ONBOARD SVM ANALYSIS OF HYPERION DATA TO DETECT SULFUR DEPOSITS IN ARCTIC REGIONS

Lukas Mandrake[(1)], Kiri L. Wagstaff[(1)], Damhnait Gleeson[(1) (2)], Umaa Rebbapragada[(3)], Daniel Tran[(1)], Rebecca Castaño[(1)], Steve Chien[(1)], Robert T. Pappalardo[(1)]

[(1)]Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, CA 91109
<first.lastname>@jpl.nasa.gov

[(2)]Laboratory for Atmospheric and
Space Physics
University of Colorado at Boulder
Boulder, CO 80303

[(3)]Tufts University
Computer Science Department 161
College Ave
Medford, MA 02155
urebbapr@cs.tufts.edu

ABSTRACT

Onboard classification of remote sensing data can permit autonomous, intelligent scheduling decisions without ground interaction. In this study, we observe the sulfur-rich Borup-Fiord glacial springs in Canada with the Hyperion instrument aboard the EO-1 spacecraft. This system offers an analog to far more exotic locales such as Europa where remote sensing of biogenic indicators is of considerable interest. Previous work has been performed in the generation and execution of an onboard SVM (support vector machine) classifier to autonomously identify the presence of sulfur compounds associated with the activity of microbial life. However, those results were severely limited in the number of positive examples available to be labeled. In this paper we extend the sample size from 1 to 7 example scenes between 2006 and 2008, corresponding to a change from 18 to 235 positive labels. We also explore nonlinear SVM kernels as an extension of our onboard capability.

Index Terms—SVM, Hyperspectral, Autonomy, Sulfur, Hyperion

1. INTRODUCTION

Remote sensing enables the otherwise impossible goal of analyzing large, inhospitable regions too distant for direct human observation and study. However, satellite-based instruments generate far more data than can ever be successfully downlinked. While several approaches have been employed to address this limitation, new instruments such as hyperspectral cameras continue to strain even the most robust operations centers [1][2]. Despite this production capability, the majority of data collected by a streaming camera is redundant or otherwise uninteresting after a region has been initially scanned. Instead, key regions with unusual or dynamic features may be selected as high priority while static areas might receive lower or no priority. While time-based events may be predicted and captured in similar manner, such occurrences are often either unpredictable or not known to exist prior to discovery. Thus an onboard detector capable of deciding when a region is interesting based on predefined criteria is of great interest. In our case, we are supporting NASA's mission to discover living systems and their evidence by creating a remote sensing detector for sulfur compounds.

Prior work on precisely this question was hampered by a lack of available labeled hyperspectral imagery [3]. We will be extending these results by including two orders of magnitude more data for both test and training sets. A more thorough treatment of this topic can be found in Mandrake et al [4].

2. TARGET SCENES AND INSTRUMENT

From above Ellesmere Island can be seen vivid yellow stains along a glacial edge. These discolorations were identified as supraglacial deposits associated with sulfur-rich springs flowing through and across a 200-meter thick glacier. The telltale yellow stain was due to the presence of elemental sulfur, which dominates the composition of the deposits. The presence of elemental sulfur, gypsum, and hydrogen sulfide gas within the flow, compounds containing sulfur in three different oxidation states, suggest the presence of a complex redox system likely biogenic in origin. This theory is supported by the presence of a rich microfauna of known and as yet unidentified microbes. This arrangement of fissured permanent ice, sulfur-bearing subsurface rock, and seeping water is potentially analogous to the surface of Jupiter's moon Europa. As elemental sulfur occurs naturally generally from either volcanic or anaerobic bacterial processes, producing a remote elemental sulfur detector is a step towards a microbial life detector.

The data available to this study consists of seven flyover reflectance images of Borup Fiord collected by the Hyperion imager aboard the Earth-Observing-1 spacecraft (EO-1) for both training and test cases. Labels were provided by hand by Damhnait Gleeson, with only a few positive (sulfur-bearing) examples present in each image along with ~760k negative examples divided between ice and rock categories. False Positives in these source images are referred to hereafter as "Likely False Positive" as we do not truly know there are no other sulfur springs in the vicinity. In addition, seven other scenes were obtained that are far from any known sulfur. False positives in these later images are referred to as Sulfur-Free False Positives and are considere to be more egregious mistakes than false positives detected near known sulfur deposits.

