An Evaluation of the Accuracy and Feasibility of Identifying Risk Factors Using Remote Sensing Techniques Along Transportation Corridors

Coler & Colantonio, Inc.
Brown University
Affiliated Research Center

submitted to
NASA Commercial Remote Sensing Program Office
John C. Stennis Space Center, MS 39529

1999
Final Report:
An Evaluation of the Accuracy and Feasibility of Identifying Risk Factors Using Remote Sensing Techniques Along Transportation Corridors

Project conducted by:
Coler & Colantonio, Inc.
101 Accord Park Drive
Norwell, Massachusetts 02061-1685

Report prepared by:
Jeff Allan
Coler & Colantonio, Inc.
and
Dr. John F. Mustard, Melanie Harlow, and Lin Li
Brown University

Report prepared for:
Brown University
Department of Geological Sciences, Box 1846
Providence, Rhode Island 02912
and
Commercial Remote Sensing Program Office
National Aeronautics and Space Administration
John C. Stennis Space Center, Mississippi 39529

July 20, 2000
# Table of Contents

Executive Summary .................................................................................................................. v  

1.0 Affiliate Background ........................................................................................................... 1  

2.0 Project Description ............................................................................................................ 1  

2.1 Current Procedures ........................................................................................................ 1  

2.2 Potential for Gains through Automation ....................................................................... 2  

2.3 Project Objectives .......................................................................................................... 2  

2.4 Accuracy Standards ....................................................................................................... 3  

2.5 Methodology ................................................................................................................ 3  

2.5.1 Thresholding ............................................................................................................. 3  

2.5.2 Mapping Anomalies ................................................................................................. 3  

2.5.2.1 Minimum Noise Fraction Transformation ......................................................... 3  

2.5.2.2 Principal Components Analysis ...................................................................... 4  

2.5.2.3 Spectral Mixture Analysis ............................................................................... 4  

2.5.2.4 Matched Filtering ............................................................................................ 5  

2.5.3 Image Algebra ........................................................................................................ 5  

2.5.4 Raster to Vector Conversion .................................................................................... 5  

2.6 Data Sets ...................................................................................................................... 5  

3.0 Project Implementation .................................................................................................... 6  

3.1 Developing Methodology ............................................................................................... 6  

3.2 Finalized Methodology .................................................................................................. 6  

3.2.1 Step 1 ....................................................................................................................... 6  

3.2.2 Step 2 ....................................................................................................................... 6  

3.2.3 Step 3 ....................................................................................................................... 7  

3.2.4 Step 4 ....................................................................................................................... 7  

3.2.5 Step 5 ....................................................................................................................... 8  

3.2.6 Step 6 ....................................................................................................................... 8  

3.3 Why Not SMA? .......................................................................................................... 8  

3.4 Why Not PCA? ............................................................................................................ 9  

4.0 Observations ................................................................................................................ 9  

4.1 Observations Based on Images .................................................................................... 9  

4.1.1 Wareham, Massachusetts .................................................................................. 9  

4.1.2 Brown University ............................................................................................... 9  

4.1.3 Cape Cod, Massachusetts ................................................................................. 9  

4.1.4 Narragansett, Rhode Island .............................................................................. 10  

4.1.5 Barnstable, Massachusetts ................................................................................. 10  

4.2 Other Observations ..................................................................................................... 13  

5.0 Metrics and Conclusions .............................................................................................. 13  

6.0 References .................................................................................................................. 15  

Appendix A. Overview of ATLAS Images ........................................................................ A-1
Appendix B. Development of GIS Coverage Files ............................................................... B-1
Appendix C. Binary File Creation......................................................................................... C-1
Appendix D. IKONOS Pricing.............................................................................................. D-1

Tables
Table 1. Initial accuracy comparison using Barnstable image. ............................................... 10
Table 2. Final accuracy comparison........................................................................................ 12

Figures
Figure A-1. ATLAS wavelength band descriptions. ............................................................... A-1
Figure A-2. Cape Cod ATLAS image. Bands 7, 5, 4 combination. ........................................ A-2
Figure A-3. Narragansett ATLAS image. Bands 7, 5, 4 combination. ................................. A-3
Figure A-4. Barnstable ATLAS image. Bands 7, 5, 4 combination...................................... A-4
Figure A-5. Wareham ATLAS image. Bands 7, 5, 3 combination. ........................................ A-5
Figure A-6. Brown ATLAS image. Bands 7, 5, 4 combination............................................ A-6
Figure A-7. Clip of Narragansett image, Bands 7, 5, 4. ........................................................ A-7
Figure A-8. Narragansett image from Step 2, identifying mainly building structures.......... A-7
Figure A-9. Narragansett binary image identifying building structures.............................. A-8
Figure D-1. Pricing list for the IKONOS imagery. Obtained from the Space Imaging
    Catalog of Products and Services, September 1999, Volume 1 Supplement.................. D-1
Executive Summary

Risk factors along transportation corridors are computed by assessing the number of structures located within a specified spatial boundary. Once the number of structures has been identified, population densities can be estimated and risk factors can be evaluated. Risk factors are then applied in the design phase to properly assess safety matters along these corridors.

The present procedures for evaluating the risk factors include capturing aerial photography and applying photogrammetric techniques to identify the structures in the photography and to determine the location of those structures. This process is applied once prior to construction activity and annually thereafter to continually evaluate the safety of the system. The federal government requires this type of risk analysis.

Coler & Colantonio, Inc., a survey, civil, and environmental engineering services company, has been actively pursuing a more cost-effective procedure for evaluating risk factors along transportation corridors. Joining together with the Brown University Affiliated Research Center Program, Coler & Colantonio was able to develop and test methods using spectral analysis techniques to identify specific manmade structures within high-resolution multispectral and thermal imagery.

The project provided useful results through building methods that utilized a minimum noise fractions transformation and matched filter. The combination of both the multispectral and thermal data provided the best results.

