AUTOMATED LAND COVER CHANGE DETECTION: THE QUEST FOR MEANINGFUL HIGH TEMPORAL TIME SERIES EXTRACTION

[†‡]B.P. Salmon, [†*]J.C. Olivier, [†‡]W. Kleynhans, [‡]K.J. Wessels and [‡]F. van den Bergh

[†]Department of Electrical, Electronic and Computer Engineering, University of Pretoria, South Africa

[‡]Remote Sensing Research Unit, Meraka Institute, CSIR, Pretoria, South Africa
bsalmon@csir.co.za

[*]Defense, Peace, Safety and Security Unit, CSIR, Pretoria, South Africa

ABSTRACT

An automated land cover change detection method is proposed that uses coarse resolution hyper-temporal satellite time series data. The study compared two different unsupervised clustering approaches that operate on the short term Fourier transform coefficients of subsequences of 8-day composite MODerate-resolution Imaging Spectroradiometer (MODIS) surface reflectance data that were extracted with a temporal sliding window. The method uses a feature extraction process that creates meaningful sequential time series that can be analyzed and processed for change detection. The method was evaluated on real and simulated land cover change examples and obtained a change detection accuracy higher than 76% on real land cover conversion and more than 70% on simulated land cover conversion.

Index Terms— Change detection, clustering, satellite, time series.

1. INTRODUCTION

The transformation of natural vegetation by practices such as deforestation, agricultural expansion and urbanization, has significant impacts on hydrology, ecosystems and climate [1]. Coarse resolution satellite data provide the only regional, spatial, long-term and high temporal measurements for monitoring the earth's surface. Automated land cover change detection at regional or global scales, using hyper-temporal, coarse resolution satellite data has been a highly desired but elusive goal of environmental remote sensing [2, 3].

A time series is a sequence of data points measured at successive time intervals. Time series analysis comprises methods that attempt to understand the underlying force structuring the data, identifying patterns, detecting changes and clustering. Subsequence clustering is performed on streaming time series that are extracted with a sliding window from an individual time series [4]. A subsequence $x_p(t)$ for a given time series $x(t)$ of length N, is given as

$$x_p(t) = \left[\, x(t_p)\; x(t_{p+1})\; \dots\; x(t_{p+Q}) \,\right], \qquad (1)$$

for $1 \le p \le N$-Q+1, where Q is the length of the subsequence. The sequential extraction of subsequences in (1) is achieved by using a temporal sliding window that has a length of Q and position p that is incremented with a natural number $\mathbb{N}$ to extract sequential subsequences $x_p(t)$ from $x(t)$. The signal processing and data mining communities have made wide use of the clustering of subsequence time series, $x_p(t)$, that were extracted using a temporal sliding window. However, it has found very limited application on satellite time series data.

Recently the data mining community's attention was brought to a fundamental limitation of clustering subsequences of a time series that were extracted with a sliding window [4]. The sliding window causes clustering algorithms to form sine wave cluster centers regardless of the data set, and clearly makes it impossible to distinguish one dataset's clusters from another. This is due to the fact that each data point within the sliding window contributes to the overall shape of the cluster center as the window moves through the time series [4]. This limitation was illustrated by using data sets from various fields, i.e. stockmarket and a random walk data sets. Keogh and Lin [4] demonstrated a tentative solution that would not suffer from the afore-mentioned limitation when the procedure was applied to a periodic data set and the sliding window position p was incremented by the exact length of the periodic cycle. Since remote sensing time series data has a very strong periodic component due to seasonal vegetation dynamics, the extracted sequential time series could potentially be processed to yield usable features. These features could enable effective subsequence clustering and potentially be used for change detection.

Land cover change in context is defined here as the assignment of subsequences that are extracted from a time series that transition from one cluster to a different cluster and

This research was funded by the CSIR Strategic Research Panel.

