PII:S0031-3203(95)00152-2

ANALYSIS OF TERRAIN USING MULTISPECTRAL IMAGES

BIR BHANU,[†,*] PETER SYMOSEK[‡] and SUBHODEV DAS[†]

[†]College of Engineering, University of California Riverside, CA 92521-0425, U.S.A.
[‡]Honeywell Technology Center, 3660 Technology Drive Minneapolis, MN 55418, U.S.A.

Abstract—Automated terrain analysis is required for many practical applications, such as outdoor navigation, image exploitation, remote sensing, reconnaissance and surveillance. In this paper, we present a hierarchical approach to analyze multispectral (MS) imagery for autonomous land vehicle navigation. The approach integrates several strategies to label various terrain classes in these images acquired using twelve spectral bands in the visible and near-infrared spectrum. At the low (pixel) level, it combines texture gradient results from specifically selected channels by varying the size of gradient operators and performing multithresholding and relaxation-based edge linking operations to obtain robust closed region boundaries. At the high (symbolic) level, it makes use of the spectral, locational, and relational constraints among regions to achieve accurate terrain image interpretation. Details of the technique with examples from real imagery collected by an autonomous land vehicle (ALV) are presented. Copyright © 1997 Pattern Recognition Society. Published by Elsevier Science Ltd.

1. INTRODUCTION

The problem of terrain interpretation using image-based techniques requires the segmentation of the terrain imagery into regions corresponding to distinct land cover classes, e.g. field, road, forest. This interpretation task is of critical importance to many practical applications such as autonomous navigation in an unstructured natural environment. As an example, if an autonomous land vehicle (ALV) is required to survive and carry out its mission of surveillance, search and rescue and munitions deployment, it should be able to perform the functions of path planning, landmark recognition, and obstacle detection/avoidance. In order to achieve these functionalities, the ALV must first interpret the imagery obtained from its sensors into various regions in terms of their land cover, trafficability, and slope. The challenges presented by this task are significantly greater than those encountered in navigation on paved highways[1] because of the great variability of the terrain.

In the past, sensors for autonomous navigation and other remote sensing operations have primarily been TV cameras that provided data from the visible portion of electromagnetic spectrum. More recently, laser radars (LADARs) are finding applications in ALV navigation. At the same time, multispectral scanners (MSSs) and synthetic aperture radar (SAR) sensors are being increasingly used in many remote sensing operations. The former provide data in the visible and infrared portions of the spectrum in a digital form and can cover

a wide spectral region. However, the use of MSSs for ALV navigation has been limited so far due to their size and weight. These sensors are particularly suitable for terrain interpretation as they can capture the spectral characteristics of the different land cover regions in the different spectral bands and can provide digital data for image analysis. As the field of perception-based autonomous navigation advances, these sensors will become important for image understanding in natural environments.

The interpretation of sensory data, including multispectral (MS) data, in remote sensing applications is often affected by the problem of insufficient spatial resolution of the sensor. The finite spatial resolution causes the geographical area subtended by a pixel to contain a mixture of terrain categories. As a result of this, the spectral response observed at a pixel is a mixture of the spectral responses of several individual land categories.[2,3] This problem, known as the mixture-pixel problem, also persists in the ALV navigation scenario. In general, the spectral response at a pixel of an arbitrary region of a multispectral (MS) image will less likely to be due to a single category of terrain as the range to the region in the scene increases. Thus, spectral features calculated for these regions (located at the farthest distance of the scene) will not match those of data collected at close ranges where the mixture-pixel problem is negligible.

A robust terrain interpretation approach requires the integration of statistical techniques and adaptive feature detection, and the use of world knowledge. The use of constraints derived from context-dependent knowledge

---

* Author to whom correspondence should be addressed.

is helpful to overcome effects of sensor measurement variability caused by errors in sensor calibration, thermal noise, atmospheric effects and the mixture-pixel problem during region-based analysis of MS images. The knowledge-based techniques for region labeling can tolerate some errors in input data and still produce meaningful results. Because of their compact rule database, these techniques can be designed to have modular structures; as new contexts for classification of features are discovered, the system is reconfigured by the addition or modification of a few rules. The contextual information from a general world model that is applicable to all natural environments, can also be used to improve the system's performance for terrain interpretation in a specific application such as the ALV navigation.

In this paper, we present a technique for the labeling of regions in terrain images that are obtained with a MSS. What distinguishes our approach to region labeling from other scene interpretation algorithms[4-6] is the combination of texture feature detection operator results obtained from selected channels. The integration scheme involves adaptive texture gradient calculation, multithresholding, relaxation-based edge linking to obtain robust region boundaries, and the use of context-dependent constraints from the knowledge base. The algorithm described in this paper is hierarchical and consists of two stages. First, a multispectral image is segmented into *macro*-level regions with a texture gradient-based technique that is optimal with respect to the spectral characteristics of *all* the desired terrain categories: sky, forest, field, and road. The algorithm defines only closed region boundaries, and these regions are classified into one of the four categories (plus unknown) with a knowledge-based approach which uses relational, locational, and spectral constraints. Second, the macro-level regions are further segmented with the texture gradient-based approach where the parameters for texture feature extraction are optimal with respect to the *individual* macro-level region, e.g. field, forest, etc. The resulting subregions are classified using locational and spectral constraints. Finally, very small *micro*-level regions are merged with neighbors and boundaries are refined to identify all the terrain regions.

The paper is organized as follows. In the next section, we discuss the previous work related to region labeling in MS imagery. Section 3 describes the principle of the design of the texture gradient-based region segmentation algorithm and the approach to knowledge-based classification of regions. Experimental results based on imagery obtained with an ALV-mounted MSS are presented in Section 4. Finally, Section 5 presents the conclusions of the paper.

2. PREVIOUS WORK

The various approaches to the problem of region labeling in segmented images have ranged from pure statistical techniques to pure knowledge-based methods. In this section, we shall be reviewing mainly the latter category with some references to the former in the context of multispectral image segmentation. Knowledge-based techniques have been found to be useful in a number of application areas such as photointerpretation,[7-9] autonomous weapon delivery systems,[10] and the labeling of arbitrary urban scenes.[4-6] These knowledge-based systems have several features in common: a database of computed image features is matched with antecedents of production rules (rules are represented as logical statements of the form "if..., then...."), and a control system that supervises rule activation.

The system developed by Nagao and Matsuyama[8] uses a knowledge base representing contextual and geometric constraints for the task of labeling regions in multispectral imagery obtained by a low-flying helicopter. This method is intended for reliable classification of vegetational regions that is independent of the time of year and uses the ratio of two distinct spectral bands to discriminate the vegetational regions from the non-vegetational regions. Within the non-vegetational regions, such as congested urban areas, this method allows reliable identification of houses and roads. The region boundaries are detected by multiple low-level image segmentation algorithms and the resultant information is archived on a blackboard shared by each of the experts of the system. Each expert is optimized for locating a specific kind of object or region. The approach is hierarchical in which the final segmentation/classification of regions is dependent on the preliminary segmentation of the regions. A region growing approach is used to identify large regions of the image with global region properties such as homogeneity of multispectral intensity levels or elementary region-average texture measures, but no measure of local texture information is used at this stage, which is critical for the detection of regions in images of arbitrary outdoor natural scenes.

Ohtas[5] developed a hierarchical region labeling scheme for color images of urban scenes. The approach derives an initial "plan" image which is then labeled before a more detailed, data-directed segmentation is carried out. The "plan" image is defined by a region-based color image segmentation algorithm. The regions of the "plan" image are sky, tree, building and road. These region categories are detected using contextual and spectral constraints in a top-down manner. The algorithm is reliable and can correctly label regions in urban outdoor scenes using only 57 rules. The algorithm employs a texture measure for region classification, but the measure is expressed as a three valued logical variable: a region is either not-textured, textured or heavily-textured. This approach is useful for the discrimination of tree regions from sky, buildings and roads, which are essentially untextured, but it is not robust enough to discriminate the textures typically found in natural terrain images.

Hierarchical image interpretation strategies have also been implemented for the automatic classification of regions of multispectral Landsat imagery. Ton *et al.*[11] demonstrated a knowledge-based segmentation and

interpretation methodology for the interpretation of township-sized (5 × 5 mile) Landsat images. The image interpretation system establishes a coarse segmentation of the imagery using spectral features alone during a stage denoted by *category-oriented segmentation*, with simple, application-independent spatial constraints. The identified terrain regions are transferred to a second stage, denoted by *image-oriented segmentation*, where a hierarchical implementation of various image processing operations establishes an accurate segmentation of the image. Domain knowledge, and spectral and spatial constraints are utilized at this stage. This approach is appropriate for satellite imagery because the coarse resolution of the imagery averages out the effects of fine image texture and tends to blur out microscale region signature variations. Thus, spectral features alone can be used for coarse region classification. For a sensor mounted on a ground vehicle, the effects of perspective and shadow and the greater spatial resolution of the sensor diminish the utility and reliability of this strategy. Instead, the application of spatial and locational restraints is required to obtain useful region category maps.

Previous work on the problem of scene labeling for remote sensing applications[12] has predominantly used classic statistical techniques with some applications of knowledge-based techniques. Landgrebe[3] and Swain[13] surveyed the state-of-the-art for this application and identified four generic approaches: (a) spectral methods, (b) spectral/temporal methods, (c) spectral/spatial methods, and (d) spectral/general scene context methods. Algorithms from category (d) do employ auxiliary information, such as digital map data, water depth maps and ownership boundaries, but very limited work has been done to determine the value of knowledge-based techniques for the classification of regions in remotely sensed imagery. Region labeling with *a priori* spatial constraints is accomplished with relaxation labeling for consistency,[14] and measures of image texture. Image texture may be measured locally, such as using the co-occurrence matrix[2,15] or globally, such as using production system grammars for image structure.[16] The major obstacle to defining a knowledge base for the interpretation of remotely sensed imagery is that it is exceedingly difficult to identify suitable structural rules that are universally valid.[13]

Research in the area of automatic labeling of agricultural and urban areas in Landsat and SPOT imagery during the last few years has been concerned with attaining improved region classification accuracy using spatial and statistical attributes. Kusaka and Kawata[17] evaluated a hierarchical, spatial feature-based approach for the detection of linear features (roads and rivers) in Landsat imagery. The approach was implemented in four stages: (a) a gradient operator was utilized to establish high-confidence edge locations for each of the images of a multispectral image; (b) a distance transform for the detected edges (for all images) was derived to establish locations at which to obtain estimates for each individual region's spectral features; (c) a region growing approach was utilized to establish region boundaries; and (d) the linear inter-region boundaries are then authenticated by cross-referencing them with a map. Meyer[18] used a series of edge-preserving smoothing preprocessing steps and a statistical classifier to distinguish and classify agricultural regions in Landsat imagery. This approach derived region classifications with 100% accuracy for a test database containing 142 agricultural fields for a classification set of 12 major classes and 12 subclasses. The approach was predicated on the use of spectral and locational constraints that were specific to the problem being addressed. Another approach[19] obtained regions with uniform color in Landsat imagery by merging appropriate pixels with spectral values within a small predefined range. Zhang et al.[20] empirically evaluated the reliability of a stochastic relaxation algorithm for region classification in Landsat imagery where a Markov–Gibbs distribution was used to establish constraints on region adjacency. The stochastic relaxation algorithm was utilized to establish a minima of a cost function representing the reliability of region labels and restraints on the neighbors of each region. These approaches perform well for the applications for which they were designed, but for an autonomous vehicle navigation application, they would require significant modifications to deal with greater variability of the imagery. This could also prohibit implementation in real-time hardware.

