High-level Data Fusion Software for SHOALS-1000TH
FY07 Annual Report

Grady H. Tuell
Optech International
7225 Stennis Airport Drive
Suite 400
Kiln, MS 39556
Phone: (228) 252-1004 FAX: (228) 252-1007 E-mail: gradyt@optechint.com

Steven E. Lohrenz
University of Southern Mississippi
Hydrographic Science Research Center
Department of Marine Science
Stennis Space Center, MS 39529
Phone: (228) 688-3720 FAX: (228) 688-1121 E-mail: steve.lohrenz@usm.edu

Award Number: *W912HZ-05-C-0061*
http://www.optechint.com ; http://www.marine.usm.edu

LONG-TERM GOALS

Optech International and the Department of Marine Science at the University of Southern Mississippi have partnered to develop and apply data fusion techniques for application to environmental mapping problems in shallow-water and coastal environments. This work has lead to: (1) a robust collaboration between industry and academia focused on the use of airborne remote sensing technologies for near-shore and coastal analysis; (2) the emergence of a data fusion paradigm wherein a larger community of researchers can understand and eventually contribute to this work; (3) dissemination of appropriate data sets into the larger community; and to (4) improved understanding of regional-scale coastal environmental processes which will be realized thorough the structured analysis of combined airborne lidar and passive spectral data.

OBJECTIVES

The objectives of this collaboration are: (1) develop a number of new data fusion algorithms and computer programs to produce coastal and environmental information from SHOALS-1000TH data; (2) achieve increased accuracy of environmental information extracted from SHOALS-1000TH data by collection and application of *in situ* oceanographic ground truth; (3) facilitate the transfer of knowledge from academia to industry regarding the collection and use of *in situ* optical data; (4) facilitate the transfer of knowledge from industry to academia related to airborne laser technology and sensor and data fusion; and (5) facilitate the education of future researchers and workforce through the integration of these concepts into graduate level courses at the University of Southern Mississippi.

APPROACH AND WORK PLAN

During this reporting period, we conducted two field data collection campaigns: in Looe Key, Florida, from November13-17, 2006; and in Thunder Bay, Michigan, from Aug 18- 20, 2007. Optech International and USM worked closely with the Joint Airborne Lidar Technical Center of Expertise (JALBTCX) to identify a location for these data collection campaigns, and coordinated resources to collect the airborne and *in situ* data. Optech International and JALBTCX operated airborne systems for CASI hyperspectral and SHOALS data, and provided personnel, expertise, and equipment for the collection of *in situ* data.

In year 1 of this project, we developed and refined the underlying fusion paradigm (the SIT data fusion model), and created software to combine lidar and passive spectral data for seafloor and beach classification, and automated extraction and attribution of shorelines. For classification, identity declarations were achieved using a supervised technique, the maximum likelihood classifier. A second, higher level technique was developed using Dempster-Shafer evidential theory, and this approach was applied to pixels and blobs in the spatial domain.

During this reporting period, we refined the operational software for these data processing steps, and implemented a new higher level fusion algorithm, a rule-based approach, using third party software to automate the generation of decision rules. We implemented this rule-based approach for the pixel level case, and developed implementation strategies to extend it to blob-level fusion. We extended existing and new data fusion concepts to produce a consistent strategy for bathymetric and topographic classifications, and we developed strategies and collected data to support rigorous accuracy assessments of the resulting classifications.

For the upcoming year, we will refine our algorithms and accuracy assessments using new data, and produce operational code to implement blob-level, rule-based classification. The resulting classification will represent higher level classification on the spatial and information and technique axes of the SIT data fusion model. We will also submit research papers from this work for peer-review and publication.

