

A variance-based Bayesian framework for improving Land-Cover classification through wide-area learning from large geographic regions



Tommy Chang*, Bharath Comandur, Johnny Park, Avinash C. Kak

Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, 47907, USA

ARTICLE INFO

Article history:

Received 10 July 2015

Revised 2 April 2016

Accepted 2 April 2016

Available online 6 April 2016

ABSTRACT

Common to much work on land-cover classification in multispectral imagery is the use of single satellite images for training the classifiers for the different land types. Unfortunately, more often than not, decision boundaries derived in this manner do not extrapolate well from one image to another. This happens for several reasons, most having to do with the fact that different satellite images correspond to different view angles on the earth's surface, different sun angles, different seasons, and so on.

In this paper, we get around these limitations of the current state-of-the-art by first proposing a new integrated representation for all of the images, overlapping and non-overlapping, that cover a large geographic ROI (Region of Interest). In addition to helping understand the data variability in the images, this representation also makes it possible to create the ground truth that can be used for ROI-based wide-area learning of the classifiers. We use this integrated representation in a new Bayesian framework for data classification that is characterized by: (1) learning of the decision boundaries from a sampling of all the satellite data available for an entire geographic ROI; (2) probabilistic modeling of within-class and between-class variations, as opposed to the more traditional probabilistic modeling of the “feature vectors” extracted from the measurement data; and (3) using variance-based ML (maximum-likelihood) and MAP (maximum a posteriori) classifiers whose decision boundary calculations incorporate all of the multi-view data for a geographic point if that point is selected for learning and testing.

We show results with the new classification framework for an ROI in Chile whose size is roughly 10,000 square kilometers. This ROI is covered by 189 satellite images with varying degrees of overlap. We compare the classification performance of the proposed ROI-based framework with the results obtained by extrapolating the decision boundaries learned from a single image to the entire ROI. Using a 10-fold cross-validation test, we demonstrate significant increases in the classification accuracy for five of the six land-cover classes. In addition, we show that our variance based Bayesian classifier outperforms a traditional Support Vector Machine (SVM) based approach to classification for four out of six classes.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Our interest in pixel-level classification of satellite images¹ is driven by the role that such classifications can play in solving

problems related to the geolocalization of everyday photographs and videos of outdoor scenes. Using spatial relationships between different land-types - say between arable land, a pond, and a nearby road - to establish correspondences between the “objects” seen in a photograph and those extracted from the satellite images can significantly reduce the candidate locations for where the photograph was taken. For obvious reasons, reliable pixel level classification of satellite data is a necessary prerequisite to the development of such solutions to geolocalization problems.

While there has been much work during the last several decades on the classification of multispectral data in satellite images Anderson (1976); Unsalan (2003), as such this work cannot directly be used for solving geolocalization problems in general.

* Corresponding author. Tel.: +17654913680.

E-mail addresses: chang177@purdue.edu (T. Chang), bcomandur@purdue.edu (B. Comandur), jpark@purdue.edu (J. Park), kak@purdue.edu (A.C. Kak).

¹ In reality, pixel-level classification of a satellite image amounts to carrying out land-type classification of an array of points in that portion of the earth's surface that is viewed in the image. Before classification, the multispectral data in a satellite image goes through a processing step called orthorectification that “maps” the array of pixels in an image to an array of latitude/longitude (lat/long for short) coordinates.

Before we explain the reasons for why that is the case, note that the traditional approaches to satellite data classification use standard statistical pattern recognition techniques, with the more recent contributions also using decision trees, random forests, neural networks, support vector machines, and so on [DeFries and Chan \(2000\)](#); [Duro et al. \(2012\)](#); [Huang et al. \(2002\)](#); [Marchisio et al. \(2010\)](#).

One thing that is common to practically all these traditional approaches is that the training and the testing data for constructing a classifier are drawn from the same satellite image. There are very few examples where authors have derived the decision boundaries in one satellite image and shown usable classification results in a different satellite image in a given geographic area.² In practically all these cases, the classifiers derived for one satellite image fail to perform adequately on the other satellite images even in the same general geographic region. And, when such results have been shown, it is usually for an adjacent area for which the satellite images were captured at the same time as for the image from which the training data was drawn.

While these traditional approaches may suffice for solving, say, land-cover resource management and forecasting problems, they come up short when solving problems related to the geolocalization of photographs and videos. The reason has to do with the fact that the data in satellite images is affected by the change of seasons, the off-nadir/elevation angle associated with a satellite view, the sun angle, and so on. The logic of the algorithms that one might use for geolocating a photograph becomes simpler if the information extracted from the satellite images is invariant to all of the aforementioned changes to the maximum extent possible. The easiest way to achieve such invariance is to use ALL of the satellite data that may be available for a given geographic region. When decision boundaries in a feature space are based on all of the data – meaning data recorded at different times of the year, with different off-nadir/elevation angles, with different sun angles, etc. – any discriminations one is able to make in that feature space are likely to possess the desired properties of invariance. Said another way, *our classifier would be able to make distinctions between the variances associated with the objects that look more or less the same around the year and the variances associated with the objects that change significantly with, say, seasons.*

An additional aspect related to using satellite imagery for solving the problems of geolocalization is that a photograph (or a video) may have been recorded anywhere in a large geographic region of interest whose area may far exceed what is typically associated with a single satellite image. This requires that the land-cover classifications be carried out for the entire ROI using all of the satellite data available for the region.

For reasons stated above, this paper addresses the problem of land-cover classification from a larger geographical perspective than has traditionally been the case in the past. We want to be able to classify all of the data in an ROI (Region of Interest) that can be much larger than the area covered by a typical single satellite image. Consider, for example, the Chile ROI shown in [Fig. 1](#). This ROI, of size 10,000 km² is covered by a total of 189 satellite images in the WorldView2 dataset. Our goal is to see if it is possible to create decision boundaries from all of this data taken

together so that the overall classification rate for the entire ROI shown in [Fig. 1](#) would be at a usable level of accuracy.

Obviously, before we can design a classifier at the level of an ROI, we must first come to grips with the data variability over the ROI. Understanding data variability at the scale of a large ROI presents its own challenges and can be thought of as a “Big Data” problem. The challenges are created by the typical fast-response and dynamic-storage needs of any human-interactive computer system that must work with very large variable-sized datasets.³ We have addressed these challenges by developing a special software tool (named PIMSIR for “Purdue Integrated MultiSpectral Image Representation” tool) that is custom designed to achieve the following:

- Rapid visualization of all of the data in an ROI
- Rapid visualization of the geographic area overlaps between the satellite images. Understanding the overlaps is important because any probabilistic modelling of the data at any given geographic *point* is predicated on how much data is available at that point through overlapping satellite views.⁴
- Rapid visualization of the variability of the spectral signatures⁵ both spatially and across the views.

This tool is run on a cloud-based cluster of five physical computing nodes, each with up to 48 cores and 256 GB of RAM, that are connected with a 10 Gb network switch. The system is supported by a network storage server with 24 TB of storage.

Even after understanding the extent of data overlap and variability, there remains the big issue of what classification strategy to use for the data. Machine learning now gives us tens of choices for classifiers and it's not always clear at the outset as to which choice would work the best for a given problem. In this paper, based on our analysis of the data overlap and of the view-to-view variability that we have seen, we chose to design ML (maximum-likelihood) and MAP (maximum a posteriori) classifiers by probabilistic modeling of NOT the spectral signatures themselves, but of the variability in the spectral signatures. We show our ROI based results obtained with this Bayesian classifier, and for comparison, also with the more commonly used SVM classifiers in [Section 8](#) of this paper. As the reader will see, this comparison justifies our intuition regarding the superiority of variance-based Bayesian classification vis-a-vis the more traditional approaches (as exemplified by SVM-based classification).

Moreover, even after a choice is made regarding the classification strategy, there remains the complex problem of how to generate on an ROI basis the positive and negative examples for the different land-cover types for training and testing a classifier. Obviously, it would be very challenging for a human to scan through an entire ROI and manually select such examples. What is needed is a human-in-the-loop random sampling strategy that has the power to yield positive and negative samples that adequately represent ROI based distributions for the different land-cover types. In our classifier training protocol, we use a random sampler based on the Metropolis-Hastings algorithm to select a small number of ROI subregions for presentation to a human and it is for the human to decide whether or not to use that subregion for generating the positive examples for a given land-cover label. (What if a randomly selected subregion is mostly over water while the human is seeking positive examples of high vegetation?) After the human has ac-

² We draw the reader's attention to Wilkinson's survey [Wilkinson \(2005\)](#) whose major conclusions are just as valid today as they were when the survey was published in 2005. This survey covered 574 experiments in satellite image classification as reported in 138 publications over a period of 15 years. Wilkinson concluded that despite the large body of published research, virtually no progress had been made in satellite image classification over the time period covered by the survey. One of the reasons he highlighted for this lack of progress was the common practice of the researchers drawing their training and testing data sets from the same satellite image.

³ By very large, we mean datasets that are hundreds of gigabytes in size.

⁴ In general, probabilistic modelling is with respect to spatial distribution of the observed data. However, in order to address view-to-view data variability issues, one can also talk about probabilistic modeling with respect to the viewing dimension.

⁵ We use the term spectral signature to refer to the 4 or 8 spectral band values at each pixel.



Fig. 1. The entire Chile ROI (Region of Interest) covering about 10,000 square kilometers.

cepted a subregion, he/she can zoom into the subregion and with quick mouse clicks mark the positive and negative examples. This entire process takes about 30 minutes for each land-cover label for a large ROI of the size that we will show experimental results on later in this paper.

In the rest of this paper, [Section 2](#) presents the details of the PIMSIR tool for understanding data overlaps and variability in all the satellite images available for a given ROI. [Section 3](#) then presents the ROI-based protocol we use for generating the positive and the negative examples for training and testing the classifiers. Subsequently, in [Section 4](#) we discuss the theoretical framework used for probabilistic modeling of the variances and how these probabilities can be used for ML and MAP classifiers for land-cover classification. [Section 5](#) briefly describes the SVM classifiers that we used to compare our Bayesian classifier with. [Section 6](#) discusses the notion of “compatible classes.” The point here is that certain types of geographical points can be expected to show predictable seasonal variations that are best handled by defining a set of compatible classes. This is particularly the case for agricultural land. So we suggest that the geographical points that possess this property be given compound land-class labels that reflect these predictable variations. For experimental validation, we first present an overview of how the experiments were conducted in [Section 7](#) and then proceed to show the actual results in [Section 8](#). The results shown demonstrate that ROI based wide-area learning

significantly outperforms single satellite image based learning using both our Bayesian and the more traditional SVM classifiers. In addition we show that our Bayesian classifier outperforms the traditional SVM based classification for four out of six classes.

2. The PIMSIR tool for understanding multispectral data overlap and variability at the scale of a large ROI

As alluded to in the Introduction, we have created a software tool called PIMSIR for the purpose of understanding data variability at the ROI level. On the one hand, PIMSIR gives us a holistic representation of all the satellite images that cover an ROI, and, on the other, it allows the data provided by multiple satellite images in any given subregion to be viewed and manipulated easily regardless of whether that subregion has to pull in information from overlapping or non-overlapping images.

What makes PIMSIR versatile is that the data structure it creates for an ROI is dynamic - in the sense that an ROI is allowed to be covered by an arbitrary number of satellite images, with arbitrary degrees of overlap between the images. As to the size of the ROI that can be accommodated in this representation, that depends on the number of images. For example, the Chile ROI covers about 10,000 km² and consists of 189 overlapping images, each

with a spatial resolution of two meters per pixel.⁶ Its corresponding PIMSIR structure takes 208 GB of disk storage. As mentioned in the Introduction, PIMSIR runs on a cloud-based cluster of high-performance physical nodes, each with up to 48 cores and 256 GB of memory. This allows a large ROI to be accommodated entirely in the fast memory of just one node. For obvious reasons, that speeds up the construction of the PIMSIR structure. However, it is not a necessary requirement that an entire ROI fit in the fast memory of a single node.

In what follows, we will describe how we create the PIMSIR structure for a given ROI. Subsequently, we will describe the information that is stored in PIMSIR for each geographical point in the ROI.

