A Comparison of Methods of Data Fusion for Land-Mine Detection

Dr Matthew Roughan[1]
Software Engineering Research Centre
Level 2, 723 Swanston St, Carlton, Vic 3053, Australia
ph: + 61 3 9282 2412
fax: + 61 3 9282 2444
email: matt@serc.rmit.edu.au

Assoc. Prof. Daniel McMichael
Cooperative Research Centre for Sensor Signal and Information Processing
SPRI Building, Technology Park, The Levels, SA 5095, Australia
ph: + 61 8 8302 5021
fax: + 61 8 3032 3124
email: dwm@cssip.edu.au

Abstract: **It has been suggested that fusion of multiple data sources may be required for the reliable detection of land mines at acceptable false alarm rates. This study validates this contention by showing that a fusion algorithm using images from two infrared wave bands (3-5μm and 8-12μm) yields significantly better false alarm rates than either by itself. The research was pursued with methodological rigour, both in experimental design and statistical assessment. The data processing involves orthographic registration, region of interest extraction, feature extraction, feature selection, classification.**

1 Introduction

Land mines pose problems for the military by restricting their mobility, and for civilian populations, who are at risk long after conflicts have passed. At current rates of clearance, more than one thousand years would be required to remove all of the existing land mines (approximately 100 million according to UN figures), not to mention the 2.5 million mines being laid each year.

Current techniques for mine clearance involve painstaking investigating of each inch of ground, by hand, or by sniffer dog. Active investigation is under way into various sensor technologies – chemical, acoustic, electro-optical and electromagnetic – to provide improved land-mine detection methods. Notably, infrared (IR) imagery can detect buried land mines by detecting associated variations in the surface temperature of the soil caused by the different thermal inertias and conductivities of the mine and soil and the diurnal temperature cycle [Clark *et al.*, 1992; Clark *et al.*, 1993; DelGrande, 1990], as can be seen in *Figure 1*.

The primary requirement of any land-mine detection system is a high probability of detection (P_d). Modern land mines may be very small, have little metallic content, and therefore be very hard to detect. As in other difficult detection applications, single-sensor systems with high probabilities of detection have correspondingly high probabilities of false alarm (P_f). At present, no *single-sensor* technique can simultaneously deliver the required probabilities of detection and false alarm, and achieve an acceptable rate of progress [Clark *et al.*, 1992; Clark *et al.*, 1993; DelGrande, 1990; Hanshaw, 1995; Mullins, 1996].

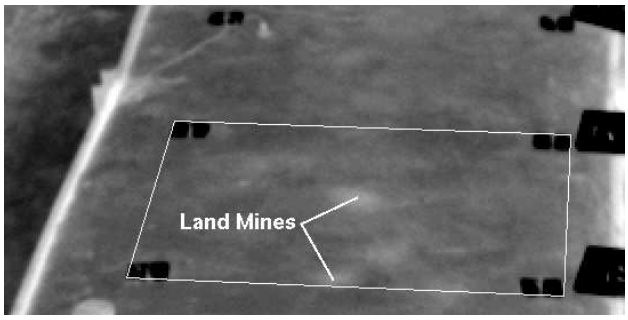

Figure 1: An example IR image containing a land mine which appears as the a lighter circular feature near the centre of the registration square. The is also a second mine near the base of the registration square.

This study aims to provide a quantitative demonstration of the benefits of using multiple sensors for the detection of buried land mines. In particular, it aims to show a reduction in the P_f through the fusion of two sets of imagery from different IR bands.

This paper is based upon three technical reports [Roughan and McMichael, 1996a,b,c] which describe the methodology for MSF [Roughan and McMichael, 1996a], the experimental design [Roughan and McMichael, 1996b], and the results [Roughan and McMichael, 1996c] of the project. The paper provides several features not seen in other land-mine detection studies:

- a more sophisticated approach to MSF;
- a thorough discussion of the methodology used;
- a well-founded experimental design;
- a statistical analysis of the results to provide strong conclusions.

The paper is divided into six sections. Following this introduction, Section 2 describes the data. Section 3 presents the methodology and implementation of the algorithms, and

[1] Formerly of CSSIP.

Section 4 presents the results. Section 5 discusses these results, and areas for future research. Section 6 summarises and concludes the paper.

2 Data

The data was provided under a contract with the Defence Science and Technology Organisation of Australia (DSTO). The data was collected from a platform approximately 4m high, located 7m from the base of the minefield, which was approximately 8m long, and 3m wide. One third of the minefield was covered in sand. Please see *Figure 2* for a schematic of the minefield.

A summary of the data provided to the Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP) by DSTO, which was in the form of raw IR imagery, plus a small number of visual range images of the minefield, is given in *Table 2*. The Agema imagery, collected using the Agema DBTI 900 system, consists of two bands of IR imagery: 3-5 μm, and 8-12 μm, referred to here as short and long wavelength imagery, respectively. The IR imagery was 12 bit, with a spatial resolution of 272x136 pixels. The visual range imagery was taken with a Kodak DCS 420 camera, and provided as 24 bit colour RGB TIFF files, with a spatial resolution of 1524x1012 pixels.

The experimental design of the Data Fusion for Land-Mine Detection project [Roughan and McMichael, 1996b] specified 18 categories of data, based on soil type, time of day, and whether the images contained extensive clutter. Classes I and II, shown in *Table 1*, were considered crucial for assessing the advantages of MSF in this application.

Class	Experimental variables			
	time	weather	soil type	clutter
class I	daylight	$*$	sand	yes
class II	daylight	$*$	sand	no

Table 1: The classification of conditions for key experiments. The "$*$" indicates that the variable has not been specified, but it is uniform for all the data.

The amount of data required was designed to be sufficient to match a requirement on the size of the statistical test of the hypothesis that this fusion method gives better performance than an analogous single-sensor algorithm [Roughan and McMichael, 1996a]. It was found that approximately 100 two-band images would be required to differentiate the P_f of two different detection methods (with a reasonable level of confidence), while 200 two-band images would be required to differentiate P_d's. In the event, the data that was actually provided is shown in *Table 2*.

The net result of registration, Region Of Interest (ROI) selection, restriction to PMN mines (see following sections), and selecting class II data was 90 ROI with a PMN or PMN2 land mine in them. This was approximately the amount of data required for detecting differences in P_f, but more data would be needed to detect differences in P_d's.

2.1 Data driven objectives

In order to focus the results of the project where there is the most available data, the specific objectives of this paper are limited to consideration of:

Data class	Number of two-band images		
	short wave	long wave	visual
class I	7×5	7×5	0
class II	18×5	18×5	20

Table 2: The quantity of data provided. Note the IR imagery is shown in terms of sets of five images, each set of which covers the $8m \times 1m$ sand-covered part of the minefield, shown in *Figure 2*.

- class II data (described above);
- detection of PMN an PMN2 mines;
- testing classification algorithms.