3. METHODS

Following the methodology originally developed by Castaño et al [3], we employed a multiclass SVM to classify each sample location using 12 (of the total 220 available) wavelengths Hyperion observes using both linear and Gaussian kernels. Only 12 bands may be examined at a time due to onboard processing limitations. The selection of these optimal wavelengths for classification purposes were evaluated by Greedy Forward Feature Selection, Recursive Feature Elimination (RFE), a domain expert's manual examination, and finally compared to Castaño et al.'s RFE based only on a single 2006 image. For initial band selection, a binary classifier was used (sulfur / non-sulfur). Using the optimal band

sets produced by each of these four methods, we trained a linear kernel 2-class (sulfur / non-sulfur), 3-class (sulfur, ice, rock), and 4-class (bright/dark sulfur, ice, rock) multiclass SVMs. Statistics on success were measured via per-image cross validation, where one of the seven source images was used as a held-out test set and the others were used to train the SVM.

We evaluated the impact of a large range of different values for the SVM hyperparameter C. The "best" result will be some set of bands (one of the four sets selected above), a C value, and the number of permitted classes. This configuration can then be coded and uploaded into EO-1 for autonomous sulfur detection. For the final comparison, we generalized this formulation to include a second hyperparameter γ that defines the width of the Gaussian kernel.

In our search for an accurate classifier, we allowed the possibility that our labeled examples were less than pure. Labeling these pixels manually is challenging even for experts. To address this, we used a label confidence evaluation method, Pair-Wise Expectation Maximization (PWEM) [5], to filter suspect labels.

4. RESULTS

4.1. Band Selection

The sets of 12 bands of interest selected are shown in Table 1. General agreement is seen between the two RFE methods and the Expert picks, with the 2006-only, single image method choosing a single anomalous high-wavelength value. The Greedy Forward method however retains some very high wavelength components which are likely due to the very low signal to noise ratio at those higher values. The Greedy Forward method will not be pursued from this point onward based on these results.

Table 1. Band Selection Results (λ nm)

RFE 2006 only	RFE	Greedy Forward	Expert Pick
426	426	426	426
436	436	436	436
446	446	446	446
456	456	456	456
506	466	556	466
516	536	566	486
526	546	576	506
536	556	586	526
546	566	596	546
556	576	1448	566
576	586	1791	586
994	596	2013	626

4.2. Expanding from 3 to 4 Classes

Our early attempt at this analysis showed abundant false positives (FP) using the entirely of the sulfur training set and a 3-class SVM ($\sim 10^5$ FP events per downlinked image). We suspected that there were multiple populations (including potentially false labels) within our positive example sulfur training set. We applied a K-means clustering method to the sulfur labels that immediately yielded two distinct populations: one with high mean reflectance and one low. We divided our sulfur labels according to the

clustering results (yielding four classes) and discovered that most of the false positives were from the "dark" sulfur class. Therefore, although this division was not obtained via manual labeling, it still provided a very useful refinement of the labeled data and, as we will show, led to improved performance. Physically, sulfur that occurs within rock rich pixels is extremely difficult to distinguish from sulfur-free rock-rich pixels, supporting this differentiation.

4.3. Filtering Poorly Labeled Examples

Armed with our three band sets of interest (RFE 2006, RFE, and Expert Pick), we investigated the mislabeling hypothesis using PWEM [5]. As shown in Figure 1, ice labels proved to be the most confident, since selecting pure ice examples in the image is quite easy. Rock was similarly well labeled with a few remote outliers. However, sulfur was highly contentious. No sulfur label exceeded 0.75 probability of consistency. Separating into bright and dark sulfur classes we immediately rise to 104 / 144 bright sulfur labels as above 0.75 confident, and 39 / 91 dark sulfur labels above 0.75 confident. At this point, we discarded any label not above 0.75 confident as too uncertain for use. We will report these results as "4class filtered." Although necessary, this step removed 40% of our already limited positive examples. Low-confidence pixels tended to occur near ice-rock boundary areas.
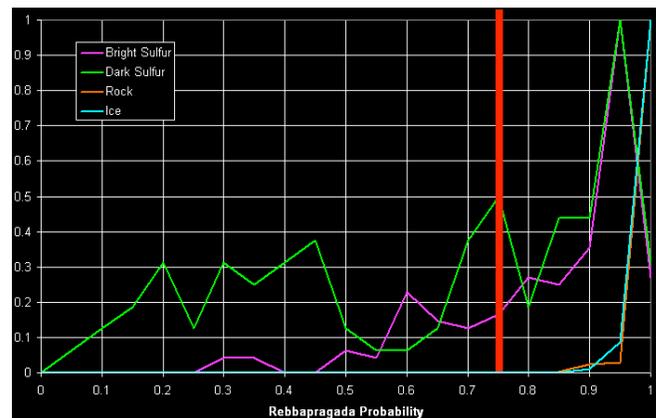

Figure 1. Histogram of label correctness. X axis is the PWEM estimate of label confidence. Everything left of the red line was filtered as too unconfident.

