A COMPARISON OF LAND USE LAND COVER CLASSIFICATION USING SUPERSPECTRAL WORLDVIEW-3 VS HYPERSPECTRAL IMAGERY

Jan Koenig and Lionel Gueguen

DigitalGlobe Inc., Product Development and Labs, Westminster, Colorado U.S.A

WorldView-3 pan-sharpened RGB image
(Image Credit: DigitalGlobe)

ROSIS georeferenced hyperspectral image
(Image Credit: German Aerospace Center (DLR))

ABSTRACT

In advance of releasing a WorldView-3 (WV-3) dataset with both VNIR and SWIR bands for research purposes, this study was conducted to provide a baseline comparison of land use/land cover (LULC) classification based on hyperspectral and 16-, 8-, and 4-bands of WV-3 imagery. We chose a well-researched area over the city center of Pavia, Italy. Results suggest that the addition of spectral information from WV-3's SWIR bands helps bridge the gap between precision/recall scores obtained with multispectral VNIR vs. hyperspectral VNIR imagery.

Index Terms— Hyperspectral, Short Wave Infrared, Classification

1. INTRODUCTION

Land use/land cover classification using multispectral and hyperspectral imagery is a long-standing research activity for information extraction (e.g., [1], [2]). Many sensors are available for use in performing an LULC classification: those equipped with 4-8 bands in the VNIR spectrum such as GeoEye and RapidEye, those with bands in the VNIR/SWIR spectrum such as Landsat and ASTER, and hyperspectral sensors with hundreds of bands such as AVIRIS. With the launch of DigitalGlobe's WV-3 satellite, very high-resolution imagery with 16 bands in the VNIR/SWIR spectrum is now available. The different characteristics of each sensor in terms of bandwidth, spectral range, and spatial resolution result in LULCs with different precision/recall scores. Understanding how different sensor characteristics affect the resulting LULC is useful in achieving the necessary quality.

WV-3 was tasked soon after launch for the purpose of evaluating the benefit of SWIR bands to LULC. We selected a small area (1.5 km x 1.5 km) over the center of Pavia, Italy, because a hyperspectral dataset was available for comparison. The hyperspectral dataset for Pavia is well-known [3], having been the focus of considerable research for assessing approaches to improve LULC classification (e.g., [4], [5]).

The purpose of this study was twofold: 1) to compare LULC classification using WV-3 16-band imagery not only to the hyperspectral but also to WV-3 8- and 4-band imagery; 2)

Table 1. Ground truth pixel counts per class.

LULC Classes	Original	Adjusted	% of original
Water	56689	55683	98.2%
Trees	6634	2246	33.9%
Meadows	2653	825	31.1%
Self-blk bricks	2356	1147	48.7%
Bare soil	5847	429	7.3%
Asphalt	7975	3131	39.3%
Bitumen	5928	4978	84.0%
Tiles	36143	22715	62.8%
Shadow	2465	1943	78.8%

to examine the features important to the classification. This initial baseline comparison focuses on differences in precision and recall for an LULC based on spectral bands only as features (no indices, textures, or spatial features were used). Considerable pre-processing involving image georeferencing and adjustment of ground truth, was required to enable the comparison.

2. DATASET

2.1. WorldView-3 Imagery

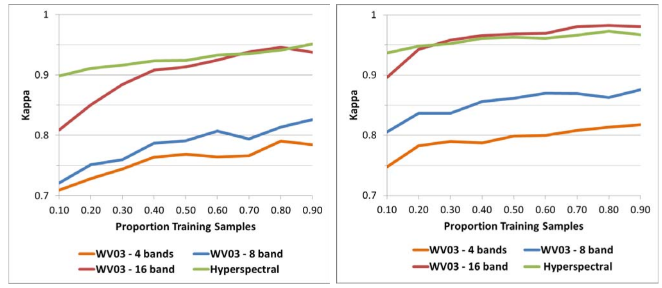
DigitalGlobe's latest satellite, WV-3, was launched on August 13, 2014. In addition to a spatial resolution of 31 cm in the panchromatic band and 1.24 m in the 8 bands of the VNIR spectrum (compared to 50 cm/2.0 m respectively for WV-2), WV-3 has 8 bands in the SWIR spectral range with a resolution of 3.7 m. It is also possible to obtain 12 bands that allow correction for Clouds, Aerosols, Vapor, Ice and Snow (CAVIS).

WV-3's SWIR bands can be grouped into 3 atmospheric windows [6] that were selected to maximize information extraction. The first SWIR band is useful for measuring soil moisture, leaf moisture, and iron content. The middle 4 bands can identify many man-made materials and the 3 bands in the upper portion of SWIR allow for mineral identification.