Class II data, and the fusion of dual-band IR data were of primary interest here, while class I data will be examined in more detail once more imagery has been collected for this data set. The study was conducted with the emphasis on the PMN and PMN2 anti-personnel mines, for the following reasons:

- Different mines have substantially different signatures – based on weight, radius, etc. – and therefore a separate classifier is required for each type.
- To simplify the statistical analysis.
- Anti-personnel mines were chosen for testing as very little anti-tank data was available. There were only two anti-tank mines in the mine field that was tested, and these were all at the extreme end of the minefield, where a number of features (poor resolution, and camera angle) made imagery of these mines hard to use.
- Of the anti-personnel mines, the PMN mine is best represented in the data.
- The PMN mines are well distributed over the mine-field.

3 Methodology

The study's methodology is simple and robust. The detection process is broken into four sequential processes:

- registration of the imagery;
- selection of the ROI;
- extraction of features from the ROI;
- classification of the ROI based upon the features.

The four stages are discussed in detail in Sections 3.1 to 3.4.

3.1 Registration

The data of interest consists of three sets of imagery - two sets of IR imagery and one set of visible imagery (to be used later). The cameras used to take this imagery each have different fields of view, orientations, and images resolutions, as shown in *Figure 3*.

The first stage of the processing, described above, is that of registration, the process of overlaying the images so that there is a spatial correspondence between the features in the different sets of imagery. In this application, the imagery was to be co-registered to a common coordinate system – a square grid

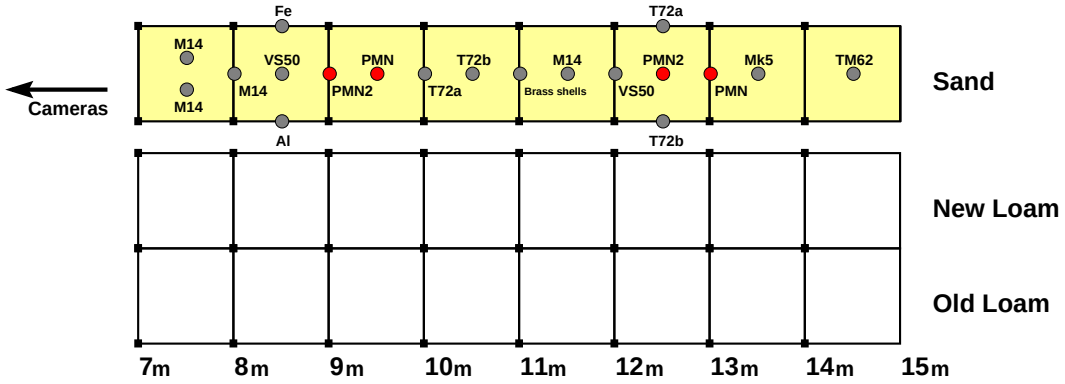

Figure 2: A schematic of the minefield, showing the types of the mines (PMN, M14, T72, VS50, Mk5, TM62) buried in the sand covered part of the field (shaded). The small black squares indicate the registration markers, while the distances along the x-axis indicate the distance from the cameras.

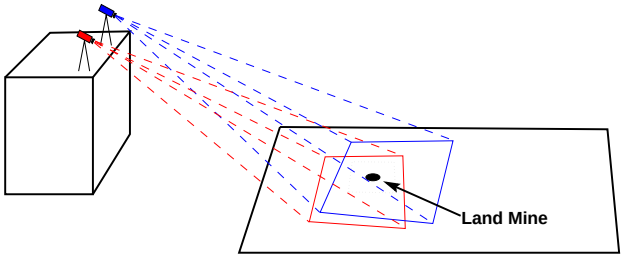

Figure 3: The camera setup. The two cameras have different fields of view, orientations and resolutions. The dotted black lines indicate a square from the grid.

placed over the minefield. This process, sometimes referred to as *rectification*, corrects the perspective distortions of the flat minefield. In general, registration is a difficult problem; however, in this instance the registration process was aided by the addition of fiducial points in the image. The fiducial markers were small square aluminium plates, placed in recorded positions on the ground before capturing the images (refer to *Figure 2*).

A graphical user interface (GUI) was developed at CSSIP to allow manual extraction of the registration points using mouse clicks on the source imagery; the points were then used as reference points to construct a transformation from the registered image to the source image:

$$[x, y] \quad = \quad [f_x(u, v), f_y(u, v)], \qquad (1)$$

where $[x, y]$ refers to coordinates in the source image corresponding to $[u, v]$ in the registered image, while $f_x(\cdot, \cdot)$ and $f_y(\cdot, \cdot)$ are the mapping functions.

The transformation from the registered image to the source image may be used to construct the registered image through resampling. Sampling is required because of the digital nature of both source and registered images. In nearest neighbour resampling [Russ, 1995, p. 203] the (x, y) coordinates obtained from the transformation in equation (1) are rounded off to find the pixel in the source image nearest to (x, y). An alternative, used in this application, is to round down (x, y) to obtain the location of the sampled pixel; that is,

$$I_{\text{reg}}(u, v) \quad = \quad I_{\text{source}}\Big(\lfloor f_x(u, v) \rfloor, \lfloor f_y(u, v) \rfloor\Big). \qquad (2)$$

where $I_{\text{source}}(\cdot, \cdot)$ is the source image transformed to a common reference frame $I_{\text{reg}}(\cdot, \cdot)$, and $\lfloor x \rfloor$ denotes the largest integer less than x.

Interpolation using neighbouring pixels can reduce artifacts in the registered images introduced by sampling [Russ, 1995, p. 205], but the minimal occurrence of such artifacts made interpolation unnecessary in this application.

The perspective transformations introduced by the cameras distorts the square grid on the mine-field into trapezoids in the images [Russ, 1995, pp. 200–201], as can be seen in *Figure 1* which shows the registration square (in white) transformed into a trapezoid.

When the registration points lie at the corners of a known rectangle the coordinate transformations $f_x(\cdot, \cdot)$ and $f_y(\cdot, \cdot)$ which reverse the perspective transformation may be performed using a *trapezoidal mapping* without calculating the perspective transformation explicitly. The trapezoidal mapping is found using the four registration points, chosen manually from the source image. These four points define a trapezoid which corresponds, in reality, to a horizontal square of the grid on the minefield. Thus pixels in the registered square can be mapped to the source image by bi-linear interpolation of the positions of the registration points.

The GUI application, programmed in IDL, allowed co-registration of the three image sources (two bands of IR, and one visible range image) by manually inputing the location of the four fiducial points, via mouse click in a window. The size of the registered image was chosen to be 300×300 pixels, in order to oversample the IR source images. *Figure 4* shows *Figure 1* after rectification.

3.2 Region of interest selection

The *regions of interest* (ROI) are subregions of the image, to which the test for a land mine is applied. The test is applied by extracting features from the ROI, as described in Section 3.3 and classifying based on features selected in Section 3.3. Restricting feature extraction to the relatively small ROI reduces the computation required to detect significant features.