Goldberg et al.[21] designed an expert system for the detection of changes in the forests of Newfoundland from Landsat imagery. The system is hierarchical, with multiple tiers of experts applying domain knowledge, in a top-down fashion, to the problems associated with foresting, forest fire damage, and disease infestation. The experts exchange information for control and hypothesis generation in a blackboard framework. Knowledge-based techniques are used for the tasks of change detection and the interpretation of changes in forestry terms, but no knowledge-based system is used for region classification.

Other researchers have evaluated Markov random field (MRF) based approaches for image texture for the task of multispectral image segmentation.[22,23] Since MRF models do not apply very well at region boundaries, approaches of this kind have problems at these locations. The computational complexity of a MRF approach can also be significant. Evaluations of the effectiveness of decision-directed learning of statistical region attributes and spatial interaction parameters have also been carried out for simple scenarios. Mardia and Hainsworth[24] quantitatively evaluated various approaches of this kind and demonstrated that, for low signal-to-noise ratios and artificially generated, single-channel imagery, these approaches can attain low misclassification rates. For three category region labeling problems, this approach produces very noisy estimates for region boundaries and requires several iterations of noise cleaning and re-estimation to achieve satisfactory estimates of region boundary locations. Bouman and Shapiro[25] describe a Bayesian

segmentation approach for multispectral images called multiscale random field (MSRF). The non-iterative approach uses a sequential maximum *a posteriori* estimation to derive the parameters of the multiscale model.

Machine learning-based methods have also been applied to the multispectral image segmentation problem. Fernandes and Jemigan[26] discuss a hierarchical, unsupervised learning approach for segmenting high resolution multispectral forestry images. Using a multiresolution texture representation, a hierarchical smoothing network aggregates pixels at five different resolutions. Texture and spectral similarity measures between and within network levels are used to inhibit smoothing between land cover classes. Segmentation performance is evaluated in terms of classification accuracy using independent and dependent samples for labeling emergent classes. The unsupervised classification method of Amadasum and King[27] performs agglomerative-type clustering of feature values computed over uniform neighborhoods. The learned mean feature vectors of the clusters are used in the segmentation of a three-band multispectral image of a terrain. This approach assumes that at least one uniform neighborhood can be found for each of the different categories present in the image. Another assumption is that feature vectors of neighborhoods representative of a particular category are similar to each other, but different from those of neighborhoods belonging to different categories. Another algorithm[28] which is intermediate between traditional supervised and unsupervised methods selects uniform and representative training samples for spatial segmentation of multispectral images.

Some researchers have addressed the problem of segmentation as an intermediate step in the analysis of multispectral images. Salvaggio and Schott[29] achieved segmentation in the course of extracting pseudoinvariant features. Their approach uses the rate-of-change information due to thresholding during the feature normalization process. Brill[30] used the bands of a multispectral image to segment three-dimensional objects, thereby overcoming the problems caused by image highlights and shadows during such segmentation. These methods are not directly applicable to terrain image analysis.

In summary, we see that the current image interpretation techniques are not general for automated terrain analysis which is required for many practical applications such as outdoor navigation, image exploitation, remote sensing, reconnaissance and surveillance. There exists a need for hierarchical analysis of terrain regions that will be capable of producing good segmentation results both globally and locally, particularly for applications in which the scene resolution changes continuously such as in outdoor navigation.

3. CONCEPTUAL APPROACH

An autonomous navigation system that is responsible for establishing the best course across an arbitrary geographical area requires evidence from many sources. Terrain regions are rarely characterized by distinct structural features such as line segments that are straight or at right angles. Even the cultural features such as buildings and bridges are difficult to be identified for day/night operations using a single source such as a TV camera because of the extreme variability of natural scenes. Therefore, as many auxiliary reference sources for terrain region discrimination and identification as are available must be used.

The signature observed by a MSS on-board an ALV for a specific terrain region will vary with the time of day, orientation of the surface relative to the imaging system, and shadowing effects. The subdivision of an image into regions where the terrain category is known versus those regions where the terrain category cannot be established defines useful information for a navigation system in that it may help the vehicle to avoid attempting to cross hazardous, non-traversable regions. Because of the degree of variability in the signatures that are measured by this sensor, it is not practical to configure a terrain region interpretation system that establishes the classification of terrain regions using pure statistical approaches as may be appropriate for some remote sensing applications. Instead, the estimated features for terrain categories obtained using MSS imagery should be utilized to identify corroborating evidence obtained from other sources such as landmark identification, active sensor (3-D) imagery, or on-board inertial reference unit. Region segmentation results also benefit the autonomous navigation system by coordinating the effective use of available computational resources in which the maximum amount of processing can be dedicated to those regions that represent the least hazard for traversal.

The focus of this paper is on segmentation and labeling of terrain images that are acquired by a MSS. The terrain interpretation algorithm is designed to label regions seen in a single frame of the MS imagery to one of several land cover categories, possibly including information regarding the slope, trafficability, and geological characteristics of the region. This information is helpful to the mobile agent navigating the terrain and exploiting the MSS data. Examples of region labels, corresponding to land covers typically encountered in a cross-country navigation scenario, are field with grass and snowberry/chokeberry vegetation, and forest with hemlock vegetation. In this paper, large regions like *field* will be referred to as the "macro-level" regions, while small regions like *snowberry* will be called the "micro-level" regions. It is important to note that an acceptable performance of the interpretation algorithm requires that the candidate set of labels for a region be small should the selection of a unique label not be possible. This requirement is posed by the navigation task for which the candidate set for the land cover should be specific enough for the mobile agent either to match the region to a digital terrain map database or to plan a course across the terrain while avoiding obstacles.

3.1. Overview of the system for terrain interpretation

Our system for terrain interpretation works by combining results from different multispectral channels in a cooperative and complimentary fashion. In contrast, most approaches to segmentation of MS imagery handle the data from individual channels in an independent manner. Those that do combine results from different channels rely on statistical properties of the image regions which are not very effective in discriminating among land covers encountered in navigation scenarios. For example, snowberry and scrub bushes have very similar image textures and are, therefore, nearly indistinguishable using Markov random field based methods. One of the key differences between segmenting images for most remote sensing applications and that for autonomous navigation is that in the latter case, the segmentation algorithm has to cope with varying resolution for different parts of the same image frame.

Due to the nature of the imaging process, a mobile platform-mounted MSS captures data that are affected to a greater extent by the mixture-pixel problem (see Section 1) due to perspective than data collected using an airborne MSS. (The latter type of sensor is typically used in most remote sensing applications.) This fundamental difference in the input data calls for a different approach to segmentation of MS imagery in the former case which is the focus of this paper. Accordingly, the terrain interpretation system described below is designed to be hierarchical in nature and knowledge-base oriented. The hierarchical structure allows the system to be more tolerant of the mixture-pixel problem; the use of region-average features at each stage of the hierarchy reduces the number of errors in region classification since averaging tends to reduce variability in features. This also reduces the computational complexity at each stage of the hierarchy compared to that in a non-hierarchical classifier as classifications at each stage are based on a reduced set of terrain categories identified by the previous stages. The knowledge database rules lend robustness to the interpretation system against the mixture-pixel problem by capturing contextual information which is difficult to express in terms of statistical rules.

The interpretation of MS imagery occurs in two distinct stages at each level of the hierarchy: segmentation and classification. Segmentation involves locating discontinuities of texture gradients that correspond to region boundaries. The classification step utilizes the knowledge base to assign a distinct label to each of the segmented regions. The regions identified at the lower level of the hierarchy are further subdivided at the higher level. These steps of the algorithm are now described in detail.

The algorithm for terrain interpretation is subdivided into five key steps:

1. *Macro-level Region Segmentation*—The segmentation of the multispectral images into large regions using texture gradient techniques.

2. *Macro-Level Knowledge-Based Classification*—The classification of the macro-level regions with a knowledge-based scheme which utilizes relational, locational, and spectral constraints.
3. *Micro-Level Region Segmentation*—The segmentation of the labeled macro-level regions with texture gradient techniques which are specifically designed for those regions.
4. *Micro-Level Knowledge-Based Classification*—The labeling of micro-level regions using contextual and spectral knowledge.
5. *Refinement of Region Boundaries*—The merging of micro-level regions with large neighboring regions and adjustment of boundaries.

Figure 1 shows the block diagram of the hierarchical symbolic approach for terrain interpretation. Optimal channels for the detection of region boundaries are selected both at the macro-level and at the micro-level.

3.1.1. Selection of optimal multispectral scanner channels.
The selection of a MSS as the sensor for remote sensing applications is motivated by the fact that such sensors are responsive to the spectral characteristics of the individual land cover categories. Thus, for optimal performance of a system deploying such a sensor, the different spectral channels of the MSS to be used for feature detection are required to be selected in a fashion which allows for maximum discrimination among the land cover classes. Consequently, the selection process is application dependent. The varying gross-scale emissivity and reflectivity of the "macro-level" regions specify the particular spectral attributes to be used for "macro-level" region boundary detection. For a given spectral waveband, the apparent contrast between "macro-level" regions varies with the specific times of day as it is dependent on the natural coloration and emissivities of the terrain categories that make up these regions. The predominant portion of a multispectral signature of an object in an outdoor scene at wavelengths shorter than 3 μm is due to the reflection of diffuse and direct solar irradiation. The measured signature of these objects in the visible and near-wave infrared (NWIR) wavebands is a function of their chemical composition and surface orientation relative to the imaging system. Absolute reflectivities of terrain categories can be established by measuring their reflected radiation simultaneously with the measurement of the reflected radiation from a reference standard and then normalizing the sample's radiant energy by that of the reference standard. Consequently, the identification of a specific waveband (i.e. an MS channel) for a selected pair of terrain categories, in which the reflectance of one is significantly greater than that of the other, should allow the design of a robust feature detector for the discrimination of boundaries between the categories. In the following, we discuss the criteria employed by our terrain interpretation system for the selection of
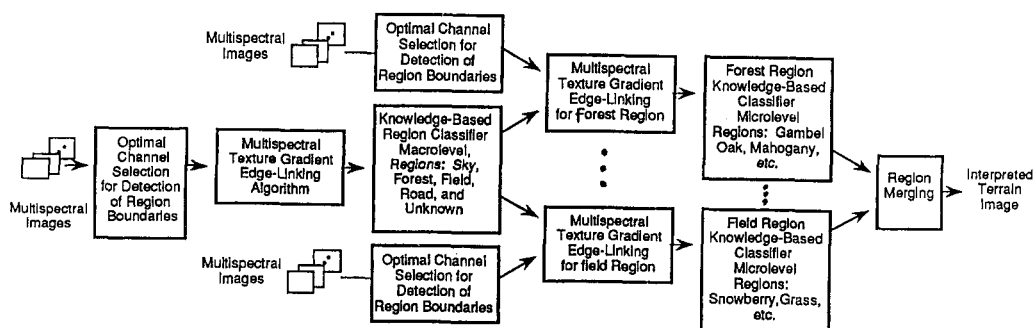
Fig. 1. Hierarchical symbolic grouping approach for the segmentation of multi-spectral images.
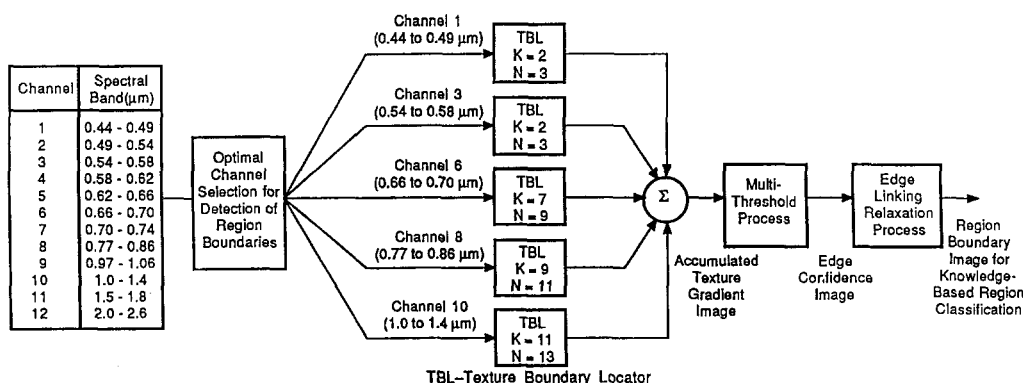

Fig. 2. The texture gradient edge linking relaxation algorithm. TBL is texture boundary locator (TBL) algorithm. K and N are window parameters used in the TBL Algorithm.

optimal MSS channels to best locate the region boundaries.