WORK COMPLETED

During fiscal year 2007, our work consisted of:

1. Conducted field data collection campaign in Looe Key, FL, in Nov. 17, 2006.
2. Conducted field data collection campaign in Thunder Bay, MI, in Aug 25- 27, 2007.
3. Created seafloor classification using blob-level data fusion based on a decision algorithm, Dempster-Shafer evidential theory, and conducted accuracy assessment.
4. Generated a proposition Look-Up Table (LUT) for automatic Dempster-Shafer fusion processing.
5. Created seafloor classification using pixel-level, rule-based algorithms and used third party free-ware to support the automation of decision rules and processing of decision tree.
6. Created topographic classification at the blob-level using automatic decision trees processing and mean-shift algorithm.

1. Field Data Collection

We collected airborne optical imagery and infrared (topographic) laser data for shoreline mapping in Bahia Honda, Florida Keys on November, 2006, using a SHOALS-1000 and a CASI-2 spectral imager.

Unfortunately, no green lidar data were collected over the land area. A sample of the passive spectral data from the Casi-2 is shown in Figure 1.

Figure 1. CASI hyperspectral image mosaic of Bahia Honda, Looe Key, FL, November 17, 2007.

We conducted a second data collection campaign in Thunder Bay, Lake Huron, Michigan, from August 25 to 27, 2007, using a SHOALS-1000 lidar and a CASI-1500 passive spectral imager. This survey covered the lakeside town of Alpena, Michigan, and the shoreline of Lake Huron. Additionally a small set of ground reflectance signatures were collected on August 27 (sand, grass, asphalt). In Figure 2, we show the SHOALS green reflectance image for this dataset.

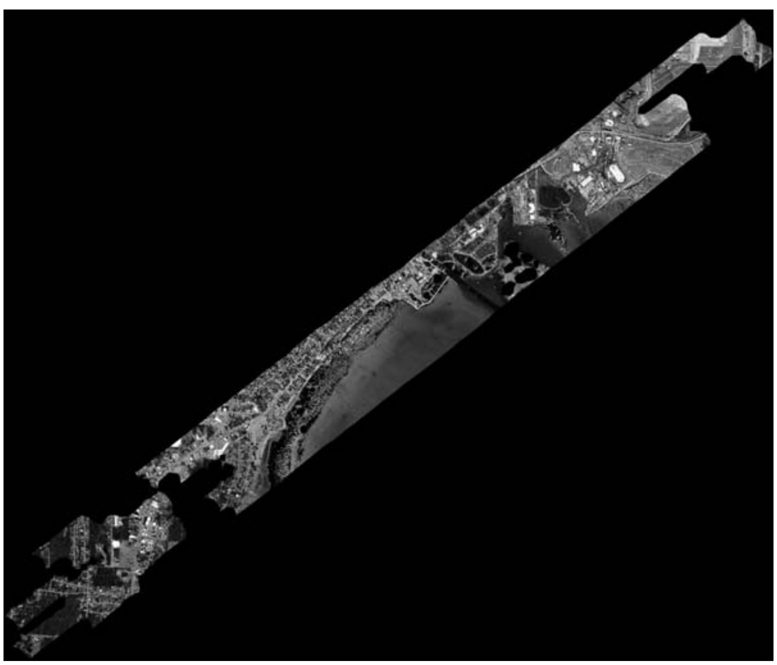


Figure 2. SHOALS green laser reflectance image of the eastern part of Alpena, Michigan, and the shoreline of Thunder Bay, Lake Huron.