In this study, the set of ROI were manually chosen to directly coincide with the positions of the land mines in the image. In fully automated systems the ROI would be selected from the image using image processing algorithms, but in order to

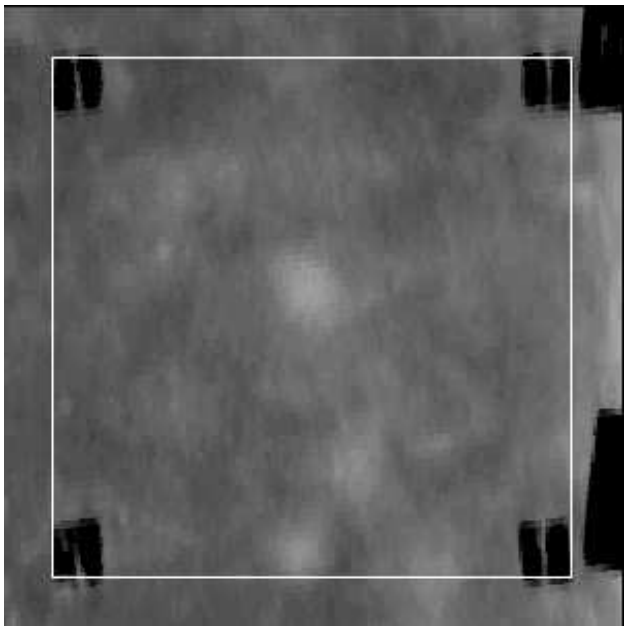

Figure 4: The registration square from *Figure 1* after rectification. The four black squares at the corners of the registration square are the aluminium registration markers.

focus the testing on the classification stage – the critical stage for MSF – ROI were extracted using ground truth information. A set of control ROI, referred to as *decoy regions of interest* (DROI), were chosen so that they contained no part of a land mine.

The ROI were chosen to fit only the PMN and PMN2 mines, as discussed in Section 2.1. The exact locations of the ROI and DROI are given in [Roughan and McMichael, 1996c].

3.3 Feature extraction and selection

Feature extraction is the next task. Finding a suitable feature set is critical to success. In this project, a set of features (described below) was tested, from which a useful subset was selected. The subset of features was initially chosen by a visual inspection of the data, but the appropriateness of these features was later verified by Fisher discrimination (refer to Section 3.3).

Feature extraction The features used in this project were those which could be sensibly extracted from a rectangular ROI. Features based on the shape of the ROI are obviously useless in this problem, where all of the ROI are square. The features extracted from the ROI fall into two categories - the gray-scale statistics of the ROI, and rotationally invariant statistics. Gray-scale statistics are commonly used as features, and their inclusion is necessary in any comprehensive testing, if only to rule them out. The rotationally invariant statistics have been specifically chosen for this problem because the mines have a circular signature in the images. The signature does not have a sharp edge, and so the standard methods for detecting circles – for example the Hough transform – will not perform well. Instead a series of statistics, designed to measure the behaviour of a *radial profile*, are used as features. All together a set of 12 features was extracted from the ROI.

The gray-scale statistics for $N \times N$ ROI are:

- The minimum intensity in the ROI:
$$ROI_{min} = \min_{x,y} ROI(x,y).$$
- The maximum intensity in the ROI:
$$ROI_{max} = \max_{x,y} ROI(x,y).$$
- The mean intensity μ in the ROI:
$$\mu = \frac{1}{N^2} \int_{x,y} ROI(x,y)\, dx\, dy.$$
- The standard deviation σ of the intensity in the ROI where
$$\sigma^2 = \frac{1}{N^2} \int_{x,y} \{ROI(x,y) - \mu\}^2\, dx\, dy.$$
- The coefficient of skewness η_3 of the intensity in the ROI:
$$\eta_3 = \frac{1}{N^2} \frac{\int_{x,y} \{ROI(x,y) - \mu\}^3\, dx\, dy}{\sigma^3}.$$
- The coefficient of excess $\eta_4 - 3$, of the intensity in the ROI where
$$\eta_4 = \frac{1}{N^2} \frac{\int_{x,y} \{ROI(x,y) - \mu\}^4\, dx\, dy}{\sigma^4}.$$

The rotationally invariant statistics used here are based on a radial profile. The radial profile is obtained by finding the mean pixel intensity at given distances from the centre of the ROI, that is:

$$r(\rho) = \frac{1}{2\pi\rho} \int_{(x,y) \in C(\rho)} ROI(x,y)\, dx\, dy,$$

where $C(\rho)$ is the circle, of radius ρ defined by

$$C(\rho) = \left\{ x, y \,\middle|\, (x - x_0)^2 + (y - y_0)^2 = \rho^2 \right\},$$

where (x_0, y_0) denotes the centre of the ROI. The radial profile itself (discretised) is not suitable for classification as it has a high dimensionality, and so features of the profile are used for classification. Given a normalised radial profile $\hat{r}(x)$, and Gaussian approximation $g(x)$ to this profile, the features used were:

- The *variational radius*:
$$\mathcal{R} = 2 \int_0^R \hat{r}(\rho)\, |\rho|\, d\rho.$$
- The mean square error between $g(\rho)$ and $\hat{r}(\rho)$:
$$\mathcal{G}_{MSE} = 2 \int_0^R \{\hat{r}(\rho) - g(\rho)\}^2\, d\rho.$$
- The chi-square goodness of fit of $g(\rho)$ to $\hat{r}(\rho)$:
$$\mathcal{G}_{\chi^2} = 2 \int_0^R \frac{\{\hat{r}(\rho) - g(\rho)\}^2}{g(\rho)}\, d\rho.$$
- The Kolmogorov goodness of fit of $g(\rho)$ to $\hat{r}(\rho)$:
$$\mathcal{G}_K = \max_{\rho} |\hat{r}(\rho) - g(\rho)|.$$

- The mean of the circular variance from the radial profile, the circular variance being given by

$$v(\rho) = \frac{1}{2\pi\rho} \int_{(x,y)\in C(\rho)} \{r(\rho) - ROI(x,y)\}^2 \, dx \, dy.$$

- The variance of the circular variance from the radial profile.

Note that the discretised versions of the above formulae are used in the implementation, and hence these features are only truly invariant under rotations through 90°.

Three of the statistics above have measured the goodness of fit of the radial profile to a Gaussian approximation. Ideally, physical modelling would provide an estimate of the expected radial profile from the collateral data, and the goodness of fit would be evaluated with respect to that.

Feature selection *Initial data exploration.* Initially the features were examined with `XGobi` [Swayne *et al.*, 1991], a program specifically designed to provide a graphical means of examining feature data. From this visual examination, the most suitable feature for classification was clearly the variational radius. *Figure 5* shows the variational radius for the short-wave images *vs* the variational radius for the long-wave images, for the class II data only.