A discussion of the optimal channel selection process is facilitated by a brief description of the MSS used in this research. The MS data is collected in twelve discrete bands[31] with a passive scanner. The passive multispectral scanner was developed by the Environmental Research Institute of Michigan (ERIM), the spectral bands for the scanner are given in Fig. 2. Also available for the ALV test site are spectral reflectance measurements obtained from the U.S. Army Engineering Topographic Laboratories (ETL).[32] They include spectroradiometer measurements of the spectral behavior of several terrain categories in the form of graphs of radiance and reflectance, with respect to a Halon reference standard, as a function of wavelength. These data are used to extrapolate the average spectral characteristics of the regions of interest. In our approach, the estimated spectral characteristics are used to identify the MS channels exhibiting the greatest discontinuity between regions for all possible pairings of the four classes (sky, forest, field and road) at the macro-level. Consequently, the best discriminating set of channels for the four classes is defined (see Table 1).

The macro-level MS channels identified for this research are very close to the set of four channels identified by a computer modeling study for the design of the ERIM active multispectral 3-D Sensor.[33] The four channels are: (a) green reflectance peak (0.52 –

Table 1. Multi-spectral channels which are optimal for the detection of region boundaries between four macro-level classes

	Sky	Forest	Field	Road
Sky	X	1	1	1
Forest	1	X	3	6
Field	1	3	X	8,10
Road	1	6	8,10	X

The optimal channel(s) for the detection of a boundary between a class of the row dimension and a class of the column dimension is found as the table entry at the intersection of the row and column.

0.55 µm), (b) red reflectance peak (0.62–0.69 µm), (c) near infrared (0.8–1.2 µm) and (d) shortwave infrared (2.0–2.35 µm). These channels roughly correspond to the channels 2, 5, 8–10 and 12, respectively, of Fig. 2. Actually, the study[33] has established the best set of up to seven channels for terrain region classification by a minimum-distance, statistical classifier for thirteen terrain categories. It has also been reported in the study that when the minimum-distance classifier used features from all four channels, a 95% probability of correct recognition (POCR) was attained. The best discriminator set for a two-dimensional classifier was a visible spectrum channel and a near-infrared channel for all images studied (POCR > 85%). These quantitative results reinforce the validity of the theoretical

Table 2. A cross-reference table for the best multi-spectral images to obtain significant region boundaries between the four micro-level region categories of field

	Grass	Snowberry	Scrub bushes	Road
Grass		8,10	8	4,5,6,8
Snowberry	8,10		10	4,5,6,8,10
Scrub bushes	8	10		4,5,6,10
Road	4,5,6,8	4,5,6,8,10	4,5,6,10	

assumption that a single channel or a combination of channels establishes a high contrast feature for discrimination between pairs of terrain regions.

Similar to the selection of channels for macro-level, the micro-level channels identified for feature detection are optimized with respect to the measured spectral features of each of the micro-level regions within a macro-level region so as to obtain the best discrimination. For example, the optimal micro-level channels identified for the macro-level field category are indicated in Table 2.

3.2. Algorithm details

The various algorithms of the terrain interpretation process are indicated in Fig. 1. The algorithms at the macro-level stages are concerned with region boundary extraction through linking of multithresholded, multispectral texture-gradient edges and region classification using contextual (i.e. spatial, spectral, relational) constraints from the knowledge base. The MS channels of Table 1 selected for this level are optimal with respect to the macro-level terrain categories (field, forest, sky, road and unknown). The micro-level obtains region boundaries in a similar fashion, except that it uses a different set of optimal channels and algorithm parameters for each of the macro-level regions. The labeling of micro-level regions uses only the spectral constraints from the knowledge base. In the following, the algorithms that constitute each of the five steps of the terrain interpretation process will be described in detail. The justification for the approach established for each step is presented together with the salient features of the algorithms. Experimental results based on these steps appear in Section 4.

3.2.1. Macro-level region segmentation.

The objective of this step is to define a coarse image segmentation which is the basis of all succeeding region classifications. The use of spatially wide windows for feature detection at this stage guarantees that the diversity of the natural terrain imagery does not cause invalid interpretations to be transferred forward to the next level of the hierarchy. The use of spectral features, that are effective for detection of critical structural features, where the feature detection operators sample the image at a resolution appropriate for the current level of the hierarchy, diminishes the likelihood of classification errors for this stage and the following

ones. This coarsely segmented/labeled image is similar to the "plan" image of Ohta.[5] To obtain a coarse representation of the input MS image, the algorithm's parameters are tuned to detect only large features. Extraction of the resulting macro-level regions is fairly reliable and is little affected by sensor noise or small changes in the parameter values. The importance of this segmentation step, called the *Multispectral Texture-Gradient Edge Linking Relaxation Algorithm*, lies in data abstraction—transforming the raw image data into a form which is suitable for a high-level, knowledge-based, labeling algorithm. A block diagram of the texture-gradient edge linking algorithm is shown in Fig. 2. The key processes used in this algorithm are the computation of texture gradients and their integration, multi-thresholding and relaxation-based edge-linking. In the following, each of these processes is described in detail.

Process A. Computation of Texture Gradients and their Integration: texture gradients are obtained by a texture boundary locator (TBL) algorithm. For most regions in natural scenes, it has been observed that on either side of the boundary between two adjacent regions there occurs a gradual transition from the mean of the image intensities of one of the regions to that of its neighbor. Thus, the TBL algorithm calculates the texture gradient of an image, which is the local rate of change of a textural attribute of the image. The textural attribute is derived from the mean, μ and standard deviation of each $N \times N$ window of the image, where N is obtained as a function of the image features to be detected. The sole requirement imposed on the gradient image is that it should have a large magnitude wherever major discontinuities occur, such as between the region categories (sky, forest, field and road) at the macro level.

In order to compute a texture gradient image, a $(2K + 1) \times (2K + 1)$ window is centered at each pixel, where K is a function of the size of the region of interest and/or the range of the scene objects from the sensor. The window geometry for an arbitrary image plane location P is shown in Fig. 3. The pixels at the centers of the four sides and the corners of the $(2K + 1) \times (2K + 1)$ window are labeled sequentially
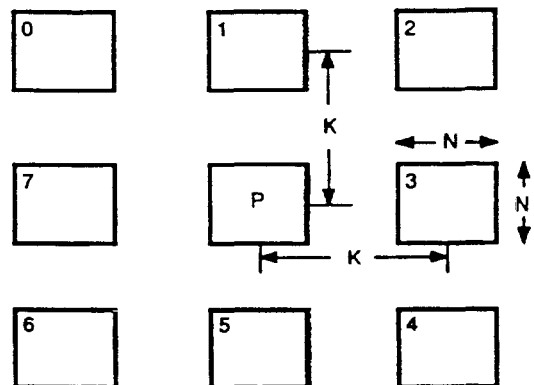

Fig. 3. Window geometry for texture boundary locator (TBL) algorithm.

beginning at the top left corner as shown in the figure. The texture gradient at a pixel is obtained as

$$\max_{0\leq i\leq 3}\left\{(\mu_i-\mu_{i+1})^2+(\sigma_i-\sigma_{i+4})^2\right\}^{1/2}, \quad (1)$$

where μ_i and σ_i are the mean and the standard deviation of the image's intensities for an $N \times N$ window (see Fig. 3).

The texture gradient is not calculated for every MS image, but only for those MS images which display sharp differences between the means of regions representing the four macro-level classes: sky, forest, field, and road. As discussed earlier and shown in Fig. 2, MS channels 1, 3, 6, 8, and 10 provide the best images for texture gradient computation out of the twelve channels that are available and are, therefore, used.

The parameters of the TBL algorithm, K and N, are set for each selected MS image according to the nominal range of the terrain class boundaries to be detected. For instance, channels 8 and 10 are identified as the best discriminators between the road and field classes. Because fields are usually located within 5–100 yards of the ALV (as a function of the ALV's current mission profile), the road-field boundary extends across several pixel samples laterally. Therefore, the parameters of the TBL algorithm used for these MS images are $K = 9, N = 11$ and $K = 11$, $N = 13$, respectively. The values of the parameters K and N used for estimating the gradient of the five images (at the macro level) for the experimental results presented in this paper are shown in Fig. 2.

The evidence for region boundaries obtained by processing each of the five images (corresponding to channels 1, 3, 6, 8, and 10) with the TBL algorithm are combined to produce a single gradient image called the accumulated texture gradient image. Because the locations of the image where each of the five gradient images attains its maximum value are essentially disjoint, the gradient images may be combined additively. Therefore, each gradient image provides maximum support for the existence of a true inter-class boundary for the region boundaries it is designed to detect and can contribute evidence for the remaining region boundaries.

Process B. Multi-thresholding: the algorithm uses multiple thresholds for the estimation of the true region boundaries. It endeavors to fill in the gaps of the highest threshold binary image by using the evidence from the edge points that are output for each of the remaining lower thresholds. In this way, the confidence that the edges of the output binary image are true region boundaries is high. This approach is more robust than traditional gradient-based techniques for the following reasons. For the traditional gradient-based approach, a gradient image is calculated with a differential window operator. The resultant gradient image is transformed into a binary image with an appropriately chosen threshold. If the threshold value is chosen too low, spurious region boundaries will appear alongside true ones in the resultant region boundary image. On the

other hand, a high threshold value will cause pixels to be missing from the region boundaries.

For an edge confidence image, pixel values range from A_1 to A_N, where A_N signifies the highest confidence that a pixel is an edge element. The confidence image locations with the intensity value A_N are those locations of the gradient image whose intensities are equal to or greater than the strictest, highest-value of the N thresholds. The thresholds are not absolute thresholds for image intensities, but are upper percentage thresholds for the cumulative distribution function of the accumulated texture gradient image's histogram. The cumulative distribution function of an image's histogram is evaluated by deriving the image's histogram, normalizing the histogram by the total number of pixels of the image and then integrating the normalized histogram bin counts from the lowest intensity value of the image. The accumulated texture gradient image is mapped to the edge confidence image according to the following steps: (a) absolute thresholds $T_1,\ldots,T_N$ (which correspond to the $P_1,\ldots,P_N$ upper percentage points of the cumulative histogram) are calculated; (b) the accumulated texture gradient image is mapped to N binary images by thresholding with each of the thresholds; (c) the N binary images are combined additively to obtain the edge confidence image. Thus, accumulated texture gradient image magnitudes greater than the highest threshold are mapped to the greatest confidence of the edge confidence image (N), and those magnitudes greater than the next highest level threshold but less than the highest level threshold are mapped to the next highest confidence ($N-1$), and so on. Three thresholds at the upper 15, 25 and 35% of the accumulated texture gradient image are found to be sufficient to define the true region boundaries. The use of a greater number of thresholds did not improve the quality of the detected boundaries significantly.

Process C. Relaxation-based Edge-linking: an edge confidence image is utilized for joining together the incomplete region boundaries.[34] The three steps of this edge-linking relaxation process are

1. Label the maximum intensity pixels of the confidence image as edge elements.
2. Identify incomplete region boundaries. If no incomplete boundaries are found, stop; otherwise go to Step 3.
3. Link the incomplete boundaries found for Step 2. based on local edge evidence and the smoothness of the boundary. Go to Step 2.

