detect due to the fact that there are often natural variations in the background that are indistinguishable from those of mines. Refer to (Sjökvist *et al.*, 2005 and Letalick *et al.*, 2006) for more discussions and examples of this kind of operators.

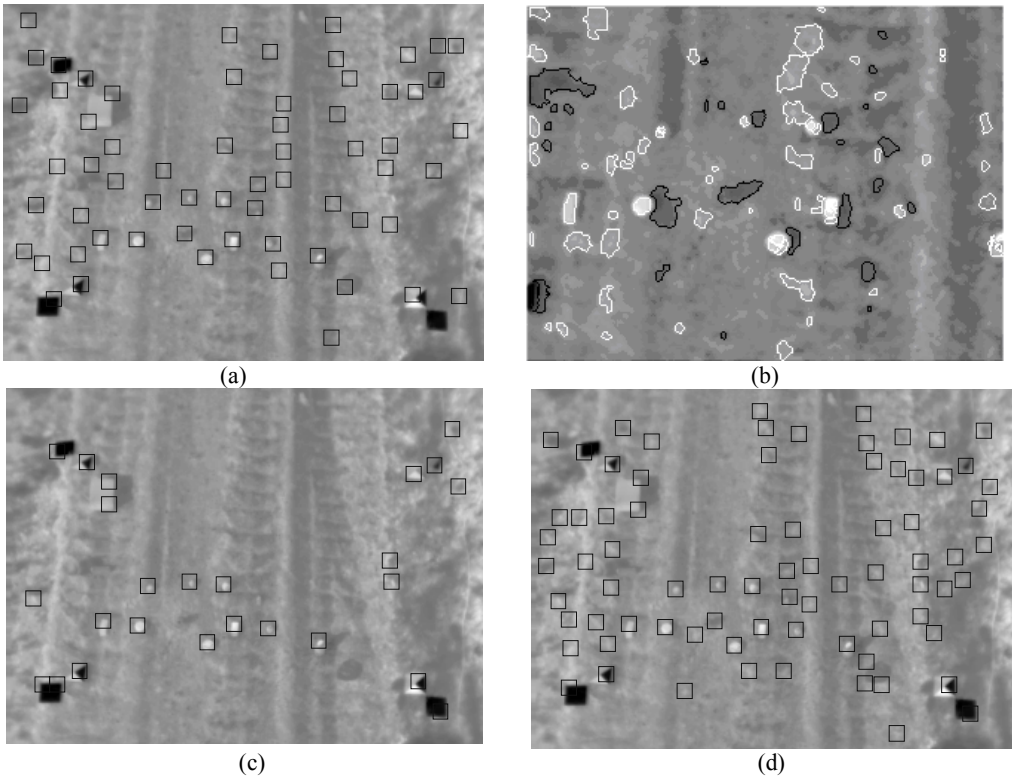


Figure 6. Examples of Thermal IR (LWIR) image analysis. The scene contains a gravel road on which a number of mines were put (the small bright blobs). (a) The result after applying the convexity operator. (b) A blob detector followed by segmentation and boundary tracing applied to the TIR image. “Hot” areas (white) represent the mines and other warm areas. Cold areas (black) are mostly shadows but may also represent other cold spots on the ground. Note that a shadow near a hot area also indicates the presence of a three-dimensional object, not only a cold area on the ground. (c) The result of applying a Laplace operator. (b) The result after having applied a curvature operator

Even in the quite uncluttered scene in Figure 6 (mines placed in the open on a gravel road), efficient detection with a low false alarm rate is difficult. In this example, the Laplace operator (that basically finds local temperature extremes) performs best but still misses a couple of objects even after careful parameter tuning. In more cluttered environments, the task is of course more difficult. Robust extraction of this kind of features requires that the contrast between the target and the background is significant compared to the natural variations. Thus, in order to be able to judge whether a particular feature is expected to be useful for helping discriminate between target and background, we must gather knowledge about the background. Figure 7 shows an example where IR images have been collected regularly (one image per minute) during several hours. It illustrates that some objects stand out from the background more or less regardless of the time of the day at which the image was acquired. This implies that applying a hot spot detector, such as a Laplace operator (of adequate size) would be very likely to pick out that target. It can generally be noticed that the possibility to detect other objects typically depends strongly on when the measurement was taken. Hence, such contrast statistics serve as a useful tool for estimating what performance could be obtained for a certain combination of target, scene and weather conditions. It should be pointed out that even if

the operators cannot be used for detection of suspicious-looking objects, they could still potentially be used for characterization of already detected objects.

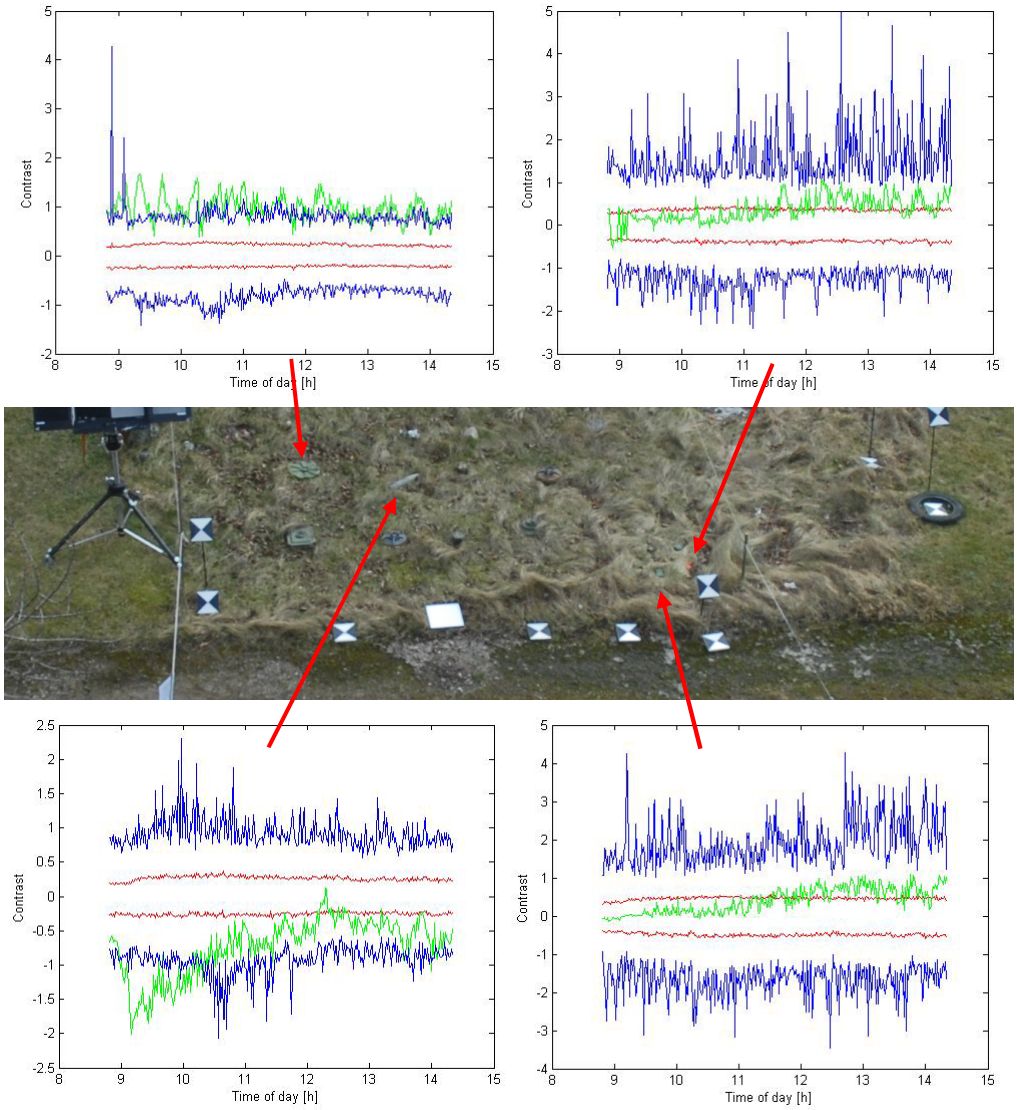


Figure 7. LWIR contrast measurements. The figure shows an example scene containing a number of mines and other objects. Images were taken with a LWIR camera once every minute, from about 9 am to about 2:30 pm. Each of the four plots shows the results for a particular object. Images were first manually segmented into object and background masks. The *green* curve shows the contrast between the object and the background pixels around it. The *red* curves correspond to mean contrast ±1 standard deviation in the background (computed with the same masks as the object) and thus depict the natural contrast variations in the background. The *blue* curves show the maximum and minimum contrast in the background region, respectively.

7 Anomaly detection

Following the assumptions discussed in Section 2.3, we expect mines to be different in some respect from the rest of the scene and also quite rare. The purpose of *anomaly detection* is to identify those objects (samples, pixels) that differ significantly from the background, without actually using *a priori* explicit knowledge about the signature or characteristics of the sought-after targets, other than it constitutes only a small portion of the dataset. The role of the anomaly detector is thus to identify "hot spots" on which subsequent analysis can be performed.

The background can be described by the probability distribution $p(y \mid x = 0)$ and the targets can be described by $p(y \mid x = 1)$. To be able to evaluate the system also $p(x = 0)$ or $p(x = 1)$ has to be known or assumed. Anomaly detection implies that $p(y \mid x = 0)$ is known while $p(y \mid x = 1)$ is unknown. The probability distribution $p(y \mid x = 0)$ describing the background is estimated from measurements or in some cases it may be known from other sources.

The decision limit has to be drawn based on this conditional probability distribution only. The system will be evaluated with some examples which hopefully are typical otherwise the limit has to be drawn based on the acceptable rate of false alarms.

7.1 Choosing background model

First, a background model has to be created. Depending on the amount of knowledge we have of the expected characteristics of the scene, some models are more appropriate than others. The models themselves can then be chosen to be *global* or *local*. A global model means that one model is created for the entire scene, whereas a local model means that different models are defined for different regions in the image, possibly creating a new model for every pixel under consideration. Both approaches have their own merits and disadvantages in terms of performance on different kinds of scenes, computational complexity, etc.

There are many models that can be used to describe the background. In this work we have considered Gaussian mixture model. The parameters to be decided are the number of components and the number of iterations. When modelling hyperspectral data which are multidimensional there are many parameters in the model which means that many pixels are needed to estimate the parameters. Thus we have used an entire image as background, excluding manually segmented targets, and other objects that do no belong in the scene. We have chosen to have 15 components and 25 iterations which is enough for parameters to converge. Thus it qualifies as a global model. A local model would only consider a small neighbourhood around the each pixel.

If the hyperspectral data has too many dimensions it seems that the model does not converge in reasonable number of iterations. Thus we have restricted the tests to 30 bands instead of the 240 available bands from our sensor. With only thirty bands the model converges in 25 iterations and the model describes the background well enough so that many of the targets appear as anomalies.

A possible problem is that if a target that is different from the background is in the training data it seems that the model adapt to include the target and thus the target will not appear as an anomaly as expected. However measurement of the mutual information between the classification of the image by the closest component of the Gaussian mixture of each pixel and the target mask reveal that there is information about the targets in the model. However at this point we have no way of dealing with this information. Thus we have restricted out tests to cases where the model is adapted only to background data without targets.

7.2 Anomaly detection using thermal history

Since the IR signature of the target relies on the thermal history, the results are very weather dependent, as discussed in Section 6.2, which makes it very difficult to detect mines using only a single image. However, if the *thermal history* is recorded, improved detection probabilities can be obtained. This type of temporal analysis requires that the scene is measured a number of times with a certain time in between (typically at least several minutes). An example of using a scene sequence is shown below. In Figure 8, a sequence of three recordings is pseudo-colour coded. Differences in colour are a manifestation of differences in thermal history. The 25 largest anomalies detected in Figure 8 are detected. These anomalies are clustered according to the thermal variations, shown in Figure 9 using colour coded labels.

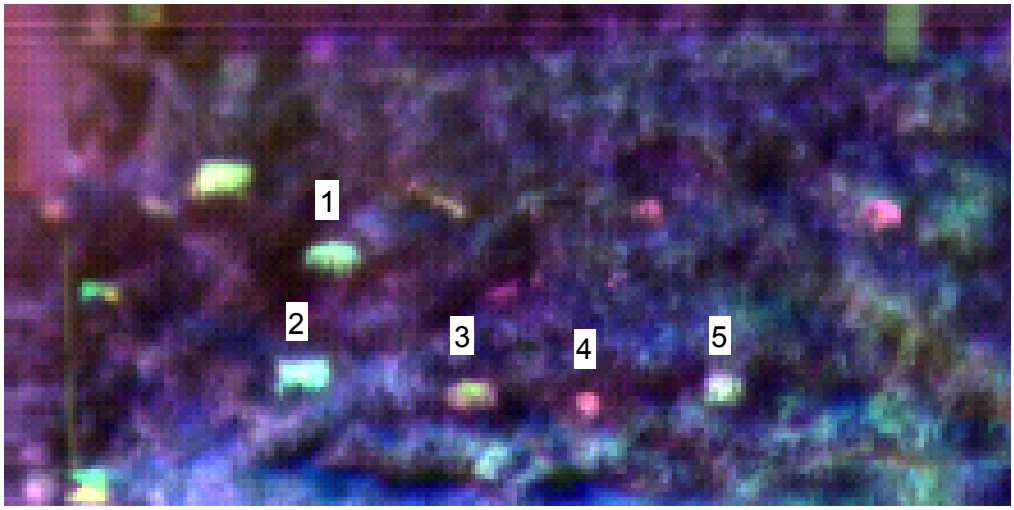


Figure 8. Three images from three different times are used in order to illustrate the thermal variation. For visualization purposes the images have been combined into a pseudo-colored RGB image.

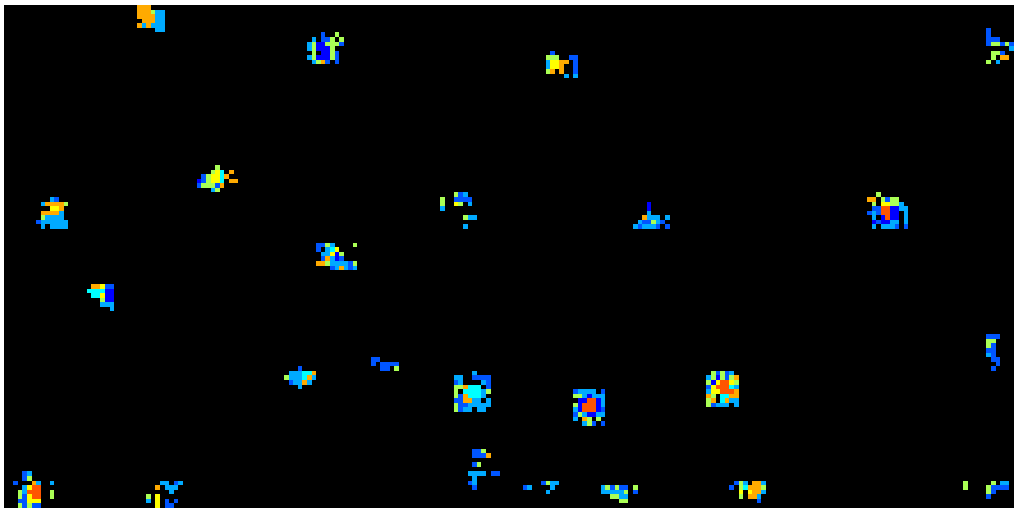


Figure 9. Detection of 25 objects in the thermal sequence illustrated in Figure 8. Detection results have subsequently been clustered to show spectral similarities between objects.

As the typical mine detection scenarios mainly considered within MOMS have not included repeated measurements over a particular region, this type of analysis has not been a major focus in this work. Instead, for results and discussions on temporal analysis for mine detection using IR data, refer to (Linderhed *et al.*, 2005).

8 Detection of mine-like objects

So far, we have discussed techniques for finding indications of there being some kind of unexpected object in the scene, e.g. by looking for spectral anomalies. In this section, we describe a number of methods for going a step further in the processing chain, to analyze each of the anomalies and to obtain detections of mine-like objects. Typically this involves using more knowledge or assumptions about the expected objects, e.g. size, shape and spectral properties, as well as about system properties, e.g. sensor resolution and viewing distance.

In the process of determining which of the anomaly pixels that can be grouped to mine-like objects, two different approaches are used. The first concerns spatial and spectral processing of the anomaly detections, to detect groups of co-located pixels that form an object. In the other approach, the anomaly detections are used to cue fusion-based analysis methods. Both feature- and decision-level fusion has been investigated.

8.1 Spatial post-processing of anomalies

The anomaly detection (Section 7) will return pixels that are anomalous. Some of the pixels are placed close to other anomalous pixels, while other is more or less solitary. In this analysis, all pixels are analyzed spatially; pixels that are close to other pixels are considered a pixel set. If the pixel set is of the correct dimensions, compared to objects that are we are looking for, it will be saved. Singular pixels and pixel sets that are too small or too large to possibly be a mine-like are disregarded. By this post-processing a lot of false detections can be removed.

8.2 Spectral post-processing of anomalies

The anomaly detection discussed in Section 7 results in a number of pixels that are anomalous, according to the background model and the particular distance measure used. However, the fact that a pixel is non-background obviously does not necessarily make it a target. In (Renhorn *et al.*, 2008) some approaches for spectral analysis of detected anomalies were presented and exemplified. It was demonstrated that the detected anomalies can be clustered into groups of similar spectral characteristics. In this section, this idea is discussed and developed further.

To illustrate the approach, consider the example shown in Figure 10. Here, the measurements of mines in backgrounds were performed under quite unfavourable conditions, with low light levels and substantial shadowing effects. A hyperspectral anomaly detection was first carried out. Then, by identifying anomalies having very similar spectral properties (spectrum distribution), groups of similar objects can be defined. This information gives an estimate of the frequency with which each object type appears in the terrain. Based on assumptions or *a priori* knowledge about the expected target density in the area, the chances of making the correct decisions are likely to increase. Anomalies that are considered irrelevant after such considerations can be fed back into the background model.

