Landmines detection using multispectral images

Ivo Fernando Fontes Linhas Guerra, 86213, MEEC
Academia Militar/ Instituto Superior Técnico
e-mail: guerra.iffl@gmail.com

***Abstract*—This paper explores the detection of landmines using multispectral images acquired in military context. The conditions in which the images are obtained have a direct influence on the methods used to perform the automatic detection of landmines through image processing techniques.**

Two methods are proposed, one using traditional classifiers and the other using Deep Learning methods, namely a Convolutional Neuronal Network (CNN). In the first methodology, classifier fusion techniques are also used to understand their potentialities. The performance was evaluated according to the number of features the type of landmine, the environment and the depth of the mine. In deep learning, a study was carried out according to the feature map and regarding the type of landmine and the environment.

A quantitative analysis shows that using traditional classifiers gives overall accuracy (OA) above 97% in indoor and outdoor environments for the detection of land mines up to a given depth tested. It has been shown that the robustness of some classifiers, when exposed to specific standards (ie. only buried mines), has a decreased performance, however, the fusion of classifiers is constant, 97.9% for 0 [mm] and 96.0% for 1- 50 [mm], suppressing this fact. The adopted deep learning methods present an increase in these values for larger mines and a decrease for smaller ones.

These experimental results shed light into the factors that influence the detection of mines and into the merits and demerits of CNN based classification compared with classical methods.

***Index Terms*— landmine, detection, classifier combination, deep learning, Convolutional Neuronal Network**

I. Introduction

THE problem of landmine clearance is timely, complex, and demanding due to a multiplicity of factors to consider at the time of detection. Because of an increasing number of war zones and conflicts worldwide, the menace of landmines and unexploded ordnances is becoming a very serious problem that is going to affect the involved countries for years to come [1]

According to the Nato Standardization Agency (NSA) [2] there are two types of demining. The first, during military operations aimed at the purpose of the military commander, namely to achieve the principle of freedom of movementThe second type of demining appears in operations outside Article Five[1] (in peacetime) where demining operations require greater precision and minimal acceptable risk. In the operational context and during military operations, the development of a methodology for the landmine detection using multispectral sensors facilitates the detection of minefields using sensors of small size and small energy consumption [3] [4].

In the humanitarian context, the advantages are essentially the minimal risk that this method presents, ie the ability to be used remotely without risk to the operator, the possibility of its use in a variety of humanitarian and peace support operations.

Thus, the motivation to develop a solution that meets existing needs, using new technologies and developing new methodologies for the well-being of the civilian population as well as for the increase of the Portuguese Army's proficiency within the scope of this matter.

II. State of Art

In this chapter we present the state of the art regarding research work in the proposed area of work. The research focused on the different existing approaches to landmine detection. These techniques were grouped into five groups / families according to their basic operating characteristics, which include electromagnetic technology, acoustic / seismic technology, mine-based explosive technology, and technologies with physical contact

A. Electromagnetic Technology

Electromagnetic technology corresponds to methods that use electromagnetism, electromagnetic spectrum, or electromagnetic induction as the basis of detection.

Images in the infrared band are often used in the detection of buried objects. Krilov [5] states that this method is based on different rates of heat release by buried objects compared to their surroundings throughout the day. Using infrared cameras, it is possible to collect the contrast between the objects and their neighborhood [6], as shown in figure 1.

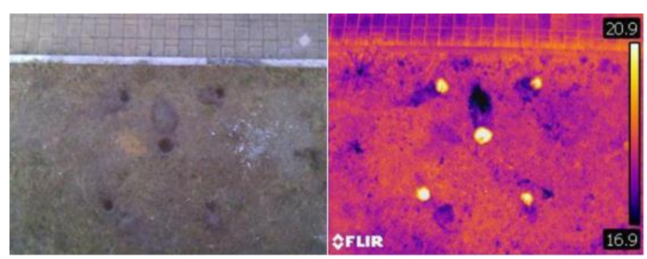


Fig. 1. Thermal infrared images obtained on a terrain with sparse and low vegetation where five objects are arranged. On the left is the actual image of the objects and on the right the contrast of the rate of release of heat between the objects and the surrounding environment (figure taken from [6]).

[1] Article 5 of the North Atlantic Treaty requires Member States to assist any member who is subject to an armed attack.

Using the same *modos operandi* of the previous method, the multi-spectral images use different bands of the electromagnetic spectrum. Makki [7] states that the purpose of this technique is to differentiate a landmine from its neighborhood into a multi-spectrum image using Visible Narrow Infra-Red (VNIR), Short Wave IR (SWIR) and Thermal IR bands. This author also makes a precise revision of past projects that used multispectral images, mentioning that knowing the spectral curve of the landmines it is possible to know in which place they are.

The methodology presented by Suganthi is based on a neural back-propagation network. Gray Level Co-occurrence Matrix (GLMC) is used as input to the neural network. The processing performed includes contrast enhancement, filtering, segmentation, feature extraction and classification.

There are also a large number of methods that use this technology in their operation such as the X-ray diffraction method using a process of emitting and receiving X-rays, the Electrical Impedance Tomography, which uses electric currents to represent the conductivity distribution of the medium to be investigated, the Magnetic Anomaly Method that is based on the detection of changes in the magnetic field of an object, with ferromagnetic characteristics, when exposed to an external magnetic and finally the method of electromagnetic induction that is based on the fact that when a time-varying magnetic field is established in the vicinity of a conductive object, an electric field is induced in the conducting object which causes a change in the charge flow within the object [4] [8].

B. *Acoustic / Seismic Technology*

These methods are based on sound waves, the phenomena of their reflection and seismic waves from the interior of the earth. The main acoustic / seismic methods are - Ground Penetration Radar, Acoustic Seismic Reflection and Rayleigh Scattering.

The Ground Penetration Radar method detects buried objects by radio wave emission in the direction of the ground and subsequent analysis of the signal that is returned [4].

The acoustic seismic approach is based on the principle that sound waves emitted by a source under the ground are reflected from the boundaries of structures and objects. Basically, low frequency sound waves 1 (less than 1 Khz) are emitted to the ground, collected later by sensors and analyzed for anomalies in their period. The dispersion of Rayleigh waves by landmines is a new method explored by Krylov [5]. This author states that the fact that these waves are efficiently dispersed by irregularities of any surface causes these properties to be exploited from the detection of landmines.