The WV-3 image was collected on January 24, 2015 with an off-nadir angle of 21.8 (see left introductory figure). Both the VNIR and SWIR bands were atmospherically compensated using DigitalGlobe's proprietary algorithm. The SWIR bands were resampled to the resolution of the VNIR bands using a nearest-neighbor algorithm and then layer-stacked to merge all features for input to the machine-learning algorithms. Note that while 3.7-m SWIR data was used for the analyses, only 7.5-m SWIR data is available commercially at the present time.

2.2. Hyperspectral Imagery

The hyperspectral image was taken on July 8, 2002 using the ROSIS-03 sensor by DLR, the German Aerospace Center. Spatial resolution is 1.3 m, spectral resolution is 4 nm, and the spectral range is 430 to 860 nm (VNIR). A sensor problem with a flightline resulted in one strip with missing data in the center of the study area. Of the 115 bands, the last 13 were removed due to excessive noise prior to the public distribution of the data. The hyperspectral dataset was also atmospherically corrected but not georeferenced. The georeferenced image (method described later) is shown in the right introductory figure.



(a) RF (b) SVM

Fig. 1. Normalized kappa as a function of the proportion of ground truth used for the training set.

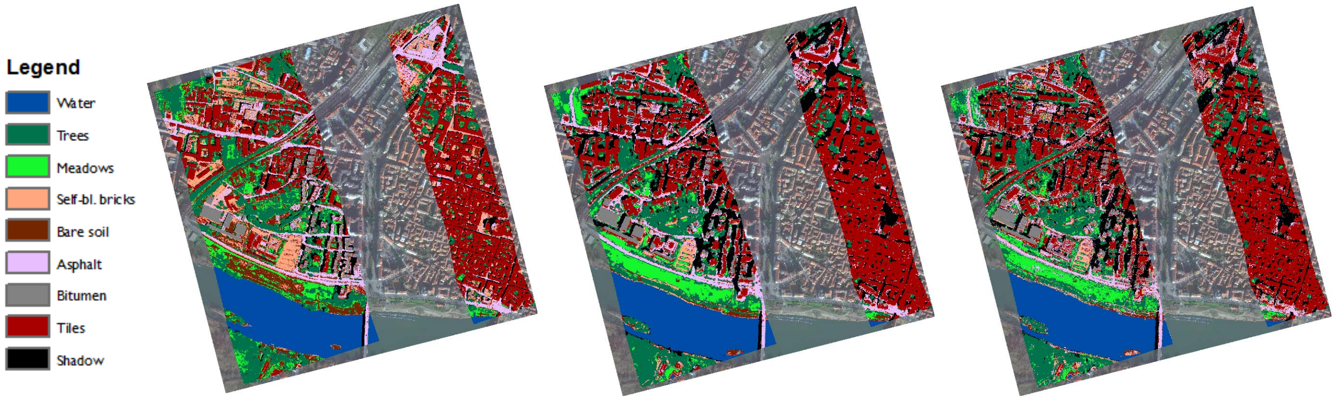
2.3. Making the datasets comparable

a-Georeferencing the hyperspectral image and ground truth

Because the WV-3 image was already georeferenced, the hyperspectral image was co-registered to the WV-3 image using ESRI's ArcGIS. Tie points were made on the ground whenever possible. However, the height of the buildings led to shifts in the rooftops so tie points were also placed at the corners of roofs. The georeferenced hyperspectral image was resampled to 1.24 m using a cubic-convolution algorithm. The same tie points used for the hyperspectral image were used to georeference the ground truth. The ground truth was resampled to 1.24 m using a nearest-neighbor algorithm.

b-LULC classes and adjustment of the ground truth

The LULC classes of interest to the classification are shown in Table 1. In addition to the natural classes of Water, Trees, Meadows, Bare soil, and Shadow, several classes are man-made materials: Self-blocking bricks, Asphalt, Bitumen, and Tiles. Temporal and seasonal differences as well a difference in viewing angles between the hyperspectral and WV-3 images resulted in the need to adjust the ground truth to be valid for both images (see Table 1). Several classes, especially Trees, Meadows, Bare soil, and Asphalt, were substantially reduced. Bare soil that was located primarily in a park along the river in the original image was a grassy meadow in the WV-3 image. Seasonal differences were seen primarily in trees and shadow. The leaf-off conditions in the WV-3 image diminished the correspondence of trees, al-



Hyperspectral WV-3 16 bands WV-3 8 bands

Fig. 2. LULC results obtained with SVM for hyperspectral and two WV-3 cases.

Table 2. Normalized kappa using RF and SVM classifiers. Spectral bands only were used as features in the LULC.

