MULTI-SENSOR MULTI-RESOLUTION IMAGE FUSION FOR IMPROVED VEGETATION AND URBAN AREA CLASSIFICATION

Uttam Kumar [a], Cristina Milesi [b], Ramakrishna R. Nemani [b], Saikat Basu [c]

[a] NASA Ames Research Center / ORAU, Moffett Field, California 94035, USA - uttam.kumar@nasa.gov (Corresponding Author)
[b] NASA Ames Research Center, Moffett Field, California 94035, USA - (cristina.milesi, rama.nemani)@nasa.gov
[c] Department of Computer Science, Louisiana State University, Baton Rouge, LA 80803, USA -
sbasu8@lsu.edu

KEY WORDS: multi-sensor, multi-resolution, linear mixture model, data fusion, classification

ABSTRACT:

In this paper, we perform multi-sensor multi-resolution data fusion of Landsat-5 TM bands (at 30 m spatial resolution) and multispectral bands of World View-2 (WV-2 at 2 m spatial resolution) through linear spectral unmixing model. The advantages of fusing Landsat and WV-2 data are two fold: first, spatial resolution of the Landsat bands increases to WV-2 resolution. Second, integration of data from two sensors allows two additional SWIR bands from Landsat data to the fused product which have advantages such as improved atmospheric transparency and material identification, for example, urban features, construction materials, moisture contents of soil and vegetation, etc. In 150 separate experiments, WV-2 data were clustered in to 5, 10, 15, 20 and 25 spectral classes and data fusion were performed with 3x3, 5x5, 7x7, 9x9 and 11x11 kernel sizes for each Landsat band. The optimal fused bands were selected based on Pearson product-moment correlation coefficient, RMSE (root mean square error) and ERGAS index and were subsequently used for vegetation, urban area and dark objects (deep water, shadows) classification using Random Forest classifier for a test site near Golden Gate Bridge, San Francisco, California, USA. Accuracy assessment of the classified images through error matrix before and after fusion showed that the overall accuracy and Kappa for fused data classification (93.74%, 0.91) was much higher than Landsat data classification (72.71%, 0.70) and WV-2 data classification (74.99%, 0.71). This approach increased the spatial resolution of Landsat data to WV-2 spatial resolution while retaining the original Landsat spectral bands with significant improvement in classification.

1. INTRODUCTION

In the geospatial terminology, land cover (LC) refers to the physical state of the Earth's surface in terms of natural environment, such as vegetation, settlement, water, etc. Vegetation and urban areas are currently among the most rapidly changing LC types and are sites of significant natural resource transformation (Lambin et al., 2001). The conversion of most productive-vegetative land and farmland to impervious surfaces (settlements) is leading to environmental degradation and ultimately influencing weather and climate of the Earth from local to global levels. Therefore, mapping and assessing the status of vegetation and urban LC classes at various scales and time frames is required for mitigation, planning, and decision-making. The LC patterns can be captured through data acquired from space-borne remote sensing (RS) satellites that facilitate observations across larger extent of the Earth's surface. With the availability of multi-resolution RS data from operational Earth observation satellites, each with unique spatial, spectral, radiometric and temporal characteristics, fusion of digital images has become a valuable tool for cost effective LC mapping. Often, integration of data from multiple sensors using numerous data fusion techniques (Castanedo, 2013) aids in delineating objects with comprehensive information due to the enhanced details. Multi-sensor image fusion seeks to combine information from different images to obtain more inferences than can be derived from a single sensor (Dong et al., 2009).

A common approach in multi-sensor data environment is to combine the geometric detail of a high spatial resolution (HSR) Panchromatic (PAN) image and the spectral information of a low spatial resolution (LSR) Multispectral (MS) image to produce a final image with the highest possible spatial information content while preserving good spectral information quality. Other recent advances such as MS and LiDAR image fusion (IEEE GRSS Data Fusion Contest, 2013), and LiDAR and SAR data have been used for many LC studies and urban research (Berger et al., 2013; Makarau et al., 2011). The multi-sensor data integration aims to provide fused data for robust operational performance, i.e., increased confidence, reduced ambiguity, improved reliability and improved classification (Rogers and Wood, 1990). It is mainly used for sharpening images, improving registration accuracy, creation of stereo data sets, feature enhancement, improving image classification, change detection, substituting missing information (e.g., clouds-VIR, shadows-SAR) in one image with signals from another sensor image, replacement of defective data, topographic mapping and map updating, flood monitoring, ice/snow monitoring, geological mapping, etc. At this resolution, image fusion is performed at pixel level, feature level and decision level. Selection of fusion technique depends on many factors: (i) the objective/application of the user; (ii) types of data that are most useful for meeting the user's need; (iii) selection of the best technique and combination of the data sources which are most successful for the user's application. In order to address a particular application, it is necessary to have apt spectral and spatial resolution, which is a constrain by availability. Availability depends on the satellite coverage, operational aspects, atmospheric constraints such as cloud cover, economic issues, etc. (Pohl and Van Genderen, 1998).

Numerous techniques to fuse PAN with MS data, MS with LiDAR data, etc. have been proposed in literatures. For a review of the state-of-the-art, status and trends in multi-source remote sensing data fusion see Khaleghi et al. (2013) and Zhang