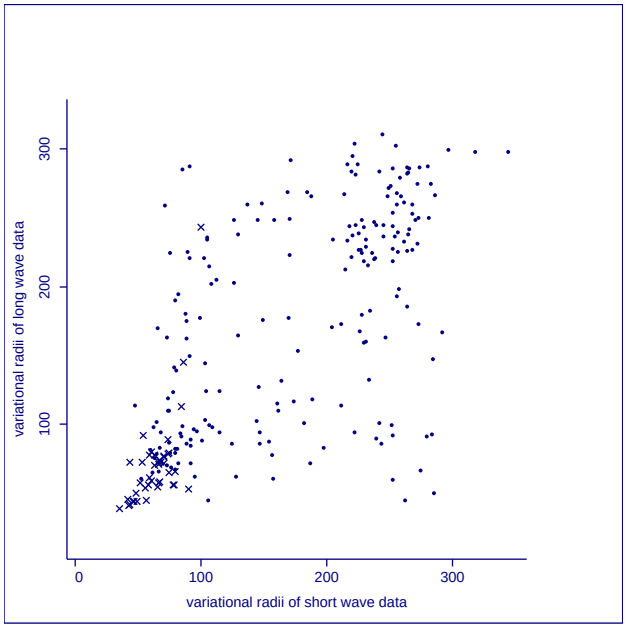

Figure 5: The variational radius for the long vs short-wave images. The circles are cases with no land mine present, while the crosses indicate cases where a land mine was present.

Fisher discrimination. The Fisher discriminant (see Section 3.4) was found for the land-mine features. It is a linear combination of the original 24 features (12 each from the long- and short-wave imagery). The component of each of the original 24 features in the discriminant is shown in *Table 3*.

Table 3 shows that the variational radius is by far the dominant feature in the Fisher discriminant. The coefficients of the variational radius in both long- and short-wave imagery are larger than any other coefficient, by a factor of 100. This supports the initial estimate of which features should be used for classification. The distribution of the Fisher discriminant for the ROI and the DROI, for class II data is shown as a scatter plot in *Figure 6*.

Feature	Component value	
	short wave	long wave
Minimum intensity	-0.00124724	-0.00625205
Maximum intensity	0.00630188	0.00333990
Mean intensity	-0.00059549	-0.00476067
Standard deviation	0.00189057	0.00264853
Coefficient of skewness	0.00419134	0.00457964
Coefficient of excess	-0.00233164	-0.00694964
Variational radius	**-0.69212500**	**-0.72161900**
Mean squared error	-0.00000081	-0.00000098
Chi-squared GOF	-0.00167863	-0.00249913
Kolmogorov GOF	-0.00009252	-0.00012151
Mean of circular var.	0.00029314	0.00075823
Var. of circular var.	-0.00077399	-0.00049984

Table 3: The Fisher discriminant components from each of the original 24 features. The variational radius is shown in bold to highlight its importance.

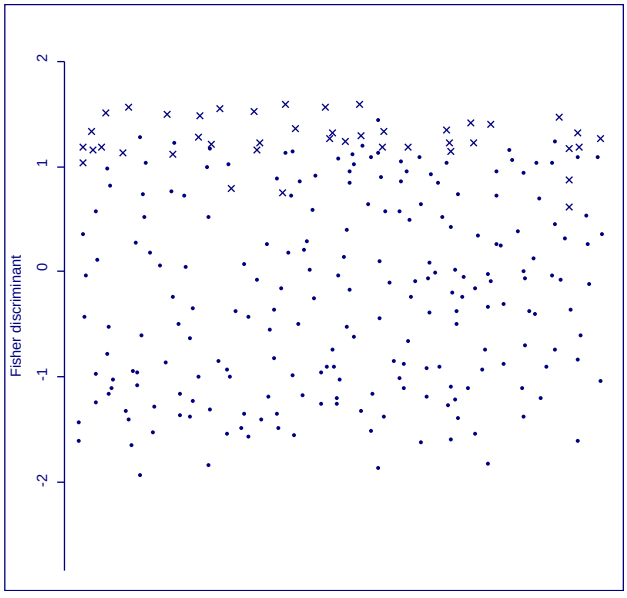

Figure 6: A scatter plot of the Fisher discriminant. The cases with a mine present are shown as crosses, while the cases without a mine present are shown as circles. The x-axis has been used to spread the data to make it easier to interpret.

3.4 Classification

The final stage of processing – classification – is discussed in this section. The ROI must be classified into two classes "land mine present," and "no land mine present," based on the features selected in the previous stage. Five classification methods were tested, two single-sensor techniques (one for each IR band) and three fusion methods. For reference to these methods please see [Duda and Hart, 1973].

The simplest form of classification is a special case of linear discrimination, *thresholding*. Linear discrimination separates

classes with a linear boundary in the feature space, thresholding being the case where only one feature is used. The two single-sensor methods (one for each IR band) use a threshold based on the best of the features extracted in the previous stage.

A naïve method of combining the results from each single-band classification is *polling*, where the results of the simple thresholding classification are combined via an AND operation, as shown in *Table 4*. Polling is simple because it separates fusion from classification, but much is to be gained by integrating them.

	Short Wave	
Long wave	No mine	Mine present
No mine	"No mine"	"No mine"
Mine present	"No mine"	"Mine present"

Table 4: A table showing the declaration of the polling method, given the individual classifications based on the two IR bands.

An approach to MSF, which integrates classification with fusion, is to perform classification on the combined features from both IR bands. Fisher discrimination (discussed above) finds the optimal linear discriminant function for the combined set of features by finding the linear transform which maximises the ratio of *between-class scatter* to *within-class scatter*. The result, the *Fisher discriminant*, shown in *Figure 6*, can then be thresholded to perform the classification.

When a linear boundary is not appropriate in a higher dimensional feature space, an alternative is the Bayes classifier, in which estimates of the probability densities of the data are found, and using these and the class priors, the most probable class is chosen. The method used here for estimating the densities is Parzen windowing. The advantages of Bayesian classifier methods are that they may yield non-linear classification boundaries, and they may be optimised through the use of hyper-parameters.

Approximate Bayes classifiers work by finding approximations to $p(\boldsymbol{x}|c)$, the probability of a set of data $\boldsymbol{x}$, given class c. From these conditional probabilities and the class priors, Bayes rule is used to find $p(c|\boldsymbol{x})$ for some particular data. The class priors are estimated as

$$p(c) = \frac{N_c}{N},$$

where N_c and N are the numbers of data in class c and the total number of data respectively.

Gaussian classification is applied by modelling $p(\boldsymbol{x}|c)$ by a Gaussian distribution:

$$p(\boldsymbol{x}|c) = \frac{1}{|2\pi\Sigma_c|^{\frac{1}{2}}} \exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{\mu_c})^T\Sigma_c^{-1}(\boldsymbol{x}-\boldsymbol{\mu_c})}{2}\right),$$

where $\boldsymbol{\mu_c}$ and Σ_c are the means and covariances of the conditional distribution, which may be estimated by

$$\boldsymbol{\mu_c} = \frac{1}{N_c}\sum_{j:y_j=c}\boldsymbol{x_j}, \quad (3)$$

$$\Sigma_c = \frac{1}{N_c-1}\sum_{j:y_j=c}(\boldsymbol{x_j}-\boldsymbol{\mu_c})(\boldsymbol{x_j}-\boldsymbol{\mu_c})^T, \quad (4)$$

where y_j is the class of data $\boldsymbol{x_j}$. Initial examinations of the data, however, would indicate that the data is non-Gaussian (refer to *Figure 5*), and therefore another Bayesian classifier was considered.