	SVM	RF
WV-3 4 bands (NRGB)	0.80	0.78
WV-3 8 bands (VNIR)	0.88	0.81
WV-3 16 bands (VNIR/SWIR)	0.98	0.93
Hyperspectral (VNIR)	0.97	0.94

though some leaf-off trees were included in the ground truth. Shadows in January are considerably longer than in July so the Shadow class required considerable adjustment. Many samples of the tile roofs were removed because they were shadowed in the WV-3 image. The off-nadir angle of WV-3 exposed walls of buildings, necessitating the removal of additional ground truth.

The adjustment of ground truth was accomplished using ESRI's ArcGIS by converting the raster to polygons and editing the polygons individually. Most adjustments involved the removal of ground truth. In a few cases, ground truth was added if close to the original ground truth and not in shadow.

3. LULC CLASSIFICATION RESULTS

3.1. Machine Learning Tools

Both Random Forest (RF) and Support Vector Machine (SVM) classifiers were used to classify the images based using the spectral bands only as features. Data values were normalized by band for input to SVM.

In order to avoid overfitting, the hyperparameters of the models were optimized for each case using grid-search functions available in scikit-learn [7]. Hyperparameters are those general parameters of a model, such as the number of trees used in an RF classifier, that are not associated with a specific training set. The performance metric was a normalized kappa using K-fold cross validation with 4 folds. For the RF classifier, the number of trees tested ranged from 15 to 300 in increments of 25 or 50 and the maximum depth of the tree varied from 8 to 32 in increments of 4. The optimum parameters for RF for the 4 LULCs were in the range of 100-200 trees with a maximum depth of 20-24.

For SVM, early experimentation revealed that a radial basis function (RBF) kernel consistently yielded superior results so the kernel type was not varied for the hyperparameter optimization runs. Gamma, the distance over which a sample has influence, ranged from 0.01 to 10 divided by the number of features. The error tolerance parameter, C, which influences the smoothness of the decision surface, varied from 0.1 to 1000. Optimum values for gamma were in the range of 1-10 and 100-1000 for C. Except for lower parameter values, the cross-validation scores were relatively insensitive to variations in the hyperparameters.

3.2. Impact of Spectral Sampling on LULC

The ground truth was split into training and test samples using an optimal proportion determined by single runs. The test runs used splits ranging from 0.1 to 0.9 in increments of 0.1; see Fig. 1 for the test results. A training set of 70% was selected for both classifiers, although any value between 60% and 80% would have yielded similar results.

Table 2 summarizes the average normalized kappa over five runs using 70% of the ground truth as training samples. Results of the SVM classifier are presented below because cross-validation scores about 2-7% higher than RF. Note that no indices, textures, or spatial features were used.

3.3. LULC classifications

LULCs using the SVM classifier are shown in Fig. 2. The hyperspectral LULC is sharply-delineated but with some confusion for Self-blocking bricks. The WV-3 LULC appears less distinct in its building footprints for a couple of reasons: 1) the off-nadir angle exposed the sides of the buildings, which

Table 3. Precision/Recall per class for the SVM classifier.

	Adj. GT				Orig. GT
LULC Classes	WV-3/4	WV-3/8	WV-3/16	Hypersp	Hypersp
Water	0.98 / 1.00	0.99 / 1.00	0.99 / 1.00	1.00 / 1.00	1.00 / 1.00
Trees	0.81 / 0.92	0.83 / 0.93	0.97 / 0.98	0.95 / 1.00	0.97 / 0.99
Meadows	0.92 / 0.82	0.94 / 0.85	0.98 / 0.98	0.99 / 0.95	0.99 / 0.97
Self-blk bricks	0.71 / 0.53	0.83 / 0.74	0.98 / 0.97	0.95 / 0.96	0.98 / 0.97
Bare soil	0.89 / 0.57	0.94 / 0.75	1.00 / 0.96	0.99 / 0.95	0.98 / 0.99
Asphalt	0.58 / 0.77	0.75 / 0.83	0.96 / 0.95	0.97 / 0.98	0.97 / 0.99
Bitumen	0.91 / 0.94	0.94 / 0.96	1.00 / 1.00	0.96 / 0.97	0.98 / 0.97
Tiles	0.81 / 0.98	0.85 / 0.98	0.96 / 0.99	0.96 / 0.99	1.00 / 1.00
Shadow	0.93 / 0.94	0.94 / 0.94	0.96 / 0.97	0.98 / 0.97	1.00 / 1.00

were classified as Tiles for lack of a better option, and 2) the low sun angle cast long shadows. The differences between an LULC based on 16-, 8-, and 4-band WV-3 images are most apparent in the increasing confusion between the Asphalt/Tiles and Meadows/Trees classes.