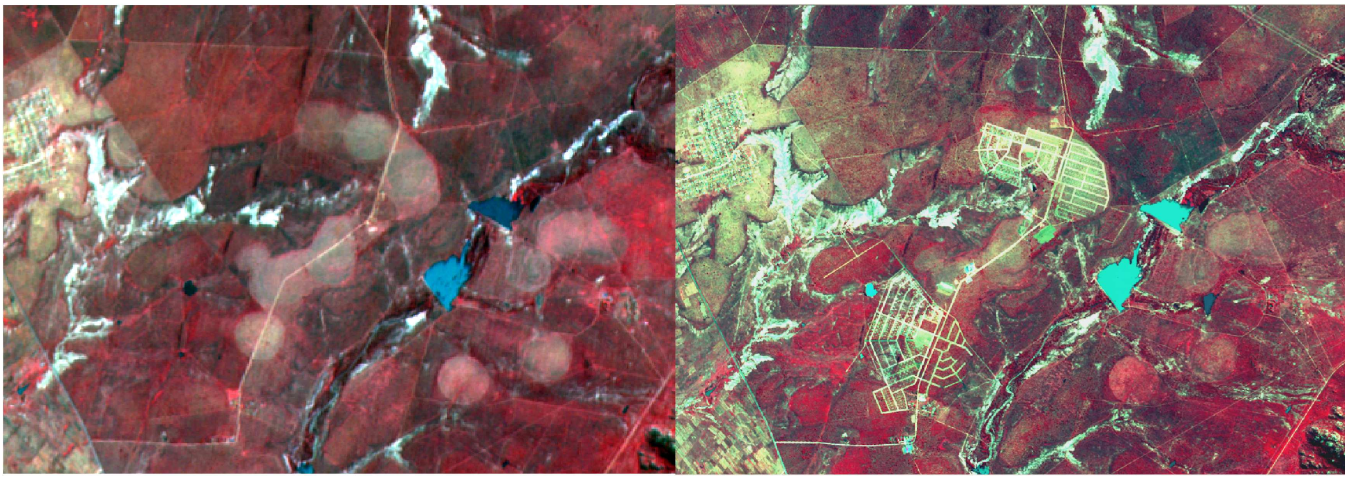

Fig. 1. SPOT2 image of 2 May 2000 of natural vegetation area in the Limpopo province (left) and a SPOT5 image of 10 May 2006 of new human settlement of same geographical area (right).

remains there for the rest of the time series.

The objective of this paper is to introduce the concept of unsupervised land cover change detection algorithm that operates on a temporal sliding window of MODIS time series data that uses a feature extraction method that does not suffer from the limitation shown by Keogh and Lin [4]. Two well-known unsupervised clustering techniques were used within a land cover change detection algorithm and were evaluated specifically on new settlement development, both real and simulated land cover change, using the 8-day composite MODIS land surface reflectance data product.

The paper is organized as follows. Section 2 presents the methodology used, while section 2.4 discusses the approach to compensating for the impairment presented with extracting subsequences from a sliding window [4]. Section 2.5 gives a brief overview of the clustering algorithm used for the unsupervised change detection and section 3 presents the results for the automated change detection on real and simulated land cover change. Section 4 presents the conclusions.

2. METHODOLOGY

2.1. Study Areas

The area of interest was the Limpopo province which is situated in the northern part of South Africa. The province is still largely covered by natural vegetation used as grazing for cattle and wildlife. The development of settlements is one of the most pervasive forms of land cover change in South Africa. The area within the province was selected where settlements and natural vegetation occur in close proximity to ensure that the rainfall, soil type and local climate were similar over both land cover types. The selected areas of interest in the study area is composed of 433.75km^2 of natural vegetation and 374.25km^2 of human settlements.

2.2. MODIS time series data

The 500-meter MODIS MCD43A4 land surface reflectance product was used as it provides nadir and bidirectional reflectance distribution function (BRDF) adjusted spectral reflectance bands. This significantly reduces noise due to anisotropic scattering effects of surfaces under different illumination and observation conditions [5]. For each pixel a time series was extracted for only the first two spectral bands from the 8-day composite MODIS MCD43A4 data set (tile H20V11) (year 2000–2008) as it was shown to have considerable class separation when the features are analyzed [6].