(2010). Many applications such as urban planning, species mapping, etc. require fusion of MS data from one sensor with MS data or hyperspectral data acquired from another sensor in order to combine different spectral bands at the higher spatial resolution. In this context, recently, attempts to fuse long wave infrared (LWIR, thermal infrared) hyperspectral data set in 84 channels (between 7.8 to 11.5 µm) at 1 m spatial resolution and HSR MS data acquired in the visible wavelength range with 20 cm spatial resolution, followed by classification of the data in to seven different land use classes for an urban area near Thetford Mines, Québec, Canada were made by many researchers in the 2014 IEEE GRSS Data Fusion Contest (IEEE GRSS Data Fusion Contest, 2014). Approaches to use principal component analysis (PCA), use of textural features, vegetation index and morphological building index were explored in a hierarchical classification framework. Other attempts to fuse and classify the above data were made through multi-scale segmentation, feature extraction with object-oriented classification, SVM classification with contextual information in a hierarchical schema, use of GLCM (Grey-Level Co-Occurrence Matrix) and 3D-DWT (Discrete Wavelet Transform) textures with different levels of success. All of the above approaches used HSR data sets (1 m and 20 cm) in which objects are easily discernible. However, when the MS data have LSR pixels representing mixture of multiple objects (forming mixed pixels), they pose difficulty in classification and interpretation. In such cases, image fusion techniques that can fuse or merge MS with MS/hyperspectral bands, such as IRS LISS-III MS bands with MODIS bands, or IKONOS MS bands with Landsat TM MS bands are required. Here, HSR bands are used to analyze the composition of mixed pixels in images obtained by LSR sensors in order to unmix them for their efficient utilization for many land-oriented applications (Zhukov et al., 1999a).

In this paper, we use linear mixture model (LMM) to fuse Landsat-5 Thematic Mapper (TM) bands with World View-2 (WV-2) bands. LMM is often used to solve the mixed pixel problem in medium and coarse spatial resolution pixels. LMM estimates the abundance or proportion of each class within individual pixels assuming that the reflectance spectrum of a mixture is systematic combination of the component's reflectance spectra in the mixture (called endmembers). The optimal solution of the mixture models can be unconstrained or constrained (when the abundance nonnegativity constraint (ANC) and abundance sum-to-one constraints (ASC) are imposed). ANC restricts the abundance values from being negative and ASC confines the sum of abundance values of all the classes to one. In LMM based fusion, first a HSR data (WV-2) is clustered using any clustering technique such as K-means clustering algorithm into user-defined number of unknown classes. This clustered image provides abundance or proportion of different classes corresponding to each pixel of the LSR Landsat data. The reflectance spectrum is obtained from Landsat pixels and the LMM equations are inverted in a moving window to obtain the individual class reflectance spectra. The initial number of clusters in the HSR data and the size of moving window are heuristic. In other words, this iteration is continued for all the pixels in the LSR bands while solving the LMM equations with proportions from HSR clustered image to obtain user-defined number of class's spectra. Finally, these individual class's spectra are assigned to the corresponding pixel's location for which class labels are known from the HSR clustered image to form HSR fused images. LMM based data fusion technique, also called unmixing based data fusion generates images that have the spatial resolution of WV-2 data and the spectral resolution provided by Landsat TM sensor. The advantages of fusing Landsat and WV-2 data are two fold. First,

spatial resolution of the Landsat bands (at 30 m) increase to WV-2 resolution (2 m). Second, WV-2 spectral bands do not encompass the SWIR wavelength in the Electromagnetic spectrum. If one were to use only WV-2 data in classification as it is of HSR, the benefits of using SWIR spectral bands would be overlooked, as WV-2 data do not have any bands in this region of the Electromagnetic spectrum. Hence integration of data from two sensors would allow two additional SWIR bands from Landsat data to the fused product. Images acquired in SWIR wavelength have advantages such as improved atmospheric transparency and material identification. Many materials have specific reflectance and absorption features in the SWIR bands that allow for their characterization, for example, urban features, such as roofing and construction materials, moisture contents of soil and vegetation, mineral exploration, etc. Snow and ice display distinctive variations in SWIR bands and energy in SWIR wavelength can even penetrate smoke, such as from a forest fire. In the present study, data fusion experiments were carried out using window sizes of 3, 5, 7, 9, and 11 on the LSR data and the number of unknown classes in the HSR clustered image were set to 5, 10, 15, 20 and 25 for each window size. Assessment of the fused bands was carried out with the original LSR bands as reference using various statistics. The best-fused data were then classified with Random Forest classifier in to three classes (substrate, vegetation and dark objects), which showed higher classification accuracies compared to classification of original Landsat bands and WV-2 bands. The paper is organized as follows: section 2 discusses linear mixture model, section 3 discusses LMM based image fusion and section 4 details the data and methods used in this analysis. Results and discussion are presented in section 5 with concluding remarks in section 6.

2. LINEAR MIXTURE MODEL (LMM)

If there are M spectral bands and N classes, then associated with each pixel is a M-dimensional vector $\mathbf{y}$ whose components are the gray values corresponding to the M bands. Let $\mathbf{E} = [e_1, ...e_{n-1}, e_n, e_{n+1} ..., e_N]$ be a $M \times N$ matrix, where $\{e_n\}$ is a column vector representing the spectral signature (endmember) of the nth target material. For a given pixel, the abundance or fraction of the nth target material present in the pixel is denoted by α_n, and these values are the components of the N-dimensional abundance vector $\boldsymbol{\alpha}$. Assuming LMM (Shimabukuro and Smith, 1991), the spectral response of a pixel in any given spectral band is a linear combination of all the endmembers present in the pixel at the respective spectral band. For each pixel, the observation vector $\mathbf{y}$ is related to $\mathbf{E}$ by a linear model written as

$$\mathbf{y} = \mathbf{E}\boldsymbol{\alpha} + \boldsymbol{\eta} \qquad (1)$$

where $\boldsymbol{\eta}$ accounts for the measurement noise. We further assume that the components of the noise vector $\boldsymbol{\eta}$ are zero-mean random variables that are i.i.d. (independent and identically distributed). Therefore, covariance matrix of the noise vector is $\sigma^2\mathbf{I}$, where σ^2 is the variance, and $\mathbf{I}$ is $M \times M$ identity matrix.

3. LMM BASED DATA FUSION

Mixed pixels in LSR images can be analyzed with spectral endmember unmixing techniques that use endmember proportions in mixed pixels rather than to reconstruct an actual HSR image in the LSR bands. Multi-sensor multi-resolution data fusion can be used for unmixing of LSR pixels and reconstruction of an image in the LSR bands with the HSR pixel

size in case of mixed pixels in the LSR bands (Zhukov et al., 1999b). The endmembers are neither obtained from an endmember library nor extracted from the images. It is assumed that features that are recognizable in the HSR image, have the same LSR signals in the central LSR pixel as in the surrounding LSR pixels in a moving window (Zhukov et al., 1995, 1999a, 1999b; Zurita-Milla et al., 2006, 2008; Mezned et al., 2010). The algorithm is briefed as follows:

(i) Classification of the HSR image – This is performed using any unsupervised classification algorithm such as K-means clustering to perform a spectral classification. The result of classification is a HSR classified map $n(x,y)$, with code n as the corresponding classes.