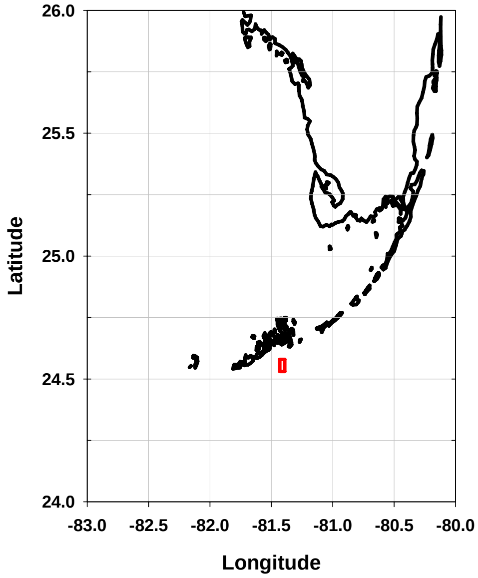
We collected *in situ* data for water column properties and bottom types during these campaigns. During the Looe Key data collection, we experienced difficulties with some instrumentation in obtaining high quality *in situ* measurements in shallow water. From this experience, we refined our field collection instrumentation and protocol and successfully achieved more stable *in situ* data in Thunder Bay in August, 2007.

In the Looe key area shown in Figure 3 (a), high temperature and very clear water pushed the equipment to its limit. Filters on the ac-9 did not perform well, possibly due to insufficient pressure to eliminate bubbles. At some locations, K_d could not be computed due to tilting of the MicroPro profiler or boat motion. To overcome these problems, we conducted extensive testing of the instrumentation prior to the subsequent deployment in Michigan. This effort included field testing of the profiling package, including the bb9, ac-9, ac-s, and SBE49, at a boat launch site on the Pearl River outside the Stennis Space Center South Gate on July 2007. This exercise led us to adjust our techniques such that ac-9 and ac-s measurements yielded comparable results. A modified field protocol was developed and used in subsequent field operations.

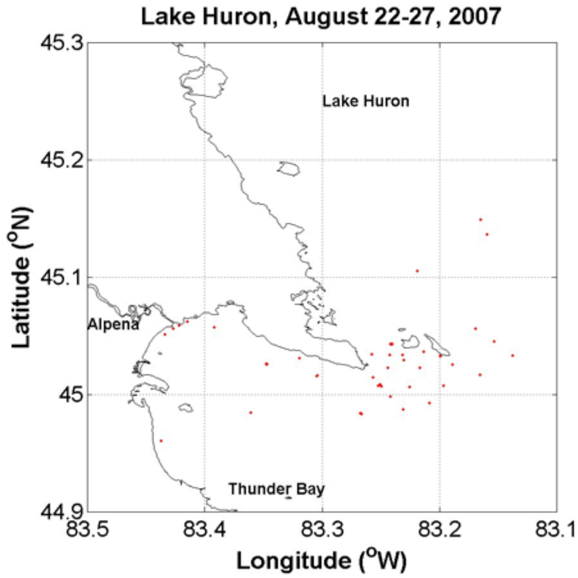
Other work to prepare for the Lake Huron field campaign included refining the field data collection protocol to include additional measurements including: water sampling for suspended particulate matter (SPM), HPLC pigment concentrations, chromophoric dissolved organic matter (CDOM), and the quantitative filter pad technique to determine particulate absorption.

In Fig. 3 (b) we show the location of *in situ* data collection site for the Lake Huron campaign. Due to our a priori lack of knowledge about large scale bottom type distributions and spatial and seasonal water quality, choices of the *in situ* data collection sites were limited to several ship-wreck sites based on discussions with researchers at the NOAA Thunder Bay Laboratory. In this area, the seafloor is relatively uncomplicated. Bottom types are mostly sand and cobble with small, scattered portions of macrophytes, silt and clay.