4.4. SVM Performance

Figure 2 shows the performance of the SVMs for various class numbers, kernel types, and mislabeled filtration strategies. On the right we see the Sulfur-Free FPOS results along the Y axis, while on the left are the Likely FPOS results. Note that there are always more Likely FPOS than there are Sulfur-Free FPOS. This makes sense, as separating sulfur from nearby sulfur-like terrain is a harder problem than separating entirely non-sulfurous terrain from sulfur-bearing surroundings. On the X axis in both plots is the F-measure, with each point representing a given SVM configuration of C, class number, set of 12 wavelength bands, and in the case of the Gaussian kernel the width hyper-parameter γ. We have separated the sets of 12 bands (Expert, original RFE result using only 2006 image, and current RFE result) into blue, red, and white

points respectively. The optimal results would be points lying at the bottom-right of each graph with zero Likely / Sulfur-Free FPOS (Y) and F-measure of one (X).

The linear kernel 2-class (sulfur / non-sulfur) SVM shows poor performance. While several configurations achieve an impressive F-measure of up to 0.80 - 0.82 for all three band selections, there are more than 10^4 Sulfur-Free FPOS and similarly for the Likely FPOS.

The linear kernel 3-class (ice, rock, and sulfur) SVM shows improvement in all band selections. Sulfur-Free FPOS were reduced by two orders of magnitude to a mean of ~40 FPOS per Sulfur-Free image. Likely FPOS remained high with a mean of ~3400 per image. The mean and maximum F-measure also increased, with a new maximum of 0.86.

The linear kernel 4-class (ice, rock, "bright" sulfur-on-ice, and "dark" sulfur-on-rock) SVM results in two consequences. The Sulfur-Free FPOS increases to a mean of ~70 pixels per image. Likely FPOS was reduced yielding a mean of ~800 false positives per image. Most of the later improvement was due to the decision to define all sulfur on rock "dark sulfur" labels as unreliable and no longer count them towards our accuracy requirement. The mean F-measure increased with a mean of ~0.9 and a maximum of 0.96. As performance was increased by filtering our positive labeled instances, a more rigorous method to remove contaminated training examples became attractive.

Employing PWEM's confidence estimation method on the 4-class data, a mean F-measure of 0.90 was maintained while reducing the mean Likely FPOS to ~600 per image. Sulfur-Free FPOS reduced further to a mean of 24 pixels per image. This implies mislabeled training examples were indeed corrupting our results with FPOS. For EO-1 Hyperion effort, we would select one of the blue or white indicators on the graph representing a particular combination of band (Expert Pick or RFE) and C hyper-parameter (3-20 is roughly comparable in performance). This linear kernel SVM would then be uploaded for operation. Unfortunately, we still expect ~500 false detections on images near Borup Fiord.

Extending to Gaussian kernels (using only Expert and RFE bands) yields the last set of graphs that achieve a mean F-measure of 0.93 and maximum F-measure of 0.98 using the 4-class filtered dataset. Mean sulfur-Free FPOS remains at ~ 50 pixels per image, while mean Likely FPOS remains high at mean ~1000 detections per image. Thus Gaussian kernels do suggest utility but are no panacea for the limitations of this approach.

5. CONCLUSIONS

We have shown that while three classes of terrain are an appropriate description of the data (ice, rock, and sulfur), in order to overcome the limited number of training positive samples we were forced to create a fourth "junk" class to contain and remove noise. This noise was determined to correspond to the majority of "dark" sulfur examples being mixed ice and rock. Since pure ice and rock examples were provided as negative labels, we in effect created both a sulfur detector and a rock-ice mixture detector simultaneously. In the absence of disambiguating additional positive sulfur examples, we were forced to instead pear down the training label set to remove confounding similarities between positive and negative labeled examples. In the end, we successfully trained a SVM where less than one FPOS event in a mean 256 x 3200 image is expected in terrain unlike the training set, ~50 FPOS in similar-to-training terrain, and an F-measure of ~ 0.9 with regard to the filtered labels. Usage of a Gaussian kernel did improve F-measures slightly but at a heavy computational cost.

6. REFERENCES

[1] S. G. Ungar et al, "Overview of the Earth Observing One (EO-1) Mission", IEEE Transactions on Geoscience and Remote Sensing, vol. 41, No. 6, doi: 10.1109/TGRS.2003.815999, June 2003.

[2] S. Chien et al, "Onboard Classification of Hyperspectral Data on the Earth Observing One Mission", IEEE WHISPERS conference, Grenoble France, August 26-28, 2009.

[3] R. Castaño et al, "Onboard Detection of Active Canadian Sulfur Springs: A Europa Analogue," International Symposium on Artificial Intelligence Robotics and Automation in Space (i-SAIRAS), Los Angeles, Feb. 25-29th, 2008.