C. *Mine-based explosive technology*

In these methods, the existence of explosives such as RDX (nitrogen-based), cyclotetramethylene tetranitramine (HMX), pentaerythritol tetranitrate (PETN) and TNT (nitrogen-hydrogen-based) are decisive for the detection of mines.

The methodology proposed by Ege [4] is given by the name of Nuclear Quadrupole Resonance (NQR). This method is described as a special radio frequency technique based on the detection of the isotope of Nitrogen (^{14}N) found in the structure of many explosives and drugs.

A different methodology, proposed by MacDonald [8], uses a neutron beam and involves distinguishing the explosive constituent of the mine from the ground. For this, neutrons are sent to the ground that will later induce the atomic nucleus of the explosive. The differences in the intensity of the received radiation may thus indicate the presence of explosive. Another approach is the detection of vapors released by explosives.

Biological detection (using dogs, mice, bacteria, among others) involves the use of mammals, insects or microorganisms in the detection of the explosive constituents of the mine. These methods have the potential to reduce false alarms on metal clusters, but there are inherent difficulties in training the animals and detecting certain types of mines [9].

Detection by chemical methods, which may be termed Fluorescent, Electrochemical, Piezoelectric, are based on the excitation of the vapors released by the explosive components of landmines [10].

D. *Technologies with physical contact*

The last step in mine detection is the hand probe. The probe operator, called sapper, through years of training and experience learns to distinguish between a mine and another type of buried object. These methods endanger human lives, depend on the experience of the operator, and must be rigorous to be approved by the community.

In the operational context and in military operations, according to the United States Army and NATO doctrine [11] Clearing Operations are designated to clear/neutralize all mines or obstacles of a given route or area.

III. Landmines and Soils

This chapter explains in a somewhat more technical and doctrinal way the explosive devices in question, which are the most common, their types, their employabilities, and then the types of soils in which they are mostly present. It should be noted that all present information has been removed from non-confidential documentation and is available for research.

A. *Landmines*

Based on the NATO doctrine and the United States Army School of Engineering [11], a mine is an explosive device used to destroy or incapacitate people or land vehicles, boats or aircraft. An area of land containing landmines arranged in patterns or randomly is called minefields. Minefields are designed to disorganize, channel, retard or deter the enemy and can be employed in three ways: terrain-oriented, for the situation and for the target. There is a high categorization of mine types, depending on the purpose for each mine. Thus, landmines can be divided into three different types: real, simulated and instructional. Real landmines according to purpose can be classified as anti-tank landmines (AT) and antipersonnel landmines (AP) [12].

.

B. *Soils*

The thermal signature of landmines depends on a set of environmental conditions, and the soil properties play a decisive role in the detection of landmines. To compare the influence of different types of soils and their composition, and even more important, to allow the implemented system to be as generalized as possible (important aspect in machine learning problems), several types of soils were used. as possible in order to try to cover all environments where landmines can be used

Generic soils of simple composition were used and a mixture of soils also were used in order to simulate low traffic roads (so-called non-tarred roads), which are very common in underdeveloped countries[2] with propensities to conflict or have left a recent conflict.

IV. Methodology

This chapter explains the procedures adopted to implement a solution to the problem. Two methodologies have been applied, a classic that follows, in general, the phases of a pattern recognition problem and another one that uses Deep Learning techniques as an essential tool.

A. Initial procedures

In this phase are acquired multi-spectral images. The obtaining of these images may be accomplished by various imaging equipment or by imaging equipment capable of obtaining images at various spectral intervals. These images are obtained according to certain parameters according to the equipment itself, the environment in which the images are obtained, or the objects to be detected. Then it is common to perform the alignment of the images, if they are misaligned or are obtained by different equipment with different fields of vision or different resolution. This alignment is performed manually, highlighting the need for the use of marks / targets of visible material in the different spectra of the images obtained to mark the minefield, for example, using marks constructed of aluminum, capable of being easily identified in the different images.

B. Extraction and Feature Selection

In this phase the regions of interest (ROI) are defined. This setting is made automatically from parameters obtained in the image itself or set manually.

The definition and respective extraction of features for the requirements of a problem is a fundamental step in automatic learning tasks, such as the classification of standards, so in this phase an analysis is done to the data obtained in the identification of ROIs in the sense of choose which features can be extracted from ROIs.

1) *First Order Gray Level Statistics.*

According to Gonzalez [13] the analysis using first order statistics (FOS) is based on the histogram of gray levels. Assuming that n_p is the number of pixels of a region of interest and L the number of gray levels of the ROI, the normalized first-order histogram is given by the probability distribution function in equation (1).

$$h(i) = \frac{\#\{(x,y) \in I(x,y) = i\}}{n_p}, \qquad 0 \le i \le L \tag{1}$$

From the normalized histogram, the various first order features such as the mean, standard deviation, variance, Entropy, symmetry, kurtosis and energy are proposed.

2) *Second Order Gray Level Statistics.*

First-order statistics reflect features that do not consider the spatial distribution of gray levels in the image and can therefore be presented as limited metrics. Second order statistics consider the spatial distribution of gray levels in the image. The Spatial Gray Level Dependency Method (SGLDM) helps to extract the gray level co-occurrence matrix (GLCM). This matrix is based on the second order conditional probability $p(i,j\,|d,\theta)$ density function that can be estimated in several directions θ and at various distances d, being i and j gray levels. These functions can be represented as GLCM according to equation 2 [14] [15]

$$\Omega(d,\theta) = p(i,j\,|d,\theta), \qquad 0 \le i,j < L \tag{2}$$

The matrices $\Omega(d,\theta)$ are the basis for calculating several statistical measures, and for each pair (d,θ) is calculated a matrix $\Omega(d,\theta)$ and a set of texture descriptors such as contrast, energy correlation, entropy and homogeneity.