Water constituent measurements (IOPs) and radiance and irradiance measurements (AOPs) were coordinated to the greatest extent possible. However, depending on the field situations, some of the measurements were omitted. For example, when the water is very shallow (less than 3-4 meters), the profiling radiometers were not used. The vertical profiling package included an ac-9 absorption/attenuation meter, ac-s absorption/attenuation meter, ECO bb9 backscattering sensor, ECO Triplet (FL3) fluorescence sensor (WET Labs, Inc.), and SBE49 CTD sensor (SeaBird, Inc.). Integration of the data acquisition was achieved using a DH-4 data logger (WET Labs, Inc.). The instruments were mounted on a single frame to provide simultaneous data collection. MicroPro II and HyperPro II profiling radiometers (Satlantic) were also used to measure subsurface upwelling radiance and downwelling irradiance. At some stations, Polycap TC 0.2 µm filters (Fisher Scientific) were placed on the ac-9 meter to provide a measure of the dissolved phase absorption and attenuation. For selected stations, water samples were collected and returned to shore for dock-side measurements with the ac-s meter. This approach was used because it was found that the ac-s meter was unstable during profiling due to ship motion. Other analyses performed on water samples included particle size distribution measurements with a LISST 100X (Sequoia, Inc.), and analyses for SPM concentration, HPLC pigment analysis, CDOM and filter pad particulate absorption. A list of measurements is given in Table 1.



(a) Looe Key area, Florida Key, FL



(b) Lake Huron area, Michigan

Figure 3. Station map for Nov. 2006 (a) and Aug. 2007 (b).

Measurements	Processed data	Number of Samples
ac-9	a & c, scattering and no-scattering corrected cast files	26
ac-s	a & c scattering and no-scattering corrected cast files a & c scattering and no-scattering corrected dock-side files	26 47
BB9	back scattering at 9 wavelengths	41
FL3	chl a, phycoerythrin, and CDOM concentrations	37
LISST	particle size distribution, % optical transmission, c(670), and LISST measurements	24
MicroPro II	Level-2 & 4 files	15
HyperPro II	Level-3 & 4 files	21
SPM	SPM concentration	25
CDOM	CDOM absorption spectrum	25
HPLC	HPLC pigment concentrations	6
Filter Pad	absorption spectrum files	6

Table 1. Pre-processed *in situ* data from Lake Huron campaign

2. Development of High-Level Data Fusion Algorithms for Bathymetric Lidar and Spectral Data

A central goal of this project is to develop operational approaches to seafloor classification wherein we progress from measured data at differing spatial scales to knowledge about spatial objects. There is no unique way to achieve this goal, and different strategies will produce differing results. During the past year, we refined the SIT data fusion paradigm and began using it as a framework for comparing strategies. We have also adopted the use of traditional classification accuracy assessments as metrics for judging which of several competing strategies is the best.

For example, in Figure 4 we illustrate the use of the SIT model to produce the seafloor classification shown in Figure 5(d) for a section of the seafloor near Fort Lauderdale, Florida. The text on the right side of the figure describes the step-by-step computations leading from the raw measured data to knowledge about the seafloor. Significantly, the application of a connected neighbor algorithm in steps 11 and 12 produces spatial blobs, rather than individual pixels. In the final step, we apply set theoretic techniques to produce a second, higher-level identity declaration on these blobs (compared to the declarations made at steps 5 and 10 from the individual sensor data streams as shown in Figure 5 (a) and (b), or to a pixel-level combination of them as shown in Figure 5(c)).

