Advanced Active Learning Methods for Robust Classification of Multi-Source Remotely Sensed Data

Presenting Author: James C. Tilton, NASA GSFC

Co-Authors: Yuhang Zhang*, H. Lexie Yang**, Edoardo Pasolli**, Zhou Zhang**, Melba M. Crawford** and Saurabh Prasad**

* University of Houston, **: Purdue University

OUTLINE

1. Introduction and motivation
2. Methodology
   - Multi-source active learning (AL) methods : Multiple kernel-AL (MKL-AL) and Ensemble MKL-AL
   - Hierarchical Segmentation (HSeg)
3. Testbed datasets
   - Experimental results on testbed datasets
4. Ongoing and future work
5. Publications

1. INTRODUCTION

Overall Goal

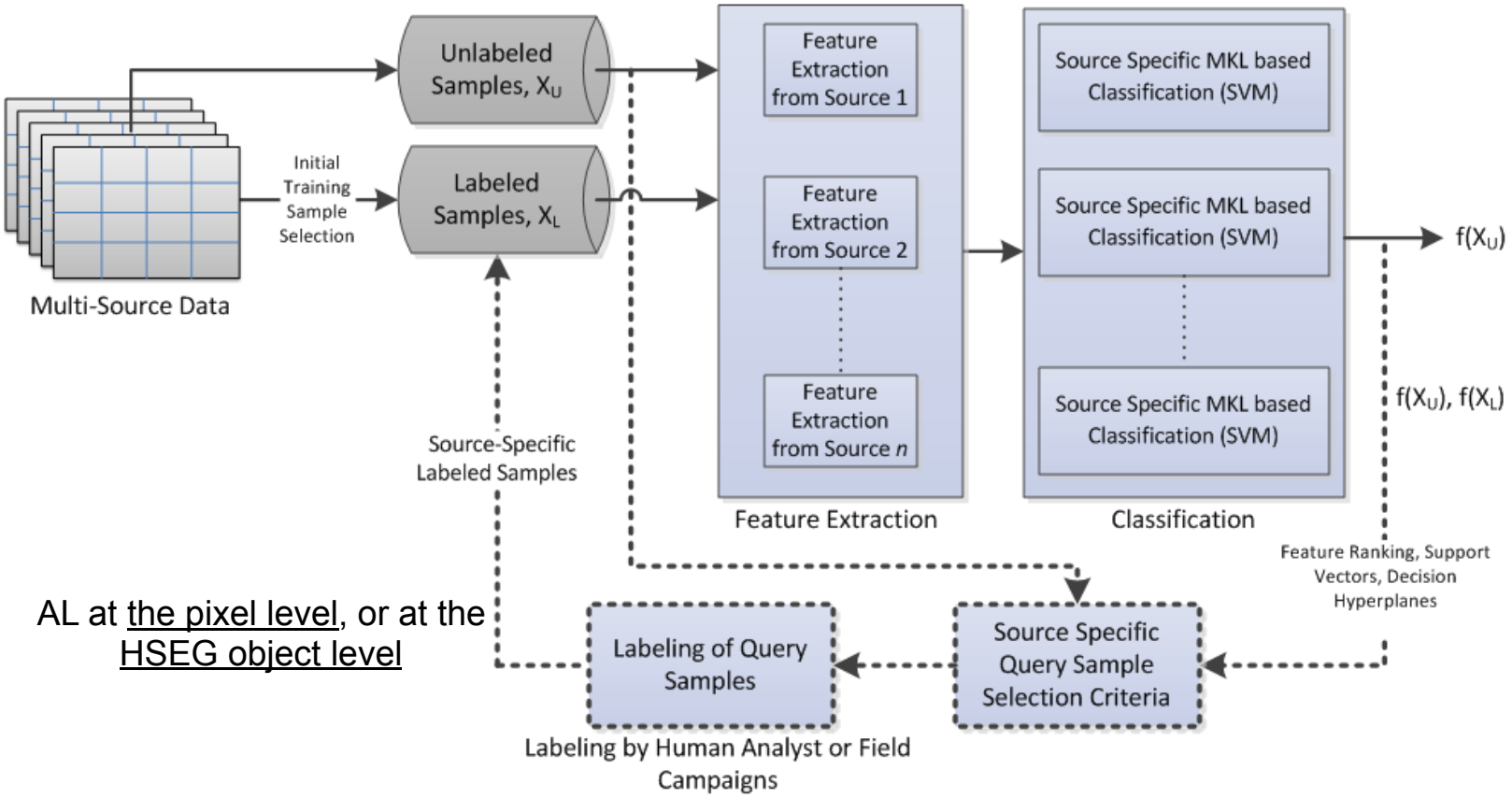
- To develop a robust classifier framework that can be applied in a multi-sensor environment using an Active Learning strategy to select samples for labeling.

Motivation for Research

- Hyperspectral and multi-sensor data require large quantities of labeled data (which are difficult and expensive to collect) to train supervised classifiers, motivating “intelligent” use of unlabeled samples.
- Large scale data remote sensing data sets exhibit variable characteristics over space and time, motivating development of adaptive classifiers.
- Traditional labeled training sets typically contain redundant samples, motivating use of Active Learning to select the best subsets.
- Optimal exploitation of multi-source data continues to be an open problem.

OBJECTIVES

- **Overall Objective:** Develop a multi-source active learning framework for geospatial data analysis. Case studies to include spatial-spectral feature extraction and analysis from hyperspectral imagery, as well as multi-sensor active learning – utilizing hyperspectral imagery and LiDAR data.



2. Methodology

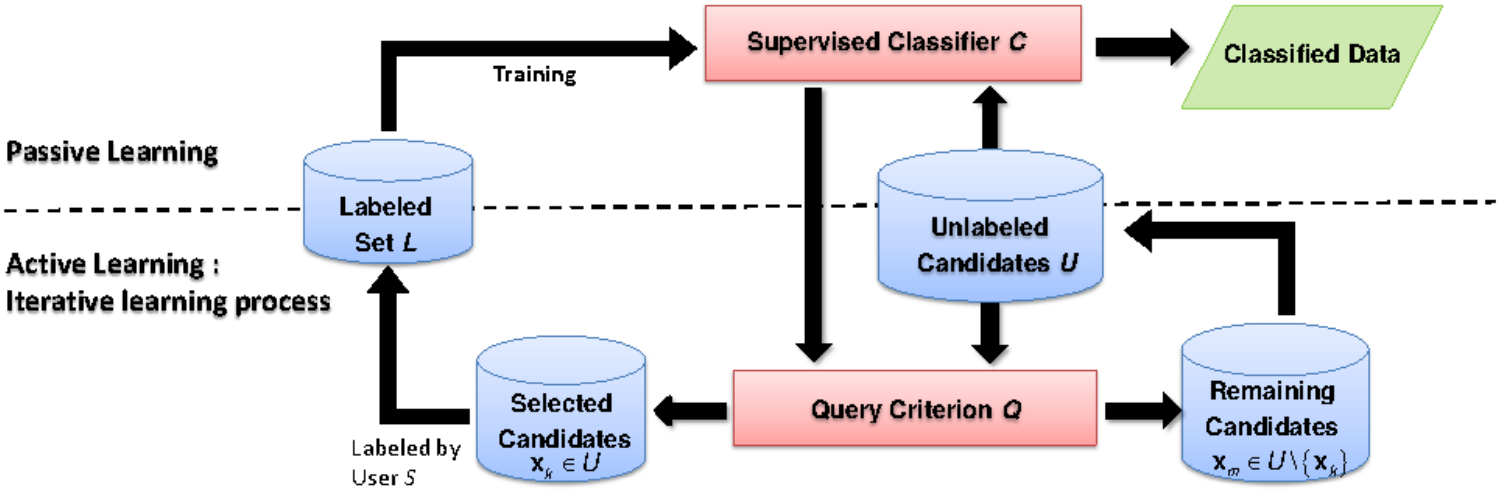
MULTI SOURCE ACTIVE LEARNING

- Introduction to Active Learning

- Proposed multi-kernel learning active learning (MKL-AL) method for multi-source image classification

BACKGROUND: ACTIVE LEARNING (AL)

- AL selects samples from the unlabeled data pool in a biased manner via query strategies that are designed to exploit properties of the classifier and the current labeled and auxiliary unlabeled data.
- Goal of AL: Obtain satisfactory classification performance with fewer labeled samples than those of conventional passive learning, where the training set is often selected randomly or manually without interaction with the classifier.

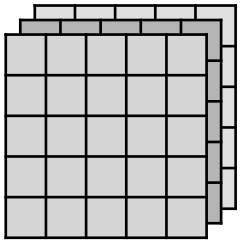


Multi-Kernel Learning (MKL) for Multi-Source Image Analysis

- Traditional SVM classifiers employ single-kernel models
  - Tuning is performed to obtain “optimal” kernel parameters via cross-validation.

- Multi-Kernel Learning (MKL) is based on a linear mixture of kernels
  - MKL can implicitly adapt the kernel to the data by learning appropriate weights of pre-determined kernels, eliminating the need to re-tune SVMs at each AL induction step.
  - Kernel alignment can further be employed to determine a good bank of kernels.
  - MKL is being investigated in single and ensemble classification frameworks for multi-source data analysis.

Flow Chart: MKL-AL Algorithm



Multi-source data

Construct kernels

$K_1^1(\mathbf{x}_i^1,\mathbf{x}_j^1),\ldots,K_m^1(\mathbf{x}_i^1,\mathbf{x}_j^1)$

$K_1^2(\mathbf{x}_i^2,\mathbf{x}_j^2),\ldots,K_m^2(\mathbf{x}_i^2,\mathbf{x}_j^2)$

$\vdots$

$K_1^p(\mathbf{x}_i^p,\mathbf{x}_j^p),\ldots,K_m^p(\mathbf{x}_i^p,\mathbf{x}_j^p)$

Solve optimal weights

$$K(\mathbf{x}_i,\mathbf{x}_j)=\sum_{m=1}^{M}\sum_{p=1}^{P}d_m^p K_m^p(\mathbf{x}_i^m,\mathbf{x}_j^m)$$

SVM classifier

Classified data

Labeled set L

Unlabeled candidates U

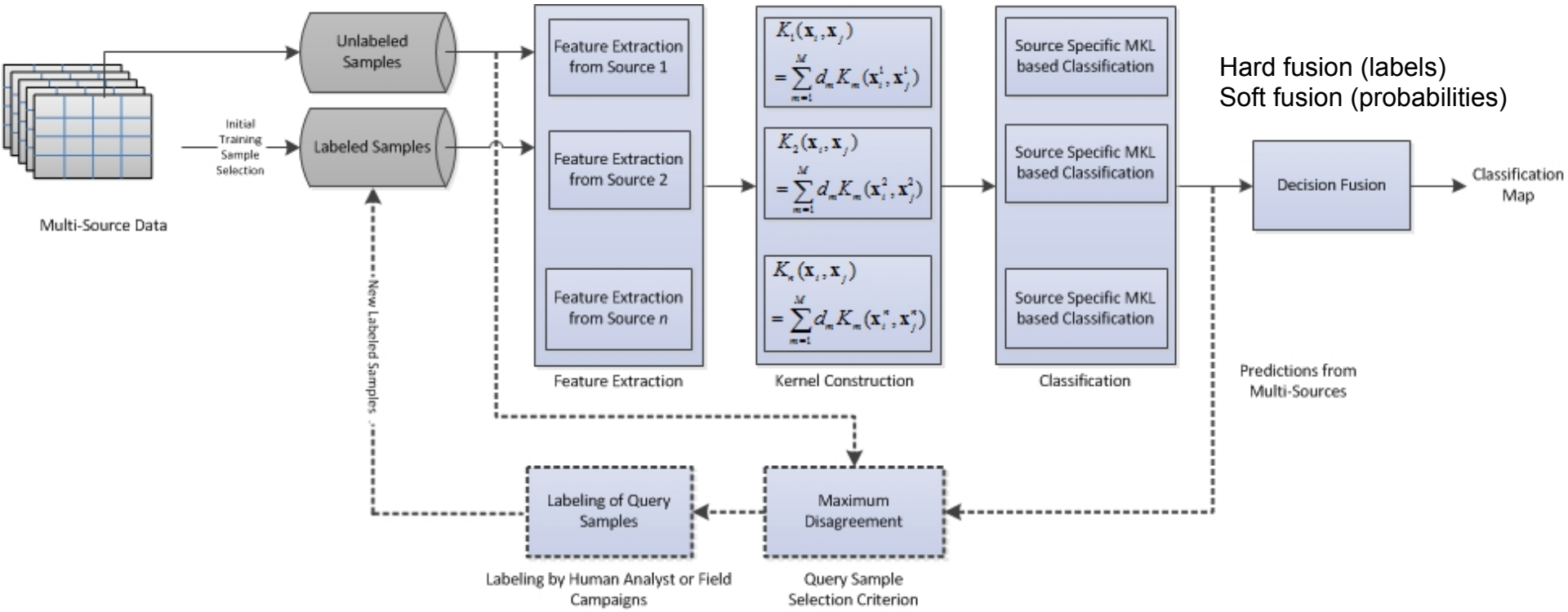
Select samples with smallest distance to hyperplane