3) *Higher order statistics*

The Gray Level Run-Lengh Method (GLRLM) method consists of counting the number of pixel sequences with the same intensity in a given direction. Primarily the primitive matrices $\Psi(\theta)$ are calculated (equation 3) from which it is possible to extract texture descriptors [16].

$$\Psi(\theta) = M(a,r|\theta), \qquad 0 \le a \le L, \qquad 0 < r \le N_r \tag{3}$$

Each element of matrix $M(a,r|\theta)$ represents the number of times occurring in the primitive ROI with gray level a and length r according to the direction θ. From this matrix it is possible to extract the following texture descriptors: Long Run Emphasis (LRE), Gray-Level Nonuniformity (GLN), Run Length Nonuniformity (RLN), Run Percentage (RP), Low Gray-Level Run Emphasis (LGRE), High Gray-Level Run Emphasis (HGRE), Short Run Low Gray-Level Emphasis (SRLGE), Short Run High Gray-Level Emphasis (SRHGE), Long Run Low Gray-Level Emphasis (LRLGE) and Long Run High Gray-Level Emphasis (LRHGE) [16].

4) *Feature selection and normalization*

In classification problems, after the feature extraction step it is common to have a feature selection process that aims to reduce the size of the data set. In this context the features selection algorithms are separated into three categories: Filters methods which extract features from the data set without regard to classification or any other method of learning as a criterion; Wrappers methods that use classifiers / learning techniques to evaluate which features are statistically relevant; Embedded methods that aim to combine the advantages of the two previous

[2] Examples of these countries are: Afghanistan, Syria, Central African Republic, Kosovo, Mali

methods [17].

One of the algorithms of the filter methods, widely used is the Relief algorithm. The original Relief algorithm estimates the quality of features according to how well their values differ between patterns that are close to each other. The family of these algorithms is especially attractive because they can be applied in all situations, it is easy to implement in problems in which several classifiers are used, it includes interactions between features and it can capture local dependencies among them that other methods cannot [18].

C. *Classification*

The problem of discriminating landmines from the background is a binary problem. In order to solve it, several classification techniques are proposed: Neural Networks, SVM Classifier, Decision Trees, KNN Classifier and Linear Classifier using PCA.

1) *Artificial Neural Networks*

Artificial Neural Networks are mathematical models governed by the principle of biological neural networks. The choice of the number of neurons that constitute the network will depend on the classification process to be performed. There are several ways to connect artificial neurons to create a neural network, but the most common is the feedforward network. Each input data in the neuron has a weight, this simply represents a floating number that is adjusted when training the network, assuming positive or negative values, in order to provide activating or non-activating influences at each input data [19].

In order to enhance these neural networks, two important extensions have been proposed, one of which is the use of multilayer neural networks, generally organized in layers, the so-called multilayer percepton (MLP) and the other extension was the introduction of differential and continuous activation functions [19].

2) *Support Vector Machines*

The Support Vector Machines (SVM) algorithm is a supervised learning methodology used for statistical classification and regression analysis. Represents an object classifier according to its features, based on the concept of planes that define decision boundaries. A decision plan separates sets of objects from different categories. For a given set of training data, SVM constructs an iterative model that will correctly predict whether a new object belongs to one category or another [19].

The simplest situation corresponds to a training set in which the data are linearly separable. However, in some classification problems the data distribution does not allow a linear separation between the classes. This problem is often solved by mapping the data to a space of greater dimensionality called a feature space where it can thus be linearly separated. This mapping occurs using kernel functions such as linear, Radial Basis Function, and Polynomial.

3) *K-nearest Neighbors*

The k-nearest neighbors (KNN) classifier is one of the simplest, most used classifiers and presents good results in solving classification problems. Given a test sample, this classifier assigns a class based on the calculation of the distance of this sample to the nearest samples of the training set. After analyzing the class of the chosen samples, through a voting system, the class with the highest absolute frequency [20].

The performance of this classifier depends essentially on the number of neighbors to consider and the metric of calculation of chosen distances that can be Euclidian, correlation, Cityblock, Chebyshev [20]

4) *Decision Trees*

Decision trees are one of the most practical and most used models, namely in research operations and decision analysis. The Decision Trees classifier performs a series of mathematical questions / comparisons about the features of a data set. Each time a response is received, a new question is raised until the classification of the data is obtained. The classifier organizes the series of questions and conditions in a tree structure [21].

Beginning with the root node, the test conditions apply to the data and continue by the appropriate branch, based on the result of that condition. When the terminal node is reached, the classification associated with that node is assigned to the test data [21].

5) *Principal Component Analysis*

According to Almeida [22] when there is a data set with a large number of dimensions, it is sometimes advisable to have the ability to reduce the number of dimensions while maintaining the same amount of information, thus facilitating the analysis of data with high dimensionality.

Principal Component Analysis (PCA) finds a linear function that allows the separation of the different classes of the training set realizing the projection of the data in the so-called eigenvectors [22].

The advantages of using dimensionality reduction methods are decreasing data size, and decreasing processing time. The size of the features dimensions is determined by each classifier with the best performance at the classification level [20].

D. *Fusion of Classifiers*

A strategy to improve the overall classification performance is to combine several classifiers into a single classifier (multi-classifier) according to the output of all classifiers used. Two plausible methods are used at this stage: methods of majority voting and methods of heavy voting.

1) *Majority Vote*

The majority vote method (MV) consider the classification obtained for each classifier used, recourse to a vote, which consists of finding which of the classifications occurs most frequently, assigning it to the multi-classifier. Given the simplicity that the majority voting method presents, the performance of each individual classifier is not considered, given this fact it was considered the next method [23].

2) *Heavy Vote*

To improve the performance of the multi-classifier, there is a method of combining classifiers that assign a dynamic weight (HV), which is proportional to the performance of each classifier. Thus, classifiers that present an individual behavior of low performance have lower weight and consequent less importance in the global classification [23].

E. System Performance Measures

The quality of the algorithm can be calculated from a confusion matrix allowing the visualization of the performance of a learning algorithm in a specific table where information about the actual classifications and those predicted according to a classification algorithm [19].

With the aid of this matrix it is possible to calculate measures that characterize the performance of the different algorithms, being these, sensitivity, specificity, precision, overall accuracy (OA), F-Score among others. It should be noted that depending on the different specificities of each classification problem, it may not be efficient or useful to use / calculate all these performance measurement measures, and there are still others that were not mentioned [24].

F. Deep Learning

Deep Learning is a current trend in data analysis and learning techniques. This technique is a type of machine learning that performs classification tasks directly from images, video, texts or sounds. In the literature, Deep Learning is characterized as an improvement to the artificial neural networks, consisting of a significant increase of layers, which provides a higher level of abstraction and improvements in the predicted data. It is thus considered as the main automatic learning tool in the general fields of computer vision and image processing [25].