The criteria employed for selecting new edge locations for Step 3, in order of precedence, are the following:

(a) the edge location of maximum edge evidence of the set of neighbors of the endpoint (if such a maximum exists), and
(b) if all the neighbors of the end point have the same edge evidence, then select the pixel location among the neighbors of the endpoint that causes the

smallest change in the *curvature* of the incomplete boundary.

Because new edge pixels are appended to incomplete boundaries until they become complete boundaries, the region boundaries defined by this algorithm are closed. After edge-thinning, region properties/features are calculated for each region, and these properties are archived in a database for use by the knowledge-based region labeling algorithm.

3.2.2. Knowledge-based classification (macro-level).

A knowledge-based region labeling algorithm is used to establish region labels for the regions identified in the preceding stage. This algorithm is implemented as a production system which cross-references features established in the current image with an archive of representative features of outdoor scenes via pattern matching. The rules of the knowledge-based region labeling algorithm are derived using *a priori* knowledge of the spectral, locational, and relational characteristics of the regions of interest for the ALV's current location. The macro-level regions are individually classified into four categories (plus unknown) which are

$$\Theta = \{\text{sky, forest, field, road}\} \qquad (2)$$

The specific features used by the region labeling scheme are,

1. Spectral features—the mean and standard deviation of each region.
2. Locational features—the location in image space of the region with respect to the expected location of the terrain classes, as a function of the imaging system's orientation.
3. Relational features—neighborhood adjacency constraints for specific pairs of terrain classes.

The likelihood or confidence of the classification of each macro-level region is calculated with a pseudo-Bayesian confidence measure. The classes {sky, forest, field, road} are labeled as c_1, c_2, c_3 and c_4, respectively. A fifth class, unknown, also exists at this stage, but no confidence measures are calculated for it. This last class correspond to regions which are not classifiable based on the current rules in the knowledge base.

The spectral features employed to classify the regions into one of the four classes are denoted by $\{f_{ij}\}$ $j = 1, \ldots, J_i$ for class c_i, where J_i is the number of spectral features and $i = 1, \ldots, 4$. If the *a priori* probability for each class of Θ, $P(c_i)$ $(i = 1, \ldots, 4)$ is known, then the conditional likelihood that a region is an instance of class c_i when the feature data $f_{i1}, f_{i2}, \ldots, f_{iJ_i}$, are observed is given by

$$p(c_i \mid f_{i1}, f_{i2}, \ldots, f_{iJ_i}) = \frac{P(c_i)p(f_{i1}, f_{i2}, \ldots, f_{iJ_i} \mid c_i)}{\sum_{k=1}^{4} P(c_k)p(f_{i1}, f_{i2}, \ldots, f_{iJ_i} \mid c_k)}, \qquad (3)$$

where $P(f_{i1}, f_{i2}, \ldots, f_{iJ_i} \mid c_i)$ is the probability density of the features f_{ij}, $j = 1, \ldots, J_i$, conditioned on the fact that they are observations of class c_i. If each of the features

of a specific class c_i $(i = 1, \ldots, 4)$ are independent, then equation (3) can be expressed in a computationally efficient form. Because of the immense variability of outdoor imagery, the features, $\{f_{ij}\}$, for the ith region are asymptotically independent as the size of the region becomes large. Thus, for macro-level regions, which are large, the features can be viewed as being independent. Under this assumption, equation (3) becomes:

$$p(c_i \mid f_{i1}, f_{i2}, \ldots, f_{iJ_i}) = \frac{P(c_i) \prod_{j=1}^{J_i} p(f_{ij} \mid c_i)}{\sum_{k=1}^{4} P(c_k) \prod_{j=1}^{J_i} p(f_{ij} \mid c_k)}. \qquad (4)$$

The initial probabilities of the classes, $P(c_i)$, $i = 1, \ldots, 4$, are all equal to 0.25 for our system.

The terms $p(f_{ij} \mid c_i)$, $j = 1, \ldots, J_i$, $i = 1, \ldots, 4$ in equation (4) are very difficult to estimate precisely for an arbitrary MS image. Instead, these density functions are approximated as piecewise uniform or linear likelihood functions for computational efficiency. One such approximation is the trapezoidal density function. Figures 4 and 5 show the densities estimated for the sky and forest, and field and road classes, respectively, using the selected MS channel image. Also presented, for comparison purposes, are the histograms of measured image intensities of these regions in the image MULTI36 (an element of the data set of MS imagery collected using ERIM passive scanner).

Currently, our approach uses the four relational constraints: ABOVE, BELOW, LEFT and RIGHT. These relational constraints are defined for all pairings of the four terrain categories, but have logical values only for those pairings where the relation makes sense. For instance, the relation {sky ABOVE grassy field} is possible (the sky is above all fields), but the relation {forest ABOVE sky} is invalid.

The knowledge-based classifier uses a pseudo-Bayesian likelihood approach for the spectral and locational constraints. Because conditional likelihoods for relational constraints are binary, i.e. $p(f \mid c_i) \approx 1$ and $p(\tilde{f} \mid c_i) \approx 0$, where $\tilde{f}$, represents the compliment set of feature f), the reasoning process carried out by terrain interpretation system at this step is essentially one of pattern matching. These knowledge-based pattern matching operations can be represented equivalently as pseudo-rules. Two of the equivalent pseudo-rules are given below.

Classify-forest pseudo-rule
    If the class *Forest* has the greatest likelihood and
    If the location of the region is in the upper half of the image and
    If the region ABOVE this region has a high likelihood of being *Sky*, then
    Classification = *Forest*

Classify-road pseudo-rule
    If the class *Road* has the greatest likelihood and
    If the location of the region is in the lower half of the image and
    If the location of the region is in the middle one-third of the image and
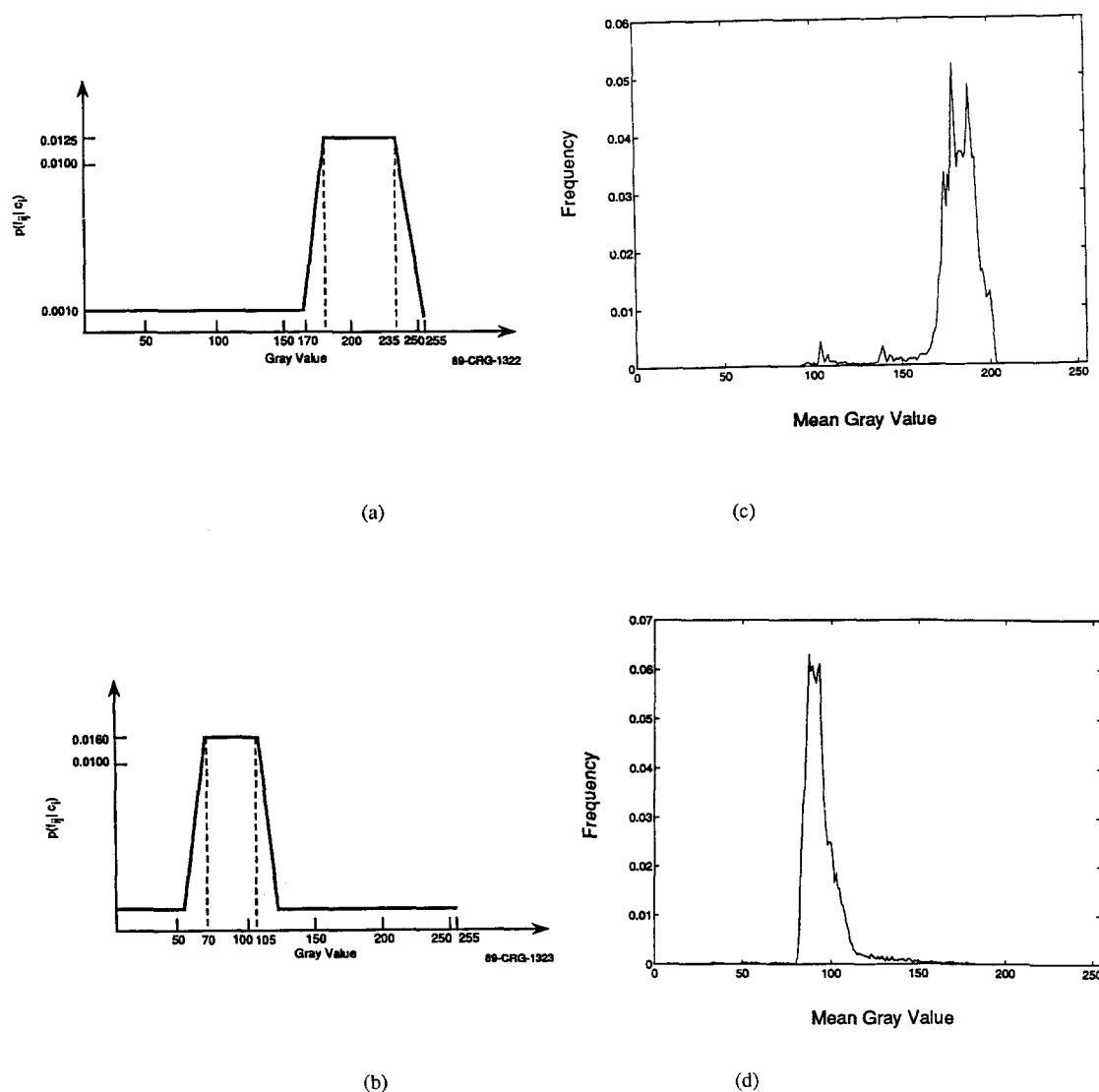
(a)

(c)

(b)

(d)

Fig. 4. Piecewise linear densities estimated for (a) sky and (b) forest regions in multi-spectral channel 1 of MULTI36. Measured image intensity histograms of 5×5 window *means* for (c) sky and (d) forest using the same data.

If the region RIGHT of the region has a high likelihood of being *Field* or if the region LEFT of the region has a high likelihood of being *Field*, then

Classification = *Road*

The performance of this step can be enhanced with the integration of auxiliary information such as digital map, digital terrain feature data (DFAD) and digital terrain elevation data (DTED) into the knowledge base, beyond the baseline spectral, locational, and relational constraints described here. The representation of this auxiliary knowledge will also be in the form of conditional probabilities, similar to that of the baseline constraints.

3.2.3. Micro-level region segmentation. Once the macro-level regions have been segmented and labeled, adjacent regions which belong to the same class are

merged. The updated labeled image consists of contiguous regions of forest, field, road, sky and unknown. The third step of the terrain interpretation algorithm is the segmentation of each macro-level image region into subregions or the micro-level regions. We have identified the spectral bands of the data which are optimal for differentiation of the micro-level regions within each macro-level region. A cross-reference table for the micro-level regions of the field category is provided in Table 2.