(ii) Computation of class contributions to the signal of the LSR pixels – This is performed by accounting the class area proportion in the LSR pixels obtained from the classified map $n(x,y)$ and the point spread function (PSF's) of the LSR channels:

$$c_m(l,s,n_0) = \sum_{n(x,y) \in n_0} DA_m(l,s,x,y) \qquad (2)$$

where $c_m(l,s,n_0)$ is the contribution of class n_0 to the signal of LSR pixel (l, s) in the LSR band m, and $DA_m(l,s,x,y)$ is a discrete approximation for the sensor PSF. The discrete PSF gives the contribution from the area of a HSR pixel (x, y) to the LSR signal of a pixel (l, s) in band m. Its sum over (x, y) is assumed to be normalised to 1. PSF accounts for the sensor related effects as well as for resampling (if the LSR image is resampled to co-register with the HSR image).

(iii) Window-based unmixing of the LSR pixels – Unmixing of the LSR data is performed consecutively in a window that is moved with the step of 1 LSR pixel size. In each window, the central LSR pixel is unmixed by an inversion of a system of linear mixture equations for all the LSR pixels in the window.

$$S_m(l,s) = \sum_{n=1}^{N} c_m(l,s,n)\, \bar{S}_m(n) + \varepsilon_m(l,s) \qquad (3)$$

where $S_m(l, s)$ is the signal of the LSR pixel (l, s) in the window, $\bar{S}_m(n)$ is the mean LSR signal for class n in the window, and $\varepsilon_m(l,s)$ is the model error. The sources of model error are radiometric noise, co-registration errors, etc. that may cause a difference of class signals in various LSR pixels from the mean signal $\bar{S}_m(n)$ in the window. Equation (3) returns a vector $\bar{S}$ that minimises norm $(C * \bar{S} - S)$ subject to $\bar{S} \geq 0$. See Mezned et al., (2007) for more details.

(iv) The inversion of linear model by least squares optimization method is carried out for each LSR band to estimate mean class signal. The inversion is most likely ill conditioned, which increases the noise in the solution and is stabilized by minimizing the cost function F with a regularization parameter that restricts large deviations of the estimated class spectra:

$$F = \sum_{(l,s) \neq (l_0,s_0)} (S_m(l,s) - \sum_{n=1}^{N} c_m(l,s,n)\, \bar{S}_m(n))^2 + \alpha \frac{NP_{low\_res} - 1}{N} \sum_{n=1}^{N} (\bar{S}_m(n) - \bar{S}'_m(n))^2$$

$$(4)$$

where N and NP_{low_res} are number of classes and number of LSR pixels in the window, α is the regularization

parameter for regularization of the solution. It limits large deviations of the LSR class signal. Radiometric noise and georegistration errors may contribute to the linear model error.

(v) Restoration of unmixed fused bands – This is performed by assigning the estimated mean class signal $\bar{S}_m^*(n)$ to the corresponding HSR classified pixels

$$H_m^*(x,y) = \bar{S}_m^*(n(x,y)) \qquad (5)$$

where $(x,y)\varepsilon(l_0,s_0)$. It is done in each window within the area of the central LSR pixel (l_0, s_0). Movement of the window over the image in the steps of 1 LSR pixel reconstructs the unmixed image.

4. DATA AND METHODS

A set of coincident clear sky Landsat-5 TM data of dimension (478 rows x 578 columns) and WV-2 data (of size 7170 rows x 8670 columns) for an area of San Francisco (SF) were used in the experiments. SF is chosen for the test site because of its urbanized landscape that has mix of buildings of different construction materials, shadows of tall buildings, roads with mixture of trees, narrow streets with pavements, urban forest, parks, etc. as shown in the false colour composite (FCC) of WV-2 data in figure 1. 2 m resolution of WV-2 data is adequate to capture small to medium sized buildings, sidewalks, streets and trees with shadows as evident from figure 1. At 30 m resolution, each 2 m WV-2 pixel is even less than 0.5% of the area within the 30 m full width half maximum of Landsat point spread function.
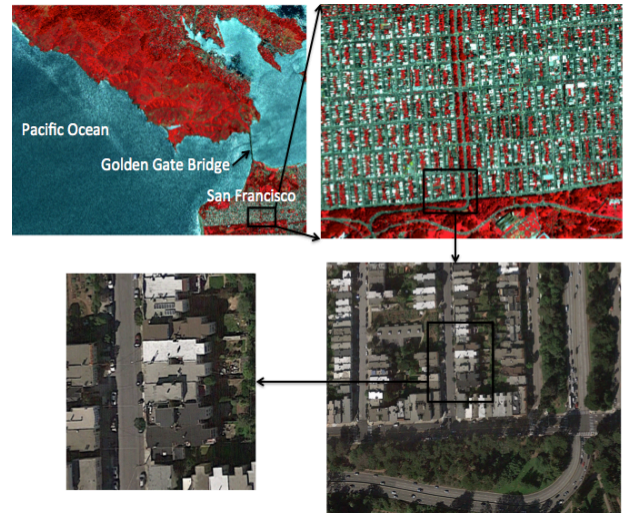

Figure 1. Study area: False colour composite (FCC) of a part of San Francisco city. Zoomed image of the urban area (marked with rectangle in inset) shows mixing of substrate with vegetation, roads, shadows and dark objects.

