Automatic interpretation of infrared and optical reconnaissance imagery.

Booth, D.[a], Reno, A., Foulkes, S., Kent, P., Hermiston, K., Lewis, S., Ducksbury, P.
DRA, St. Andrews Road, Malvern, Worcs, WR14 3PS, England.

ABSTRACT

Progress is reviewed on the development of an all source image interpretation system which exploits complementary evidence from a range of experts. This co-operation may occur between feature detectors in different bands, between detectors searching for different types of feature, or between different types of detector of the same feature. Algorithms for detecting vehicles in infrared linescan imagery give a low missed detection rate but have been found to respond falsely to: roads fragmented by trees; structures such as cylindrical storage tanks; and to corners of man made objects, such as buildings. False alarms are reduced by applying algorithms which detect subclasses of false alarms reliably i.e. buildings and storage tanks. In addition, both are features of interest in themselves, and are useful primitives in the identification of sites. The integration of depth (in the form of disparity maps) is examined as a means of reducing false building detections. Outputs from the feature detectors are combined using a simple rule-based approach. A surface based model matching technique is examined as a means of classifying the remaining vehicle candidates.

Keywords: Feature detection, image interpretation.

1. INTRODUCTION

This paper reviews progress on the development of an all source image interpretation system with which to provide cues for human photographic interpreters and image intelligence for a strategic intelligence fusion test bed. The features of military interest could be summarised as any sign of human activity; past or present. Many of the features, such as airfields, military installations and ports, form extended regions over which the application of automatic recognition techniques would be directed at identifying sub-features (or primitives) and the spatial relationships that normally exist between them. The detection of several primitive features, such as vehicles and buildings, is examined, and evidence from a number of such detectors is exploited to their mutual benefit. Preliminary work on the development of a facility for recognising extended sites is also described.

Several themes run throughout this work:

1. The exploitation of complementary evidence. This can occur between sensors (e.g. infrared and optical), between detectors of different feature classes (e.g. vehicle and building), and between different types of detector of the same feature (e.g. edge and region based approaches).
2. Ease of integration of additional imagery cues.
3. Employment of a progressive focus of attention strategy (comparatively inexpensive filters reduce the search area to be examined by more sophisticated algorithms).
4. Exploitation of prior and domain knowledge to invoke appropriate algorithms and define their inter-relationships.
5. And in the longer term, exploitation of algorithm performance measures for identifying domain changes and for providing the weights necessary for evidence fusion.

We consider the Jaguar reconnaissance system which is comprised of an infrared linescan (IRLS) and five optical cameras. The IRLS sensor is a roll stabilised BAe 401 with 1.5 inch lens and having a 120 degree field of view (nadir +/- 60 degrees). The optical cameras are F95s. One is forward looking and the others are arranged in a strip under the aircraft such that together they give 180 degree coverage with very little overlap. The outer cameras each have a 3 inch lens, while the inner ones (those used here) have 1.5 inch lenses. At present the imagery is recorded on wet film, however, in anticipation of future systems, rudimentary position and attitude information is assumed.

Algorithms for detecting vehicles in infrared linescan imagery were found to respond falsely to roads that had been fragmented by trees, to structures such as cylindrical storage tanks, and to corners of man made objects, such as buildings. False alarms are reduced by using algorithms which detect a subclass of the false alarms reliably, in this case buildings and storage tanks. In addition, both are features of interest in themselves, and are useful primitives in the identification of extended sites. The integration of depth (in the form of disparity maps) is examined as a means of reducing false building detections, and the results are combined using a simple rule-based approach. The remaining vehicle candidates are classified using a surface based model

a. Correspondence: D.M. Booth; E-Mail dmbooth@dra.hmg.gb; Telephone: +44 (0)1684 894529; Fax: +44 (0)1684 896490

matching technique. Automatic image registration techniques are examined as a means of relating information extracted from the IR and optical bands, and remove global affine distortions between optical frames prior to disparity map construction.

2. SYSTEM ARCHITECTURE AND DATABASE

Figure 1 illustrates the straightforward but flexible design of the proposed system. An ellipse denotes a control mechanism; a rectangle denotes an information store; and the flow of data is indicated by directed lines. Solid and dashed lines signify mandatory and optional data flow, respectively.

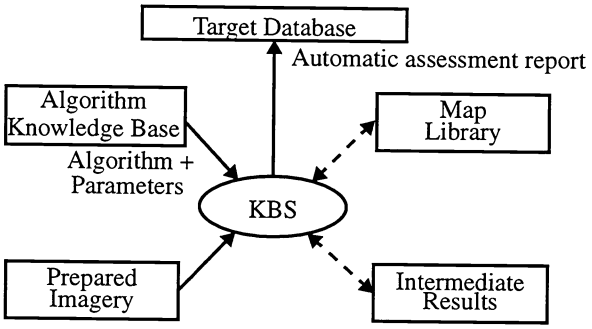

Figure 1 - The Analysis of Prepared Imagery

Examining each of the components in turn:

1 The prepared imagery store contains both raw and preprocessed imagery.
2 The algorithm knowledge base is a repository of algorithmic information. It will include the algorithms themselves, information on their parameterisation, usage and scheduling (if necessary).
3 The map library is a repository of map specific information generated both internally and externally to the system.
4 The intermediate results store provides a general purpose scratch area during processing.
5 The target database contains information pertaining to the potential targets discovered in imagery.
6 At present the knowledge-based system adopts a simple rule based approach.