There are no studies in the literature of the application of Deep Learning in the detection of mines in multi-spectral images, however it is known that in deep learning, most of the algorithms use the CNNs [26].

The reason CNNs are currently the most investigated machine learning algorithms is that they preserve spatial relationships when input images are filtered. These spatial relationships are of crucial importance in detecting differences in images, for example in medical imaging analysis this information is used to distinguish, among other things, lung tissue from a cancerous tissue [25].

1) Convolutional Neural Network

The networks used in Deep Learning have more layers compared to the classical neural networks, particularly the CNNs can be constituted by tens or hundreds of layers, each one being trained and responsible for detecting different features in each image. In practice what happens is that filters are applied at different resolutions to each training image and the output of each convoluted image is used as input to the next layer. These filters begin to produce simple features such as brightness or corners / lines and increase complexity for unique features that define the object [25].

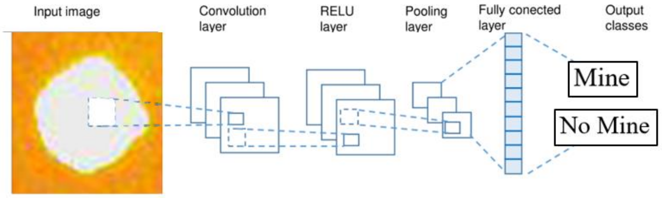


Fig. 2. Example of the classification task in the detection of landmines according to a CNN scheme. (figure adapted from [26])

As can be seen in figure 2, the flow of a CNN is initiated with the input of an image being exposed to layers of features extraction via convolutional layers, Rectified Linear Unit (RELU) layers and Pooling layers. The output of this transformation later feeds a final layer called the Fully Connected Layer that assigns the values or probabilities, thus classifying the input image into the class with the highest value or probability [26].

V. Results and discussion

This chapter presents the results of the two methodologies applied to the detection of landmines. Firstly, the procedures for acquiring multi-spectral images are described. The qualitative and quantitative classification results for the images obtained in the fields constructed in the laboratory (indoor) and in the fields (outdoor) constructed in the Military Academy in Amadora (AAMA) are presented below. The results are compared according to the depths to which the landmines were buried, the type of soil in which they were buried and the type of classifier. At the end of the chapter the results of the Deep Learning study of the CNN network are presented and the performance of this method compared to a classical methodology is analyzed.

A. Image acquisition process

The images were obtained using imaging equipment from the Military Academy, acquired during the execution of the FUSIMIL and FIVE projects. These equipments are, a Quest Condor3 VNN-618 (Multi-Spectral Camera) with 640 × 494 [pixels] resolution, (400-670); (670-850); (850-1000) [nm] spectral band and Sony ICX-618 CCD sensor, 1/4 ", 4.08 [µm] and a FLIR T440bx (thermal infrared camera) with 320 × 240 [pixel] resolution and RGB + (7500 -13000) [nm] spectral band.

In order to carry out both experiments, it was necessary to construct a metal structure to support the two image acquisition cameras at a fixed and predetermined height, and also the construction of containers, in this case plastic, for the placement of the different soil types and the respective thermal insulation between the soils and the containers.

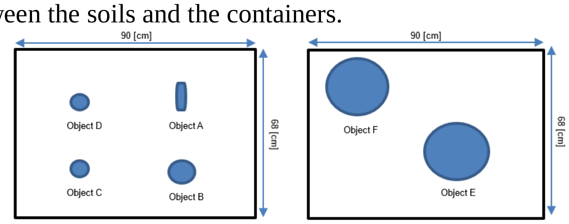


Fig. 3. Diagram of the minefields built for the acquisition of images. On the left, the diagram for the AP mines. On the right is the diagram for the AT mines.

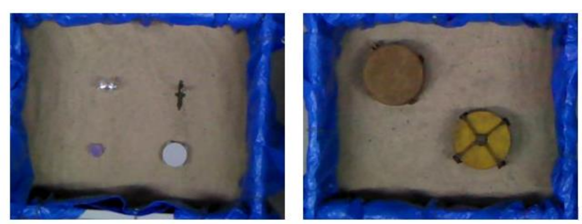


Fig. 4 Practical representation of the minefield diagram with its surface objects. On the left, the diagram for the AP mines. On the right is the diagram for the AC mines.

After the cameras were positioned and the remote communication systems were operational different soil types were introduced into the various containers, and two main diagrams were developed for the construction of the minefields.

The first diagram was used for the experiments with AP landmines and the second one for the experiments with AT landmines as shown in the figure 3 and 4.

B. Data acquisition

The data acquisition process corresponds to using two previously mentioned equipment. The images obtained from the top of the metal structure at a height of 2.3 [meters] show the entire contents (soil and buried objects) inside the container

Images taken from the FLIR T440bx infrared camera have an image resolution of 320 × 240 [pixels] in which each acquisition corresponds to an image of the visible spectrum and the corresponding image in thermal infrared spectrum. This one has a temperature scale which was always fixed according to the type of soil to be used. The choice of the values of the temperature scale was made mainly according to the general conditions of the environment in which the images were inserted.

The images obtained from the Quest Condor3 VNN-618 multi-spectral camera have an image resolution of 640 × 494 [pixels] and each acquisition of this equipment corresponds to three different images. The first (Channel 0) corresponds to an image of the visible spectrum in gray levels, with wavelengths in the range 400-670 [nm], the last two images (channel 1 and channel 2) correspond to two spectra belonging to VNIR with channel 1 being in the range 670-850 [nm] and channel 3 being in the range 850-1000 [nm].

Regarding the settings made in this camera, in indoor there is a need to set high values in the exposure time due to the low light in the room. This increase in exposure time and high gain consequence correspond to an increase of the present noise, mainly in the image of channel 2 having consequently the decrease of useful information present in this spectral band. Regarding the outdoor environment, it is possible to assign lower values in the settings due to the increase of luminosity compared to the interior of the room. It is considered an effective reduction of noise especially in channel 2. However, in outdoor environment there is a need for a constant change of these values in relation to the current meteorological conditions.