Atmospheric reflectance of the Landsat bands were converted to surface reflectance correcting for atmospheric effects by means of the 6S code implementation in the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) atmospheric correction method (Masek et al., 2006). WV-2 data were acquired a few minutes after the Landsat-5 TM data acquisition on May 1, 2010 for an area near the Golden Gate Bridge. The spectral range of first four bands (Blue, Green, Red

and NIR) of Landsat data have a good correspondence with the WV-2 bands 2, 3, 5 and 7 in terms of wavelength range in the Electromagnetic spectrum and therefore they have a similar mixing space. Table 1 lists the spectral properties of Landsat-5 and WV-2 data. WV-2 data were converted to Top of Atmosphere Reflectance values using the python program (available at https://github.com/egoddard/i.wv2.toar) in GRASS GIS 7.1. It is to be noted that WV-2 data have 8 spectral bands, out of which only 4 bands have spectral overlap with the Landsat-5 TM bands. Coordinate comparison of the HSR and LSR data sets at many random pixels corresponding to same spatial location did not reveal any systematic image registration error. A quantitative assessment of the quality of the fused images was done at the level of LSR images. The fused images were degraded (*FHSRD*) to the original LSR image (*OLSR*) and the quality of the degraded images were assessed by comparing them with the original LSR image using Pearson product-moment correlation coefficient (cc or *r*), root mean square error (RMSE) and ERGAS (Erreur Relative Globale Adimensionalle de Synthése / Relative Dimensionless Global Error) (Wald, 2002).

Table 1. Spectral properties of the Landsat and WV-2 bands

Landsat			WV-2	
Band	Band	Wavelength (μm)	Band	Wavelength (μm)
1	Blue	0.45 – 0.52	Coastal	0.40 – 0.45
2	Green	0.52 – 0.60	Blue	0.45 – 0.51
3	Red	0.63 – 0.69	Green	0.51 – 0.58
4	NIR	0.77 – 0.90	Yellow	0.59 – 0.63
5	SWIR1	1.55 – 1.75	Red	0.63 – 0.69
6	TIR	10.4 – 12.5	Red Edge	0.70 – 0.74
7	SWIR2	2.09 – 2.35	NIR1	0.77 – 0.895
8	PAN	0.52 – 1.90	NIR2	0.86 – 0.95

Considering P = number of observed M-dimensional pixel vector,

$$ r = \frac{\sum_{p=1}^{P}(OLSR_p - \overline{OLSR})(FHSRD_p - \overline{FHSRD})}{\sqrt{\sum_{p=1}^{P}(OLSR_p - \overline{OLSR})^2}\sqrt{\sum_{p=1}^{P}(FHSRD_p - \overline{FHSRD})^2}} \qquad (6) $$

r ranges from −1 to 1. 1 implies that a linear equation describes the relationship between *OLSR* and *FHSRD* perfectly, with all the data points lying on a straight line for which *FHSRD* increases as *OLSR* increases. $r = -1$ infers that all data points lie on a line for which *FHSRD* decreases as *OLSR* increases and $r = 0$ means there is no linear correlation between *OLSR* and *FHSRD*. RMSE (Nascimento and Dias, 2005) is defined as:

$$ RMSE = \sqrt{\frac{1}{P}\sum_{p=1}^{P}(OLSR_p - FHSRD_p)^2} \qquad (7) $$

RMSE is calculated for each band separately. Smaller the RMSE, better the unmixing result and higher is the accuracy. The ERGAS index (as given by (8)) equals zero when the degraded fused image is equal to the original image. Therefore, low ERGAS values indicate high image fusion quality.

$$ ERGAS = 100\frac{HSR}{LSR}\sqrt{\frac{1}{M}\sum_{m=1}^{M}(rmse_m^2/\mu_m^2)} \qquad (8) $$

where,
- HSR is the resolution of the high spatial resolution image;
- LSR is the resolution of the low spatial resolution image;
- *M* is the number of spectral bands involved in fusion;
- $rmse_m$ is the root mean square error computed between the degraded fused image and the original image for band *m*;
- μ_m is the mean value of band *m* of the reference image.

The fused Landsat bands that retained the spectral property of the original Landsat bands (before fusion) were classified using Random Forest (RF) classifier with global endmembers (class signatures). RF are ensemble methods based on tree-type classifiers that uses bagging to form classification tree (Breiman, 2001; Gislason, 2006; Walton, 2008). RF is light in implementation and is distinguished from other bagging approaches in that at each splitting node in the underlying classification trees, a random subset of the predictor variables is used as potential variables to define split (Breiman and Cutler, 2010; Na et al., 2010). In training, it creates multiple CART (Classification and Regression Tree) trained on a bootstrapped sample of the original training data, and searches only across randomly selected subset of the input variables to determine a split for each node. While classification, each tree casts a unit vote for the most popular class at input. The output of the classifier is determined by a majority vote of the trees. The classified outputs were assessed using error matrix by computing producer's accuracy, user's accuracy, overall accuracy, kappa statistic and Quality index (Q) (Rutzinger et al., 2009) given by (9):

$$ Q = \frac{||TP||}{||TP|| + ||FP|| + ||FN||} \qquad (9) $$

where, TP = true positive, FP = false positive and FN = false negative cases.

To obtain the class signatures (endmembers) for classification, global mixing spaces were sampled by using a spectrally diverse LC and diversity of biomes with 100 Landsat ETM+ scenes (Small and Milesi, 2013). This defined a standardized set of spectral class's endmembers of substrate (S), vegetation (V), and dark objects (D). Substrate includes soils, sediments, rocks, and non-photosynthetic vegetation; vegetation refers to green photosynthetic plants; and dark objects encompass absorptive substrate materials, clear water, deep shadows, etc. For simplicity, we refer substrate, vegetation and dark objects as "S", "V", and "D" (SVD) in the rest of this paper. The SVD endmember coefficient, in addition to dates and locations of each subscene are available at http://www.LDEO.columbia.edu/~small/GlobalLandsat/. The estimates obtained from the global endmembers have been compared to fractional vegetation cover derived vicariously by linearly unmixing near-coincidental WV-2 acquisitions over a set of diverse coastal environments, using both global endmembers and image-specific endmembers to unmix the WV-2 images. The strong 1:1 linear correlation between the fractions obtained from the two types of images indicate that the mixture model fractions scale linearly from 2 m to 30 m over a wide range of LC types (Small, 2004). These endmembers were used to classify the fused data. The overall methodology is summarized below:

1.) Atmospheric correction of Landsat bands and conversion of atmospheric reflectance to surface reflectance.
2.) Conversion of WV-2 bands to Top of Atmosphere (TOA) reflectance.

