CONF-9510310--1

IMAGE EXPLOITATION USING MULTI-SENSOR/NEURAL NETWORK SYSTEMS

E. C. Uberbacher, Y. Xu, R. W. Lee, C. W. Glover, M. Beckerman, , R. C. Mann

Intelligent Systems Section
Computer Science and Mathematics Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee USA

To Be Presented At:

AIPR-95 WORKSHOP, "TOOLS AND TECHNIQUES FOR MODELING AND SIMULATION"

Cosmos Club
Washington, D. C.

October 11-13, 1995

***Research sponsored by the U. S. Department of Energy, under Contract No. DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.**

Image exploitation using multi-sensor/neural network systems

E. C. Uberbacher, Y. Xu, R. W. Lee, C. W. Glover, M. Beckerman, and R. C. Mann*
Intelligent Systems Section, Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6364

ABSTRACT

We have developed and evaluated a tool for change detection and other analysis tasks relevant to image exploitation. The tool, visGRAIL, integrates three key elements: (1) the use of multiple algorithms to extract information from images - feature extractors or "sensors", (2) an algorithm to fuse the information - presently a neural network, and (3) empirical estimation of the fusion parameters based on a representative set of images. The system was applied to test images in the RADIUS Common Development Environment (RCDE). In a task designed to distinguish natural scenes from those containing various amounts of human-made objects and structure, the system classified correctly 95% of 350 images in a test set. This paper describes details of the feature extractors, and presents analyses of the discriminatory characteristics of the features. visGRAIL has been integrated into the RCDE.

Keywords: image exploitation, computer vision, pattern recognition, neural network, sensor fusion

1. INTRODUCTION

As the volume of routinely collected imagery increases dramatically, the ability of image analysts (IAs) to scrutinize the data and produce reports in a timely and reliable fashion is being pushed to the limits. Image exploitation (IE) combines a number of different clues in order to characterize images of sites of interest according to a profile of criteria determined by the particular application. Significant effort is being expended on developing algorithms and corresponding computing environments to support image exploitation tasks [1, 2]. This on-going work is generating significant advanced capabilities in model- and context-based image understanding, image registration, model construction, change detection, and several other important image analysis functions, including a common development environment for integration and evaluation of new functionality.

A characteristic feature of the applications targeted by IE efforts is the very large volume of data. The infrastructure necessary to handle large amounts of digital imagery presents significant technical challenges. Key issues are indexing of visual information and scalability of imagery analysis methods to high data rates and volumes.

The level of automation is driven by specific application requirements. For example, routine screening of selected image areas in large imagery databases for detection and classification of certain changes can potentially be highly automated, whereas the construction of site models remains a more interactive process.

Higher level analysis of imagery, e.g., the recognition of objects, tracking of objects in time-varying imagery, etc., is fundamentally a process that relies on the integration of data and information from multiple sources. These can consist of various feature extractors,

* Send correspondence to R. C. Mann, Oak Ridge National Laboratory, Bldg. 6025, P. O. Box 2008, Oak Ridge, TN 37831-6364, e-mail: mannrc@ornl.gov

e.g., edge detectors, surface estimators, etc.; spatio-temporal filters; Hough transforms; morphological filters; graph-matching modules to determine similarities with prior information, such as CAD databases; and many others. Contrary to earlier efforts to develop reliable image understanding (IU) methods, there is currently increasing focus on using any available contextual information, i.e., information other than the original image pixel values, in order to improve the performance and reliability of IU algorithms [3, 4, 5]. A number of computer vision research and development groups have achieved remarkable successes with various applications of context-based vision, including aerial image exploitation [6], surveillance and monitoring-type tasks for advanced traffic management systems [7], and others involving real-time analysis of video [8]. The use of functional, linguistic, or other type of context can result in greatly improved performance. The associated loss of flexibility or less wide-spread applicability of a vision system is a small price to pay.

Context is commonly derived from sources such as site models for aerial image analysis, CAD models for robotic inspection and object recognition, verbal descriptions of scenes, or knowledge about the function of objects in a scene. Notwithstanding the effectiveness of context-based vision systems, the manner in which context of any kind can be used systematically and optimally in some sense remains an open issue. Use of context depends, of course, on the problem at hand. Several image analysis approaches and tools such as perceptual grouping, and model-based object recognition make implicit or explicit use of context.

The use of contextual information derived from other physical or logical sensors has been an active field of research and development, most commonly referred to as sensor fusion [9]. Work in the area of decision fusion [10] is closely related to these efforts, often building (explicitly or implicitly) on previous work in statistics (e.g., [11]). For many practical applications it is impossible to obtain *a priori* all the necessary information in order to design optimal fusion schemes. Recent work on using empirical data, obtained from finite samples to estimate fusion mappings provides guidance for the design of systems that can learn from empirical data and improve their performance with time [12, 13, 14, 15, 16]. The integration of such machine learning algorithms into computer vision systems is an active field of research.