This project was defined as a positive step toward understanding how remotely sensed imagery could provide information more quickly than traditional methods. This imagery also has potential use for conceptual route layout and for determining areas of change.
1.0 Affiliate Background

Coler & Colantonio, Inc., was founded in 1986 to provide high-quality survey, civil, and environmental engineering services. Since its inception, the firm has experienced a dramatic growth in the number and diversity of its staff. Coler & Colantonio, Inc., currently employs over 150 professionals with education and expertise in such specialized topics as automated mapping and facilities management, pipeline surveying and engineering, hydrographic surveying and engineering, construction management, transportation engineering, land information, planning, municipal water and wastewater engineering, environmental permitting, and hazardous waste remediation. The firm has provided consulting services to a variety of clients, including municipalities, architects, corporations, development firms, law firms, utility companies, engineering firms, and private individuals. Because of its knowledgeable staff and equipment resources, Coler & Colantonio, Inc., is in the unique position of being able to offer its clients highly specialized precision engineering services. These services are provided both nationally and internationally.

The firm takes full advantage of available computer-assisted data collection, modeling, design, and drafting through sophisticated numerical, design, and graphics software. The corporate strategy of Coler & Colantonio, Inc., involves integrating these advances in spatial engineering and computer technology with the knowledge and experience of its staff. By applying creative engineering and spatial solutions, the firm can ensure the completion of projects in a timely, cost-effective manner.

2.0 Project Description

Coler & Colantonio, Inc., has been actively pursuing a more cost-effective procedure for evaluating risk factors along transportation corridors. These risk factors are computed by assessing the number of structures located within a specified spatial boundary. Once the number of structures has been identified, population density can be estimated and risk factors can be evaluated. Risk factors are then applied in the design phase to properly assess safety matters along these corridors.

2.1 Current Procedures

The present procedures for evaluating the risk factors include capturing aerial photography and then having an engineer apply photogrammetric techniques to identify the structures in the photography and to determine the location of those structures within an accuracy of +/- 1 foot. This process is applied once prior to construction activity and annually thereafter to continually evaluate the safety of the system. This type of risk analysis is required by the federal government. When the adjacent population increases substantially around the transportation facility, the operating company may need to upgrade its systems to compensate for the increased risk to the public.

Coler & Colantonio estimates the costs for this procedure to be approximately $240/mile (including aerial photography and manual labor). It is estimated that at present the industry
has over 300,000 miles of transmission pipeline connected to a 1.2-million-mile natural gas
distribution network serving over 175 million customers. One such customer oversees
approximately 26,600 miles of transmission pipeline, and it can cost several million dollars
each year to re-evaluate the risk factors along this particular transportation system.

2.2 Potential for Gains through Automation

With a shrinking workforce and expanding facilities, companies are facing an ever-increasing
workload with less time and budget to complete analyses. At present, the acquisition of the
aerial photography produces the greatest cost to the risk assessments. Of the estimated
$240/mile cost, $200/mile is the cost of the aerial photography. The remainder of the cost is
labor based. An engineer needs approximately one workweek (40 hours) to assess 50 miles
along a transportation corridor for an estimated cost of $40/mile.

By using remotely sensed imagery (other than aerial photography), the overall costs have the
potential to be significantly decreased.

2.3 Project Objectives

The key objective of this project was to determine the presence of manmade structures along
a transmission pipeline from a remote sensing product. The presence of structures would then
be identified in an aggregate structure per mile product. Specifically identifying the size and
function of individual manmade structures was beyond the scope of this project and would
not meet Coler & Colantonio's needs. The aggregate structure per mile product can be
compared against safety requirements mandated by the federal government.

Our goals were as follows:

1) Assess if the use of remotely sensed data could lead to significant savings over the
current costs of $240/mile.

2) Evaluate whether the capabilities of prospective sensors (such as the IKONOS
1-meter panchromatic and 4-meter multispectral) would meet the quality standards of
the present procedures used by the company. Accomplish this evaluation by
simulating the data that are expected to be imaged by the proposed high-resolution
sensors.

3) Evaluate the utility of thermal infrared bands (which are not currently considered for
orbital instruments) to enhance the procedures or results.

The successful location of structures from any one scene was evaluated using the following
criteria:

- The number of structures correctly located.
- The number of objects falsely identified as structures (commission).
- The number of structures missed (omission).
- The spatial accuracy of the located structures.
The transportation corridor used for assessment in this study was an area that was evaluated by Coler & Colantonio. The accuracy and precision assessments were based on these data.

2.4 Accuracy Standards

Coler & Colantonio currently obtains an accuracy of +/- 1 foot from the engineer's photogrammetric technique. The firm’s project experience tells us that locational accuracy of +/- 1 meter would still be valid enough for use in evaluating risk factors along transportation corridors.

To be commercially successful, the structures would need to be correctly identified 95% of the time. This high percentage is driven by two competing factors: 1) satisfying the Department of Transportation’s safety requirements and 2) the pipeline operators' need to be sure that only the pipes that truly move into a new zone are upgraded. Zones are areas identified by the number of risk factor that can be contained within them.

2.5 Methodology

Four methods were explored through this project: thresholding, mapping anomalies using several spectral analysis procedures, combining data, and converting data into a vector dataset.

2.5.1 Thresholding

Thresholding was used to exclude known unwanted data (e.g., clouds, water) and applying spectroscopic principles to model and remove background components, including vegetation (green, non-green), soils, and other materials. Parameters were developed on the basis of fundamental knowledge of their spectra and tailored to perform with the spectral resolution of systems.

2.5.2 Mapping Anomalies

Mapping anomalies is a technique using mixture modeling or orthogonal subspace projection. Again the objectives were to accommodate typical background materials in the modeling and to allow unmodeled materials (e.g., manmade structures) to be partitioned into the error. The four procedures for mapping anomalies were Minimum Noise Fraction (MNF), Principal Components Analysis (PCA), Spectral Mixture Analysis (SMA), and Matched Filtering.