After the acquisition of all the images by the equipment indicated in the previous section, it was necessary to pre-process them, to constitute the data set. By visual inspection it was verified that from a given depth, in these tests, it was not possible to detect buried objects in any of the different tested specimens, and the images acquired from that limit depth were taken from the data set, starting of the assumption that from these depths the system does not detect buried objects. For each case, the limit depths are respectively: Indoor / outdoor fields AP diagram: 5 [mm] for all types of soils. Indoor / outdoor fields AT diagram: 10 [mm] for river sand, sea sand and organic soil and 100 [mm] for granular soil

For the creation of the multi-spectral image, the images of the visible and infrared spectrum obtained by the FLIR T440 camera were selected, the image of the visible spectrum in gray levels of channel 0, the image of channel 1 and the image of channel 2, of the Quest Condor3 VNN-618 camera were also selected obtaining a 7-dimensional image (Table I).

TABLE I
SPECTRAL BANDS OF EACH COMPONENT OF THE MULTI-SPECTRAL IMAGE CONSTRUCTED

Dimension N. º	Spectrum	Input/Processing
1, 2, 3	Visible (RGB)	No
4	TIR (gray scale)	Yes
5	Visible (gray scale)	Yes
6	VNIR (670 – 850 [nm])	Yes
7	VNIR (850 – 1000 [nm])	Yes

As input data set dimension 1, 2 and 3 (visible image) do not enter the processing, being only present as control image and comparison with the amount of information that the other specs offer. The fact of inserting dimension number 5 in the processing corroborates the non-introduction of the visible image and is due to the large amount of information obtained by these images in the detection of partially buried mines

Then, the images were rotated without the same orientation, the downscale process was performed, the resolution of the images obtained by the multi-spectral camera was reduced, and then the alignment of all the images was obtained, obtaining a final data set where all dimensions are aligned, with a final resolution of 240 × 180 [pixels]. The following figures represent examples of the final multi-spectral image constructed.

Fig. 5. Example of multi-spectral image, for Sea sand, in indoor environment, with objects on surface, and AP diagram

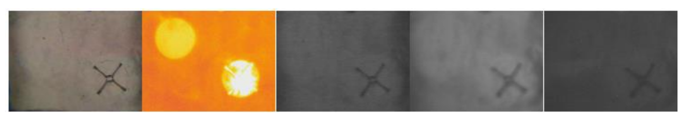


Fig. 6. Example of the multi-spectral image, for River sand, in indoor environment, with objects buried at depth 1mm, and AT diagram

Next, a sliding window algorithm was implemented, which consists of extracting multi-spectral ROIs from a given multi-spectral image to obtain multiple images in which the feature extraction is performed to each of the multi-spectral ROIs. For the AP diagram the multi-spectral ROI size was set to 10 pixels and ROIs were extracted 2 by 2 pixels, for the AT diagram was defined as multi-spectral ROI size 80 pixels and ROIs were extracted 2 by 2 pixels. For the purpose of evaluating the performance of classification systems, a set of data is required. From the totality of multi-spectral ROIs and the need for the constitution of a balanced data set it was possible to obtain a final data set present in the following table.

TABLE II
FINAL DATA SET

Environment	Diagram	Data Set	Train Set	Test Set
indoor	AP	10262	8723	1539
	AC	29984	25487	4497
outdoor	AP	7056	5998	1058
	AC	11694	9940	1754

Given the large data set and number of ROIs as validation method, we used holdout validation using 15% of the data set as the test set, with the rest being 85% reserved for the training set.

C. Extraction and Feature Selection

The next step is the extraction and selection of features. First, second and higher order features were extracted from each multi-spectral ROI making a total of 264 features. Then the features were normalized, and because this process is not of extreme importance it was decided doing a simple normalization with a maximum value of 1 and a minimum of 0.

By getting the normalized data, the next step was to perform a feature selection experiment using a filter method. Because this methodology relies on multiple classifiers, the use of other types of methods that are based on the performance values of the classifiers would make the process significantly longer, complex, and require a longer processing power time. With this type of methods, it was possible to perform a study of the selection of features without having to perform multiple training sessions in the classifiers.

Using the Relief algorithm, it is verified that there is no uniformity in the type of features that has the greatest weight of importance. It can be observed that the higher-order features assume a greater importance in the case of AT mines. An analysis of the spectra already shows the fact that the most important spectrum according to this algorithm is thermal infrared. In all cases, this spectrum was the one that obtained higher average weight in relation to the other spectra and this analysis is confirmed by visual inspection of the different dimensions of the multi-spectral image constructed.

D. Classification results

From the final data set, several classifiers were trained, being those that, in preliminary tests and using the toolbox classification learner, obtained higher OA values. These classifiers are SVM with Cubic core, SVM with Gaussian core, Fine KNN, Medium KNN, Fine Tree and Bagged Tree (Set of decision trees). We also used a simple neural network with two layers, where the inputs are the features. The configuration parameters of each of the classification processes are given in the following table.

TABLE III
SUMMARY OF THE CLASSIFIERS USED IN THIS METHODOLOGY, ITS TYPE AND THE PARAMETERS USED IN EACH.

Classifier	Type	Parameters
Decision Tree	Fine Tree	Max divisions: 100 Criterium: Gini
SVM	Cubic SVM	Kernel: Cubic Scale: Automatic
	Gaussian SVM	Kernel: Gaussian Scale: 4.1
KNN	Fine KNN	Neighbors: 1 Distance: Euclidian
	Medium KNN	Neighbors: 10 Distance: Euclidian
Ensemble	Bagged Trees	Learning: Tree Number of Trees: 30
Neural Network	-	Input: 264 features Hidden Layer: 10 Output layer: 1 Output: 2 classes

After all the classifiers were correctly configured, and training the classification algorithms with {26, 66, 132, 198, 264} features was finished, they were evaluated. It should be noted that, for a more efficient analysis, the Accuracy / Overall Accuracy metric was used to simplify the analysis of the results in each case.

1) Performance evaluation in indoor environment

For the indoor environment results were obtained for the detection of AP mines and for AT mines. For the AP diagram, from a data set constituted by 10262 multi-spectral ROIs the following results were obtained.