The Parzen windows classifier approximates the conditional probability $p(\boldsymbol{x}|c)$ by a sum of *window* functions. In this case the windows functions $w(\boldsymbol{x}, \boldsymbol{x_j})$ used were spherical Gaussians, centred on $\boldsymbol{x_j}$, giving

$$p(\boldsymbol{x}|c) = \frac{1}{N_c(2\pi\sigma^2)^{\frac{d}{2}}}\sum_{j:y_j=c}\exp\left(-\frac{(\boldsymbol{x_j}-\boldsymbol{x})^T(\boldsymbol{x_j}-\boldsymbol{x})}{2\sigma^2}\right).$$

Parzen windows would seem a much more appropriate method than the Gaussian classifier for approximating the distribution of data in *Figure 5*, because it can accommodate the non-Gaussian distributions evident in the data.

The parameter σ is a *hyperparameter* of the Parzen windows classifier. The performance of the classifier depends on the hyperparameter. Cross-validation is the process of using *separate* sets of training and testing data to optimise its value. The Parzen window approximation is found for the training data, and the numbers of errors on the test data made by the Parzen window classification are calculated (making the hard classification boundary at $p(c|\boldsymbol{x}) = 0.5$), for a range of hyperparameter values σ. The result can be seen in *Figure 7* which shows the number of errors made on the training set. In land-mine detection, failed detections are critical. It can be seen that higher values of σ result in both higher detection and false alarm rates. Simultaneous optimisation of these two error rates e_D and $e_{\mathrm{P_f}}$ requires a loss function $L(e_D, e_{\mathrm{P_f}})$, to be defined. The hyper-parameter estimate is the value which minimises the loss function.

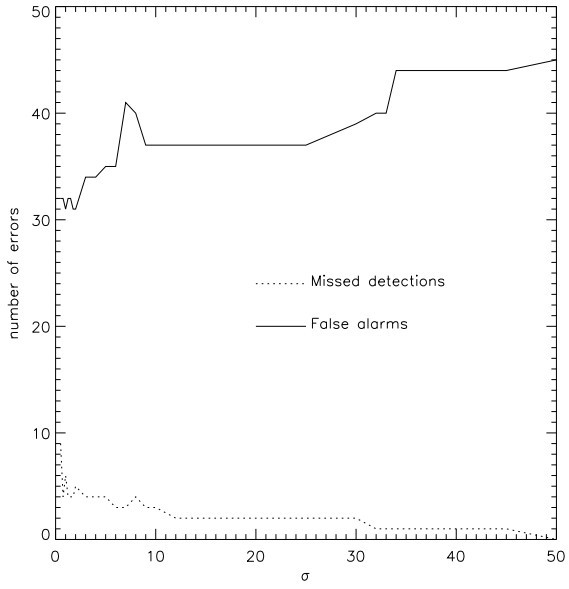

Figure 7: The number of errors as a function of the hyperparameter. The number of test samples were 45 and 217 for the missed detection, and the false alarms respectively.

An alternative is to maximise the likelihood of the test data, $\mathcal{L} = p(\boldsymbol{Y}|\boldsymbol{X})$ where $p(\boldsymbol{Y}|\boldsymbol{X})$, is the probability that the correct classes $\boldsymbol{Y}$ are chosen for the test data $\boldsymbol{X}$. The likelihood

function is simplified by assuming independent experiments, and the log is taken to give the average log-likelihood:

$$\log \mathcal{L} = \log \left(\prod_{j=1}^{N_{\text{test}}} p(y_j|\boldsymbol{x}_j) \right) \quad (5)$$

$$= \sum_{j=1}^{N_{\text{test}}} \log \left(p(y_j|\boldsymbol{x}_j) \right). \quad (6)$$

When, for an outlying datum $\boldsymbol{x}_j$, $p(y_j|\boldsymbol{x}_j) \simeq 0$, the log becomes large and negative, and can outweigh all of the other contributions to the function. The moderated cross validation (MCV) score has been used here in order to reduce the weight of outliers. *Figure 8* shows the MCV score given by:

$$\sum_{j=1}^{N_{\text{test}}} \log \left(\frac{p(y_j|\boldsymbol{x}_j) + \epsilon}{1 + N_{\text{classes}}\epsilon} \right),$$

where N_{classes} is the number of classes (for this application $N_{\text{classes}} = 2$) and ϵ is a small positive constant; three values of ϵ: $0.005, 0.01$ and 0.02, have been used to show that the result is not sensitive to its value near to the maximum of the MCV score. *Figure 8* shows the MCV score as a function of the hyper-parameter. The optimal values are $\sigma = 32, 34, 35$, for each of the three values of ϵ. In the testing performed here the median of these values, $\sigma = 34.0$, is used.

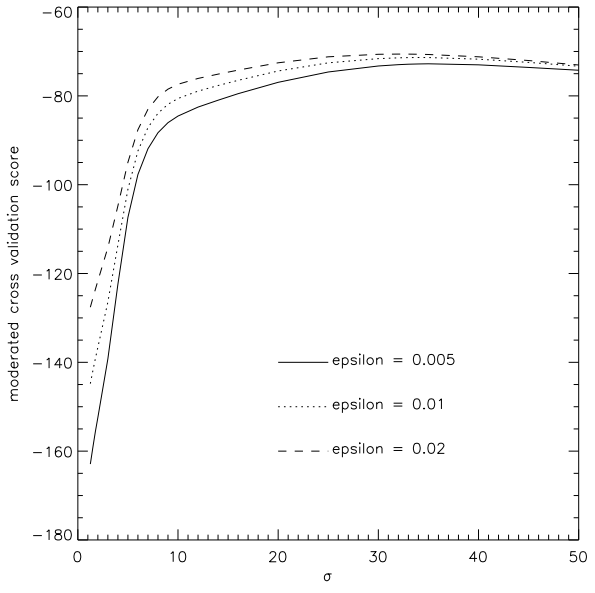

Figure 8: The moderated cross validation score as a function of the hyperparameter.

Once the hyperparameter has been chosen, the classification boundary can be set. Given only two classes, as in this application, the decision between classes may be made at a fixed threshold $p(c|\boldsymbol{x}) = k_\beta$, where k_β is chosen such that the probability of failing to detect a land mine which is present, $p(c = \text{'no mine'}|y = \text{'mine present'}) \leq \beta$, for some fixed number β. *Figure 9* shows the boundaries for $\beta = 0.04, 0.05$ and 0.06.