Selected candidates $\mathbf{x}_k \in U$

Remaining candidates $U \setminus \{\mathbf{x}_k\}$

Labeled by user S

ENSEMBLE MKL-AL



HSeg Background:

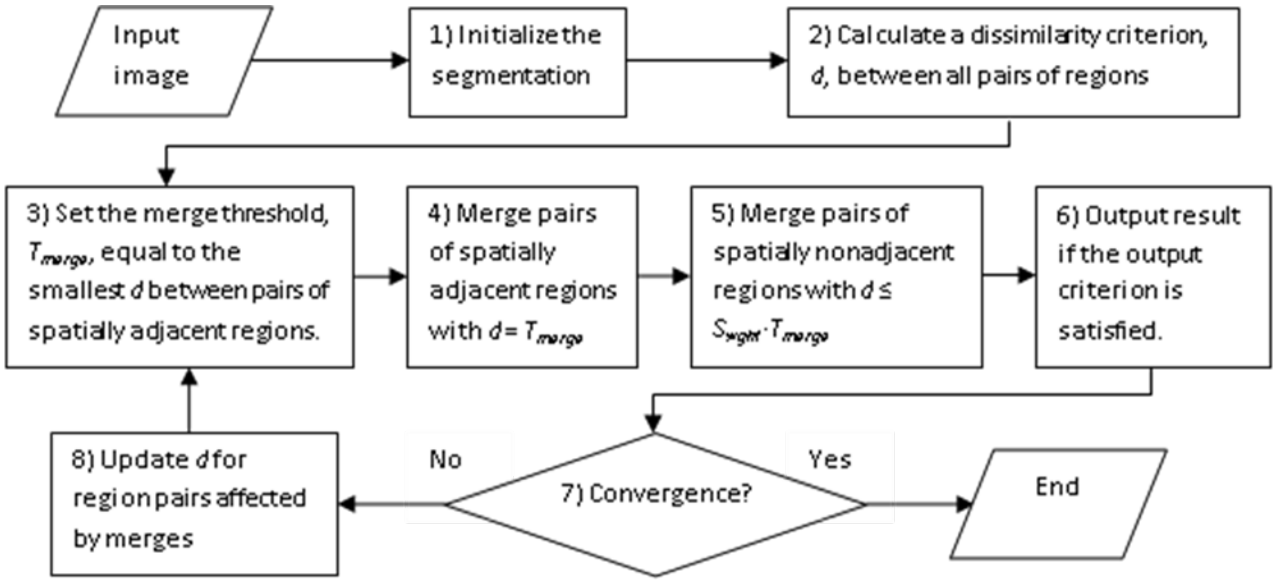
- HSeg performs image segmentation through a form of best merge region growing.

- Algorithm based on the basic hierarchical step-wise optimization approach (HSWO) described in: J.-M. Beaulieu and M. Goldberg, "Hierarchy in picture segmentation: A stepwise optimal approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 2, pp. 150-163, Feb. 1989.

- HSWO finds individual closed-connected region objects.

HSeg Background (cont'd):

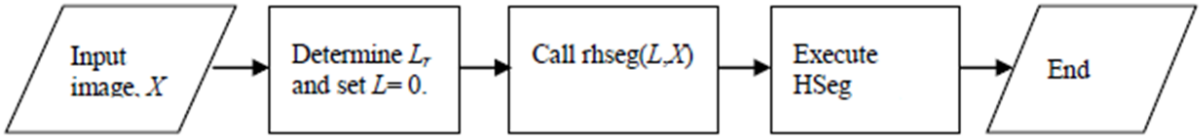
- HSeg modifies HSWO by also aggregating spectrally similar but spatially separated region objects into groups of region objects – called region classes. The HSeg Flowchart:



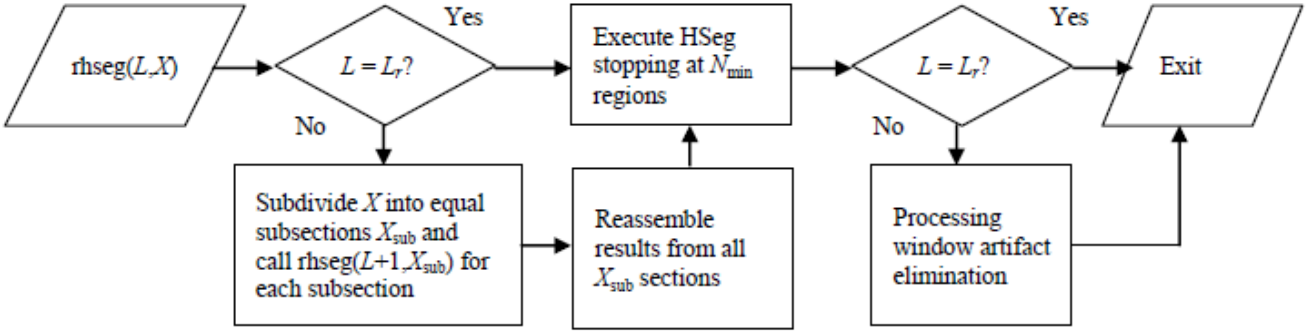
- S_{wght}, ranging from 0 to 1, controls the relative importance of merges between adjacent regions versus non-adjacent regions.

HSeg Background (cont'd):

- The RHSeg approximation of HSeg has an efficient parallel implementation useful for processing large images:



- L_r is determined as the number of times the input image must be subdivided to achieve a small enough image size for efficient processing with HSeg.
- The rhseg(L,X) function:



- N_{min} is equal to ¼ the number of pixels in the subimage processed at the deepest level of recursion.

HSeg Background (cont'd):

Both HSWO and HSeg produce a hierarchical set of image segmentations that:

1. Consist of segmentations at different levels of detail, in which
2. The coarser segmentations can be produced from merges of regions from the finer segmentations, and
3. The region boundaries are maintained at the full image spatial resolution

The HSeg algorithm is fully described in:

J. C. Tilton, Y. Tarabalka, P. M. Montesano and E. G., "Best Merge Region Growing Segmentation with Integrated Non-Adjacent Region Object Aggregation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4454-4467, Nov. 2012.

- This is version 1.59 of HSeg/RHSeg.

Problem Noted with HSeg/RHSeg version 1.59

Problem:

Large homogeneous regions with gradual gradients aren't readily formed.

Observation:

The boundaries between HSeg (or HSWO) subregions of large homogeneous regions do not correspond to any visibly apparent boundary – There is no "edge" between these subregions.

Idea:

Can edge information be utilized to influence the HSWO/HSeg region growing process to encourage the merging together of large homogeneous regions with gradual gradients?

Frei-Chen Difference Operator

The Frei-Chen Edge Difference Operator consists of nine convolution masks:

$$G_1 = \frac{1}{2\sqrt{2}}\begin{bmatrix} 1 & \sqrt{2} & 1 \\ 0 & 0 & 0 \\ -1 & -\sqrt{2} & -1 \end{bmatrix} \quad G_2 = \frac{1}{2\sqrt{2}}\begin{bmatrix} 1 & 0 & -1 \\ \sqrt{2} & 0 & -\sqrt{2} \\ 1 & 0 & -1 \end{bmatrix} \quad G_3 = \frac{1}{2\sqrt{2}}\begin{bmatrix} 0 & -1 & \sqrt{2} \\ 1 & 0 & -1 \\ -\sqrt{2} & 1 & 0 \end{bmatrix}$$

$$G_4 = \frac{1}{2\sqrt{2}}\begin{bmatrix} \sqrt{2} & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & -\sqrt{2} \end{bmatrix} \quad G_5 = \frac{1}{2}\begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \quad G_6 = \frac{1}{2}\begin{bmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

$$G_7 = \frac{1}{6}\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix} \quad G_8 = \frac{1}{6}\begin{bmatrix} -2 & 1 & -2 \\ 1 & 4 & 1 \\ -2 & 1 & -2 \end{bmatrix} \quad G_9 = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

The edge value (EV) is computed from the nine convolution masks as follows:

EV = $\sqrt{M/S}$, where $M = \sum_{k=1}^{4} (G_k)^2$ and $S = \sum_{k=1}^{9} (G_k)^2$

INCORPORATION OF EDGE INFORMATION INTO HSEG

- Edge information is incorporated at three different stages:

1. An initialization stage in which the edge information directs a fast first-merge region growing process to quickly merge connected areas with edge values <= edge_threshold (set by user), and

2. The normal HSWO/HSeg best merge region growing stage in which the edge information influences the best merge decisions.

3. In performing processing window artifact elimination in RHSeg: merge pairs of regions across the processing window seams that have low edge values along the seams.

Incorporation of Edge Information Into HSeg

A number of approaches were investigated for having the edge information influence the best merge decisions in the HSWO/HSeg best merge region growing stage. A simple extension of HSeg was found to be quite effective:

HSeg version 1.61: Add to the original HSeg one region feature value, E_{max}, the maximum value of E_{value} for all pixels in the region. The edge dissimilarity value, E_{dissim}, is assigned as the maximum of E_{max} for the two regions being compared.

Normalize the value of E_{dissim} to range from 0.0 to 1.0, by computing

$$E\downarrow dissim\uparrow' = (E\downarrow dissim - \min\nolimits_{\top} I\,[E\downarrow value\,]\,)/(\max\nolimits_{\top} I\,[\,E\downarrow value\,] - \min\nolimits_{\top} I\,[E\downarrow value\,]\,)$$

where $\min\nolimits_{\top} I\,[E\downarrow value\,]$ is the minimum value of E_{value} over the entire image, I, and $\max\nolimits_{\top} I\,[E\downarrow value\,]$ is the maximum value of E_{value} over the entire image.

ESTO Earth Science Technology Office

INCORPORATION OF EDGE INFORMATION INTO HSEG

Compute an edge factor, E_f, as follows:

$$E_f = (S_{wght} + (1.0 - S_{wght})((1.0 - E_w) + E{\downarrow}dissim{\uparrow}' E_w))/S_{wght}$$

where E_w as a user settable parameter that controls the weighting of the edge feature (ranging from 0.0 to 1.0), and S_{wght} is the HSeg "spectral clustering weight."

The combined region dissimilarity is the computed as $C_{dissim} = R_{dissim}{}^*E_f$, where R_{dissim} is the dissimilarity between the region pair for the original version of HSeg.