TABLE IV
RESULTS OF DIFFERENT CLASSIFIERS FOR INDOOR ENVIRONMENT AND AP DIAGRAM

Classifier	Type	Precision (OA) [%]				
		26	66	132	198	264
Tree	Fine	84.4	84.2	85.8	86.1	87.0
SVM	Cubic	90.4	95.0	94.9	94.9	96.4
	Gaussian	92.3	96.5	96.4	94.9	97.6
KNN	Fine	89.7	93.7	92.6	93.8	94.1
	Medium	88.5	91.9	91.5	92.3	92.6
Ensemble	Trees	94.1	95.6	95.1	95.8	96.4
Neural Network		75.0	87.9	89.4	86.6	90.4

For the AC diagram, from a set of data consisting of 29984 multi-spec ROIs the following results were obtained.

TABLE V
RESULTS OF DIFFERENT CLASSIFIERS FOR INDOOR ENVIRONMENT AND AT DIAGRAM

Classifier	Type	Precision (OA) [%]				
		26	66	132	198	264
Tree	Fine	91.9	91.5	94.4	94.1	93.8
SVM	Cubic	98.4	98.8	98.8	99.0	99.0
	Gaussian	97.2	97.6	97.9	98.0	98.4
KNN	Fine	97.9	98.0	98.3	98.3	98.4
	Medium	97.3	97.1	97.8	97.1	97.9
Ensemble	Trees	98.5	98.7	98.8	99.1	99.1
Rede Neuronal		94.6	97.3	97.4	97.9	97.8

From the analysis of the results presented in the tables IV and V it is observed that using only a single classifier, results are quite promising in the detection and the possibility of reducing the dimensionality of the data set. It has been found that, in general comparative terms, higher precision values are obtained in the detection of AT mines in relation to the AP mines. In general, it was verified that the Ensemble Bagged Tree classifier was one of the classifiers that obtained the best results, with the decision Trees achieving the worst performance

For the AP diagram, in quantitative terms, the maximum accuracy results of 97.6% were obtained for the Gaussian SVM classifier with a characteristic vector of 264, but it is possible to reduce dimensionality with the achievement of precision values of 96.5% and 95.6% for the Gaussian SVM classifier and for Ensemble classifier respectively, both with the 66.

For the AT diagram, results very close to the total detection were obtained, with maximum results of 99.1% and 99.0%

accuracy for Ensemble classifier and Cubic SVM respectively and for a characteristic vector of 264. Interestingly decreasing the number of features in 50 % we obtained the same maximum results as described above

2) Performance evaluation in outdoor environment

For the outdoor environment results were obtained for the detection of AP mines and for AT mines. For the AP diagram, from a set of data consisting of 7056 multi-spectral ROIs the following results were obtained.

TABLE VI
RESULTS OF DIFFERENT CLASSIFIERS FOR OUTDOOR ENVIRONMENT AND AP DIAGRAM

Classifier	Type	Precision (OA) [%]				
		26	66	132	198	264s
Tree	Fine	86.7	86.7	85.0	84.3	85.4
SVM	Cubic	94.8	93.9	97.0	96.0	94.0
	Gaussian	95.7	94.9	97.5	97.3	97.1
KNN	Fine	93.3	93.2	94.9	95.1	94.7
	Medium	92.3	91.5	92.5	92.5	93.3
Ensemble	Trees	94.5	93.9	95.4	94.7	95.1
Neural Network		74.0	89.6	84.4	84.4	81.8

TABLE VII
RESULTS OF DIFFERENT CLASSIFIERS FOR OUTDOOR ENVIRONMENT AND AP DIAGRAM

Classifier	Type	Precision (OA) [%]				
		26	66	132	198	264
Tree	Fine	90.5	92.4	90.9	93.9	92.3
SVM	Cubic	98.3	98.5	97.9	98.5	98.7
	Gaussian	97.2	96.8	96.7	97.0	97.0
KNN	Fine	97.5	97.9	97.9	97.6	97.4
	Medium	95.9	96.0	96.3	95.9	96.1
Ensemble	Bagged	98.2	98.2	98.4	98.2	98.2
Neural Network		93.6	93.8	96.3	96.8	95.4

From the analysis of the results presented in the tables VI and VII it is observed that these prove and corroborate the results obtained in indoor environment, obtaining, however, a general decrease in the values of precision. This was mainly due to the reduction of the data set. In indoor, 40246 ROIs were used from 45 multi-spec images and in the outdoor area 18750 ROIs (about 46.5% of the number used in indoor) were used. As in the indoor environment, higher accuracy values for the detection of AT mines in relation to the AP mines.

For the AP diagram, maximum accuracy results of 97.5% were obtained for the Gaussian SVM classifier with a characteristic vector of 132. This result may show that some of the features extracted in outdoor environment do not have statistical value / information and may be to impair the detection by the classifiers.

For the AT diagram, maximum overall accuracy results were obtained greater than in the AP diagram, with maximum results of 98.7% for SVM Cubic with a feature vector of 264. Overall, decreasing dimensionality, precision values do not suffer large changes being very close to the different classifiers.

3) Performance evaluation as a function of depth, compared to multi-classification.

Using the classification models described above, a study of the performance of the classifier was performed based on the depth and compared to the results of a multi-classification by majority vote (MV) and heavy vote (HV).

For this study, we selected random ROIs from multi-spectral images with objects at 0 [mm] and objects buried between the minimum depth of 1 [mm] and maximum 50 [mm]. As a test set, we considered 8512 multi-spec ROIs, not used in the training and test of the previous section, randomly chosen from both diagrams and both environments, for surface objects and for buried objects.

The non-use of the two-layer neural network was due to several reasons. The first one, due to the low OA values obtained in the previous section, the second due to the complexity of using the previously trained model to reconcile new predictions with the same and associate them with the predictions of the remaining classifiers and finally the third reason, the study conducted in the next section, in deep learning of a CNN being a more complex neural network.

Table VIII shows the accuracy values obtained for each of the classifiers depending on whether the mines are buried (1-50 [mm]) or surface (0 [mm]).

TABLE VIII
OA RESULTS FOR EACH OF THE CLASSIFIERS DEPENDING ON THE DEPTH OF THE LANDMINES

Classifier	Type	Precision (OA) [%]	
		0 [mm]	1-50 [mm]
Decision Tree	*Fine Tree*	87.4	73.6
SVM	*Cubic SVM*	96.5	92.0
	Gaussian SVM	89.4	87.2
KNN	*Fine KNN*	95.3	92.5
	Medium KNN	94.6	90.1
Ensemble	*Bagged Trees*	93.4	92.7

From the analysis of the results it is verified that, as expected, the detection of the objects at the surface obtains better results than the detection of the buried objects. This is mainly due to the greater amount of information, collected by the features, in the multi-spectral ROIs at 0 [mm].