2.5.2.1 Minimum Noise Fraction Transformation

The Minimum Noise Fraction is a transformation based on maximization of the Signal to Noise Ratio (SNR). The MNF transformation has the ability to provide an optimal ordering of images in terms of image quality (Green et al., 1988), which could prove useful when trying to automate the entire structure identification method. The methods developed for this project used the tools available in the ENVI (the Environment for Visualizing Images) remote sensing software package developed by Research Systems, Inc. ENVI's MNF is a two-stage
process. The first stage involves noise-whitening, where a noise covariance matrix is estimated for all the image bands specified. It is not necessary to use all the bands collected by the Airborne Terrestrial Applications Sensor (ATLAS) (see Section 2.6). For this project, three different combinations of the ATLAS bands were used: only the reflectance (multispectral) bands, only the thermal bands, and all the bands. The noise from the image was determined by analyzing a homogenous area. This procedure relies on the theory that the signal at any point in the image is strongly correlated with the signal of neighboring pixels, while the noise shows only weak spatial correlations (Green et al., 1988). Large bodies of water in several of the project’s images were used to derive the covariance matrix. In images lacking bodies of water, the next best homogenous area was chosen, typically uniform vegetated areas, woods, or open grass spaces (sports fields).

The second stage involves the use of a standard PCA of the noise-whitened data. The inherent dimensionality of the data is determined by examination of the final eigenvalues and the associated images (ENVI User's Guide, 1997).

2.5.2.2 Principal Components Analysis

The principal component transformation is a technique designed to remove or reduce redundancy in multispectral data, where the principal component data values are simply linear combinations of the original data values (Lillesand and Kiefer, 1994). This procedure was also available in ENVI, and the ENVI User's Guide describes it as being useful for enhancing the information content, for segregating noise components, and for reducing the dimensionality of the data sets.

The difference between MNF and PCA is in that MNF reflects the real SNR while the PCA doesn't because of the unequal noise variance incurred in the different bands.

2.5.2.3 Spectral Mixture Analysis

The spectral mixture analysis was performed using ENVI. The SMA model assumes that remotely sensed spectroscopic measurements are mixed signatures. The principal objective of SMA is to define a coherent set of spectral endmembers that are representative of physical components on the surface and that model the spectral variability inherent in a given scene (Mustard and Sunshine, 1999). Essentially, this process allows a user to perform a type of sub-pixel analysis by determining key signatures (endmembers) contained within the scene. Tompkins et al. (1997) describes the SMA as having three strengths:

1) SMA is a physically based model that transforms radiance or reflectance values to physical variables, which are linked to the sub-pixel abundance of endmembers within each pixel.

2) SMA provides a means to detect and represent components that occur entirely at a sub-pixel level, such as sparse vegetation in an arid environment.

3) SMA provides quantitative results that can in turn be incorporated into models of the process governing the distribution of material within the image scene.
The SMA procedure in ENVI is referred to as Linear Spectral Unmixing. For this project, the SMA was performed in an attempt to identify the key spectral features within the images, with hopes that the manmade structures would be one of these key features. The structural features were also used as an endmember in the SMA.

2.5.2.4 Matched Filtering

Matched filtering is another spectral analysis operation tested in this project. The Matched Filter maximizes the response of a known endmember and suppresses the response of the unknown background (ENVI User's Guide, 1997). It is used to identify specific materials in a scene based on the input spectral signatures. In this situation, a user-defined image endmember spectral signature was chosen, thus producing a spatial image matching the identified spectral signature.

2.5.3 Image Algebra

Image Algebra involves combining data using specific spectral parameters and thresholds of these parameters to aggregate anomalies.

2.5.4 Raster to Vector Conversion

The product resulting from the imagery must be converted from the raster data format to a vector data format, such as an ESRI (Environmental Research Systems Institute, Inc.) coverage or shape file, to be easily understood by Coler & Colantonio staff and to be compatible with the firm’s in-house geographic information system (GIS). This conversion involves creating a classified or binary image from the results, creating polygons from the grouped raster values, and outputting a point file that identifies the center of each polygon.

2.6 Data Sets

The data utilized in this project were acquired by the ATLAS. The ATLAS is a 15-channel airborne multispectral sensor system with 6 visible to near infrared (VISNIR) channels, 2 of 3 active short-wave infrared (SWIR) channels, and 6 thermal infrared (TIR) channels (Figure A-1). The sensor was developed for NASA’s Commercial Remote Sensing Program Office and is flown on the NASA Stennis Learjet 23. Four ATLAS images were used in this project. Two images were collected with a ground resolution of 7.5 meters per pixel. These images were Cape Cod (Figure A-2), taken on July 9, 1999, at 1:30 pm Eastern Standard Time (EST) and Narragansett (Figure A-3), taken on May 28, 1997, at 10:13 am EST. Three images were collected with a ground resolution of 2.5 meters per pixel. These images were Barnstable (Figure A-4), taken on July 7, 1999, at 1:19 pm EST, Wareham (Figure A-5), taken on July 8, 1999 at 9:45 am EST, and Brown (Figure A-6), taken on July 7, 1999, at 9:45 am EST.

The Cape Cod and Barnstable images were identified for inclusion in this project because Coler & Colantonio had previously obtained risk factor information of those areas that could be used for comparison and analysis.
For the three 1999 flights, color infrared (CIR) photography was collected simultaneously. This photography produced images of a much higher resolution than the ATLAS imagery. Therefore, this imagery proved useful in ground truthing to determine the accuracy and precision of the results.

NASA had previously provided Brown University with ATLAS imagery that was used in this project. This imagery was important to the project because it not only provided visible and IR imagery but also included the thermal IR data.