Thus, an adjacent region is treated as a non-adjacent region for $E{\downarrow}dissim{\uparrow}' = 1.0$, and treated normally as an adjacent region for $E{\downarrow}dissim{\uparrow}' = 0.0$, with gradations in-between for $0.0 < E{\downarrow}dissim{\uparrow}' < 1.0$.

Incorporation of Edge Information Into HSeg

- Frei-Chen edge difference operator result:

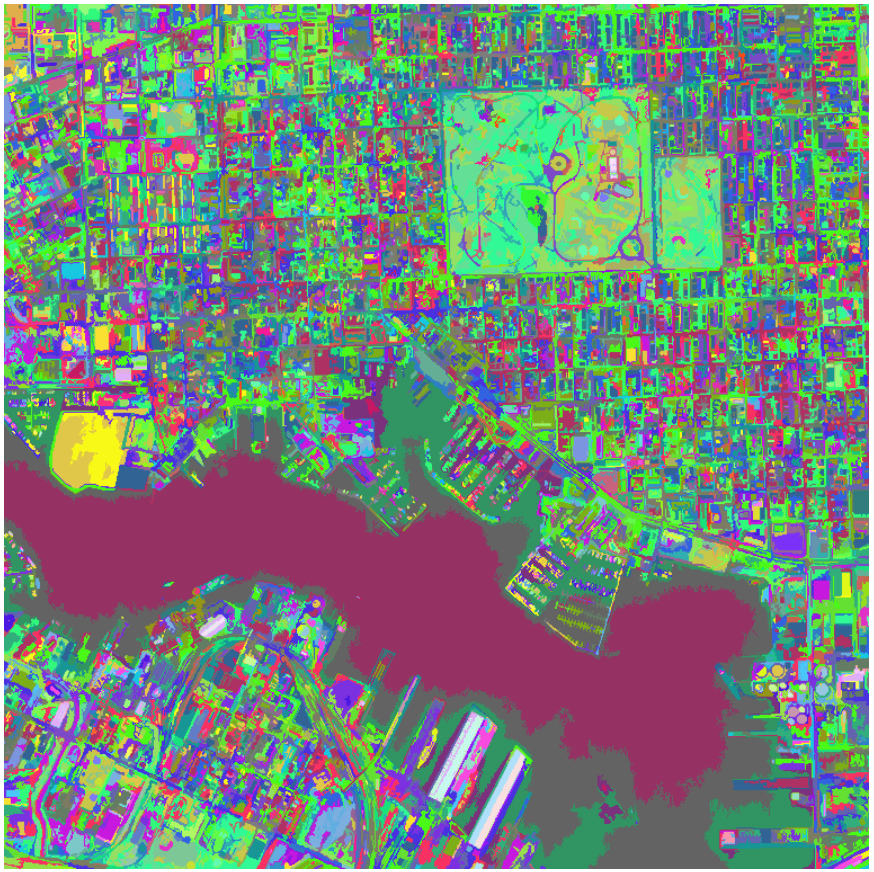
Ikonos data, 768x768 pixels, Patterson Park/Inner Harbor area of Baltimore, MD.

Frei-Chen edge difference operator result, maximum over spectral bands, thresholded at 0.07.

Incorporation of Edge Information into HSeg

- HSeg v. 1.59 result:

**Ikonos data, 768x768 pixels,
Patterson Park/Inner Harbor area of Baltimore, MD.**

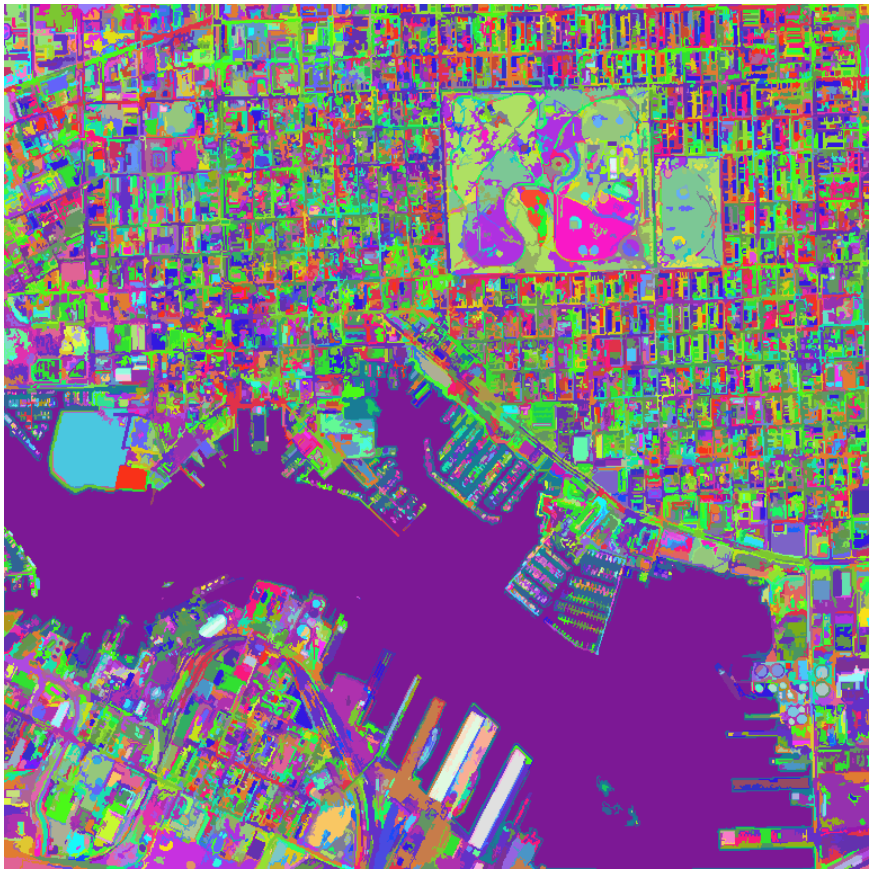


**HSeg v. 1.59 result,
at global dissimilarity 0.371,
155 region classes and 9871 region objects.**

INCORPORATION OF EDGE INFORMATION INTO HSEG

- HSeg v. 1.61 result – edge initialization only:

Ikonos data, 768x768 pixels, Patterson Park/Inner Harbor area of Baltimore, MD.

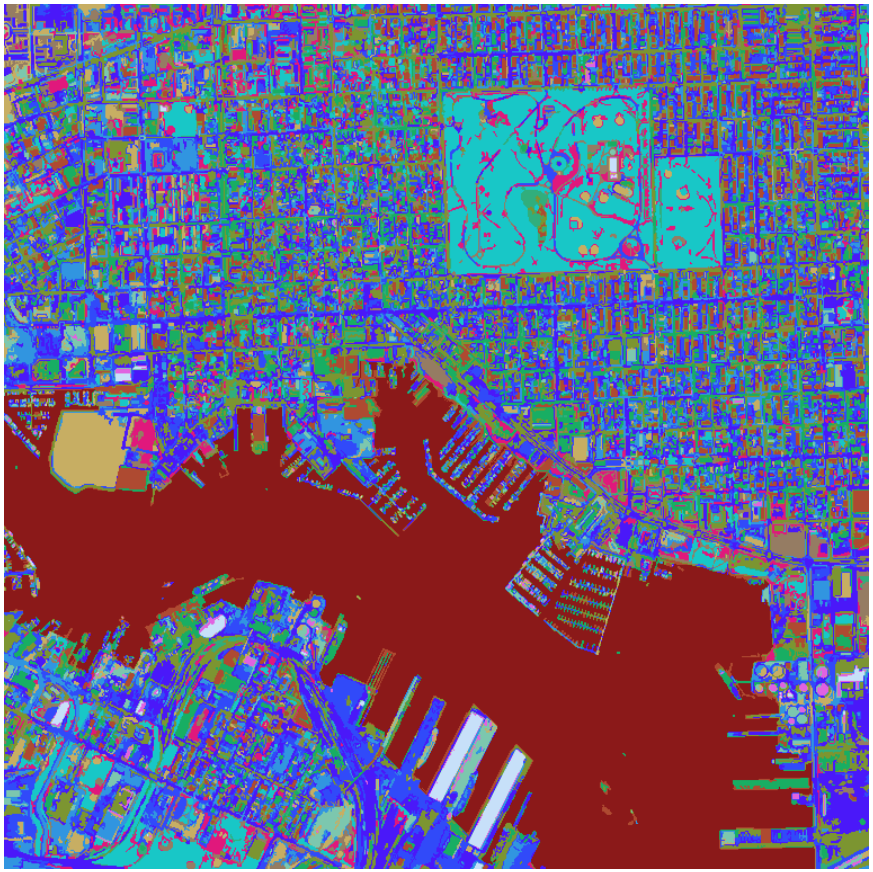


HSeg v. 1.61 result, E_t=0.05 and E_w=0.0, at global dissimilarity 0.371, 192 region classes and 9954 region objects.

INCORPORATION OF EDGE INFORMATION INTO HSEG

- HSeg v. 1.61 result – E_w=1.0:

Ikonos data, 768x768 pixels, Patterson Park/Inner Harbor area of Baltimore, MD.



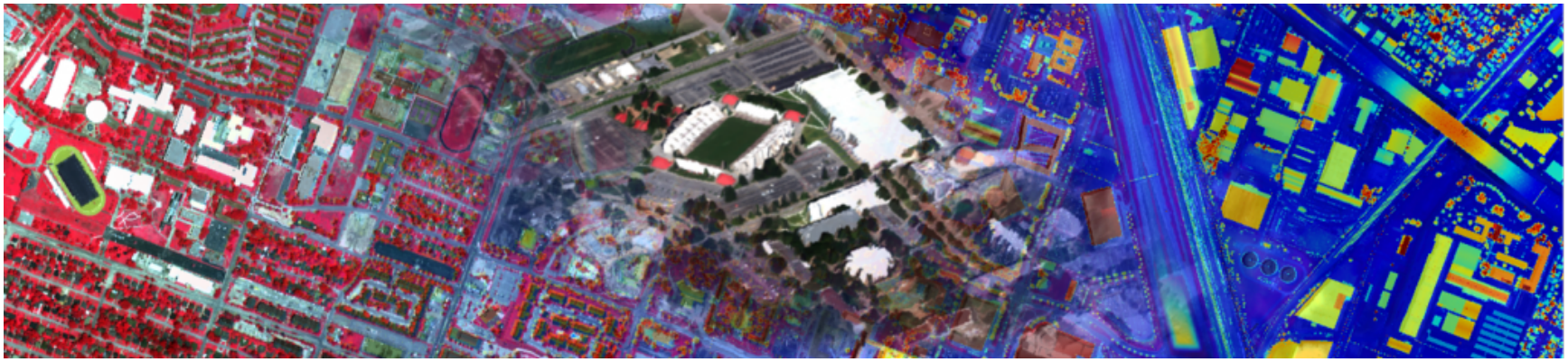
HSeg v. 1.61 result, E_t=0.05 and E_w=1.0, at global dissimilarity 0.371, 15 region classes and 14513 region objects.