The strategy that, according to the literature, allows to improve the overall performance of the classification is the combination of several classifiers. The following table shows the results obtained by the two classifier fusion methods.

TABLE IX
OA RESULTS FOR THE TWO METHODS OF MULTI-CLASSIFICATION

Classifier	Type	Precision (OA) [%]	
		0 [mm]	1-50 [mm]
Fusion	MV	97.2	95.9
Fusion	HV	97.9	96.0

The data obtained in this section allow an analysis of the robustness of the classifiers with respect to the depth and the detection of objects. The decision tree classifier is the one that obtains the greatest discrepancy between the two cases, in the order of 13.8% difference. The SVM classifier with the cubic polynomial kernel is the one that obtains the best results for the average of the two cases but presents less robustness when applied to ROIs from images with buried objects. It is also verified that the Ensemble classifier, although not the one that obtains greater general precision, behaves quite effectively and robustly to the selected subsets of different ROIs, presenting only a 0.7% decrease in OA when compared to the depths.

The implementation of methods of fusion of classifiers promotes quite promising results, being that this increase is more significant in the case of buried objects

E. Deep Learning results

The implementation of a deep learning method for mine detection presents itself as an innovative study not yet addressed in the literature. Thus, the methodology and parameters used in this method and the construction of CNN itself have undergone several changes and empirical experiments to obtain the best possible results.

The network configuration built for the experience, in the feature learning step is composed by: one input layer (size of ROI / dimensions) and three sets of: one convolution layer (filter size, number of filters), one batch normalization layer (used to normalize activations and propagation in the network), one ReLU layer and one pooling layer (down-sampling operation, not used in the third set). In the classification step we have: one fully Connected Layer (this layer is responsible for connecting all the neurons responsible for the features to classify the image), one Soft max layer (normalizes the output) and one classification layer. If we fix the first argument of the convolution layer (filter size) with a value of 3×3 pixels, the 2nd argument, the number of filters, refers to the number of neurons connected to the same input region and thus determines the number of feature maps, which can be varied several times in order to perform a comparative study of OA as a function of the feature map. Table X presents the results of the Deep Learning CNN network. Denote that The "n/computed" means that the time required (due to the iterations number) for the training and test this network is too high (more than 7h for training).

TABLE X
BEST OVERALL ACCURACY RESULTS AND NUMBER OF FILTERS FOR DIFFERENT ENVIRONMENTS (INDOOR/ OUTDOOR) ACCORDING TO THE DIFFERENT TYPES OF MINES (AP AND AT)

Deep Learning (CNN)				
1st layer	8 filters	16 filters	64 filters	256 filters
2nd layer	16 filters	32 filters	128 filters	512 filters
3rd layer	32 filters	64 filters	256 filters	1024filters
	Overall Accuracy			
Indoor AP	82.4	82.7	84.7	86.1
Indoor AT	95.5	97.8	96.7	n/computed
Outdoor AP	79.6	82.0	83.4	82.0
Outdoor AT	99.0	99.1	99.1	n/computed

It is seen that the best results are also related to the detection of AT landmines and that the ideal number of filters / features to be implemented is 64/128/256 respectively for each of the 3 different convolutional layers implemented. There is a 12-20% OA difference between detection of AP mines and AT mines, justified by the fact that AT landmines are larger than AP, allowing to extract more textural information. Regarding the processing time, it was found that the AT input ROIs had the size of 80×80 pixels, while for AP landmines the size of 10×10 pixels making the processing time approximately 5 times lower for ROIs of smaller size. This is because as we define 3×3 pixels the size of the filters for both input images, it is easy to understand that it is more time consuming to compute a filter over an entire 80×80 pixels AT image than an image of 10×10 pixels AP image.

VI. CONCLUSIONS

The issue of demining continues to be complex and demanding today, given the numerous factors to be considered at the time of detection, and allied to this complexity, there is still the constant danger slope that the sapper responsible for the inactivation and removal of the landmines

In the execution of the classification, the obtained results demonstrate a superior overall performance of the SVM classifiers in relation to the others, being this type of classifier is widely used in binary problems. From the results in the indoor environment it is concluded that there is a relevant dichotomy related to the Ensemble Bagged Trees classifier that obtained better performance and the decision tree that obtained the worst performance. It is concluded that for problems of two classes and given the variety of characteristics, the ensemble methods, which use techniques that combine several decision trees, produce better results than the use of only one decision tree. Regarding the outdoor environment, it is concluded that laboratory tests corroborate the results obtained in the billboard, however these are slightly lower, justified by the 46.6% decrease in the dataset used. Still on the classification, higher accuracy values are detected in the detection of AC mines than in AP mines, due to the size of AC mines and consequent increase in ROI, which contain more useful information in AC ROIs than in AP ROIs.

Regarding the depth-based study, better results were detected in the detection of surface / partially buried objects (0 [mm]) than in the detection of buried objects to the boundary depths. In a first analysis the results obtained for the new standards are slightly lower than those obtained by the test set used in the training phase.

The built CNN was based on a generic configuration with the change of certain parameters to elaborate a study in function of the same ones. It is verified that the approach of varying the number of filters in the convolution layer had the advantage of perceiving which feature set indicated when solving a given problem. In practice, it was found that a high number of filters for AC mines makes the network ineffective in terms of its time and performance, due to the enormous number of characteristics that this network would generate. There is thus a need to balance the number of filters for each specific problem in order to avoid network configurations that lead to high processing times, which are impossible to handle with traditional computing power.

As final considerations, the results obtained are quite promising, for both methodologies, it is verified that there is the practical potential of using the classifiers fusion. It has been shown that the use of a CNN in this type of problem needs to be well adjusted to the problem, the size of the input image / ROI, the number of filters and special attention to the size of the same. It was also verified in CNN that it is possible to find an optimal configuration to solve the problem in diagram AT.

