An operational framework for object-based land use classification of heterogeneous rural landscapes

Article in International Journal of Applied Earth Observation and Geoinformation · February 2017
DOI: 10.1016/j.jag.2016.09.012

3 authors, including:

Gary R Watmough
Aarhus University
13 PUBLICATIONS 77 CITATIONS

Clare C Sullivan
University of Wisconsin–Madison
17 PUBLICATIONS 337 CITATIONS

Some of the authors of this publication are also working on these related projects:

Understanding Social-Ecological Systems View project

Land cover and poverty in Assam: A remote sensing perspective View project

All content following this page was uploaded by Gary R Watmough on 13 October 2016.

The user has requested enhancement of the downloaded file.
Manuscript Number: JAG-D-16-00216R3

Title: An operational framework for object-based land use classification of heterogeneous rural landscapes

Article Type: Research Paper

Keywords: object-oriented classification; social-ecological systems; high spatial resolution data; heterogeneous landscape classification; Kenya land use; multi-scale segmentation; GEOBIA; landscape heterogeneity,

Corresponding Author: Dr. Gary R watmough, Ph.D.

Corresponding Author's Institution: Ecoinformatics and Biodiversity

First Author: Gary R watmough, Ph.D.

Order of Authors: Gary R watmough, Ph.D.; Clare Sullivan; Cheryl A Palm

Abstract: The characteristics of very high resolution (VHR) satellite data are encouraging development agencies to investigate its use in monitoring and evaluation programmes. VHR data pose challenges for land use classification of heterogeneous rural landscapes as it is not possible to develop generalised and transferable land use classification definitions and algorithms. We present an operational framework for classifying VHR satellite data in heterogeneous rural landscapes using an object-based and random forest classifier. The framework overcomes the challenges of classifying VHR data in anthropogenic landscapes. It does this by using an image stack of RGB-NIR, Normalised Difference Vegetation Index (NDVI) and textural bands in a two-phase object-based classification. The framework can be applied to data acquired by different sensors, with different view and illumination geometries, at different times of the year. Even with these complex input data the framework can produce classification results that are comparable across time. Here we describe the framework and present an example of its application using data from QuickBird (2 images) and GeoEye (1 image) sensors.

Opposed Reviewers:
An operational framework for object-based land use classification of heterogeneous rural landscapes

Abstract

The characteristics of very high resolution (VHR) satellite data are encouraging development agencies to investigate its use in monitoring and evaluation programmes. VHR data pose challenges for land use classification of heterogeneous rural landscapes as it is not possible to develop generalised and transferable land use classification definitions and algorithms. We present an operational framework for classifying VHR satellite data in heterogeneous rural landscapes using an object-based and random forest classifier. The framework overcomes the challenges of classifying VHR data in anthropogenic landscapes. It does this by using an image stack of RGB-NIR, Normalised Difference Vegetation Index (NDVI) and textural bands in a two-phase object-based classification. The framework can be applied to data acquired by different sensors, with different view and illumination geometries, at different times of the year. Even with these complex input data the framework can produce classification results that are comparable across time. Here we describe the framework and present an example of its application using data from QuickBird (2 images) and GeoEye (1 image) sensors.

1. Introduction

There is an increasing interest from development agencies in the use of remotely sensed satellite data as part of monitoring and evaluation programmes. The International Fund for Agricultural Development (IFAD) is currently exploring how remote sensing satellite data can contribute to evaluating the impact of projects on the landscape. The Millennium Villages Project (MVP) collected remotely sensed satellite data to support the on-going monitoring and evaluation during the 10 year project (www.millenniumvillages.org). The underlying concept is that, remote sensing data can provide information on landscape characteristics that are important for human livelihoods (Watmough et al. 2016). These characteristics are often not adequately covered in traditional household surveys. This novel application of remotely sensed satellite data often requires very high spatial resolution (VHR - defined here as a spatial resolution of less than 2m) data due to small land parcels and complex management structures (Fassnacht et al 2006; Verburg et al. 2009; Ozdogan and Woodcock 2006). However, VHR data pose challenges for land use classification of heterogeneous rural landscapes. It has, so far, not possible to develop generalised and transferable land use classification definitions and algorithms. There are few past studies documenting how VHR data can be used for classifying land use in complex anthropogenic rural landscapes. We present an
operational framework for classifying VHR satellite data in heterogeneous rural landscapes using an object-based and random forest classifier. The framework overcomes the challenges of VHR data in anthropogenic landscapes by using image stack of RGB-NIR, NDVI and textural bands in a two-phase object-based classification. It is flexible enough to be applied to data acquired by different sensors, with different view and illumination geometries, at different times of the year, producing classification results that are comparable across time.

Agriculture is an important economic activity in rural areas of developing countries which means they are characterised by complex land use mosaics (Ellis 2011). This is due to division of family land through inheritances creating ever smaller land parcels and land management practices such as land sharing (Adams 2012) and agro-forestry (Clough et al. 2011). Agro-forestry and land sharing have been suggested as ways of balancing the need for biological conservation and increased agricultural yields (Chappell and LaValle 2009; Melo et al. 2013; Tscharntke et al. 2012) and are often encouraged as part of rural development projects. Understanding and monitoring the impact of such management practices requires landscape level information (Balmford et al. 2012; Phalan et al. 2011). Remotely sensed data derived from MODIS and Landsat can be used to characterise global and regional land use (Friedl et al. 2010) and classification parameters can often be generalised across large spatial extents. However, it is not possible to generate fine resolution characterisations of agroforestry or monitor fine-scale changes within these systems using such data (Monfreda et al. 2008; Nol et al. 2008; Ellis et al. 2006).