The objective of the work described in this report was to evaluate the performance of a multi-sensor/neural net pattern recognition system when applied to aerial image exploitation, specifically, reliable detection of new construction activity in sites of interest. The system was intended originally and used successfully for genome informatics applications [17, 18]. The approach involves (1) the use of multiple algorithms to extract information from images including contextual information - feature extractors or "sensors", (2) an algorithm to fuse the information - a multi-layer feed-forward neural network, and (3) a set of well-characterized images to train the overall system.

The paper is organized as follows. Section 2 provides additional background and a description of the set of image analysis problems to be solved. Section 3 describes the feature extractors used as well as the training procedure for the system. Results are presented and discussed in section 4, and conclusions and open issues are presented in section 5.

2. BACKGROUND AND PROBLEM DESCRIPTION

The multi-sensor/neural net system GRAIL (Gene Recognition and Analysis Internet Link) was originally developed by ORNL researchers in response to a critical need in the Genome

Program for reliable methods that could locate in DNA sequences highly variable patterns of interest that occur very infrequently (genes) in extremely large amounts of data [17]. GRAIL is a pattern recognition system that combines statistical and syntactic methods for analyzing complex input patterns. A multi-layer feed-forward neural network receives inputs from N sensors that measure different characteristics of the signals or data sets to be analyzed. The net acts as a classifier and assigns the input pattern to a given number of classes [18]. The net is trained using a set of known data patterns that are representative of the application domain. In its simplest form, GRAIL operates with sensors that give signals that are either "on" or "off", indicating the presence or absence of a particular pattern or characteristic. Sensors can also supply "analog" input signals to the integrating network. The reason for using multiple sensors is that for environments of reasonable complexity not any single sensor can reliably detect and classify a pattern of interest. The neural net represents a systematic mechanism to integrate the information from multiple sources to form a combined best estimate of the true classification decision. We interpret the term "sensor" in a broad sense. It can encompass a real physical sensor device and/or an algorithm that computes a feature of a signal acquired by a physical sensor. The term "logical sensor" has been used in this context. Hence several or all of our logical sensors may operate on data from one physical sensor.

GRAIL has been used since 1991 routinely by over 3000 researchers world-wide through electronic mail and client/server access. Usage has been increasing steadily. It is currently the standard method for finding genes in DNA sequence. It has won a 1992 RD100 Award recognizing the 100 most technologically significant new products of the year. Several versions of GRAIL, e.g. different sensors and different network architectures, are being tested in additional molecular biology applications. The ORNL GRAIL server handles an average of 7000 analysis requests every month.

Large amounts of data, and rare and highly variable patterns of interest are also characteristic of the problem considered in this paper: to detect change in the form of new construction activities in aerial imagery. Such activities can include excavations or other movement and accumulation of soil, movement of construction tools and materials into an area of interest, erection of structures, etc. We assume that there are reasonably accurate site and camera models so that image coordinates can be related effectively to world coordinates. We assume further that there is a set of images that are characteristic for the kind of changes to be detected, as well as a set of images that do not contain the changes to be detected. For the work described in this paper, we formulated the problem of finding new construction as equivalent to finding sufficient evidence for significant structure to appear in a well-defined area of interest in the imagery collected over time. This particular problem was intended to be a paradigm for other applications to be discussed later in this report.

Work to date has focused on designing feature extractors (sensors) capable of providing measures for structure, on estimating the parameters of the fusing network based on empirical data, and on integrating an initial version of the visGRAIL software into RCDE.

3. SENSORS AND TRAINING OF THE SYSTEM

Figure 1 shows representative examples of imagery with and without human-made structure. One key feature of the visGRAIL system design is the capability to integrate edge- and region-based features. This is particularly important to allow for detection of early signs of change due to construction during which not many changes manifest themselves in sufficiently long straight lines. It is also a capability that is important for

performing detailed shape analysis of items in the areas that have been determined to reflect change.