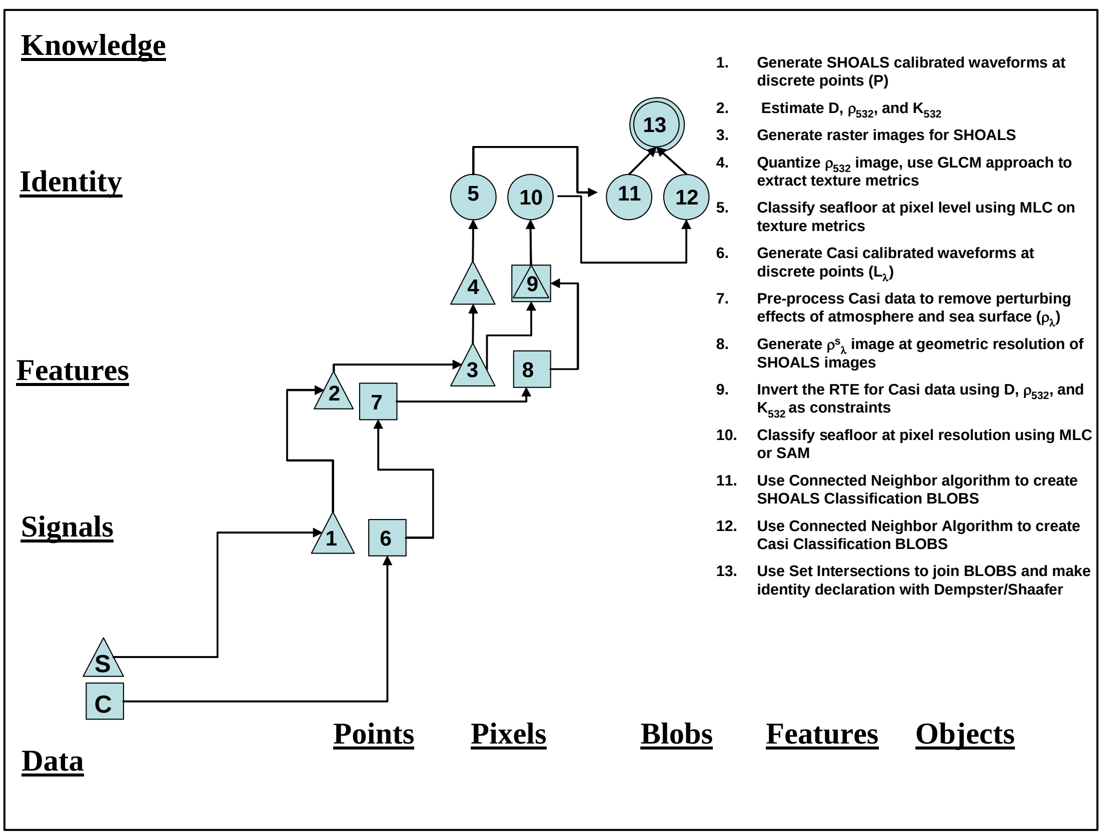


Figure 4. Use of SIT model to illustrate high-level data fusion classification of the seafloor.

(a) CASI Seafloor Reflectance MLC

Overall accuracy : 69.22%

(b) SHOALS MLC

Overall accuracy : 84.92%

(c) Dempster-Shafer Pixel level

Overall accuracy : 85.37 %

(d) Dempster-Shafer Blob level classification

Overall accuracy : **89.25%**

Sand Hard Bottom Inner Reef OuterReef

Figure 5. Classification images and overall accuracies for seafloor near Fort Lauderdale, Florida.

In Figure 5, we note the highest accuracy is achieved in the highest-level classification process (step 13 in Figure 4). We had hoped to generate precisely this result: increasing levels of abstraction on both the information axis and spatial axes in the SIT model lead to improved knowledge about the spatial environment.

However, the classification achieved at step 13 is not perfect. Inspection of Figure 5(d) reveals significant misclassifications in very shallow water. This error probably arises from a cascade of problems at multiple levels, but one possible improvement is to replace the decision algorithm at step 13. This replacement can be understood as a change in the overall processing flow within the technique axis of the SIT data fusion paradigm. The present algorithm at step 13, a Dempster-Shafer algorithm, applies Bayesian statistics to individual classifications produced at steps 5 and 10, and uses a maximum likelihood classifier to produce the required a priori probabilities at those two steps. In this reporting period we have explored replacing the Dempster-Shafer algorithm with a rule-based decision algorithm. In Year 1 we had successfully used a rule-based approach for topographic classification.

The largest challenge in implementing a rule-based approach is in generating the decision rules. Other researchers have created programs to generate rules automatically, and C4.5 is an example of third party software that can be used to generate rules from a decision tree. During this year, we investigated the use of this program for rule generation, and conducted classifications of the same dataset shown in Figure 5.