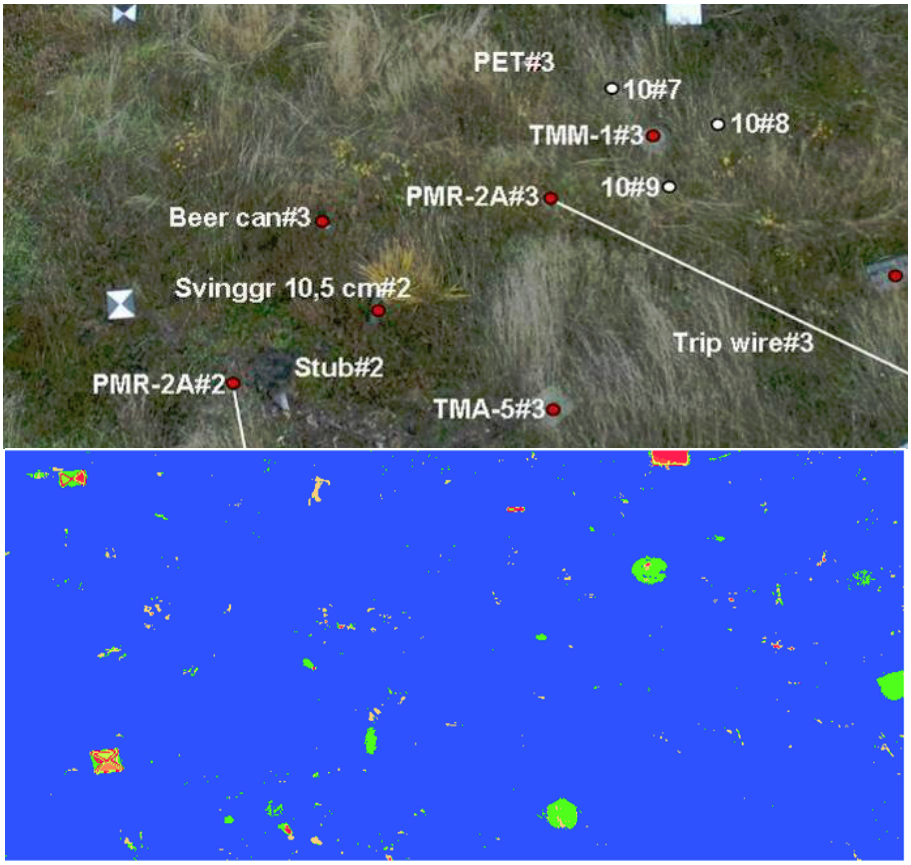


Figure 10. Top: An example scene. Bottom: Detected anomalies. The colour corresponds to the estimated spectral similarity between anomalies. Note that most of the mines are found to belong to the same cluster, i.e. have similar spectral characteristics ("colour"). The difference in perspective between the images is due to different positions and orientations of the visual and hyperspectral sensor during the measurements.

8.3 Feature-level fusion with the EM-MML algorithm

This technique has been developed at FOI (Westberg *et al.*, 2008). The purpose is to segment the sensor data into background and possible targets, i.e. mine-like objects. This is done by fusing data from the available sensors and then modelling the data as a number of Gaussian components, resulting in a so called Gaussian Mixture Model (GMM). Some GMM will describe the mine-like targets and some will describe the background. The method uses the EM algorithm to automatically find the optimal GMM with respect to the Minimum Message Length (MML) criterion. The purpose of the MML criterion is to prevent an over-fitting of the model to the available data. The method also involves a number of measures for comparing the segmentation results obtained with different feature sets, as well as for determining whether the extracted segments are mine-like objects (e.g. have adequate physical size).

The data consists of features extracted from different sensor data, and so far ladar and IR have been combined with quite promising results. The features are first ranked according a performance criterion using the various clustering measures mentioned above, and then added (in this order) to the EM-MML framework one by one, as long as the clustering quality keeps improving. The method has been tested on a dataset in which target objects have been manually segmented, and thus provides a way of judging the performance quantitatively. The initial results obtained with this method can be summarized as follows:

- Blind segmentation of data works rather well, not only in the sense that the quality is good when judged by the eye, but also when compared to ground truth.

- An efficient criterion for computer aided selection is needed; the most natural choice of true class could probably be based on analyzing segments with respect to number of samples, shape and size, e.g. using a spatial scatter measure.

The approach have been tested on 14 scenes, where 8 contained a mine, an IED or a ammunition box and 6 contained a rock, branches or just ground. All mines were segmented, but not always into one single segment and the false objects were not detected. An post-processing of the segments have to added, to get clear segments of mine-like objects.

The processing framework is iterative in that the solution gradually converges to the final state, and is thus quite computationally demanding. This makes it feasible primarily for relatively small datasets, e.g. as a tool for more in-depth analysis of already detected regions of interest (anomalies). Moreover, the method is designed to combine data from different sensors, and hence requires well-registered sensor data.

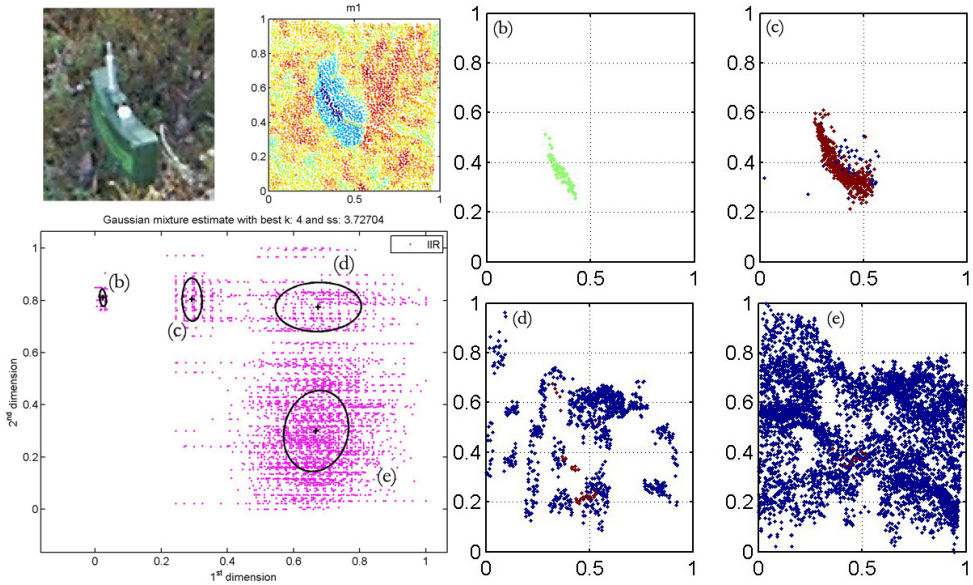


Figure 11. An illustration of the GMM data segmentation. The two images in the upper-left corner show the mine (visual camera data and ladar intensity, respectively). The plot below illustrates the clustering in the feature space that, in this example, resulted in four clusters. The ladar data points belonging to each of these clusters are shown in the plots to the right.

8.4 Feature-level fusion with an SVM classifier

In this section, we describe a technique for combining information from different sensors through supervised classification based on a Support Vector Machine (SVM). We start with a short summary of the SVM. More information on this subject can be found in (Cristianini and Shawe-Taylor, 2000).

The SVM is one among many types of classifiers that represents the classification library by a compact model (decision boundaries). A useful property of the SVM is that it maximizes the margin between classes, which leads to increased robustness against changes in the features. This, in turn, leads to a system with robust and generic properties. In practice, the SVM has means to adapt to the variations in data that are due to different appearance, such as aspect angle, slight changes in illumination, etc. This adaptation is facilitated by *kernel functions* that map input data (or features) to a new *feature space*. In this feature space, nonlinear decision boundaries are represented by simple hyper planes.

The optimal set of support vectors can take some time to compute, particularly for large libraries in high dimensions. Once the support vectors are computed, however, the classification is relatively easy to obtain. When the data is classified, also the classification

probability needed for efficient data fusion is estimated. These *class posterior probabilities* $P_p(X_i)$ are assumed available as an output of the classification program (SVM) – one per class, $p = 1, 2, \ldots, q$ (number of classes). A confident classification requires that one class posterior probability is significantly larger than all others.

8.5 Decision-level fusion

Decision-level fusion can be done in many ways, following a number of different frameworks, e.g. probabilistic (Bayesian approach), fuzzy or Dempster-Shafer fusion. Within MOMS, we have investigated a Bayesian approach. Starting with a set of class posterior probabilities $P(p|X_{ik})$ of having class p given data X_{ik} from object i and sensor k, we may combine this information in a number of ways, forming for example *weighted mean*, *median* or *product* over all sensors, then decide for the class with greatest resulting P. The weighted mean approach offers a way to express the degree to which we should trust the results for each sensor. The weights could then contain the estimated probability that a mine has been detected, as well as factors corresponding to the estimated reliability of each sensor under the particular circumstances. For example, referring to the IR contrast measurements in Section 6.2, *a priori* information or estimates concerning the expected background variations at a particular time of the day could be taken into account in this manner. Having many sensors, the median, resembling a voting rule, should be most robust against accidental sensor failures. The product rule, relating to Bayes' formula on conditional probability, is most sensitive to variations in participating probabilities and should therefore be constrained by adding a safe lower limit to the factors, like P+0.1.

Figure 12 shows the results of an initial test of the decision-level fusion concept, obtained with data from a scene containing 26 mines as test objects. For each of the four available sensors an SVM-based classification was performed, thus resulting in four individual classification outputs. A number of different ways of combining the respective outputs were tested (e.g. product and median operators) and the results, although very limited, indicate that decision-level fusion can improve the quality in the final detection result, compared to using the sensors individually. More work on decision-level fusion aspects is planned for the remainder of the MOMS project.

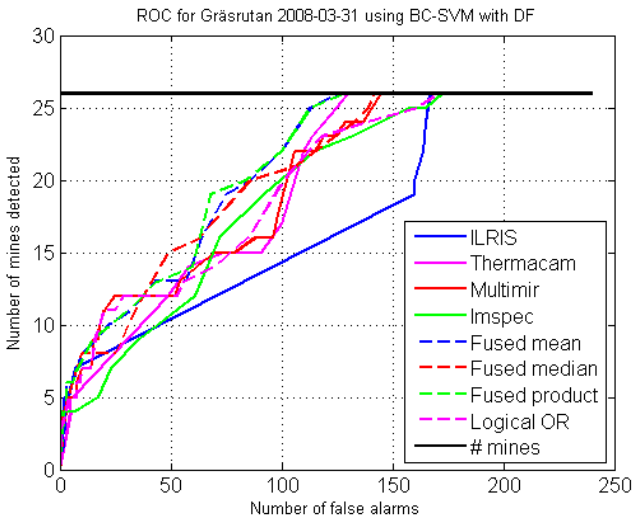


Figure 12. ROC curve for binary (mine) detection using decision-level fusion of results obtained through SVM-based classification of data from four different sensors. Here very limited attention should be paid to the actual values for the false alarm rate; they were computed using a ground truth mask that did not cover the entire objects. Hence detected pixels outside this mask were interpreted as false alarms, even though they were really on the target. The reason for this is that the masks were used for training the classifiers and thus were made small to assure that no background was included.

9 Mine recognition

Up to this point, we have mostly focused on techniques for judging whether there is probable mine present or not. Now we turn to the problem of analyzing each detected mine (-like object) in order to try to identify what kind of mine it actually is.

9.1 Model-based 3D recognition

The goal in model-based 3D recognition is to determine whether one (or several) of a given number of target models are present in the scene. Some approaches use *local shape descriptors*. Such approaches have an advantage in that only small parts of the object may suffice to perform a successful recognition. In cluttered environment this is indeed a desirable property. However, for noisy or sparse data such techniques may have problems as the robustness of the estimated shape features deteriorates. A popular technique is the so-called Spin Image approach (Johnson, 1997).

There are also techniques that use the entire shape of the object in the matching process. One advantage is that no local shape descriptors have to be computed, which makes them more suitable for noisy data. The Data-Aligned Rigidity-Constrained Exhaustive Search (DARCES) technique (Chen, 1999) is based on an exhaustive search among the possible positions and orientations of a particular object in the scene. In theory, the exhaustive search should make it possible to find a particular target even if parts of it are occluded. The search space is limited by imposing spatial constraints in the matching process but it still is a very time-consuming technique.

In MOMS it has been found that reliable land-mine matching based on techniques like the ones mentioned above will need 3D measurements with accuracy beyond what is possible to obtain today with most operational systems. Instead, a 3D target recognition technique developed at FOI and originally intended for vehicle recognition (Grönwall *et al.*, 2006) was considered the most successful. It involves estimating the size and orientation of a segmented object through simple geometrical assumptions and partitioning of the object into geometrical primitives (rectangles). The size and orientation estimates are then used to initialize a Least-Squares fitting procedure with a CAD model. This technique is different in nature from those discussed above in that it requires that the scene is segmented into object and background before the matching can be performed. While target/background segmentation directly in noisy 3D data from cluttered scenes is quite error-prone, we should use the fact that there are other sensors that can be used to detect possible targets, i.e., provide the first steps of target/background segmentation. This means that 3D-matching could then be applied to those segments. Figure 13 and Figure 14 illustrate the techniques. The first step is to estimate the dimensions of the object, Figure 13, by fitting a rectangle to the object data. The dimension estimates are used to select library models of corresponding dimensions. This means that only library models of relevant size are selected for CAD model matching, this reduces the number of matches that needs to be performed. In the second step the object data are fitted to the CAD model by iterative Least-Squares fitting. Examples of matching results are shown in Figure 14.

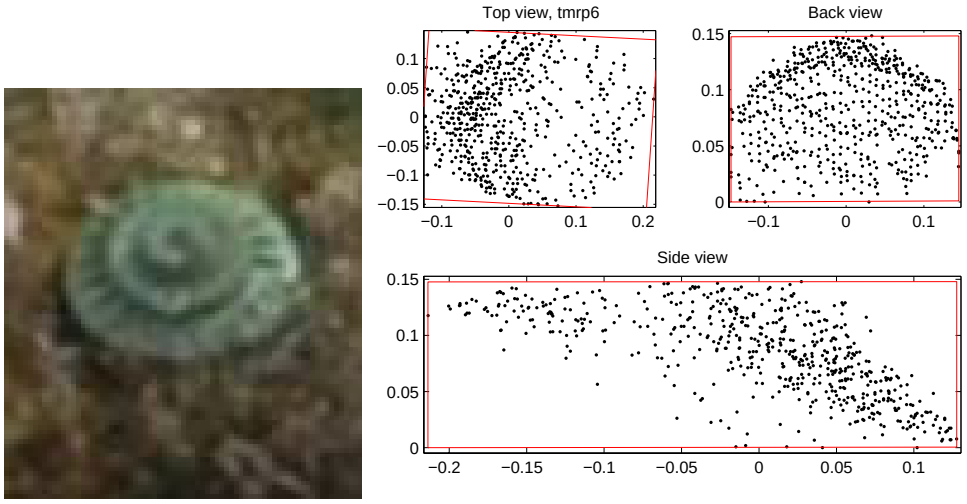


Figure 13. Image of a TMRP6 mine (left) and result of dimension estimation of the detected mine (right). Axes in meters.

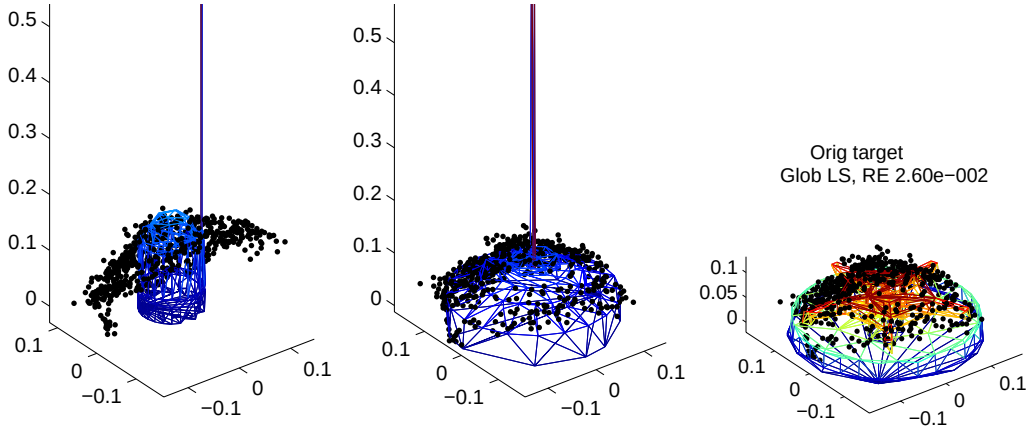


Figure 14. Example of match with CAD models of AT-2 (left), TMRP-6 (middle) and AT-47b (right). Axes in meters.

9.2 LBP

The LBP (Local Binary Pattern) operator was introduced in (Ojala, 1996) as a technique for capturing local shape in 2D images. The local binary pattern at a certain pixel is obtained by comparing the intensity of this pixel with those of its neighbours. Representing every pair-wise comparison between a certain pixel and a neighbour with one bit means that we can represent the shape within a 3×3 pixel neighbourhood with eight bits (eight pair-wise comparisons), which then can be stored as a 8-bit integer value (0-255). See Figure 15 for an example.

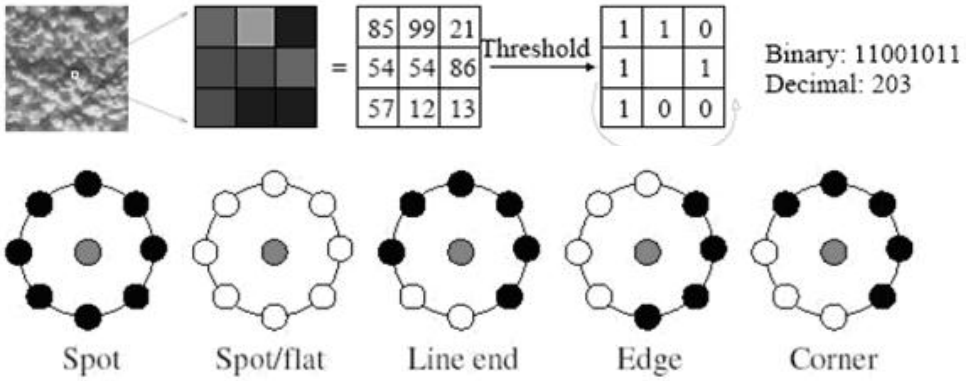


Figure 15. Top row: Example of an LBP calculation. Bottom row: Example of texture primitives detected by LBP, (white circles represent ones and black circles zeros).

By computing the LBP at a large number of (possibly all) pixels corresponding to the object under study, we can create a histogram for all those 8-bit LBP integers with each bin in the histogram corresponding to a particular spatial pattern, a *micro-pattern*. Hence, the shape of the histogram (i.e., the number of times each micro-pattern occurs) carries information about the texture of the object. Recognition of a certain object is then typically performed in a nearest-neighbour fashion by finding the stored prototype LBP representation that is most similar to the object under study. In order to increase the robustness against occlusion, noise, etc, the object is typically divided into smaller pieces, each with its own LPB histogram. Since its introduction, the LPB concept has been extended, for example to consider different neighbourhood sizes (Ojala *et al*., 2002). An example is shown in Figure 16.