Precision and recall scores, averaged over five runs, are given in Table 3. The general trend is an increase in the score with the number of bands. The hyperspectral LULC produced with the adjusted ground truth yielded somewhat lower scores than with the original ground truth (also in table). Results for the WV-3 16-band image are similar to the hyperspectral for this particular set of classes. Notable is the improvement in precision/recall with increasing bands for the man-made materials: Self-blocking bricks, Asphalt, and Tiles.

3.4. Importance of Spectrum Sections

The features important for the RF LULC classifications are easily obtained in scikit-learn. Some Water and Tile samples were removed from the ground truth to balance class sizes because feature importances are influenced by the frequency that a node in the decision tree is used for classification and thus, imbalanced classes bias the feature importances.

In order to make the feature importances for each case comparable, results for each sensor/band combination are normalized to 1.0 using the maximum feature importance of the relevant LULC (see Fig. 3). As seen in the figure, for LULC classifications with only VNIR bands (hyperspectral, WV-3 8 bands, WV-3 4 bands), the red, red edge, and NIR bands are the most important. With the addition of SWIR bands for the WV-3 16 bands case, the NIR and SWIR bands in the upper end of the spectrum stand out in importance.

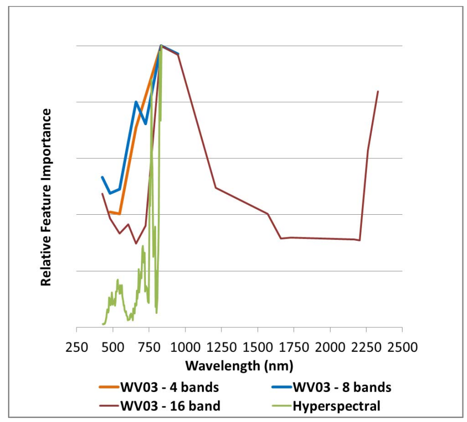


Fig. 3. Relative feature importance for the RF classification for each case. Results are normalized to the maximum feature importance for each sensor/band combination.

4. CONCLUSION

Because of the availability of hyperspectral data and ground truth, the city center of Pavia, Italy, was selected as an area to explore the benefit of added spectral capabilities of DigitalGlobe's WV-3 satellite to LULC classification. For the SVM classifier, the baseline comparison, using only spectral bands as features, yields similar results for the 16-band WV-3 imagery and hyperspectral images. The advantage of 102 narrow hyperspectral bands is balanced by the advantage of WV-3's SWIR bands. Precision and recall scores for hyperspectral and WV-3 16-band LULCs exceeds those of the 8-band and 4-band, primarily for man-made materials.

Results using the RF classifier show a similar trend, though the precision/recall scores are lower by 2-7%. RF scores would likely improve with the use of indices, textures, or spatial features. Relative importances of the features determined by the RF classifier suggest that, in addition to bands in the red to NIR spectrum, the upper end of the SWIR range contains additional information that improves precision/recall for this particular set of classes. Future work will address comparisons of the LULC integrating band ratios, spatial features, feature embedding, and feature selection. The release of the Pavia 16-band WV-3 image to the Geoscience and Remote Sensing community for scientific research will allow further investigation into the benefits of SWIR.

5. REFERENCES

[1] J. Townshend, C. Justice, and W. Li et al., "Global Land Cover Classification by Remote Sensing - Present Capabilities and Future Possibilities," *Remote Sensing of the Environment*, vol. 35, pp. 243–255, 1991.

[2] M. Friedl, D. McIver, and J. Hodges et al., "Global Land Cover Mapping from MODIS: Algorithms and Early Results," *Remote Sensing of the Environment*, vol. 83, pp. 287–302, 2002.

[3] P. Gamba, "A Collection of Data for Urban Area Characterization," in *Proc. IEEE International Geoscience and Remote Sensing Symposium*, 2004.

[4] F. Dell'Acqua, "Exploiting Spectral and Spatial Information for Classifying Hyperspectral Data in Urban Areas," in *Proc. IEEE International Geoscience and Remote Sensing Symposium*, 2003.

[5] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J.s R. Sveinsson, "Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles," in *Proc. IEEE International Geoscience and Remote Sensing Symposium*, 2008.

[6] N. Longbotham, F. Pacifici, B. Baugh, and G. Camps-Valls, "Prelaunch Assessment of WorldView-3 Information Content," in *Proc. IEEE Conference Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, WHISPERS*, 2013.

[7] F. Pedregosa, G. Varoquauz, and A. Gramfort et al., "scikit learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.