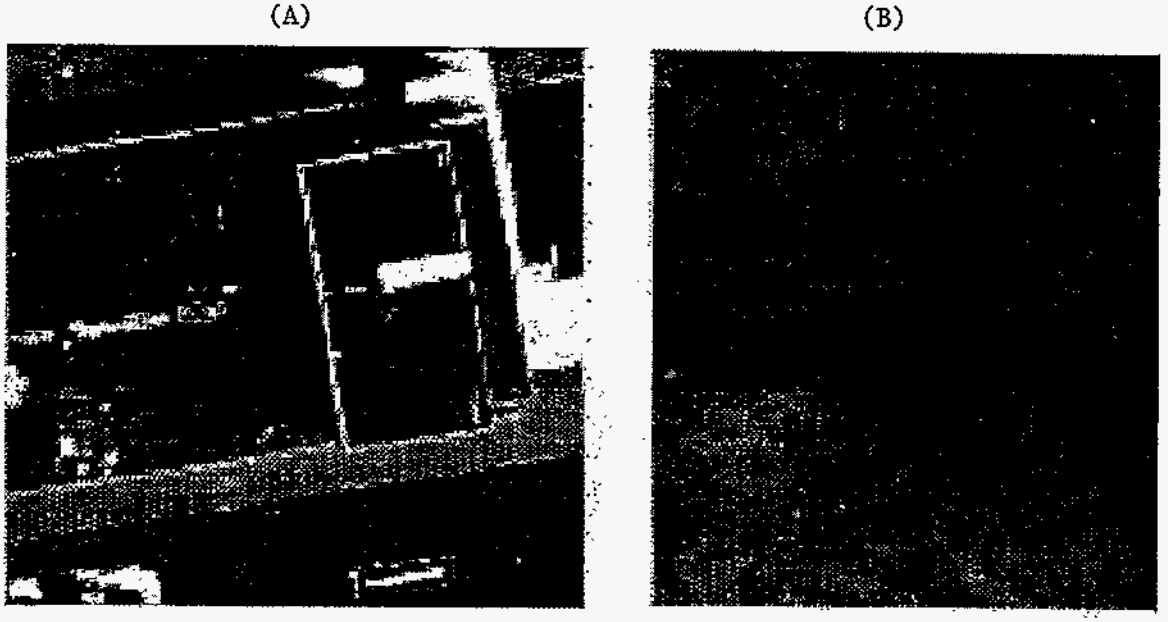


Figure 1: Examples of Aerial Imagery with (A) and without (B) evidence for human-made structure.

Since our goal was to design and develop an initial version of visGRAIL so that it can be integrated into the RCDE testbed, we used, whenever possible, algorithms that were also implemented in the current RCDE. For example, the Canny operator was used for edge detection.

For the region-based analyses, we developed a new efficient method for segmenting gray value images [19] into connected homogeneous regions. This new algorithm constructs a minimum spanning tree (MST) representation of the gray value image and reduces the region partitioning problem to a MST partitioning problem, thereby reducing the computational complexity of the segmentation problem. The MST corresponding to an image is partitioned into subtrees that represent homogeneous regions by minimizing the sum of variations of gray levels over all subtrees under the constraints that each subtree has at least a specified number of nodes, and that two adjacent subtrees have significantly different average gray levels. Our results show that the algorithm produced consistently good segmentations and that it is insensitive to noise.

The feature extraction proceeds as follows: the output produced by the Canny operator is subjected to morphological filters (erosions, dilations), and then a MST representation is computed by linking edgels. Similarly, region boundaries are computed from the output of the segmentation algorithm, and a MST representation is generated. A one pass check, essentially an "and" operation, is performed to reconciliate edgel information in the MST derived from the Canny operator output and from the segmentation algorithm. The resulting MST is then partitioned into subtrees of specified minimal size. Image areas of interest are determined by circumscribing polygons to the subtrees.

Within the areas of interest the following features are computed from the Canny-based edge map and from the region-based map: (1) length of longest line; (2) number of lines longer

than N pixels; (3) sum of squared lengths of lines longer than N pixels; (4) sum of squared lengths of parallel lines longer than N pixels; (5) sum of squared lengths of near orthogonal line pairs longer than N pixels; (6) - (10) corresponding features from region-based edge map; (11) longest symmetry line; (12) number of symmetry lines longer than N pixels; (13) sum of squared lengths of intersecting symmetry lines longer than N pixels. The results reported here were achieved with N = 30. N can also be determined by using size context information and the ground sampled distance (GSD) associated with the sensing event.