1.1 Challenges of Using VHR data

VHR sensors allow for small land parcels and ground features to be detected and classified which is ideal for monitoring complex landscapes. However, VHR data are also characterised by a series of challenges. The H-resolution problem occurs when a single ground feature is characterised by multiple image pixels. This increases the likelihood that pixels comprising a single ground feature may have different spectral reflectance values (Strahler et al. 1986). This problem, combined with the often low number of spectral bands available on VHR sensors, reduces the spectral separation of different ground features in VHR satellite sensor data. This is because it is more likely that pixels comprising one ground feature will share spectral characteristics with pixels comprising different ground features.

VHR sensors have small swath-widths and, to maximise repeat visit cycles the sensors have off-nadir viewing capabilities. This creates challenges associated with the ‘sun-surface-sensor geometry’ (Wulder et al. 2008) and often results in problems such as shadow (Dare 2005). It can also cause
problems associated with sun glint, where the orientation of ground features reflects light directly towards the sensor (Kay et al. 2009). Furthermore, the limited number of spectral bands and the H-resolution problem (Strahler et al. 1986) can further exacerbate the sun-sensor-geometry problems. The above issues pose challenges for the use of VHR data in operational settings, especially when images are required for a particular scene during a particular time-window (for example to coincide with important project deliverables). It is often the case that anniversary images are not available from the same sensor leaving two options open; (i) using an image from non-anniversary dates, or; (ii) using an image from an entirely different sensor. Neither of these options enable the development of general class rules that can be universally applied. This is because the spectral characteristics of ground features will differ across scenes, time periods and sensor data (Kohli et al. 2013; Walker and Blaschke 2008).

Geographic Object Based Image Analysis (GEOBIA) has been developed as an approach for dealing with some of these issues by segmenting groups of pixels into image objects. These objects are subsequently used as the minimum mapping unit (MMU) rather than the pixel (Addink et al. 2012; Blaschke et al. 2014). GEOBIA has been found to improve classification accuracy in a range of different sensors and applications as compared to pixel-based approaches (Whiteside et al. 2011; Myint et al. 2011; Hamada et al. 2013; Yan et al. 2006; Tehrany et al. 2013). However, without a targeted approach to multiresolution segmentation in complex rural landscapes the objects resulting from image segmentation are often unrepresentative of all target ground features.

2. A framework for classifying VHR data in heterogeneous landscapes

The framework is designed to overcome several challenges of using VHR satellite data to classify heterogeneous anthropogenic rural landscapes. The main focus was to minimise the number of changes required when classifying multiple images covering the same landscape. First, the framework breaks the classification problem down to focus on target ground features and maximise the differences between these in the image stack. Second, to reduce the likelihood of over- or under-segmentation, the framework splits the classification problem into two distinct phases. This is one of the key aspects of the framework as the classification process and subsequent accuracy in GEOBIA is determined by the segmentation results (Gao et al. 2013; Kim et al. 2009; Myint et al. 2011). Image segmentation parameters that are focussed on particular ground features can result in objects that are unrepresentative of other ground features within the image. Therefore, phase-one of the classification focuses on features with distinct spectral and spatial characteristics such as water, roads and buildings. Phase-two of the classification focusses on features that have subtle differences such as different types of vegetated ground features.
The framework is split into ten steps (Figure 1), including pre-processing (steps 1-4) and two-phases of classification. The data are standardised to reflectance values (step-1) before the user identifies target ground features and how they differ from one-another spectrally (step-2). If one or more target ground features are smaller than four pixels in size, the data are pan-sharpened (step-3). If spectral separation of classes is low, a series of additional bands are created (such as NDVI and texture) to maximise the differences between target ground features. Classification phase-one (step-5) uses GEOBIA multi-resolution segmentation and fuzzy ruleset methods to classify ground features. Ground features classified in Phase-1 are masked from the original image (Step-6). Phase-two (step-7) classification focusses on ground features that have overlapping spectral and spatial characteristics. As the differences between some of these classes can be subtle, a machine learning algorithm is used rather than user-defined membership functions. In step-8 samples are defined for each of the target features. Each object is associated with a series of spatial and spectral characteristics which are used in step-9 to train a random forest classifier. In step-10 the error of the classification can be checked to ensure that the class accuracy is acceptable. If not step 7 through 10 can be repeated until the accuracy is satisfactory. R code is provided on GitHub for steps; one, three, four, six, eight, nine and ten (https://github.com/gary-watmough/VHR_LandUse_Classification).

2.1 A Two-phase approach

In GEOBIA, the coarsest level of multiresolution segmentation can be applied to every pixel within an image. Subsequent segmentations are nested within the parent object. This means that, child objects are constrained by the spatial extent of parent objects, and if a ground feature overlaps two or more image objects it cannot be segmented into a single object in later segmentations. Segmentation parameters designed for a specific ground feature often results in poor characterisation of other ground features. For example, varying the compactness parameter can return image objects that are more representative of target ground features. If a target class is square, such as a building, then the compactness ratio can be weighted towards one to prioritise square features, if it is elongated then it can be weighted towards zero. This can have a significant impact on the resulting objects and how they represent the target ground features. It can result in over- or under-segmented objects for other target features. We found many examples in the literature of multiresolution segmentation approaches that varied the scale parameter but kept the compactness and shape parameters constant (Mui et al. 2015, Myint et al. 2011). In our experience this resulted in objects that were poor characterisations of all target ground features.