Given a raw WorldView2 Multispectral Image (MSI), the first necessary preprocessing step is to apply the Top-of-Atmosphere (ToA) reflectance correction [Udipi and Comp \(2010\)](#) to the images in order to normalize out the view-angle (with respect to the sun angle) variability. Then, the corrected image goes through the orthorectification process that maps the pixels to geo lat/long coordinates.⁷

After all of the corrections mentioned above have been applied to the data, creating the PIMSIR structure involves the following steps:

1. Construct a bounding box for an ROI.
2. Rasterize the bounding box with a matrix of sampling points, taking into account implicitly the spatial resolution desired. The results we show in this section are for the case when each cell in the bounding box represents a 2m x 2m area.
3. Scan the bounding box and, for each cell, calculate its geo lat-long coordinates.
4. Fetch the list of satellite images for the lat-long coordinates at a sampling point. Project the bounding-box sampling point into each image and:
 - Record the four MSI pixels that are nearest to the projected point in the satellite image. (The point projected into an image will, in general, not correspond exactly to any of the pixel locations in the image.)
 - Apply bilinear interpolation to the spectral signatures at the four nearest neighbors and return this answer for the image in question.
5. Pool together the spectral signatures collected from all the satellite images that see the geo-point and store them in a compact data structure whose pointer is held by the bounding-box point in question.

Each location in the PIMSIR array points to a linked list of nodes, the number of nodes being equal to the number of images that can see that location. At each node, we allocate N bytes for the spectral signatures extracted from the corresponding satellite image, where the value of N is declared in the header segment

⁶ Most of the satellite images in this dataset contain four band data: Red, Blue, Green, and Near Infrared (NIR). We plan to incorporate satellite images with arbitrary number of bands in the next version of PIMSIR.

⁷ The satellite images used for the experiments in this paper were orthorectified by Ryan Smith and Nathan Campbell of GeoEye using the functionality provided by the publicly available GDAL library. The orthorectification formulas that convert pixel coordinates of the raw satellite images into lat/long coordinates utilize a DEM (Digital Elevation Map) model of the portion of the earth's surface that corresponds to the satellite image. These formulas are iterative and involve various approximations for warping the image data onto the earth's surface, in addition to the interpolation needed to create a uniformly sampled array in the lat/long coordinates. The errors in orthorectification computations are exacerbated by any errors in the DEM model. For this study, the GMTED2010 DEM at 7.5 arc-sec resolution was used. One could achieve better orthorectification using carefully selected ground control points, but this was not done for the images used in this study. The ToA reflectance correction is based on an implementation of the algorithm in [Udipi and Comp \(2010\)](#) by Craig Stutts of ARA.

of the file that carries the .pimsir suffix. In our current implementation, we use $N = 17$ bytes per satellite image. Four of these bytes are reserved for the spectral value in each of the four primary bands, and one byte reserved for the pointer to a file that contains the meta data for that satellite image.⁸

For an ROI of size, say, 10,000 km², the resulting representation may take a couple of hundred gigabytes - the precise value depending on how many satellite views need to be accommodated. Should this amount of memory not be available as RAM, we can always resort to using the virtual memory address space with the help of POSIX functions like `mmap` with some loss of performance with regard to the time taken for construction of the integrated representation.

Given the PIMSIR representation for an ROI, one can then investigate how the data varies with respect to any of the independent variables such as the view angle, the time of the year, etc., as the reader will see in the rest of this paper. What is even more significant, this representation allows for fast construction of probability distributions of special signatures over an ROI. Subsequently, these distributions can be plugged directly into various machine learning algorithms for the learning of classification boundaries.

2.1. Viewing image overlaps with PIMSIR

PIMSIR makes it very convenient to view how much data is available in the different subregions of an ROI. This is an important issue since the applicability of the different types of approaches for ROI-based pixel and object classification depends on the extent of the data that is available at the ground level. To elaborate, suppose all of the geographical points in a subregion were covered by, say, a single satellite view, that subregion would not lend itself to the application of multi-view logic as described in [Section 6](#).

Using the PIMSIR representation, this section shows the data variability results for the Chile ROI. PIMSIR for the Chile ROI was constructed using all 189 satellite images that intersect with this ROI. Just for the purpose of visualization, what is displayed in [Fig. 1](#) is the RGB part of the spectral signature in just the first satellite image at each point of the ROI-based bounding box. The satellite images relevant to each sampling point are sorted by their names (that, in general, may be considered to be arbitrary strings). We refer to the display that is constructed from all the first images in this sorted list at each sampling point as the '1-overlap display' (in order to distinguish it from a more general N -overlap display to be shown presently). For the N -overlap display, we show pixels at only those geographic sampling points where there exists data in all of the first N overlapping images in the sorted list - the pixel value chosen for display is the RGB value in the N th image. It is in this manner that [Fig. 2](#) shows the first 16 overlap displays. The image at the upper left corner is for the 1-overlap display; it is the same image that was shown earlier in [Fig. 1](#). The second image in the top row is for the 2-overlap display, and so on. Note that what's being displayed at each sampling point in an N -overlay display for any N has no bearing whatsoever on the data variability calculations at the point.

The blocky artifacts in [Fig. 2](#) (and in [Fig. 1](#) also) are caused by several factors: (1) The image-to-image variability that can be attributed to the different look angles for the sensors aboard the satellites, sun angle variations, and the time of the year when the data was recorded; (2) Small errors in the application of the Top-of-Atmosphere correction to the pixels; and (3) small errors in the image rectification process.

⁸ This obviously limits us to a maximum of 256 overlapping satellite images for any geo-point. We have yet to see a case where that condition would be violated. Nonetheless, in order to "future-proof" PIMSIR, we plan to use two bytes for the pointer to the metafile in the next version of PIMSIR.

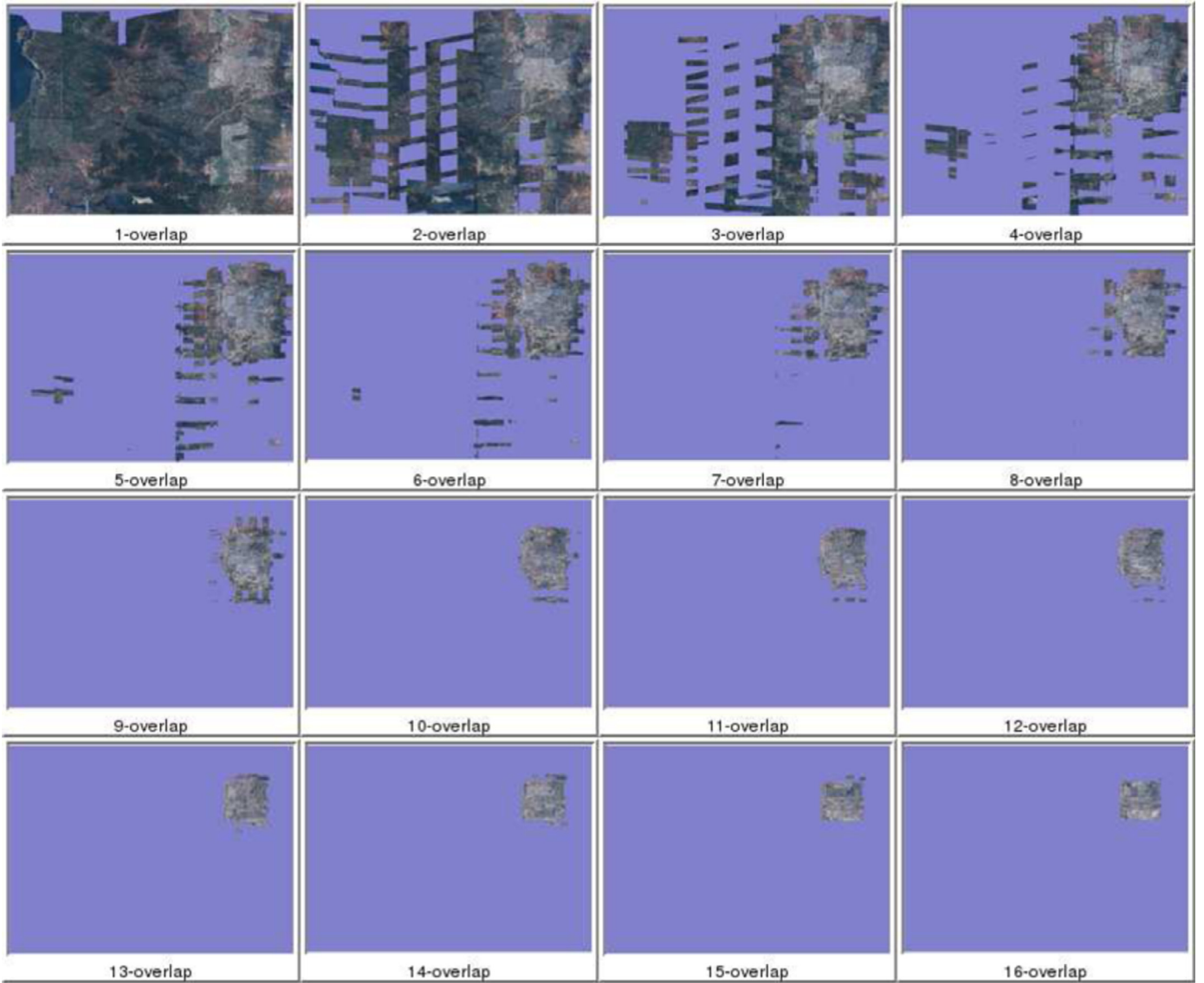


Fig. 2. A collage of the first 16 (out of 49) overlaps: 1-overlap, 2-overlap, ..., 16-overlap.

When we examine the overlaps in all of the satellite images for the Chile ROI, we go from one extreme where there is only a single image at a given geo-point to the other extreme where we have 49 satellite images looking at the same geo-point. Since it would occupy too much space to show all of these overlaps, in what follows we will show the results for just the case of 49 overlaps. Fig. 3 shows the portions of the Chile ROI where we have data from 49 images. At each cell of the ROI-based bounding box, we arbitrarily selected the RGB from one of the 49 satellite images for constructing the display in Fig. 3. Note again that the goal of this display is merely to indicate where we have a total of 49 views looking at the same geo-point. The spectral variability results that are shown in the next subsection correspond to this case of the geo-points covered by these 49 images.

2.2. Viewing data variability heat maps with PIMSIR

We now show results of our spectral variability investigation over the set of geo-points depicted in Fig. 3. Recall, we have a total of 49 satellite images contributing MSI data to each of these points. We will report on the data variability as produced by the following three sources:

1. Seasonal changes
2. View angle changes
3. Content changes

However, before showing the variability as caused separately by each of these three sources, we will first show the overall data variability through what we refer to as variability heat maps. Fig. 4 shows an example of such a heat map. The values depicted in Fig. 4 correspond to maximum spectral range, r_{max} , at each point. This parameter is calculated using the formulas:

$$r_i = \max(SS_i) - \min(SS_i) \quad (1)$$

$$r_{max} = \max_i(\{r_i\}) \quad (2)$$

In these formulas, SS_i is the set of 49 spectral measurements for a band i at a given point. The value $\max(SS_i)$ is the maximum of 49 such band values at a given point and $\min(SS_i)$ is the minimum of the same set of spectral signatures. Therefore, the value of r_i is the largest variation within the i th band across all 49 satellite images. And the largest of these variations among the four spectral bands is the maximum range, denoted r_{max} . Note that these calculations are carried out after the spectral values are normalized to