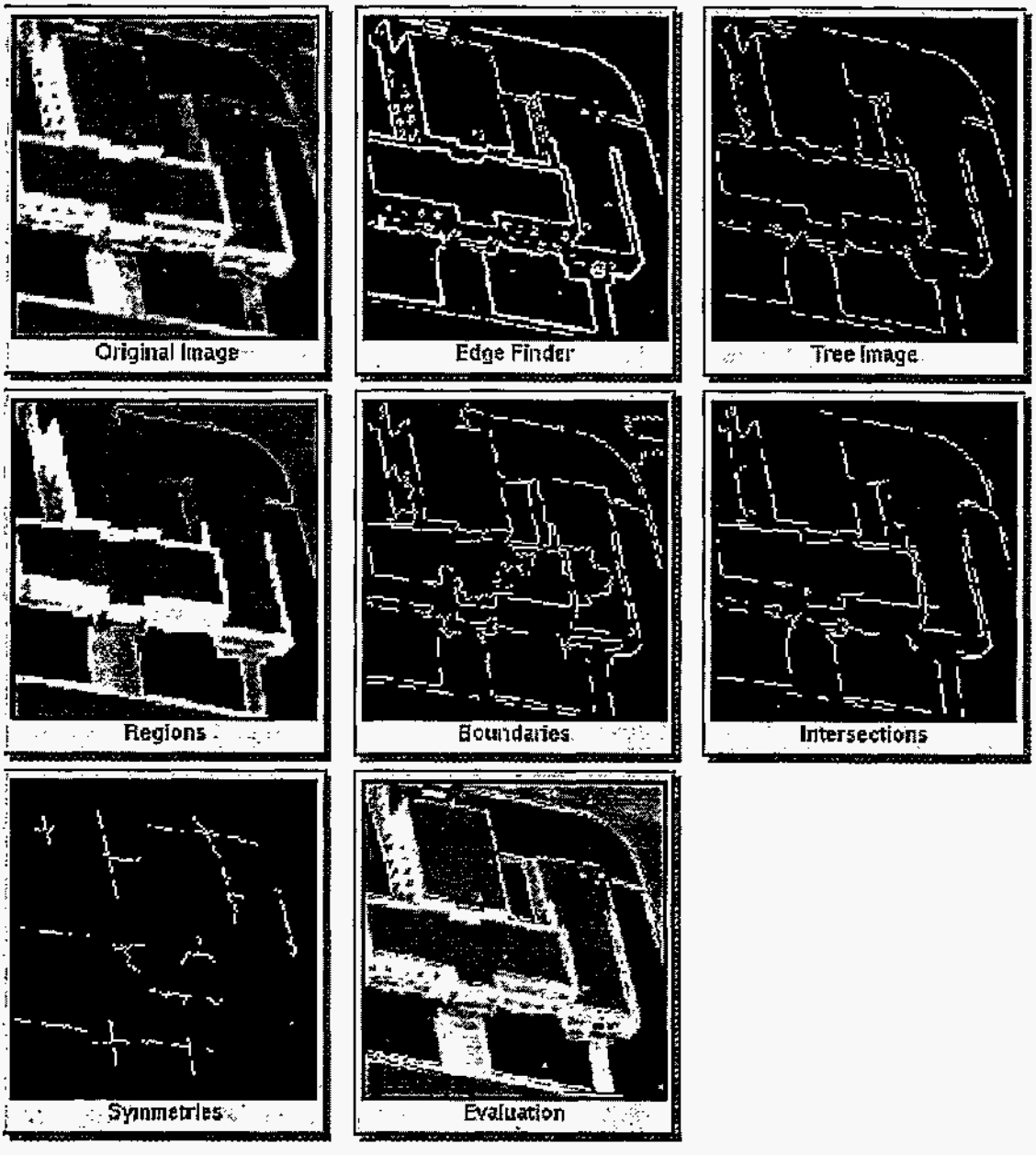


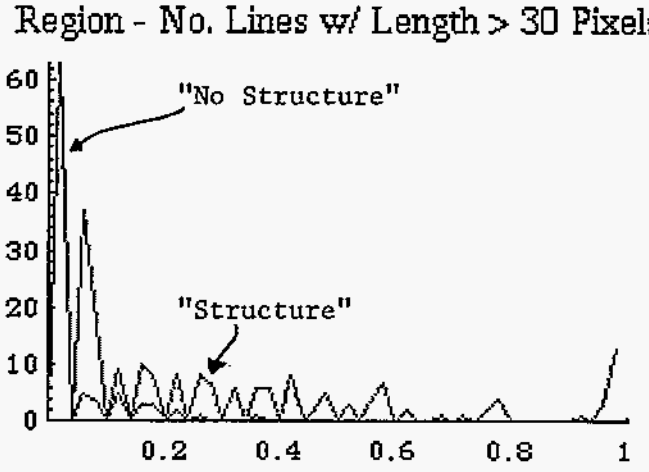
Figure 2: Illustration of visGRAIL analysis steps: from top left to bottom right the panels show the original image snippet, the result of the Canny operator, the edge tree representation, the result of segmentation, region-based boundaries, the integrated edge map, symmetry lines, and the outline of the region scored by visGRAIL (in this example as "structure"). The classified regions are color-coded (not visible in this reproduction).

Figure 2 shows a typical example of imagery that we have been able to analyze with this system and the results from different stages of the analysis. There is a wider spectrum of features that can be extracted from the region-based analysis, e.g., shape parameters, size parameters, etc. As part of our on-going work we are now extracting more information from the regions in order to support different exploitation tasks.