3.) Pixel to pixel registration of Landsat and WV-2 bands.

4.) Clustering of WV-2 data into 5, 10, 15, 20 and 25 unknown classes.

5.) Generation of linear mixture equations with Landsat pixels as LSR reflectance (**y** in equation (1)), and class proportions (**α** in equation (1)) from WV-2 clustered image with window size of 3, 5, 7, 9 and 11, for each Landsat band separately.

6.) Solving the inverse optimization problem in equation (1) while minimizing the cost function to estimate class's reflectance (**E** in equation (1)) for each LSR pixel in the moving window.

7.) Selection of optimal number of clusters in the HSR data and choice of appropriate window size for the LSR band by assessment of the degraded fused bands with reference to the original Landsat bands using cc, RMSE and ERGAS to obtain best LMM based fused data.

8.) Classification of the fused data with global endmembers into SVD classes using RF followed by accuracy assessment.

5. RESULTS AND DISCUSSION

150 experiments of LMM based Landsat and WV-2 data fusion were carried out with 5, 10, 15, 20 and 25 clusters and 5 different window sizes. Figure 2 shows FCC generated from Landsat bands (4-3-2), WV-2 bands (7-5-3), fused band (4-3-2) with 5 clusters and 3 x 3 window, and fused bands (4-3-2) with 25 clusters and 5 x 5 window. Visual comparison of these fused images showed that they were very similar. However, in figure 2 (b) large areas of the Pacific Ocean (in the eastern region of the image) have similar reflectance as that of the San Francisco urban region (in the south-eastern area of the image). This can be misleading as deep water looks similar to urban surfaces. In general, the fused images preserved the spatial patterns of WV-2 images while remaining spectrally similar to the Landsat TM images. There are some pixels in the water area (Pacific Ocean) in figure 2 (c), where pixels show salt and pepper effects. These pixels have composition of two or more classes and are difficult to unmix because WV-2 clustered image have very few of these pixels by proportion in the unmixing window. Since fused image should be as identical as possible to the original LSR image once degraded back to its original resolution, so the fused bands were degraded to 30 m using a mean filter. The quality of the degraded fused images was then assessed by comparing them with the original Landsat TM image. The ERGAS values revealed that LMM based data fusion succeeded in synthesizing the spectral information of the Landsat TM bands at the WV-2 resolution. The evaluation of the fused bands revealed several important aspects:

(i) Comparison of statistical distribution of the original Landsat and fused bands demonstrated a significant improvement in sharpness and radiometric accuracy of the fused bands in comparison to the original bands.

(ii) Cc (which were found to be statistically significant at 0.99 confidence level, p-value $< 2.2e^{-16}$) between degraded fused bands and original Landsat bands were higher for smaller window sizes and kept on decreasing as the window size increased. It concludes that smaller window size is able to retain the spectral variability and larger window size averages the spectral values in the neighbourhood. This also indicates that solution of the unmixing equations stabilizes with few equations because of smaller window size and the regularization parameters might not be important in these cases.

(iii) RMSE and ERGAS values between degraded fused bands and original Landsat bands increased monotonically with the increase in window size.

(iv) Cc between degraded fused bands and original Landsat bands were lower for smaller number of clusters in the HSR image and improved as the number of clusters increased in the HSR image.

(v) RMSE and ERGAS values decreased with the increasing number of clusters in the HSR image. Therefore, cc and RMSE/ERGAS are inversely proportional.
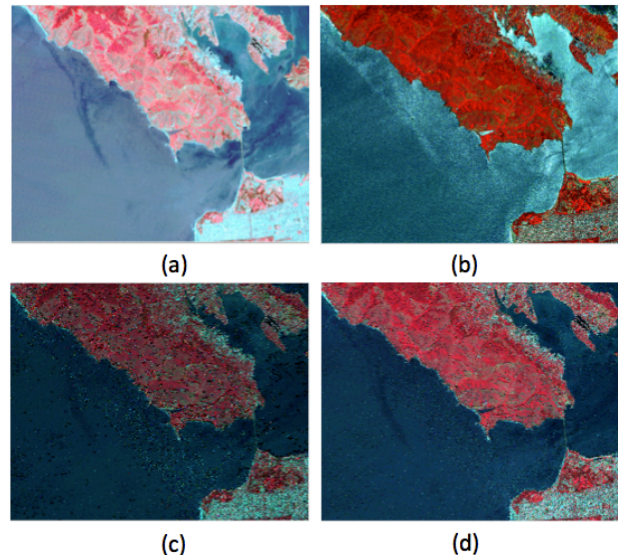

Figure 2. (a) FCC of the original spectral bands, (b) FCC of the WV-2 bands, (c) FCC of the bands obtained from 5 clusters and 3 x 3 windows, and (d) FCC of the bands obtained from 25 clusters and 5 x 5 window.