Given a space-time window and, if desirable, particular features and sensors of interest, the KBS will identify available imagery, and, via the algorithm knowledge base, identify and schedule the appropriate algorithms. These will include target detection, image registration, construction of terrain models, evidence fusion etc. Any map information or digital elevation models that may be required are retrieved by the KBS from the map library. As each algorithm is applied, the results of the analysis can be stored as intermediate, together with an appropriately completed interim assessment report. An audit trail must be maintained because different reporting categories may include common features, and hence, classification decisions may change as new sources of evidence arise. After processing the available data, automatic assessment reports are generated for each target hypothesis and dispatched to the target database for verification.

3. THE DETECTION OF LAND FEATURES

Figure 2 outlines the algorithm configuration that has been used in this work. Briefly, the automatic image registration techniques provide a means of relating information extracted from the infrared and optical sensors, and also remove global distortions between optical frames prior to the construction of a local disparity map and change detection. As a precursor to this, the infrared linescan imagery requires destriping. The focus of attention mechanism is intended to detect localised regions of activity (particularly small encampments) over which more computationally intensive feature detection algorithms can be applied. Detection algorithms have been developed for locating vehicles, buildings and cylindrical storage tank-like features. These are features of interest in themselves, and are useful primitives in the identification of extended sites. The integration of depth (in the form of disparity maps) is examined as a means of reducing false building detections. Finally, the remaining vehicle candidates are recognised using a surface based model matching technique.

4. PREPROCESSING

Preprocessing of infrared linescan imagery typically involves destriping and rectilinearisation. For imagery acquired in digital form, scan lines map directly onto lines of pixels in the image, making striping readily quantifiable and distortions in general easier to correct[1]. Imagery from the Jaguar reconnaissance cycle is in hard copy form and must be scanned prior to analysis. Imperfect alignment between detector and image samples means that algorithms relying on line-to-line comparisons break down. In such cases scan line noise removal is best performed using Fourier techniques. Here the infrared linescan imagery was destriped by applying a narrow 'bow tie' filter to the image's Fourier transform in a direction approximately perpendicular to

that of the striping, and the resulting hole filled by interpolation. The results were not appealing visually as there was some smoothing, and detector saturation caused some further artefacts to be introduced. However, the improvement was sufficient to enable the image registration algorithm to operate satisfactorily.

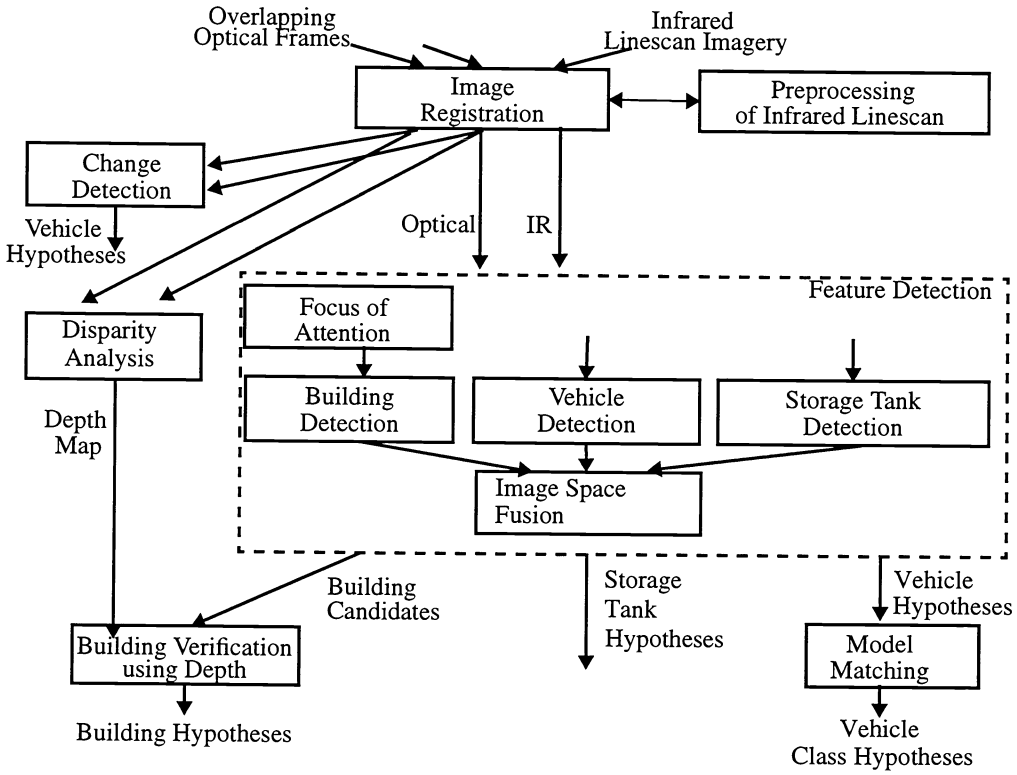

Figure 2 - The detection of land-based features in Jaguar reconnaissance imagery

5. IMAGE REGISTRATION

For the registration of unimodal imagery we have used the total pixel-by-pixel intensity difference as a measure of misalignment. This was used in preference to other measures (such as correlation) because it permits the application of efficient least-squares estimation methods, and hence the ability to reach an iterative solution given favourable starting conditions (i.e. the initial alignment of images is assumed to be consistent with that achievable using annotation data). Metrics based on intensity difference can be susceptible to changes in illumination because the best alignment of the images may no longer correspond to the minimum intensity difference. This is overcome in part by using statistical measures of image intensity to correct for global illumination differences between the images.

The technique above has been applied to the registration of images of different modalities by using edge features as a common representation upon which to base the matching process. However, the binary representation of edges means that vast amounts of image information is neglected. Two potentially more robust algorithms based on phase and entropy have been investigated.