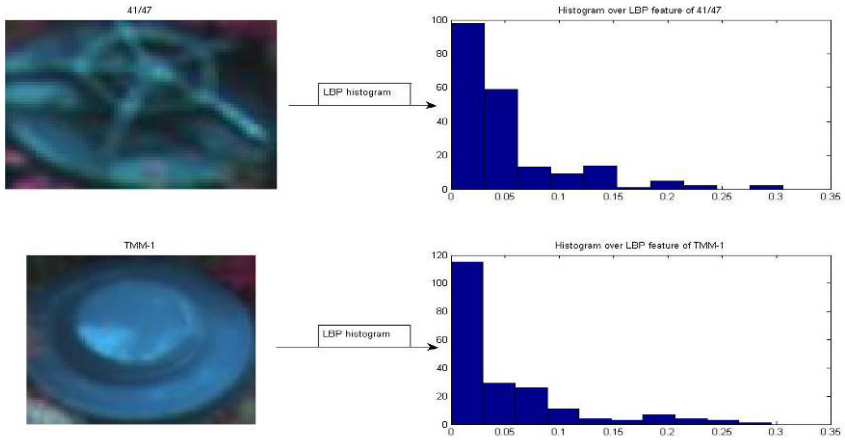


Figure 16. Histogram of LBP features. Top-left: Landmine"41/47", Top-right: histogram over LBP features of "41/47". Bottom-left: Landmine"TMM-1", Bottom-right: histogram over LBP features of "TMM-1".

9.3 SIFT

The Scale-Invariant Feature Transform (SIFT) was introduced in (Lowe 1999). The idea behind SIFT is to describe local shape with localized image gradient histograms. Since SIFT is only applied locally, it is inherently designed for robustness against occlusion, which is indeed a relevant aspect when detecting mines in cluttered environments.

First, a set of *key points* is detected by smoothing the image on different scales and finding local maxima in the derivative of the image with respect to scale and position. The gradients within a number of regions within the neighbourhood of each key point are computed and stored as a *gradient histogram*. This captures the intensity variations (the

pattern) at the key point. At each key point, the dominant direction is estimated, that allows for normalizing the features according to orientation, thus producing orientation invariance. A normalization of the magnitude of the histogram bins also provides intensity invariance.

Again, retrieving the correct match from a database of prototype objects is performed through nearest-neighbour matching, where the object under study is tested to each of the objects in the database. In this comparison, pair-wise correspondences between SIFT features are first established. Since some of these correspondence pairs may be wrong, only those matching pairs that display *geometrical consistency* are kept and used for identifying the object.

9.4 Shape contexts

Assume that we have a set of points located on prominent features of the object, e.g., edges, corners or other types of interest points. The *shape context* (Belongie *et al*., 2002) at a particular point captures the distribution of the remaining points relative to it. The rationale is that two similar shapes will have similar shape contexts. By solving an optimal point-to-point assignment problem, a transformation is estimated that aligns the two similar shapes. A dissimilarity measure between the shapes is then computed. As with LBP above, recognition is performed by finding the stored prototype object whose descriptor is most similar to the object under study (nearest neighbour).

9.5 Data fusion for mine recognition

Analogous to the mine detection problem we can combine different sources of information in order to increase the chance of recognizing the mine type.

9.5.1 Feature-level fusion with a SVM classifier

The SVM-based mine *recognition* applied within MOMS is very similar to the SVM-based mine detection discussed in Section 8.4. The difference is that before, we had a *two*-class problem (mine/not mine) whereas we now have a *multi*-class problem, where each class corresponds to a particular type of mine. Where one hyper plane was enough for the two-class problem, a whole set of hyper planes is now needed in the multi-class case, for dividing the feature space into, say, *r* regions. In practice, this can be done by using a single hyper plane for every distinct class pair, together with a strategy to combine them into a classification rule. The class library is thus represented by *r(r-1)/2* support vectors of the same dimension as the data plus equally many scalar bias terms. Alternatively, a number of classifiers can be constructed, each of which separates one particular class from the union of all other classes.

10 Performance assessment

In this section we briefly describe the data used for performance assessment and the sensors used for obtaining the data. Moreover we give a short explanation of two common ways of presenting the performance: the Receiver Operator Characteristics (ROC) and the confusion matrix.

10.1 Sensor data for assessment

In order to acquire real sensor data of mines in different terrain types, a number of experiments have been carried out within MOMS. Data have been collected for a number of scenarios, during different seasons, at different times of the day and using different sensors to obtain an adequate data set for developing and evaluating mine detection methods. The results presented in this report mainly stem from four environments

1. Forest
2. Road and road embankment
3. Clear-cut area of forest
4. Grass field

The data acquisition experiments for the first three environments were all set up in the facilities of the Swedish Demining Center (SWEDEC) and were carried out during different seasons. The grass field experiments were carried out in a temporary test site at FOI of about 15×15 m^2 in which several dummy mines and clutter objects were deployed. From the available data, a number of assessment data sets were defined, see Table 3 below.

Table 3. Short description of the datasets used for evaluation.

Dataset	Season	Time of day	Type of scene	Note
A	Apr	Morning	Sprigs, forest	
B	Apr	Afternoon	Sprigs, forest	Same scene as A, but changed view
C	Apr	Noon	Clear-cut forest	
D	Apr	Afternoon	Clear-cut forest	Same scene as C, but changed view
E	Apr	Noon	Road and embankment	
F	Oct	Afternoon	Clear-cut forest	
G	Oct	Afternoon	Road and embankment	
H	Mar	Morning	Grass field	
I	Mar	Noon	Grass field	

10.1.1 Ground truth – object masks and annotation

In order to enable quantitative performance analysis, object masks were created for the sensor images in the datasets. Wherever possible, the masks marked the visible portion of each object in the scene. In some cases, the exact location of some objects could not be determined – they were simply not discernable – which then only allowed for marking their approximate location. Instead of being simple binary masks, each blob (corresponding to an object) was assigned a unique number so that particular objects could be easily identified and extracted.

10.2 Sensors used for the data acquisition

This section contains a short description of each of the sensors used for collecting the data. More information, e.g. technical specifications, for the sensors can be found in (Letalick *et al.*, 2007 and Larsson *et al.*, 2007). The sensors are:

1. **Thermal camera (LWIR) – SC3000**
   The ThermaCam3000 (SC3000) Quantum Well Infrared Photo detector (QWIP) system is a Long-wave IR (LWIR) sensor. The SC3000 operates in the 8-9 μm wavelength band and contains a FPA of quantum well type with 320×240 pixel resolution.
2. **Multi-spectral camera (VIS-NIR) – Redlake**
   Redlake MS3100 is a multi-spectral high resolution 3-chip digital camera. The camera is based on a colour separating prism with different coatings and filters. Three CCD sensors are used to acquire images with three spectral bands: 525-575 nm (green), 640-690 nm (red) and 770-830 nm (NIR). The spatial resolution of the sensor is 1392×1040 pixels.
3. **Multi-spectral camera (SWIR-MWIR) – MultiMIR**
   The Multimir is a multi-spectral sensor using a spinning filter wheel containing four optical band pass filters. Two of the transmission bands are in the short-wave IR (SWIR) domain (1.5-1.8 μm and 2.1-2.5 μm), while the other two reside in the mid-wave IR (MWIR) domain (3.5-4.1 μm and 4.5-5.2 μm). The spatial resolution of the sensor is 384×288 pixels.
4. **Hyper-spectral camera (VIS-NIR) – Imspec**
   The Imspec sensor is a hyperspectral camera for visible to NIR (396-961 nm) with 240 spectral bands. The spatial resolution is 1024×1024 pixels. The system consists of three main parts; a CCD camera (for detection), Imspector (the dispersive element) and a scanning mirror for forming images from several line scans.
5. **Imaging 3D laser radar (SWIR) – Optech ILRIS-3D**
   The ILRIS is a laser scanner operating in the Short-wave IR (SWIR) domain at 1541 nm. Although the minimum spot step of the mirrors is quite low (26.6 μrad) the effective resolution is limited by the laser footprint (about 1.3 cm at distance of 30 m).
6. **Nikon D200 (VIS)**
   The Nikon D200 is a Digital SLR camera hosting a CCD detector of 10.2 Mpixels. The D200 was mostly used as a reference sensor, to enable masking and object annotation of the sensor data.

10.3 Receiver Operating Characteristics (ROC)

Assume that a particular algorithm, say, a mine detector, produces a value representing the degree to which a particular object is believed to be a mine. In order to produce a crisp (binary) decision about whether this object should, or should not, be classified as a mine, this value has to be thresholded. If the value is above the threshold, the decision is, say, positive ("mine detected") and if the value falls below the threshold, the decision is negative ("no mine detected"). A particular choice of threshold will result in probabilities of making a correct and erroneous decision, respectively. By varying the detection threshold and measuring how the detection and false alarm rates vary, we capture the so called Receiver Operating Characteristics (ROC) of the detector. In other words, the ROC illustrates the trade-off between correct decisions and erroneous decisions that a particular threshold implies.

What we would like to see is a distinct separation between the two classes (mine/not mine), so that the system would be able to make correct decisions without a very careful choice of threshold value. In practice, however, the problem is often more complicated. This is illustrated in Figure 17 below. The success with which the target can be discriminated depends on the separation between the two distributions as well as the size of the variance. The ROC curve corresponding to this detection example is shown in Figure 18. Refer to (Renhorn *et al.*, 2008) for more theory and discussion about ROC analysis.

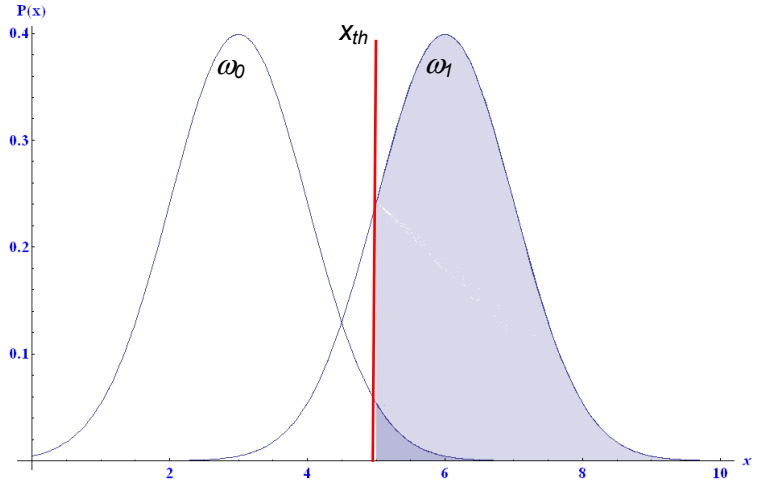


Figure 17. Probability distribution for a normal distributed signal without and with the target present. The blue shaded area represents the area above the threshold.

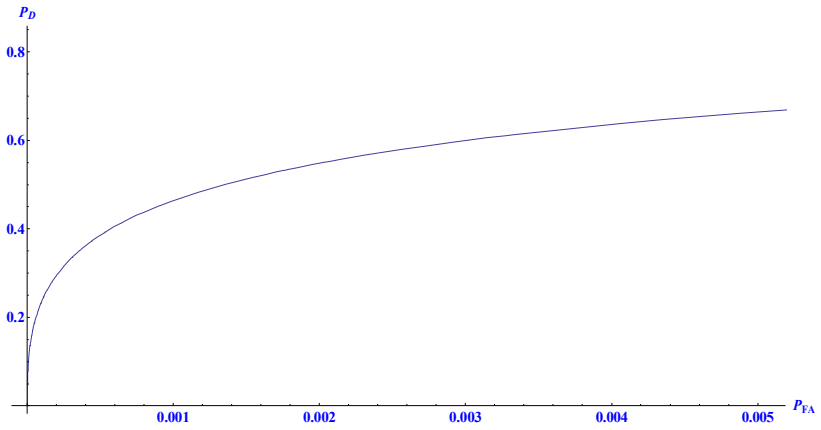


Figure 18. Receiver operating characteristic curve (ROC) for the detection example in Figure 17.

10.4 Confusion matrix

In the case of object recognition, the goal is to recognize which class, if any, a particular object belongs to. The results from such a study is often represented in terms of a confusion matrix that shows into what class (if any) a particular object has been classified. In the confusion matrix we define the recognition results as true-positive (TP), true-negative (TN), false-positive (FP) and false-negative (FN) recognitions, see Table 4. In fact, the TP (True-Positive) cell contains the amount of mines in test data which are correctly recognized; or FP (False-Positive) is the amount of non-mines which are classified as mines. In an ideal situation, all results should be determined to be true-positive or true-negative recognitions.

Table 4: The relationships between predicted target class and true class. This is a confusion matrix for target classification.

	Target Class	
True Class	Yes	No
Yes	True-Positive(TP)	False-Negative(FN)
No	False-Positive(FP)	True-Negative(TN)

If the recognition algorithm can determine the mine type, the true positive results can be studied further. In Table 5 the confusion matrix for a part of Test case 4 is presented. The training set consisted of 6 mines from dataset C and 9 unknown objects from dataset D were used for testing. This confusion matrix shows how good the algorithm can discriminate land mines that it has seen before.

Table 5. Example of a part of a confusion matrix for target type recognition. Four recognition algorithms are tested (LBP, SC, SIFT and SIFT*).

	Landmine classification											
Mines in Scene	**TMM-1**				**PMR-2A**				**Grenade**			
	LBP	SC	SIFT	SIFT*	LBP	SC	SIFT	SIFT*	LBP	SC	SIFT	SIFT*
TMM-1	2	1	1	2								
PMR-2A					1	1	1	1				
Grenade									0	0	0	0

11 Results

In this section we present signal processing results obtained with a number of selected methods aiming at solving a particular signal processing task, according to the framework in Section 3. The considered algorithms are the following:

- Optimal sensor design
- Anomaly detection
- Detection of mine-like objects
- Mine recognition

11.1 Optimal sensor configuration

In this section we show an example of how the mutual information can be used to find the spectral bands that convey most information about the presence or absence of a target. There are some bands that have very small mutual information and thus convey little information about the presence or absence of a target and other bands with more mutual information.

Figure 19 shows the scene used in this example. Figure 20 shows the manually segmented target and background mask. The spectral values from the sensor are 16 bit integers but only a smaller range of values actually occur in image in this example, see the histogram in Figure 21.

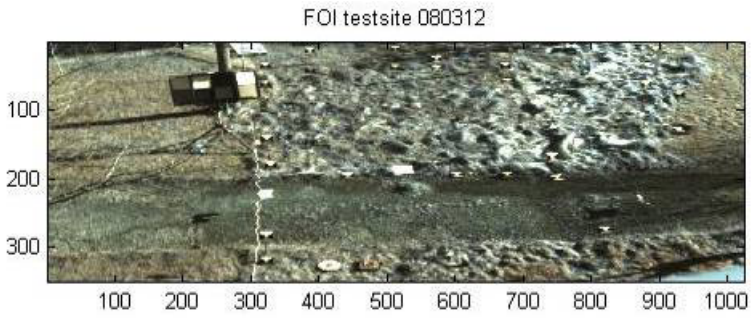


Figure 19. An image of the scene from which hyperspectral is collected for this example.

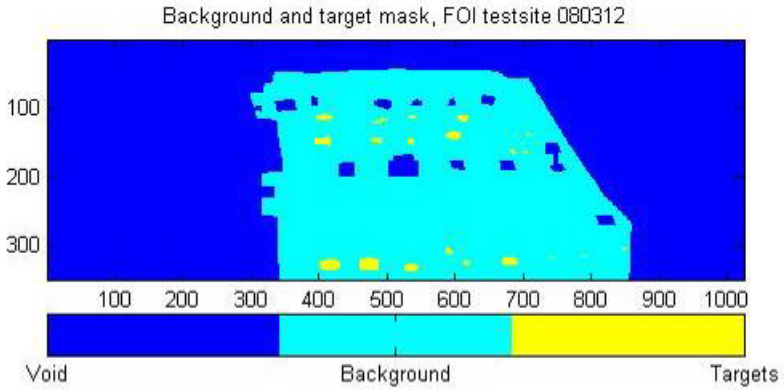


Figure 20. The target/background of the current example.

Figure 22 shows the mutual information between each of the thirty bands and the target mask. There are a few bands (5-8 and 27, 28) which convey significantly more information about the presence of a target that most other bands.

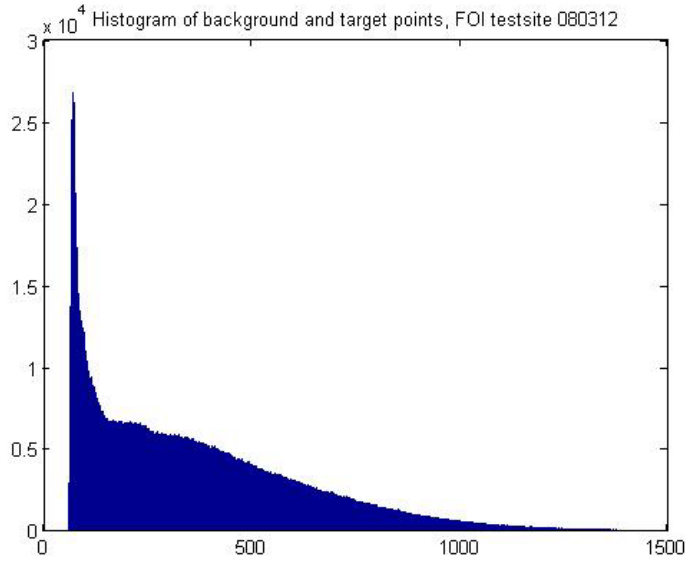


Figure 21. A histogram of all the spectral values in the image. The values are integers.

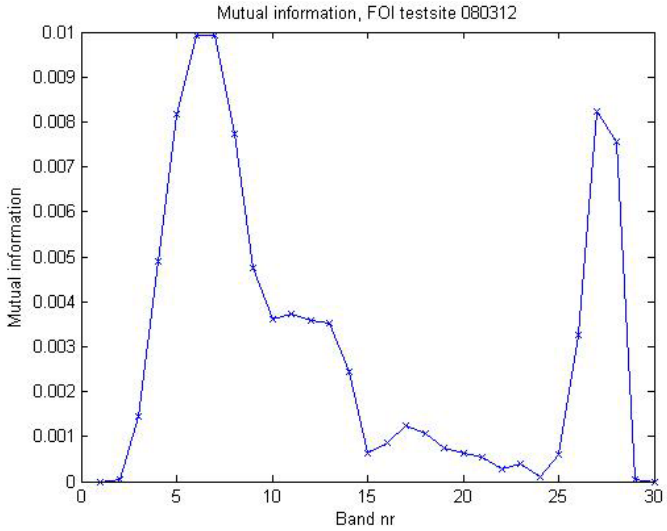


Figure 22. The mutual information between each of the thirty bands and the target mask. There are some bands with considerable higher mutual information then than other bands.

We will also show how the mutual information can be used to find the quantisation level where the quantised data will have the largest mutual information with the target mask. To be able to estimate the needed probability distributions the set of values for the stochastic variables must not be too large. Hence we only consider two quantisation levels in this example. It is possible to consider several levels simultaneously. Figure 23 shows the mutual information between the quantised sensor data and the target mask. In this example there is a range of quantisation levels (200-400) that give roughly the same mutual information. Figure 24 shows the values of spectral band six quantised to convey maximum information about the targets. The target mask is overlaid on the quantised sensor data. The image shows that even though the band conveys some information, it is not enough to discriminate the targets from the background by it self. Figure 25 shows the image for band 24.