In order to train the system, we developed a set of tools that allow one to divide a large site image into 256x256 snippets (with no or variable overlap that can be adjusted by the user). The snippets are presented to the user for classification. This process results in a batch file that contains the snippet file identification as well as its classification. For each snippet the 13-dimensional feature vector discussed earlier is computed. We use a commercially available neural net package (Professional II/Plus by NeuralWare, Pittsburgh, PA) to train a backpropagation network. The resulting net is then used to regions of interest in new imagery.

4. RESULTS

By using standard principal component analysis we evaluated the linear separability properties of the feature set with respect to a training and test set of 700 image snippets extracted from aerial imagery of the Lockheed Martin, Denver, site. Figure 3 shows some representative examples of individual feature distributions for the two classes ("structure", "no structure").



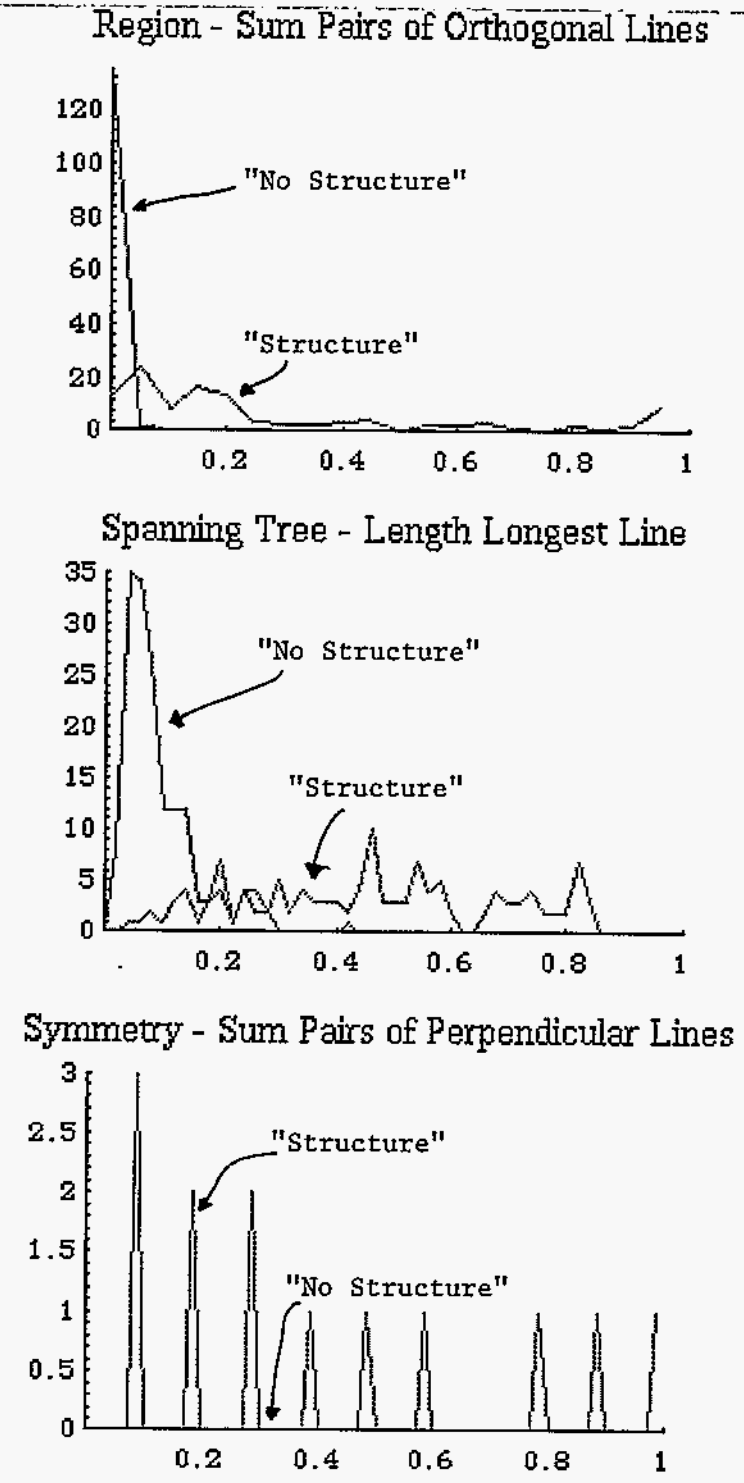
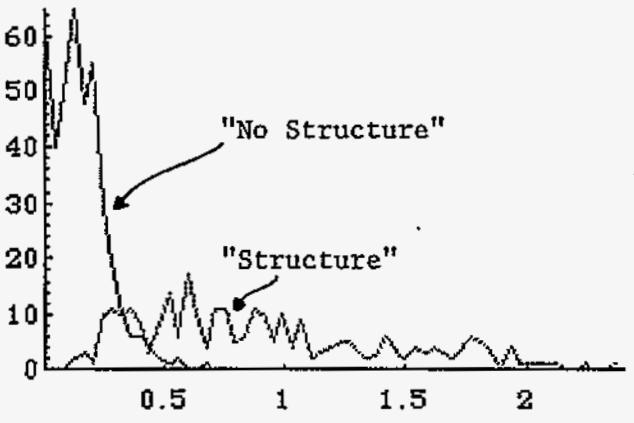