Running segmentation with varying shape and compactness parameters causes a new set of problems for which our framework overcomes. The problem is that, parameterising a segmentation
to generate objects that are an accurate representation of a particular ground feature often results in a reduced ability to classify other ground features accurately. For example, if a first level segmentation targets large elongated objects such as rivers and roads the compactness parameter can be weighted at 0.1 and scale 20. Elongated road objects will result, but vegetation objects will also tend to be elongated and will include multiple types of vegetation in a single object. Segmentation can be altered to address this by using additional bands to differentiate between vegetation types. However, the original road objects will likely be compromised. Instead, the vegetation objects can be segmented in a subsequent step. But, because the multi-resolution segmentation algorithm is nested, child objects are bounded by parent objects. So woodland can be segmented from agriculture but if the parent objects were over-segmented this will remain throughout the analysis. To overcome this problem the framework is split into two distinct phases: Phase-one targets objects that have distinctive and consistent shapes such as water, roads and buildings. Rivers and Roads are elongated ground features and the compactness ratio in segmentation can be lowered to allow for such objects to be created. This approach results in objects representative of the rivers, streams and roads but not of ground features with variable and inconsistent shape characteristics such as agricultural fields and patches of woody cover. Buildings are often so small in these locations and spectrally distinct from surrounding features that they can be segmented in phase-one even though they are not elongated. Phase-two allows the segmentation to begin again with pixels and have a focus on different ground features. Without this hierarchical, two-phase approach the segmentation parameters would be compromised and return a series of objects that were poor characterisations of any ground features.

3 Case Study: Sauri Millennium Village, Kenya

The framework was tested using QuickBird and GeoEye data for a site in Kenya which was part of the Sauri Millennium Village (Mutuo et al. 2007). There was an operational requirement to produce land cover maps of the site to support the data collected in household surveys at baseline (2004) and year five (2009) as well as in 2011 to support an additional field campaign. No automated LU change detection was required in this project. Instead, stand-alone classification products were required to support monitoring and evaluation activities of the wider project. The image characteristics are presented in Table 1. The images represent well the operational challenges faced when using this type of data. Images are often required for specific years to represent important milestones in a project. No QuickBird data was available in 2011 and instead a GeoEye image from December was used. This meant that the characteristics of the 2004 and 2009 were similar but those of 2011 were significantly different.
3.1 Pre-processing steps

The framework, as applied to Kenya data, is presented in Figure 2 as a schematic diagram. The thresholds and parameters used in phase-one and phase-two are shown in Figures 3-5 and Table 3. In step-1 the standard data products (both four-band multispectral and panchromatic), which were georeferenced and normalized for topographic relief (known as LV2A) by the image vendor, were converted from Digital Number (DN) values to radiance using the approach in Krause (2005). The radiance values were converted to top-of-atmosphere (TOA) reflectance using the approach in Updike and Comp (2010) - this was performed in R and the code is available from GitHub (https://github.com/gary-watmough/VHR_LandUse_Classification). In step-2 we identified the following target ground features; homesteads, woody cover (forest and small agroforestry plots) agriculture, grassland, bare land (non-vegetated), scrubland, roads and water bodies. Homesteads were comprised of several small buildings, the majority of which were less than two pixels in size in the 2 m resolution multispectral data.

3.1.1 Pansharpening

Panchromatic data supplied by the image vendor were used to pan-sharpen the multispectral data. This produced an image stack of four bands (RGB and NIR) with the spectral characteristics of the multispectral data and the spatial resolution of the panchromatic band (Vrabel 1996). Experimentation with different pan-sharpening algorithms in ENVI 4.7 (Excelis Visual Information Systems, Boulder, Colorado) found the Gram-Schmidt Sharpening algorithm (Laben and Brower 2000) returned finer resolution pixels with consistent spectral characteristics. The Gram-Schmidt algorithm returns pan-sharpened data with 4 bands rather than 3 bands that are produced when using an Intensity-Hue-Saturation approach (Aiazzi et al. 2006).

3.1.2 Additional Features

Much of the woody vegetation had similar spectral characteristics to long-term agricultural fallows. It was possible to visually identify the differences between woody and long term fallow due to textural differences. However, finding a spectral difference between the two using the four multispectral bands was not possible. Therefore, step-4 was required to generate three additional features (NDVI, Texture1 and Texture2) that could be used to aid classification. The NDVI differentiated between different non-woody features such as agriculture, bare and grassland fields that had a small vegetation signal. Texture1 (T1) was the GLCM variance of the red band with a moving window size of 11x11 pixels and Texture2 (T2) was the GLCM mean of the red band using a 7x7 pixel moving window. The texture bands were derived by experimenting with combinations of
bands, window size and texture types. T1 helped to separate woody vegetation cover from scrubland fallow and agriculture and T2 helped separate small buildings from surrounding bare land. The additional features were generated in R 3.2.0 and the code is available from GitHub (https://github.com/gary-watmough/VHR_LandUse_Classification).