In Figure 6 we show two pixel level classification images produced with a rule-based approach. Figure 6 (a) was produced using 9 channels of seafloor reflectance data from the CASI system, and 16 information feature images produced from the SHOALS data. Figure 6 (b) was generated from 9 channels of seafloor reflectance from the CASI system, seafloor reflectance from SHOALS. We have not yet completed accuracy assessments on these images but will complete this work early in the next phase of research. We will also extend this approach to the blob level on the information axis and use the mean shift algorithm as a blob generator.

3. DEVELOPMENT OF DATA FUSION ALGORITHMS FOR TOPOGRAPHIC LIDAR AND SPECTRAL DATA

In Year 1, we applied automated decision tree algorithm for land-cover classification and used C4.5 for this purpose. In this reporting period, we replaced C4.5 with the commercial version, See5. See5 works on any machine but it must have an internet connection because the processing to compute the rules is done on the company’s server. The principle advantage to this approach is fast processing speed because the only file uploaded to server is the training data, a relatively small file.

After settling on a technique for executing the decision tree algorithm in ENVI’s program language IDL, and investigating a way for the user to specify training locations based on ROIs, we completed automated generation of decision rules to accomplish a pixel-level fusion for the topographic case. We then extended this approach to the blob-level case using the mean shift algorithm to generate the blobs. For example, we show in Figure 7(a), 3848 blobs generated on a hyperspectral image and in 7(b) the blob-level classification image. In the next phase of research we will test this approach will multiple datasets and make the rule-generation step more convenient by designing a user interface.