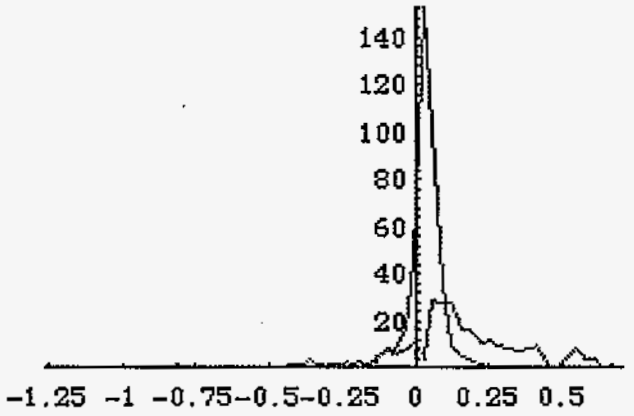
Figure 3: Distributions of selected features for the "structure" and "no structure" classes. Normalized dimensionless feature values on the abscissa, frequency on the ordinate.

Figure 4 shows representative feature distributions after projection of the features values onto the axes determined by the eigenvectors of the covariance matrix (principal component analysis).

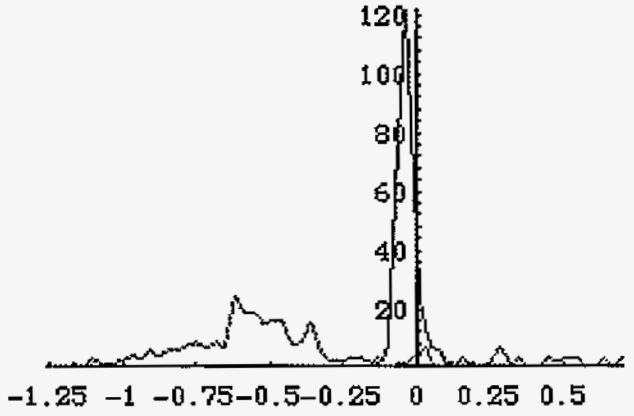
EigenVector 1



EigenVector 2



EigenVector 3



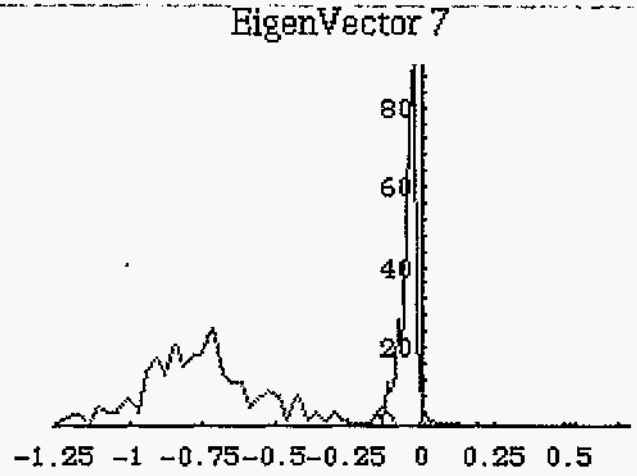


Figure 4: Distribution of features after projection onto representative eigenvectors of the covariance matrix (principal component analysis). The projection on eigenvector 7 shows the well-known phenomenon that components corresponding to low magnitude eigenvalues can contribute significantly to the discrimination. Normalized dimensionless feature values on the abscissa, frequency on the ordinate.

The optimal linear classifier results in an overall correct classification rate of 80%. We determined the optimal size of the neural network classifier by standard cross-validation [20]. Figure 5 shows the results of this process.