Phase based algorithms provide a locally contrast invariant measure of image match that is more robust to illumination variation. Phase values (which are related to local Fourier phase) are obtained by measuring the responses of filters derived from single cycles of sine and cosine functions. In the basic form the phase methods can only match images of the same modality. However, the methods have been further developed to be invariant to contrast inversion such as might occur when matching infrared and optical imagery. The technique matches the images by measuring the similarity of underlying variations in intensity (which effectively correspond to edges). However, there is no requirement to actually extract edges from the images: the phase response provides a continuous measure of the likelihood of a pixel being an edge. The phase measures discard underlying constant intensity values (which can be expected to vary between images) but retains a highly descriptive representation of the intensity variation within the image. Phase based methods are not as effective at matching multi-modal imagery. There is an implicit assumption that the changes in intensity (on which the matching is based) are linearly related between the images. This assumption is approximately correct in local areas, e.g. in the immediate vicinity of edges, but the property does not in general hold over larger areas in images of different modalities. Despite this and other more general problems, the technique has produced useful results.

The mutual information (or entropy) metric[2] assumes that the distribution of intensities within the images are in some way mutually dependent. For example, any image generally consists of distinct regions of different intensity: in another modality it is assumed there will be a similar distribution of distinct regions albeit with different intensities. When aligned these regions will coincide pixel-by-pixel between the images. This causes the co-occurrence probabilities of intensity values between the images to tend to more extreme values (i.e. high or low probabilities when compared with random chance). The mutual information measure quantifies this dependency which is assumed to be maximal when the images are in alignment. It should be noted that this measure does not apply the same limitation as phase registration in that there are no implicit assumptions about the relationship between intensity values between the images. However, the form of the entropy measure is a discrete sum over the distribution of pixel intensity values: although it is simple to evaluate how well the images are aligned it is difficult to determine how the alignment may be improved. In practice a number of approximations are required (such as randomly sampling a small number of pixels from the images) in order to produce a computationally tractable algorithm.

Initial experiments with entropy based registration algorithms have shown promising results. The greater the difference in the imaging mechanism (e.g. frequency in the electromagnetic spectrum) the less mutual information will be contained in the imagery reducing the likelihood of successful registration. Additionally the algorithm implicitly assumes that each individual pixel in the image is indicative of the local area: this property does not hold in highly textured or noisy images such as SAR. Nevertheless, the mutual information registration technique provides a useful alternative to phase based methods: mutual information techniques match images using regions of constant intensity whereas phase methods register using intensity variation.

6. VEHICLE DETECTION

Vehicle detection[1] is a 5-stage process, incorporating pre-processing, vehicle detection, segmentation, feature extraction and classification. Pre-processing incorporates scan line noise removal and correction of across heading geometrical distortion which would otherwise introduce unnecessary variability into vehicle and non-vehicle models. Vehicle detection and segmentation are achieved using two detection filters. The first is an approximation to a difference of two Gaussians filter that detects localised bright regions that contrast greatly with their surroundings. The second filter (which also has the dual role of an object segmenter) is based on a difference of two medians filter, which identifies possible vehicles by subtracting a local background model. Finally, locally adaptive thresholding is applied to extract the shape of the objects. The matched filter acts as an enabler to the co-median filter. Thus the desirable properties of both filters are utilised, i.e. robustness to false alarms and retention of shape information. Features are extracted from a local neighbourhood centred on the object of interest. Five types of feature have been examined: shape, object grey level distribution, object to background contrast, texture and aspect. A morphological closing operator is applied to the thresholded image to improve performance of the shape features. After an investigation into the performances of different classifiers and classifier arrangements, a Fisher linear discriminant was decided upon. Performance is illustrated in Figure 3.

7. SEGMENTATION IMPROVEMENT

The acquisition of quality target segmentations is crucial to the success of subsequent characterisation tasks. Most image segmentation techniques are based on thresholding. Even if it is locally adaptive, thresholding is a point operation. This leads to pixel drop out, and ragged boundaries etc. The Geman and Geman[3] relaxation algorithm (and some derivatives) have been examined as a means of improving the quality of segmentations. In common with other relaxation algorithms, a decision as to whether a given pixel is target or non-target is influenced by the classifications of its neighbours. In addition, the Geman and Geman algorithm has the facility to include prior knowledge of the spatial relationships between adjacent pixels, and also maintains consistency with the underlying data: there is a cost associated with deviating from an initial segmentation. Given the reasonable quality the segmentations produced by our adaptive co-median filter, this should alleviate one of the problems encountered previously, that of convergence to an implausible solution, usually the merging of target and neighbouring background features.

Two versions of the relaxation algorithm have been implemented[4]. The first is essentially a sophisticated smoothing algorithm which is applied as a pre-processing stage prior to segmentation. This is similar to the Synthetic Aperture Radar despeckling algorithm proposed by White[5]. The main difference is in the choice of noise model. The second version is based on the full implementation of Geman and Geman's algorithm. However, the computational overheads involved in minimising the cost function using simulated annealing are sufficient that the process has been approximated by a Kohonen - Multi-Layer Perceptron hybrid neural network (trained to approximate the mapping).