3. Testbed Datasets

- Dataset 1: University of Houston, urban area mapping
- Dataset 2: Corpus Christi, seagrass mapping

Multi-sensor Testbed Dataset 1: University of Houston

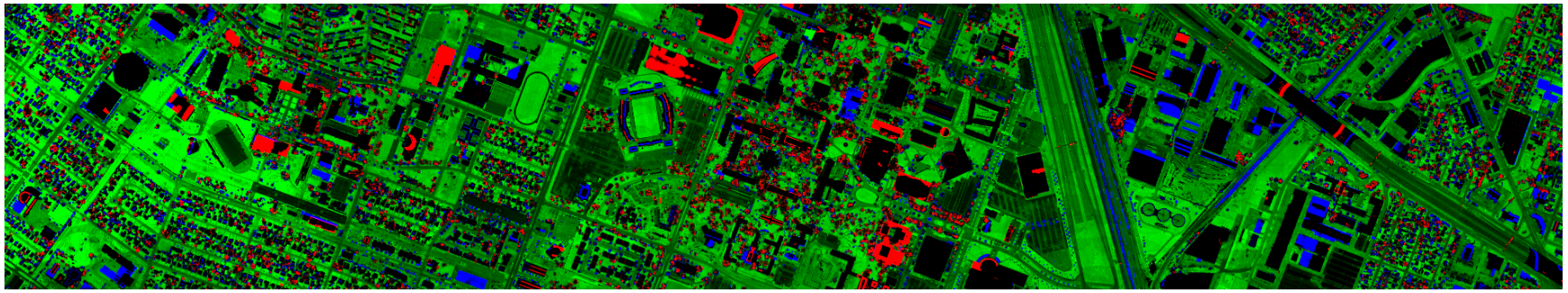
- UH-NCALM Multi-sensor data were processed and provided for analysis by algorithms developed in this research
  - Hyperspectral data: acquired by ITRES CASI sensor – Vis-VNIR
    - Samples/Lines/Bands: 349/1905/144.
  - LiDAR DSM (standard product extracted from point cloud).
  - LiDAR pseudo-waveform (extracted from point cloud)
    - Samples/Lines/Bands: 349/1905/80.
  - Provides a challenging classification scenario, including clouds in the image.



Multi-sensor Testbed Dataset 1: University of Houston

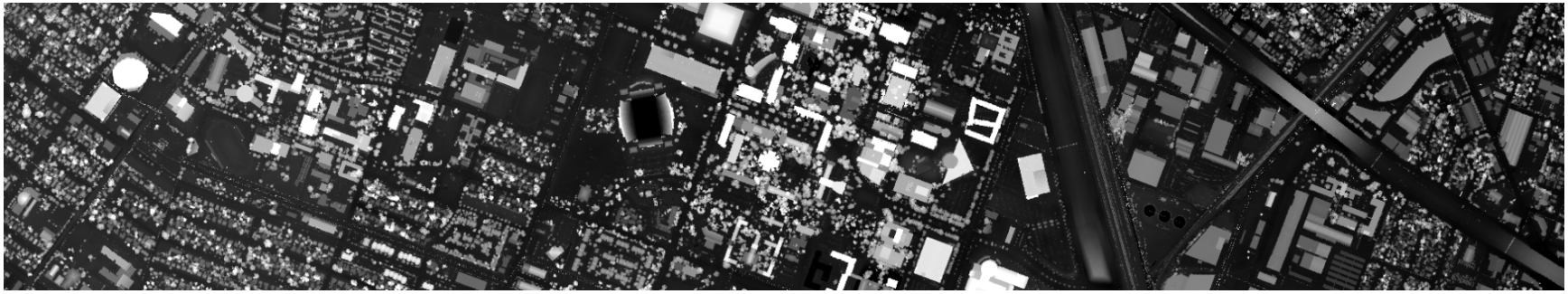
Hyperspectral: True Color Image

LiDAR Pseudo-Waveform: RGB Composite (Red: 20m, Green: 10m, Blue: 15m)

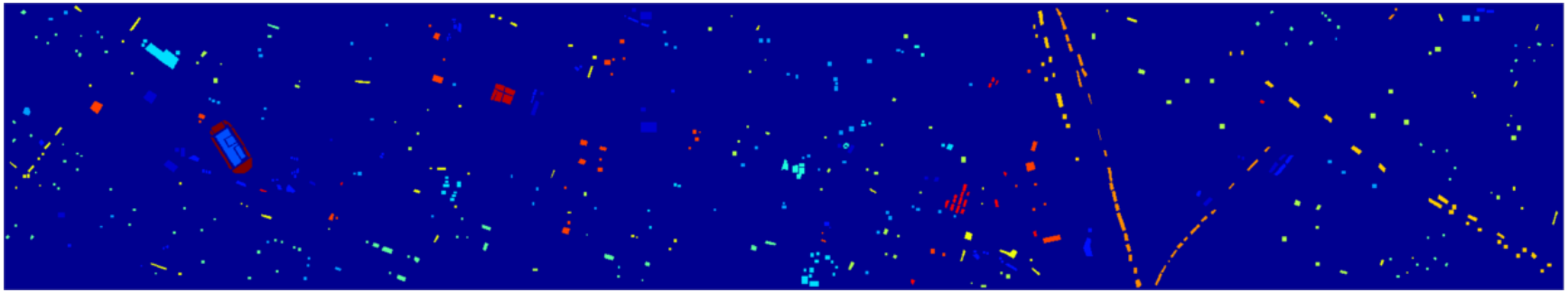


Multi-sensor Testbed Dataset 1: University of Houston

LiDAR DSM



Ground Reference Map



Healthy grass	Stressed grass	Synthetic grass	Trees	Soil	Water	Residential	Commercial
Road	Highway	Railway	Parking Lot 1	Parking Lot 2	Tennis Court	Running Track	

Multi-sensor Testbed Dataset 2: Corpus Christi, Seagrass Mapping

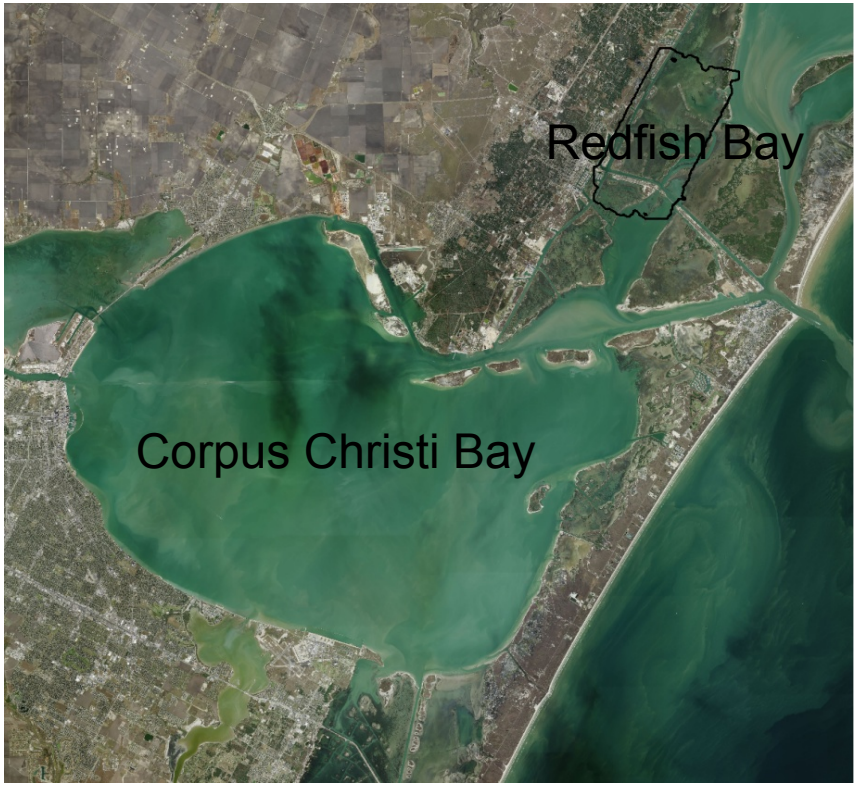
- Extent of seagrass habitat is an important indicator of ecosystem health in coastal environments
  - Seagrass beds are often negatively impacted by human activity such as shipping and dredging
  - Significant effort has recently been focused on reintroduction of seagrasses, resulting in the need to monitor recovery

- Monitoring seagrass via remote sensing
  - Hyperspectral data provide capability for mapping submerged vegetation in shallow environments with good clarity
  - Bathymetric LiDAR data provide relevant information on topography and possibly on submerged vegetation

- Objective: Classification of multi-sensor airborne data over coastal seagrass beds.

Multi-sensor Testbed Dataset 2: Corpus Christi, Seagrass Mapping

Study area

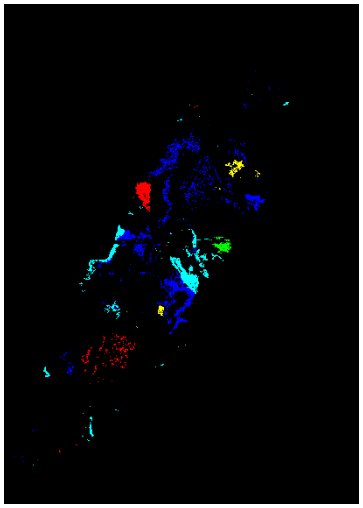


- Redfish Bay, Texas
- Coordinates: 27°54'47.01"N
  97°6'25.73"W

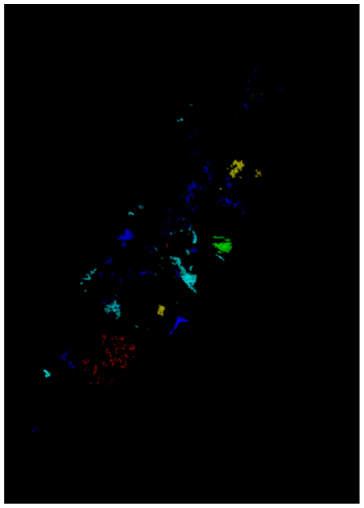
- Data acquisition:
  Hyperspectral image and LiDAR point cloud data collected by NCALM on September, 2012.

Multi-sensor Testbed Dataset 2: Corpus Christi, Seagrass Mapping

- Number of ground reference points extremely limited
- Ground reference information extended via spatial-spectral segmentation:
  - Hyperspectral data were spatially/spectrally clustered using HSEG
  - Photo-interpretation using very high resolution (5 cm) color images employed to remove incorrectly labeled pixels

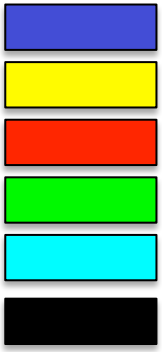


Extended ground reference after HSEG



Extended ground reference after screening

Legend:

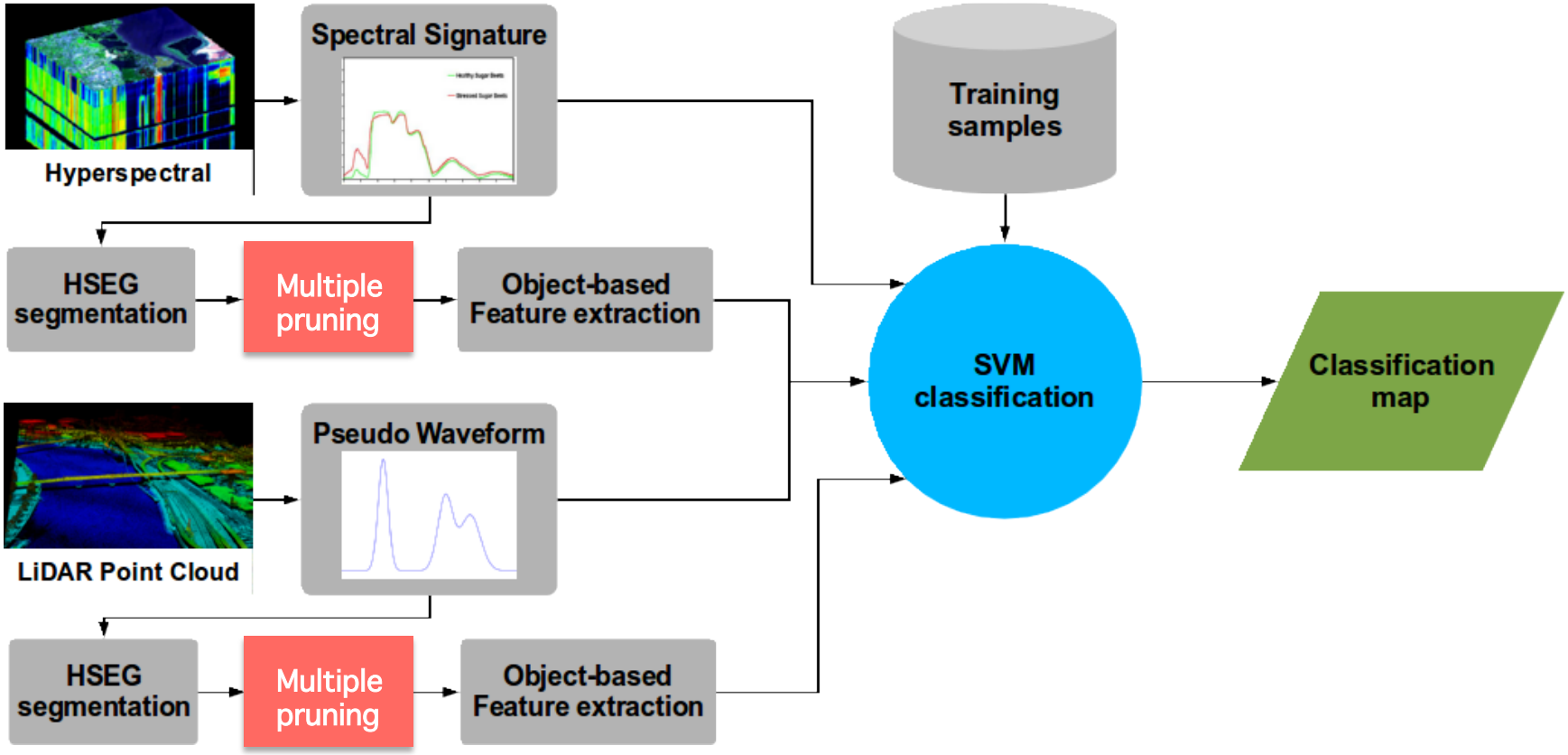


Thalassia: 45,424
Drift Algae: 11,800
Halodule: 11,012
Syringodium: 11,504
Water: 37,640
Undefined

Extraction of Features from Remote Sensing Data

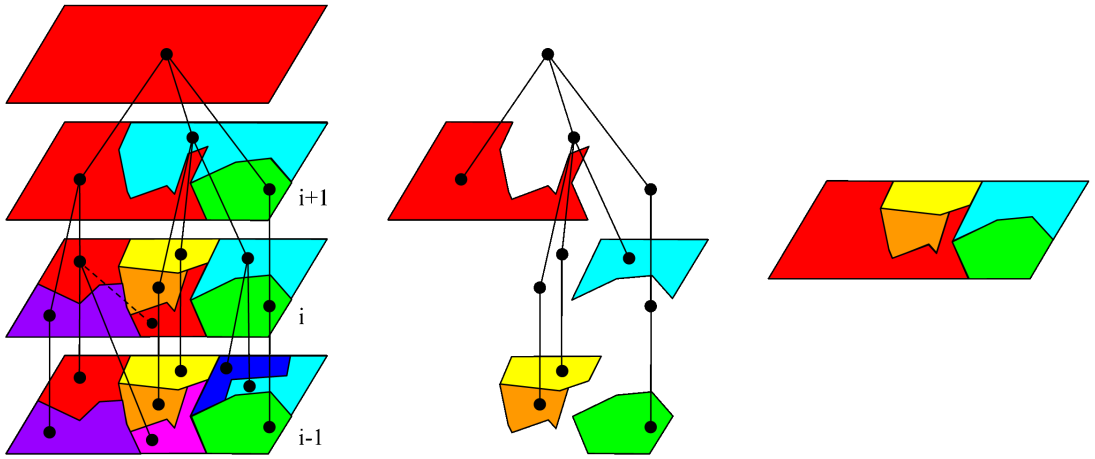
- Features from LiDAR data
  - Digital elevation maps and related features
  - Pseudo-waveforms from discrete return LiDAR data
  - Full-waveform signatures
- Spatial features
  - Object-based texture
  - Morphological features

SPATIAL FEATURES : OBJECT-BASED TEXTURE

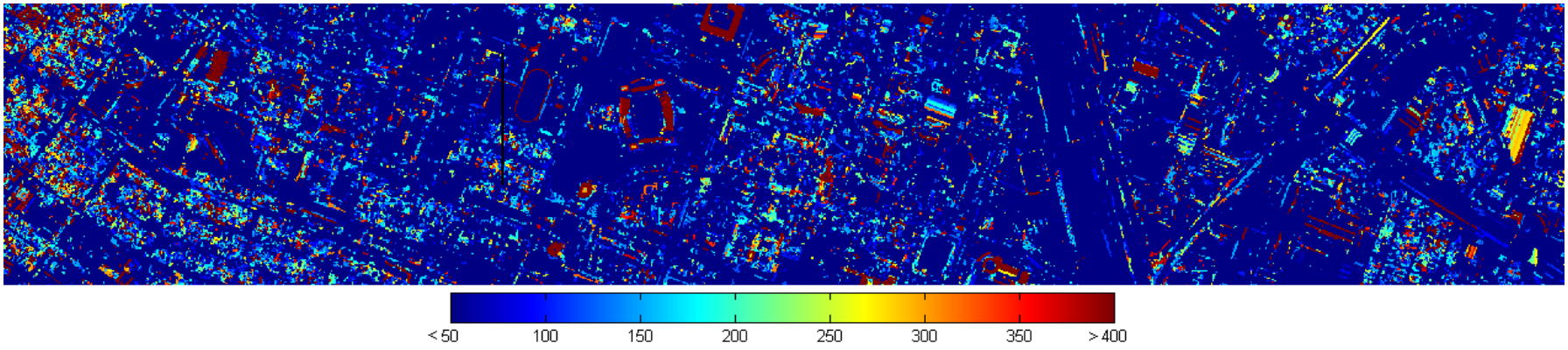


- Object-based feature extraction departs from traditional window-based approach extracting features from potentially irregular spatial regions.

SPATIAL FEATURES : OBJECT-BASED TEXTURE



Pruning level map



SPATIAL FEATURES : MORPHOLOGICAL FEATURES

- Morphological attribute filters process images by removing connected components that do not fulfill a given criterion.

- Morphological features are extracted from “thickened” and “thinned” variants of images (akin to erosion/dilation).

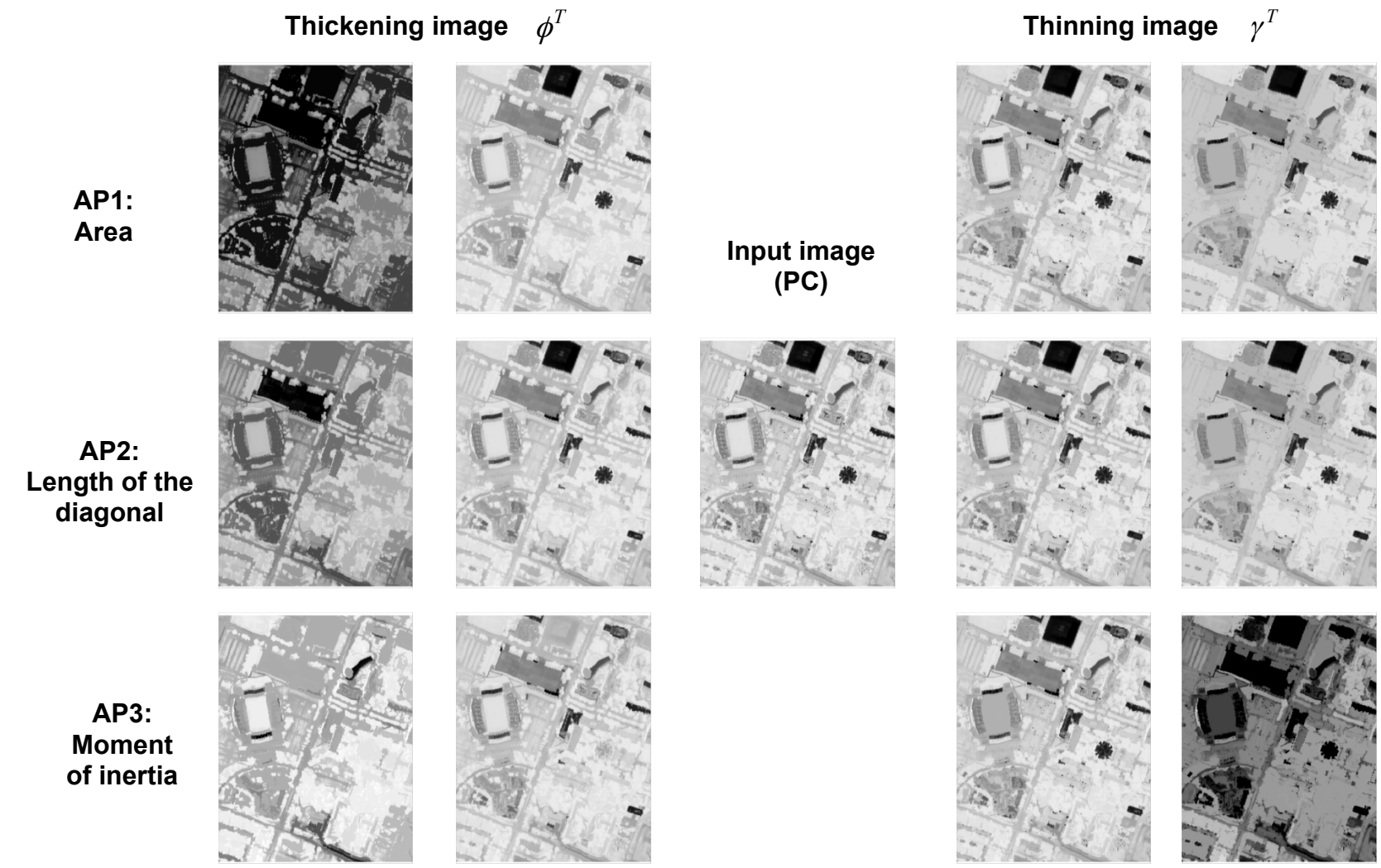
- Morphological attribute profiles (APs) consist of *n* morphological attribute thickening ϕ^T and *n* attribute thinning operators γ^T as given by

$$AP(f) = \{\phi_n^T(f),\ldots,\phi_1^T(f), f, \gamma_1^T(f),\ldots,\gamma_n^T(f)\}$$

  where f is the input image.

- Attributes can be geometric (e.g. area, shape, length of the perimeter, image moments), textural (e.g. range, standard deviation), etc.