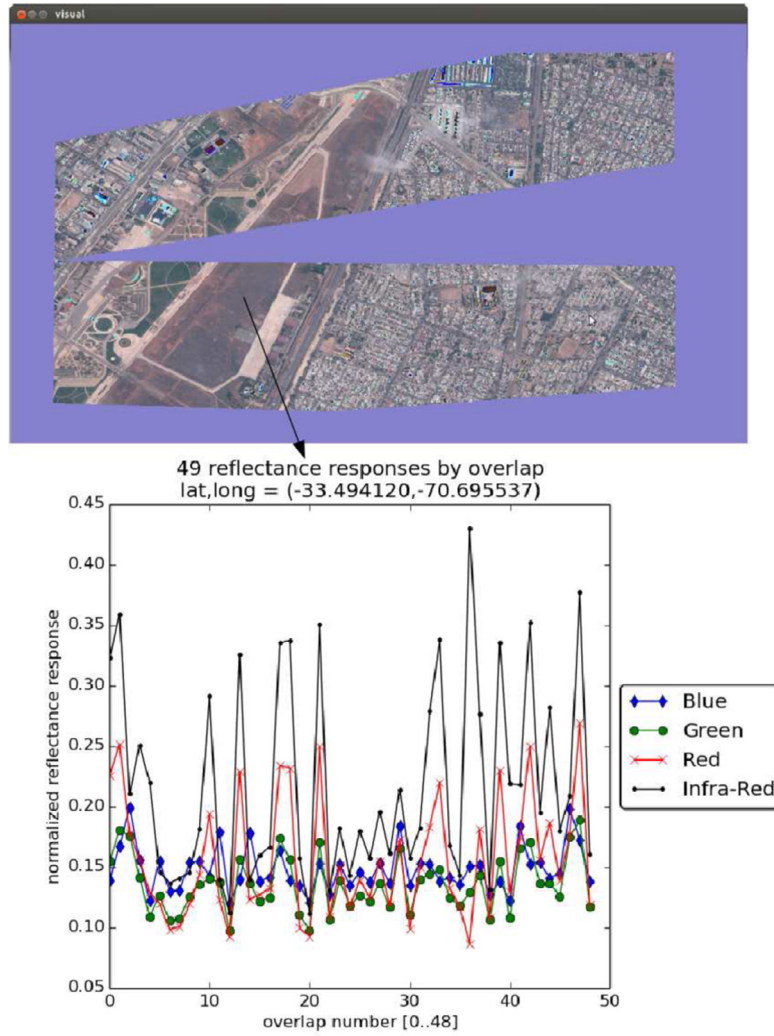


Fig. 3. The 49-overlap display shows the area with 49 overlapping satellite images. What is shown in the lower plot are all the spectral signatures (that is, the Top-of-the-Atmosphere reflectance measurements) for a particular geo-location within the area. The full range of the Y-axis is [0,1].

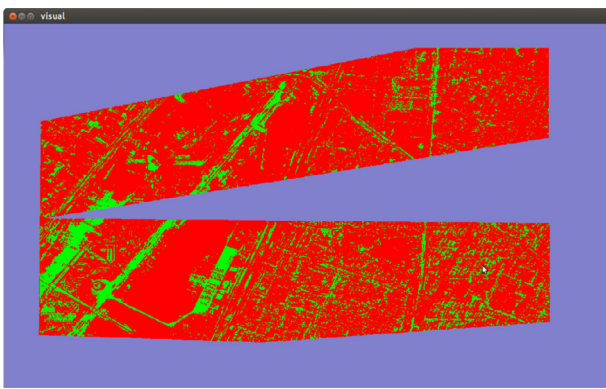


Fig. 4. Variability heat map for 49 overlapping satellite images in the region shown in Fig. 3: Green = maximum range $r_{\max} < 0.15$.

lie between 0 and 1.0. The data normalization for each image is a part of what is accomplished by the Top-of-Atmosphere correction mentioned earlier in Section 2. The red areas in Fig. 4 correspond to the heat map values above 0.15. On account of the data normalization, this means that we have more than 15% within-band variation at the points that are shown as red in Fig. 4. The points that

are shown as green have their within-band variability under 15%. Rapid visualization of the variability allows us to systematically try many different cutoff thresholds and, in this case, 0.15 was chosen as it produced the most visually informative heatmap.

Comparing the images in Figs. 3 and 4, one can infer that the orthorectification errors in the 49 satellite images are sufficiently small and can be ignored - *this is an important by-product of this study since one is always concerned about the quality of orthorectification*. If this error had been large, the green strands in Fig. 4 would not correspond to the roads in Fig. 3. Also note that the spectral signatures are significantly “invariable” across the 49 images for the road pixels. In other words, for this particular area in the ROI, the points of the earth’s surface that correspond to the roads produce the same spectral signatures regardless of when the satellite image was taken and the look angle for the image.

We can draw the same conclusion in urban areas. Figs. 5 and 6 show an urban area covered by 25 overlapping satellite images near the center of the city of Santiago. This area contains many tall buildings. While the major roads in Fig. 5 are easily visualized in the heat map of Fig. 6, notice also the lining up of some of the finer detail in the heat map and how it matches the scene structure shown in Fig. 5. This leads us to conclude that even in dense urban areas the orthorectification errors are not so large as to render our integrated representation completely useless. It would be

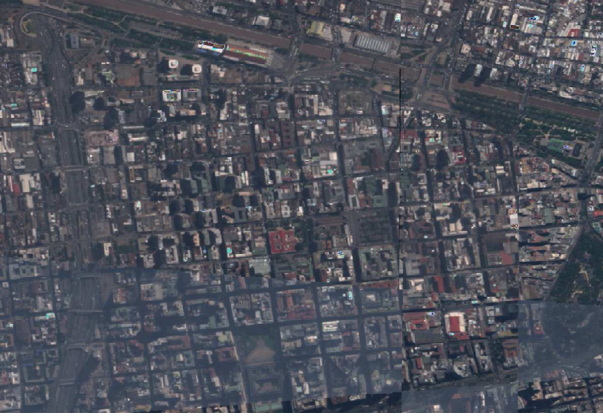


Fig. 5. A dense urban area containing many tall buildings in the city of Santiago in Chile. The area is centered at lat-long: $(-33.4372, -70.6526)$. This area is covered by 25 overlapping satellite images taken at different times and from different view angles.

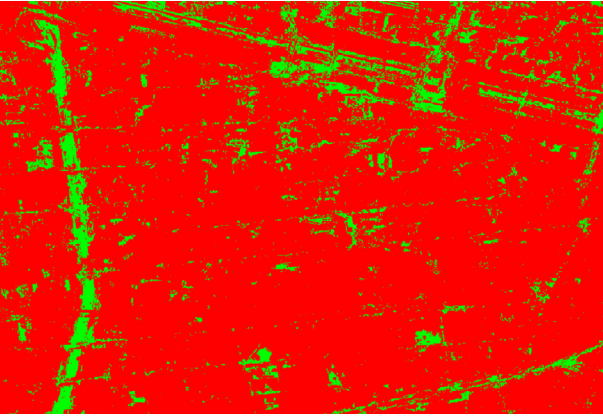


Fig. 6. Variability heat map for the dense urban area shown in Fig. 5. While the major roads are clearly visible in the heat map, note also the lining up of several pixels in the heat-map in a way that corresponds to the finer structure in Fig. 5. The variability that is visible can be attributed to the occlusions caused by tall buildings. Green = maximum range $r_{max} < 0.15$.

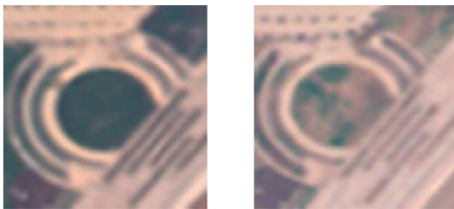


Fig. 7. Examples of variation due to seasonal change. The same geo-location may need to be labeled as light vegetation and soil at different times of the year.

safe to attribute the variability that is visible in Fig. 6 to the different occlusions caused by tall buildings in the satellite images from different view angles.

Fig. 7 shows an example of data variation in a vegetation covered area as caused by seasonal change and Fig. 8 shows variation due to view-angle change. For the latter case especially, even when the actual land-cover remains the same, the occlusions created by tall structures cause the spectral signatures collected for a geo-point to vary from one image to another. Data variations may also be caused by a piece of land being under development during the time when the images were recorded. In this case, the same area can completely change from one land type to another. Fig. 9 shows such an example.



Fig. 8. Example of data variation due to view-angle change. In this case, the small areas in the middle of the two image patches shown should correspond to the same geo-location. But due to different view angles, the area in the right image is occluded by part of the tall building that can be seen in the left image. This results in different spectral signatures for the same geo-point.

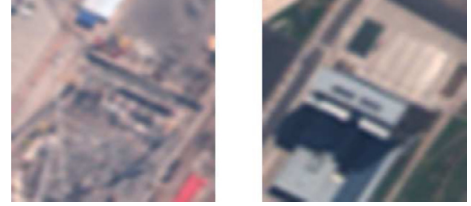


Fig. 9. Example of land-cover change caused by urban development. The area in the middle of the image patches shown changes from road/parking lot (left image) to building (right image).



Fig. 10. The three image patches cropped from three different satellite images centered at the same geo-location over a road surface.

While all the previous examples illustrated data variability caused by different sources, it's also good to know that there can exist areas where the data stays constant, more or less. Fig. 10 shows the same road area in three different images (selected randomly from the 49 satellite images for the subregion shown earlier in Fig. 3). In Fig. 11 we plot the spectral signatures with respect to the view angles and with respect to time at a particular geo-location at the center of the three image patches in Fig. 10.

The left plot of Fig. 11 shows the variability with respect to the look angle and the right shows the variability with respect to the month in which the data was recorded. As the reader can see, there is much less variation in the data for the point chosen. This result is consistent with the overall data variability heat map shown in Fig. 4.

3. An ROI-based protocol for generating the ground-truth for training and testing a classifier

We now present a data annotation procedure for creating positive examples for the different land-types for the training of the classifiers presented in the next section. Since ROI based wide-area learning involves a very large amount of data, and since it would be much too frustrating for a human to have to scan all of the images that cover an ROI for identifying the positive examples, the procedure we present samples the ROI and throws up small patches, along with their contexts, for a human to see and categorize.

The system starts out by showing the entire ROI (as, for example, shown in Fig. 1) to the annotator. The annotator can zoom in and out in order to become familiar with the general content of the ROI. At any given geo-point, the annotator can also scan across

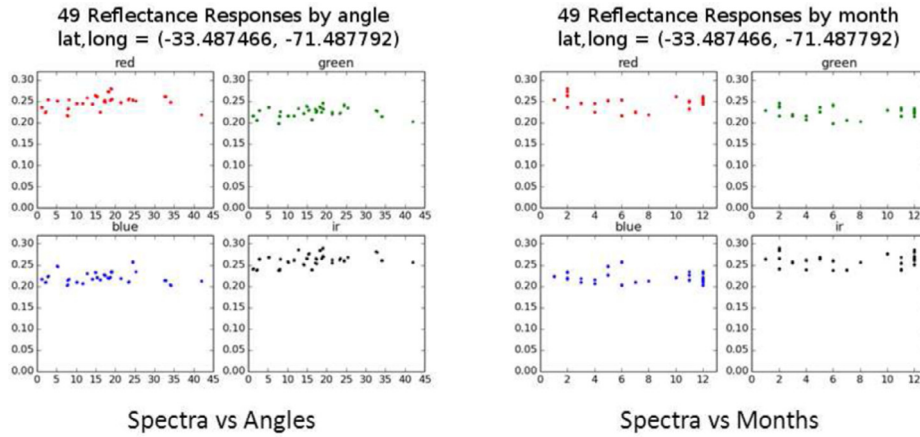


Fig. 11. Road spectral signature versus the off-nadir view angle and versus the month of the year in which the image was recorded.

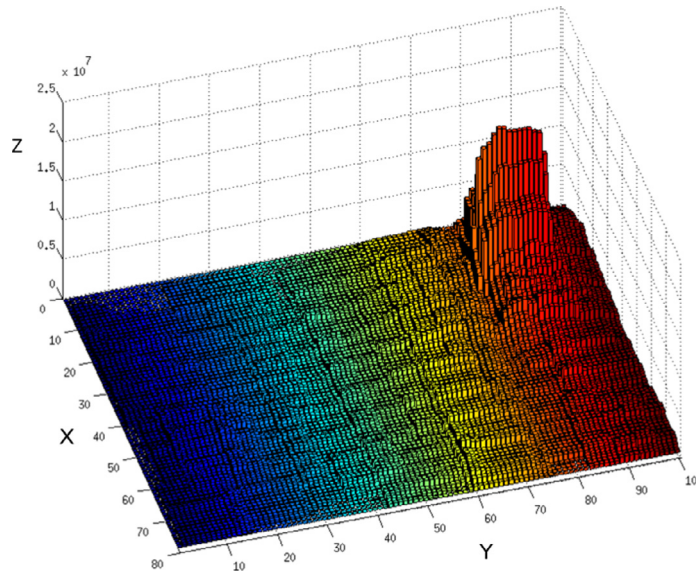


Fig. 12. An example of the satellite image overlap distribution used in a sampler based on the Metropolis-Hastings algorithm. This distribution is for the Chile ROI. The X and Y axes have units in km and the Z axis indicates the number of all 2m-by-2m cells from all overlapping images in a given 1km-by-1km area.

all of the overlapping views. Subsequently, the system throws up M randomly chosen subregions for each land-type for the human to annotate, M is typically a small user-specified integer. (We used $M = 3$ for the experimental results in Section 7.) The size of such subregions is set at system initialization time.⁹ The randomization is carried out by the Metropolis-Hastings algorithm that is applied to the image coverage histogram of the sort shown in Fig. 12.¹⁰

The annotator then visually examines the geographical content in the vicinity of the drawn subregion and is given the option of rejecting the subregion if it lacks the land-type for which the annotator is seeking positive examples.