The window size should be small so that the fused images have variation in the dynamic range of the pixels and are consistent with the variability in the LSR pixels. A limitation of the window size is that the number of unknown class's signature to be resolved should be ≤ the number of pixels in the window size. Therefore, with 3 x 3 window size, a maximum of 9 class's signatures could be retrieved. However with large window size, the variability in the fused image diminished, because each system of equation results in a unique solution. It is obvious that if the size of the window is equivalent to the image size, then all the pixels of one class identified in WV-2 data will have the same spectral reflectance regardless of their spatial location in the scene and will result in a fused band with very low dynamic range. The LMM based fusion is realistic in the condition when the spectral classes are identified in the HSR image. A homogeneous class in HSR image but heterogeneous in LSR image will lead to averaging of its LSR signature over the class area in the window while utilizing all the spectral information in the LSR data by imposing ANC and ASC (Zurita-Milla et al., 2008). However, varied LSR signals of classes located in different windows are treated independently without averaging of the LSR signatures. Therefore, LMM based fusion can be applied in applications where the classes of interest are known a priori. These observations are in agreement with MERIS and Landsat TM data merging (Zurita-Milla et al., 2008), Landsat TM visible and IR bands with Landsat TIR band fusion (Zhukov et al., 1999a), and Landsat ETM+ and ASTER data fusion (Mezned et al., 2007). There are some limitations to LMM based fusion method such as the number of classes, their

separability in the HSR image and the window size, which should be small enough to reduce spatial averaging of the LSR class signatures, at the same time, it should have sufficient number of LSR pixels for a stable inversion of the LMM equation. The optimal combination of clusters and window size for best fused output was found to be 25 and 5 x 5. Although still higher number of clusters in the HSR data was possible, it was found that ERGAS was nearly becoming stable. Finally, the fused bands were classified into three classes – SVD with RF. Figure 3 shows classified output from Landsat TM bands at 30 m (a), WV-2 data classification at 2 m (b), and fused bands at 2 m (c). It shows that the WV-2 classified image and fused classified image exhibit more details than the Landsat classified image. Many of the pixels belonging to substrate class in the lower right portion of the image (the SF city area) have been classified properly in the WV-2 data and fused data classification. Those areas are mix of urban features and urban trees forming mixed pixels in Landsat data. Also, extent of vegetative areas has been mapped with greater spatial accuracy in both WV-2 data and fused data classification, although visually both classified images show relatively similar proportions of the vegetation class. A careful observation of the FCC in figure 1 (top left image) and figure 2 (b) shows that in fact, substrate which comprises soils, sediments, rocks, and non-photosynthetic vegetation are found in the ocean close to land areas and along the sand beaches with shallow water near
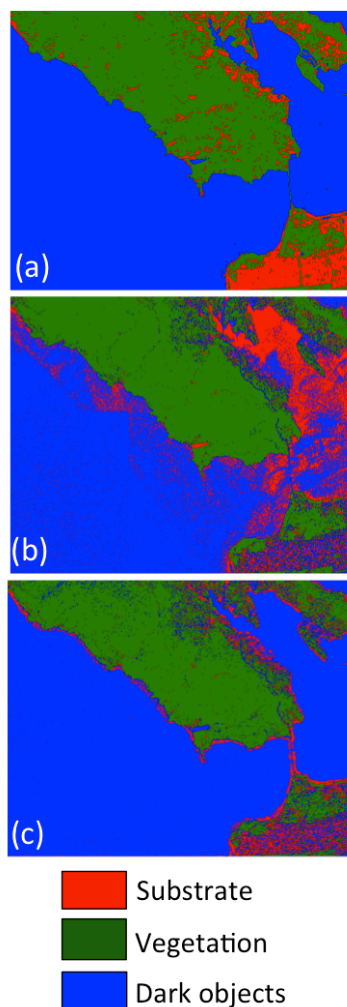

Figure 3: (a) Classified images from Landsat, (b) WV-2 data, and (c) fused data.
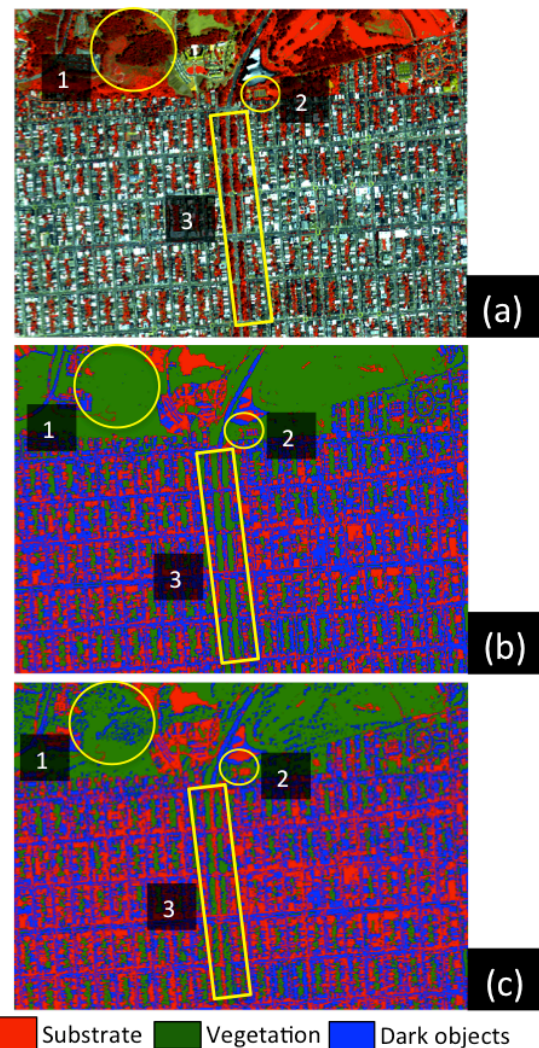

Figure 4: (a) FCC of a downtown area in San Francisco, (b) classified image of the original WV-2 data, (c) classified image of the fused Landsat and WV-2 data.