8. FOCUS OF ATTENTION

Feature detection and recognition are computationally intensive processes which, for practical reasons, ought to be directed at local regions rather than at an image as a whole. A useful strategy is therefore to process an image with a computationally inexpensive interest operator to identify potential features on which to focus the attention of more sophisticated algorithms. Here images are scanned for regions which hold the potential for containing buildings. As buildings can be characterised by the presence of corners, a filter is chosen which gives a maximal response at the vertex of an approximately right-angled corner in

the image. The Kitchen-Rosenfeld detector[6] yields such a response at the zero-crossing contours of the image convolved with the Laplacian of a Gaussian filter. The edge representation defined by the contours of the image convolved with Laplacian of a Gaussian coincides with the vertex of a corner although it is displaced around the corner itself[7]. The edge representation defined by the zero-crossing contours of the second directional-derivative in the direction of the edge normal generally does not pass through the corner vertex. The procedure we have chosen to detect corners therefore involves computing the Laplacian of a Gaussian of the image and the Kitchen-Rosenfeld corner response. The corner response is accepted only where there is a significant edge response from the Laplacian. The scale of the Gaussian determines the scale of the corners that are detected.

The approach above has been applied to several grey-level aerial images[8] (Figure 5a). In these results, the response image of the corner detector was convolved with a wide Gaussian to form regions of high corner density. The edges of these regions were computed and are shown overlaid on the original images.

9. BUILDING DETECTION

9.1. Edge Based Building Detection

This is a three stage process based on the work of Krishnamachari and Chellappa[9] which involves straight edge extraction, line merging, and line grouping. Edge elements in the image are detected using any conventional detector and then parametric representations for lines of edges are derived using a random Hough Transform. Line merging is designed to join small line segments that are likely to belong to the same line. This is a multi-pass algorithm which merges on the basis of line separation and angular difference. Unfortunately, the result of a merge will not necessarily be representative of the underlying image, as the potential exists for unrelated lines to be merged and be replaced by a hybrid which reflects neither of the image features. Line grouping is based on the optimisation of a cost function concerning the spatial relationships between line features that are likely to characterise a building. The cost function favours lines that are collinear, parallel or orthogonal.

The algorithm was tested with various images but in most cases its performance was quite poor. One reason may have been the large number of parameters involved, and the fact that they are domain dependent. The effect on parameters of some variables (such as aircraft height) may be predictable, but others such as general image quality and background clutter characteristics are less so. Image quality influences the parameterisation of the line grouping stage. In good quality, high contrast imagery, line merging can be a costly overhead, while in other cases it is essential though potentially error prone. The background clutter, such as that present in any urban environment, is likely to exhibit a sufficiently systematic grouping of features to produce false alarms. For example, car parks and road patterns may produce false detections, and the exploitation of depth cues would be invaluable. In addition, buildings or building like objects that are in close proximity will interfere and tend to reinforce and propagate errors. As a consequence, this edge based algorithm is best suited to the detection of isolated, or small groups of buildings, on a background that is essentially natural. Desert terrain is one such example. See also Figure 5.

9.2. Region Based Building Detection

In brief, an image is segmented into regions which exhibit a planar variation of intensity. These form a set of building candidates which are accepted or discarded on the basis of a number of characteristics: area; compactness; contrast with surroundings; Fourier shape descriptors and global shape entropy. Classification is achieved by computing the probability that each of the features is representative of a building (given class conditional probability density functions estimated from training data), and these probabilities are then combined using the Bayes theorem to give an overall classification probability, and classified accordingly. Some hole filling may be necessary depending on future requirements. The algorithm is described in[8] and some results are shown in Figures 4b and 5c.

The main weakness with this approach is that the region classification stage depends heavily on the results of the segmentation. Segmentation fails when regions are highly textured, exhibit non-planar intensity variation, or are very small. Textured regions are often severely fragmented by the segmentation process. A more goal driven approach to segmentation is now being investigated.

10. STORAGE TANKS / CIRCULAR FEATURES

Several types of military target could be characterised, at least in part, by the presence of a circle or ellipse. These include POL (petroleum, oil and lubricants) storage tanks, artillery dug-outs, radar domes and dishes. Davies[10] describes a number of Hough Transform based algorithms for detecting circles and ellipses in imagery.

The relatively straightforward detection algorithm used here is applied to an edge map. This is produced by a standard Sobel edge detector (including nonmaximal suppression and thresholding with hysteresis) followed by a fractal discriminant which removes the ragged lines generated by image texture. The Hough transform then considers each edge element to lie on the circumference of a circle. Given the position and orientation of the edge element, and the radius of the circle, two votes are cast as to the position of the circle centre, one on either side of the edge element (assuming no knowledge of the relative brightness of target and background). Votes are cast by all edge elements. Those pixel centres having significantly more votes than their

neighbours are the ones most likely to represent circles.

To reduce false alarms from clutter, separate searches are made for objects of each target radius. However, in order to handle noise both in the image and in the edge detection procedure, and also allow for slightly oblique viewing angles, pixels local to the potential centre are allocated a vote of reduced weighting. Peaks are identified on the basis of the minimum acceptable proportion of a circle that must be present in order to constitute a hit. This algorithm is quite crude, even so, given the simple parametric model of a circle, and that circles do not tend to occur in imagery naturally, circles can be detected quite reliably. False alarms are usually caused by textured background rather than other man made objects. Figures 3b and 4a show the algorithm in operation.

Ellipse detection is also a practical proposition, particularly if annotation data is available for identifying viewing angles and hence constraining search spaces. This is particularly true of infrared linescan imagery where the pixel position along a scan can determine the viewing angle.

11. DEPTH CUES

Many man-made structures, such as car-parks and road segments and other non-elevated features, can often be mistaken for buildings on the basis of two-dimensional data. Depth can provide confirmation that a particular feature genuinely represents a three-dimensional structure.