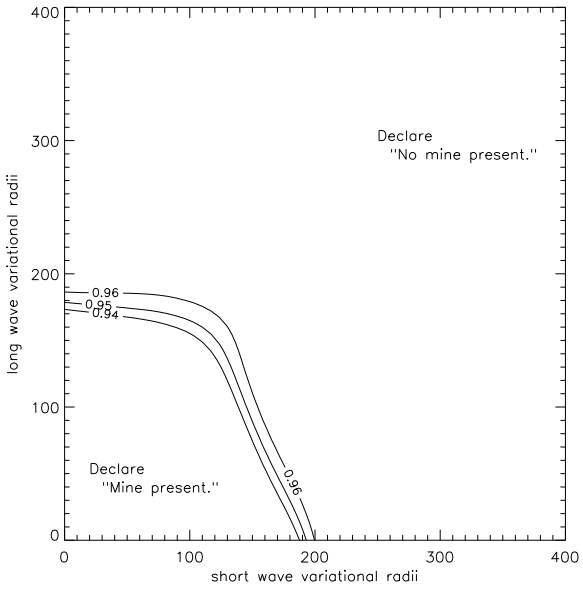

Figure 9: The boundaries between classes. The contours, labelled by the probability of detection $1 - \beta$, are the lines $p(c|\boldsymbol{x}) = k_\beta$, where k_β is chosen to to give required probability of detection.

4 Results

This section describes the results of the land-mine detection methods discussed above. Three approaches are used in describing the performance of the algorithms. The first, described in Section 4.1, is non-statistical, but is included to show the relationship between the false alarm rate and P_d for the five methods. The second section, Section 4.2 provides an intuitive statistical assessment of the five algorithms, followed in Section 4.3, by a formal assessment using the generalised likelihood ratio test, which was described in [Roughan and McMichael, 1996b].

The five methods of classification assessed here are two single-sensor methods (one for each IR band), polling, Fisher discrimination and the Parzen windows classifier, with hyperparameter $\sigma = 34.0$. In these results only the class II data was considered, and only PMN type mines detected, for the reasons outlined in Section 2.1.

4.1 Assessment using Receiver Operating Characteristic curves

A traditional method for comparing detection algorithms is the Receiver Operating Characteristic (ROC) curve which shows the detection and false alarm rates over a range of threshold levels, illustrating the trade-off between the two error rates. The ROC curves are shown in *Figure 10*, for the five methods of detection.

The figure does not show the statistical relationship between the five methods, but it does provide an idea of their relative performance, at different threshold levels. Our tentative conclusion is that the two single-sensor methods are worse than the MSF methods, and of the MSF techniques Parzen windowing is the best.

A less standard, but more revealing, form of ROC curve, is

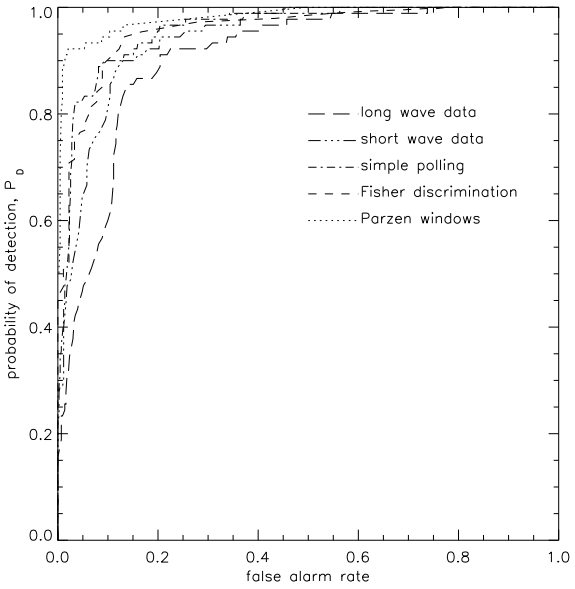

Figure 10: The ROC curves for the different classification methods, showing the P_f *vs* P_d for various methods of classification.

shown in *Figure 11(a)*. The axis have been switched, and the x-axis now shows $\beta = 1 - P_D$, the probability of a missed detection. The logarithmic x-axis gives more detail in the important interval $P_D = (0.8, 0.99)$.

4.2 Confidence interval analysis

In this section the results of the land-mine detection feasibility study are analysed using standard frequentist confidence intervals. *Table 5* shows the 80% confidence intervals for the P_f, given an estimated rate of detection, $94, 95$ and 96%. The confidence intervals provide an intuitive measure of the relative performance of the different algorithms.

The results in *Table 5* are graphically displayed in *Figure 11* (b) which shows the 80% confidence intervals for the false alarm rate on the ROC curves of *Figure 11(a)*. A similar graph of the 80% confidence intervals for the P_d's, shown in [Roughan and McMichael, 1996c], illustrates the fact that there was insufficient data to determine a difference between P_d's of the methods.

4.3 Generalised likelihood ratio test

The previous method for analysing the results is somewhat informal. It presents simple, intuitive measures of the algorithmic performance. This section uses a more sophisticated approach, and can provide stronger conclusions. By forming a likelihood based test, the exact confidence in the results can be assessed. In particular, these results show that the confidence in the assessment that the Parzen windows classifier has a lower P_f than a single-sensor classifier is better than 99.998% over a range of probabilities of detection.

The starting point, for a mathematical analysis, is the distribution of the difference of the two false alarm rates $e_1 - e_2$, given two estimated P_f's, $\hat{e}_1$ and $\hat{e}_2$. The false alarm rates, estimated by $\hat{e} = \frac{N_E}{N}$ are shown in *Table 5*, for each classifier.

Considering the error process as a sequence of Bernoulli trials, with parameter e, the distribution of errors follows the binomial distribution:

$$p(N_E|e, N) = \frac{N! e^{N_E} (1-e)^{N-N_E}}{N_E!(N-N_E)!},$$

where N_E is the number of errors, given N trials. As in [Roughan and McMichael, 1996b], the normal approximation to this distribution, $\mathrm{N}(eN, e(1-e)N)$, is used. The distribution for the estimated error rate $\hat{e} = \frac{N_E}{N}$ can therefore be written

$$p(\hat{e}|e, N) \sim \mathrm{N}(e, e(1-e)/N).$$

The difference of two normally distributed random variables is itself normally distributed, and given a difference in the underlying decision probabilities $\delta = e_1 - e_2$, this distribution is

$$p(\hat{e}_1 - \hat{e}_2|\delta, N) \approx \frac{\exp\left(-\frac{(\hat{e}_1 - \hat{e}_2 - \delta)^2}{2\sigma_{12}^2}\right)}{\sqrt{2\pi}\sigma_{12}},$$

where

$$\begin{aligned} \sigma_{12}^2 &= e_1(1-e_1)/N_1 + e_2(1-e_2)/N_2 \\ &\approx \hat{e}_1(1-\hat{e}_1)/N_1 + \hat{e}_2(1-\hat{e}_2)/N_2. \end{aligned} \quad (7)$$

Using Bayes theorem, and uniform priors for the error rates e_1 and e_2, the distribution of δ given the estimated error rates can be written

$$p(\delta|\hat{e}_1 - \hat{e}_2, N) \approx \frac{\exp\left(-\frac{(\hat{e}_1 - \hat{e}_2 - \delta)^2}{2\sigma_{12}^2}\right)}{\sqrt{2\pi}\sigma_{12}}.$$

The generalised likelihood test, described in the [Roughan and McMichael, 1996b], forms the ratio of the likelihood of a hypothesis H_0 that the error rates are the same, to the likelihood of an alternative hypothesis H_1 that they are different. This ratio, or rather twice its negative log, is a good statistic, by which to make a quantitative assessment of the significance of results.