the Golden Gate Bridge. The reflectance of settlements in the city is very similar to sand and low depth surface water in the ocean. This phenomenon is very evident in figure 3 (c) with red pixels (substrate class) in the fused classified image. Dark objects, which encompass deep water, shadows, absorptive substrate materials appear to be mixed with substrate in the WV-2 classified image; a larger part of the Pacific Ocean in the first quadrat of the image in figure 3 (b) are classified as substrate. This is not surprising and is expected as seen in figure 2 (b) where the corresponding area (deep water) has very similar reflectance as that of the substrate in bottom right region. However, after fusion of the Landsat and WV-2 bands, the difference in substrate (urban pixels) and dark objects (deep water in the Pacific Ocean) is apparent in figure 2 (d). A detailed analysis of the urban areas (buildings/concrete structures/roads with presence of vegetation) showed that the classified fused output was better in discriminating substrate and dark object classes. For example, location "1" highlighted in figure 4 (a) shows shadows from dense trees, location "2" shows building with small structures on the terrace and location "3" shows a tarred road with vegetation on both sides. All 3 locations have been misclassified as either absence of presence of dark object class in the WV-2 classified output (shown in figure 4 (b)). These features have been properly classified in the

fused data as shown in figure 4 (c). Table 2 indicates the producer's and user's accuracies. The overall accuracy and Kappa for fused data classification (93.74%, 0.91) was higher than Landsat data classification (72.71%, 0.70) and WV-2 data classification (74.99, 0.71). For any particular class, if the reference data has more pixels with correct label, the producer's accuracy is higher and if the pixels with the incorrect label in classification result is less, its user's accuracy is higher (Mingguo et al., 2009). There was not much improvement in accuracy for WV-2 data classification compared to Landsat data classification except for vegetation class. Producer's accuracy increased for substrate (12.6%), vegetation (14.3%), and dark objects (27%), and user's accuracy increased for substrate (18%), vegetation (26.7%), and dark objects (27.6%) in fused classified output compared to Landsat data classification. Fused data classification was intended to improve classification accuracies by correctly classifying pixels that were misclassified in the WV-2 data classification and Landsat data classification (which also suffers from the problem of mixed pixels).

Table 2. Accuracy assessment for the classified Landsat data, classified WV-2 data and classified fused data

Data	Landsat bands		Fused bands				
Class ↓	PA*	UA*	PA*		UA*		
Substrate	77.54	77.47	71.23	↓	69.44	↓	
Vegetation	75.11	71.68	85.10	↑	87.22	↑	
Dark objects	65.71	68.73	67.99	↑	68.94	↑	
Average	72.79	72.63	74.77	↑	75.2	↑	
	Landsat WV-2 fused data						
	PA*		UA*				
Substrate	90.15	↑	95.56		↑		
Vegetation	89.39	↑	98.33		↑		
Dark objects	92.61	↑	96.36		↑		
Average	90.72	↑	96.75		↑		

* PA – Producer's Accuracy; UA – User's Accuracy.

Q index for Landsat and WV-2 classified data for SVD classes were 0.63, 0.60, 0.54 and 0.63, 0.75, 0.55 respectively. For the fused classified data, Q index for SVD classes were much higher (0.88, 0.92, 0.89). The LMM based fusion and classification represented the land surface as independent constituents with different landscape properties such as urban and vegetation. It characterized the fraction of illuminated vegetation, substrate or impervious materials and the shadowed or nonreflective surfaces such as water, roofing tar, etc. High substrate fractions are rational estimates of the impervious surface in developed land in temperate and tropical regions as pervious surfaces are mostly covered by some kind of vegetation and exposed substrate are most likely impervious. Small and Lu (2006) argue to use vegetation as a proxy for pervious surface because vegetation cannot thrive on impervious surface, so presence of vegetation implies presence of some amount of pervious surface. Therefore, using detectable vegetation as an indicator of permeable surface can account for the range of natural and built surfaces. It may be noted that classification of objects that exhibit high degrees of spectral heterogeneity representing variable endmembers with high intra-class spectral variation is beyond the scope of the current study. Therefore, a more detailed classification with higher number of classes while accounting for the different types of roofs and concrete features with the ground knowledge of vegetation cover at various phenological stages would give more insights to the fusion and further to the classification

process. Field data collection while considering the local topography would decrease minor errors in classification of substrate and dark objects. This difference is anticipated to be more with lower vegetation cover than at dense vegetation sites, appreciating the fact that vegetation from the image is modelled only for the portion that is illuminated by sunlight and the shaded portions of the canopy are likely to be assigned to the dark objects. One potential limitation of this study is that the number of LC classes considered is only three. Additionally, more classes have to be included in classification and the fusion approach has to be examined with different combination of multi-sensor data sets. Future directions of this study involve investigation on the applicability of LMM based fusion on the large repository of the time series Landsat-5 TM data with WV-2 data that would be useful for applications such as vegetation mapping, soil moisture estimation, biomass estimation, etc. LMM based fusion of MODIS 250 m (2 bands) product with MODIS 500 m (7 bands) product would be useful for continental to global scale LC mapping. Of course, fusion of LSR data must be performed within the same time frame as that of the HSR data acquisition. It is acknowledged that the fused bands are only an approximation of the Landsat bands at 2 m spatial resolution and LMM based fusion synthesized images closest to those the corresponding Landsat-5 TM sensor would observe at the WV-2 resolution level.

6. CONCLUSION

In this paper, Landsat-5 TM and WV-2 data fusion through linear mixture model was attempted. The outcome of this study based on the quantitative and inter-comparative assessment revealed that correlation between original and fused bands were higher for smaller window size and decreased with the increasing window size. Higher number of clusters in the high spatial resolution data showed higher correlation and lower RMSE and ERGAS values between fused and original Landsat data. Classification of the fused bands at 2 m spatial resolution showed higher producer's, user's and overall accuracies compared to the classification of Landsat and WV-2 data.

REFERENCES

Berger, C., Voltersen, M., Eckardt, R., Eberle, J., Heyer, T., Salepci, N., et al., 2013. Multi-modal and Multi-Temporal Data Fusion: Outcome of the 2012 GRSS Data Fusion Contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(3), pp. 1324-1340.

Breiman, L., 2001. Random Forests. *Machine Learning*, 40, pp. 5-32.

Breiman, L., and Cutler, A., 2010. Breiman and Cutler's random forests for classification and regression. Version 4.5–36, Repository – CRAN.

Castando, F., 2013. A Review of Data Fusion Techniques. *Hindawi Publishing Corporation, The Scientific World Journal*, 2013, pp. 1-19.

Dong, J., Zhuang, D., Huang, Y., and Fu, J., 2009. Advances in Multi-Sensor Data Fusion: Algorithms and Applications. *Sensors*, 9, pp. 7771-7784.