The depths of objects in a scene can be recovered by employing a stereo imaging system: the relative distance, or disparity, between corresponding features in two overlapping image frames is approximately linearly dependent on its depth. Close objects yield higher disparities than more distant ones. Thus, matching features in the two images allows the relative depth of an object to be recovered. Generally, there is a trade-off between the complexity of the monocular and matching processes. At one extreme, object recognition can be performed before matching, whilst towards the other, matching can be performed on more primitive features such as edges or peaks in image intensity.

Global affine distortion between frames can be removed using the registration techniques described earlier. Residual local disparities are determined by region based matching. Two algorithms have been examined: one based on correlation coefficient, and the other on variance of residuals. The former offers superior performance but is much more expensive to compute. However, the overheads can be reduced by computing the correlation coefficient for regions centred on every pixel in the image for a given disparity in one batch, and keeping a record of the best match for each pixel. This approach means that correlation coefficient can be computed in terms of convolutions, which may be supported by hardware accelerators or perhaps carried out in Fourier space. In addition, the reference frame is constant and therefore many statistics need be computed only once, and normalisation can be omitted (Figure 5b).

The variance of residuals metric could also be computed in terms of convolutions. Alternatively, adopting a pixel-by-pixel approach, a measure of how likely it is that the disparity at a particular pixel location is the same as that of the previous one can be obtained by a statistical comparison of the corresponding variances of residuals (using an F-test or B-distance, for example). If they are considered the same, then the range of possible disparity values need not be explored. A computational saving of up to 85 percent over an exhaustive search has been achieved without degrading performance enough to hinder subsequent use.

The pixel representation is converted into regions by thresholding with hysteresis: pixels are accepted providing that they are above a lower threshold, and are connected by other accepted pixels to a pixel which is above an upper threshold.

12. FUSION

12.1. Rule Based Fusion

We share the development approach of Eklundh[11] who advocates the implementation of vision systems and their application to real, natural environments. These experiments may reveal quite powerful selection strategies, rather than the often used fusion by averaging approach which may be adopted by accident if not by design. The aim is to build complex systems step by step by adding operational models, thereby adding more and more competencies. The use of confidence measures is also advocated in order to signal the failure of a vision agent and consequently that a new ranking is required (i.e. the problem domain has changed). Experiments have shown the following:

1 The vehicle detector has a very low missed detection rate but is susceptible to false alarms on the corners of buildings, on storage tanks, and on roads fragmented by trees.

2 The storage tank detector operates with some reliability and, because circles do not arise in nature that frequently, its false alarm rate is low. In particular, does not respond to vehicles.

Conclusion: if an object is detected by the vehicle detector and the storage tank detector it is almost certainly a storage tank.

3 The region based building detector finds most buildings but can also mistakenly extract ground features.

Conclusion: Buildings are larger than vehicles. If a vehicle candidate is a subregion of a building candidate then the vehicle candidate is a false alarm.

The logic above is demonstrated in Figure 4.

4 The tall object detector is based on region matching and is unreliable over flat regions such as fields. The other feature detectors operate at their best under the same conditions. The tall object detector works reasonably well over featured terrain. In particular, it will generate false alarms rather than miss detections.

Conclusion: The tall object detector can be used as an enabler to building and storage tank hypotheses (Figure 5).

The logic above was carried out in image space using simple masking and intersection operations. This is quite acceptable given the clear prioritisation between vision agents, and has proved to be fast and reliable. Masking has been used to remove potential vehicles given over-riding information from the storage tank and building detectors. Circles are represented completely by a parametric model and these were projected back onto the image as a mask, thus erasing any potential vehicle beneath. Similarly, regions (surfaces) identified by the building detector can be turned into masks since there is practically no danger that a vehicle lies somewhere within that area. Building verification by the tall object detector is less straightforward because of occlusion, and the behaviour of the correlation metric as it moves across surface boundaries, can introduce errors. As a temporary measure, if a potential tall object intersected with a potential building, then the building hypothesis was accepted. This is acceptable because as long as the tall object detector does not fail to respond to a building, then nothing is lost by the operation. In general, as long as agents generate false alarms rather than miss detections, they provide evidence that is easy to handle. Rankings may vary for different terrain types and for different sensors, and this must be recognised either from prior knowledge, from a domain classification scheme, or from algorithm performance measures.

12.2. Evidence Based Fusion

In the example above, the priority ranking between the detectors was quite clear cut. The relationship between the vehicle detector and the building detector is particularly so, and it is difficult to see how an evidence fusion approach could improve reliability. However, this sort of priority ranking will not work when supporting evidence is arriving from several sources. Ducksbury, Booth and Radford[12] have applied Pearl-Bayes networks to the vehicle detection problem described previously. Supporting evidence is provided by vehicle track and shadow detectors. The spatial arrangement of the vehicles provides the means for marginal detections to reinforce one another. Spatial relationships are also exploited by allowing a combined belief in support of a vehicle to feedback to the detectors and influence their sensitivity within the local neighbourhood.

Over the terrain being considered, vehicle detection and classification have been reasonably successful. Evidential reasoning does not have a major effect when the vehicle classification results are reliable (and certain) but is more significant when the classifier decisions are more marginal. In such cases, the vehicle classifications have been reinforced by the supporting evidence.

More generally, the detection algorithms described in this report could produce estimates of their own performance. The vehicle and region based building detector both output detection probabilities. Hough transform and energy based techniques, such as the storage tank detector and edge based building detector, can be adapted to produce confidence measures on the basis of the height of the target peak in relation to its neighbouring histogram bins. The matching metrics used for computing stereo disparity both have formal measures of statistical significance.