3.2 Classification Phase-1

In step-5, five levels of segmentation and fuzzy rulesets (Figures 4-6) were used to classify ground features with distinct spectral and spatial characteristics (Building, Road, Water). Multiresolution segmentation was used (implemented in the eCognition Developer 9 software environment) to generate hierarchical levels of objects each focussed on a specific ground feature (Level-1 on water, Level-2 roads, Level’s 3-5 buildings). Weightings applied to each band in the segmentation were determined by experimentation. Level-1 segmentation was applied to all pixels (Segmentation Object Domain in Figures 4-6) and fuzzy classification targeted large water bodies, but was also able to classify some large buildings. Subsequent levels of analysis were applied to specific objects from the previous level classification outputs. For example, level-2 was applied to objects that were unclassified in level-1, whilst, level-three was applied to vegetated objects from level-two (Figure 3-5). Temporary holding classes (Bare, Vegetated, Woodland) were used during Phase-one classification. These objects were not representative of the target ground features. For example, the vegetated objects in level-one and two were relatively large objects that contained non-vegetated ground features such as small buildings and small bare fields that were partially covered by tree canopy or surrounded by vegetation. These objects required further segmentation to prevent omitting small target ground features. The use of holding classes was important as it decreased the complexity and helped focus each level of segmentation on a particular target ground feature.

At each level, prior to commencing classification some objects classified in the preceding level were copied across. The only time this did not apply was if a particular class from a parent level had been segmented to create smaller objects. For example, in level-three, vegetated objects from level-two were segmented. Thus, Water, Building and Road classes were carried forward but vegetated objects from level-two were not as they were segmented into smaller component objects in level-three.

3.3 Classification Phase-2

A second classification phase targeting different vegetation types was performed using multiresolution segmentation. This phase generated objects representative of ground features such as woodland, agricultural fields, grassland, bare-land and scrubland (step-7). Phase-two (Table 3) was comprised of three levels of multi-resolution segmentation. The first segmentation was applied to all
pixels in the image and resulting objects classified as vegetation or non-vegetation using a threshold of mean NDVI > 0 or mean NDVI ≤ 0 respectively. The second segmentation was applied to objects classified as vegetation in level-one. This second level created objects that were representative of the different vegetated ground features. Level two objects were classified as vegetation if the threshold of NDVI > 0 was met. All non-vegetated objects in level-2 including non-vegetated objects from level-1 were passed to the level-3 segmentation. This final segmentation aimed at separating small vegetated ground features such as individual trees from non-vegetated objects. Samples were selected from within the image (step-8) using expert knowledge of the study site for the following classes; agriculture, grassland, non-vegetated (bare-land), scrub-land, woody.

3.3.1 Random Forest Classification

Phase-two classification was conducted using a random forest classifier (Step-9) in R 3.2.0 (R Development Core Team) using the ‘randomforest’ package (Liaw and Wiener 2014). The random forest (RF) classifier is an ensemble learning algorithm first developed by Breiman (2001). It is increasingly used for performing LULC classifications (Ghimire et al. 2012; Lawrence et al. 2006) as it can run efficiently on large data sets. The RF model was applied to fourteen features generated for each image object (Table 2). The samples were exported from eCognition into R 3.2.0 and split into training and test sample sets using the ‘sample’ function prior to building the model. The training set was 70% of the total samples and the test or ground truth set was the remaining 30% of samples. All samples were selected from within the image by team members that had an extensive knowledge of the site. An iterative classification process was followed, whereby classes with large commission errors were resampled and the RF repeated. The final trained model was applied to all objects using the ‘predict’ function in R 3.2.0. The code for the random forest model is provided on GitHub (https://github.com/gary-watmough/VHR_LandUse_Classification).

The out-of-bag error statistics were used to identify classes that were confused (Step-10). Any class with an error greater than 20% was re-classified by returning to step-7. If the image objects were found to be poor characterisations of particular ground features steps-7 through 10 were repeated. If the objects were good characterisations of the ground features, the sampling in step-8 was repeated. This iterative classification continued until all classes had errors that were below 20%.

4. Results

Overall accuracies were 90.5% for the 2004 image; 89.6% for 2009 and 89.6% for 2011 with Kappa coefficients of 0.8788, 0.8679 and 0.8655 for 2004, 2009 and 2011 respectively (Table 4 – 6). The producers and users accuracies for each class were consistent with the results to the internal Out-of-
Bag (OOB) errors (not shown) supporting the findings of Rodriguez-Galiano et al. (2012) that OOB is a robust measure of RF classification error. Non-vegetated, Grassland and Woody cover had the highest accuracies in all three images whilst the Agriculture and Scrub classes had lower accuracies. The error matrices indicated that, in each of the images, Scrubland was often misclassified as Woody cover or Agriculture which was expected as scrubland was fallow agricultural land. Early fallow shared spectral and textural characteristics with agricultural plots and older fallow shared characteristics with woodland due to the rapid regrowth of semi-natural vegetation in fallow areas of Sauri.

4.1 Classification Phase-1

The scale, shape and compactness parameters of the segmentation were the same for each image because the aim was to produce objects that had consistent sizes and shapes in each classification. This was achieved by maintaining the same scale, shape and compactness values and varying the band combinations and weightings applied to these bands. The similarities in image acquisition characteristics meant that channels and weightings used in all levels of phase-one segmentation remained the same in 2004 and 2009. Phase-one, level-1 segmentation in 2011 was the same as that in 2004 and 2009 because the target ground feature was water which was spectrally distinct from other ground features. The material of building roofs did not change in most cases. However, the differences in view and illumination angles in 2011 compared to 2004/2009 images resulted in buildings having different spectral characteristics. Therefore, segmentation parameters were varied to generate objects with shapes and sizes that were consistent with those in 2004 and 2009. In all three images, the T\textsubscript{2} band is used most often in the membership functions for building as it highlighted pixels with saturation caused by pitched metal roofs.