Gislason, P. O., Benediktsson, J. A., and Sveinsson, J. R., 2006. Random Forests for land cover classification. *Pattern Recognition Letters*, 27, pp. 294-300.

IEEE GRSS Data Fusion Contest, 2014, Fusion of Hyperspectral and LiDAR Data. http://hyperspectral.ee.uh.edu/?page_id=459

IEEE GRSS Data Fusion Contest, 2014, Multiresolution Fusion of Thermal Infrared Hyperspectral and VIS Data. http://cucciolo.dibe.unige.it/IPRS/IEEE_GRSS_IADFTC_2014 _Data_Fusion_Contest.htm#get_the_data_section

Khaleghi, B., Khamis, A., Karray, F. O., Razavi, S. N., 2013. Multi-sensor data fusion: A review of the state-of-the-art. *Information Fusion*, 14(1), pp. 28-44.

Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., et al., 2001. The causes of land-use and land-cover change: Moving beyond the myths. *Global Environmental Change*, 11, pp. 261-269.

Makarau, A., Palubinskas, G., and Reinartz, P., 2011. Multi-sensor data fusion for urban area classification. In: *Joint Urban Remote Sensing Event*, Munich, Germany, 11-13 April 2011, pp. 21-24.

Masek, J. G., Vermote, E. F., Saleous, N., Wolfe, R., Hall, F. G., Huemmrich, F., Gao, F., Kutler, J., and Lim, T. K., 2006. A Landsat surface reflectance data set for North America, 1990-2000. *IEEE Geoscience and Remote Sensing Letters*, 3(1), pp. 68-72.

Mezned, N., Abdeljaoued, S., and Boussema, M. R., 2007. Unmixing based Landsat ETM+ and ASTER image fusion for hybrid multispectral image analysis. In: *International Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007: Sensing and Understanding Our Planet*, Barcelona, Spain, 23-28 July 2007, pp. 3074-3077.

Mezned, N., Abdeljaoued, S., and Boussema, M. R., 2010. A comparative study for unmixing based Landsat ETM+ and ASTER image fusion. *International Journal of Applied Earth Observation and Geoinformation*, 12S(2010), pp. S131-S137.

Mingguo, Z., Qiangguo, C., and Mingzhou, Q., 2009. The Effect of Prior Probabilities in the Maximum Likelihood Classification on Individual Classes: A Theoretical Reasoning and Empirical Testing. *Photogrammetric Engineering and Remote Sensing*, 75(9), pp. 1109-1116.

Na, X., Zhang, S., Li, X., Yu, H., and Liu, C., 2010. Improved Land Cover Mapping using Random Forests Combined with Landsat Thematic Mapper Imagery and Ancillary Geographic Data. *Photogrammetric Engineering and Remote Sensing*, 76(7), pp. 833-840.

Nascimento, J. M. P., and Dias, J. M. B., 2005. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Transaction on Geoscience and Remote Sensing*, 43(4), pp. 898-910.

Pohl, C., and Van Genderen, J. L., 1998. Multisensor image fusion in remote sensing: Concepts, methods and applications. *International Journal of Remote Sensing*, 19(5), pp. 823-854.

Rogers, R. H., and Wood, L., 1990. The history and status of merging multiple sensor data: an overview. *Technical Papers, ACSM-ASPRS Annual Convention, Image Processing and Remote Sensing*, 4, pp. 352-360.

Rutzinger, M., Rottensteiner, F., and Pfeifer, N., 2009. A Comparison of Evaluation Techniques for Building Extraction From Airborne Laser Scanning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2(1), pp. 11-20.

Shimabukuro, Y. E., and Smith, A. J., 1991. The least-squares mixing models to generate fraction images derived from remote sensing multispectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 29(1), pp. 16-20.

Small, C., 2004. The Landsat ETM+ spectral mixing space. *Remote Sensing of Environment*, 93, pp. 1-17.

Small, C., and Lu, J. W. T., 2006. Estimation and vicarious validation of urban vegetation abundance by spectral mixture analysis. *Remote Sensing of Environment*, 100, pp. 441-456.

Small, C., and Milesi, C., 2013. Multi-scale Standardized Spectral Mixture Models. *Remote Sensing of Environment*, 136, pp. 442-454.

Wald, L., 2002. *Data Fusion Definitions and Architectures: Fusion of Images of Different Spatial Resolutions*. Paris, France: Ecole des Mines de Paris.

Walton, J. T., 2008. Subpixel Urban Land Cover Estimation: Comparing Cubist, Random Forests, and Support Vector Regression. *Photogrammetric Engineering and Remote Sensing*, 74(10), pp. 1213-1222.

Zhang, J., 2010. Multi-source remote sensing data fusion: status and trends. *International Journal of Image and Data Fusion*, 1(1), pp. 5-24.

Zhukov, B., Oertel, D., and Lanzl, F., 1995. A Multiresolution Multisensor Technique for Satellite Remote Sensing. In: *International Geoscience and Remote Sensing Symposium, 1995. IGARSS '95: Quantitative Remote Sensing for Science and Applications*, Firenze, Italy, 1, pp. 51-53.

Zhukov, B., Oertel, D., Lanzl, F., and Reinhackel., G., 1999a. Unmixing-Based Multisensor Multiresolution Image Fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3), pp. 1212-1226.

Zhukov, B., Oertel, D., Lanzl, F., and Reinhackel, G., 1999b. Unmixing-Based Multisensor Multiresolution Image Fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3), pp. 1212-1226.

Zurita-Milla, R., Clevers, J. G. P. W., and Schaepman, M. E., 2008. Unmixing-Based Landsat TM and MERIS Data Fusion. *IEEE Geoscience and Remote Sensing Letters*, 5(3), pp. 453-457.

Zurita-Milla, R., Guillen-Climent, M. L., Clevers, J. G. P. W., and Schaepman, M. E., 2006. Unmixing-based Landsat and MERIS image fusion for land cover mapping over the Netherlands. In: *ISPRS Commission VII WG VII/6 Mid-term Symposium on Remote Sensing: From Pixels to Processes*, Enschede, The Netherlands, May 8-11, 2006.