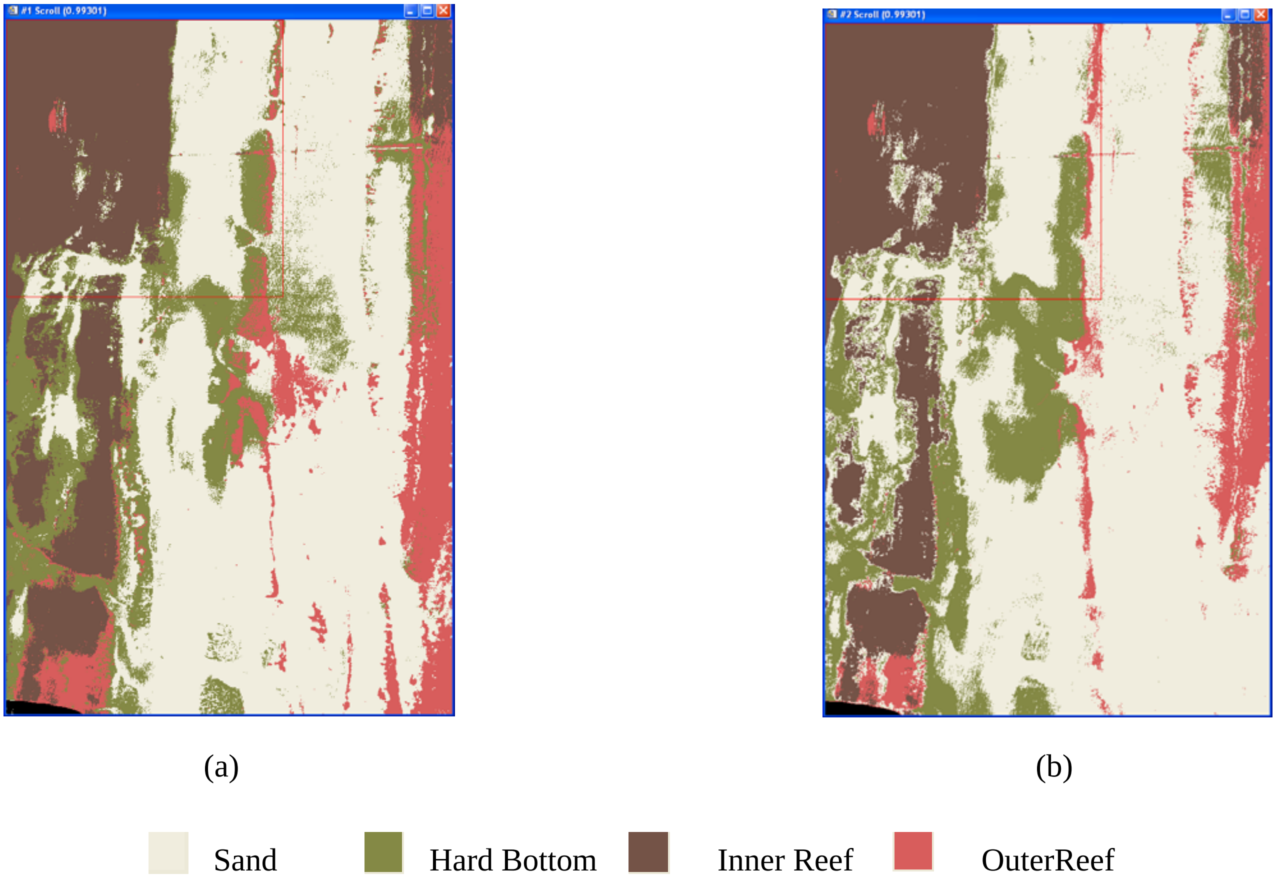
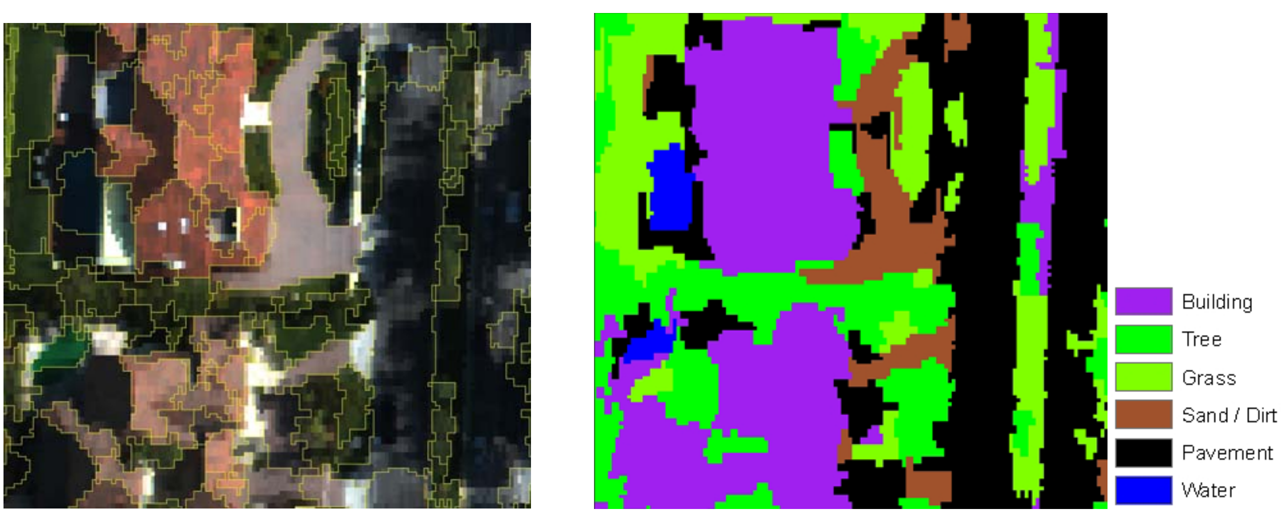


Figure 6. Classification using automatic decision tree program. (a) classification using 9 seafloor reflectance channels from CASI and 16 features from SHOALS, (b) classification from 9 CASI seafloor reflectance images and SHOALS seafloor reflectance.