Morphological features: Extended Multi Attribute Profiles (EMAPs)

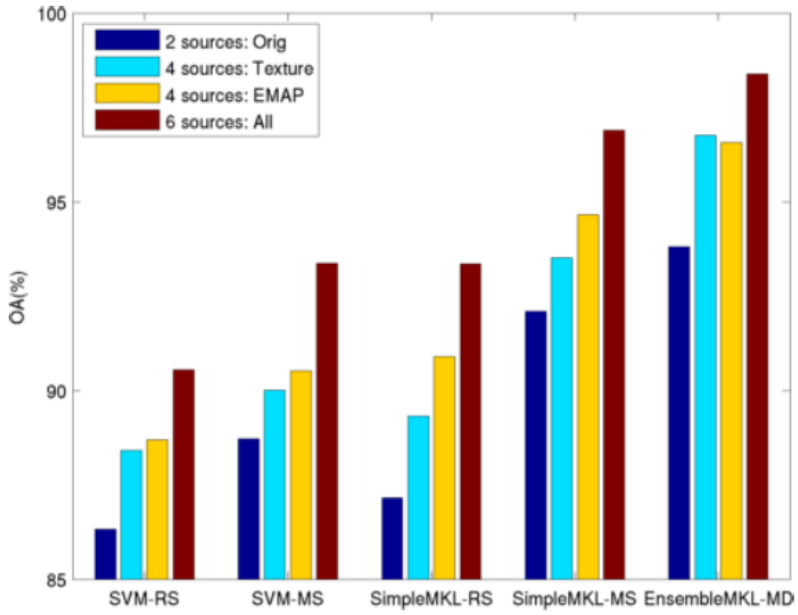


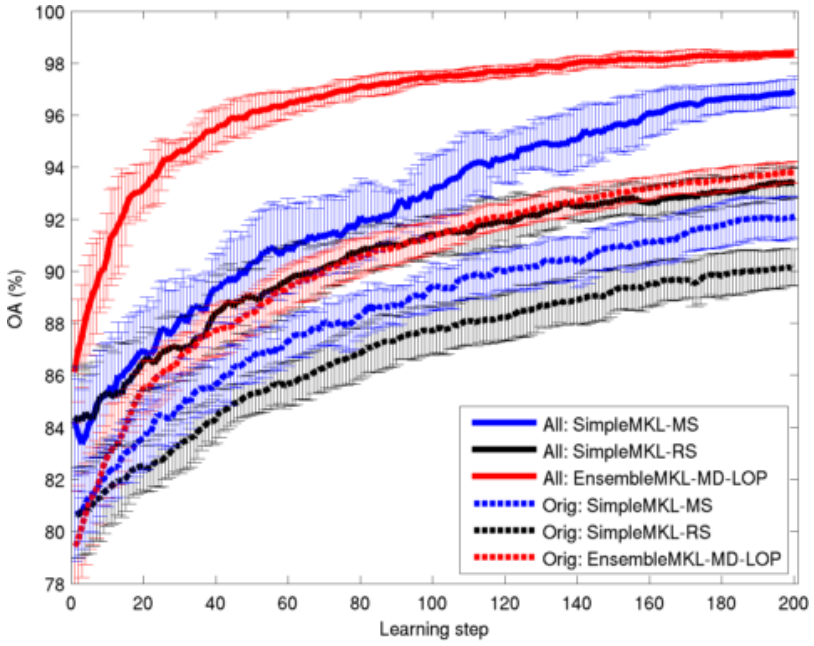
EXPERIMENTAL RESULTS

- Ensemble MKL-AL method
  - UH multi-source dataset
  - CC multi-source dataset

VALIDATION RESULTS WITH UH DATA: ENSEMBLE MKL-AL

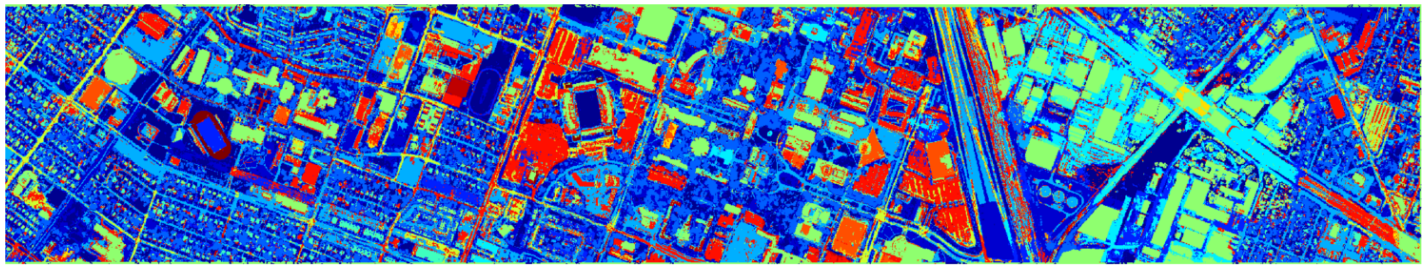
- Multiple feature extraction
  - Original hyperspectral and LiDAR pseudo-waveform
  - Object-based texture features
  - Extended multi attribute profiles (EMAPs)