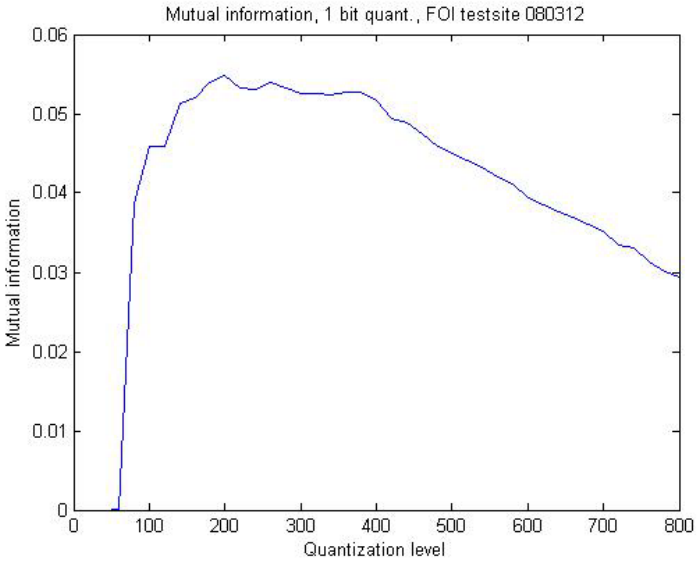


Figure 23. The mutual information between the target mask and the pixels where each spectral value has been quantised into two levels. Every band has the same quantisation level. There is a wide span of quantisation levels that give roughly that same mutual information.

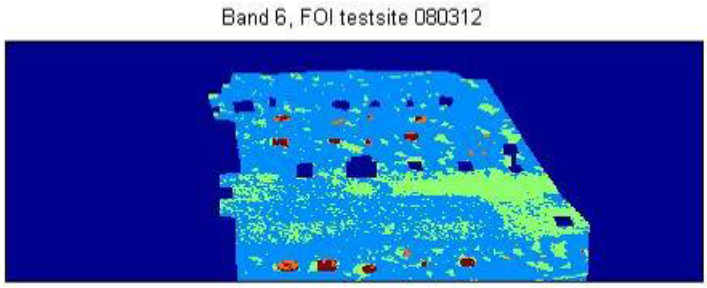


Figure 24. The data in band six with binary quantisation and the void/background/target classification overlaid.

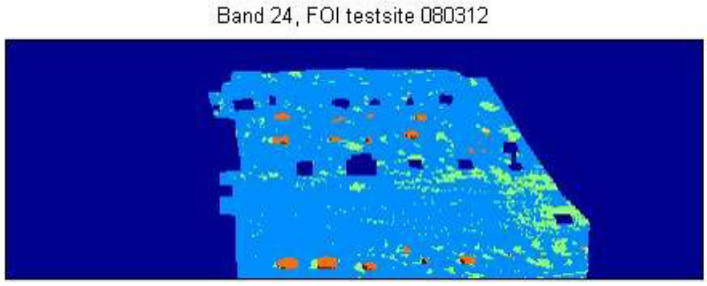


Figure 25. The data in band 24 with binary quantisation and the void/background/target classification overlaid.

11.2 Anomaly detection

In this section we show three examples of anomaly detection using hyperspectral data. For each of the images a Gaussian mixture model with 15 components has been adapted to the image. Then we have determined the decision limit on the anomaly values giving maximum mutual information. The optimal decision limit is compared with a limit giving a small percentage of the most unlikely values. In these three examples these two limits are similar. Thus anomaly detection based only on the background data without any specific knowledge about the targets used in these examples give about the same amount of information as a decision where the limit has been optimised for the specific targets used in these examples.

For all three scenes in this section there are some targets that are clearly anomalous with respect to the Gaussian mixture model and other targets that can not be discriminated from the background without too many false alarms. However the ROC curve showing the first detected pixel on each target shows that all targets in these examples have at least one anomalous pixel.

The examples are datasets A, D and H described in Section 10.

Dataset A

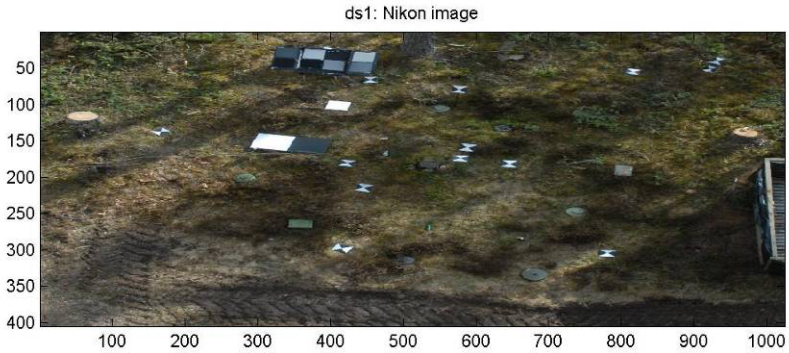


Figure 26. Dataset A, Nikon image.

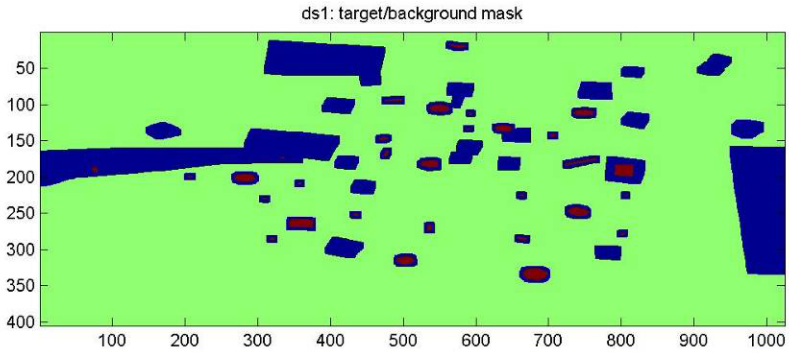


Figure 27. Dataset A, Target/Background mask. Green – background, red targets and blue – void.

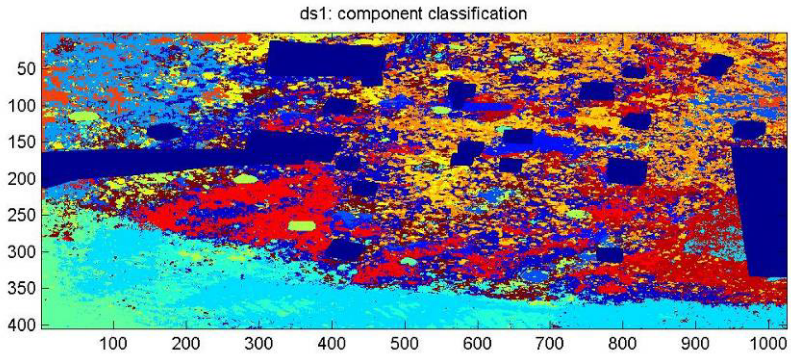


Figure 28. Dataset A. The Imspec image classified into the fifteen components of a Gaussian mixture model.

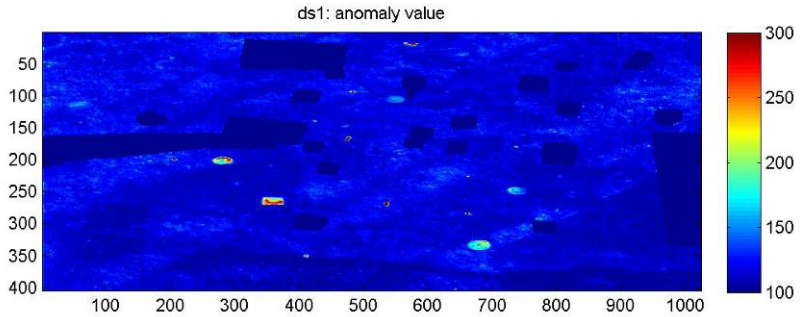


Figure 29. Dataset A The anomaly values obtained by comparison with a Gaussian mixture model.

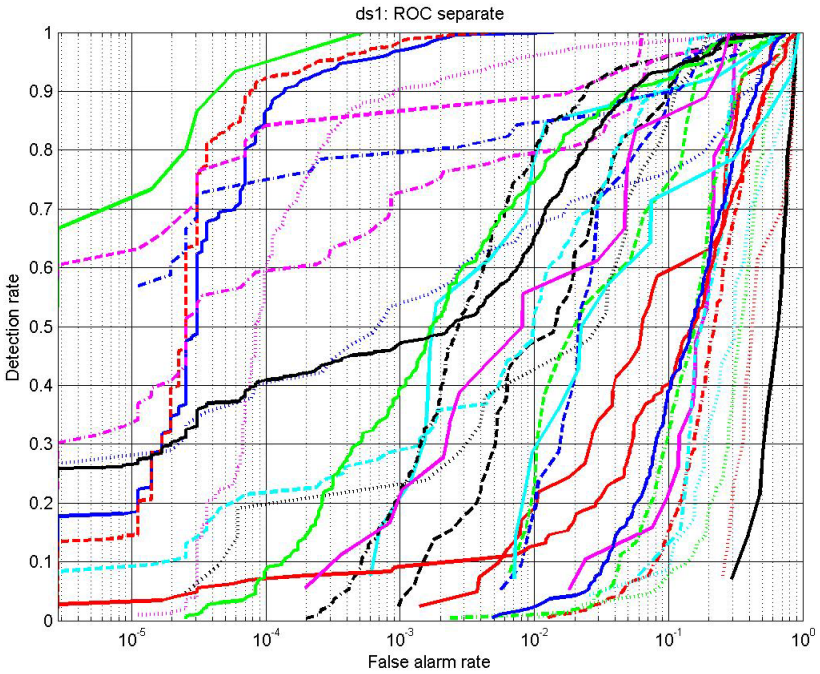


Figure 30. Dataset A. The ROC curve for each target separately. There are some targets that partly can be detected without false alarm while some other targets cannot be detected without many false alarms.

The ROC curve shows that some of the objects can be detected without any false alarms while most will have at least some false alarms and some objects have many false alarms.

Figure 32 shows that a decision limit around 146 gives maximum mutual information. There is also an anomaly limit around the same limit. There are only few pixels with larger distance than this limit to the background.

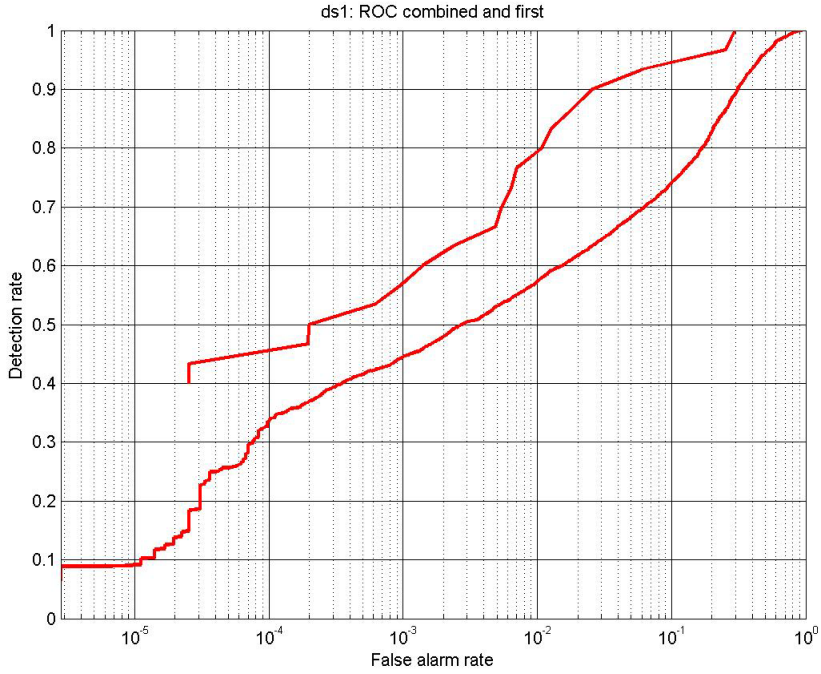


Figure 31. Dataset A. ROC curve for first pixel on each target (upper curve) and for all targets combined (lower curve)

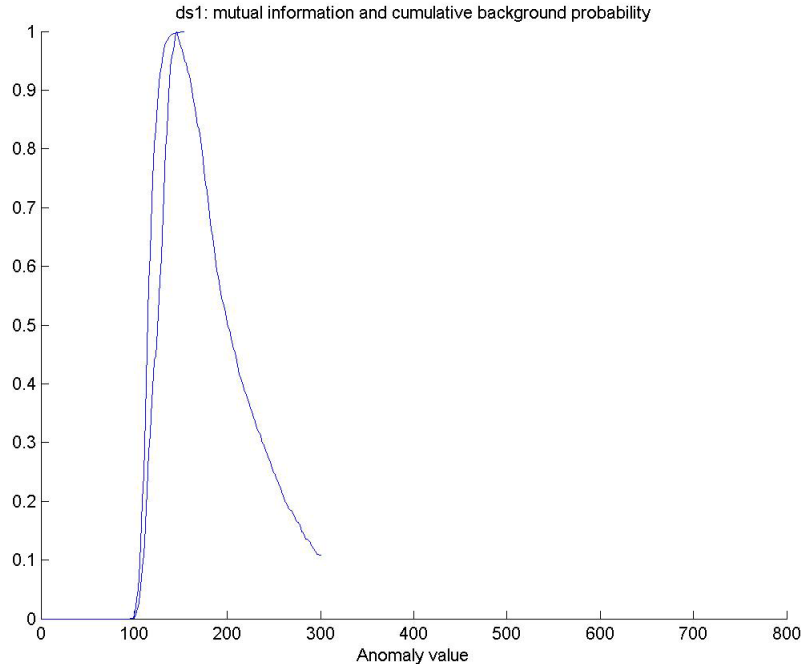


Figure 32. Dataset A Comparing how the mutual information (rightmost curve) varies with anomaly value and how the anomaly probability (middle curve) varies with the anomaly value.

Dataset D with the same figures.

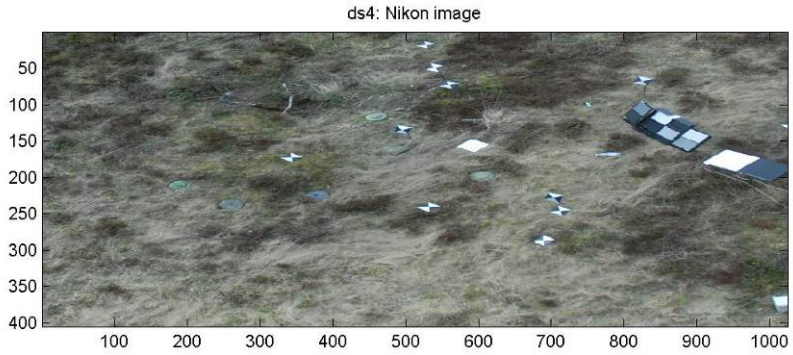


Figure 33. Dataset D, Nikon image.

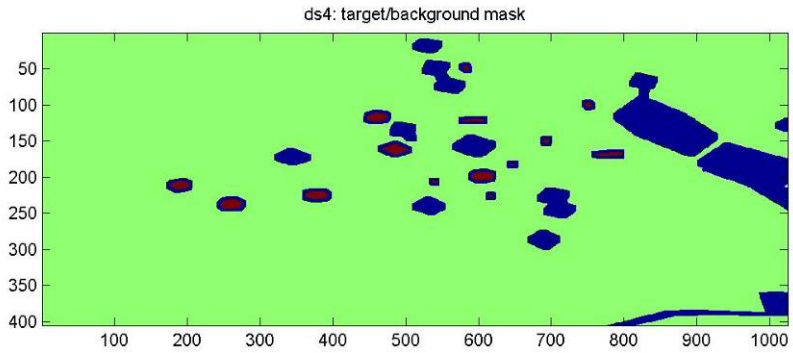


Figure 34. Dataset D, Target/Background mask. Green – background, red –targets and blue – void.

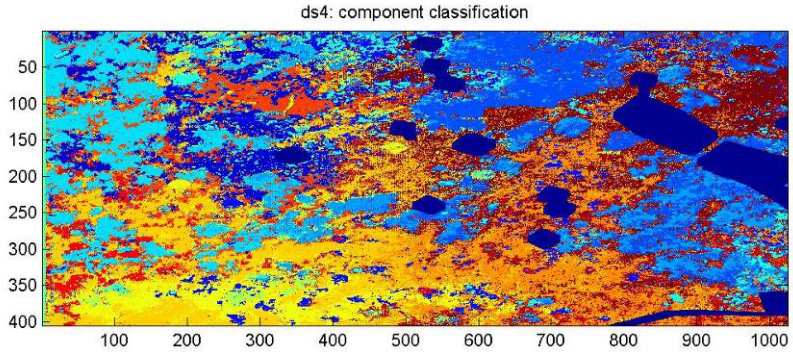


Figure 35. Dataset D. The classification of the Imspec image into the fifteen classes of the Gaussian mixture model.

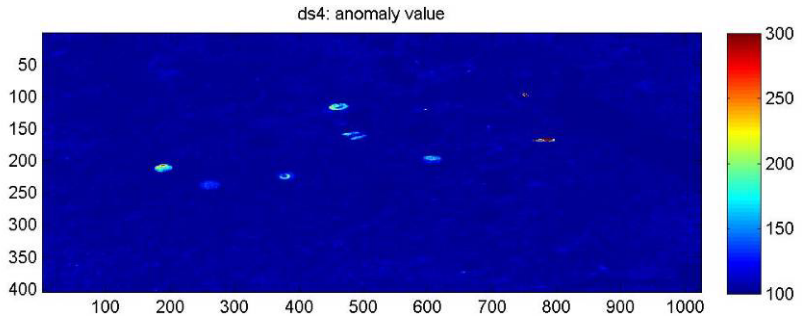


Figure 36. Data set D, anomaly values.