Road features were classified using the mean red reflectance and length/width ratio in each image (Figure 3-5). The threshold used for the red channel was the same in 2004 and 2009 but brighter in 2011. This was because unsealed dirt roads had high levels of reflectance in the red channel and therefore would need to be changed if asphalt roads were also present in the landscape. Some bare ground and bare agricultural fields had the same reflectance values in the red channel as unsealed roads. The inclusion of the length/width ratio (Figure 3-5) ensured that only elongated objects with high red reflectance values were selected as roads. This was possible because of the use of a compactness ratio of 0.3 in the segmentation which meant that roads appeared as large elongated objects compared to the relatively smaller more compact bare fields. Using the default compactness parameter of 0.5, which others have done in the past (Myint et al. 2011; Miu et al. 2015), resulted in roads being split into multiple smaller objects which could not be differentiated from bare fields.
4.1.1 Temporary Holding Classes

Temporary holding classes were used throughout phase-one. They included vegetated and bare classes and had two purposes. The first was to group objects together that required further segmentation at a later point. For example, the vegetated class used in Level-1 to Level-4 grouped objects that were mainly vegetated but contained small buildings partially covered by vegetation. All of the objects classed as vegetated throughout the process were passed to Level-5 segmentation which focused on segmenting very small buildings from vegetated areas. The second reason for using temporary classes was to reduce errors of commission in Phase-one. For example, the bare class was used to remove objects that were clearly bare ground from being considered as buildings or roads. This was an important class to include as some bare areas had spectral characteristics that were similar to roads and buildings. In the 2011 classification there were fewer temporary classes required because; (i) there was less vegetation in the image due to the December acquisition, and; (ii) the spectral signatures of classes such as roads, buildings and bare land were more distinct, likely due to the acquisition being in the December dry period. Depending on the number of bands and the spectral similarities between them, it is recommended that users consider using temporary classes in Phase-one.

4.2 Classification Phase-2

The aim of phase-two was to generate objects that were representative of agriculture, grass, bare land, scrubland and woody cover. Three levels of segmentation were used for each image (Table 3) and the same parameters used for each. We experimented with varying the segmentation parameters for each image to accommodate varying ‘sun-sensor-geometry’ characteristics but found little difference in the output objects in our study site. The variable importance (Figure 7) from each classification indicated that the three additional bands (NDVI, T_1, T_2) had a significant impact on the phase-two classification accuracy. NDVI was the most important variable in all three classifications.

5: Discussion

We have created an operational framework because we are aware that further development will be required. But a framework will allow the development of common language and idea development. This will help elucidate the requirements of RS classification to those considering using it for monitoring and evaluation as well as those more used to using coarser resolution data. The framework resulted in land use classifications for the study site at three snapshots in time (Figure 6). Although different classes were targeted by Kohli et al. (2013), they demonstrate the challenges in developing transferable and generalizable classification rulesets using VHR data and GEOBIA. When
classifying slum settlements in three subsets of the same image, Kohli et al. (2013) achieved accuracies of 47-68%, which are relatively low considering that the ‘sun-surface-geometry’ was the same. However, the results in Kohli et al. (2013) highlight that the more focussed the classification is on a particular image or subset the lower the likelihood is of generalising the class descriptions for application to other regions. In Kenya, we classified nine land use classes in three different images with varying ‘sun-surface-geometry’ characteristics and achieved approximately 90% accuracy in each image. These results indicate that the framework is flexible enough to accurately segment and classify multiple ground features across heterogeneous landscapes that have a range of spectral and spatial parameters.

This framework produces stand-alone LU classification products. Currently it cannot be used for automated change detection for three reasons. The first is that pixel-to-pixel geolocation accuracy between different images is currently not possible due to the varying view angles of different image acquisitions. The second is that the ‘sun-sensor-geometry’ problem can result in different shadow and saturation characteristics in each image. This can mean that ground features have different appearances between images which will cause confusion in an automated change analysis algorithm. The third is that often it is not possible to acquire anniversary images and, as such, a change from one image to the next is not indicative of a ‘true’ LU change. It could simply be that the acquisitions have captured different periods in the agricultural cycle. Furthermore, some ground features can appear and disappear between images due to changes in over-hanging vegetation canopies. For example, this can give the false impression that a building has appeared or disappeared between images. The parameters of our framework can be varied to ensure that the same set of land use classes are classified in each image over time. But, comparing between two or more classification products requires human/visual inspection and knowledge of the local conditions. Currently, automated change analysis cannot be run on these data products. Therefore, future work should focus on how to generate a change detection algorithm that can identify ‘true’ land use change.

The GEOBIA analysis was applied to an image stack that contained the multispectral data, NDVI and two texture bands. Similar approaches were used by Esch et al. (2014) and Mui et al. (2015). The additional features help to segment the image into objects representative of the target ground features. The three used here (NDVI $T_1$ and $T_2$) contributed significantly to the segmentation of image objects and the random forest classification (Figure 7).

Within each step of the framework the user should identify the most appropriate course of action for the specific imagery and classification problem. Pan-sharpening (step-3) does not have to be performed if the ground features are comprised of multiple pixels in the original resolution, step-4
could use features such as principal components analysis, different biophysical parameters, composite indices and variations in textural analysis (texture type, band used moving window size).