For each of the subregions thus accepted, the system randomly generates up to 20 1km-by-1km patches. To give the reader a sense of the size of these patches, for 2m/pixel data, each patch is represented by a 500×500 array of pixels that can easily be displayed

fully in a typical 1024×768 computer monitor display. This allows the annotator to clearly see and recognize the land-types within a patch and its vicinity. The annotator can also zoom into a patch in order to resolve any ambiguities regarding the land-type of a pixel or a group of pixels. When the annotator sees a blob of pixels in which all pixels appear to possess the same land-type, he/she can draw a polygon around the blob and mark the entire blob as constituting a set of positive instances for that land-type. The annotator can also quickly check the overlapping satellite images and decide whether to mark the blob within the same polygon in all the overlapping images as well. In this manner, the annotator collects around 2000 samples from the images within the patches and their vicinity in each subregion for each land class. When the parameter M is set to 3, that makes for a total of 6000 positive instances for each land-type collected from three subregions.

Obviously, one must exercise some care with regard to the size of the subregions used. If this size is too small, the samples collected may not capture sufficient variability needed for properly training a classifier. Fig. 13 shows some marked subregions within the Chile ROI that were used for training the classifiers described in the next section. The subregions shown in the figure are of size 64 square kilometers, which is roughly 25% of the size of a typical satellite image. The size of a typical satellite image is about 10000

⁹ As we will mention in greater detail later, we typically set the size of a subregion to be approximately 64 square kilometers.

¹⁰ Drawing samples according to the image coverage distribution yields a larger amount of training data in less time. The greater the number of overlaps at a geo-point, the larger the number of training samples that can be labeled at once at that geo-point. For Metropolis-Hastings sampling, we used the Perl-based implementation presented in Kak (2014).



Fig. 13. The subregions (shown in white outlines, each of size 8km-by-8km) used in the Chile ROI for extracting the positive instances for the six different land-types. They include: water, active crop field, dormant crop field (aka soil), building, road, and tree (aka forest). There are three subregions per land-type and a subregion may be shared by more than one land-type.

$\times 7000$ pixels (or $20\text{km} \times 14\text{km}$) at 2 meters per pixel spatial resolution.

4. A probabilistic framework for modeling within-class and between-class variations

In this section, we describe in detail the Bayesian framework that we have developed for land-type classification. Instead of directly using feature vectors for classification, we model the variations in the feature space for each class and use these models for land-type classification. The intuition behind this approach is that different classes exhibit different variations in their spectral signatures (and therefore in any derived feature space). Our results presented in Section 8 show that this reasoning is well-founded.

We group the variations into two categories, within-class and between-class variations. To understand this distinction, without loss of generality, consider one particular class of land-type, for example ‘Trees’. One can definitely expect inter-image and inter-view differences in spectral signatures for pixels belonging to this class due to different categories of trees, different angles of satellite views, seasonal changes and so on. For example, perpetually green trees such as pine trees are likely to produce the same spectral signature across seasons, whereas other trees are more likely to exhibit large variations in their spectral signatures in different seasons. We refer to such variations as *within-class variations*. Correspondingly, differences in the spectral signatures between two different land-types, such as between buildings and trees, are referred to as *between-class variations*. It is important to realize that the nature of within-class variations itself might be different from class to class.

We chose a Bayesian framework for classification as it offers a practical and powerful way of modelling both within-class and between-class variations in a single unified framework. In addition, as mentioned earlier, we aim to develop a system that can fuse information from multiple views and the Bayesian framework easily lends itself to fulfilling this objective.

The foundations of our Bayesian framework lie in the prior contributions by Moghaddam Moghaddam et al. (2000), Moghaddam et al. Moghaddam (2002), and Aksoy and Haralick Aksoy and Haralick (2000; 2001). In Moghaddam (2002); Moghaddam et al. (2000), the authors use the difference in feature vectors for face recognition and show good performance. In Aksoy and Haralick (2000; 2001), the authors use a similar probabilistic approach for image retrieval. They use a likelihood-based similarity measure to

classify the difference between two feature vectors (images) as ‘relevant’ or ‘irrelevant’. The authors demonstrate better performance using this metric when compared to a geometric-distance based metric. In those works, a difference between two feature vectors can belong to one of two classes, a relevant or irrelevant class. For example, in Moghaddam (2002); Moghaddam et al. (2000) all variations between any two images are grouped together irrespective of the identity of the human subject.

Our Bayesian framework differs from these prior contributions in that we model the probability distributions of the within-class and between-class variations for each land-type class separately. This helps us to capture the intuition that the nature of variations themselves could be different for each class. For example the variations within the ‘Trees’ class due to seasonal changes or due to different varieties of trees are very different from the variations within the ‘Buildings’ class due to the usage of different construction materials.

Before discussing the details of our Bayesian framework, in what follows, we will first present details about the feature space and the target classes used for classification. We will also introduce the symbolic notation used later in this section.

Feature space - The Normalized Difference Vegetation Index (NDVI) was developed specifically to identify and characterize vegetation in aerial images DeFries and Townshend (1994); Kerdiles and Grondona (1995). Essentially it exploits the difference in how plants respond to incident light in the photosynthetically active region (PAR) and infrared region. The index is based on easy-to-compute spectral band ratios that show good performance for classifying vegetation. More recently in Marchisio et al. (2010); Wolf (2010) the authors have extended the NDVI like features and applied them for general land-type classification. We use these features for our framework. In subsequent discussions, we will refer to these features as Band Difference Ratios (BDR). Our dataset consists of 4-band satellite images (Red, Blue, Green, and NIR), which yields a 10-dimensional feature vector for each geo-point. The 10 dimensions correspond to the six band ratios (two bands taken together at a time) plus the four spectral signatures themselves.

Since each geo-point in our PIMSIR representation can receive data from multiple overlapping views, each geo-point can have multiple 10-dimensional feature vectors associated with it. This is a significant departure from conventional methods that focus on selecting the best visually representative satellite image for each geo-point. Our probabilistic framework enables us to easily fuse data from all available views at the same time.

Target classes - We use six land-type classes, namely Buildings, (Asphalt) Roads, Active Crop Field, Soil, Water, and Trees. The class ‘Active Crop Field’ refers to both crop fields and areas that contain grass and shrubs. Obviously the framework can be extended to an arbitrary number of classes. We have selected classes that one would expect in any meaningful attempt at land-type classification.

Similarity measures for classification - In general, Bayesian classifiers make a probabilistic assessment of the class assignment of a test sample based on its “distance” from the distributions for the target classes. In this paper we have investigated two different similarity measures and analyzed their performance on land-type classification. In one case, we only model the within-class variations, and, in the second, we use a similarity measure that incorporates both within-class and between-class variations.

Notation - We denote a feature vector by \vec{F} . The difference between two feature vectors \vec{F}_a and \vec{F}_b will be represented by $\vec{\Delta}_{ab} = \vec{F}_a - \vec{F}_b$. C_i^W will denote the within-class variations for the i th class and C_i^B will denote the between-class variations for the i th class (ie., the variation between class i and all the other classes taken together).

4.1. ML and MAP classifiers using only within-class variations

For this case, we do not take the between-class variations into account. We only model the within-class variations. The underlying assumption is that any given $\vec{\Delta}_{ab}$ must belong to one of the C_i^W for some i .

Training - For each class, we model the probability distribution of the within-class variations as a unimodal multivariate Gaussian distribution with a zero-mean and an estimated covariance matrix. Given two feature vectors \vec{F}_a and \vec{F}_b that belong to two geo-points of the same class i , both $\vec{\Delta}_{ab} = \vec{F}_a - \vec{F}_b$ and $\vec{\Delta}_{ba} = \vec{F}_b - \vec{F}_a = -\vec{\Delta}_{ab}$ belong to the within-class variations of class i . Therefore the probability distribution of the within-class variations will have a zero mean value. Obviously one can use more complex distributions than the Gaussian distribution to model the variations. Our experimental results validate our assumption of a Gaussian distribution. The covariance matrix for each class is estimated from its positive training samples. For the i th class, the probability distribution of the within-class variations is denoted as

$$p(\vec{\Delta}|C_i^W) = \mathcal{N}(0, E_i^W), \quad (3)$$

where E_i^W denotes the estimated covariance matrix for the within-class variations of class i and C_i^W is as described in the Notation in Section 4.

Testing - For a candidate geo-point P_k , we compute the 10-dimensional feature vector for each available overlapping view. For each such view, we then compute the difference between the corresponding feature vector of P_k and each feature vector in our training database. The maximum a-posteriori probability that one such difference vector $\vec{\Delta}$ belongs to the within-class variations of class i is given by

$$S_i(\vec{\Delta}) = P(C_i^W|\vec{\Delta}) = \frac{P(\vec{\Delta}|C_i^W)P(C_i^W)}{P(\vec{\Delta})} \quad (4)$$

Recall that in this case we model only the within-class variations. Using the above stated assumption that $\vec{\Delta}$ must belong to one of the C_i^W for some i , the denominator in Eq. 4 is given by

$$P(\vec{\Delta}) = \sum_i P(\vec{\Delta}|C_i^W)P(C_i^W) \quad (5)$$

Since the denominator term in Eq. 4 is the same for each class, we do not have to compute it explicitly. If we assume equal prior probabilities for each class, then Eq. 4 reduces to a Maximum Likelihood measure as shown below.

$$S_i(\vec{\Delta}) = P(\vec{\Delta}|C_i^W). \quad (6)$$

We compute this probability for each training feature vector from each class. The class label assigned to this view of P_k is then the class label of the training feature vector that maximizes this probability.¹¹ Since a geo-point can have multiple overlapping views and thus multiple feature vectors, it is possible to assign multiple class labels to the same geo-point. Thus an interesting problem with using multi-view data is to find an optimum way to combine multiple class labels from multiple feature vectors for a particular geo-point. In Section 6, we present a practical and novel approach to handle this problem.

Shown in Algorithms 1 and 2 is a brief algorithmic description for designing the classifier based on only the within-class variations.

¹¹ Since we are modeling the within-class variations as unimodal Gaussians, this calculation is done efficiently through the computation of Mahalanobis distances using the covariances for the different classes. Note that this cannot be done for the case where we model both the within-class and between-class variations.

Algorithm 1 Training using Within-Class Variations.

M_i is a matrix whose columns correspond to the spectral signatures of the samples of the i^{th} target class.

```

1: procedure TRAINWITHINCLASS
2:   for  $i \in \{\text{Target Land Classes}\}$  do
3:     Compute 10-D feature vectors from  $M_i$ 
4:      $\Delta_i \leftarrow$  A matrix of pairwise differences between feature
       vectors. Each column of  $\Delta_i$  is one such difference vector.
5:      $N_i \leftarrow$  number of columns of  $\Delta_i$ 
6:      $E_i^W \approx \frac{1}{N_i} \Delta_i \Delta_i^T$ 
7:     Calculate  $E_i^{W-1}$  and  $\det(E_i^W)$ 
```

Algorithm 2 Testing using Within-Class Variations.