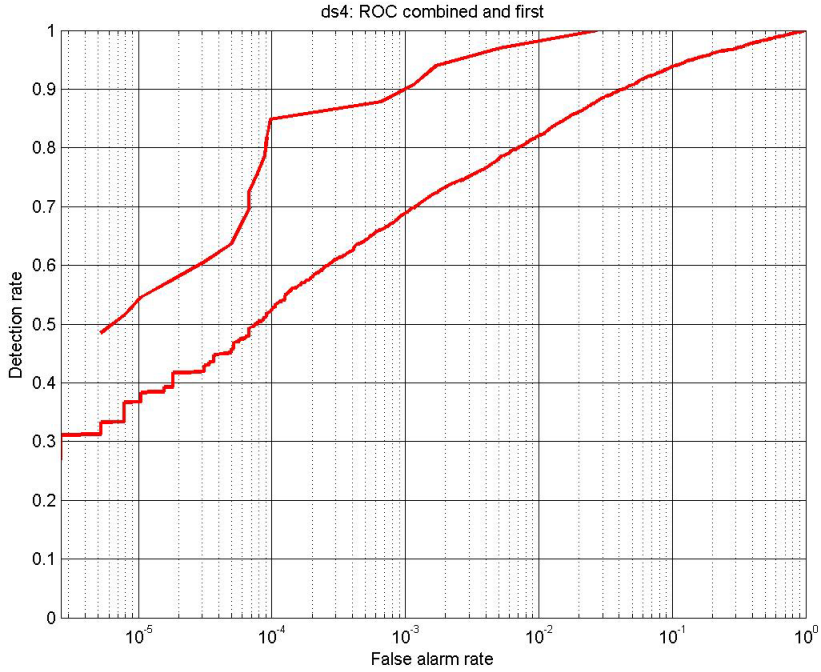


Figure 37. Data set D, ROC curve with first pixel on each target (upper curve) and all target pixels combined (lower curve).

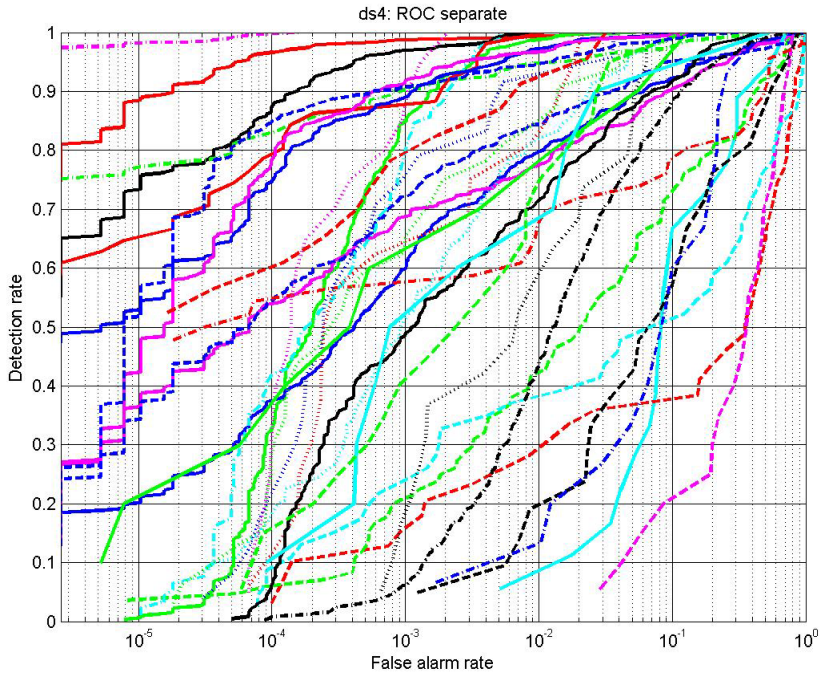


Figure 38. Dataset D, The ROC curve for each object separately.

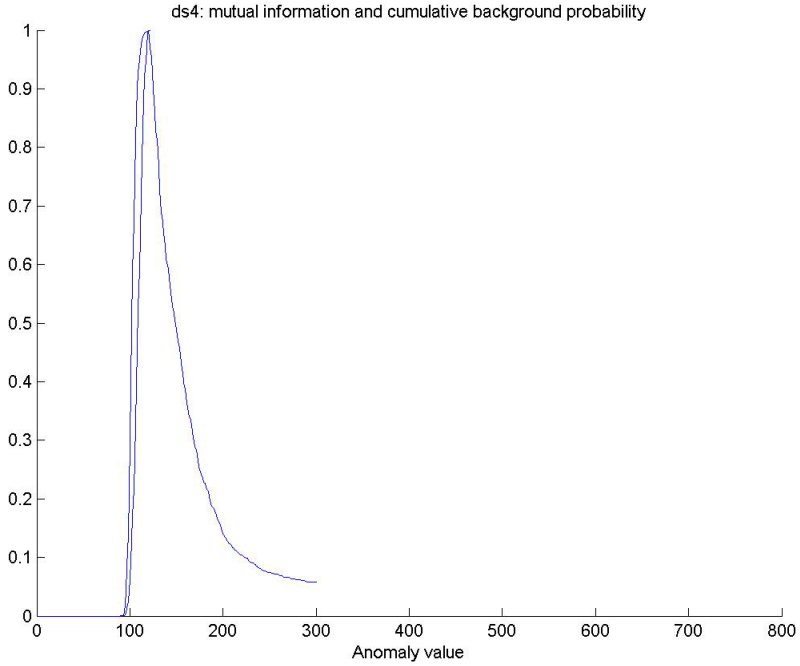


Figure 39. Dataset D, Comparing how the mutual information (rightmost curve) varies with the anomaly value and how the probability (middle curve) for anomaly varies with the anomaly value.

Dataset H

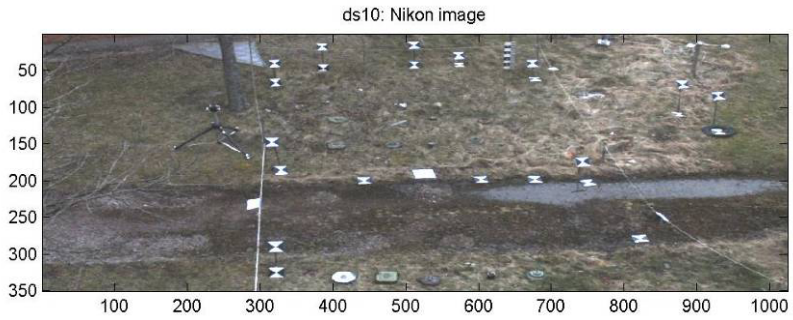


Figure 40. Data set H, Nikon image.

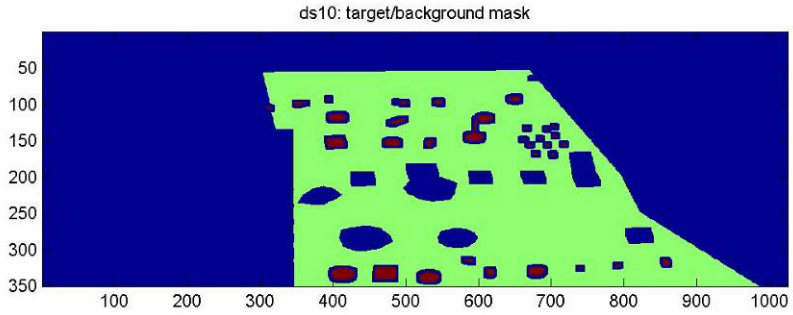


Figure 41. Data set H, Target/Background mask.

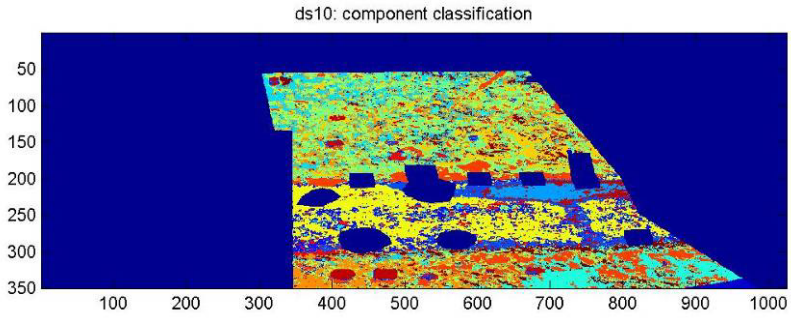


Figure 42. Data set H, The classification into the fifteen components of the Gaussian mixture model.

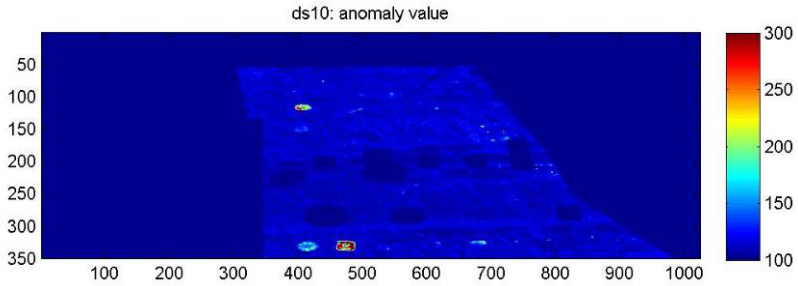


Figure 43. Data set H, The anomaly value obtained by comparison with the Gaussian mixture model.

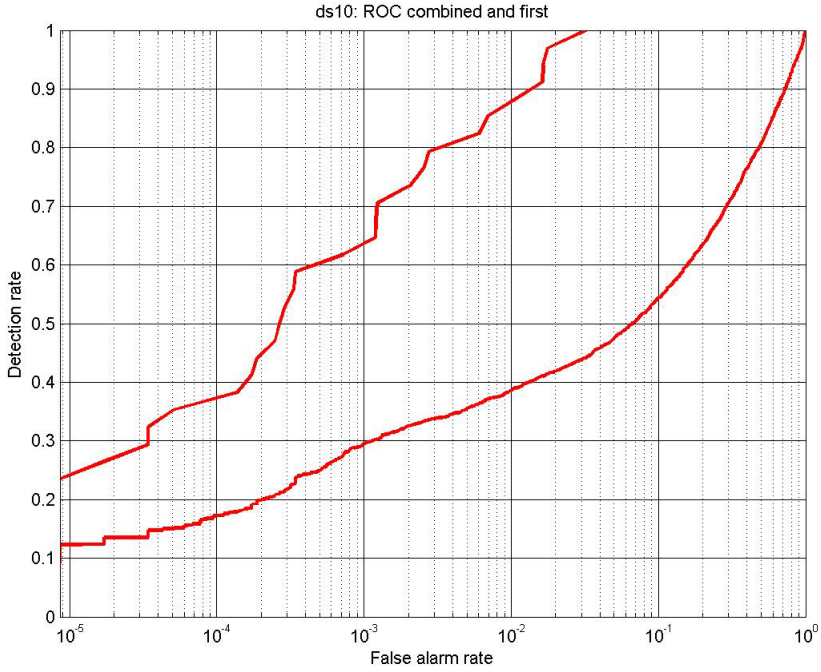


Figure 44. Data set H, The ROC curve with the first pixel on each target (upper curve) and with all target pixels combined (lower curve).

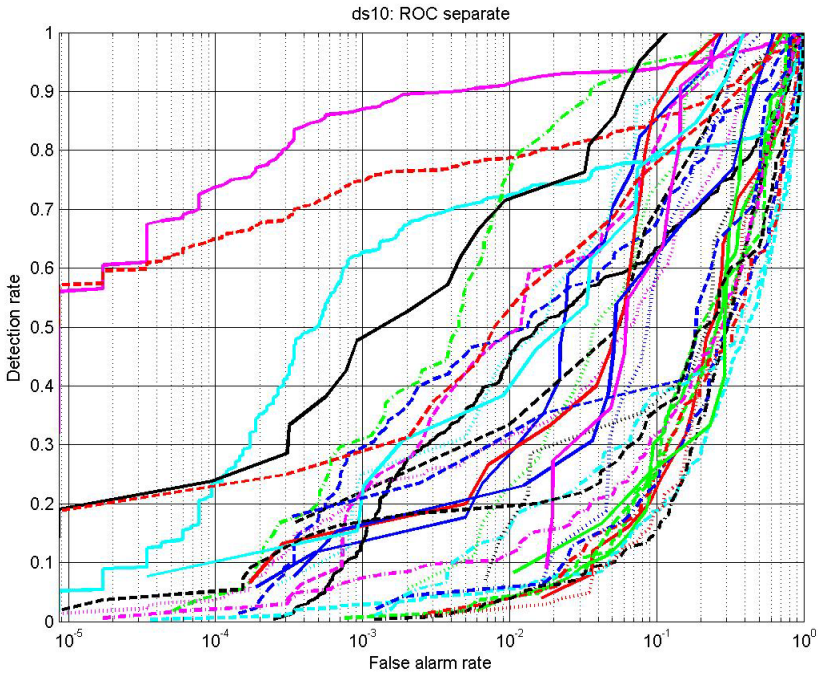


Figure 45. Data set H, The ROC curve with each target separately.

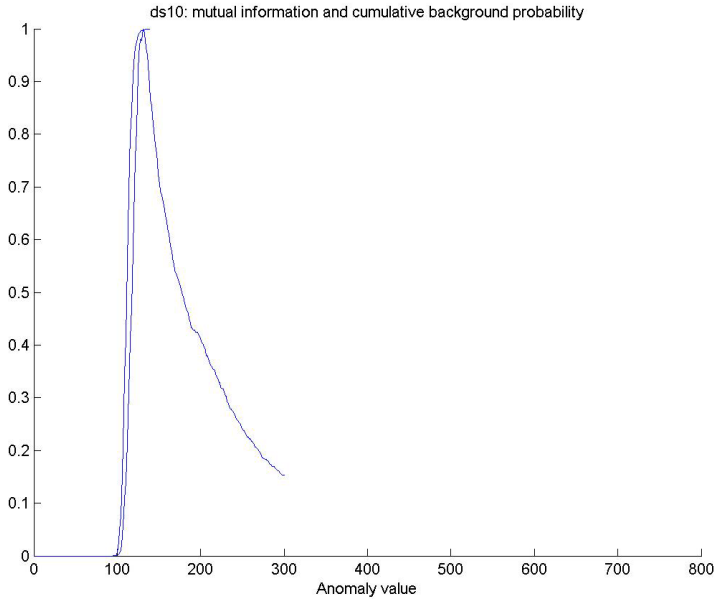


Figure 46. Data set H, Comparing how the mutual information (rightmost curve) varies with the anomaly value and how the probability of anomaly (middle curve) varies with the anomaly value.

Table 6 shows the decision limit giving maximum mutual information. The table also shows the limit where a given high percentage of the background values are below the limit. The values with larger distance than the limit are considered anomalies. The table also shows the mutual information between target mask and anomaly value, component number and with both of them together.

In other examples where the background data have contained some targets the background model has adapted to include the targets even though the targets are different from the rest of the background. Thus the anomalies are no longer anomalies but rather belong to one of the components. This is clear because the component number has much higher mutual information with the target mask then in the case described above where the background model is adapted to pure background. It seems that the model still has the information that

the anomalies are anomalies but encoded as a part of the model. If it is possible to detect which component contain the anomalies then it would probably not be a problem to have the targets as part of the background model.

Table 6. A comparison of the mutual information for the three data sets A, D, and H.

	Data set A	**Data set D**	**Data set H**
max mut info	0.0228	0.0281	0.0347
index for max mut info	146	120	131
p = 0.90	127	109	119
p = 0.95	130	110	121
p = 0.99	138	115	126
p = 0.999	152	121	136
mut info anom value	0.0299	0.0326	0.0443
mut info component nr	0.0174	0.0102	0.0367
mut info anom+comp.	0.0341	0.0351	0.0611
p = 0.90	0.0125	0.0148	0.0202
p = 0.95	0.0149	0.0166	0.0243
p = 0.99	0.0209	0.0247	0.0331
p = 0.999	0.0221	0.0277	0.0330

In the next example we have measured the mutual information between the target mask and each of the sensor bands separately. Figure 47 shows a visual image of the scene and Figure 48 shows the true target/background mask. A Gaussian mixture model with 15 components was adapted to the background pixels of the scene. Figure 52 shows the mean spectrum and Figure 53 shows the covariance matrix of each of the 15 components. In Figure 49 and Figure 50 the corresponding ROC curves are presented. Some targets are anomalous without any false alarms while some other targets only can be detected with a significant amount of false alarms.

Figure 47. A visual image of the scene.

Figure 56 and Figure 57 show an example of selecting bands where the bands are selected by a selection from the Gaussian mixture model instead of selecting bands from the data and then estimating the background model. This approach will not lead to better approximation but the actual result using the fewer selected bands could be worse because the model with all bands can benefit from the bands that are later removed.

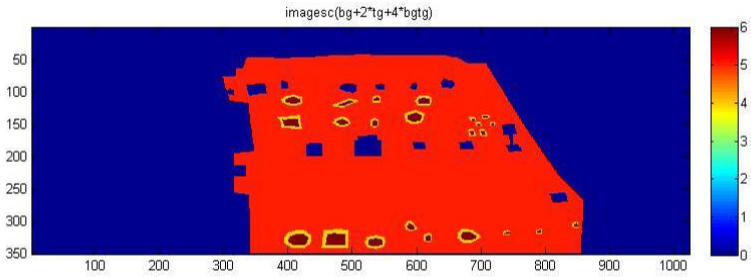


Figure 48. The target/background mask for the scene shown in Figure 47. Blue – void, red – background, dark red – targets, yellow – guard.

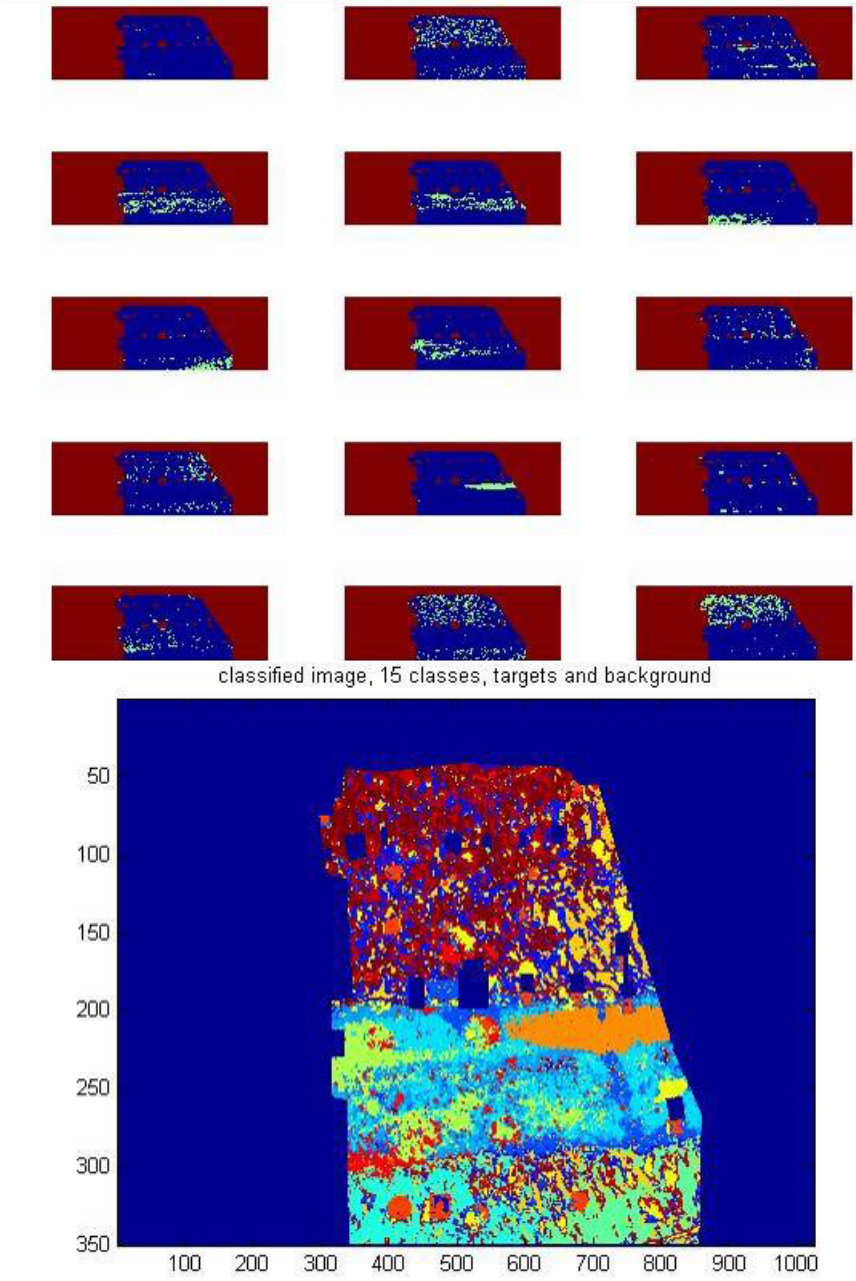


Figure 49. The image shows the classification of the image into the fifteen components of the Gaussian mixture model. Above are the pixels in each of the fifteen components displayed separately (red – void, blue – background, green – the pixels classified as the component). A component with few pixels gathered in few positions may be objects while components with many pixels spread all over the image probably belong to the background.