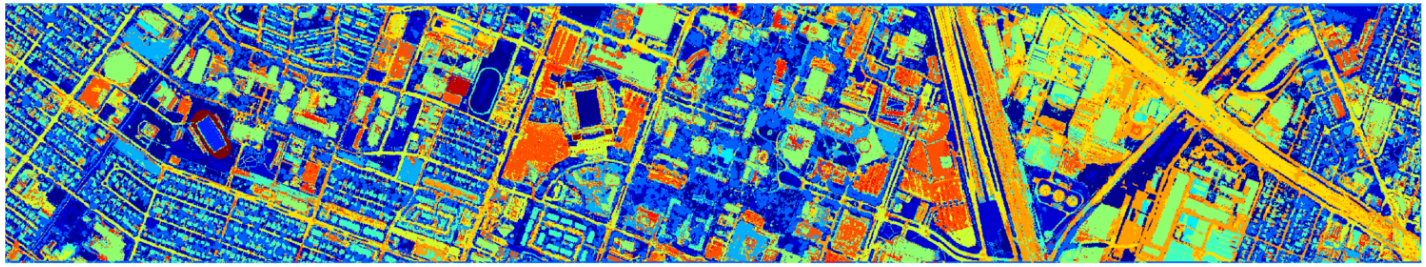


Validation Results With UH Data: Classification Maps

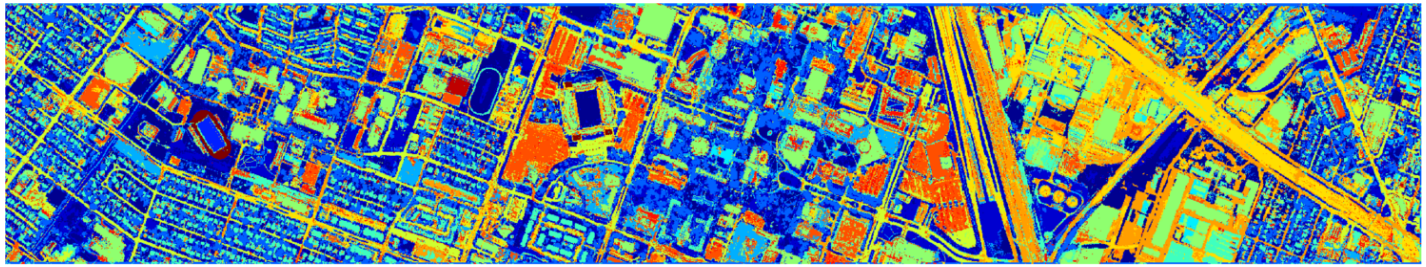
SimpleMKL-AL
2 sources



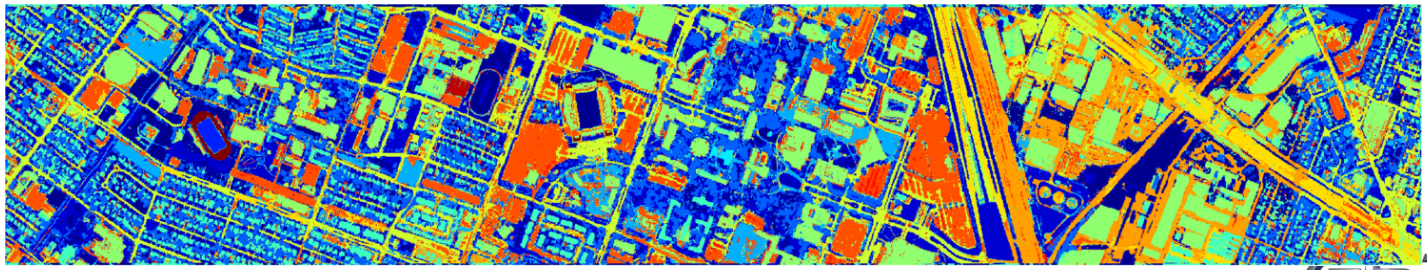
Ensemble MKL-AL
2 sources



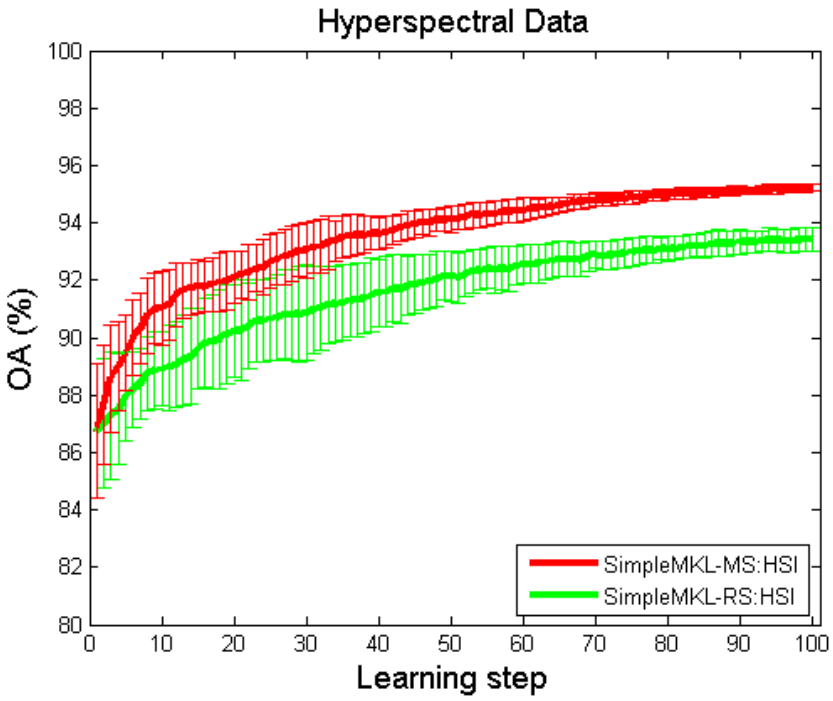
SimpleMKL-AL
6 sources

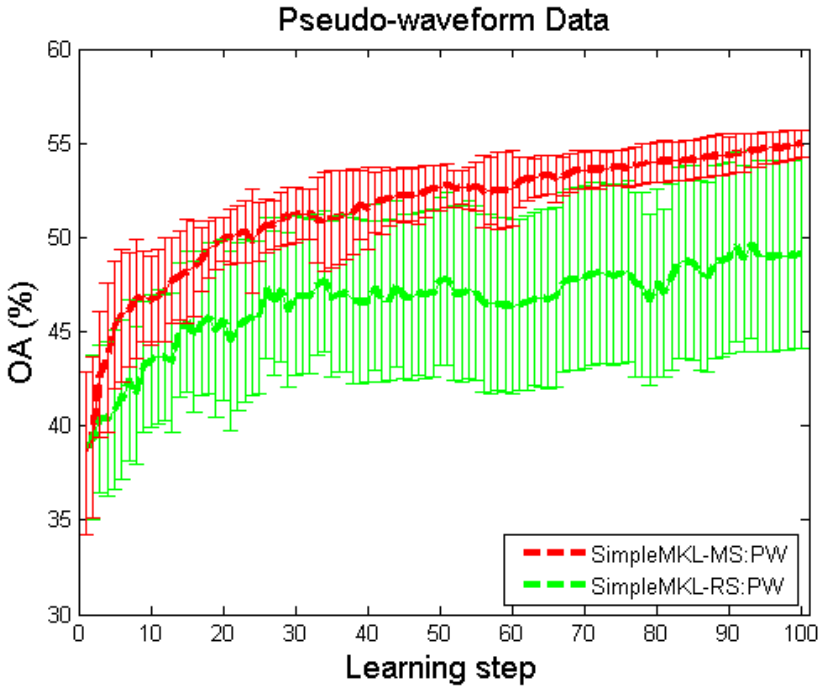


Ensemble MKL-AL
6 sources

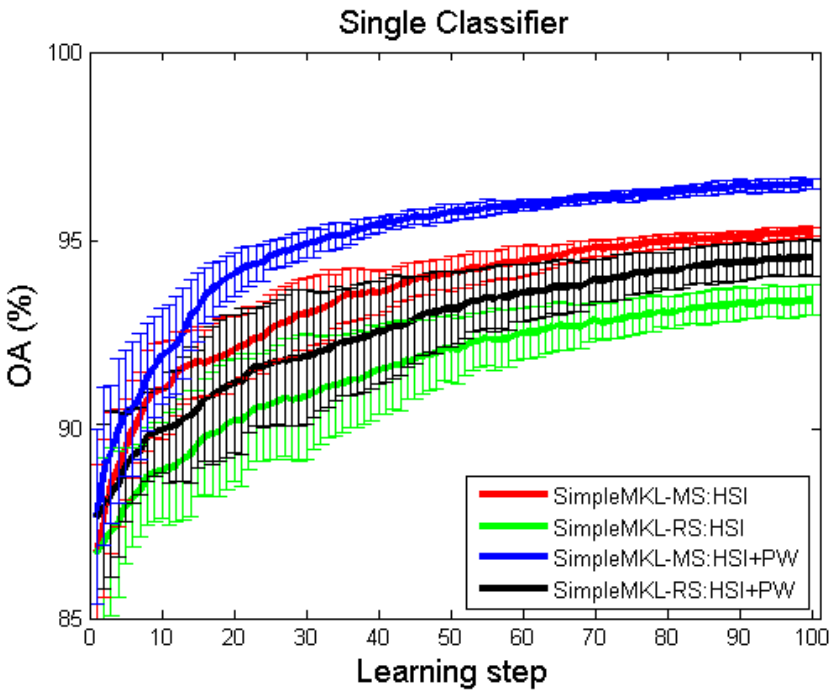


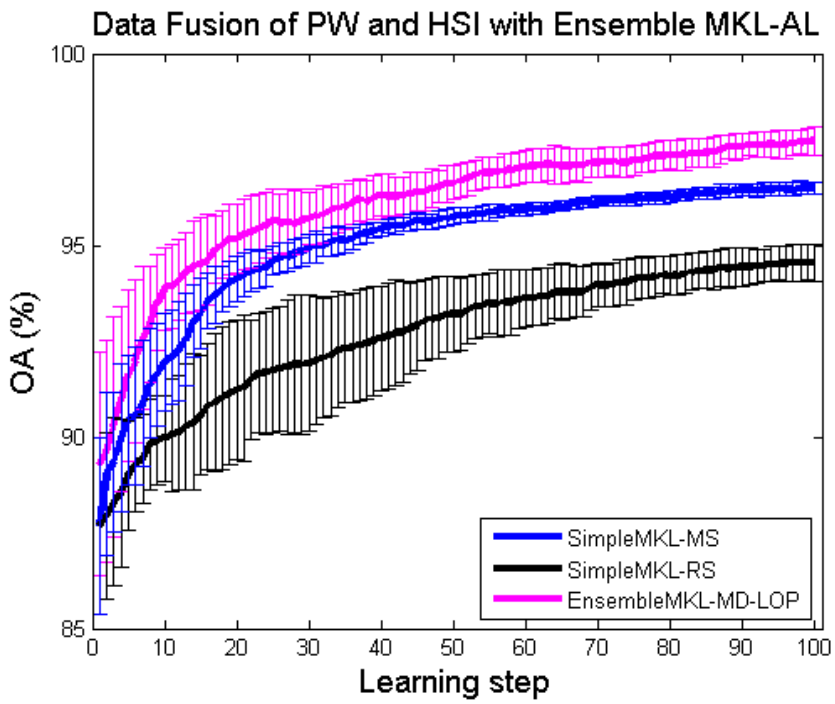
Validation Results With CC Data: Single Source (HSI and PW)



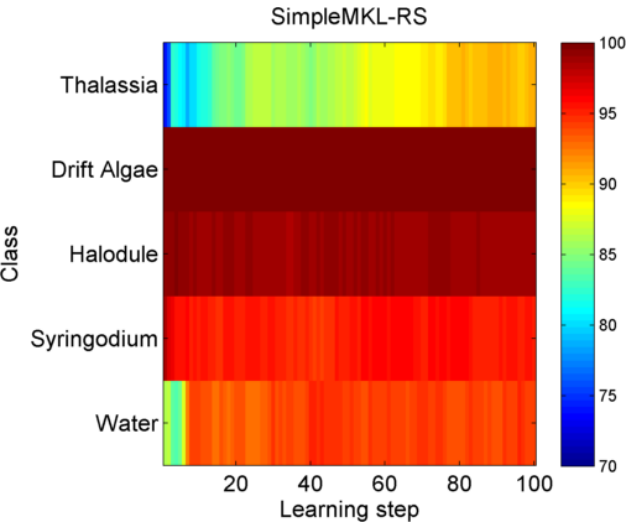


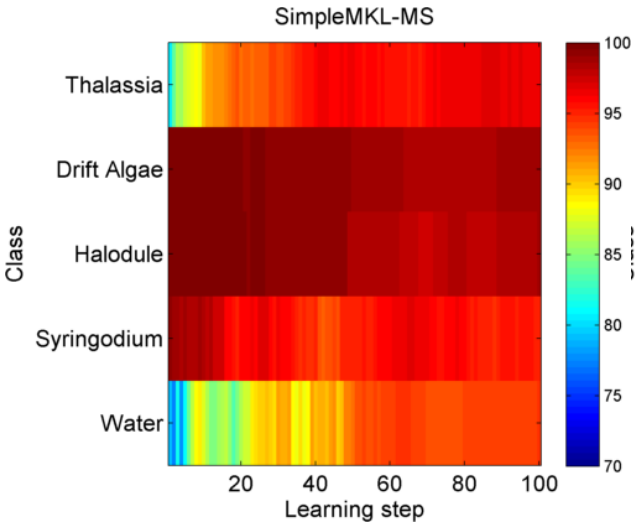
Validation Results With CC Data: Multiple Source (HSI + PW)

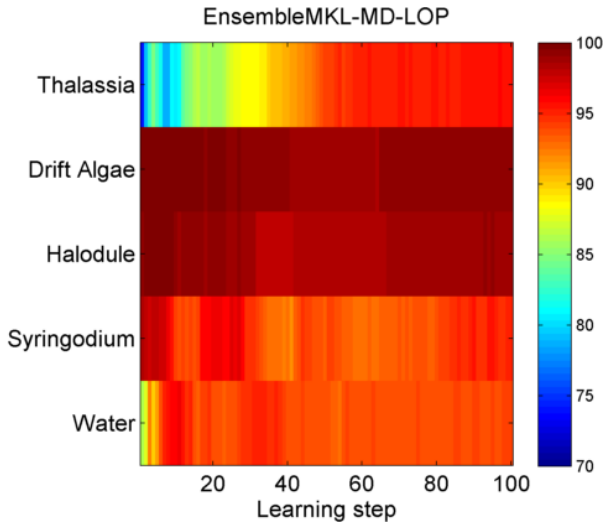


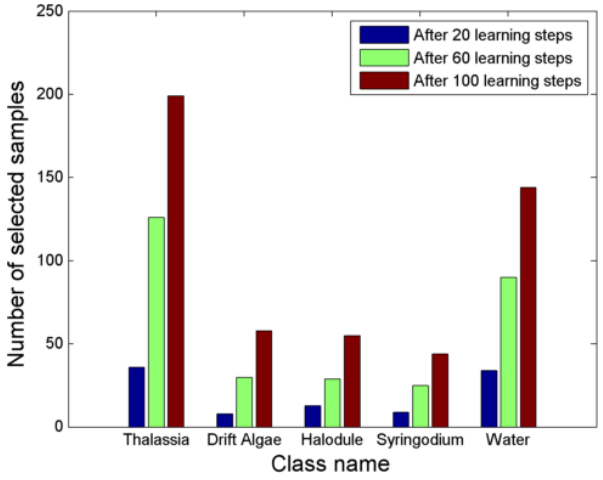


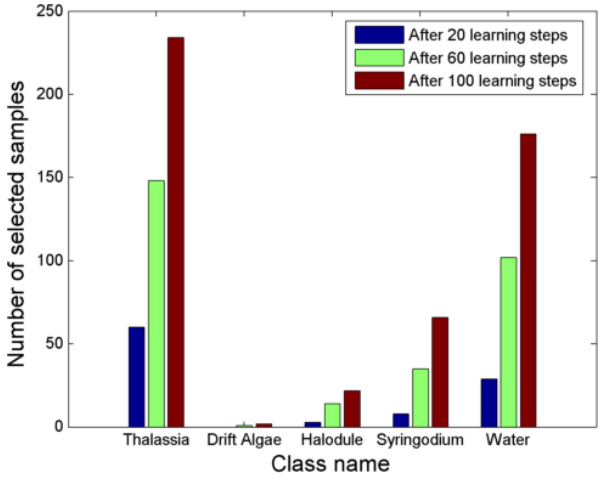
Validation Results With CC Data: Class-Specific Accuracies and Selected Samples