P_k is a test geo-point. M_k^o is the spectral signature for P_k from the o^{th} overlapping view.

```

1: procedure TESTWITHINCLASS
2:   for  $o \in \{\text{Overlapping views}\}$  do
3:      $\vec{F}_k^o \leftarrow$  Compute 10-D feature vector from  $M_k^o$ 
4:     for  $\vec{F}_i \in \{\text{Training feature vectors}\}$  do
5:        $\vec{\Delta}_{k,i}^o \leftarrow \vec{F}_k^o - \vec{F}_i$ 
6:        $S(\vec{\Delta}_{k,i}^o) \leftarrow P(\vec{\Delta}_{k,i}^o|C_i^W)$ 
7:        $\text{label}_k^o \leftarrow \text{argmax}_i S(\vec{\Delta}_{k,i}^o)$ 
8:    $\text{label}_k \leftarrow$  Combine  $\text{label}_k^o$ 's as described in Section 6.
```

Note: Repeat Algorithm 2 for each P_k in the test data set. Additionally, we can use matrix operations to implement the above algorithm using a single for-loop.

4.2. ML and MAP classifiers using both within-class and between-class variations

In this strategy, we incorporate both within-class and between-class variations in the similarity measure. We again model the between-class variations as unimodal multivariate Gaussian distributions. One could argue that between-class variations are better represented by a more complex multi-modal distribution. For this paper, as a starting point, we use the simpler unimodal distribution for computational convenience. However, extending this part of the classifier design to multi-modal distributions for between-class variations is one of our future goals.

Training - The within-class variations are modeled as in the previous case. The probability distribution of the between-class variations for the i^{th} class, is denoted as

$$p(\vec{\Delta}|C_i^B) = \mathcal{N}(0, E_i^B), \quad (7)$$

where E_i^B denotes the estimated covariance matrix for the variations between the feature vectors of class i and those of the other classes and C_i^B is as described in the Notation in Section 4.

Testing - For each view of a test candidate geo-point P_k we compute the difference between the corresponding feature vector of P_k and all the training feature vectors. The MAP probability that one such difference vector $\vec{\Delta}$ belongs to the within-class variations of class i is defined as

$$S_i(\vec{\Delta}) = P(C_i^W|\vec{\Delta}) = \frac{P(\vec{\Delta}|C_i^W)P(C_i^W)}{P(\vec{\Delta})} \quad (8)$$

Now comes the key difference between this strategy and the previous strategy. Since we explicitly model the between-class variations, therefore $\vec{\Delta}$ can either belong to the within-class variations of class i or the between-class variations of class i . Hence we

can rewrite the similarity measure as

$$S_i(\vec{\Delta}) = \frac{P(\vec{\Delta} | C_i^W) P(C_i^W)}{P(\vec{\Delta} | C_i^W) P(C_i^W) + P(\vec{\Delta} | C_i^B) P(C_i^B)} \quad (9)$$

To this view of P_k we can then assign the class label of the training feature vector that maximizes the above probability.

However, it is possible that P_k does not belong to any of the classes under consideration. Thus we can go a step further with the observation that for $\vec{\Delta}$ to actually be a within-class variation of class i , we require

$$P(C_i^W | \vec{\Delta}) > P(C_i^B | \vec{\Delta}) \quad (10)$$

Using Eq. 10, Eq. 9 reduces to

$$S_i(\vec{\Delta}) > \frac{1}{2} \quad (11)$$

If the above condition is not satisfied for any class i , then geo-point P_k can be assigned a label ‘Other’. Note that this could be used to identify geo-points that belong to unmodeled classes such as clouds.

We use the majority voting notion as mentioned in Section 6 to combine the information from multiple class labels.

We present a brief algorithmic implementation of the above described strategy in Algorithms 3 and 4.

Algorithm 3 Training using both Within-Class and Between-Class Variations.

\mathbf{M}_i is a matrix whose columns correspond to the spectral signatures of the samples of the i^{th} target class.

- 1: **procedure** TRAINBETWEENCLASS
 - 2: Estimate E_i^W 's as in Algorithm 1
 - 3: **for** $i \in \{\text{Target Land Classes}\}$ **do**
 - 4: Compute 10-D feature vectors from \mathbf{M}_i .
 - 5: Calculate all possible pairwise differences between the \vec{F} 's belonging to class i and the \vec{F} 's belonging to the remaining classes.
 - 6: Create matrix Δ_i such that each column of Δ_i is one such difference vector.
 - 7: $N_i^B \leftarrow$ number of columns of Δ_i
 - 8: $E_i^B \approx \frac{1}{N_i^B} \Delta_i \Delta_i^T$
 - 9: Calculate $E_i^{B^{-1}}$ and $\det(E_i^B)$
-

Algorithm 4 Testing using both Within-Class and Between-Class Variations.

P_k is a test geo-point. M_k^o is the spectral signature for P_k from the o^{th} overlapping view.

- 1: **procedure** TESTBETWEENCLASS
 - 2: **for** $o \in \{\text{Overlapping views}\}$ **do**
 - 3: $\vec{F}_k^o \leftarrow$ Compute 10-D feature vector from M_k^o
 - 4: **for** $\vec{F}_i \in \{\text{Training feature vectors}\}$ **do**
 - 5: $\vec{\Delta}_{k,i}^o \leftarrow \vec{F}_k^o - \vec{F}_i$
 - 6: $S(\vec{\Delta}_{k,i}^o) \leftarrow P(C_i^W | \vec{\Delta}_{k,i}^o)$
 - 7: $label_k^o \leftarrow \text{argmax}_i S(\vec{\Delta}_{k,i}^o)$
 - 8: $label_k \leftarrow$ Combine $label_k^o$'s as described in Section 6.
-

Note: Repeat Algorithm 4 for each P_k in the test data set. Additionally, we can use matrix operations to implement the above algorithm using a single for-loop.

5. Support vector machines for classification

Since the variance based Bayesian framework presented in Section 4 is novel - novel in the context of satellite imagery - we must compare it with the current practice in satellite image classification. Considering that Support Vector Machines (SVM) are widely used for classifying satellite data, our Section 8 will compare the wide-area classification results obtained with our Bayesian framework and with SVM. As we will show in that section, using the same training and testing data sets, the Bayesian framework of Section 4 outperforms SVM for four out of six classes. This section provides a brief overview of how we have used SVM in our comparison studies.

Note that, in contrast to our variance based framework in which classification rules are applied to the differences of feature vectors, SVM as popularly used are meant to be applied directly to feature vectors. We should also point out that since we apply the ToA reflectance correction to the data (Section 2), the resulting reflectance values are already normalized to the 0–1 range. By their definition, the BDR features that we extract have values between –1 and 1. Thus additional data normalization is not required for the SVM classifiers.

We trained two different SVM classifiers, one with a Radial Basis Function (RBF) kernel and the other with a linear kernel. For multi-class classification, the former uses a “one-against-one” approach Kner et al. (1990). Thus if we have n classes, the system uses $\frac{n(n-1)}{2}$ classifiers, each of which acts on two classes. For the linear kernel SVM, we use a “one-versus-the-rest” approach. We discuss the performance of our Bayesian classifier vis-a-vis both types of SVM classifiers in Section 8.

6. Combining multi-view class labels - compatible classes

For the reasons discussed in Section 2, we can expect that the different satellite views of the same geo-location will yield different classes for some of the points on the earth. While some of these variations in the class labels can be attributed to classification errors, others are due to genuine seasonal differences in the land-cover types. We now present the notion of “Compatible Classes” for dealing with the latter case.

Say that for a particular geo-point, two different views give rise to the two labels ‘Soil’ and ‘Active Crop Field’. We can think of them as compatible class labels. Such geo-points can be classified as belonging to a super-class ‘Agricultural Land’. Choice of such compatible classes can be based on prior knowledge about the geographic area under consideration. For example, in Taiwan where paddy fields are very common, ‘Water’ can be considered as compatible with ‘Active Crop Field’. The notion of compatible classes enables us to bring in information from meta-data, such as the time of year when the satellite images were taken, in order to make better judgments on the compatibility of multiple class labels.

When the number of target classes is large, one could try to design a system that automatically learns such rules of compatibility. This would require collecting ground-truth labels in a more extensive and elaborate manner and then using data mining techniques to detect any regular/periodic relationships between different class labels. For example, seasonal alternations between the ‘Active Crop Field’ and the ‘Soil’ labels will exhibit periodicity that could be learned.

In cases where labels obtained from multiple views cannot co-exist, that is, when the labels are incompatible, ordinarily we apply majority voting to the labels from each view - unless the data at that location lends itself to the following logic: When several views are available at a geo-point, if the temporally earlier views

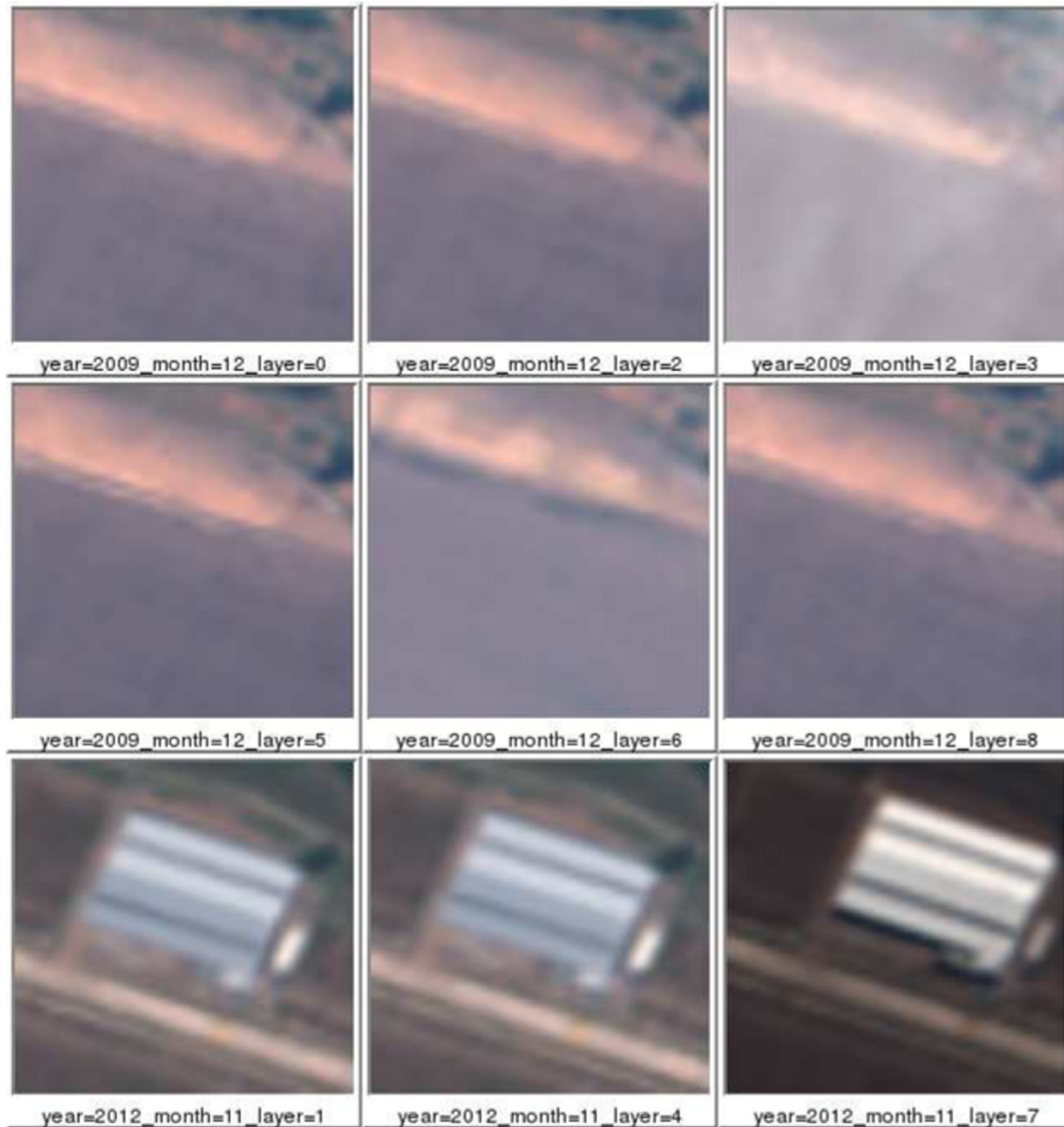


Fig. 14. 'Buildings' (last row) and 'Soil' (first two rows) are incompatible land-types. This particular location has nine overlaps. The overlaps were contrast-enhanced for better visualization and sorted by year and month.

consistently¹² give a vegetation label to a geo-point and the later views consistently call it a building or a road, we could safely assume that the land-type has changed permanently and we could just use the more recent label for that location.