3.0 Project Implementation

3.1 Developing Methodology

Lin Li, a Brown University graduate student, first began working with the Wareham image, while Coler & Colantonio and the ARC Coordinator narrowed down the specific areas of interest. The Wareham image section was chosen because it was rather simple. This image had very few manmade structures: two subdivisions and some roads. One subdivision had average single-family homes and garages, and the other subdivision had mobile homes. The mobile homes tended to be smaller in size (square footage) and closer together than the single-family homes. This image was used as a test to see if the manmade structures could be identified within the image and to see how clearly each could be identified. It was determined that the structures could be identified fairly clearly; therefore, the work was continued using the other ATLAS imagery.

3.2 Finalized Methodology

This section describes the technical steps taken in processing ATLAS imagery to derive structure density results. The level of human interaction is also discussed. The procedures that were focused on included the MNF, Matched Filter, and use of image algebra.

3.2.1 Step 1

First a homogenous area within the image must be identified. This area is used to derive the noise covariance matrix. A homogenous area can be a large field (with vegetation) or a body of water, for example. For this project, bodies of water were used whenever possible. When using this procedure in ENVI, a person should be involved in the identification process; however, this step could be automated if a special filter were designed to identify this region.

Next, the Forward MNF transform is run and processing moves to Step 2.

3.2.2 Step 2

A person should be involved in the analysis of the MNF images. This person needs to identify which image displays the houses well, which image displays roads well (or other interference), and which images are noisy. Generally the 2nd or 3rd MNF component
identified the homes well and the 4th or 5th MNF component best identified the roads; however, this result was not always consistent. The last few MNF components are too noisy to provide any useful data for further analysis.

If suitable images showing houses and roads are not found, processing moves to Step 3. If suitable MNF component images are found – one image that displays houses well and one image that displays roads well – processing moves to Step 4.

3.2.3 Step 3

The matched filter maximizes the response of a known endmember and suppresses the response of the unknown background. An ideal result depends on one or a set of endmembers.

To select the endmember, a person should decide where the house endmember should be chosen and which endmember is good for optimizing the number of houses identified. This activity has the potential for being automated if these areas were chosen prior to the procedure (i.e., in the beginning or in previous years).

The Matched Filter operation is carried out by using selected endmembers. A person should be involved in picking out the images that have better quality (i.e., show the houses well). Processing then moves to Step 4 to finish the image analysis.

3.2.4 Step 4

This step uses at least two MNF component images: one image that displays houses well and one image that displays roads well. More than two images that can be processed in this step, depending on how many images are chosen from a set of the matched filter images. At least two images are selected because the images that identify the structures (houses) well often have some roads identified, but the images identifying roads clearly may not always show structures (houses). The road image therefore can be used as a filter and the image displaying houses can be processed to remove the roads and to keep the structures (this procedure can be extended to operate on more than two binary images). This step corresponds to a logic operation of two binary images (resulted from thresholding). A person should be involved in deciding the digital number thresholds for the house and road images to produce the binary images. However, some type of automation could be attempted for this step. By examining the histogram of the image of interest and selecting all values within three standard deviations of the mean, a relatively accurate threshold can be identified. Sometimes it can work with the mean + 2.5*(STD). In the cases used in this project, most values for defining which pixels to keep are generally greater than 230.

Processing then moves to Step 5.
3.2.5 Step 5

A binary image in which all pixels identified as structures are highlighted (value=1) and those identified as other are dark (value = 0) is the result of the image algebra performed in step 4. The binary image can then be converted from its raster form into a vector coverage identifying pixel boundaries of the structures. Of the software programs available to the Brown ARC, ESRI's ArcInfo seemed to provide the best conversion (based on ease of use and accuracy).

To create a vector file, the binary image (a Tiff file) is imported into ArcInfo, which creates a grid coverage. A grid is the file format ArcInfo uses to handle raster data. Next, the grid is converted to a polygon file, which accomplishes the raster to vector conversion. Because no filtering was done to the image to weed out pixel areas too small or large to be considered a structure of interest (i.e., a house), a filter-like operation was performed using ArcInfo. A new attribute was added to the polygon attribute table (.pat) and a value of 1 or 0 was attached depending on the perimeter size of the polygon. The polygons that did not meet the criteria were given an attribute value of 1. Part of the ArcInfo operation of building the topology for its polygon coverage files included identifying a label point. A command was used to move the label point to the center of the polygon. Next the point was extracted to create a point file. The final product consisted of a point file indicating the center of each identified structure.

(Note: The specific commands used in this step are listed in Appendix B.)

3.2.6 Step 6

The initial comparison (Table 1) used image interpretation to manually assess whether the procedure was identifying homes. This assessment was performed by overlaying the point file on the original image and by using aerial photography to identify correct house locations along with commission and omission of locations. The final comparison was performed by both Coler & Colantonio and by Brown (results are shown in Table 2).

Coler & Colantonio manually created a point file identifying house locations interpreted from an orthophoto image. This point file and the one created at Brown from the ATLAS imagery were then compared.

3.3 Why Not SMA?

Spectral mixture analysis was not applied in this image processing procedure. However, evidence shows that the matched filter is explicitly equivalent to SMA (Settle, 1996, and Tu et al., 1997). The difference in the two approaches is a constant difference in their magnitudes. SMA can be viewed as an a posteriori version of the matched filter (Chang, 1998).
3.4 Why Not PCA?

The motivation for this project was to sort structures (houses) from backgrounds (such as vegetation, roads, and other interference). It is assumed that a structure has a spectral feature distinct from its background or interference; i.e., has a distinct variance contribution to the total variance of the image. If noise levels are the same for all image bands, the structure may be sorted out according to the order of the variance of the transformed image bands. The variance reflects the SNR, which in this case is image quality. However, the distinct variance contribution from the structure may be overwhelmed by the noise level in another band because of an unequal noise level in the different bands. Thus in the PCA results, the principal component with large variance may not show good quality (i.e., can be very noisy) but may include the useful signal as mentioned in Section 2.5.2.2. MNF can deal with this problem by adjusting the noise in different bands into the same level. In the final MNF component, the PCA with large variance corresponds to the PCA with the greater SNR (i.e., good quality image). Thus by ordering images in terms of SNR, a structure can be sorted out from its background or interference.