The general principles for the segmentation of micro-level regions are the same as that of macro-level regions. However, the notable differences between the two are

- A distinct group of MS images is used,
- The parameters that define the sensitivity of the TBL algorithm are revised, and
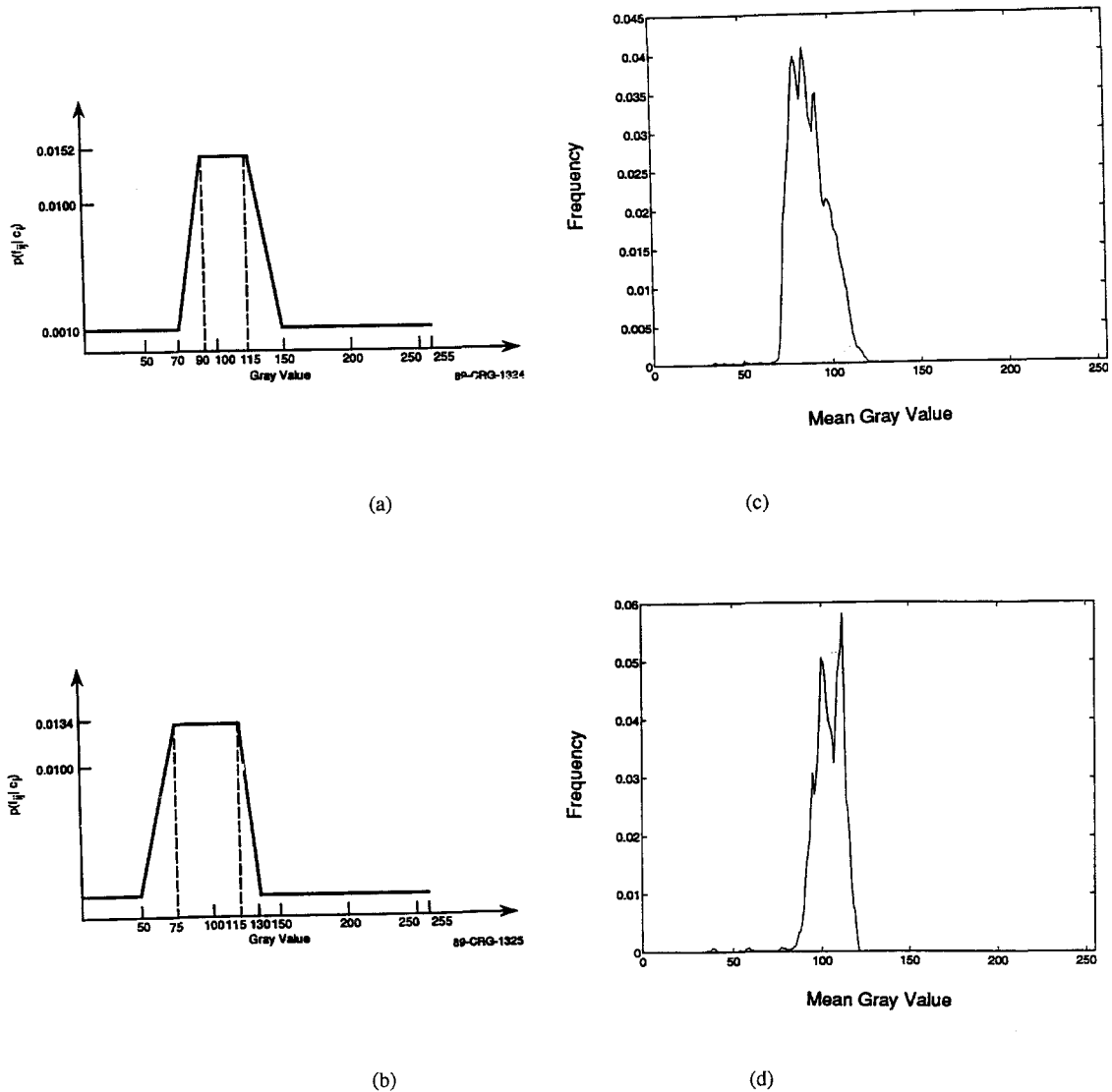
(a)

(c)

(b)

(d)

Fig. 5. Piecewise linear densities estimated for (a) field and (b) road regions in multi-spectral channel 1 of MULTI36. Measured image intensity histograms of 5×5 window *means* for (c) field and (d) forest using the same data.

- Region merging is done after the MS texture gradient edge linking relaxation algorithm has produced an edge image to combine small regions resulting from this algorithm. These small regions are obtained because the parameters of the TBL algorithm are tuned at this step to detect higher resolution features.

The set of MS images used for subregion boundary detection for each of the four terrain classes (excluding the unknown class) are those that exhibit sharp discontinuities between subregions of interest. The parameters of the TBL algorithm are selected to correspond to the range of macro-level image regions and the resolution required for the detection of the micro-level features. Range to a specific macro-level region is derived from knowledge of the height of the camera above ground level and the DTED or a flat earth assumption if no DTED is available. The parameters of the TBL algorithm are given in Section 4.

Region merging, as mentioned above, is performed after the spectral features for different regions have been obtained using the luminance image (the Y image of the NTSC standard for color imagery transmission). These features are utilized in a series of hypothesis tests for the estimation of equivalence of statistical distributions of image intensities for neighboring regions. The derivation of hypothesis test procedures where the mean and variance of the micro-level classes are unknown is very complicated and no generally accepted methodology exists. This category of statistical hypothesis test is known as the Behrens–Fisher Problem in the statistical literature[35] and although diverse strategies have been proposed for scenarios of this kind, controversy still

rages regarding the appropriateness and usefulness of these test procedures. The statistical hypothesis test procedure derived for establishing micro-level region equivalence is heuristic and is realized as a sequence of three hypothesis tests. These three hypothesis tests are: (1) an F-test for the equivalence of the variance of the samples' image intensities, (2) a t-test for the equivalence of the means of the samples' image intensities, and (3) a test of the equivalence of the distributions of the sample measurements, which is called the Kolmogorov–Smirnov Test. These hypothesis test procedures are described below.

A hypothesis test is a standard statistical estimation approach for discriminating between mutually exclusive hypotheses. The disjoint spaces represented by the hypotheses are Ω_0 and Ω_1 and the traditional notations for the hypotheses themselves are:

$$H_0 : \quad \theta \in \Omega_0, \\ H_1 : \quad \theta \in \Omega_1, \tag{5}$$

where random samples (X) are drawn from a distribution which involves the unknown parameter θ. The subset of the sample space S for which the hypothesis H_0 will be rejected is denoted as the *critical region*. For a specific hypothesis testing strategy, one is interested in the likelihood of rejecting H_0, which is denoted as $\pi(\theta)$ for all $\theta \in \Omega$, where $\Omega_0 \cup \Theta_1$. The function $\pi(\theta)$ is called the power function of the test. Thus, if C denotes the critical region of the test, then the power function, $\pi(\theta)$, is determined by the relation:

$$\pi(\theta) = \Pr(X \in C \mid \theta) \quad \text{for } \theta \in \Omega \tag{6}$$

The likelihood of an incorrect decision is minimized by minimizing $\pi(\theta)$ for the region Ω_0. In many problems, an upper bound $\alpha_0(0 < \alpha_0 < 1)$ is specified and tests are considered for which $\pi(\theta) \leq \alpha_0$ for every $\theta \in \Omega_0$. The constraint on this category of error is defined as the level of significance of the test. For the region labeling approach, the three individual hypothesis tests are carried out at a level of significance of 0.05.

First Test for Merging—The F-Test: this test for the equivalence of the variances of the image intensities of the micro-level regions is specified[35] by the equation:

$$r(x,y) = \frac{\sup_{(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2) \in \Omega_1} g(x,y \mid \mu_1,\mu_2,\sigma_1^2,\sigma_2^2)}{\sup_{(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2) \in \Omega_0} g(x,y \mid \mu_1,\mu_2,\sigma_1^2,\sigma_2^2)}, \tag{7}$$

where $r(x,y)$ is the discriminant of the hypothesis test; x the vector of image intensity samples from region 1; y the vector of image intensity samples from region 2; $g(x,y \mid \mu_1,\mu_2,\sigma_1^2,\sigma_2^2)$ is the joint likelihood of the samples from the neighboring regions; μ_1,σ_1^2 represent the mean and variance of the first region, respectively; and μ_2,σ_2^2 represent the mean and variance of the second region, respectively.

If image intensity samples from each region are viewed as being Gaussian random variables whose mean is μ_i, $i = 1,2$, and variance is σ_i^2, $i = 1,2$, then the preceding equation can be rewritten as

$$r'(x,y) = \frac{\sum_{i=1}^{m} (x_i - \overline{x_m})^2 / (m-1)}{\sum_{i=1}^{n} (y_i - \overline{y_n})^2 / (n-1)}, \tag{8}$$

where $\overline{x_m}$ is the estimated mean of the image intensity samples (m) for the first region $(= \frac{1}{m} \sum_{i=1}^{m} x_i)$, and $\overline{y_n}$ the estimated mean of the image intensity samples (n) for the second region $(= \frac{1}{n} \sum_{i=1}^{n} y_i)$.

The attributes of the statistical distributions of the image intensities that we wish to test are

$$H_0 : \sigma_1^1 = \sigma_2^2, H_1 : \sigma_1^1 \neq \sigma_2^2. \tag{9}$$

It can be shown that under the hypothesis H_0 the discriminant $r'(x,y)$ is distributed as an F random variable of m-1 and n-1 degrees of freedom. Because the percentage levels of the F-distribution having m-1 and n-1 degrees of freedom can be tabulated numerically, the discriminant threshold can be calculated to obtain a prerequisite level of significance. The hypothesis test of equation (9) specifies that the critical region is obtained as two subregions, the region where $r'(x,y) \leq c_1$ and the region where $r'(x,y) \geq c_2$. If misclassification errors for either of these regions are viewed as being equally undesirable, then the discrimination levels are obtained as the $\frac{1}{2}\alpha_0$ and $1 - \frac{1}{2}\alpha_0$ percentage levels of an F-distribution of m-1 and n-1 degrees of freedom. Thus, the hypothesis test procedure is as follows: reject the null hypothesis H_0 if $r'(x,y) < c_1$ or $r'(x,y) > c_2$, where c_1 and c_2 are the $\frac{1}{2}\alpha_0$ and $1 - \frac{1}{2}\alpha_0$ percentage points of an F-distribution with m-1 and n-1 degrees of freedom.

Second Test for Merging—t-Test: It is used for the equivalence of the means of the image intensity of micro-level regions. This discrimination function is calculated only if the variances of the image intensities are established to be statistically equivalent because the derivation of the hypothesis test is predicated on the samples being measurements obtained from independent Gaussian distributions with unknown means and one known variance. The hypotheses for this test are

$$H_0 : \quad \mu_1 = \mu_2, \\ H_1 : \quad \mu_1 \neq \mu_2. \tag{10}$$

The discriminant of the hypothesis test can be established from the ratio of the joint likelihoods of the samples for the distinct hypotheses' subspaces. The ratio of the joint likelihoods can be transformed into the following equation:

$$r'(x,y) = \\ \frac{(m+n-2)^{1/2} \mid \overline{x_m} - \overline{y_n} \mid}{\left(\frac{1}{m} + \frac{1}{n}\right)^{1/2} \left[\sum_{i=1}^{m} (x_i - \overline{x_m})^2 + \sum_{i=1}^{n} (y_i - \overline{y_n})^2 \right]^{1/2}}. \tag{11}$$

If $\mu_1 = \mu_2$ (the hypothesis H_0 is true), then the discriminant function $r'(x,y)$ has a t-distribution of $m+n-2$ degrees of freedom. The t-distribution function is symmetrical and displays equal likelihoods for the coordinates $\pm x$. Thus, the hypothesis test is specified

equivalently as follows: reject H_0 if $\mid r'(x,y) \mid > c$, where c is the $1 - \frac{1}{2}\alpha_0$ percentage point of a t-distribution of $m+n$-2 degrees of freedom.

Third Test for Merging—Kolmogorov–Smirnov Test: this hypothesis test is a non-parametric decision strategy because the derivation makes no requirements on the categories of distributions to be used for the test. The hypothesis test is derived on the basis of the established statistical behavior of empirical distribution functions. For this test, we assume that an array of samples $X_1, \ldots, X_m$ is taken from a distribution for which the distribution function $F(x)$ is unknown, and an independent array of samples $Y_1, \ldots, Y_m$ is taken from another distribution for which the distribution function $G(x)$ is also unknown. The hypothesis test is based on the assumption that $F(x)$ and $G(x)$ are continuous functions and the hypothesis test establishes the statistical equivalence of the distribution functions, or a decision is made between the hypotheses:

$$H_0 : \quad F(x) = G(x), \text{ for } -\infty < x < \infty$$
$$H_1 : \quad \text{Hypothesis } H_0 \text{ is true.} \qquad (12)$$

For each number $x(-\infty < x < \infty)$, the value of the sample distribution function, $F_m(x)$, is defined to be the proportion of observed values in the sample which are less than or equal to x. In other words, if exactly k of the observed values in the sample are less than or equal to x, then $F_m(x)=k/m$.