Fig. 14 shows an example of incompatible land-types. In the example, 'Soil' and 'Buildings' are incompatible. Fig. 15 shows an example of compatible land-types. In the example, 'Soil' and 'Active Crop Field' can fall into a more general class such as 'Agricultural Land'.

The accuracy of using temporal consistency to combine labels will be affected by how well the orthorectified overlapping views line up together. This is especially true for pixels near the boundaries of objects as, for instance, the edge pixels of the building in Fig. 14. One can use additional higher level logic such as morphological operations and spatial filters to clean up such noisy pixels. In any case, from the standpoint of what is needed for geolocalization, the benefits of developing classifiers that are invariant to

inter-view and inter-image variations far outweigh the effects of such noisy pixels.

7. Experiments

Our main goal is to compare two different strategies for land-cover classification from satellite images: (1) We train the classifier using the data from a single satellite image and then test the classifier on a wide-area basis. And (2) We train the classifier using the wide-area training samples as made available by the Metropolis-Hastings sampler and then test the classifier on a wide-area basis.

We want to make the above mentioned comparison first with the variance-based Bayesian classifiers of Section 4 and then with the more conventional SVM classifiers of Section 5. For the variance-based Bayesian approach, we will also compare the results obtained using only within-class variations and the results obtained using both within-class and between-class variations.

For learning based on a single satellite image, we draw 2000 samples per class from a single image. For wide-area learning, on the other hand, generating the training and the test data is more complex and, as stated previously, we follow the ground-truth data

¹² By 'consistent' we mean that there are no temporal changes in the class labels.



Fig. 15. ‘Active Crop Field’ (overlap1 and overlap2) and ‘Soil’ (overlap0) are compatible land-types. They can be grouped into a more abstract class such as ‘Agricultural Land’. This particular location has three overlaps. The overlaps were contrast-enhanced for better visualization and sorted by year and month.

collection procedure described in Section 3. To make that explanation specific to the experiments reported in this section, we use three different subregions¹³ that do not overlap with the image used for the single-satellite-image based learning. That the subregions returned by the sampler not overlap with that single satellite image is important because the total ground truth used for wide-area learning contains all of the samples drawn from the three subregions, as picked by the Metropolis-Hastings sampler, and the samples from the single satellite image used for single-satellite-image based learning. As is the case for the single-satellite-image learning experiment, we draw 2000 samples per class from each of the three subregions, for a total of 6000 samples per class from the three subregions. When pooled with the 2000 samples per class drawn from the aforementioned single satellite image, we end up with a total of 8000 samples per class for wide-area learning.

For each classifier used, we use the stratified 10-fold cross-validation to assess its performance. Specifically, we divide the ground-truth data into 10 parts and use nine parts for training and one part for testing. This is done for all the ten ways in which the data can be partitioned in such a manner. Each part in the ten-part division of the ground-truth data consists of equal number of random samples drawn from each of the three subregions and from the single satellite image for each land-type. As a result, each classifier in each run of the 10-fold cross-validation test is trained with 7200 ($8000 \times (9/10)$) training samples from each class.

We also ran the Bayesian and SVM classifiers on a region from which no samples were collected for training the classifiers - we refer to this region as the “unseen region” and any data collected from this region as the “unseen data.” This region, presented in Fig. 16, is about 3.3km by 3.3km and is viewed in six satellite images whose intersections with the region are shown in Fig. 17. We assess the performance of the classifiers on this region both qualitatively (that is, visually) and quantitatively. The quantitative assessment was carried out with the help of a small ground-truth data set we created separately for this region. This dataset consists of roughly 2000 samples per class.

We use the same stratified 10-fold cross validation strategy discussed above for evaluating the SVM classifiers described in Section 5. Results for both the single-satellite-image and wide-area learning cases are presented in Section 8.

8. Results and comparisons

This section shows the results obtained with the different classifiers. As the reader will recall, our main goal is to compare the wide-area based learning of the classifiers with the single-satellite-



Fig. 16. An unseen region (3394m by 3356m) not sampled from the satellite images used for training the classifiers.

image based learning of the same. Our additional goal is to compare the performance of the variance-based Bayesian formalism for classification of Section 4 with the performance of the SVM approach outlined in Section 5.

In what follows, we will first take up the issue of wide-area learning versus single-satellite-image based learning in Section 8.1. This will be followed by a comparison of the Bayesian based classification with SVM in Section 8.2. Section 8.3 will present the results obtained on the unseen data as defined and specified in Section 7. Note that for the computation of average accuracies and confusion matrices, majority voting was used to combine class labels from overlapping views.

8.1. Wide-area learning versus single-satellite-image based learning

In this and the next subsection, we compare the performance of the classifiers based on the two different approaches outlined in Section 4. In one, we only consider within-class variations and, in the other, we also include the between-class variations.

8.1.1. Learning based on Bayesian modeling of within-class variations

Table 1 is the average confusion matrix obtained when we train on only one single satellite image. (Average is over the ten data sets in the 10-fold cross validation). We model only the

¹³ As mentioned toward the end of Section 3, a subregion is what is returned by the Metropolis-Hastings algorithm based sampler at each randomly selected point chosen by the sampler. It is roughly of size 64 sq km.



Fig. 17. All six overlaps in the unseen region. Images were contrast-enhanced for better visualization.

Table 1

Average Confusion Matrix using a Single Satellite Image for Training with only Within-Class Variations.

	ActiveCropField	Soil	Trees	Water	Buildings	Roads
ActiveCropField	60.875	0.0	39.087	0.0	0.0375	0.0
Soil	0.0	38.038	0.1125	0.0	59.975	1.875
Trees	1.113	0.0	97.838	0.0	1.050	0.0
Water	0.0	0.0	0.0	41.087	57.125	1.788
Buildings	0.55	1.613	0.475	0.4625	81.287	17.925
Roads	0.2125	0.25	0.0875	0.0	4.838	94.612

Table 2

Average Confusion Matrix using a Wide Area Region for Training with only Within-Class Variations.

	ActiveCropField	Soil	Trees	Water	Buildings	Roads
ActiveCropField	64.000	0.0	34.138	0.0	0.0	1.863
Soil	0.0	99.987	0.0125	0.0	0.0	0.0
Trees	0.0875	0.025	99.775	0.0	0.0	0.1125
Water	0.0	0.0	0.0375	99.938	0.0	0.025
Buildings	0.0125	6.362	0.0	0.0125	64.250	29.363
Roads	0.0875	0.625	0.0125	0.0	0.0375	99.237

Table 3

Average Confusion Matrix using a Single Satellite Image for Training with Both Between-Class and Within-Class Variations.

	ActiveCropField	Soil	Trees	Water	Buildings	Roads
ActiveCropField	62.188	0.0	37.763	0.0	0.05	0.0
Soil	0.0	37.175	0.0875	0.0	60.888	1.850
Trees	0.875	0.0	98.050	0.0	1.075	0.0
Water	0.0	0.0	0.0	38.888	59.587	1.525
Buildings	0.55	1.450	0.475	0.4625	82.412	16.962
Roads	0.225	0.25	0.05	0.0	5.188	94.287

Table 4

Average Confusion Matrix using a Wide Area Region for Training with both Between-Class and Within-Class Variations.

	ActiveCropField	Soil	Trees	Water	Buildings	Roads
ActiveCropField	67.125	0.0	31.962	0.0	0.0	0.9125
Soil	0.0	99.987	0.0125	0.0	0.0	0.0
Trees	0.125	0.025	99.763	0.0	0.0	0.0875
Water	0.0	0.0	0.0375	99.950	0.0	0.0125
Buildings	0.025	6.487	0.0	0.0125	66.338	27.137
Roads	0.1	0.8625	0.0125	0.0	0.075	98.950

within-class variations in this case. The average classification accuracy obtained was 68.88%.

Table 2 is the average confusion matrix obtained when we train on multiple satellite images that cover a wide area. (Average is over the ten data sets in the 10-fold cross validation). Again, we model only the within-class variations in this case. The average classification accuracy was 87.86%.

We observe good improvement in classification accuracy for five out of six classes, with significant improvement for the ‘Soil’ and ‘Water’ classes. We also see that the classifier exhibits some confusion between the ‘Active Crop Field’ class and the ‘Trees’ class. This is understandable as our feature space is derived from spectral signatures. Texture-based features and spatial features can be used in an additional step to separate these two classes. But it is encouraging that the vegetation classes exhibit minimal confusion with man-made classes such as buildings and roads. For the ‘Buildings’ class alone, we notice a decrease in classification accuracy. However, comparing Tables 1 and 2, we observe a decrease in confusion between the ‘Buildings’ class and the two vegetation classes. Confusion for the ‘Buildings’ class is now mostly with the ‘Roads’ class. One possible reason for this is the fact that the building rooftops

come in a wide variety of colors. When we collect samples for the ‘Buildings’ class from multiple regions, that increases the range of rooftop colors in the positive examples of buildings and this range can intersect with the spectral signatures in the positive examples for the roads. Since our feature space is purely spectral in nature, this overlap in the spectral signatures can be expected to cause confusion between some of the buildings and the roads. One solution to this problem would be to train on separate classes of buildings on the basis of the colors of the rooftops. In future extensions of the research reported in this paper, one could conceive of augmenting the spectral features with spatial features in order to better separate the building and the road classes Fauvel et al. (2013).

8.1.2. Also modelling the between-class variations

Table 3 is the average confusion matrix obtained when we model both within-class and between-class variations in classifier training based on a single satellite image. The average classification accuracy obtained in this case was 68.76%.

And Table 4 is the average confusion matrix obtained when we model both within-class and between-class variations in wide-area based classifier training. The average classification accuracy obtained in this case was 88.69%.

Table 5
Average Confusion Matrix Using a Single Satellite Image for an SVM classifier with RBF Kernel.

	ActiveCropField	Soil	Trees	Water	Buildings	Roads
ActiveCropField	82.450	0.0	17.550	0.0	0.0	0.0
Soil	0.1625	99.838	0.0	0.0	0.0	0.0
Trees	1.887	0.0	98.112	0.0	0.0	0.0
Water	0.0	0.0125	0.0125	95.213	0.0	4.763
Buildings	0.4125	10.088	0.0	0.25	81.862	7.388
Roads	1.113	8.363	0.275	0.0	8.088	82.162

Table 6
Average Confusion Matrix using a Wide Area Region for a SVM classifier with a RBF kernel.

	ActiveCropField	Soil	Trees	Water	Buildings	Roads
ActiveCropField	98.575	0.0	1.425	0.0	0.0	0.0
Soil	0.0	99.963	0.0375	0.0	0.0	0.0
Trees	6.862	0.025	93.112	0.0	0.0	0.0
Water	0.0	0.0	0.025	98.725	0.0	1.250
Buildings	0.2125	6.537	0.0125	0.625	84.025	8.588
Roads	0.5125	2.312	0.5375	0.0125	3.475	93.150

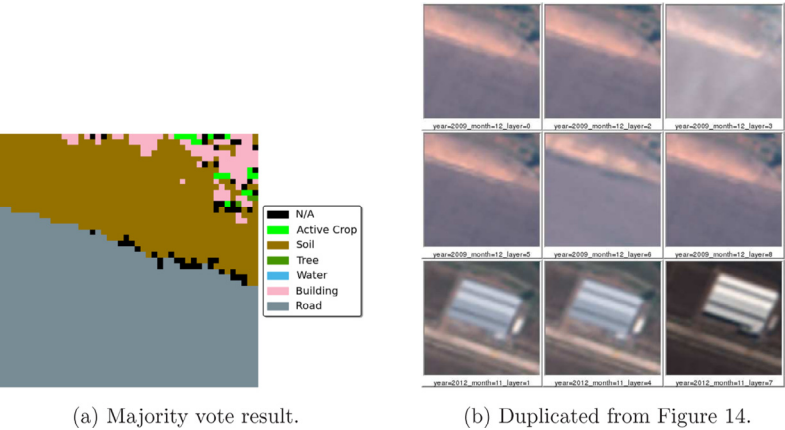


Fig. 18. Majority vote result from Fig. 14.