(a) blobs on hyperspectral image (b) classification of blobs

Figure 7. Blob-level automatic rule-based classification for topographic case.

RESULTS

We collected CASI, SHOALS, and *in situ* data in Looe Key, Florida, and in Thunder Bay, Michigan. We shared these data within the team and with other interested researchers. Using the *in situ* data, the radiometric inversion algorithms for both the CASI and SHOALS systems were tested and refined.

We implemented higher-level fusion procedures in both the spatial and the information contexts, and extended the SIT model to accommodate a new level, knowledge, representing a state where conflicts between two identity declarations have been resolved.

We demonstrated, for a single dataset, an increase in overall classification accuracy arising from the application of higher-level fusion techniques.

We developed an early version of an automated Dempster-Shafer fusion program using built-in or user defined look-up tables to generate propositions.

We implemented a classification strategy using automatically generated decision rules to produce a rule-based classification for the bathymetric case, at the pixel level.

For the topographic case, we successfully implemented rule-based classification at the blob level.

IMPACT AND APPLICATIONS

National Security

This project has the potential to produce changes in how shallow waters and the coastal environment are mapped and monitored. One possible spin-off is that the active/passive data fusion approach will prove to be an effective way to find small objects on the beach and in shallow water.

Economic Development

This project will develop new tools for the commercial application of airborne remote sensing for mapping and monitoring the littoral zone. Also, the collaboration between industry and academia will facilitate development of a workforce educated in the use of these technologies. Together, these two impacts should improve the utility of this approach and increase the demand for active/passive systems.

Quality of Life

Data fusion techniques developed in this NOPP project will be of immediate value to the U.S. Army Corps of Engineers and NOAA for the regional-scale mapping and monitoring of coastal ecosystems. Ultimately, algorithms for airborne data will be improved through the use of *in situ* data. This work will also lead to the development of new algorithms for the mapping and monitoring of spatial distributions of water column constituents, and the development of a commercial capability for this purpose.

Science Education and Communication

At present, no university in the world has a program of instruction which covers bathymetric lidar in detail. This NOPP partnership strengthens the commitment made by Optech International to assist USM in developing this capability, and USM's commitment to become an academic leader in this area. Aspects of this work will be integrated into the curricula for the graduate Marine and Hydrographic programs at USM, and will be the focus of work for one or more graduate students. This, in turn, will lead to the presentation of these results at conferences and the publication of peer-reviewed papers.

TRANSITIONS

Quality of Life

As part of JALBTCX, the U.S. Army Corps of Engineers and NOAA presently operate an airborne system capable of simultaneously measuring airborne active and passive data. The algorithms and software developed in this NOPP project will improve their ability to combine, process, and analyze these data for regional-scale coastal mapping applications.

Science Education and Communication

The data fusion paradigm developed in this partnership, and the datasets generated, will be available to all interested researchers. Because data fusion is an active area of research at many universities, we expect these data to be used in a number of graduate theses and other research projects.

RELATED PROJECTS

Subsequent to the award of this NOPP project, Optech International received a contract from the U.S. Office of Naval Research (ONR) to use bathymetric lidar combined with passive spectral data for seafloor classification as it relates to the detection of mines. This project is named: Counter-mine Lidar UAV-based System (CLUBS). The two projects are distinct, but have some commonality. For example, the data fusion paradigm developed in the NOPP project has been used to formulate, present, and quantify the work in CLUBS, and to produce operational software for rapid environmental assessment. The operational software produced in CLUBS is called the Optech Rapid Environmental Assessment (REA) Processor. USM will serve as a sub-contractor to Optech International on CLUBS. Also subsequent to the award of this NOPP project, Optech International and USM were awarded, as partners, a contract to design and build the Coastal Zone Mapping and Imaging Lidar (CZMIL). CZMIL has adopted the SIT data fusion model from this NOPP project, and has based the software design on the algorithms developed in both NOPP and CLUBS.