The sample distribution function $F_m(x)$ can be regarded as the distribution function of a discrete distribution which assigns probability $1/m$ to each of the m values $x_1, \ldots, x_m$. Thus, $F_m(x)$ will be a step function with a jump of magnitude $1/m$ at each point x_i $(i = 1, \ldots, m)$.

If the distribution function of the image intensity samples is $F(x)$, then for any x, $-\infty < x < \infty$, the probability that any particular X_i will be less than or equal to x is $F(x)$. Therefore, it follows from the law of large numbers that, as $m \rightarrow \infty$, the proportion $F_m(x)$ of observations in the sample which are less than or equal to x will converge in probability to $F(x)$. A stronger result is the Glivenko–Cantelli lemma, which states that $F_m(x)$ converges to $F(x)$ uniformly over all values of x.

The third hypothesis test is derived from a statistic obtained from the sample distribution functions $F_m(x)$ and $G_n(x)$ for sample measurements $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$, respectively. The statistic is denoted D_{mn} and is obtained from the following equation:

$$D_{mn} = \sup_{-\infty < x < \infty} \mid F_m(x) - G_n(x) \mid \qquad (13)$$

If the hypothesis H_0 is true (the distribution function $F(x)$ is equivalent to the distribution function $G(x)$), then the sample distribution functions will tend to be close to each other. In fact, when H_0 is true, it follows from the Glivenko–Cantelli lemma that:

$$p \lim_{\substack{m \to \infty \\ n \to \infty}} D_{mn} = 0 \qquad (14)$$

The hypothesis test is based on a theoretical result regarding the statistical behavior of the statistic D_{mn}. If the hypothesis H_0 is true, then the statistical attributes of D_{mn} are defined by the following equation:

$$\lim_{\substack{m \to \infty \\ n \to \infty}} \Pr\left(\left[\frac{mn}{m+n}\right]^{1/2} D_{mn} \leq t\right)$$
$$= 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1}\exp(-2i^2t^2). \qquad (15)$$

Thus, the decision strategy is to reject the hypothesis H_0 if the magnitude of the statistic D_{mn}, scaled by the factor $(mn/(m+n))^{1/2}$, is greater than the α_0 percentage level of the function of the right-hand side of equation (15).

If all three hypothesis tests do not negate (a very conservative approach) the equivalence of the statistical distributions of the samples obtained from the adjacent regions, then the regions are designated as measurements from the same terrain category. The estimation of the equivalence of the density functions of the samples is carried out for all pairs of regions obtained from the preceding stage, and then all pairs of regions that remain after the first iteration of region merging, and so on, until no changes occur. Finally, the boundary image is revised to account for the merged regions.

3.2.4. Knowledge-based classification (micro-level).
The region labeling scheme at the micro-level is identical to that at the macro-level, except that only spectral features are used for micro-level region classification. This approach is utilized because the immense variability of arbitrary outdoor scenes prevents the specification of valid locational and relational constraints for all scenarios. Therefore, the micro-level knowledge-based classifier derives terrain classifications by associating with each region that region category whose *a posteriori* Bayesian likelihood is the highest of all classes. Thus, if $P(f_i \mid c)$ denotes the probability that a region possesses a feature f_i given that the segment belongs to class c, the region is classified as belonging to class c^* if c maximizes $p(c \mid f_1, f_2, \ldots, f_N)$, where $p(c \mid f_1, f_2, \ldots, f_N)$ is defined by the following equation:

$$p(c \mid f_1, f_2, \ldots, f_N) = \frac{\prod_{i=1}^{N} p(f_i \mid c)P(c)}{\sum_{j=1}^{J} \prod_{i=1}^{N} p(f_i \mid c_j)P(c_j)}, \qquad (16)$$

where N is the number of database (independent) features employed for the classification of a region at the micro-level to class c, and J is the number of micro-level classes for the specific macro-level region under consideration. The c^* that maximizes the above expression is the Bayes classification for the micro-level region.

3.2.5. Refinement of region boundaries.
Because the multispectral texture-gradient edge linking relaxation algorithm (Fig. 2) can yield false region boundaries where the texture gradient strength at the true structural

boundaries of the image is low, a region editing step is required. The small regions which share the same classification as neighboring larger regions are merged to the larger region if the degree of disparity between the mean and the variance of the regions under consideration does not test to be significant (utilizing the hypothesis test procedure described in Section 3.2.3). After the application of all appropriate spectral constraints from the knowledge base, neighboring micro-level regions that share the same terrain classification are merged to form larger regions. This process continues until every pair of adjacent micro-level regions are elements of disjoint terrain categories. The region categories employed for micro-level classification for fields are: gamma grass, snowberries, scrub bush (e.g. gambel oak, yucca, etc.) and gravel

Fig. 6. Display of 12 channels of multi-spectral scanner image MULTI36.

road. The region categories utilized for micro-level region classification of forest regions are: gamma grass, scrub bush (e.g. gambel oak) and trees (e.g. juniper, mountain mahogany, etc.). The region categories employed for micro-level classification of road regions are: gravel road and grass. Sky regions are not segmented into subregions.

4. EXPERIMENTAL RESULTS

The terrain interpretation system is implemented on two platforms. The macro- and micro-level segmentation steps are implemented in the C programming language on a SUN 4 computer under the SUNOS UNIX 4.1.1 operating system. The knowledge-based region labeling steps at the macro- and micro-levels are implemented in a pattern-based reasoning language package, written in Common Lisp, on a Symbolics 3670. Data is transferred from one computer to the other by means of a ChaosNet File Transfer Protocol (FTP). The MS imagery input to the system are obtained from the Collage I database generated with the ERIM passive MSS at the Martin-Marietta ALV Test Site on 3rd July 1985.[31]

Figure 6 displays the twelve MS channels from the image MULTI36 ("Segment O") of the Collage I data set. These are shown sequentially going from left to right across the figure, where the first row begins at the top left corner of the figure, the next row is the one below it, etc. As a result of the scanner electronics malfunction, no data is available for channel 9 (third row, first column of Fig. 6). Edge Linking Relaxation results for image MULTI36 are shown in Fig. 7. The
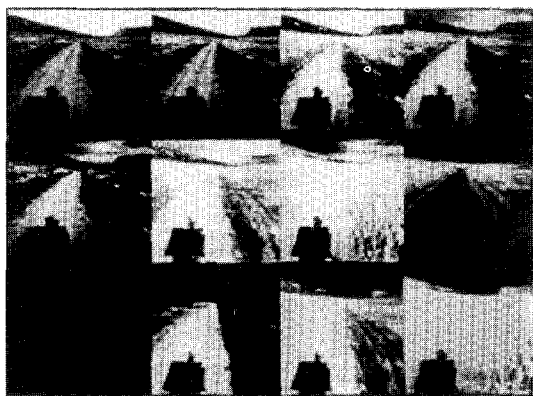

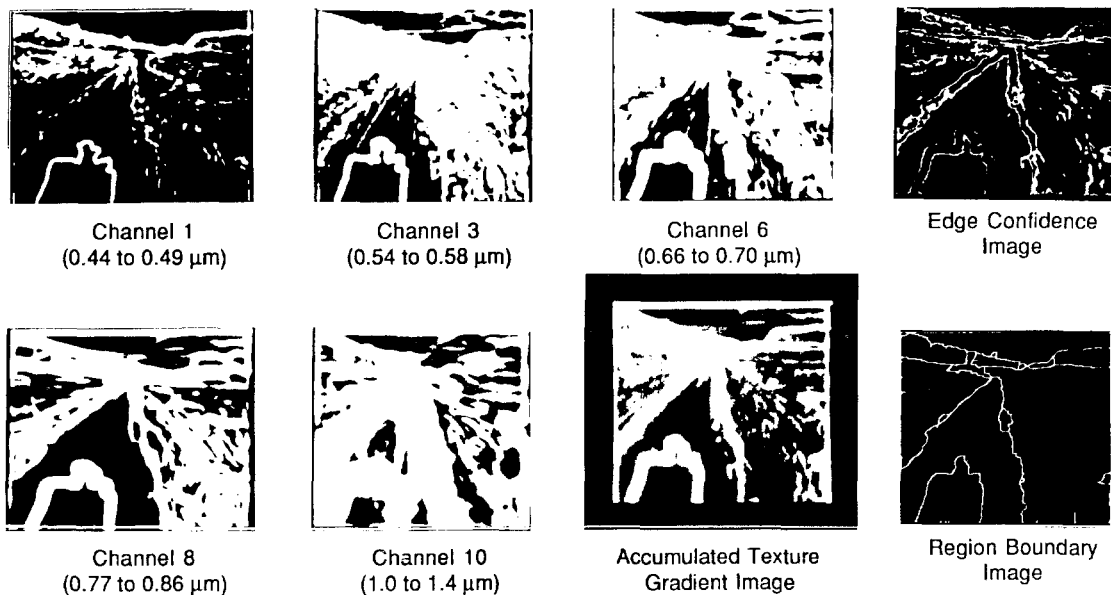|			
Channel 1 (0.44 to 0.49 µm)	Channel 3 (0.54 to 0.58 µm)	Channel 6 (0.66 to 0.70 µm)	Edge Confidence Image
Channel 8 (0.77 to 0.86 µm)	Channel 10 (1.0 to 1.4 µm)	Accumulated Texture Gradient Image	Region Boundary Image

Fig. 7. Texture gradient images for channel 1, 3, 6, 8 and 10 (image MULTI36). As described in Fig. 2, the accumulated texture gradient image, edge confidence image and region boundary image results at the macro-level are also shown.

figure shows the texture gradient images for MSS channels 1, 3, 6, 8, and 10 and the additive combination (normalized) of the five texture gradient images (the accumulated texture gradient image in Fig. 2). The estimated image textural qualities are represented as intensity images, where the image intensity is proportional to TBL gradient magnitude. The significant region boundaries are very sharp in the accumulated texture gradient image, while the individual texture gradient images demonstrate varying degrees of significant region boundary sharpness, primarily as a function of range. When the accumulated texture gradient image is subjected to multi-thresholding, the result is an edge confidence image. The edge confidence image is depicted in an encoded pseudo-color display in Fig. 7. The region boundary image is obtained from the edge confidence image by iterative continuation of unclosed boundaries. The optimal representation for the
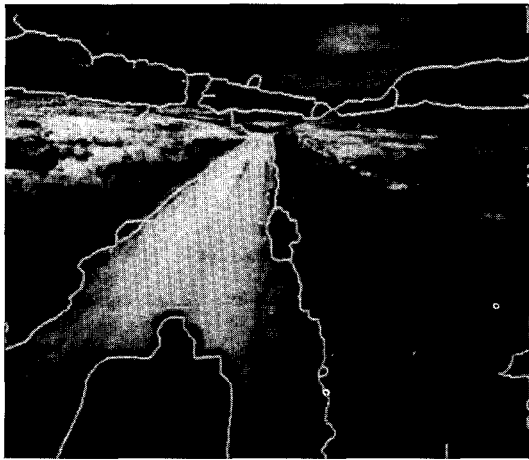
existence of region boundary locations is determined by the relaxation-based edge linking algorithm to derive the macro-level region boundaries.