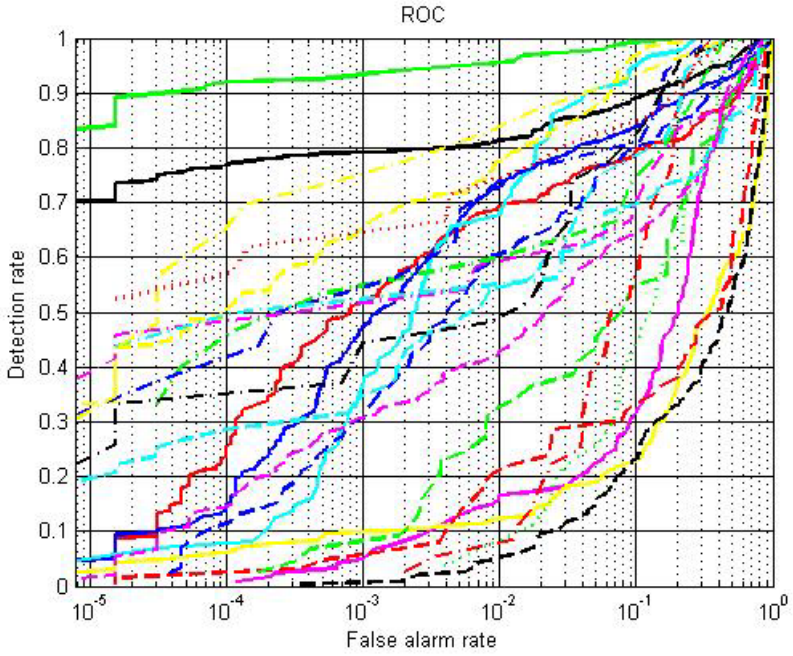


Figure 50. The ROC curve with each target separately.

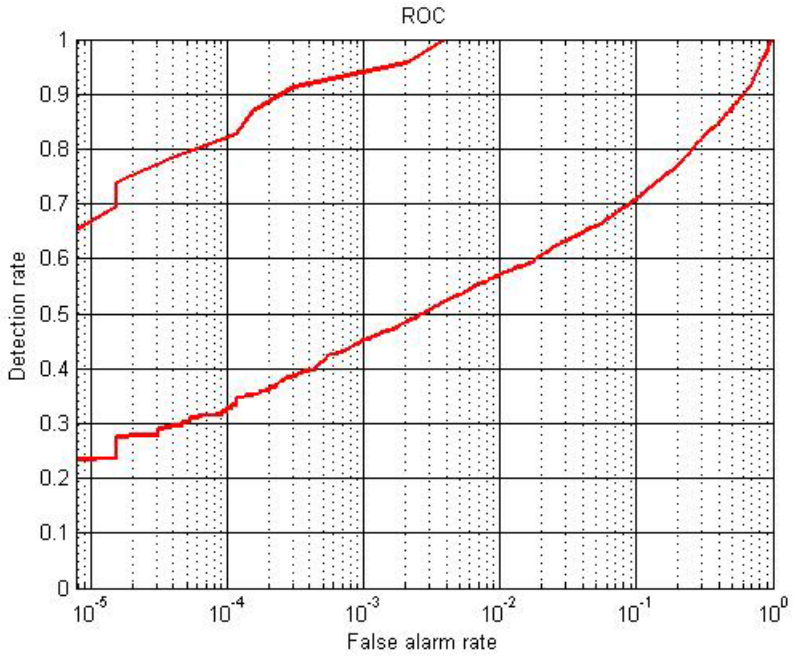


Figure 51. The ROC curve for the first pixel on each target (upper curve) and for all target combined (lower curve).

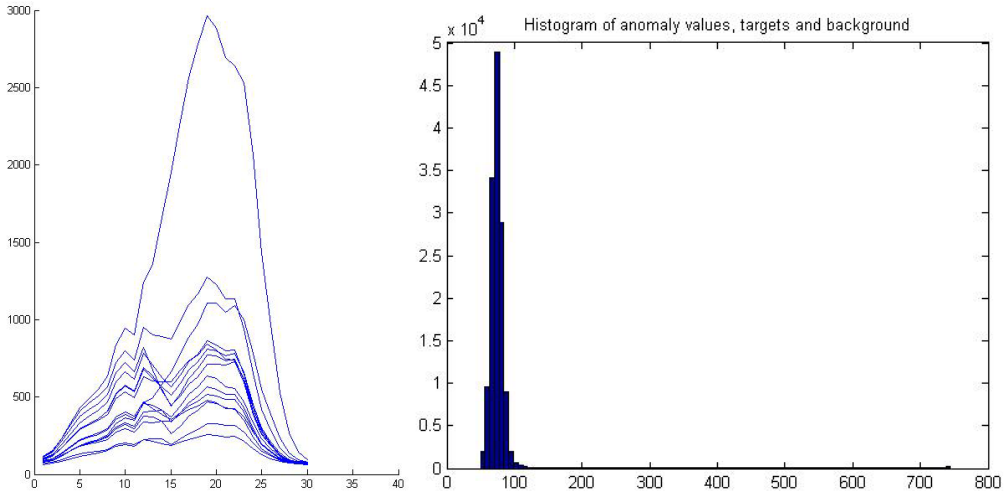


Figure 52. Left: The mean spectrum for each of the fifteen components. Right: A histogram of all spectral values in the image.

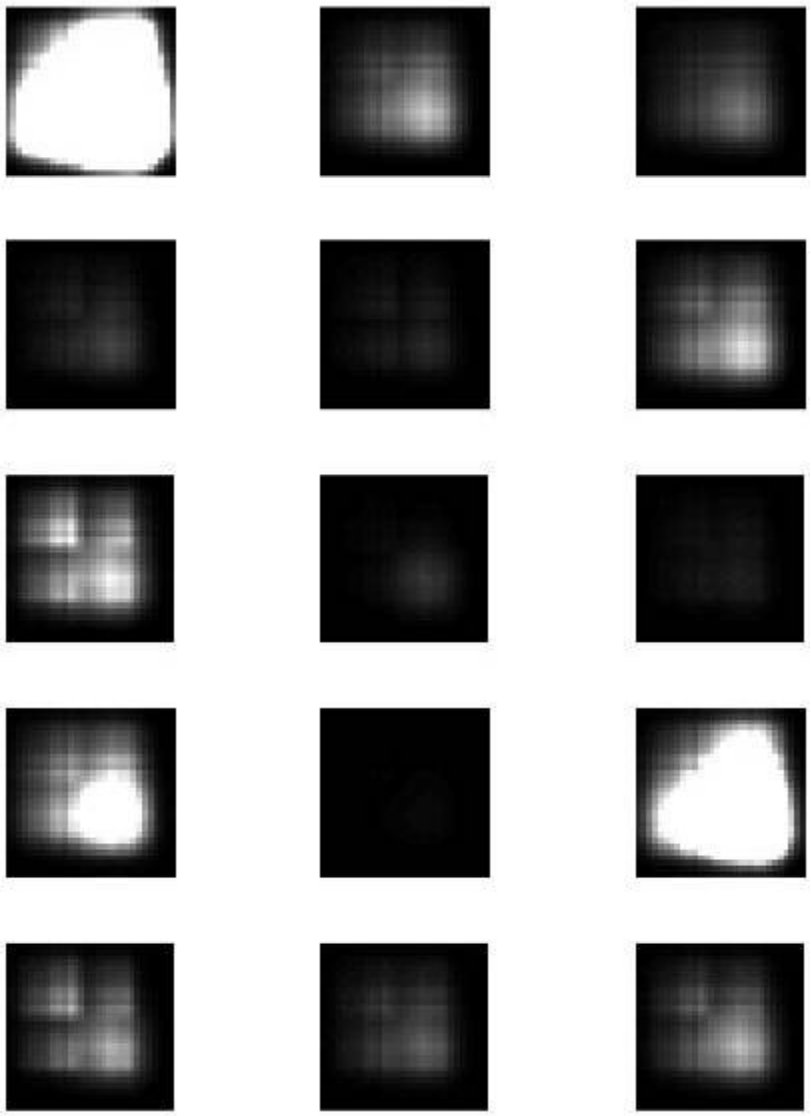


Figure 53. The figure shows the covariance matrices of the Gaussian mixture model.

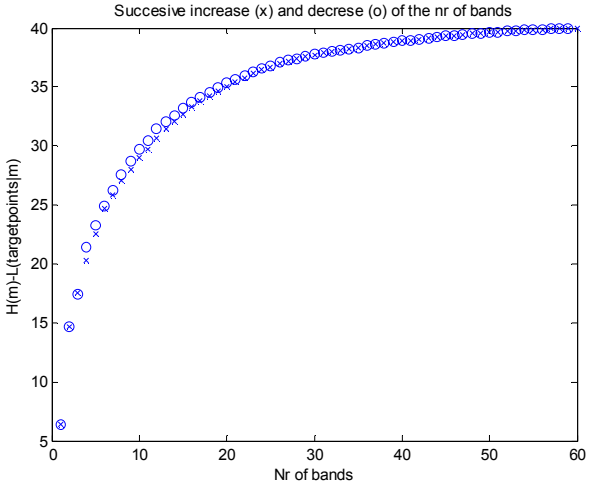


Figure 54. An example with successively increasing and decreasing the number of bands.

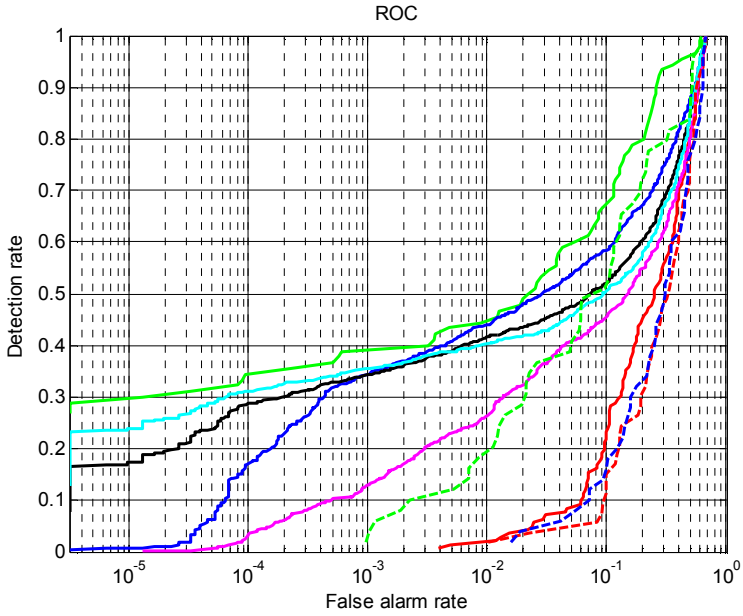


Figure 55. The ROC curve for the example.

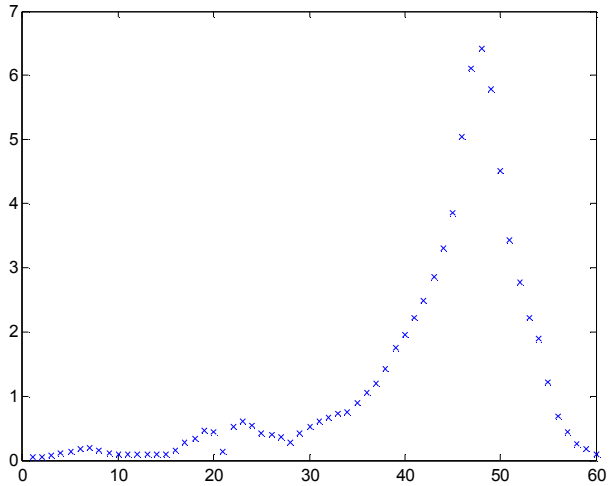


Figure 56. The difference between the estimated coding length of the nine targets and the estimated entropy of the background for each of the sixty bands separately. There are a few bands in which the targets appear as anomalous.

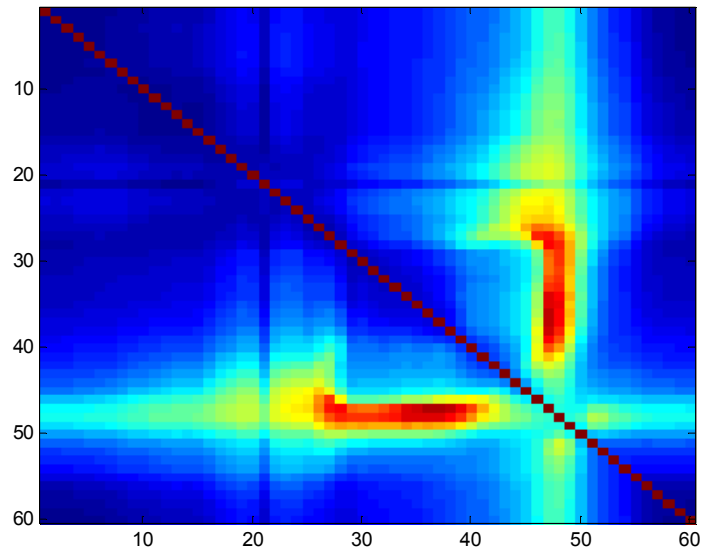


Figure 57. The difference between the estimated coding length of the nine targets and the estimated entropy of the background for each pair of two bands (blue low values, red – high values). There are a few pairs of bands in which the targets of this example appear as anomalies.

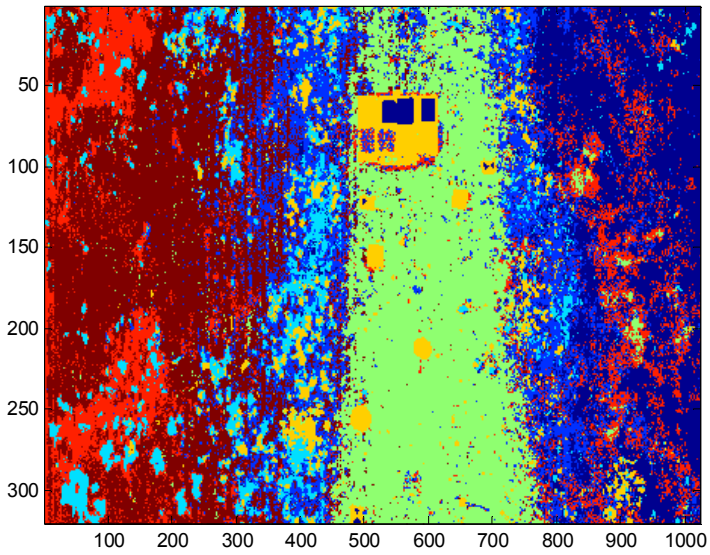


Figure 58. The component that each pixel is closest to.

11.2.1 From detected pixels to detected targets

The anomaly image together with some given limit on the anomaly value will give a set of detected pixels. If one consider every connected set of detected pixels as a potential objects there are still many detections in every image. Some of these groups of detections can be discarded because of their size. They are either too small or too large to be a reasonable target and can thus be discarded.

It is also possible to characterize the detected objects by their spectral distribution given that the spectrums are quantized and binned to a limited number of spectrums. Then it is possible to compare the detections and find those that are similar and those that are essentially different. In this way it is possible to both catalogue and let an operator classify the detections. It is also possible to look for similar objects with a spatial distribution that may indicate that they are either targets or harmless.

11.3 Mine recognition

Basically, recognition can be used for producing results on different levels. One is retrieving those targets that are most similar to the current object (*ranking*), without making a crisp decision. In order words, we can use it to create hypotheses that could for example be fed to the operator for further analysis or decision-making. Another level of result is to define a *recognition threshold*, which can be used for either discarding or accepting the hypothesis. This would enable the system to correctly classify completely unknown objects as "not previously encountered", whereas the ranking approach would always pick the most similar object and pass on the decision-making to the operator.

11.3.1 Spatial object recognition

In this section results from experimental evaluations of three recognition methods, LBP, SC and SIFT are presented. For fair performance evaluation of recognition process, five different test cases have been defined associated with Table 3, see Table 7. We present our results in confusion matrices, see Appendix A. The results are summarized below.

Table 7. The test cases and reference to Table 3.

Test case	Description
1	Train on data from gravel road (E), test on data from forest and grass environment (A-D).
2	Train on data from forest and grass environment (A, C), test on data from gravel road, forest and grass environment (B,D,E).
3	Same scene in training and test, forest environment. Train on data from one perspective (A) and test on data from another perspective (B).
4	Same scene in training and test, grass environment. Train on data from one perspective (C) and test on data from another perspective (D).
5	Train on data from gravel road, forest and grass environment (A, C, E), and test on data from forest and grass environment (B, D).

In LBP, a thresholded chi-square distance is employed as classifier. The threshold is dependent on training data whereas diversity of size and shape of objects are important. Thus, at each test case the corresponding threshold is obtained experimentally even if the discrepancy of all threshold values is not strong.

The best performance is achieved by a special sub-blocking method with respect to the size of the objects. Sub-blocking of small object e.g. TM-10 yields difficulties where the size of each cell is not big enough for having optimal block size. Since our datasets also contain small land-mines, e.g. TM-10 and TM-49 with insignificant visible pixels, the sub-blocking is implemented for the cases where the image has more than a certain number of pixels. The image is then divided into nine sub-blocks with a small overlap; otherwise the LBP feature vector will be a copy of the entire image nine times, in order to keep same feature vector length.