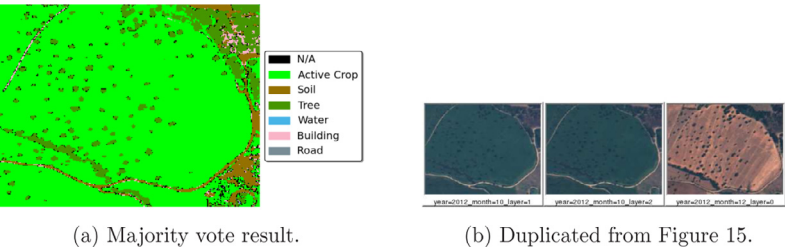


Fig. 19. Majority vote result from Fig. 15.

By comparing Tables 1 and 3, as well as Tables 2 and 4, we observe that we do not get any significant improvement by modelling the between-class variations. One reason for this might be that between-class variations span a more complex multi-modal distribution and hence the unimodal approximation described in Section 4 might not capture the information in between-class variations. If our speculation is correct, using the Expectation Maximization algorithm to model between-class variations with a mixture of Gaussians would be one way to resolve this issue. A similar observation was made by the authors of Moghaddam (2002) in the context of face recognition. Figs. 18 and 19 show the classification results using majority voting for the images shown in Figs. 14 and 15.

8.1.3. SVM classifiers

We now show classification results obtained using the SVM classifiers described in Section 5. We will use ‘SVM-RBF’ to denote

the SVM classifier with an RBF kernel and ‘SVM-Linear’ to denote the SVM classifier with a linear kernel.

Table 5 is the average confusion matrix obtained using the SVM-RBF classifier by training on a single satellite image. From the entries shown in the table, we conclude that the average accuracy in this case is 88.93%. Table 6 is the average confusion matrix obtained using the SVM-RBF classifier by training on a wide area region. The average accuracy in this case is 94.59%. Table 7 is the average confusion matrix obtained using the SVM-Linear classifier by training on a single satellite image. The average accuracy in this case is 88.33%. Table 8 is the average confusion matrix obtained using the SVM-Linear classifier by training on a wide area region. The average accuracy in this case is 93.76%.

Comparing Tables 5 and 6, we note that wide-area learning provides good improvement in average classification accuracy for five out of six classes when we use a SVM classifier with an RBF ker-

Table 7

Average Confusion Matrix Using a Single Satellite Image for an SVM classifier with a Linear Kernel.

	ActiveCropField	Soil	Trees	Water	Buildings	Roads
ActiveCropField	78.700	0.0875	21.212	0.0	0.0	0.0
Soil	1.812	94.050	0.0125	0.0	0.025	4.100
Trees	1.587	0.2375	98.175	0.0	0.0	0.0
Water	0.0	0.0125	0.0375	94.600	0.0	5.350
Buildings	0.325	6.325	0.0	0.0875	87.963	5.300
Roads	0.5125	8.025	0.4375	0.0	14.488	76.537

Table 8

Average Confusion Matrix using a Wide Area Region for a SVM classifier with a Linear kernel.

	ActiveCropField	Soil	Trees	Water	Buildings	Roads
ActiveCropField	97.213	0.6125	2.175	0.0	0.0	0.0
Soil	0.0	96.375	0.0125	0.0	0.125	3.487
Trees	5.787	0.075	94.112	0.0	0.0	0.025
Water	0.0	0.0125	0.025	99.075	0.0	0.8875
Buildings	0.2	5.825	0.0125	0.225	85.425	8.312
Roads	0.375	2.400	0.3625	0.0	6.487	90.375

nel, with significant improvements for the 'Active Crop Field' and 'Roads' classes.

Comparing [Tables 7](#) and [8](#), we see that when we use a SVM classifier with a linear kernel, wide-area learning improves average classification accuracy for 4 out of 6 classes, with significant improvements for the 'Active Crop Field' and 'Roads' classes. **Thus for both our Bayesian framework and for SVM classifiers, the results indicate that wide-area learning is better than learning that is based on just a single satellite image.**

8.2. Bayesian versus SVM

For comparing the performance of the Bayesian with the SVM classifiers, we only consider the wide-area learning paradigm. This comparison tells us that for the four classes 'Soil', 'Trees', 'Water' and 'Roads', the Bayesian framework exhibits greater average classification accuracies than the SVM classifiers.

Using [Tables 2](#) and [8](#) to compare the Bayesian classifier and the SVM-Linear classifier, we observe that the Bayesian classifier increases the classification accuracies by roughly 6% for the 'Trees' class, 9.8% for the 'Roads' class and 3.7% for the 'Soil' class. For the 'Water' class, both classifiers yield high accuracies and we only observe a marginal increase ($< 1\%$). Using [Tables 2](#) and [6](#) to compare the Bayesian and SVM-RBF classifiers, we observe that the Bayesian classifier yields an increase of roughly 7% in the classification accuracy for the 'Trees' class, an increase of roughly 6.5% for the 'Roads' class and an increase of roughly 1.2% for the 'Water' class. For the 'Soil' class, both classifiers yield very high accuracies ($> 99\%$) and the increase is very marginal. We do note a decrease in classification accuracies for the 'Active Crop Field' and 'Buildings' classes. We believe that the Bayesian classifier loses out to the SVM classifier for the case of 'Active Crop Field' because the former is based on differences of the feature vectors as opposed to the feature vectors directly. Since the variations within the 'Trees' class are similar to the variations within the 'Active Crop Field' Class when using only spectral features, there is confusion between the two classes for the Bayesian approach. With regard to the reduction in accuracy observed for the 'Buildings' class, see the discussion at the end of [Section 8.1.1](#).

At first glance, it appears as if the average accuracies of the SVM classifiers are better than those of the Bayesian classifiers by around 4–5%. However we believe that the per class comparison of accuracies described above presents a more complete and ac-

curate picture than merely comparing the total average accuracies. We have used an equal number of testing samples for each class. Therefore despite the better performance of our Bayesian framework for four out of six classes, the lower accuracy rates for the remaining two classes bring down the overall average accuracy values for the Bayesian classifiers.

We used randomization-based significance testing and the Student's t-test to evaluate the statistical significance of the differences in performance between the Bayesian and SVM classifiers [Kak \(2015\)](#). In what follows, we present the results of these tests in a comparison of the wide-area trained Bayesian classifier using both within-class and between-class variations with the wide-area trained SVM classifier with an RBF kernel. The corresponding average confusion matrices are shown in [Tables 4](#) and [6](#).¹⁴

For each type of classifier (Bayesian or SVM), we obtain a confusion matrix in each run of the 10-fold cross validation test described in [Section 7](#). For each land-type class, we can therefore create a 10 element vector consisting of the classification accuracies in each of the ten runs. This is shown in [Table 9](#). To elaborate, the first row, labeled "Active Crop Field (B)" in the first column, corresponds to using a Bayesian classifier based on both within-class and between-class variations. Similarly, the second row of this Table, labeled "Active Crop Field (S)", corresponds to the 10 classification accuracies for the "Active Crop Field" class obtained with 10 runs of the SVM-RBF classifier. The remaining rows of the table are for the other land-type classes. We use these values for statistical significance testing. As is normally done in such testing, we fix the significance value α at 0.05. If the p-value for any observed difference in performance is less than α , that observation is deemed as statistically significant.

[Table 10](#) shows the result of randomization-based significance testing. Comparing the p-values to α , we see that the performance differences in all classes are statistically significant except for the Soil class. This makes sense since as shown in [Table 9](#), the Soil accuracies in the samples from both the classifiers are very similar and tied at 100% in 8 out of 10 runs. Thus the observed performance improvements for the 'Trees', 'Roads' and the 'Water' class when comparing [Tables 4](#) and [6](#) are statistically significant. Since

¹⁴ The conclusions were similar when we compared the Bayesian classifier using only the within-class variations with the SVM-RBF classifier. When we compared the linear SVM classifier against both versions of the Bayesian classifier, all observed performance differences were deemed statistically significant.

Table 9
Classification accuracy samples from Bayesian (B) and SVM (S) classifiers. B = Wide-Area Bayesian classifier trained with both within-class and between-class variations. S = Wide-Area SVM classifier with RBF kernel.

Land-type class (Classifier)		10 fold cross-validation accuracies								
ActiveCropField (B)	65.750	69.000	66.500	67.375	65.625	67.500	67.500	69.750	64.375	67.875
ActiveCropField (S)	98.500	99.000	98.375	98.375	98.500	98.625	98.125	98.750	98.375	99.125
Soil (B)	100	100	100	99.875	100	100	100	100	100	100
Soil (S)	99.875	100	100	99.875	100	100	99.875	100	100	100
Trees (B)	99.875	99.875	99.875	99.875	99.875	99.625	99.750	99.500	99.500	99.875
Trees (S)	93.250	91.750	92.125	93.625	93.375	93.500	93.125	92.375	95.125	92.875
Water (B)	100	99.875	100	100	99.875	100	100	99.750	100	100
Water (S)	98.875	99.000	98.875	98.875	99.125	98.250	98.500	99.000	98.000	98.750
Buildings (B)	68.125	65.500	64.500	65.500	66.625	66.375	68.625	67.750	65.125	65.250
Buildings (S)	84.000	85.625	83.125	84.375	84.625	84.125	84.625	84.375	84.125	81.250
Roads (B)	98.750	98.875	99.250	98.750	99.500	99.000	98.750	98.750	98.750	99.125
Roads (S)	92.875	92.500	93.375	93.625	92.500	94.250	92.750	92.625	94.625	92.375

Table 10
Result of Significance Testing with Randomization.

Class	ActiveCropField	Soil	Trees	Water	Buildings	Roads
p-value	0.00195	0.50000	0.00195	0.00195	0.00195	0.00195
p-value < 0.05 ?	Yes	No	Yes	Yes	Yes	Yes
Better classifier	SVM	–	Bayesian	Bayesian	SVM	Bayesian

the Student’s t-test yielded similar results we will not show them here for the sake of brevity.

These results thus provide credible evidence to support our intuition that different land-classes not only possess different spectral signatures, but, even more importantly, exhibit different types of variations in their spectral signatures. Modeling these variations using a Bayesian framework provides us with a robust novel classifier that outperforms state-of-the-art approaches like SVM classifiers in identifying as many as four out of six important land types in satellite images.

Next, we show results on unseen data using the Bayesian and SVM classifiers.

8.3. Results on unseen data

8.3.1. Bayesian classifiers

As mentioned in Section 7, we tested our Bayesian classifiers on unseen data, that is, data that was not used in the training and the testing involved in the previously described 10-fold cross-validation test. The unseen data image was shown earlier in Fig. 16. Fig. 20 shows the classification output for this dataset.

Again as mentioned previously, for quantitative analysis, a small ground-truth dataset was collected from a portion of this unseen image. Table 11 shows the average confusion matrix obtained when the wide-area trained classifiers are applied to this unseen data. Table 12 shows the average confusion matrix obtained when the single-satellite-image classifier is applied to the unseen data. We modeled only the within-class variations for these tests.

Comparing Tables 11 and 12, we see that, for three out of six classes, wide-area learning of the classifiers gives us superior results compared to the case when the classifiers are trained on data collected from just a single satellite image. The wide-area-learning improvements we get for the ‘Active Crop Field’ and ‘Soil’ classes are very significant. However, we also notice that the performance of wide-area learning decreases for the remaining three classes. So, for the unseen data example of Fig. 16, choosing between wide-area learning and single-satellite-image learning for classification would appear to be a toss-up at first thought. However, upon further thought, all that this comparison tells us is that the statistical properties of the data actually used for training were not truly reflective of all of the within-class variations in the unseen data.

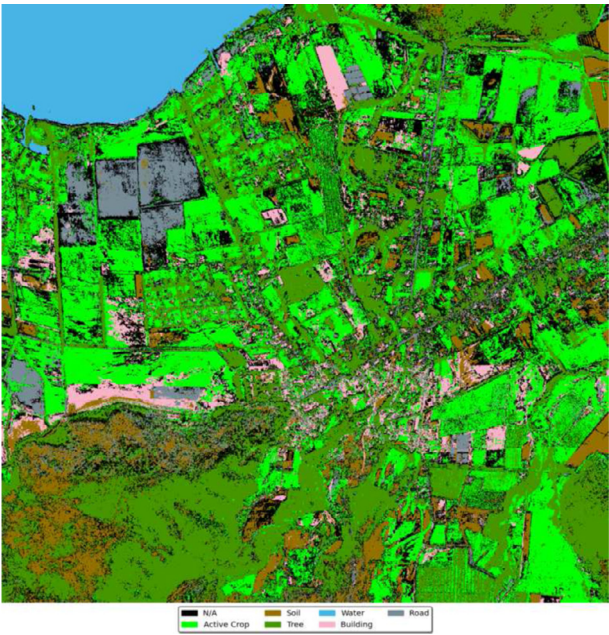


Fig. 20. Majority vote result using wide-area learning with the Bayesian classifier on Fig. 16.