Figure 8 is the luminance image (the Y image of the NTSC television standard for color imagery transmission) of the MS image MULTI36 over which the macro-level region boundaries are superimposed. For the purpose of visual verification of the results, the color image without superimposed macro-level region boundaries is presented in Fig. 9(a). Comparing Figs 8 and 9(a), we find that the segmentation of all terrain classes, sky, forest, field and road, are good except for the left and right forks of the road. These fork-region boundaries are not detected because they are located at too great a distance from the MSS. The range of distances at which road boundaries may be detected is a function of the parameters of the TBL algorithm for MS images, 8 and 10, which are the best channels for detecting road boundaries (see Table 1). Because the edges of road boundaries for these channels are not sharp, the parameters of the TBL algorithm are set for wide boundaries ($K = 9$, $N = 11$ for channel 8 and $K = 11$, $N = 13$ for channel 10). When road regions are located at a great distance from the MSS, such as the two forks of the road, both borders of the road lie inside the texture gradient measurement window (see Fig. 3). Therefore, the texture in their vicinity will appear to be homogeneous and the TBL will not detect their boundaries. According to specifications,[36] the image processing requirements for vehicle guidance at a velocity of 10 km/h are 15 m for the MSS. Because the distance to the fork in the road is approximately 30 m from the vehicle for MULTI36, the failure to detect the forks is acceptable according to the specifications. This lack of information can be supplanted with other sources, such as a knowledge base enhanced with elevation map, land cover map, or range sensor information, and temporal evidence,[37] for cross-

Fig. 8. Macro-level region boundaries superimposed on the luminance image for MULTI36.

Fig. 9. Segmentation and region labeling results for image MULTI36: (a) color image corresponding to image MULTI36 (Red=channel 6, Green=channel 3 and Blue=channel 1); (b) labeled macro-level regions for multi-spectral scanner image MULTI36. Labeled regions, r=road, g=field, f=forest, s=sky and u=unknown.

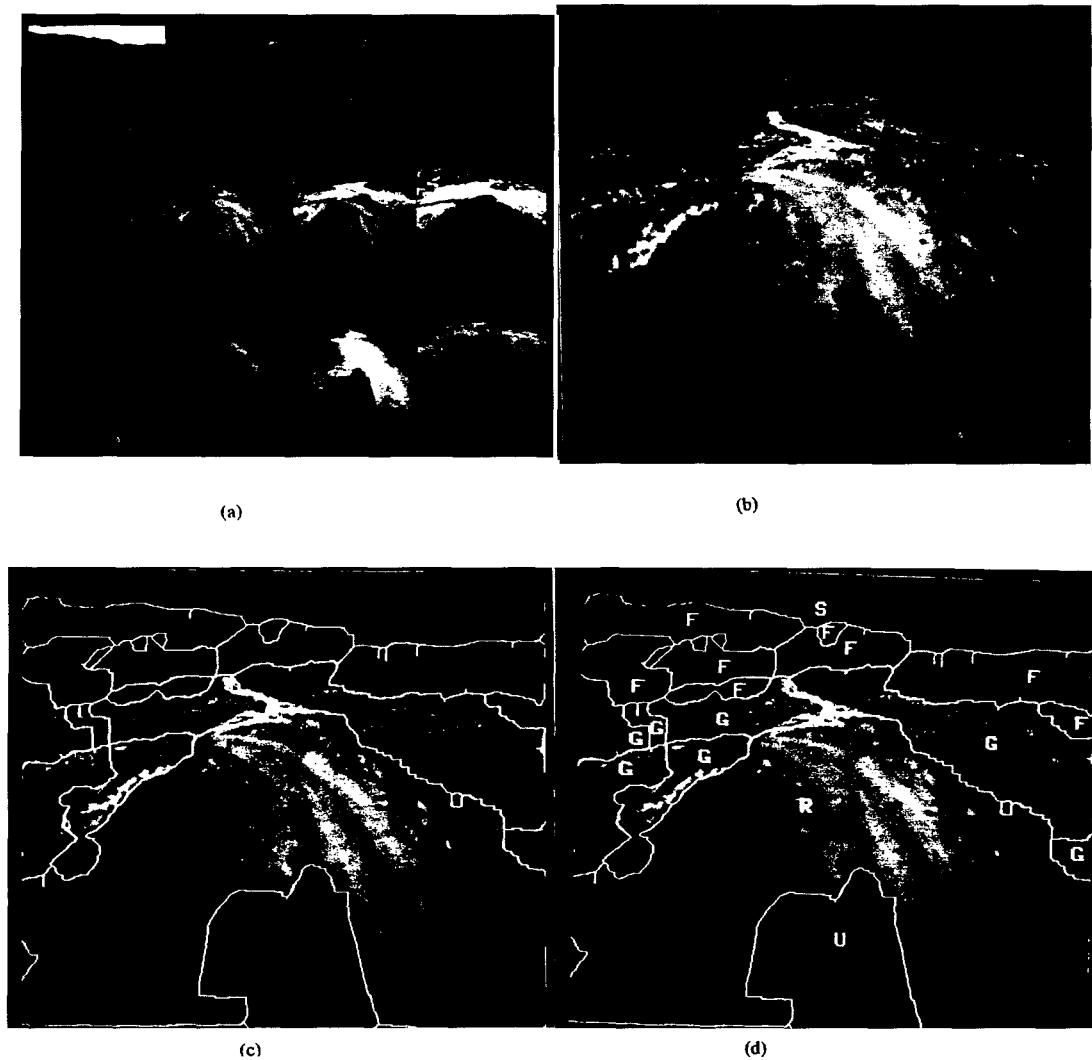(c)   (d)

Fig. 10. Segmentation and region labeling results for image MULTI12: (a) display of 12 channels of multi-spectral scanner image MULTI12 (pseudo colored); (b) luminance image (pseudo colored); (c) macro-level region boundaries superimposed on the luminance image; (d) labeled macro-level regions for multi-spectral scanner image MULTI12. Labeled macro-level regions, R=road, G=field, F=Forest, S=sky and U=unknown.

country navigation. The results of macro-level region segmentation of MULTI36 are presented in Fig. 9(b).

A second set of macro-level region segmentation results are presented in Fig. 10. Figure 10(a) shows the twelve channels of the MSS imagery MULTI12 ("Segment V") from the Collage I database and Fig. 10(b) is the luminance image corresponding to MULTI12. The extracted macro-level region boundaries and the labeled regions superimposed on the luminance image are shown in Fig. 10(c) and (d), respectively. Note that the foothills are accurately segmented and that the occluding borders of hills in the foreground are detected as well.

The results of the micro-level region segmentation for the field region using MULTI36 are demonstrated in Figs 11–13. During this segmentation process, the field regions of Fig. 9(b) are further subdivided into subregions of snowberries, scrub bushes, red clay-like

soil, and grass. The sources of evidence for this stage are MSS channels 5, 8, and 10 (see Table 2). We have used only one channel (channel 5) from channels 4, 5 and 6 to reduce the amount of computation. The parameters K and N for the TBL algorithm for these MS images are: MSS channel 5, $K = 2$, $N = 3$; MSS channel 8, $K = 3$, $N = 5$; and MSS channel 10, $K = 3$, $N = 5$. Figure 11 shows the texture gradient results on the selected channels, accumulated texture gradient images, edge confidence image and region boundaries obtained after the application of the Edge Linking Relaxation algorithm. Small regions are merged with adjacent regions using the technique as described in Section 3.2.3. The results after region merging are shown in Fig. 12(a). Figure 12(b) shows the labeled regions at the micro-level. The labels are grass, scrub bushes, snowberries and road. As mentioned in Section 3.2.4, only spectral features are used for labeling the regions. The final

Channel 5
(0.62 to 0.66 μm)

Channel 8
(0.77 to 0.86 μm)

Channel 10
(1.0 to 1.4 μm)

Accumulated
Texture Gradient
Image

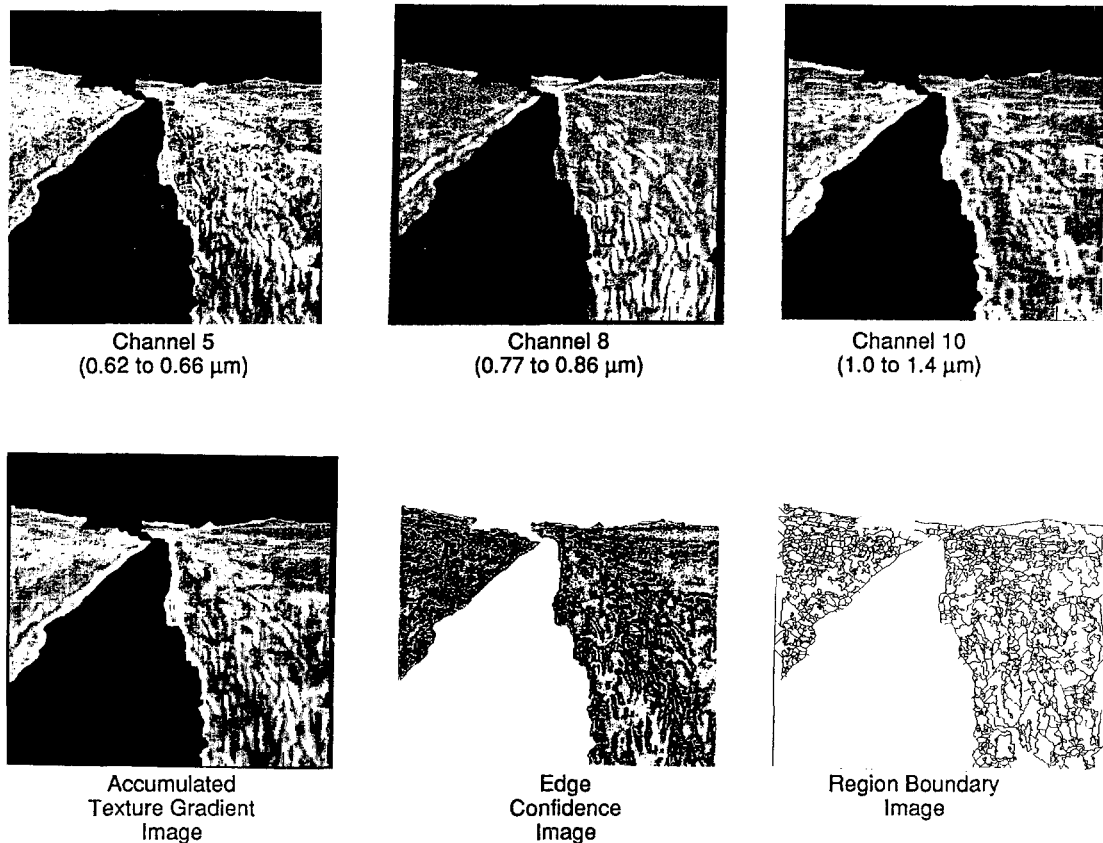Edge
Confidence
Image

Region Boundary
Image

Fig. 11. Texture gradient images, for channels 5, 8 and 10 (image MULTI36), used for the segmentation of micro-level region called "field". As described in Fig. 2, the accumulated texture gradient image, edge confidence image and region boundary image results at the micro-level are also shown for the field region.

(a)

(b)

Fig. 12. Micro-level terrain segmentation for the "field" region (image MULTI36): (a) micro-level regions; (b) labeled regions, mustard=grass, yellow=scrub bushes, tan=snowberries, blue=road.

region classification results are shown in Fig. 13; Fig. 13(a) shows the image MULTI36 and Fig. 13(b) shows the labeled macro- and micro-level regions (for the field region) superimposed on the image.