4.0 Observations

4.1 Observations Based on Images

4.1.1 Wareham, Massachusetts

The Wareham image (Figure A-5) was used as a testing image to help Lin Li develop the methodology best applicable to this dataset and to test its feasibility. The only observation made was that a methodology based on spectral information could be developed and that structures were identifiable using image analysis techniques.

4.1.2 Brown University

The Brown image (Figure A-6) was also used for testing while the real areas of interest were chosen. This image was chosen as a test because it had so many structures and because it is a highly urban area. The image contained an older city neighborhood with university athletic and education buildings, roads, full-grown trees, and residences. Although no counting was performed, it was obvious to Lin Li that this area was not ideal for the methodology developed. Depending on how the methods were applied, either too many of the image features were being identified as building structures (commission) or many of the structures were not being identified (omission).

4.1.3 Cape Cod, Massachusetts

This image (Figure A-2) was chosen as one of the areas of interest by Coler & Colantonio because of present transmission corridors. This was not a highly urban image, but it did contain some neighborhoods, both old and new, along with some roads and a canal. Unfortunately, the methods were not very successful when applied to this image. The main problem was distinguishing between roads and building structures in the image. The cause of
this problem was determined to be that these manmade these features were not spectrally distinct. If there had been repeat coverage of this area at different times within a day, there may have been thermal changes that would have helped to distinguish these features spectrally.

4.1.4 Narragansett, Rhode Island

This image (Figure A-3) produced adequate results. It appeared that most buildings were being identified even though the ground resolution was 7.5 meters. The results were also varied because of the solar illumination across the image. It was found that the accuracy variations across the image could be overcome by dividing the image into sections of similar illuminations (3 sections). Also, the large body of water in the image provided an excellent source to build the covariance matrix in Step 1.

The images shown in Figure A-7, Figure A-8, and Figure A-9 display a portion of the results using the Narragansett image.

4.1.5 Barnstable, Massachusetts

This image (Figure A-4) contained a coastal wetland, mature vegetation, major and minor roadways, and old and new residential development. Of all the images tested, this one appeared to show the best results - clear buildings identified within the imagery. The Narragansett image also showed clear buildings; however, the ground resolution in this image is smaller, and therefore it was determined that the results may be more accurate. Also, the utilized area of this image did not have strong illumination effects across the imagery to affect the results.

Table 1 displays the first results of analysis of a random area in the image, which was performed to determine whether the structures of interest (homes) were being identified. This test was performed on the three wavelength combinations: VISNIR (visible to near infrared), thermal, and VISNIR + thermal (all ATLAS bands). (Within these combinations, different thresholds were used to create the final binary images as discussed in Section 3.2.3 Step 3. The lower values in each category indicate a greater numerator value in the equation (mean + N*(STD)). N may equal 2.5 or 3). An explanation of each specific file can be found in Appendix C. The final results determined from this procedure were as follows:

1) The structures could be identified.
2) The number of structures correctly identified decreased as the threshold value increased.
3) The most accurate results were determined using both the VISNIR and thermal data.
Table 1. Initial accuracy comparison using Barnstable image.

<table>
<thead>
<tr>
<th>Wavelength Combinations</th>
<th>File</th>
<th># of Structures in Airphoto</th>
<th># of Structures Correctly Identified</th>
<th># of Structures Falsely Identified</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>VISNIR</td>
<td>F</td>
<td>279</td>
<td>198</td>
<td>34</td>
<td>71%</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>279</td>
<td>159</td>
<td>24</td>
<td>57%</td>
</tr>
<tr>
<td>Thermal</td>
<td>C</td>
<td>279</td>
<td>242</td>
<td>108</td>
<td>87%</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>279</td>
<td>211</td>
<td>86</td>
<td>76%</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>279</td>
<td>168</td>
<td>47</td>
<td>60%</td>
</tr>
<tr>
<td>VISNIR+ Thermal</td>
<td>B</td>
<td>279</td>
<td>252</td>
<td>86</td>
<td>90%</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>279</td>
<td>215</td>
<td>82</td>
<td>77%</td>
</tr>
</tbody>
</table>

The File column represents the differing thresholds used to create the final image and is described in Appendix C.

Table 2 shows the results from our final, more detailed comparison. This comparison involved using the point files derived from the binary images and the point file manually created from the orthorectified aerial photography obtained by Coler & Colantonio. Because the initial study (Table 1) showed that the best results were obtained using all ATLAS bands, only this set of data was used. Therefore only a comparison of this combination was used. Because of the imperfect georeferencing of the ATLAS imagery to the orthophotography and the possibility that the center of each structure was not identified similarly, six distances were used to measure the accuracy of the structures identified in the imagery. Areas 1 and 2 were random areas used for testing.
Table 2. Final accuracy comparison.