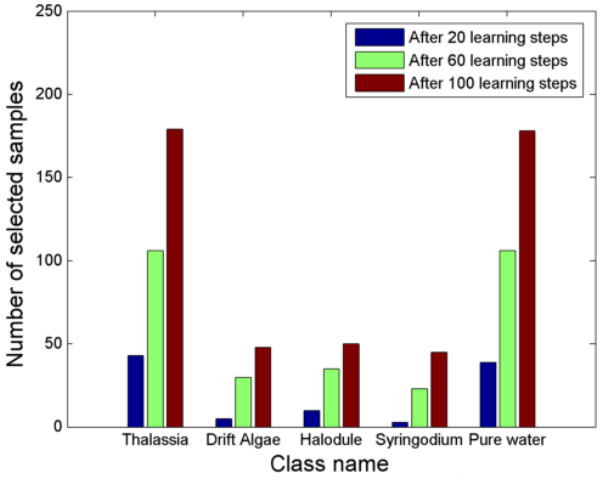






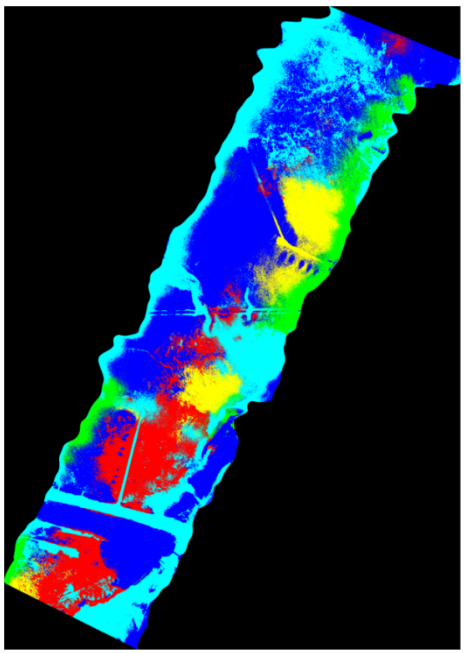




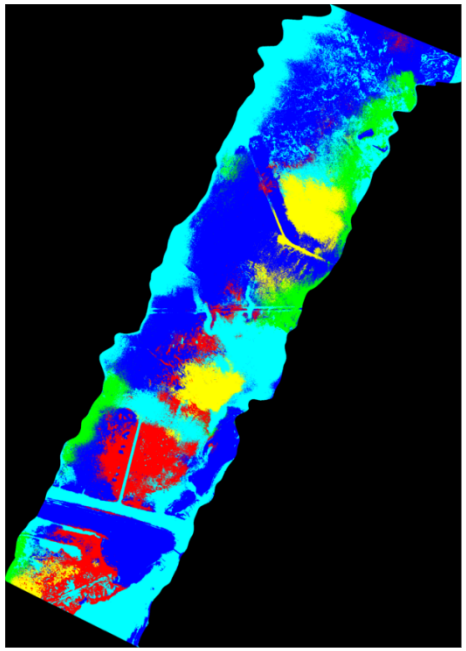


Validation Results With CC Data: Classification Maps

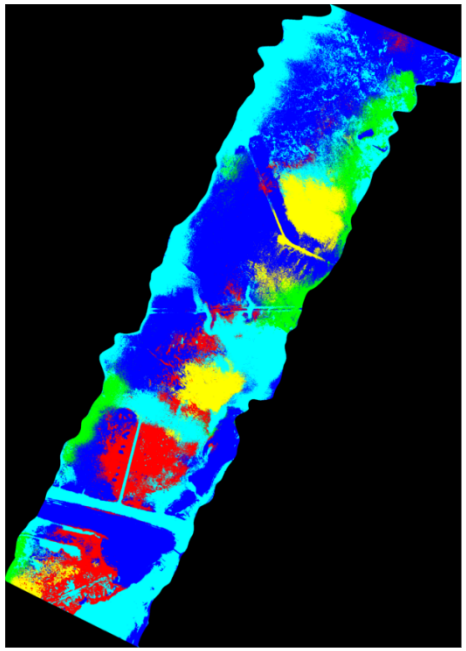
Data fusion of PW and HSI



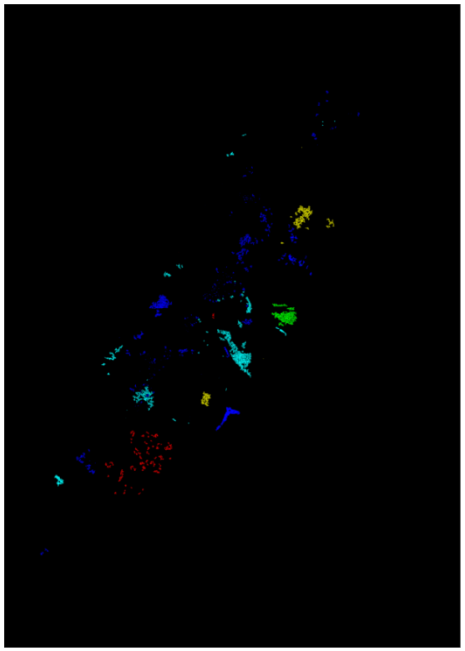
SimpleMKL-RS



SimpleMKL-MS



EnsembleMKL-MD-LOP



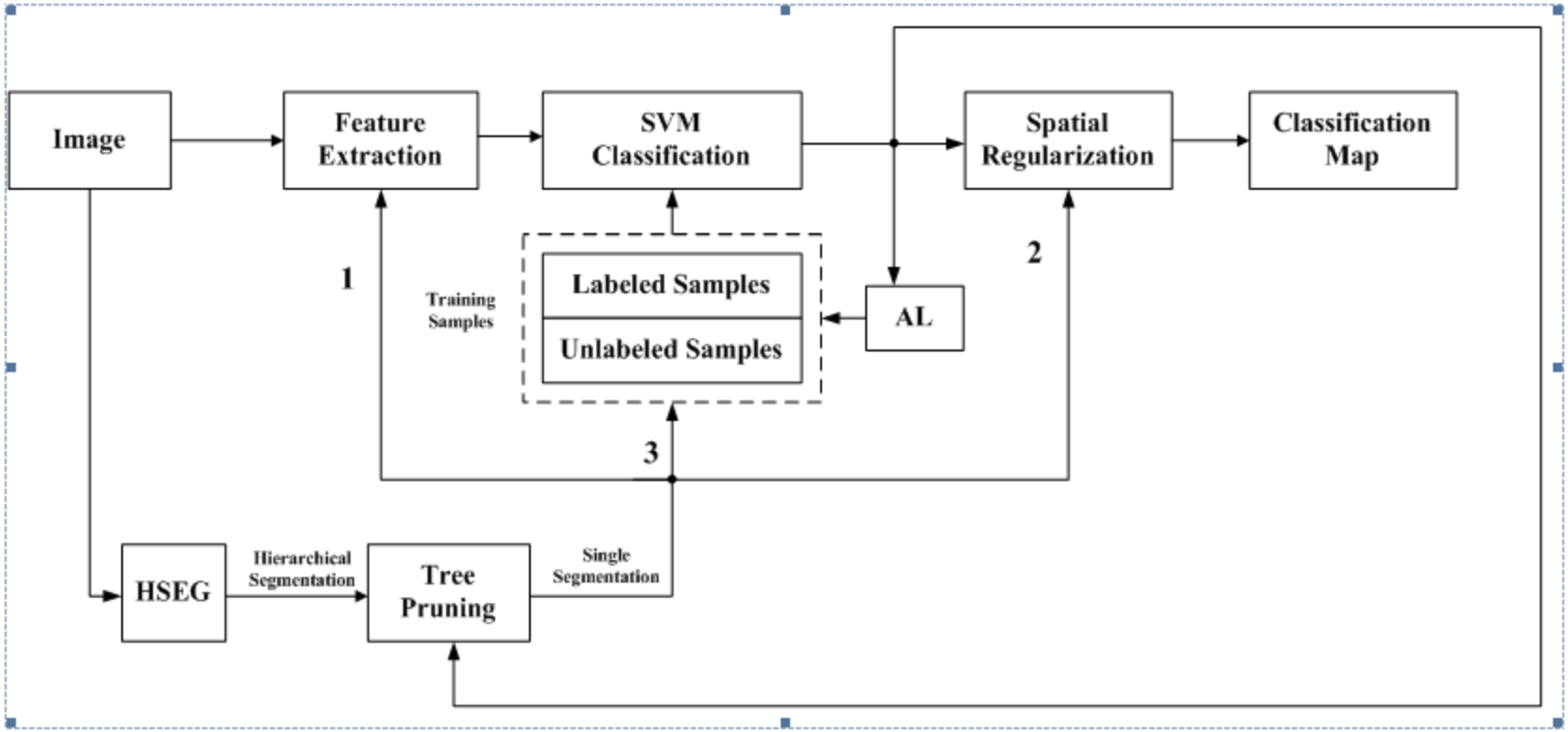
Ground Reference Map

Thalassia **Drift Algae** **Halodule** **Syringodium** **Water**

4. Ongoing and Future Work

- Full integration with the HSeg algorithm for texture feature extraction
  - –Incorporating HSeg into the query step of the proposed AL framework.
- Extension to other multi-sensor datasets

HSEG AND AL INTEGRATION FRAMEWORK



1. **Method 1: Adding Features**
2. **Method 2: Regularization**
3. **Method 3: Adding Samples**

HSeg and AL Integration Framework: Proposed Methods

- Method 1: Adding Features
  - Spatial features (mean and std.) extracted from HSeg are added to the original spectral features. Then, both spectral and spatial features are used to do the classification.
- Method 2: Regularization
  - Only the spectral features are used to do the classification, and then the HSeg-based classification map is used to refine the classification results.
- Method 3: Adding Samples
  - Spatial information is used to extend the training set by using a semi-supervised approach. We add both labeled and unlabeled samples into the training set during each iteration. The labeled samples are selected by using the breaking ties (BT) criterion.

SUMMARY

- MKL-AL is highly effective in single and multi-source scenarios
  - Adapts the kernel to the dataset at hand.
- Ensemble-MKL-AL results in substantial improvements
  - Multi-view active learning and decision level fusion results in induction of highly informative samples, and results in superior classification performance, particularly in multi-source scenarios.
- Multiple views can be generated by different sensors, different feature-types, spectral subsets, and other approaches that result in diverse views of the same scene.
- HSEG is instrumental, not only in object level feature extraction, but also for object-based active learning
  - Object level feature extraction has been implemented and validated. We are currently developing object level active learning.

5. Publications

Conference Publications to Date

1. H.H. Yang, Y. Zhang, S. Prasad, M.M. Crawford, "Multi-Kernel active learning for robust geo-spatial image analysis," in *Proc. of the 2013 IEEE Geoscience and Remote Sensing Symposium*, July 22-26, Melbourne, Australia, 2013.
2. H.L. Yang and M.M. Crawford, "Learning a joint manifold with global-local preservation for multitemporal hyperspectral image classification, in *Proc. of the 2013 IEEE Geoscience and Remote Sensing Symposium*, July 22-26, Melbourne, Australia, 2013.
3. X. Zhou, S. Prasad, M.M. Crawford, "Wavelet domain multi-view active learning for hyperspectral image analysis," in *Proc. of the 2014 IEEE Workshop on Hyperspectral Image and Signal processing: Evolution in Remote Sensing,* July 24-27, Lausanne, Switzerland, 2014.
4. E. Pasolli, H.L. Yang, and M.M. Crawford, "Combining active and metric learning for hyperspectral image classification," in *Proc. of the 2014 IEEE Workshop on Hyperspectral Image and Signal processing: Evolution in Remote Sensing,* July 24-27, Lausanne, Switzerland, 2014.
5. J.C. Tilton and E. Pasolli, "Incorporating edge information into best merge region-growing segmentation," in *Proc. of the 2014 IEEE Geoscience and Remote Sensing Symposium*, July 13-18, Quebec, Canada, 2014.
6. X. Zhou, S. Prasad, M.M. Crawford, "Wavelet domain active learning for robust classification of full-waveform LiDAR data," in *Proc. of the 2014 IEEE Geoscience and Remote Sensing Symposium*, July 13-18, Quebec, Canada, 2014.
7. X. Zhou, Y. Zhang, H.L. Yang, S. Prasad, J. Jung, M.M. Crawford, M. Starek, A. Singhania, J. Fernandez-Diaz, A. Lord, "Seagrass mapping via active learning using airborne hyperspectral and LiDAR measurements," in *Proc. of the 2014 IEEE Geoscience and Remote Sensing Symposium*, July 13-18, Quebec, Canada, 2014.

JOURNAL PUBLICATIONS TO DATE

1. J. Jung, E. Pasolli, S. Prasad, J.C. Tilton, M.M. Crawford, "A framework for land cover classification using discrete return LiDAR data: Adopting pseudo-waveform and hierarchical segmentation," *in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 2, pp. 491-502, Feb. 2014.
2. Y. Zhang, H.L. Yang, S. Prasad, E. Pasolli, J. Jung, M.M. Crawford, "Ensemble multiple kernel active learning for robust classification of multi-source remote sensing data", submitted to *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
3. Y. Zhang, S. Prasad, "Locality preserving composite kernel feature extraction for multi-source geospatial image analysis," accepted pending revisions in the *IEEE Journal of Selected Topics in Earth Observations and Remote Sensing*, 2014.
4. X. Zhou, S. Prasad, M.M. Crawford, "Wavelet domain active learning for remote sensing image analysis," manuscript in progress for the *IEEE Transactions on Geoscience and Remote Sensing*.
5. E. Pasolli, J. Jung, S. Prasad, J.C. Tilton, M.M. Crawford, "Spatial-based ensemble approach for fusing hyperspectral and discrete return LiDAR data," manuscript in progress.
6. E. Pasolli, H.L. Yang, and M.M. Crawford, "Active-metric learning for classification of hyperspectral images," manuscript in progress.

ACRONYMS

AL	Active Learning
EMAPs	Extended multi attribute profiles
HSWO	Hierarchical Step-Wise Optimization
HSeg	Hierarchical Segmentation
LOP	Linear Opinion Pool
MD	Maximum Disagreement
MKL	Multi Kernel Learning
MS	Margin Sampling
MV	Majority Voting
RS	Random Sampling
RHSeg	Recursive Hierarchical Segmentation
SVMs	Support Vector Machines