Twice the negative log of the generalised likelihood ratio is given by

$$\lambda = \frac{(\hat{e}_1 - \hat{e}_2)^2}{\sigma_{12}^2}.$$

Given hypothesis H_0 is true, λ is distributed as a χ^2 variate with one degree of freedom. Hence the probability that $\lambda \geq x$, given H_0, is given by

$$p\{\lambda \geq x|H_0\} = \frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{e^{-\frac{t}{2}}}{\sqrt{t}} dt. \quad (8)$$

We choose the size of the test α, which is the probability that the test 'gets it wrong' when hypothesis H_0 is true. This determines the critical value $\chi^2(\alpha)$ which is determined by $\alpha = p\{\lambda \geq \chi^2(\alpha)|H_0\}$. The critical value $\chi^2(\alpha)$, the upper α percentile point of the chi-square distribution, above which H_0 is rejected, determines the test:

$$H = \begin{cases} H_0, & \text{if } \lambda < \chi^2(\alpha); \\ H_1, & \text{if } \lambda \geq \chi^2(\alpha). \end{cases}$$

The Parzen windows classifier has been shown to be the best of the MSF methods above, and the short-wave classifier the best of the single-sensor methods, and so we concentrate on a comparison between the two. The number of tests for

Classification	Required P_d								
	$\hat{P}^D = 94\%$			$\hat{P}^D = 95\%$			$\hat{P}^D = 96\%$		
	P^{FA}_{min}	$\hat{P}^{FA}$	P^{FA}_{max}	P^{FA}_{min}	$\hat{P}^{FA}$	P^{FA}_{max}	P^{FA}_{min}	$\hat{P}^{FA}$	P^{FA}_{max}
long wave	0.311	0.339	0.369	0.331	0.360	0.390	0.379	0.408	0.440
short wave	0.181	0.204	0.231	0.230	0.255	0.284	0.268	0.294	0.324
polling	0.166	0.188	0.214	0.177	0.200	0.226	0.181	0.204	0.231
Fisher discrimination	0.114	0.133	0.156	0.144	0.166	0.190	0.250	0.277	0.305
Parzen	0.068	0.087	0.119	0.080	0.103	0.134	0.104	0.131	0.165

Table 5: The estimates and 80% confidence intervals of the P_fs for five different classification methods, for three estimated probabilities of detection, $\hat{P}^D$.

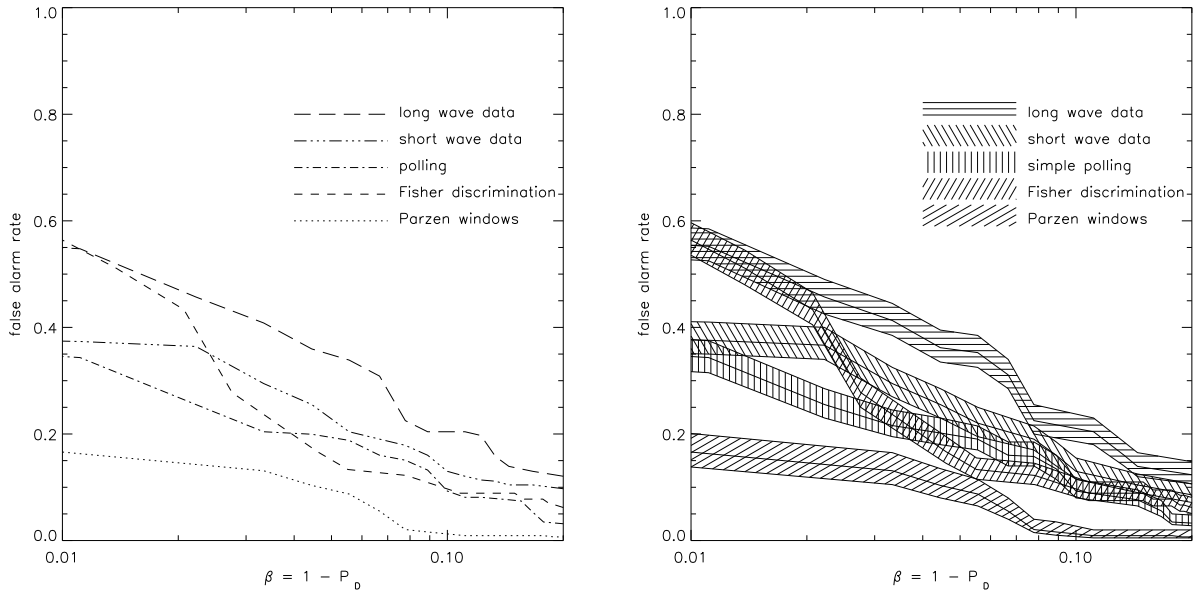

Figure 11: (a) The ROC curves for various methods of classification. (b) The 80% confidence intervals for the P_f rates *vs* the P_d.

the single-sensor and Parzen windowing examples above are 435, and 217 respectively. Parzen windowing has half as many because half of the data is needed for cross-validation. *Figure 12* shows the value of λ. Over the entire range $\lambda > 17.8$ and hence H_1 will be accepted by any test with size 2.0×10^{-5} or more. To rephrase, **the hypothesis that the two error rates are different can be accepted with 99.998% confidence.**

5 Discussion

The best of the three MSF methods tested was the Parzen windows classifier, which more than halves the estimated P_f of a single sensor, with no reduction in the detection probability. The classifier requires a computationally-intensive one-time-only calculation of the class boundaries, but the ongoing computation required is negligibly more than that for two single-sensor classifiers. The reward for doubling the number of sensors is a P_f which is more than halved.

5.1 Further work

There are many possible incremental improvements to the work already performed here. The methodology has already been developed for these tasks. This section discusses the improvements, and methods of approaching them.

Accurate estimates of the probability of detection. An obvious requirement is a comparison of the P_d's of different algorithms. Significantly more data will be required to provide estimates of P_d's accurate enough to discriminate between algorithms. The difficulty arise because the P_d's that are to be estimates are close to each other and to 1.