Step-5, which was split into five levels of analysis in the Kenyan images, can be varied to accommodate the complexity of the landscape, the characteristics of target ground features and the ‘sun-surface-sensor geometry’. For example, if ground features are relatively large and pan-sharpening is not applied the number of required segmentation levels may be lower and the scale parameters may be larger. Any ground features classified in step-5 should be masked from the image in step-6 prior to phase-two classification in step-7 to ensure that the image objects are representative of the ground features.

Varying the segmentation parameters generated objects in phase-one classification that were representative of distinctive target ground features. However, other ground features such as woodland and agricultural fields were poorly characterised by the resulting image objects. Therefore, the framework used a second phase of segmentation and classification. Phase-two classification (step-7) could again vary the number of segmentation levels depending on the target features and image characteristics. Step-8 can use ground samples if these are available rather than selecting them from within the image and step-9 could use a different classification algorithm other than random forest.

The spectral characteristics of buildings and bare ground were more distinguishable in the GeoEye image of Kenya acquired in December 2011. This meant that six membership functions were required to classify buildings. However, this number rose significantly in the QuickBird data for 2004 and 2009. Buildings and bare ground were more difficult to separate spectrally and additional intermediate classes were required to separate the objects that were most likely to be buildings from those which were not. This is an important consideration as often the user has a limited amount of control over the image characteristics when purchasing VHR data. For example, the combination of view and illumination angles in the 2004 QuickBird image for Kenya resulted in sun glint, saturation and pixel bleeding in buildings with pitched tin roofs.

7. Conclusion

The characteristics of VHR satellite data are encouraging development agencies to consider using it for monitoring and evaluation activities. In operational settings, such as these, it is often the case that users of VHR have no choice but to use images that; (i) have large variations in view and illumination angles; (ii) were acquired in different seasons, and; (iii) were acquired by different sensors. These issues mean that it is not possible to develop transferable classification procedures.
We developed a framework to classify land use in VHR satellite data in an agricultural landscape and applied it to three images acquired for a site in Kenya. The framework minimised the number of changes required to the classification process and reduced the time taken to classify images by taking a prescriptive approach to image analysis. A novel two-phase approach to image segmentation and classification was taken and spectral differences between classes was maximised by adding NDVI and two textural bands to the image stack. The work presented here can help to focus approaches to VHR land use classification in other sites and begin a dialogue that can help to develop a set of guiding principles for the classification of VHR data in rural settings. Future work will further develop the framework by applying it to images acquired by different VHR sensors and in different study sites.

Acknowledgements

This research was funded by the Earth Institute Fellows Program at Columbia University and the Millennium Villages Project (MVP). Additional Funding was received from the Ecosystem Services and Management at the International Institute of Applied Systems Analysis (IIASA) and the International Fund for Agricultural Development (IFAD). We thank Charlotte Marcinko, Ian McCallum and Steffen Fritz for comments on a previous draft of the manuscript. We also thank three anonymous reviewers for helpful comments on a previous version of the manuscript.

References


from Assam, India. *World Development*, 78; 188-203.


Figure and Table Captions

Figure 1: Steps to classify a very high spatial resolution image in a human modified rural landscape. Solid lines indicate required steps and dashed lines indicate optional steps.

Figure 2: Schematic of the classification framework applied to the three images in Kenya.

Figure 3 phase 1 classification process for 2004 QuickBird image see Table 2 for parameter definitions.

Figure 4 phase 1 classification process for 2009 QuickBird image see Table 2 for parameters definitions.

Figure 5 phase 1 classification process for 2011 GeoEye image see Table 2 for parameter definition.

Figure 6: Example of the multispectral segmentation and classification result for the 2004 and 2009 QuickBird and 2011 GeoEye images. Shows how the object sizes varied from large elongated objects (roads and river) to small and compact objects (buildings and ponds).

Figure 7 Important variables from Phase 2 random forest classification for all images.

Table 1 Characteristics of QuickBird and GeoEye images used in Kenya.

Table 2 Definitions and descriptions of parameters used in phase 1 and phase 2 classification.

Table 3 Segmentation parameters used in phase 2 of the analysis. These parameters were the same for each of the three images (2004, 2009, and 2011).

Table 4 Confusion matrix for 2004 QuickBird Phase 2 classification.

Table 5 Confusion matrix for 2009 QuickBird Phase 2 classification.

Table 6 Confusion matrix for 2011 GeoEye Phase 2 classification.
Figure 1 Framework steps

Preprocessing

Step 1: Standardise Data to reflectance

Step 2: Identify Target ground features
  - How do features differ spectrally?
  - How big are target features?

Step 3: Pan-sharpen
  - If ground features similar spectrally generate additional features

Step 4: Additional Features
  - Composite Indices, principal components, textural bands

Phase 1 Classification

Step 5: Phase 1 Classification
  - Classify feature’s with distinct spectral and spatial characteristics
  - Multiresolution segmentation
  - User-defined membership

Step 6: Mask Phase 1 Classes

Phase 2 Classification

Step 7: Phase 2 Classification
  - Classify feature’s with similar spectral and spatial characteristics
  - Multiresolution segmentation

Step 8: Collect samples for target features

Step 9: Random Forest classifier
  - Train model with spectral and spatial characteristics of image objects