[4] L. Mandrake et al, "Onboard Detection of Natural Sulfur on a Glacier via an SVM and Hyperion Data," 2009 IEEE Aerospace Applications, March 7-14th, 2009.

[5] U. Rebbapragada et al, "Improving Onboard Analysis of Hyperion Images by Filtering Mislabeled Training Data Examples," 2009 IEEE Aerospace Applications, March 7-14th, 2009.
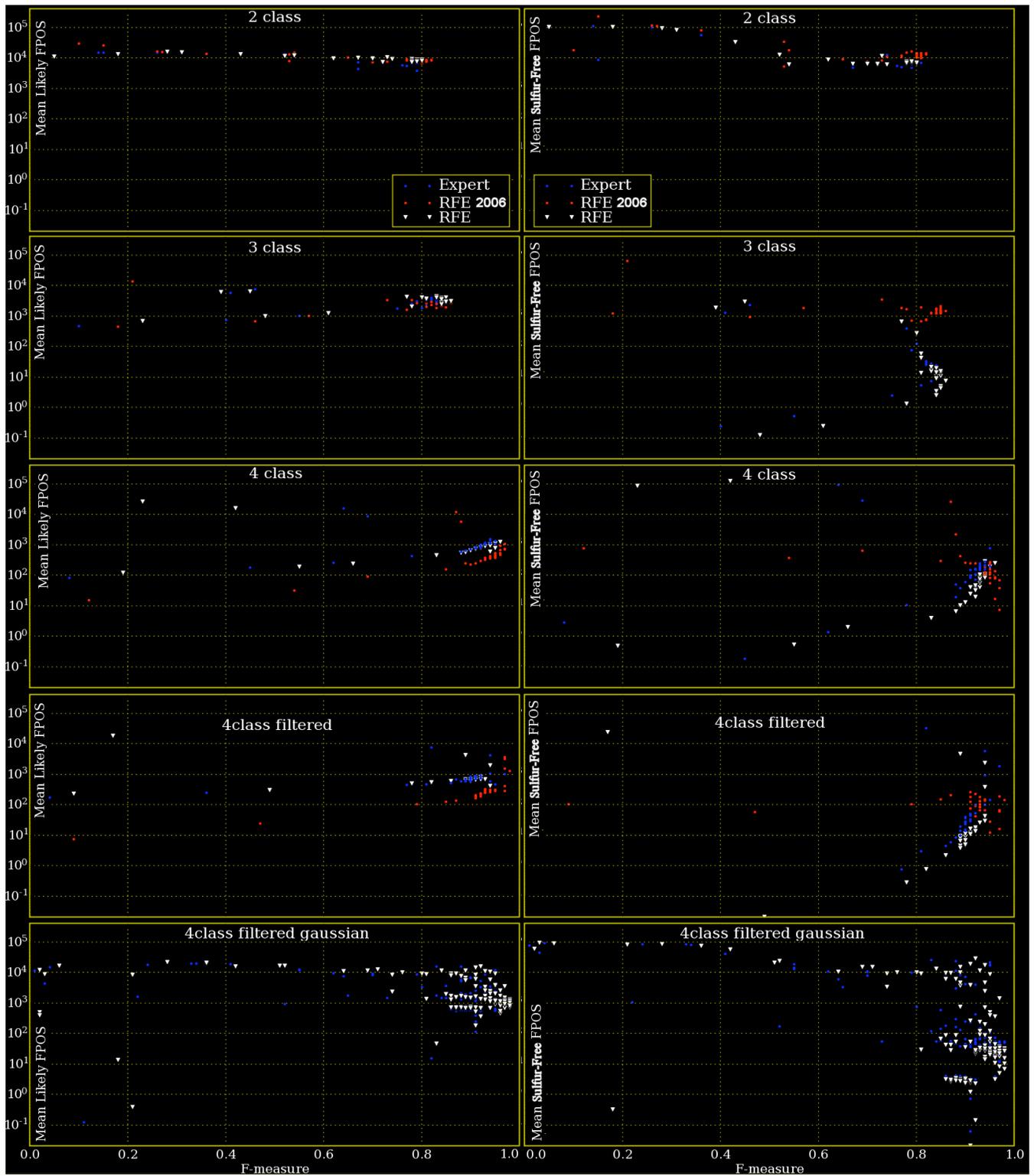
Figure 2. SVM performance for 2-4 classes, with and without mislabel filtration, using linear and Gaussian kernels. X axis are F-measure, while Y axis are # false positives. Left graphs show false positives on training images, right axis on far-from-source sulfur free images.