Comparing classifier performance on unseen segments of data, as we just did with the help of Tables 11 and 12, would be an important part of the protocol we plan to develop in follow-up research for ensuring that the number of subregions used for wide-area learning is sufficient to represent all of the statistical variations in the entire ROI.

8.3.2. SVM classifier

In this section we present the results of wide-area learning using the SVM-RBF classifier on the same unseen dataset as in Section 8.3.1. Fig. 21 shows the classification output for this dataset.

A quantitative evaluation was also carried out on the small ground-truth dataset collected from this unseen data.

Table 11

Average Confusion Matrix for Unseen Data using a Wide Area Region for Training with Within-Class Variations.

	ActiveCropField	Soil	Trees	Water	Buildings	Roads
ActiveCropField	43.937	0.0	56.063	0.0	0.0	0.0
Soil	0.0	99.682	0.0	0.0	0.0	0.3181
Trees	16.078	0.0522	83.800	0.0	0.0	0.0696
Water	0.0	0.0	0.1309	99.866	0.0035	0.0
Buildings	0.0	5.349	0.047	0.0	66.312	28.291
Roads	0.0294	17.404	0.0	0.0	0.0	82.566

Table 12

Average Confusion Matrix for Unseen Data using a Single Satellite Image for Training with Within-Class Variations.

	ActiveCropField	Soil	Trees	Water	Buildings	Roads
ActiveCropField	31.521	0.0	68.479	0.0	0.0	0.0
Soil	0.0	75.736	0.0	0.0	24.264	0.0
Trees	6.110	0.0	93.423	0.0	0.4669	0.0
Water	0.0	0.0	0.0071	96.730	3.263	0.0
Buildings	4.537	4.670	4.537	4.537	75.107	29.297
Roads	0.295	0.5605	0.0	0.0	3.451	95.693

Table 13

Average Confusion Matrix for Unseen Data using a Wide Area Region for a SVM classifier with a RBF kernel.

	ActiveCropField	Soil	Trees	Water	Buildings	Roads
ActiveCropField	100.0	0.0	0.0	0.0	0.0	0.0
Soil	0.0	100.0	0.0	0.0	0.0	0.0
Trees	63.745	0.0	36.255	0.0	0.0	0.0
Water	0.0	0.0	0.3154	97.375	0.0	2.310
Buildings	0.0	7.267	0.047	0.166	67.822	24.699
Roads	0.393	14.110	0.0	0.0	16.726	68.771

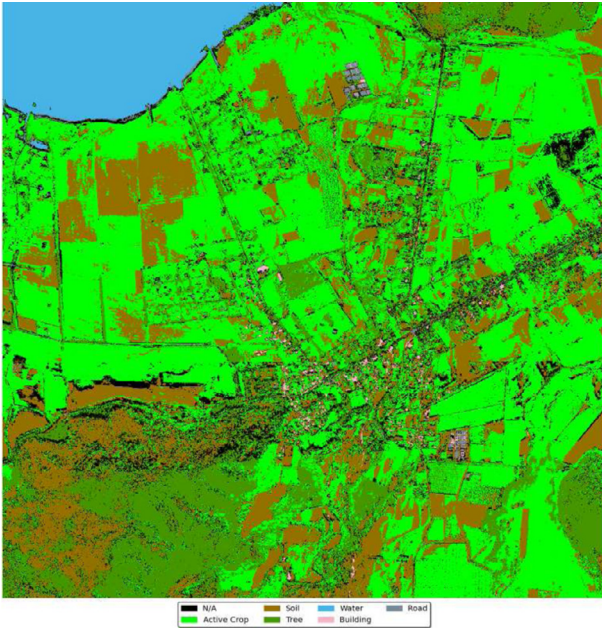
**Fig. 21.** Majority vote result using wide-area learning with the SVM-RBF classifier on Fig. 16.

Table 13 shows the average confusion matrix obtained for this experiment.

Comparing Tables 11 and 13, we can see that for wide-area learning, the Bayesian classifier shows better performance than the SVM classifier for the ‘Trees’, ‘Roads’ and ‘Water’ class. We ob-

serve comparable performances for the ‘Soil’ and the ‘Buildings’ classes. For the ‘Active Crop Field’ class, we notice lower accuracy rates for the Bayesian classifier. We have already discussed possible reasons and solutions for this in Section 8.2 and at the end of Section 8.1.1. These observations are in sync with our observations in Section 8.2 and thus lend additional credibility to the benefits of our wide-area based Bayesian learning paradigm.

Finally we would like to briefly revisit the discussion on orthorectification errors in Section 2.2. The variability heat maps in that section as well as our results presented here lend credence to the fact that our approach can handle orthorectification errors to a good extent. Obviously, extreme misregistration between two overlapping images will create confusion for our proposed classifiers. As mentioned in Section 3, we address this issue during the data annotation process by visually inspecting the overlapping satellite images beneath the marked polygons and excluding images that are significantly misaligned. Subsequently, we collect pixels within the polygon only from the remaining subset of the overlapping images. Also note that the effect of misregistration depends on the size of the target class/object as well. For example, in our experiments, we observed that one should be more careful when collecting training data for small individual objects (such as, say, red-roof buildings) compared to collecting data for extended objects such as large water bodies or forests. Additionally, for each pixel, as long as a sufficient percentage of overlapping images are aligned reasonably well, combining class labels using majority voting provides reasonable protection against misclassification.

9. Conclusions and future work

This paper first presented the PIMSIR tool for creating an integrated representation for all satellite images, including those that

are overlapping, that cover a large geographic area. We showed how this tool can be used to understand the data variability in the images - variability that runs across the images wherever they are overlapping, spatial variability, and temporal (and seasonal) variability. Subsequently, we showed how one's understanding of this variability can be used to design Bayesian classifiers that are trained on a wide-area basis, as opposed to the more traditional approach of training classifiers using the data in just a single image. A necessary requirement of wide-area learning is statistical sampling of all of the satellite data for generating the ground truth. The overall training framework we provided included a Metropolis-Hastings random sampler for generating the points to be used for training and testing in proportion to the extent of satellite coverage over the different parts of the geographic area. Despite the fact that we used only a small number of subregions thus selected for generating the ground truth, we showed that wide-area based learning classifiers gave us better classifications for five out of six classes compared to the traditional approach of using a classifier based on the data extracted from just one image.

Using the same training and the testing data sets, we also compared the performance of the variance-based Bayesian classifiers with the more traditional SVM classifiers. The conclusion here was that the variance-based framework led to superior classifications for four out of six land-cover classes.

With regard to future extensions of this research, a very important issue that remains to be addressed is testing for sufficiency of the ground-truth data as yielded by the Metropolis-Hastings based random sampler. We want the ground-truth data to represent all of the diversity - speaking statistically, of course, - that exists in an ROI for each land-cover class. At the same time, we do not wish to collect any more ground-truthed data than is necessary on account of the high cost of the human labor involved in supplying the human judgments. We therefore need some sort of a feedback loop that uses the "test on unseen data" sort of strategy presented in Sections 7 and 8.3 to evaluate the quality of the data collected for ground-truthing up to a given point so that a decision can be made whether or not to collect additional data.

For another future direction of investigation, we believe that, in the variance-based Bayesian classifier, we would get superior results when we also incorporate between-class variations provided we create a multi-modal model for such variations. As we mentioned earlier in the paper, perhaps the best future approach would be to use the EM algorithm for creating such a multi-modal model.

For yet another future direction, we think our results would become even more impressive if we could enrich the spectral-signatures based feature space with spatial features. It is impossible to take into account the spatial context of a geo-point when classification is carried out just on the spectral signatures at each point. As we saw in Section 8, it is rather easy for the building pixels to be confused with the road pixels since we can expect the spectral signatures of certain kinds of rooftops (say, the flat roofs that are made with concrete slabs) to yield nearly the same signatures as the roads. Using spatial and, perhaps, texture properties of the blobs that surround the pixels would be one way to mitigate such inter-class confusions.

Acknowledgment

Supported by the Intelligence Advanced Research Projects Activity (IARPA) via Air Force Research Laboratory, contract FA8650-12-C-7214. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes not withstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, AFRL, or the U.S. Government.

References

- Aksoy, S., Haralick, R., 2000. Probabilistic vs. geometric similarity measures for image retrieval. *IEEE Conf. Comput. Vis. Pattern Recognit. CVPR* 357–362.
- Aksoy, S., Haralick, R.M., 2001. Feature normalization and likelihood-based similarity measures for image retrieval. *Pattern Recognit. Lett.* 22 (5), 563–582.
- Anderson, J.R., 1976. A land use and land cover classification system for use with remote sensor data, 964. US Gov. Printing Office.
- DeFries, R., Chan, J.C.W., 2000. Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data. *Remote Sensing Environ.* 74 (3), 503–515.
- DeFries, R.S., Townshend, J.R.G., 1994. NDVI-derived land cover classifications at a global scale. *Int. J. Remote Sensing* 15 (17), 3567–3586.
- Duro, D.C., Franklin, S.E., Dubé, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing Environ.* 118, 259–272.
- Fauvel, M., Tarabalka, Y., Benediktsson, J.A., Chanussot, J., Tilton, J.C., Mar. 2013. Advances in spectral-spatial classification of hyperspectral images. *Procee. IEEE* 101, 652–675.
- Huang, C., Davis, L., Townshend, J., 2002. An assessment of support vector machines for land cover classification. *Int. J. Remote Sensing* 23 (4), 725–749.
- Kak, A.C., 2014. Algorithm: RandomPointGenerator. <http://search.cpan.org/~avikak/Algorithm-RandomPointGenerator-1.01/lib/Algorithm/RandomPointGenerator.pm>.
- Kak, A.C., 2015. Evaluating information retrieval algorithms with significance testing based on randomization and student's paired t-test (a tutorial). <https://engineering.purdue.edu/kak/Tutorials/SignificanceTesting.pdf>.
- Kerdiles, H., Grondona, M., 1995. NOAA-AVHRR NDVI decomposition and subpixel classification using linear mixing in the argentinean pampa. *Int. J. Remote Sensing* 16 (7), 1303–1325.
- Knerr, S., Personnaz, L., Dreyfus, G., 1990. Single-layer learning revisited: a step-wise procedure for building and training a neural network. In: *Neurocomputing*. Springer, pp. 41–50.
- Marchisio, G., Pacifici, F., Padwick, C., July 2010. On the relative predictive value of the new spectral bands in the worldview-2 sensor. In: *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, pp. 2723–2726.
- Moghaddam, B., June 2002. Principal manifolds and probabilistic subspaces for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 24, 780–788.
- Moghaddam, B., Jebara, T., Pentland, A., 2000. Bayesian face recognition. *Pattern Recognit.* 33 (11), 1771–1782.
- Unsalan, C., 2003. Multispectral satellite image understanding. The Ohio State University Phd thesis.
- Updike, T., Comp, C., 2010. Radiometric use of worldview-2 imagery. Technical Note, Digital Globe, Inc., Revision 1.0. [http://global.digitalglobe.com/sites/default/files/Radiometric_Use_of_WorldView-2_Imagery%20\(1\).pdf](http://global.digitalglobe.com/sites/default/files/Radiometric_Use_of_WorldView-2_Imagery%20(1).pdf).
- Wilkinson, G., 2005. Results and implications of a study of fifteen years of satellite image classification experiments. *IEEE Trans. Geosci. Remote Sensing* 43 (3), 433–440.
- Wolf, A., 2010. Using worldview-2 vis-NIR multispectral imagery to support land mapping and feature extraction using normalized difference index ratios. In: *Proc. SPIE 8390, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII*, 83900N (May 8, 2012).