Step 10: Examine Error
Figure 2 Classification schematic

Click here to download high resolution image

<table>
<thead>
<tr>
<th>Input Data</th>
<th>Analysis Step and Details</th>
<th>Output from Analysis step</th>
</tr>
</thead>
<tbody>
<tr>
<td>DN Data, MS and Pan</td>
<td><strong>Step 1:</strong> Radiometric &amp; TOA Atmospheric Correction</td>
<td>Image Stack in Reflectance Units</td>
</tr>
<tr>
<td></td>
<td><strong>Step 2:</strong> Homesteads; Agriculture; Woody; Bare land; roads; &amp; waterbodies</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Step 3:</strong> Pan-sharpen MS with Gram Schmidt</td>
<td>Pan-sharpened RGB &amp; NIR bands</td>
</tr>
<tr>
<td></td>
<td><strong>Step 4:</strong> Spectrally similar classes. Create NDVI, T1 &amp; T2</td>
<td>Pan-sharpened Image Stack: RGB, NIR, NDVI, T1, T2</td>
</tr>
<tr>
<td></td>
<td><strong>Step 5:</strong> Phase 1 Classification - GEOBIA with 5 object levels. Target: Building, Road, Water.</td>
<td>Classified Objects: Building, Road, Water</td>
</tr>
<tr>
<td></td>
<td><strong>Step 6:</strong> Mask Building, Road, Water from image stack</td>
<td>Pan-sharpened Image Stack with Building, road and water masked.</td>
</tr>
<tr>
<td></td>
<td><strong>Step 7:</strong> Phase 2 Segmentation - Agriculture, Bare land, Woody</td>
<td>Objects for sampling</td>
</tr>
<tr>
<td></td>
<td><strong>Step 8:</strong> Phase 2 Sampling - Collect object samples for classes</td>
<td>Training Samples for Random Forest: Agriculture, Bare land, Scrubland, Woody</td>
</tr>
<tr>
<td></td>
<td><strong>Step 9:</strong> Phase 2 Classification - Random Forest Classifier</td>
<td>Classified Objects</td>
</tr>
<tr>
<td></td>
<td><strong>Step 10:</strong> Examine OOB Error</td>
<td></td>
</tr>
</tbody>
</table>
Figure 6 classification output example for 2004, 2009, and 2011.
Figure 4 Phase 1 2009 outline method
Click here to download high resolution image
Figure 5 Phase 1 2011 outline method
Click here to download high resolution image
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisition Data (dd/mm/yyyy)</td>
<td>12/09/2004</td>
<td>05/09/2009</td>
<td>29/12/2011</td>
</tr>
<tr>
<td>Mean Sun Azimuth (°)</td>
<td>80.0</td>
<td>72.4</td>
<td>136.5</td>
</tr>
<tr>
<td>Mean Sun Elevation (°)</td>
<td>66.8</td>
<td>67.6</td>
<td>56.86</td>
</tr>
<tr>
<td>Mean off-nadir view angle (°)</td>
<td>12.2</td>
<td>5.4</td>
<td>11.1</td>
</tr>
<tr>
<td>Resolution MS (m)</td>
<td>2.4</td>
<td>2.4</td>
<td>2.1</td>
</tr>
<tr>
<td>Resolution Pan (m)</td>
<td>0.6</td>
<td>0.6</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Table 1: Image information
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Definition</th>
<th>Phase 1</th>
<th>Phase 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>µB/G/R/NIR</td>
<td>Mean Reflectance</td>
<td>Mean reflectance of all pixels within an image object in the blue/green/red/NIR band</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>µNDVI</td>
<td>Mean NDVI</td>
<td>Mean normalised difference vegetation index of all pixels within an image object.</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>µT₁</td>
<td>Mean Texture 1</td>
<td>Mean intensity of Grey-Level Co-Occurrence Matrix of red band with 7x7 window within an image object</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>µT₂</td>
<td>Mean Texture 2</td>
<td>Mean intensity of Grey-Level Co-Occurrence Matrix of red band with 11x11 window within an image object</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>µBright</td>
<td>Mean Brightness</td>
<td>Sum of mean values in Blue and Green bands divided by the total number of bands (2 in this case).</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>σB/NIR/R</td>
<td>Standard Deviation of reflectance</td>
<td>Standard Deviation of reflectance of all pixels within an image object in the Blue/NIR/Red band.</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>L/W</td>
<td>Length/Width</td>
<td>Object length divided by the object width</td>
<td>•</td>
<td>•</td>
</tr>
<tr>
<td>GLDV</td>
<td>GLDV Red</td>
<td>Object-level Grey-Level Difference Vector of the red band.</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>GLCM&lt;sub&gt;R/N/G&lt;/sub&gt;</td>
<td>GLCM texture</td>
<td>Object-level Grey-Level Co-Occurrence matrix of the red/NIR/Green band.</td>
<td>•</td>
<td></td>
</tr>
<tr>
<td>Object level</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>Scale</td>
<td>20</td>
<td>10</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Shape</td>
<td>0.01</td>
<td>0.01</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Compactness</td>
<td>0.3</td>
<td>0.5</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>Bands* (weight)</td>
<td>B (0)</td>
<td>B (0)</td>
<td>B (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G (1)</td>
<td>G (0)</td>
<td>G (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R (2)</td>
<td>R (1)</td>
<td>R (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N (1)</td>
<td>N (1)</td>
<td>N (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>V (2)</td>
<td>V (1)</td>
<td>V (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T₁ (1)</td>
<td>T₁ (2)</td>
<td>T₁ (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>T₂ (0)</td>
<td>T₂ (0)</td>
<td>T₂ (2)</td>
<td></td>
</tr>
<tr>
<td>Object domain</td>
<td>Unclassified</td>
<td>L1 Vegetation</td>
<td>L2 Unclassified</td>
<td></td>
</tr>
</tbody>
</table>