A loss function. Introduction of a loss function for the two possible types of errors: missed detections, and false alarms, would allow cross validation to optimise the quantity of interest to the user

Extraction of other features. Any number of alternative features could, some exploiting the roundness of mine shadows.

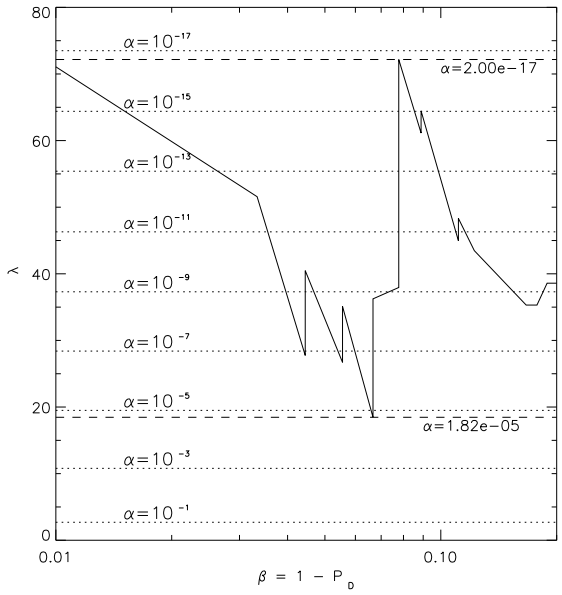

Figure 12: The generalised likelihood ratio statistic λ, for a comparison between Parzen windowing and the short-wave single-sensor classifier, for a range of estimated probabilities of detection. The dotted lines show the size α of the tests, while the dashed lines show α for the maximum and minimum values of λ.

Fusing of other sensors. There are many other sensors which could be considered for integration into a land-mine detection system. The prime candidate at this stage is multi-spectral (MS) imagery from the visible band, which could be used for surface clutter mitigation.

(i) *MS imaging sensors:* The obvious role of MS imagery in detecting *buried* land mines is clutter mitigation: to reduce the effects of clutter on the classification of IR data. Another use is to improve the systems own assessment of the uncertainty of it's classifications.

(ii) *Collateral sensors:* Another set of data which is important is data from collateral sensors, such as the temperature history, or the soil moisture. These may have a significant effect on the IR imagery, and should be taken into account when processing this imagery. The ideal method for incorporating this data is through a mathematical model of the heating/cooling process.

(iii) *Other sensors:* There are several other sensors being considered for inclusion in a land-mine detection system. For instance metal detectors, and ground penetrating radar.

Testing of other data classes and mine types. Wider ranging and more realistic data are needed.

A further point here is that the data vary considerably with range from the cameras, which use a fixed angular field of view. Hence the data should be classified by range, as well as the other classification variables stated here.

This study considered only PMN mines. Each type of mine has its own signature, and hence may require a different classifier. Thus the different mine types – in particular anti-vehicle mines – must be considered before any general performance assessment can be made.

Effects of the weather. The experimental design was not specific about the meteorological conditions under which the experiments should be conducted, but did require these conditions to be consistent. More effective methods of classifying the data according to meteorological conditions have been instituted in the follow-on study.

6 Conclusion

This paper has presented a comprehensive analysis of the data provided to CSSIP in the buried land-mine detection project. The analysis has included:

- a description of the data to be processed;
- the methodology used to process the data;
- details of the implementation of the algorithms;
- results of the methodology, as applied to the data;
- an analysis of the results.

This study has confirmed the value of multisensor fusion in the detection of land mines with a confidence of 99.998%. Furthermore, the estimated P_f is more than halved through the use of two sensors. There is insufficient data to make any such assertion about the P_d's, but the ROC curves suggest that data-fusion techniques perform better than single-sensor techniques.

The study shows that, of the methods tried, the Parzen windows classifier is best for fusing the information from the two IR bands. Parzen windowing has the lowest P_f of the MSF techniques, and can be applied with negligibly more computation than two single-sensor detectors.

This paper *does not* provide a method for automatically detecting buried land mines in IR imagery, nor is this its intent. The study has focussed on the classification of the ROI, rather than the automatic selection of the ROI. Furthermore, the paper has been limited to showing the advantage of fusing the two IR bands for class II data, because this was the only complete data available. Thus the clear avenue for future work arising from this project is the automatic identification of ROI.

Acknowledgement

The authors are grateful to the Defence Science and Technology Organisation of Australia (DSTO), who funded this work through Contract Acceptance and Purchase Order No. 476872. In particular, our thanks go to Karl Fueloep, Bob Seymour and their associates for their support and their efforts in collecting the data.

REFERENCES

[1] G.A. Clark, J.E. Hernandez, S.K. Sengupta, R.J. Sherwood, P.C. Schaich, M.R. Buhl, R.J. Kane, M.J. Barth, and N.K. DelGrande, (1992): Computer vision and sensor fusion for detecting buried objects. In *Twenty-Sixth Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 466–471, Los Alamitos, CA, USA, 1992. IEEE Comput. Soc. Press. Pacific Grove, CA, USA.

[2] G.A. Clark, S.K. Sengupta, M.R. Buhl, R.J. Sherwood, P.C. Schaich, N. Bull, R.J. Kane, M.J. Barth, D.J.

Fields, and M.R. Carter, (1993): Detecting buried objects by fusing dual-band infrared images. In *Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 135–143, Los Alamitos, CA, USA, 1993. IEEE Comput. Soc. Press. Pacific Grove, CA, USA.

[3] N.K. DelGrande, (1990): Temperture evaluated mine position survey (temps) application of dual-band infrared methodology. volume 2073 of *Proceedings of the SPIE*, pages 365–384. SPIE, 1990.

[4] R.O. Duda and P.E. Hart, (1973): *Pattern Classification and Scene Analysis*. John Wiley & Sons, Inc., New York, 1973.

[5] T. Hanshaw, (1995): Multi-sensor fusion for the detection of mines and "mine like" targets. In *Detection Technologies for Mines and Minelike Targets*, Volume 2496 of *Proceedings of the SPIE*, pages 152–158. SPIE, 1995. Orlando, FL, USA.

[6] J. Mullins, (1996): One false step... *New Scientist*, pages 32–36, 4th May 1996.

[7] M. Roughan and D.W. McMichael, (1996a): Data fusion for land mine detection: Preliminary report. Technical Report 9/96, CSSIP, 1996.

[8] M. Roughan and D.W. McMichael, (1996b): Data fusion for land mine detection: Experimental design. Technical Report 10/96, CSSIP, 1996.

[9] M. Roughan and D.W. McMichael, (1996c): Data fusion for land mine detection: Feasibility study. Technical Report 18/96, CSSIP, 1996.

[10] J.C. Russ, (1995): *The Image Processing Handbook*. CRC Press, Boca Raton, 1995.

[11] D.F. Swayne, D. Cook, and A. Buja, (1991): *User's Manual for XGobi, a Dynamic Graphics Program for Data Anaylsis Implemented in the X Window System*. Bellcore, December 1991. Bellcore Technical Memorandum.