In SIFT, an interesting question was how to define a robust classifier for matching images from matching points; where the matching points are already extracted. The first idea was to count the number of matching points and select the largest set of matching points in trained data. But the result was not convenient due to object similarities; where different objects in our dataset have same shape e.g. circle or square shape likewise same edges as well. Therefore another criterion is needed.

In this work two hypothesis are investigated to classify the image matching; First the average of distances between matching descriptors are measured (denoted by SIFT) and furthermore the distance of the first matching descriptor is considered (denoted by SIFT*). It is also mentionable that some land mines in our datasets have too low resolution to extract enough key-points; this fact may cause that no matching points are obtained in image matching process.

In SC, a specified number of points are collected from the edge points (100 points in this work). Collecting the edge points of land mines is crucial in SC algorithm. Some objects are too small to find 100 edge points and some objects have too large edge point sets, which demand sampling models to recollect 100 points. This sampling model should cover the entire edge information.

The other issue in using SC in this work is execution time. Extracting SC features is fast but the matching algorithm includes an assignment problem, solving by Hungarian method, which is most time consuming. The SC algorithm is an iterative method and convergence is usually achieved after six iterations where each iteration needs one second to operate. Therefore finding a matching of an object among *Nr* training data and *Ns* test data will take $6 \times Nr$ seconds and $6 \times Nr \times Ns$ seconds, respectively. In this test, SC is only used in two test cases, i.e., test case 3 and 4.

Considering each test case; the training set includes all visible mines in the scene, while that test data includes all objects (including mines and non-mines). Therefore the system will be able to identify an ordinary detected object to any known type (trained) of mines or even more non-mine objects as well. For instance in Test case 4 the illumination conditions and the angle of view vary between the training and test scenes due to different data collection times (noon and afternoon). For this test case the algorithms will be able to recognize all mines in dataset D after seeing dataset C, but there are miss-classifications of non-mine objects.

For instance LBP works better, than the other methods, for recognizing small mines (TM-10) and this is due to the novel sub-blocking method using in this work. However no Grenade is recognized even if the system has seen it before in the training data, due to occlusion effects in the training dataset.

To conclude, our tests indicate that it is possible to recognize small mines from low resolution data with various illumination conditions and view angles. Five test cases have been studied to evaluate the four recognition methods. We have recognized difficulties in threshold settings for the tested methods; therefore number of FP and FN are large. The thresholds can probably be defined after further tests. Another factor that has increased the number of FP in this test is the fact that we used all manually detected objects in the scenes. In an application the mine-like objects that will be recognized, will be the output from an anomaly detector. Our tests show that the anomaly detector will remove some of the FP objects from the data set that will be passed on for recognition.

11.3.2 Spectral object recognition with SVM

In order to investigate the potential of using spectral properties for mine recognition, some initial experiments with SVM-based classification (Section 8.4) were carried out on datasets A and B (refer to Table 3). In the experiments, the LIBSVM library for Matlab (Chang, 2009) was used. Three test cases were considered:

a) *Test generalization performance when lighting conditions are similar between training and testing.*
   In practice: train on a sub-set of mine objects in one scene (dataset A), test with all other mine objects in the same scene

b) *Test generalization performance when lighting conditions are different between training and testing.*

In practice: train on a sub-set of mine objects in one scene (dataset A), test with all other mine objects in another scene acquired six hours later (dataset B).

c) *Test re-identification performance in different lighting conditions*
In practice: train on a sub-set of mine objects in one scene (dataset A), test with the same objects but seen from another angle and under different lighting conditions (dataset B)

A number of blobs each corresponding to one of five different objects – TMM-1, 41/47, TM-10, TM-49 and PMR-2A – were selected for training. The SVM produces a recognition probability for each of the classes and by applying and varying a threshold, different recognition performance was obtained. For each particular threshold value a confusion matrix (Section 10.4) was calculated. An object was classified as belonging to the class to which the majority of the pixels on the target were classified. An example of a confusion matrix is shown in Table 8. Example of a confusion matrix for SVM-based mine recognition corresponding to case b) above. The results were obtained with a combination of hyperspectral and laser data.

		Classification result					
		TMM-1	*41/47*	*TM-10*	*TM-49*	*PMR-2A*	*Unknown*
True class	*TMM-1*	1					1
	41/47		1				
	TM-10			1			2
	TM-49				3		
	Other	2					2

By varying the threshold and keeping track of the number of correct and erroneous decisions, respectively, an ROC curve is obtained (Figure 59). The recognition was tested with laser intensity and hyperspectral data, respectively, as well as with a combination of hyperspectral and laser data. In order to compensate for changes in the lighting conditions, each feature band was normalized.

.

Table 8. Example of a confusion matrix for SVM-based mine recognition corresponding to case b) above. The results were obtained with a combination of hyperspectral and laser data.

		Classification result					
		TMM-1	*41/47*	*TM-10*	*TM-49*	*PMR-2A*	*Unknown*
True class	*TMM-1*	1					1
	41/47		1				
	TM-10			1			2
	TM-49				3		
	Other	2					2

By varying the threshold and keeping track of the number of correct and erroneous decisions, respectively, an ROC curve is obtained (Figure 59). The recognition was tested with laser intensity and hyperspectral data, respectively, as well as with a combination of hyperspectral and laser data. In order to compensate for changes in the lighting conditions, each feature band was normalized.

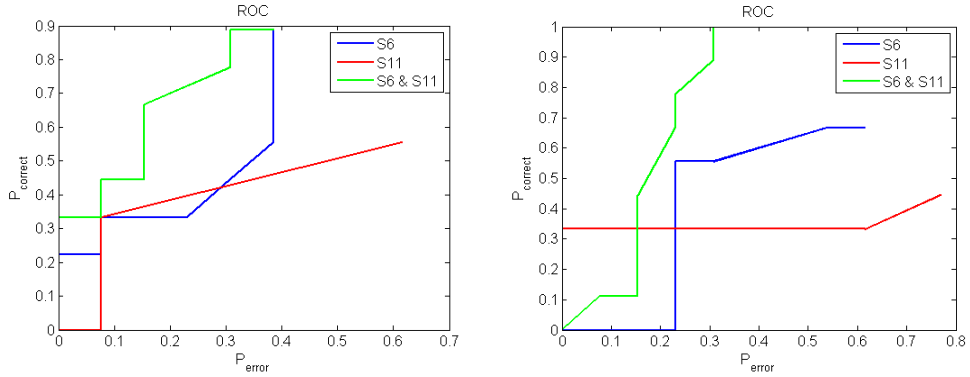


Figure 59. ROC examples showing the performance of SVM-based mine recognition using hyperspectral and laser data. Left: Case a). Right: Case b) (see explanation above).

Although the extent of the experiment was rather limited, some interesting observations were made. For example, the results generally improved significantly when using a combination of hyperspectral and lidar information, compared to using the sensors separately. The majority of the objects belonging to a class for which the classifier was not trained usually ended up in the TMM-1 class, as the colours were quite similar. Generally, the classifier had difficulties recognizing the small TM-10 mines, partly because it was very difficult to spot them and thus very difficult to obtain an accurate ground truth mask for training the classifier on those objects. A conclusion is that spectral information is indeed useful for discriminating between objects, but that it should be combined with spatial information for improved robustness. After all, spectral-based recognition only finds similar "colours" and if the appearance of the mines changes, due to environmental influence (sunlight, dirt, etc) or due to somebody re-painting them, the spectral classifier may not recognize them.

11.3.3 Mine recognition by CAD model matching

Two mine candidates that were detected in the anomaly analysis have been tested for mine classification and mine type recognition using the model-based 3D recognition approach. The mine candidates were compared with the models in two different ways, first the dimensions are compared and second, the point cloud the target is compared with the faces-representation of the CAD models, see Figure 60.

For our two candidates, only candidate 1(a TMRP-6) have a corresponding library model. For the other candidate, a TMA-5, a CAD model is not present. This means that the TMA-5 cannot be recognized, we can only see how similar it is to our models.

The mine candidates' dimensions were first compared to the dimensions of the models. The first candidate's dimensions where most similar to the TMRP-6 , followed by AT-2 and AT-47b. This is also the most similar models in the library, with a correct classification of a TMRP6. The second candidate's dimensions where most similar to the FRDM13, followed by PMR-2A and TMRP-6. The FRDM13 is the model that is most similar to a TMA5 mine. Results of dimension estimation are shown in Figure 61.

In the CAD matching step the first candidate is correctly recognized as a TMRP-6 and the second candidate is matched as a FRDM13, which is the most similar library model, results are shown in Figure 62.

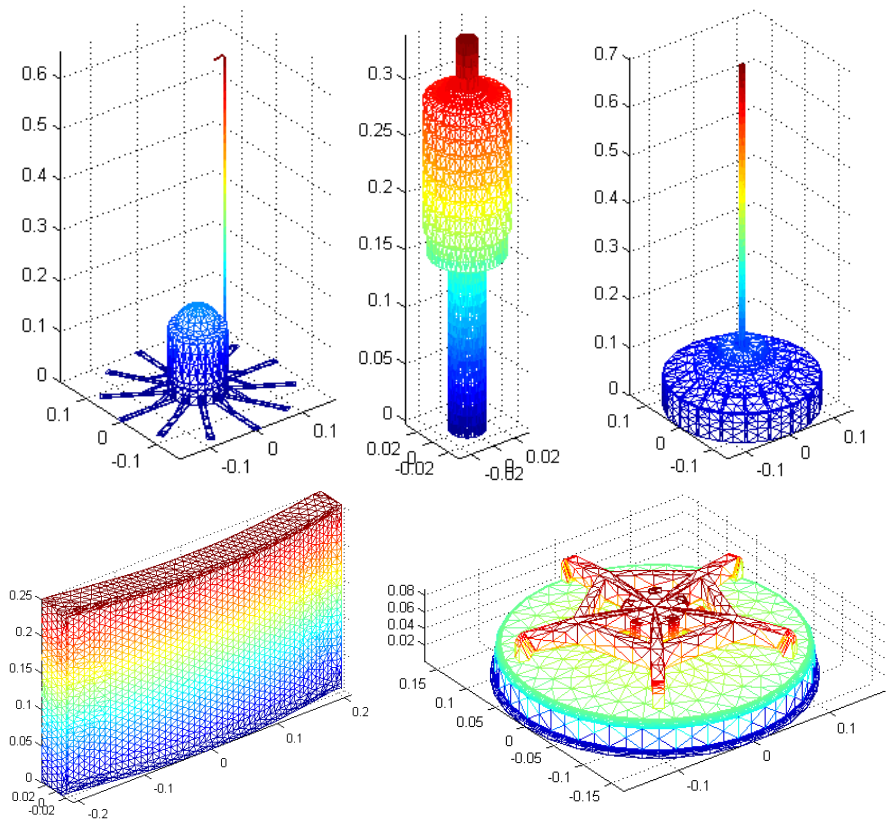


Figure 60. Model of the mines used for matching with target data; AT-2 (top, left), PMR-2A (top, middle), TMRP-6 (top, right), FRDM13 (bottom, left), AT-47b (bottom, right). Axes in meters.

The dimension comparison is a fast table-lookup comparison that is possible to perform in real time (results are presented immediately in Matlab). The CAD matching is time consuming. First, the models cannot hade too many faces as each target sample is compared with each face of the model. Investigations of vehicle and mine models indicate that most models have enough dense representation by 200-100 faces. Today a typical matching takes 30 seconds in Matlab. The matching problem is possible to redesign to parallel computing, which will reduce the computation time dramatically. We believe that CAD matching is possible to perform in real time or near-real time. Also, this type of matching is intended for few, selected objects rather than massive testing of all objects in a scene.

This approach has shown good results for vehicle recognition and these tests indicate that the approach can be useful also for mine recognition. For mine recognition maybe the first step is sufficient, after that the operator and/or a demining team perform further analysis.

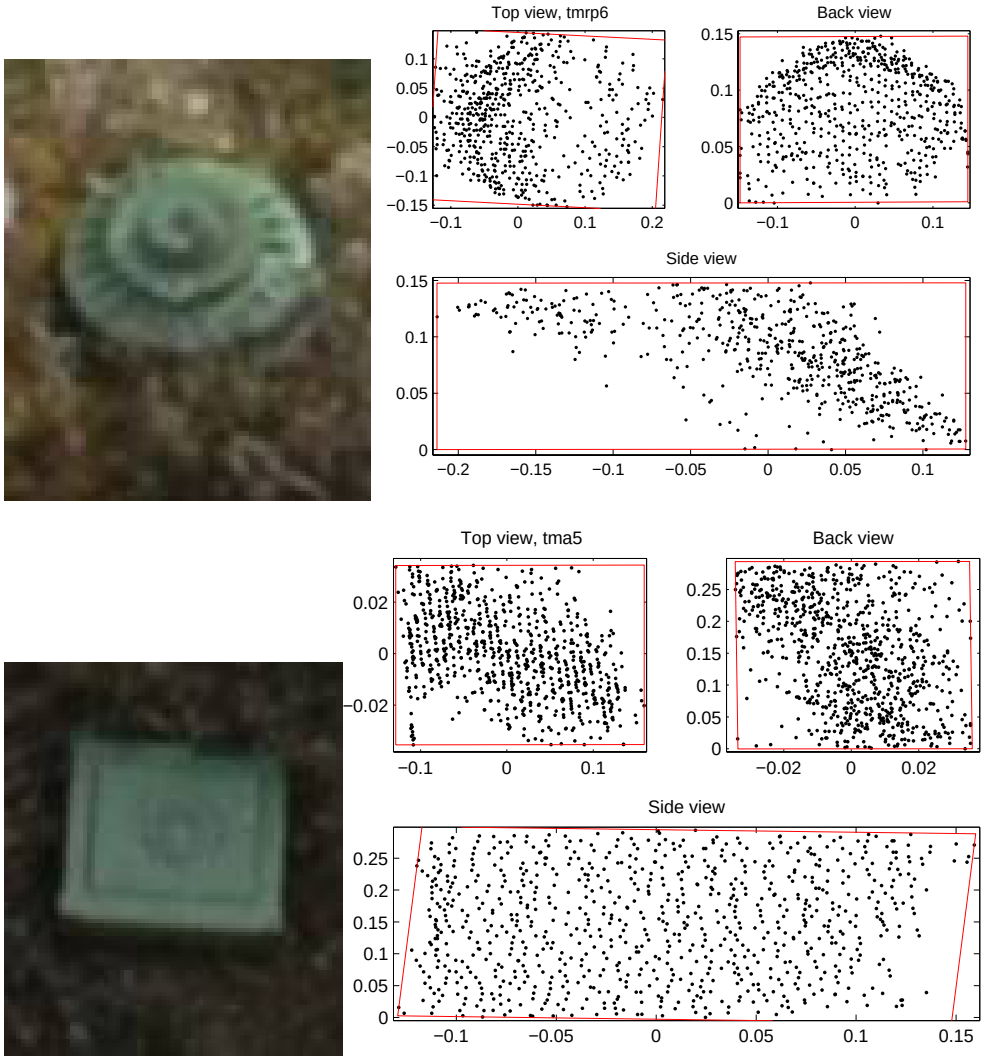


Figure 61. Possible mines that were analyzed using the model-based 3D recognition approach. Top: image of the scene, right: dimension estimation of mine candidate 1 (a TMRP6), bottom: dimension estimation of mine candidate 2 (a TMA5).

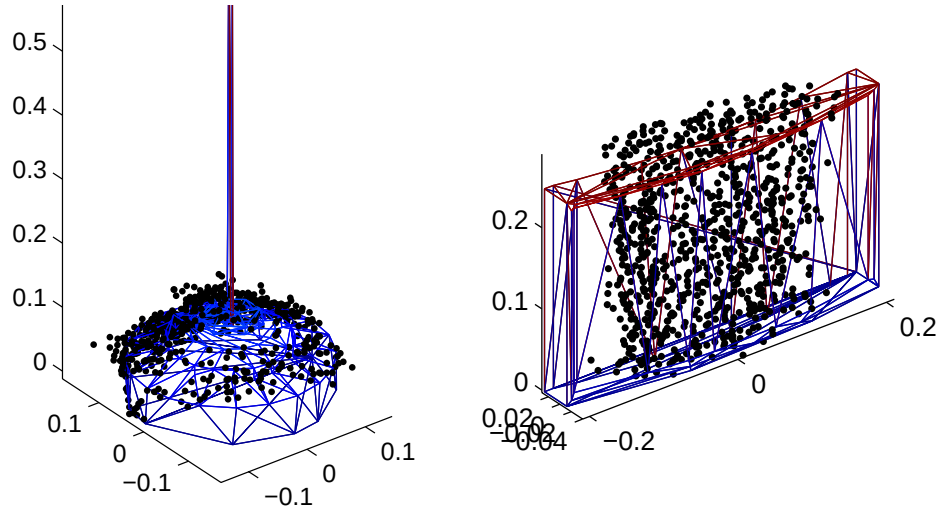


Figure 62. CAD matching of mine candidate 1 (TMRP6) with TMRP6 model (left) and of mine candidate 2 (TMA5) with FMRD13 model. Axes in meters.

12 Conclusions and discussion

12.1 Sensor design/configuration

Information theoretic measures like entropy and mutual information are useful when configuring the sensor and detector system. It is possible to compare different configurations regarding how much information they convey about the presence or absence of targets in the scene. It is also possible to characterize the sensor system without considering a specific detection algorithm. In many cases it is possible to estimate how much information the sensor data convey about the scene. If there are interesting amounts of information one has to construct a detection algorithm that can retrieve the information.

It seems that no sensor on its own can give enough information about the scene. Thus, it is necessary to be able to attribute the sensor data to specific locations in the scene so that the information from different sensors can be used together to improve the description of the scene and the objects in the scene and thus improve the possibilities to detect targets.

Many of the described techniques detect anomalous objects by comparing the appearance locally in the sensor data. If signatures of interesting objects are collected to be used for recognition then radiometric calibration is necessary to be able to recognize the objects since most certainly the environment will be different.

12.2 Occlusion effects

Apart from choices concerning sensors and signal processing techniques, the overall probability of actually detecting mines in a real scene is strongly influenced by *occlusion* effects. We found that many objects often escaped detection due to the fact that they were heavily occluded (by grass, sprigs, leaves, etc.). Although the targets were not always entirely concealed by the foreground, the sensor resolution was not enough to accurately capture information about the objects. Several objects were in fact so occluded that they were impossible to discern, even though their locations within the dataset/images were well known. Part of the explanation for this is that for several objects the viewing angle was not steep enough to allow for a clear view of the object.

Occlusion effects are usually lowest for a nadir-looking system. For practical reasons, we were not able to test that case. Instead we tested scene aspect angles of 35-70 degrees (90 degrees being the nadir direction), which is a somewhat harder case. The level of occlusion can also be lower by collecting data from multiple views but this, on the other hand, demands very accurate registration.