<table>
<thead>
<tr>
<th>Distance</th>
<th>Structures C&amp;C Found &amp; Brown Did Not Find</th>
<th>Structures Brown Located Incorrectly</th>
<th>Structures Brown Identified Correctly</th>
<th>% Identified Correctly</th>
<th>Total Brown Structures</th>
<th>Total C&amp;C Structures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area1 (File A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>within 1'</td>
<td>105</td>
<td>139</td>
<td>0</td>
<td>0</td>
<td>139</td>
<td>105</td>
</tr>
<tr>
<td>within 5'</td>
<td>104</td>
<td>138</td>
<td>1</td>
<td>1</td>
<td>139</td>
<td>105</td>
</tr>
<tr>
<td>within 10'</td>
<td>103</td>
<td>137</td>
<td>2</td>
<td>2</td>
<td>139</td>
<td>105</td>
</tr>
<tr>
<td>within 15'</td>
<td>101</td>
<td>135</td>
<td>4</td>
<td>4</td>
<td>139</td>
<td>105</td>
</tr>
<tr>
<td>within 20'</td>
<td>98</td>
<td>132</td>
<td>7</td>
<td>7</td>
<td>139</td>
<td>105</td>
</tr>
<tr>
<td>within 50'</td>
<td>45</td>
<td>79</td>
<td>60</td>
<td>57</td>
<td>139</td>
<td>105</td>
</tr>
<tr>
<td>Area2 (File A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>within 1'</td>
<td>226</td>
<td>292</td>
<td>0</td>
<td>0</td>
<td>292</td>
<td>226</td>
</tr>
<tr>
<td>within 5'</td>
<td>225</td>
<td>291</td>
<td>1</td>
<td>0</td>
<td>292</td>
<td>226</td>
</tr>
<tr>
<td>within 10'</td>
<td>222</td>
<td>288</td>
<td>4</td>
<td>2</td>
<td>292</td>
<td>226</td>
</tr>
<tr>
<td>within 15'</td>
<td>219</td>
<td>285</td>
<td>7</td>
<td>3</td>
<td>292</td>
<td>226</td>
</tr>
<tr>
<td>within 20'</td>
<td>213</td>
<td>279</td>
<td>13</td>
<td>6</td>
<td>292</td>
<td>226</td>
</tr>
<tr>
<td>within 50'</td>
<td>99</td>
<td>165</td>
<td>127</td>
<td>56</td>
<td>292</td>
<td>226</td>
</tr>
<tr>
<td>Area1 (File B)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>within 1'</td>
<td>105</td>
<td>220</td>
<td>0</td>
<td>0</td>
<td>220</td>
<td>105</td>
</tr>
<tr>
<td>within 5'</td>
<td>101</td>
<td>216</td>
<td>4</td>
<td>4</td>
<td>220</td>
<td>105</td>
</tr>
<tr>
<td>within 10'</td>
<td>98</td>
<td>213</td>
<td>7</td>
<td>7</td>
<td>220</td>
<td>105</td>
</tr>
<tr>
<td>within 15'</td>
<td>98</td>
<td>213</td>
<td>7</td>
<td>7</td>
<td>220</td>
<td>105</td>
</tr>
<tr>
<td>within 20'</td>
<td>93</td>
<td>208</td>
<td>12</td>
<td>11</td>
<td>220</td>
<td>105</td>
</tr>
<tr>
<td>within 50'</td>
<td>34</td>
<td>149</td>
<td>71</td>
<td>68</td>
<td>220</td>
<td>105</td>
</tr>
<tr>
<td>Area2 (File B)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>within 1'</td>
<td>226</td>
<td>291</td>
<td>0</td>
<td>0</td>
<td>291</td>
<td>226</td>
</tr>
<tr>
<td>within 5'</td>
<td>226</td>
<td>291</td>
<td>0</td>
<td>0</td>
<td>291</td>
<td>226</td>
</tr>
<tr>
<td>within 10'</td>
<td>220</td>
<td>285</td>
<td>6</td>
<td>3</td>
<td>291</td>
<td>226</td>
</tr>
<tr>
<td>within 15'</td>
<td>219</td>
<td>284</td>
<td>7</td>
<td>3</td>
<td>291</td>
<td>226</td>
</tr>
<tr>
<td>within 20'</td>
<td>210</td>
<td>275</td>
<td>16</td>
<td>7</td>
<td>291</td>
<td>226</td>
</tr>
<tr>
<td>within 50'</td>
<td>91</td>
<td>156</td>
<td>135</td>
<td>60</td>
<td>291</td>
<td>226</td>
</tr>
</tbody>
</table>
4.2 Other Observations

The combination of both thermal and multispectral data types provided the best results, and the accuracy heavily relied on the use of thermal data. Limits on the accuracy were due to there not being enough spectral differences or thermal difference detected between some structures, making them difficult to distinguish or causing false identification.

Identified limitations of using only spectral information included the following:

1) When interference (roads) has a similar spectral response as the structures (houses), it is difficult to remove the roads from the house image.

2) Unknown sources having spectral features similar to the house causes false identification.

3) One house may be identified with more than one point in the final data product, causing overestimation. This error could be eliminated if a filter were developed to remove or combine locations within a specific radius.

The time of day may have been an issue with some of the imagery. If the imagery were acquired in an earlier part of the morning, there might be larger differences spectrally and thermally between manmade features within the imagery. Multiple acquisitions within a day may be required to provide more spectral information.

There were difficulties in distinguishing roads from structures. This problem can be alleviated by incorporating road vectors in the model when identifying the structure center points. One could identify and remove all points that intersect or are within a set tolerance of the road vectors.

Furthermore, the poor identification of the structures using MNF may result from the following:

1) The image noise isn't additive, but multiplicative. If this is the case, one should try to perform a log operation before applying MNF and check whether the results have improved.

2) When one obtains the covariance matrix of noise from an area of vegetation, the obtained covariance of the noise may not result from the noise only but may include the variability of vegetation reflectance. It is also possible that the signal (house) and background (vegetation) are highly correlated, thus the MNF does not work.

5.0 Metrics and Conclusions

Flight time for acquisition of ATLAS data costs $3,200 per hour. Ideally, the sensor can collect approximately 230 miles (200 nautical miles) of data each hour, resulting in a minimum data collection cost of $14/mile. Excluded from this price are the costs of travel for the plane's crew and the cost of ferrying to the locations of interest, since these costs may
change with time. However, of the estimated $240/mile cost of the current methods (see Section 2.1), $200/mile is the cost of the aerial photography. The ATLAS cost of $14 verses $200 for aerial photography suggests great savings in the raw cost of data per mile.