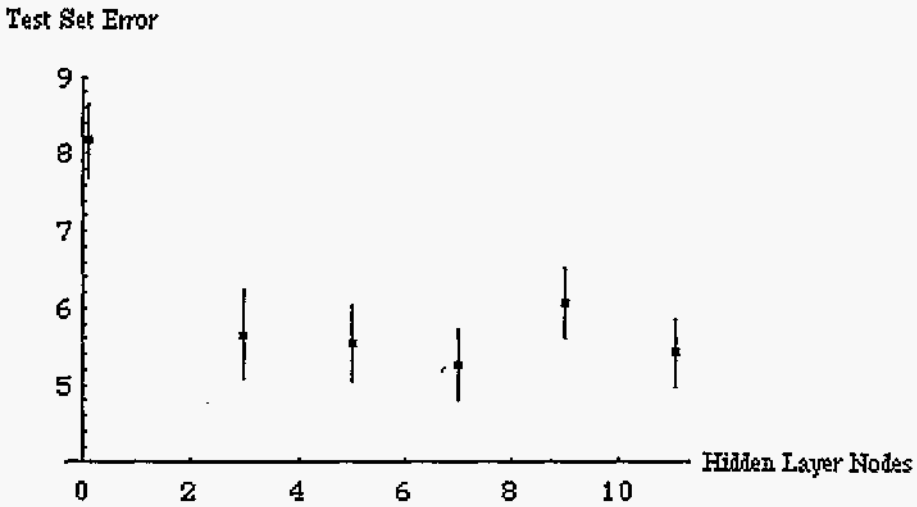


Figure 5: Cross-validation results - test set error as a function of the number of nodes in hidden layers of the backpropagation network. A set of 700 image snippets was divided in 10 sets of 70 images each, 9 of which were used for training. The 10th set was used for testing. Test set error is measured in non-dimensional units representing the sum of differences between the desired (1 for structure, 0 for no-structure) and actual values when scoring the samples in the test set. The error bars indicate one standard deviation (over the 10 independent tests).

With the network determined by this process, i.e., seven nodes in the hidden layers, we obtained the confusion matrix shown in Table 1.

	structure	no structure
structure	99.5%	0.5 %
no structure	4.0 %	96.0 %

Table 1: Confusion matrix obtained with backpropagation network using 13 input nodes, seven hidden layer nodes, and one output node.

Software modules that implement this methodology were developed and integrated into a system called visGRAIL. A first version of visGRAIL was demonstrated successfully in June 1995 at the National Exploitation Laboratory (NEL). This version was not integrated into the RCDE. As of September 22, 1995 we have a new version of visGRAIL that is integrated with the RCDE.

5. DISCUSSION

Computer vision is fundamentally a process that integrates a number of different clues and pieces of information, some generated from the data stream itself, some from other sources (including context of various kind), in order to solve a specific task or set of tasks. The experience base generated by decades of research and applications of computer vision systems shows that (1) robust vision systems can only be developed for tasks that are sufficiently constrained so that the vision algorithms do not operate outside of the parameter window designed for the task (often, this excludes even moderately complex tasks involving any kind of unstructured or natural environment); (2) approaches that rely only on "bottom-up" analyses cannot solve tasks of practical complexity; (3) the use of context greatly improves the performance of computer vision algorithms; (4) since for many applications it is impossible to obtain all a priori information needed by computer vision algorithms, the application of machine learning methods can broaden the domain in which computer vision algorithms can be applied reliably.

The visGRAIL project addresses two major research issues in computer vision: (1) the integration of contextual information into analysis algorithms, and (2) the integration of machine learning methods into analysis algorithms. Effective resolution of both issues is generally considered by the computer vision research community as paramount to developing and fielding reliable and reasonably flexible vision systems.

During the initial phase of this project, we have concentrated our efforts on developing the modules necessary for a proof-of-principle demonstration. The specific task area was defined in the course of interactions with potential users. The current version of visGRAIL represents a useful tool that can be adapted to new tasks by integrating new feature extractors and specific contextual information, and by training the neural network with imagery representative of the task to be accomplished.

Although the work during this reporting period has shown that the visGRAIL approach is viable, several issues remain to be resolved. For the current version of visGRAIL training with a representative set of images is performed once before the system is applied to imagery of interest. Future versions will include options for performance monitoring during operation and re-training. Shape-based features have not been exploited sufficiently with the current set of sensors. The current version of visGRAIL uses contextual information only in a rudimentary form, e.g., GSD, spatial location of area of interest. Much more context is available and can be captured by adding appropriate sensors to the system. For many important image exploitation tasks it is necessary to process and analyze additional spectral bands, e.g., SAR and IR. The visGRAIL approach is well-suited to

analyze imagery from these sensors, and to fuse information extracted from multiple spectral bands, including data from hyperspectral sensors.