13. CHANGE DETECTION

In urban regions feature detectors may become confused: there is insufficient background to generate representative statistics; and objects are in such close proximity that they become indistinguishable. However, as the presence of urban regions should be available either from maps or from automated segmentation techniques[13], a more useful cue might be signs of military activity, that is, scene changes.

Registration of imagery originating from a single sensor is comparatively straightforward provided that the view points are similar, and illumination and weather conditions remain constant. These constraints are usually satisfied when overlapping frames of imagery are taken during a single fly by. Once registered, scene changes, and in particular moving objects, can be highlighted by image subtraction, although a certain amount of noise and structure will exist in the difference image: the registration transformation is a 'best' global approximation, and any feature for which it is not representative, will register as a change. As well as being well above the ground plane, trees and other vegetation can move sufficiently to generate a substantial amount of noise. A moving object normally appears as two distinct regions in the difference image. One at the initial position, and another (oppositely signed) difference at the final position. The magnitude of the difference will depend on the contrast between the object and the background at both positions.

Figure 6 demonstrates the automatic detection of moving objects using essentially the same vehicle detection algorithms as those described previously. Assuming the target is always brighter than the background enables the initial and final positions of the target to be determined (and marked by black and white surrounding boxes, respectively). Automatic matching is necessary to associate pairs of initial and final positions, and hence track vehicle movements.

14. VEHICLE RECOGNITION

A 3-D model matching technique which maximises mutual information between surfaces[2] has been used for the recognition and pose estimation of vehicle candidates residing in the optical reconnaissance imagery. However, the algorithm often failed to converge when the estimate of the target position was several metres or more in error. In addition, the metric failed to provide a measure of mismatch that facilitated reliable clutter rejection. To overcome these problems, and generally improve recognition confidence, the mutual information metric has been extended to include edge information.

Mutual Information is a quantity based on the entropy of an observed object. Entropy in turn can be interpreted as a measure of disorder of a dataset. Views of an object which show many surfaces at different orientations to the light source (and therefore show different shading intensities) display high entropy measures. The Viola & Wells objective function[2] consists of three entropy terms. The first and second measure the single entropies of the image and model 2-D representations over the spatial extent of the model. The third term measures the joint entropy between the model and image over the same region. When the Mutual Information maximises, the model is translated and rotated into an orientation which presents a variety of surfaces to the observer which describe the collocated region of the model and target image well. The transformation T is sought which maximises the Mutual Information between model and target image vehicles.

$$T = \max\left(H(u(x)) + H(v(T(y))) - H(u(x),v(T(y)))\right) \quad (1)$$

where T is the transformation of the model in space. H are the single and joint entropies, u(x) is the dot product between line-of-sight and surface normal over the model extent and v(y) is the pixel intensity of the target image over the model extent.

Two sets of randomly chosen points were selected from both the target and model surfaces. The first set from each image was used to approximate the entropy probability density functions (PDFs) with which the second set was used to measure the Mutual Information. This is a technique is known as the Parzen window method which reduces the processing time and removes the need for the variable PDF to be known *a priori*.

With knowledge of the observing aircraft's position with respect to the target, the ground plane was derived and the model assumed to move on a flat earth with the usual six degree of freedom reduced to three (movements across the ground plane and heading). This produced a three-dimensional search space. Cross sections of this space with model heading correct and constant for the car model are shown in Figure 7b.

The car surfaces can be seen to misalign quite easily in view of both its near vertical observation angle and the lesser spatial extent of its surfaces. The noisy nature of the surfaces are due to the random sampling of the Parzen window method.

The algorithm was seen to converge from randomly selected start positions in a manner summarised in Table 1. As can be seen the convergence of the car model to the correct pose is achieved reliably with small translational offsets. Initial heading of the model is determined from a histogram of edge-directions and found to be accurate to within +/- 10 degrees.

Table 1: Convergence success using Mutual Information and surface-edge metric

Start offset radius (m)	final pose x (m)		final pose y (m)		final pose h (°)		convergence success	
	MI	E-S	MI	E-S	MI	E-S	MI	E-S
1.0	0.87	0.05	1.37	0.04	1.83	1.74	50%	100%
2.0	0.88	0.04	1.45	0.05	2.11	2.39	20%	100%
3.0	0.40	0.05	1.38	0.06	2.33	1.51	30%	100%

To provide an improved method of converging to the correct peak of Mutual Information, edge information was applied. Both the target and model images were automatically edge detected and thresholded to leave binary images consisting solely of major edge points. From a random sampling of these edge points the distance to the nearest edge point in the other image (along the edge normal) was summed. The original intention was for an additional Mutual Information measure, which used these distances, to globally align the edges of both model and target vehicles. However, in formulating the appropriate Mutual

Information term, it was realised that it was simply minimising the distance between model and image edge points. Examination of this simple term revealed the broad convex surface required for long-range gravitation (Figure 7c). Using this minimisation of edge-separation alone revealed that it quickly allowed convergence to the broad peak but was inaccurate in the final stages of convergence and produced oscillations about the peak.

The use of surface alignment alone to determine the pose and recognition of vehicles has proven inadequate. Surface comparison has not allowed the model to 'feel around' in space and gravitate towards the appropriate function maximum. Conversely, the edge alignment metric has the ability to gravitate the model to the appropriate region of the target image but has difficulty finely positioning the tank when there. An ideal metric for model alignment would combine the positive capabilities of both edge and surface metrics. This was done using the function shown in Equation 2.

$$I(u(x),v(T(y)) = H(u(x)) + H(v(T(y))) - H(u(x),v(T(y))) + [\ 1 - ((|c_i|+|d_i|)\ /((c_{max} + d_{max})\ ] \quad (2)$$

where c_i is the distance between randomly sampled edges points in the model image to the nearest edge points in the target image (measured along the edge normal at those points) and d_i vice versa.