12.3 Data fusion and registration

Fusion on the signal/pixel level requires very accurate data registration. From the experiments carried out within MOMS so far, it can be concluded that such accuracy is difficult to obtain with a distributed sensor system, at least for small targets (AP mines). In fact, pixel correspondence between sensor images will probably require a common detector array or arrays situated very close to each other.

Fusion on the decision-level, on the other hand, will cope considerably better with a less accurate registration, as the different sensor data streams are processed individually and only the final outputs are combined. Further signal processing work in MOMS will focus more strongly on decision-level fusion.

12.4 Anomaly detection

Based on the results obtained in MOMS, it can be concluded that anomaly detection is a very useful tool for detecting possible mines. The real benefit is that the anomaly detector only has to be trained with background data, not with targets, and based on a background model it high-lights any part of the scene that deviates from the background model. Obviously, not only mines are anomalous but also other man-made objects, natural clutter and local variations in the background may be detected. As a first indication, however, this is very valuable information, since processing power can then be directed towards those anomalies, in order to determine whether they are likely to be mines or not. The success of the anomaly detection hinges on always maintaining an accurate background model and the fact that there must be a contrast between the targets and the background. In practice also, the visible portion of the object should be more than a couple of pixels, in order to be able to remove the irrelevant anomaly pixels that will always appear.

Creating a global Gaussian mixture model background model in Matlab, using 15 Gaussian components, 30 spectral bands and a sensor resolution of about 0.5 Mpixels, takes a couple of minutes. Still, we estimate that by using fewer bands (maybe 3-5) and having a dedicated chip executing the algorithm, it would be fully possible to meet real-time constraints (10-20 images/s). This also means that this processing could take place close to the sensor and only a limited amount of data (corresponding to a number of detected anomalies) need then be transmitted. Further, incremental update of the Gaussian mixture background model demands less execution time compared to creating the model from scratch.

The proposed approach for anomaly detection can be adjusted to detect various objects that differ from the natural background, for example IEDs and other man-made objects.

12.5 Spatial feature extraction

The more information we can extract about possible targets, the better the chances to make correct decisions concerning the nature of these objects. Reliable extraction of different kinds of spatial features would be very useful in this respect. During the work within MOMS, it was found to be very challenging to extract such features, given the sensor data available. In Section 6.2, using 2-D IR imagery to estimate object properties, such as convexity and curvature, was discussed. The conclusion from the experiments performed is that it was difficult to obtain reliable feature values from other objects than relatively large objects in quite clutter-free neighbourhoods. From a computational viewpoint, however, these techniques are quite attractive as they are often convolution-based and the amount of numerical operations per frame needed to compute the desired features is known beforehand. This makes them suitable for hardware implementations close to the sensor, e.g. even onboard the platform carrying the sensors.

Another option, also considered within MOMS, is to extract spatial features from 3D data (see Section 6.1). For example, it was concluded that finding surfaces may indeed help find human-made objects, but that the range noise level and range resolution limitations of the current sensor made it difficult to obtain consistent results.

12.6 Supervised classification for detection and recognition

The common denominator of all supervised classification techniques is that they must all be presented with samples of the targets they are supposed to detect. Since the scene conditions (contrast, light levels, shadow, occlusion, etc.) may change significantly, it is important that the properties used for representing the objects are stable enough so that the system can reliable detect them even under new conditions. A general problem in the mine

detection task is that the mine signatures often look very different (even signatures from the same mine under different conditions) and it is hence difficult to define useful and robust features.

From a system perspective, a mine detection based solely on supervised classification cannot be recommended; it is risky to rely on that our target database is kept up-to-date and contains information about all the possible threats the system may encounter.

Nevertheless, such a technique can run in parallel with the anomaly-based detection and report whenever the system encounters an object that is very similar to a target with which it was trained. Certain techniques can be implemented so that real-time demands can be met. Of course, this depends on the number of target models that have to be considered. Often the training phase is quite computationally demanding, whereas the testing phase is often much faster.

12.7 Spatial resolution of the sensors

In practice, in order to be able to extract the relevant information about potential targets in the scene, the sensor must have a sufficiently good spatial resolution. Again, consider anomaly detection that often results in a number of unwanted detection that we want to get rid of. Therefore, in practice, the spatial resolution of the sensor must allow for having several pixels on the target. For a relatively large object, e.g. an AT mine, the pixels should correspond to a resolution on the target of maybe about 2-3 cm, to enable the removal of small, irrelevant objects (Section 8.1). Also for evaluating spectral similarities between objects (Section 8.2), the resolution must be good so that there are enough data for computing sufficient statistics (histograms) for each object.

For mine *recognition* based on spatial properties, the sensor resolution should be significantly better than 2 cm, probably around 5 mm or below. Even at that resolution, it may be difficult to distinguish (small) objects from each other.

During the work with 3D data acquired with the current laser radar system, it was found that the range resolution capabilities of this system were too poor. Large amounts of erroneous ghost points appeared as a result of the sensor's inability to resolve reflecting surfaces at short distances, e.g. mine behind grass. A laser radar system emitting a shorter pulse would improve the range resolution and subsequently the chances of detecting objects behind occlusion. Nevertheless, the laser radar sensor was often able to capture intensity contrasts between mines and background due to its active mode, then basically having the sensor act as a high-resolution 2D imaging device equipped with its own light source. Indeed, the intensity data from this sensor showed to be very useful, but it has to be combined with other features in order to obtain a more robust detection system.

To summarize, the spatial resolution basically governs with what level of information decisions can be made (see Figure 2). From the experiments carried out within MOMS, the following rules of thumb could be formulated

- *low resolution* (>10 cm) is likely to result in relatively poor performance, as the expected number of "clean" mine pixels will be quite small, thus making it difficult to match spectral signatures and to estimate object size
- *medium resolution* (5-10 cm) gives the ability to detect anomalies and possibly to detect suspicious-looking pixels through matching of spectral signatures
- *high resolution* (2-3 cm) enables us to clean up the detections, define objects and to detect mine-like objects
- *very high resolution* (<0.5 cm) is probably needed to be able to distinguish between different mines based on their spatial appearance.

12.8 Active versus passive sensing

A system for detection of small ground objects, like land mines, would benefit from including an active imaging sensor, preferably operating at several wavelengths or a broader range of wavelengths. In addition to providing night-time capabilities, such a system would also probably result in reduced problems caused by uneven and unpredictable illumination of the scene (e.g. shadows), which would be very favorable from a signal processing point of view.

12.9 Operator aspects

The anomaly detection and the supervised approaches can be updated under a mission, to adapt to the current conditions in the area of interest. At the first trials in a new environment there is likely to be a higher level of false detections. Through an extra training phase, supervised by a skilled operator, the algorithms can be tuned to the new environment and the false alarm rate can be lowered while retaining the mine detection rate.

A critical stage in any automatic data analysis system is to determine suitable thresholds below which objects are discarded from further analysis. Lowering the thresholds results more true detections/recognition, but at the prize of increasing the false alarms or erroneous classifications. However, an alternative to setting hard thresholds in this stage is to let the entire system – including hardware, software and operator – process as many objects as possible within a certain given time, starting with the most suspicious-looking one and continuing down the list of decreasingly interesting objects.

13 References

Andersson Pierre, Tolt Gustav, “Detection of vehicles in a forest environment using local surface flatness estimation in 3-D laser radar data”, SSBA, Linköping, 2007

Belongie Serge, Mailk Jitendra, Puzicha Jan, “Shape Matching and Object Recognition Using Shape Contexts”, IEEE PAMI, Vol. 24, No. 24, April 2002

Chan Kevin, “Registration of 3-D laser radar data and hyperspectral imagery for target detection”, FOI-R--2101--SE. November 2006

Chang Chih-Chung, Lin Chih-Jen,”LIBSVM : a library for support vector machines, 2001”. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm March 2009.

Chen Chu-Song, Hung Yi-Ping, Cheng Jen-Bo, “RANSAC-Based DARCES: A New Approach to Fast Automatic Registration of Partially Overlapping Range Images”, IEEE PAMI Vol. 21, No. 11, pp. 1229-1234, November 1999

Cristianini Nello, Shawe-Taylor John, “An Introduction to Support Vector Machines”, Cambridge University Press, Cambridge, 2000.

Grönwall Christina, Gustafsson Fredrik, Millnert Mille, “Ground Target Recognition Using Rectangle Estimation”, IEEE Transactions on Image Processing 15(11): 3400-3408 2006

Johnson Andrew Edie, “Spin-Images: A Representation for 3-D Surface Matching”, PhD Thesis, Robotics Institute. Pittsburgh, PA: Carnegie Mellon University, 1997

Larsson Håkan, Chevalier Tomas, Gustafsson Frank, “3-D structure and reflectance measurements – A system analysis of Lidar Optech ILRIS-3D”, FOI-R--2116--SE, September 2007

Larsson Håkan, Karlsson Kjell, Lindell Roland, Letalick Dietmar, Nilsson Pär, Svensson Thomas, “Measurement report from MOMS field trial in Eksjö April 2008”, FOI-D--0314--SE, 2008

Letalick Dietmar, Chevalier Tomas, Larsson Håkan, Nelsson Claes, Nyberg Sten, Steinvall Ove, Tolt Gustav, “MOMS – Analysis and evaluation of experimental data”, FOI-R--2012--SE, June 2006

Letalick Dietmar, Grönwall Christina, Hallberg Tomas, Larsson Håkan, Renhorn Ingmar, Tolt Gustav, "MOMS - Data collection and evaluation", FOI-R--2328--SE, September 2007

Linderhed Anna, Sjökvist Stefan, Nyberg Sten, Uppsäll Magnus, Grönwall Christina, Andersson Pierre, Letalick Dietmar, “Temporal analysis for land mine detection”, Proc. of IEEE. International symp. on image and signal processing and analysis, p. 389-394, Zagreb, Croatia, September 2005

Lowe David G, “Object recognition from local scale-invariant features”, International Conference on Computer Vision, Corfu, Greece, pp. 1150-1157, September 1999

Ojala T, Pietikäinen M & Harwood D, “A comparative study of texture measures with classification based on featured distribution”, Pattern Recognition, 29(1):51-59, 1996

Ojala T, Pietikäinen M & Mäenpää T, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7):971-987, 2002

Rahman Zia-ur, Jobson Daniel J, Woodell Glenn A, Hines Glenn D, “Image enhancement, image quality, and noise”, Photonic Devices and Algorithms for Computing VII, Proc. SPIE 5907, 2005

Renhorn Ingmar, Cronström Staffan, Larsson Håkan, Lindell Roland, Svensson Thomas, Tolt Gustav, Wadströmer Niclas, ”Mine detection and classification using optical sensors”, FOI-R--2496--SE, March 2008

Sjökvist Stefan, Abrahamson Staffan, Andersson Pierre, Chevalier Tomas, Forssell Göran, Grönvall Christina, Larsson Håkan, Letalick Dietmar, Linderhed Anna, Menning Dennis, Nyberg Sten, Renhorn Ingmar, Severin Mattias, Steinvall Ove, Uppsäll Magnus, Tolt Gustav, ”MOMS multi optical mine detection system - initial report”, FOI-R--1721--SE, 2005

Svensson Thomas, Ahlberg Jörgen, Allard Lars, Björklund Svante, Carlsson Leif, Cronström Staffan, Fagerström Jan, Karlsson Nils, Lindell Roland, Renhorn Ingmar, Wadströmer Niclas, “Multi- och hyperspektral spaning 2006-2008 – slutrapport”, FOI-R--2642--SE, December 2008

Steinvall Ove, Carlsson Leif, Letalick Dietmar, Renhorn Ingmar, Tolt Gustav, Wadströmer Niclas, “MOMS – System concept ideas”, FOI-R--2576--SE, Sept 2008

Tolt Gustav, Larsson Håkan, “Waveform analysis of lidar data for targets in cluttered environments”, SPIE Europe Optics/Photonics in Security & Defence; Electro-Optical Remote Sensing, Detection, and Photonic Technologies and their Applications, Florence, Italy, 2007

Westberg Daniel, Tolt Gustav, Grönwall Christina, “A sensor fusion method for detection of surface laid land mines”, FOI-R--2488--SE, January, 2008

A Confusion matrices for spatial object recognition

In the sequel we define *Nr* as the number of training objects and *Ns* as the number of test objects.

A.1 Test case 1

Training data: data set E, *Nr* = 13 mines.

Test data: A, B, C, D, *Ns* = 43 (26 mines , 17 non-mines).

Target Classification

Test case 1	**Landmines**			**Not Landmines**		
	LBP	**SIFT**	**SIFT***	**LBP**	**SIFT**	**SIFT***
Known (trained) Land mines	17	15	16	9	9	13
Unknown objects	1	11	9	16	7	4

Observations:

For some small objects no matching points can be found for SIFT, therefore *Ns* (SIFT) = 42.

LBP works better for recognition of small mines (TM-10) and this is due to the novel sub-blocking method using in this work. The number of FP and FN with LBP are less than for SIFT and SIFT* due to threshold selection difficulties with SIFT and SIFT*. To obtain a reasonable threshold for SIFT(*) was difficult. Therefore the unknown-objects which are often non-mines are mostly misclassified as known-mines.

The number of “FN” is high in this test case, due to various types of environments in the test scenes (road, forest and clear-cut forest scenes).

Landmine classification

Landmines in Scene	TMM-1			TM-10			TM-49			TMA-5			Grenade		
	LBP	SIFT	SIFT*	LBP	SIFT	SIFT*	LBP	SIFT	SIFT*	LBP	SIFT	SIFT*	LBP	SIFT	SIFT*
TMM-1	6	6	7												
TM-10				9	8	8									
TM-49							1	0	0						
TMA-5										1	0	0			
Grenade													0	1	1

Test case 2

Training data: A, C, *Nr* = 11.

Test data: B, D, E, *Ns* = 39 (24 mines, 15 non-mines).

Target Classification

Test case 2	Landmines			Not Landmines		
	LBP	**SIFT**	**SIFT***	**LBP**	**SIFT**	**SIFT***
Known (trained) Land mines	19	16	17	5	6	6
Unknown objects	9	11	11	6	4	3

Observations:

LBP works better for recognition of small mines (TM-10) and this is due to the novel sub-blocking method using in this work. The number of FP and FN with LBP are less than for SIFT and SIFT* due to threshold selection difficulties with SIFT and SIFT*. To obtain a reasonable threshold for SIFT(*) was difficult. Therefore the unknown-objects which are often non-mines are mostly misclassified as known-mines.

The number of “FN” is high in this test case, due to various types of environments in the test scenes (road, forest and clear-cut forest scenes).

The number of “FP” is high in this test case, due to shape similarity of different types of land-mines. For instance the unknown test object TMPR-6 (not in trained data) is very similar to TMM-1 and TMA-5, and it is often misclassified as one of those objects.

Landmine classification

Landmines in Scene	TMM-1			TM-10			41/47			TMA-5			PMA-2A			Grenade		
	LBP	SIFT	SIFT*	LBP	SIFT	SIFT*	LBP	SIFT	SIFT*	LBP	SIFT	SIFT*	LBP	SIFT	SIFT*	LBP	SIFT	SIFT*
TMM-1	6	6	6															
TM-10				9	7	8												
41/47							1	0	0									
TMA-5										1	0	0						
PMA-2A													1	1	1			
Grenade																1	2	2

Test case 3

Training data: A, *Nr* = 5.

Test data: B, *Ns* = 14 (8 mines, 6 non-mines).

Target Classification

Test case 3	**Landmines**				**Not Landmines**			
	LBP	**SC**	**SIFT**	**SIFT***	**LBP**	**SC**	**SIFT**	**SIFT***
Known (trained) Land mines	8	6	7	7	0	3	2	2
Unknown objects	3	2	3	3	3	3	1	1

Observations:

Our expectation of Test case 3 was similar to Test case 4. But in fact the result from the confusion matrix is not as good as for Test case 4. The main reason is the large change of illumination conditions between data set A and B, which is much stronger than in Test case 4. Furthermore, due to similarity between unknown objects and trained objects (e.g. stub and TMM-1) the number of FP is larger than Test case 4.

Landmine classification

Landmines in Scene	TMM-1				41/47				Grenade				TM-10			
	LBP	SC	SIFT	SIFT*	LBP	SC	SIFT	SIFT*	LBP	SC	SIFT	SIFT*	LBP	SC	SIFT	SIFT*
TMM-1	1	1	2	2												
41/47					1	0	0									
Grenade									1	0	1	2				
TM-10													5	5	4	4

A.2 Test case 4

Train data: C, *Nr* = 6.

Test data: D, *Ns* = 9 (7 mines, 2 non-mines).

Target Classification

Test case 4	Landmines				Not Landmines			
	LBP	**SC**	**SIFT**	**SIFT***	**LBP**	**SC**	**SIFT**	**SIFT***
Known (trained) Land mines	6	4	3	4	1	3	4	3
Unknown objects	0	0	0	0	2	2	2	2

Observations:

Based on the confusion matrix of Test case 4; six mines from dataset C and nine unknown objects (including two non-mines and six mines) from dataset D are trained and tested on. One should note that the illumination conditions and the scene aspect angle vary between the scenes, but the type of scene is the same (‘Clear-cut forest’). Hence due to similarity of type of scenes, this test case has the best result among the test cases.

Landmine classification

Landmines in Scene	TMM-1				PMR-2A				Grenade				TM-10				TMA-5			
	LBP	SC	SIFT	SIFT*	LBP	SC	SIFT	SIFT*	LBP	SC	SIFT	SIFT*	LBP	SC	SIFT	SIFT*	LBP	SC	SIFT	SIFT*
TMM-1	2	1	1	2																
PMR-2A					1	1	1	1												
Grenade									0	0	0	0								
TM-10													2	2	0	0				
TMA-5																	1	0	1	1

A.3 Test case 5

Training data: A, C, E, *Nr* = 23.

Test data: B, D, *Ns* = 23 (17 mines, 6 non-mines))

Target Classification

Test case 5	**Landmines**			**Not Landmines**		
	LBP	**SIFT**	**SIFT***	**LBP**	**SIFT**	**SIFT***
Known (trained) Land mines	13	11	11	4	5	5
Unknown objects	3	6	6	3	0	0

Observations:

The thresholding effect is obvious in this test case where the number of TN becomes zero for SIFT and SIFT*. LBP works better then the other methods due to using the novel sub-blocking method and also better thresholding on the classifier. Therefore the number of FP and FN with LBP are less than SIFT and SIFT*.

Landmine classification

Landmines in Scene	TMM-1			TM-10			41/47			TM-49			PMA-2A			Grenade		
	LBP	SIFT	SIFT*	LBP	SIFT	SIFT*	LBP	SIFT	SIFT*	LBP	SIFT	SIFT*	LBP	SIFT	SIFT*	LBP	SIFT	SIFT*
TMM-1	3	3	4															
TM-10				6	5	5												
41/47							1	0	0									
TM-49										1	0	0						
PMA-2A													1	1	1			
Grenade																1	2	1