Although this project was not entirely successful, methods may still be developed to work with remotely sensed data to provide savings on the current cost of $240/mile. One of the purposes for proceeding with this project was to identify whether the IKONOS satellite imagery could be used with the defined methods. The present IKONOS satellite has one panchromatic band with a 0.82-meter ground resolution and four multispectral bands with a 4-meter ground resolution. To successfully apply the methods identified in this project, it was determined that thermal imagery needed to be used. Therefore, IKONOS imagery does not provide enough spectral information.

The IKONOS data products provided through Space Imaging are only guaranteed to be as spatially precise as 4 meters at a cost of $170/mile (see Figure D-1), which is still a savings from the $200/mile for the aerial photography.

Some positive aspects identified from this project were as follows:

1) The methods developed to use the ATLAS imagery work very well in areas where there is relatively sparse density of structures (houses).

2) The combination of spectra-based and morphology-based methodology may improve the result since it is very evident that the structure shapes are very different from the background (roads, etc.).

3) A spectral database of typical materials found on the roofs of the structures and on the surfaces of the roads may help to identify the proper endmembers, which could improve the result.

Some suggestions for Coler & Colantonio's future use of digital remotely sensed data for identifying structures along transmission lines are as follows:

1) The data may have a greater potential use for conceptual route layout.

2) The data provides a practical, efficient way to obtain a first-order estimate of house numbers in large areas.

3) Use of multiple years may help to identify possible growth when determining areas of change and therefore investigation.

4) The developed methods could be used to identify roads.

The proposal indicated that for these methods to be commercially successful, the structures would need to be correctly identified 95% of the time. This level of accuracy was not achieved. However, this project has been defined as a positive step toward understanding how remotely sensed imagery can provide information more quickly than traditional methods. Additionally, although this method requires constant user interaction and input, it still may be faster than the original methods.
6.0 References


Space Imaging Catalog of Products and Services, Sept 1999. Volume 1 Supplement, Copyright © 2000 by Space Imaging, Thornton, Colorado, 15 pp..


Appendix A. Overview of ATLAS Images

ATLAS
(Airborne Terrestrial Applications Sensor)

<table>
<thead>
<tr>
<th></th>
<th>0.45 - 0.52</th>
<th>0.52 - 0.59</th>
<th>0.58 - 0.68</th>
<th>0.63 - 0.68</th>
<th>0.68 - 0.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>0.76 - 0.89</td>
<td>1.51 - 1.76</td>
<td>2.07 - 2.38</td>
<td></td>
<td>8.20 - 8.52</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>12</td>
<td></td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>13</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td>11.39 - 12.02</td>
</tr>
</tbody>
</table>

Electromagnetic Spectrum

Figure A-1. ATLAS wavelength band descriptions.
Figure A-2. Cape Cod ATLAS image. Bands 7, 5, 4 combination.
Figure A-3. Narragansett ATLAS image. Bands 7, 5, 4 combination.
Figure A-4. Barnstable ATLAS image. Bands 7, 5, 4 combination.
Figure A-5. Wareham ATLAS image. Bands 7, 5, 3 combination.
Figure A-6. Brown ATLAS image. Bands 7, 5, 4 combination.
Figure A-7. Clip of Narragansett image, Bands 7, 5, 4.

Figure A-8. Narragansett image from Step 2, identifying mainly building structures.
Figure A-9. Narragansett binary image identifying building structures.
Appendix B. Development of GIS Coverage Files

(Raster to Vector to Points)

Export tiff and tifw files for import into ARC/Info. These files should be binary.

>ARC

Usage: CREATEWORKSPACE <workspace>
- may need to create a new workspace
- will create a directory within the directory it was launched in
Arc: createworkspace newworkspace
Arc: w
  - w = workspace
  - will list the directory you're working in (like a pwd in unix)
Arc: w ..
  - will move you up one directory
Arc: w newworkspace
  - will move you into the workspace created

Usage: KILL <geo_dataset> {ARC | INFO | ALL}
- the best way to remove any files created with ARC/Info
Arc: kill file all

Usage: IMAGEGRID <in_image> <out_grid> {out_colormap_file} {in_band}
  {NEAREST | BILINEAR | CUBIC} {DEFAULT | SQUARE}
- used to input tiff file, will read tifw automatically
Arc: imagegrid inimage.tif outimage

Usage: GRIDPOLY <in_grid> <out_cover> {weed_tolerance}
  - to convert the grid to a polygon file
  - raster to vector conversion
Arc: gridpoly outimage outpoly

Usage: BUILD <cover> {POLY | LINE | POINT | NODE | ANNO.<subclass>}
  - builds topology, assigning labels to polygons, to node, from node, etc.
  - may not have to do this step
Arc: build outpoly

Usage: CREATELABELS <cover> {id_base}
  - creates labes for polygons if this was not done using the gridpoly command
  - may not have to do this step
Arc: createlabels outpoly
Usage: PROJECTDEFINE <COVER | GRID | FILE | TIN> <target>
- allows user to define the coordinate system that applies to the coordinates

Arc: projectdefine cover outpoly
Define Projection
Project: project utm
Project: units meters
Project: datum wgs84
Project: zone 19
Project: parameters

Usage: DESCRIBE <geo_dataset>
- will give information about the file, # of polygons, # of points, Geo-coordinates, if there have been recent changes, etc.
Arc: describe outpoly

Usage: CENTROIDLABELS <cover> {OUTSIDE | INSIDE}
- moves the label points into the center of the polygons
Arc: centroidlabels outpoly inside

Usage: ADDITEM <in_info_file> <out_info_file> <item_name> <item_width> <output_width> <item_type> {decimal_places} {start_item}
- Add a new attribute to help in editing out large/small polygons.
Arc: additem outpoly.pat outpoly.pat CODE 2 2 i

Open INFO, to edit attribute tables. When within INFO all typing must be in CAPITALS.