Despite the results obtained, it is important to mention that given the complexity of the problem, it is still early to generalize and start to implement a system in one of these methodologies This is mainly due to the origin of the data set, which is done in a controlled environment

A. Future work

After the termination of a research work, there are always paths that have not been trodden as well as ideas that arise during the same. Some of these are feasible in the medium / long term: Increase and diversify the number of landmine images either buried or to the surface and obtain them in an operational context, and if possible in real situations; Test the system for all depths, even if by visual inspection it is not possible to verify differences between the signal transmitted by the objects and the ground; Carry out a study focused mainly and exclusively on the thermal infrared spectrum; At the practical and operational level, to test these methodologies integrated in a surveillance UAV system in cooperation with the Portuguese Air Force; Finally, in order to aid detection, the integration of this system with another method of detection of landmines

VII. Referências

[1] I. Makki, R. Younes, C. Francis e M. Zucchetti, “A survey of landmine detection using hyperspectral imaging,” *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 124, pp. 40-53, 2017.

[2] J. H. Eriksen, “Standardization Agreement, Countermine Operations in Land Warfare,” Nato Standardization Agency, Brussels, Belgium, 2002.

[3] J. Florez e C. Parra, “Review of sensors used in robotics for humanitarian demining application,” em *IEEE Columbian Conference on Robotics and Automation,* Colombia, 2016.

[4] Y. Ege, A. Kakilli, O. Kılıç, H. Çalık, H. Çıtak, S. Nazlıbilek e O. Kalender, “Performance Analysis of Techniques Used for Determining Land Mines,” *International Journal of Geosciences,* 2014.

[5] V. Krylov, “Detection of buried land mines using scattering of Rayleigh waves,” em *27th International Conference onNoise and Vibration Engineering (ISMA 2016),* Leuven, Belgium, 2016.

[6] J. Pimenta, “Identificação de minas terrestres em imagens de infravermelho térmico,” Dissertação de Mestrado Instituto Superior Técnico, U. Lisboa, Lisboa, 2015.

[7] I. Makki, R. Younes, C. Francis e M. Zucchetti, “Mathematical Methods for Hyperspectral Imaging in Landmine Detection,” em *Transactions of the American Nuclear Society,* vol. 112, San Antonio, Texas, 2015.

[8] J. MacDonald, “Alternatives for Landmine Detetion,” RAND, Santa Mónica, Califórnia, 2013.

[9] A. Mahoney, C. Cox e B. Weetjens, “Reinforcement for Operational Mine Detection Rats,” *the journal of Conventional Wepons Destruction,* vol. 17, pp. 58-62, 2013.

[10] L. Robledo, M. Carrasco e D. Mery, “A survey of land mine detection technology,” *International Journal pf Remote Sensing,* vol. 30, pp. 2399-2410, 2009.

[11] Deportment of US Army, “Explosive Hazard Operations,” US Army Enginneer School, EUA, pp. 2.1-2.11, 2007.

[12] Regimento de Engenharia N. 1 do Exército Portguês, “Manual escolar do curso de explosivos, destruições, minas e armadilhas,” Regimento de Engenharia Nº1 - Centro de Treino em Explosivos e Contramedidas, Espinho, 2001.

[13] W. Gonzalez e R. Woods, “Digital Image Processing,” Prentice Hall, New Jersey, 2008.

[14] M. S. Priya e G. M. Nawaz, “Matlab Based Feature Extration and Clustering Images using K-Nearest Neighbour Algorithm,” *iJact,* vol. 2, pp. 1121-1126, 2016.

[15] R. M. Haralick, “Statical and structural approches to texture,” *Proceedings of the IEEE,* vol. 67, pp. 786-804, 1979.

[16] M. M. Galloway, “Texture analysis using gray level run lenghts,” Computer graphics and image processing, Maryland, EUA, pp. 172-179, 1975.

[17] Z. M. Hira e D. F. Gillies, “ A review of feature selection and feature extraction methods applied on microarray data,” Advances in bioinformatics, 2015.

[18] N. Morono e A. Betanzos, “Filter Methods for Feature Selection – A Comparative Study,” em *Intelligent Data Engineering and Automated Learning - IDEAL, 8th International Conference*, Birmingham, UK, pp. 178-187, 2017.

[19] N. Macari, “Analysis of a machine learning algorithm and corpus as a tool for managing the ambiguity problem of search engines,” Master of Science, Fakultat Informatik, Technische Universitat Dresden, 2010.

[20] M. F. Gonçalves, “Classificação do Coberto Vegetal em Ambiente Militar,” Dissertação de Mestrado, Instituto Superior Técnico, U. Lisboa, Lisboa, 2014.

[21] T. Mitchell, “Decision Trees Learning,” Machine Learning, McGraw-Hill Education, 1 ed, pp 52-79, 1997.

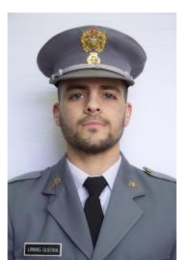
[22] L. Almeida, “PCA-Notes An introduction to principal components analysis,” Instituto Superior Técnico, U. Lisboa, Lisboa, 2015.

[23] C. Orrite, M. Rodriguez, F. Mart e M. Fairhurst, “Classifier Ensemble Generation for the Majority Vote Rule,” em *13th Iberoamerican congress on Pattern Recognition: Progress in Pattern Recognition, Image Analysis and Applications*, Havana, Cuba, pp. 340-347, 2008.

[24] A. R. Webb e K. D. Cospsey, “Statistical pattern recognition,” Chichester: John Wiley & Sons, 2011.

[25] H. Greenspan, B. Ginneken e R. Summers, “Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique,” *IEEE Transactions on Medical Imaging,* vol. 35, nº 5, pp. 1153-1159, Maio 2016.

[26] J. Ker, L. Wang, J. Rao e T. Lim, “Deep Learning Applications in Medical Image Analysis,” *Special Section on fodt Computing Techniques for image analysis in the medical industry current trends, challenges and solutions,* vol. 6, pp. 9375-9389, 2018.



Ivo Fernando Fontes Linhas Guerra was born in Mirandela, Portugal, on March 4, 1994. He lived ther until he graduated from high school.

He joined the Military Academy in September 2012 at the age of eighteen.

In the Military Academy he concludes the first four years of a seven-year course, and the he attended two years of Electrical and Computer Engineering in Instituto Superior Técnico (Lisbon). He is currently attending his final year of the course with the second lieutenant rank, in Mafra, after which he will become a Signals lieutenant Officer of the Portuguese Army.