As capabilities for digital imagery exploitation increase, it becomes increasingly important to be able to retrieve information from digital image databases in an efficient manner. Most desirable is the capability to perform content-based image retrieval. In its general form, this problem is, of course, equivalent to the general image understanding problem. The multi-sensor/neural net approach can be used to develop a set of tools that allow for content-based indexing and retrieval of imagery. Examples may include queries such as "find all images that contain more than k vehicles in a certain area", "find all images that show evidence for new construction in a selected area", etc. When applied to video imagery, these tools will allow for annotating video scenes that contain, for example, evidence for the presence of human faces, motions of certain characteristics, etc.

In conclusion, we have shown that the multi-sensor/neural net approach, previously applied successfully to data analysis problems in the Genome Program, leads to effective methods for exploitation of aerial imagery. It can integrate context information and machine learning into reliable and flexible computer vision systems.

6. ACKNOWLEGMENTS

This work was supported by the Office of Nonproliferation and National Security, U.S. Department of Energy under contract No. DE-AC05-840R21400 with Lockheed Martin Energy Systems, Inc.

7. REFERENCES

[1] *Proceedings of the 1994 ARPA Image Understanding Workshop*, Morgan Kaufman Publishers, San Francisco, November 1994, ISBN 1-55860-338-7

[2] *Proceedings of the First RADIUS Program Workshop*, October 1994.

[3] T. M. Strat, W. D. Climenson, "RADIUS: Site Model Content" , *Proceedings of the 1994 ARPA Image Understanding Workshop*, Nov. 1994, pp 277-285.

[4] T. M. Strat, M. A. Fischler, "The Role of Context in Computer Vision", *Proceedings of the Workshop on Context-Based Vision*, IEEE Computer Press, Los Alamitos, CA, June 1995, pp 2 - 12.

[5] *Proceedings of the Workshop on Context-Based Vision*, IEEE Computer Press, Los Alamitos, CA, June 1995

[6] P.Burlina, R. Chellappa, C. L. Lin, X. Zhang, "Context-Based Exploitation of Aerial Imagery", *Proceedings of the Workshop on Context-Based Vision*, IEEE Computer Press, Los Alamitos, CA, June 1995

[7] H. Buxton, S. Gong, "Advanced Visual Surveillance Using Bayesian Networks", *Proceedings of the Workshop on Context-Based Vision*, IEEE Computer Press, Los Alamitos, CA, June 1995

[8] A. F. Bobick, C. Pinhanez, "Using Approximate Models as a Source of Contextual Information for Vision Processing", *Proceedings of the Workshop on Context-Based Vision*, IEEE Computer Press, Los Alamitos, CA, June 1995

[9] R. C. Luo, M. G. Kay, "Data Fusion and Sensor Integration: State-of-the-art 1990s", in *Data Fusion in Robotics and Machine Intelligence*, M. A. Abidi, R. C. Gonzalez, eds, Academic Press, 1992

[10] B. V. Dasarthy, *Decision Fusion*, IEEE Computer Society Press, Los Alamitos, CA, 1994

[11] C. W. J. Granger, "Combining Forecasts - Twenty Years Later", *Journal of Forecasting*, 6, 167 - 173, (1989)

[12] N. S. V. Rao and E. M. Oblow, "Majority and Location-Based Fusers for System of PAC Learners," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 24, No. 4, pp. 713-727 (1994).

[13] N. S. V. Rao, E. M. Oblow, C. W. Glover, G. E. Liepins, "N-Learners Problem: Fusion of Concepts," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 24, No. 2, pp. 319-327 (1994).

[14] N. S. V. Rao, V. Protopopescu, R. C. Mann, E. M. Oblow, S. S. Iyengar, "Learning Algorithms for Feedforward Networks Based on Finite Samples," *IEEE Transactions on Neural Networks* (September 1995).

[15] N. S. V. Rao, "Fusion Methods for Multiple Sensor Systems with Unknown Error Densities," submitted to *Journal of Franklin Institute* (1995).

[16] N. S. V. Rao, "Nearest Neighbor Rules Approximate Feedforward Networks," submitted to *IEEE Transactions on Neural Networks* (1995).

[17] E. C. Uberbacher and R. J. Mural, "Locating Protein-Coding Regions in Human DNA Sequences by a Multiple Sensor-Neural Network Approach," *Proceedings of the National Academy of Science*, Vol. 88, pp. 11261-11265 (December 1991).

[18] R. J. Mural, J. R. Einstein, X. Guan, R. C. Mann and E. C. Uberbacher, "An Artificial Intelligence Approach to DNA Sequence Feature Recognition," *Trends in Biotechnology*, Vol. 10, pp. 66-69 (January/February 1992).

[19] Y. Xu, E. C. Uberbacher,"2D Image Segmentation Using Minimum Spanning Trees." submitted to *Image Vision and Computing*, 1995

[20] B. Efron, R. Tibshirani, "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuray", *Statistical Science*, Vol. 1 No. 1, 54-77, (1986)

DISCLAIMER