Arc: info
ENTER USER NAME> arc
ENTER COMMAND > SELECT OUTPOLY.PAT
ENTER COMMAND > ITEM
- shows all attributes labels, check to see if CODE is there
- step not necessary
ENTER COMMAND > LIST
- lists all features and attributes
- step not necessary
ENTER COMMAND > RESELECT FOR PERIMETER LE 30
- reselects all features with these attributes within the file that was selected
ENTER COMMAND > CALC CODE = 1
- calculated the codes of the reselected features to equal 1
ENTER COMMAND > NSELECT
- selects all other features that weren't selected before in the selected file
ENTER COMMAND > RESELECT FOR PERIMETER GT 200
- reselects features in the presently selected group
ENTER COMMAND > CALC CODE = 1
ENTER COMMAND > Q STOP

Now all polygons we didn't want have the CODE = 1 value.

Arc: build outpoly
- need to rebuild topology once the attributes have been altered
Usage: CREATE <out_cover> {tic_bnd_cover}
- creates a new file with the old file extents (uses the .tic)

Arc: create outpoly newpolyfile

Arc: arcedit
- allows us to edit the features and view them

Usage: DISPLAY <device> {option}
{SIZE {FRAME | CANVAS} <width> <height>}
{POSITION <xy>}
{TOP}
Usage: DISPLAY <device> {option}
{SIZE {FRAME | CANVAS} <width> <height>}
{POSITION <xy | UL | UC | UR | CL | CC | CR | LL | LC | LR>
{SCREEN | THREAD <thread>}
{xy | UL | UC | UR | CL | CC | CR | LL | LC | LR}
{TOP}
Usage: DISPLAY <device> {option}
{SIZE {FRAME | CANVAS} <width> <height>}
{POSITION <ABOVE | BELOW | LEFT | RIGHT | CENTER>
{SCREEN | THREAD <thread>}}
{TOP}
Usage: DISPLAY COLOMAP <total_colors> {static_colors}
Usage: DISPLAY COLOMAP DEFAULT
- we need to define the display device as being the computer terminal your on

Arcedit: display 9999
Arcedit: mape outpoly
- need to define the mapping extents as being that of the coverage extents

Usage: EDIT <cover> {feature_class}
Usage: EDIT <info_file> INFO
- specify the coverage to be edited

Arcedit: ec outpoly

Usage: EDITFEATURE <NONE | ARC | NODE | LABEL | TIC | LINK | POLYGON |
ANNO{.subclass} | SECTION.subclass | ROUTE.subclass |
REGION.subclass | GROUP.subclass>
- select the features we want to edit

Arcedit: ef label
Usage: DRAWENVIRONMENT {edit_cover {REVERT}}
ALL {ON | OFF}
ARC {ON | OFF | IDS | ARROWS | INTERSECT}
NODE {ON | OFF | IDS | ERRORS | DANGLE | PSEUDO}
LABEL {ON | OFF | IDS}
TIC {ON | OFF | IDS}
ANNO{.subclass} {ON | OFF | level...level}
LINK {ON | OFF}
POLYGON {ON | OFF | IDS | FILL}
REGION.subclass {ON | OFF | IDS | FILL}
SECTION.subclass {ON | OFF | IDS | ARROWS |
MEASURES | POSITIONS}
ROUTE.subclass {ON | OFF | IDS | ARROWS |
ROUTEERRORS | MEASUREERRORS}
GROUP.subclass {ON | OFF}
- define the drawing environment
Arcedit: de label
Arcedit: draw
- will the draw the labels (or selected features and de)
Arcedit: select CODE ne 1
- selects all features where CODE is not equal to 1, therefore we want to keep it
Arcedit: drawselect
- will redraw the selected features, in a highlighted color (yellow)

Usage: PUT <cover>
- will put the selected features into the new file
Arcedit: put newpolyfile
Appendix C. Binary File Creation

**VISNIR + Thermal: ATLAS bands 1-8, 10-15**

This procedure included running Matched Filter and MNF. The resultant images were narrowed down to include image 1, which highlighted roads and buildings best, and image 4, which highlighted mostly roads.

**File A**
Using only the Image 1.
Model: retain all values greater than the mean plus three times the standard deviation
> (mean) + 3 * (std)

**File B**
Using image 1 and 4.
Model: image 1 minus all values less than the mean in image 4, the mean was 99

**Thermal: ATLAS Bands 10-15**

This procedure did not involve using MNF after Matched Filter. The resultant images involved using three images in the model to create the final binary image. Image 1 emphasized houses and roads; in image 2 the houses were brighter than the roads, and in image 3 the roads were most clearly defined.
Model: if (%1 >= #) and (%2 >= #) and (%3 <= #) then; %4 = 255; else; %4 = 0; endif; (%4 = final binary image)

**File C**
For this model the # was 2 times the standard deviation plus the mean.
# = (mean) + 2 * (std)

**File D**
For this model the # was 2.5 times the standard deviation plus the mean.
# = (mean) + 2.5 * (std)

**File E**
For this model the # was 3 times the standard deviation plus the mean.
# = (mean) + 3 * (std)

**VISNIR: ATLAS Multispectral Bands 1-8**

This procedure included running Matched Filter and MNF. Three images were chosen for the final model to create the binary image.
Model: if (%1 >= #) and (%2 >= #) and (%3 >= #) then; %4 = 255; else; %4 = 0; endif;
**File F**
For this model the # was 2 times the standard deviation plus the mean.

\[ # = (\text{mean}) + 2 \times (\text{std}) \]

**File G**
For this model the # was 2.5 times the standard deviation plus the mean.

\[ # = (\text{mean}) + 2.5 \times (\text{std}) \]
Appendix D. IKONOS Pricing

Figure D-1. Pricing list for the IKONOS imagery. Obtained from the Space Imaging Catalog of Products and Services, September 1999, Volume 1 Supplement.