*B = Blue, G = Green, R = Red, N = NIR, V = NDVI, T₁ = GLCM mean red band mean with 7x7 moving window, T₂ = GLCM variance red band with 11x11 moving window
Table 4 2004 image confusion matrix

<table>
<thead>
<tr>
<th>Sample/Class</th>
<th>Ag</th>
<th>Grass</th>
<th>Non_veg</th>
<th>Scrub</th>
<th>Woody</th>
<th>Class Total</th>
<th>Users</th>
<th>Commission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ag</td>
<td>677</td>
<td>29</td>
<td>27</td>
<td>19</td>
<td>13</td>
<td>765</td>
<td>0.88</td>
<td>0.12</td>
</tr>
<tr>
<td>Grass</td>
<td>31</td>
<td>352</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>401</td>
<td>0.88</td>
<td>0.12</td>
</tr>
<tr>
<td>Non_veg</td>
<td>10</td>
<td>2</td>
<td>716</td>
<td>0</td>
<td>1</td>
<td>729</td>
<td>0.98</td>
<td>0.02</td>
</tr>
<tr>
<td>Scrub</td>
<td>22</td>
<td>11</td>
<td>0</td>
<td>273</td>
<td>41</td>
<td>347</td>
<td>0.79</td>
<td>0.21</td>
</tr>
<tr>
<td>Woody</td>
<td>10</td>
<td>6</td>
<td>0</td>
<td>38</td>
<td>640</td>
<td>694</td>
<td>0.92</td>
<td>0.08</td>
</tr>
<tr>
<td>Sample Total</td>
<td>750</td>
<td>400</td>
<td>750</td>
<td>336</td>
<td>700</td>
<td>2936</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Producers</td>
<td>0.9</td>
<td>0.88</td>
<td>0.95</td>
<td>0.81</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Omission</td>
<td>0.1</td>
<td>0.12</td>
<td>0.05</td>
<td>0.19</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 5 2009 Image Confusion Matrix

<table>
<thead>
<tr>
<th>Sample/Class</th>
<th>Ag</th>
<th>Grass</th>
<th>Non_veg</th>
<th>Scrub</th>
<th>Woody</th>
<th>Class Total</th>
<th>Users</th>
<th>Commission</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ag</td>
<td>641</td>
<td>21</td>
<td>8</td>
<td>55</td>
<td>42</td>
<td>767</td>
<td>0.84</td>
<td>0.16</td>
</tr>
<tr>
<td>Grass</td>
<td>46</td>
<td>375</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>425</td>
<td>0.88</td>
<td>0.12</td>
</tr>
<tr>
<td>Non_veg</td>
<td>6</td>
<td>0</td>
<td>740</td>
<td>0</td>
<td>4</td>
<td>750</td>
<td>0.99</td>
<td>0.01</td>
</tr>
<tr>
<td>Scrub</td>
<td>53</td>
<td>1</td>
<td>0</td>
<td>335</td>
<td>55</td>
<td>444</td>
<td>0.75</td>
<td>0.25</td>
</tr>
<tr>
<td>Woody</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>9</td>
<td>597</td>
<td>614</td>
<td>0.97</td>
<td>0.03</td>
</tr>
<tr>
<td>Sample Total</td>
<td>750</td>
<td>400</td>
<td>750</td>
<td>400</td>
<td>700</td>
<td>3000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Producers</td>
<td>0.85</td>
<td>0.94</td>
<td>0.99</td>
<td>0.84</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Omission</td>
<td>0.15</td>
<td>0.06</td>
<td>0.01</td>
<td>0.16</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample/Class</td>
<td>Ag</td>
<td>Grass</td>
<td>Non_veg</td>
<td>Scrub</td>
<td>Woody</td>
<td>Class Total</td>
<td>Users</td>
<td>Commission</td>
</tr>
<tr>
<td>--------------</td>
<td>-----</td>
<td>-------</td>
<td>---------</td>
<td>-------</td>
<td>-------</td>
<td>-------------</td>
<td>-------</td>
<td>------------</td>
</tr>
<tr>
<td>Ag</td>
<td>623</td>
<td>35</td>
<td>4</td>
<td>49</td>
<td>0</td>
<td>711</td>
<td>0.88</td>
<td>0.12</td>
</tr>
<tr>
<td>Grass</td>
<td>6</td>
<td>213</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>224</td>
<td>0.95</td>
<td>0.05</td>
</tr>
<tr>
<td>Non_veg</td>
<td>6</td>
<td>16</td>
<td>744</td>
<td>6</td>
<td>6</td>
<td>778</td>
<td>0.96</td>
<td>0.04</td>
</tr>
<tr>
<td>Scrub</td>
<td>26</td>
<td>14</td>
<td>0</td>
<td>320</td>
<td>35</td>
<td>395</td>
<td>0.81</td>
<td>0.19</td>
</tr>
<tr>
<td>Woody</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>56</td>
<td>320</td>
<td>406</td>
<td>0.87</td>
<td>0.13</td>
</tr>
<tr>
<td>Sample Total</td>
<td>663</td>
<td>279</td>
<td>750</td>
<td>434</td>
<td>448</td>
<td>2574</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Producers</td>
<td>0.94</td>
<td>0.76</td>
<td>0.99</td>
<td>0.74</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Omission</td>
<td>0.06</td>
<td>0.24</td>
<td>0.01</td>
<td>0.26</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>