5. CONCLUSIONS

This paper has presented a technique for the detection of structural region boundaries in terrain images using a

mobile platform-mounted MSS sensor. The problem of region classification is solved using knowledge-based techniques. To reliably label each region in the input image, it has been argued that the terrain interpretation system must make use of *a priori* information as much as possible because of the immense variability of outdoor scenes. The *a priori* information can assume various forms, for instance, the nominal location of each terrain class in the image as a function of the imaging
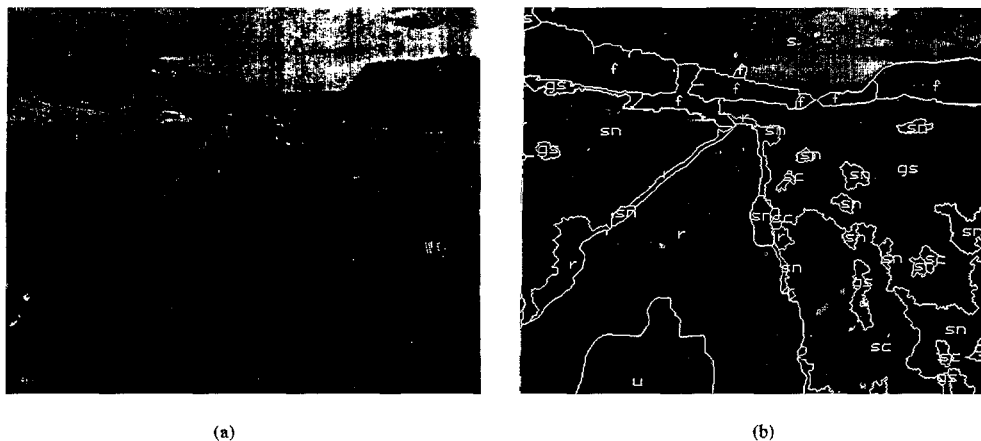
Fig. 13. Terrain segmentation for image MULTI36: (a) collage I color image—MULTI36; (b) terrain region classifications. Labeled regions, s=sky, f=forest, r=road, g=grass, sn=snowberry, sc=scrub bushes, u=unknown.

system's orientation. The reliability of the overall approach is enhanced by the application of constraints on the region labeling process where the constraints are derived from the *a priori* information about the geographic region where the vehicle is traveling. Experimental results showing extracted macro-level and micro-level region boundaries demonstrate the efficacy of the method. Since the knowledge base can include *a priori* information about terrain elevation, land cover, geological data, time of day, season, and the like, the described work should be useful in the development of a totally autonomous system navigating in an unstructured natural environment. The hierarchical, region labeling strategy discussed in this paper can be used to establish reliable scene classification maps for other application areas such as remote sensing and photointerpretation, because application-specific constraints also exist for these problems. The requirements and practices of these application areas and the specific attributes of the geographic locations where the systems are used can be mapped to constraints on the applicable labels to be used at each stage of the hierarchy. Thus, the same baseline terrain interpretation system whose functionality has been demonstrated in this research can find usage in diverse areas without requiring major modifications.

REFERENCES

1. E. D. Dickmanns and B.D. Mysliwetz, Recursive 3-D road and relative ego-state recognition, *IEEE Trans. Pattern Analysis Mach. Intell.* **14**(2), 199–213 (1992).
2. C. A. Harlow, M.H. Trivedi, R. A. Conners and D. Phillips, Scene analysis of high resolution aerial scenes, *Optical Engng* **25**(3), 347–355 (1986).
3. D. A. Landgrebe, Analysis technology for land remote sensing, *Proc. IEEE* **69**(5), 628–642 (1981).
4. M. D. Levine and A. M. Nazif, Low level image segmentation: An expert system, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-6**(5), 555–577 (1985).
5. Y. Ohta, *Knowledge-based Interpretation of Outdoor Natural Scenes.* Pitman, London (1985).
6. S. M. Rubin, Natural scene recognition using locus search, *Comput. Graphics Image Process.* **13**, 298–333 (1980).
7. D. M. McKeown, Jr. Knowledge based aerial photo interpretation, *Photogrammetria* **39**, 91–123 (1984).
8. M. Nagao and T. Matsuyama, *A Structural Analysis of Complex Aerial Photographs*, Plenum Press, New York (1980).
9. W. A. Perkins, T. J. Laffey and T. A. Nguyen, Rule-based interpretation of aerial photographs using the Lockheed Expert System, *Optical Engng* **25**(3), 356–362 (1986).
10. L. Sauer and J. Taskett, Cultural feature and syntax analysis for automatic acquisition, *SPIE Conf. Processing of Images and Data from Optical Sensors* **292**, 270–276 (1981).
11. J. Ton, J. Stricklen and A.K. Jain, Knowledge-based Segmentation of Landsat Images, *IEEE Trans. Geoscience Remote Sensing* **GE-29**(2), 222–232 (1991).
12. R. M. Hord, *Remote Sensing Methods and Applications.* Wiley, New York (1986).
13. P. H. Swain, Advanced interpretation techniques for earth data information systems, *Proc. IEEE* **73**(6), 1031–1039 (1985).
14. J. A. Richards, D. A. Landgrebe and P. H. Swain, Pixel labeling by supervised probabilistic relaxation, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-3**, 181–191 (1981).
15. A. Rosenfeld, C-Y. Wang and A. Wu, Multispectral texture, *IEEE Trans. Systems Man Cybernet.* **SMC-12**(1), 79–84 (1982).
16. J. M. Brayer, P. H. Swain and K. S. Fu, Modeling earth resources with satellite data, in *Syntactic Pattern Recognition Applications*, K. S. Fu, ed. Springer, Berlin (1977).
17. T. Kusaka and Y. Kawata, Hierarchical classification of LANDSAT TM image using spectral and spatial information, *Int. Geoscience and Remote Sensing Symposium (IGARSS '91)*, 2187–2190 (June 1991).
18. P. Meyer, Segmentation and symbolic description for a classification of agricultural areas with multispectral scanner data, *IEEE Trans. Geoscience and Remote Sensing* **GE-30**(4), 673–679 (1992).
19. D. D. Giusto, L. Parodi and G. Vemazza, Accurate segmentation of multispectral remote sensing images,

Proc. 6th Scandinavian Conf. Image Analysis, 936–939 (June 1989).

20. M. C. Zhang, R. M. Haralick and J. B. Campbell, Multispectral image context classification using stochastic relaxation, *IEEE Trans. Systems Man Cybernet.* **SMC-20**(1), 128–140 (1990).

21. M. Goldberg, D. G. Goodenough, M. Alvo and G. M. Karam, A hierarchical expert system for updating forestry maps with landsat data, *Proc. IEEE* **73**(6), 1054–1063 (1985).

22. A. Rangarajan, R. Chellappa and B. S. Manjunath, Random fields and neural networks with applications to early vision, in *Artificial Neural Networks and Statistical Pattern Recognition: Old and New Connections*, I. K. Sethi and A. K. Jain, Eds. Elsevier Science, New York (1991) .

23. C. A. Therrien, An estimation-theoretic approach to terrain image segmentation, *Comput. Graphics Image Process.* **22**, 313–326 (1983).

24. K. V. Mardia and T. J. Hainsworth, Spatial thresholding method for image segmentation, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-10**(6), 919–927 (1988).

25. C. Bouman and M. Shapiro, Multispectral image segmentation using a multiscale model, *Proc. IEEE Int. Conf. Acoustics Speech Signal Process.*, 565-568 (March 1992).

26. R. A. Fernandes and M. E. Jernigan, Unsupervised multiscale segmentation of multispectral imagery, *Proc. IEEE-SP Int. Symp. Time-Frequency and Time-Scale Analysis*, 547–550 (October 1992).

27. M. Amadasun and R. A. King, Low-level segmentation of multispectral images via agglomerative clustering of uniform neighborhoods, *Pattern Recognition* **21**(3), 261–268 (1988).

28. H. Hanaizumi, H. Okumura, H. Tsubaki and S. Fujimura, A supervised classification algorithm for remotely sensed multispectral images by using spatial segmentation, *Proc.*

Second Int. Symp. Noise and Clutter Rejection in Radars and Imaging Sensors, 136–141 (November 1989).

29. C. Salvaggio and J. R. Schott, Automated segmentation of pseudoinverse features from multispectral imagery, *Proc. SPIE Three-Dimensional Imaging and Remote Sensing Imaging* **902**, 118–127 (January 1988).

30. M. H. Brill, Object-based segmentation and color recognition in multispectral images, *Proc. SPIE Image Understanding and the Man-Machine Interface II* **1076**, 97–103 (January 1989).

31. J. A. Allison, COLLAGE: A collection of sensor images for the ALV test area, Technical Report, Martin Marietta Corporation (December 1985).

32. J. N. Rinker, J. P. Henley and M. B. Satterwhite, Terrain data base—Air photo analysis, Martin Marietta ALV Test Site, Technical Report, Martin Marietta Corporation (January 1986).

33. D. Zuk and L. Harmon, Active multispectral 3-D sensor design, Technical Paper from Environmental Research Institute of Michigan, Ann Arbor, Michigan (November 1986).

34. W. Au, S. Mader and R. Whillock, Scene analysis, Third Triannual Technical Report to Center for Night Vision and Electro-Optics, Contract No. DAAL01-85-C-0429, Honeywell Systems and Research Center, Minneapolis (July 1986).

35. M. DeGroot, *Probability and Statistics*. Addison-Wesley, Reading, Massachusetts (1975).

36. J. Lowrie, The autonomous land vehicle—Second quarterly report, Technical Report, Martin Marietta Corporation (September 1985).

37. B. Bhanu, P. Symosek, J. Ming, W. Burger, H. Nasr and J. Kim, Qualitative target motion detection and tracking, *Proc. DARPA Image Understanding Workshop*, 370–398 (May 1989).

About the Author — BIR BHANU received the S.M. and E.E. degrees from Massachusetts Institute of Technology, the Ph.D. degree from the Image Processing Institute, University of Southern California and the M.B.A. degree from the University of California, Irvine. Since 1991, Dr Bhanu has been a Professor and Director of Visualization and Intelligent Systems Laboratory at the University of California, Riverside. Prior to that he was a Senior Honeywell Fellow at Honeywell Systems and Research Center in Minneapolis. He has been the principal investigator of various programs from ARPA, AFOSR, ARO, NASA, NSF and other agencies and industries. He has five patents and over 150 reviewed publications in the areas of computer vision, image processing, pattern recognition, artificial intelligence and learning. He was the General Chair for the IEEE Conference on Computer Vision and Pattern Recognition held in June 1996 at San Francisco, California.

About the Author — PETER F. SYMOSEK was born in Lawrence, MA, on 22 September 1953. He received the B.S. degree in Electrical Engineering from Merrimack College, North Andover, MA, in 1978, and M.Sc. and Ph.D. degrees in Electrical Engineering, with minors in Computer Science and Applied Mathematics, from Brown University, Providence, RI, in 1980 and 1985, respectively. From 1978 to 1985 he was a Research Assistant at the Division of Engineering of Brown University. The research projects he was involved with during this time were concentrated in the fields of boundary finding for outdoor scenes, and the derivation of algorithms for object classification and for estimating the location and orientation of objects, employing specular image data. Since 1985 he has been working at Honeywell, Inc., Minneapolis. At Honeywell, he has been doing research on algorithm design and analysis for autonomous outdoor robotics, space image processing, and sensor/processor trade off studies for implementation of multisensor algorithms for surveillance and tactical applications.

About the Author — SUBHODEV DAS received his B.Tech. (Hons.) degree in Electronics and Electrical Engineering from the Indian Institute of Technology, Kharagpur, in 1984, his M.S. degree in Electrical Engineering from the University of Hawaii in 1986, and his Ph.D. degree in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign in 1991. Between 1991 and 1994, he was with the College of Engineering at the University of California, Riverside. Currently, he is a research staff member at PEB Inc., Princeton, working in real-time computer vision. His research interests include intelligent systems, computer vision, machine learning, human-computer interaction, real-time computing, parallel and distributed processing, and applications of artificial intelligence, signal and image processing. He is a member of the IEEE.