Cross-sectional surfaces of the 3-D search space for the car example is shown in Figure 7d. Notice that due to the independent scaling of the surface and edge terms, the edge surface dominates the summation with the surface terms producing most effect at or near the maximum of the edge surface when course translational optimisation has been achieved. This is exactly what is required and manifests itself as more peaked surfaces at the maximums in Figure 7d.

The new metric is tolerant of edge fragmentation and the edges do not require labelling.

15. CONCLUSIONS

In carrying out the ground work for a multi-source image interpretation system the intention has been to tackle the detection and recognition problems as a whole, and in particular reach a position where complementary evidence from different types of vision agent is exploited to reduce classification uncertainty. In addition, having a baseline system (or analysis strategy) in place allows weaknesses in individual algorithms to be placed in context. In particular, we have demonstrated that even relatively crude algorithms can provide discriminatory information.

Future work: general purpose algorithms for segmentation and classification of objects are being developed. There is a particular emphasis on shape characterisation (and model matching) and constraining the problem by incorporating knowledge of the viewing parameters.

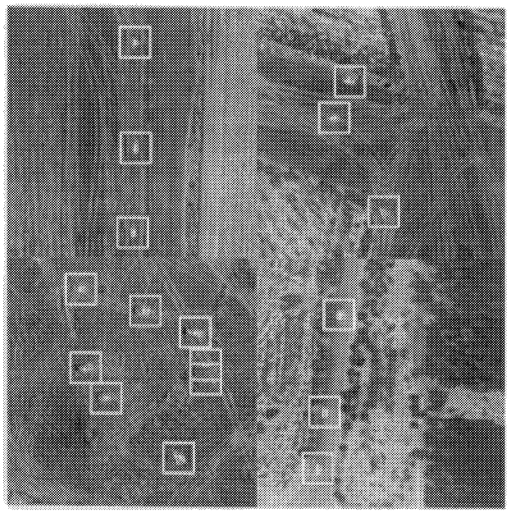

Figure 3a. Vehicle detections in IRLS.

Figure 3b. False vehicle detections. Recognised errors are marked with crosses.

4a

4b

Figure 4a. Part of an optical frame of Jaguar reconnaissance imagery with detected storage tanks marked.

Figure 4b. Potential buildings (dark grey) and vehicles (light grey). Conflicting evidence is coded white. In such regions the vehicle hypothesis is nullified.

Figure 4c. Candidate vehicles after buildings and storage tanks have been deleted.

4c

Figure 5a. Part of a frame of optical imagery.

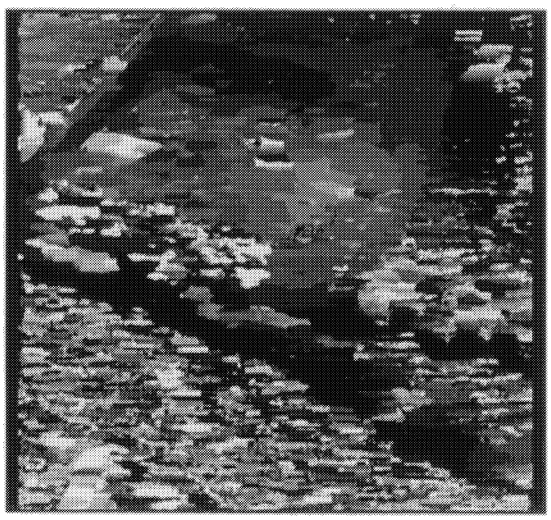

Figure 5b. A disparity map constructed from two consecutive frames of optical imagery. Global affine distortion has been removed automatically. White indicates maximum disparity, black is minimum.

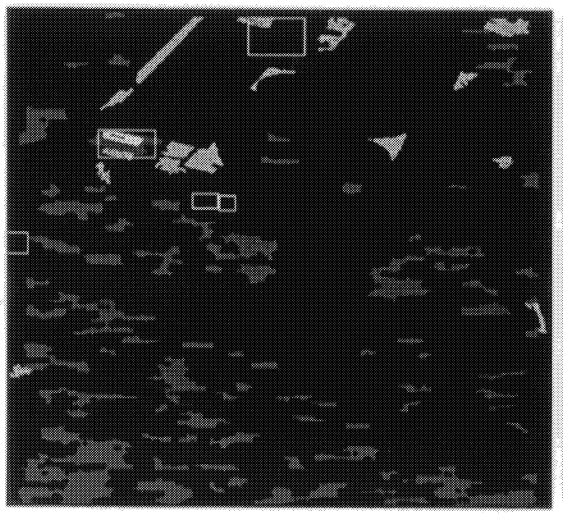

Figure 5c. Fusion of building hypotheses with depth information. Dark grey indicates regions of high disparity, light grey corresponds to region based building hypotheses, and white indicates a correspondence between the two. Rectangles indicate edge based building hypotheses.

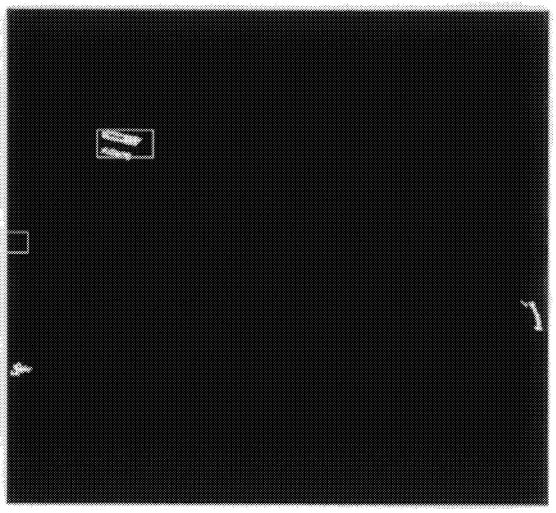

Figure 5d. A simple 'enabling' algorithm provides building hypotheses for the two algorithms when associated with depth information.

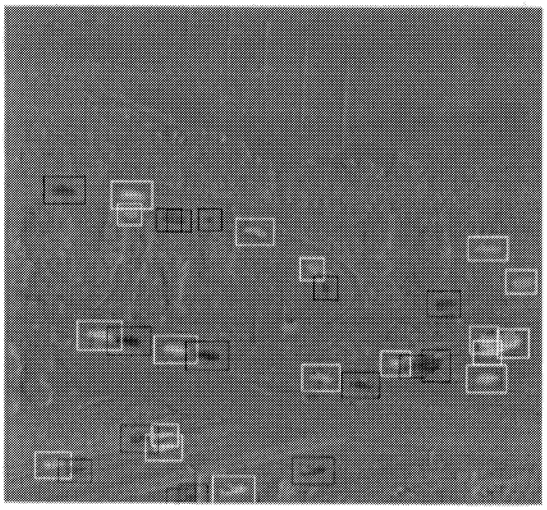

Figure 6. Automatic registration of overlapping frames of optical imagery followed by the highlighting of changes caused by potential vehicles.

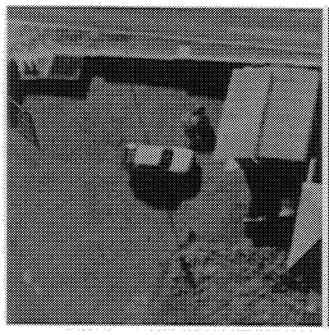

Figure 7a. Target vehicle image

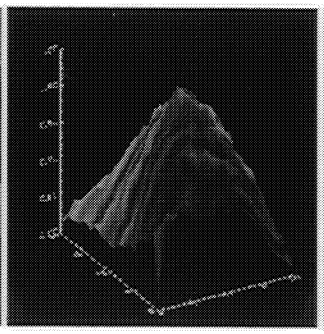

Figure 7b. Car M1 surface

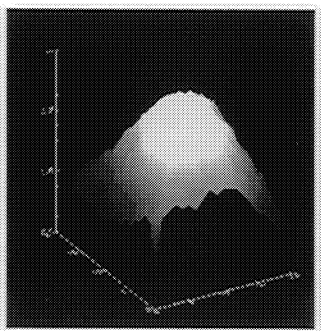

Figure 7c. Edge measure

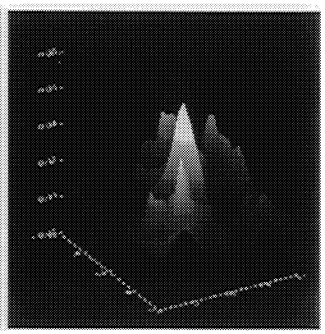

Figure 7d. Surface-edge measure

REFERENCES

1 Booth, D.M. and Radford, C.J. "The detection of vehicles in downward looking infrared linescan imagery," *Proc.12th ICPR, Jerusalem, Israel*, 1994.

2 Viola, P. and Wells, W. "Alignment by Maximisation of Mutual Information", *Proc. of 5th Int. Conf. on Computer Vision, Boston, USA*, pp16-23, 1995.

3 Geman, S. and Geman, D. "Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, **PAMI-5**, pp.721-741, 1984.

4 Foulkes, S. and Booth, D. "Improved target segmentation using Markov Random Fields, artificial networks and parallel processing techniques," to appear in *Proc. Image exploitation and target recognition*, SPIE Aerosense, Orlando, Florida, 1997.

5 White, R.G. "A simulated annealing algorithm for radar cross-section estimation and segmentation," *Proc. SPIE Int. Conf. on Applications of Artificial Neural Networks V*, Orlando FL, April 1994.

6 Kitchen, L. and Rosenfeld, A. "Grey-level corner detection," *Pattern Recognition Letters*, 1:95-102. Dec., 1982.

7 Berzins, V. "Accuracy of Laplacian edge detectors," *Computer Vision, Graphics and Image Processing*, **CVGIP**, 27:195-210, 1984.

8 Reno, A. "Detecting buildings in aerial images using shape descriptors," to appear at *IEE Int. Conf. on Image Processing and Applications*, Dublin, July 1997.

9 Krishnamachari, S., and Chellappa, R. "An energy minimisation approach to building detection in aerial imagery," *Proc. of ICASSP*, 1994.

10 Davies, E.R. *Machine Vision: theory, algorithms and practicalities*, Academic Press, London, 1990.

11 Eklundh, J. "Issues in active vision: attention and cue integration/selection," *Proc. British Machine Vision Conf.*, University of Edinburgh, 9-12th September, 1996.

12 Ducksbury, P.G., Booth, D.M. and Radford, C.J. "Vehicle detection in infrared linescan imagery using belief networks," *Proc. IEE 5th Int. Conf. Image Processing and its Applications*, Edinburgh, July 1995.

13 Ducksbury, P.G. "Parallel texture region segmentation using a Pearl Bayes Network," *Proc. British Machine Vision Conference*, University of Surrey, 21-23rd September 1993.