2.3. Data sets: Validation, Simulated and Real land cover change

The unsupervised clustering methods' generalization accuracy was assessed on a validation set. This validation set is composed of time series that were extracted from the MCD43A4 product. The time series were selected using visual interpretation of SPOT2 images in the year 2000 and SPOT5 images in the year 2006 to map areas of change and no change in land cover type during the study period. The total number of time series (pixels) available for each class is given in Table 1.

Information on known land cover change is generally very limited [7], thus the land cover change was also simulated. Land cover change was simulated by concatenating a set of time series from the natural vegetation class to another set of time series from the settlement class and vice versa. As a control, testing sets containing no land cover change were

Table 1. Number of pixels per land cover type.

Study Area	Time Series Data Set	Simulated Change Time Series
Vegetation	1235	500
Settlement	997	500

also created by concatenating the same land cover type time series to each other. Hence there were two testing data subsets based on concatenating time series of different combination of time series:

- subset 1: spliced two different land cover classes with each other;
- subset 2: spliced two of the same land cover class with each other.

These two subsets were used to produce a matching matrix to test if the unsupervised methods can detect change reliably in an automated fashion on subset 1, while not falsely detecting change for subset 2. The number of simulated land cover change time series available for the analysis process is also given in Table 1.

2.4. Feature extraction - Subsequence Time Series

In this section a method is shown that will create usable features from time series $x_p(t)$ extracted from MODIS data. The fixed acquisition rate [8] of the MODIS product and the seasonality of the vegetation in the study area makes for an annual periodic signal $x(t)$ that has a phase offset that is correlated with rainfall seasonality and vegetation phenology. The Fast Fourier Transform (FFT) of $x_p(t)$ was computed, which decomposes the time sequence's values into components of different frequencies with phase offsets. Because the time series $x_p(t)$ is annually periodic, this would translate into frequency components in the frequency spectrum that have fixed positions. This can be viewed as a fixed location for each of the features for the clustering algorithm in the feature space regardless of the sliding window position in time, which overcomes the main disadvantage to a sliding window [4]. Because of the seasonal attribute typically associated with MODIS time series and the slow temporal variation relative to the acquisition interval, the first few FFT components dominate the frequency spectrum.

Keogh and Lin [4] suggested that the sliding window position p should be shifted by a complete periodic cycle [4], but by computing the magnitude of all the FFT components removes the phase offset, which makes it possible to compensate for both the restrictive position p of the sliding window and the rainfall seasonality. The features $X_p(f)$ for the clustering method were extracted from the sliding window $x_p(t)$ by the methodology discussed above as

$$X_p(f) = |\,\mathcal{F}(\,x_p(t)\,)\,|, \qquad (2)$$

where $\mathcal{F}(\cdot)$ is the FFT function. The mean and annual FFT components from (2) were considered as it was shown in [6] that considerable class separation can be achieved from these components.

2.5. Unsupervised change detection

The clustering method was required to process subsequences of time series data and detect land cover change as a function of time. Land cover change is declared when consecutive subsequences that are extracted from one MODIS time series, transitions from one cluster to another cluster and remains in the new assigned cluster for the rest of the time series. The temporal sliding window was designed to operate on a subsequence of the time series to extract information from two spectral bands from the MODIS product. These features were analyzed with two different clustering techniques: Ward and K-means.

The Ward clustering algorithm was used as a agglomerative hierarchical clustering method, that produces a nested hierarchy of clusters of discrete objects according to some kind of proximity matrix [9]. The Ward clustering method was used because it provided the highest cophenetic correlation coefficient when compared to minimum, maximum and average link clustering [10]. The K-means method creates an unnested partitioning of the data points with K clusters.

3. EXPERIMENTAL RESULTS

3.1. Clustering accuracy - No change validation set

The clustering algorithms were tested on all the *no change* time series in the validation set and, the experimental accuracies were reported in Table 2. Each entry in Table 2 gives the average clustering accuracy calculated over 48 independent experiments (standard deviation in parentheses) using cross validation. The K-means outperformed the Ward clustering in overall clustering accuracy by 2.04% (Table 2). The more significant result is the low standard deviation obtained from the K-means algorithm.

Table 2. Classification accuracy of the validation set for the clustering methods, with standard deviation in parenthesis.

	Ward algorithm	K-means
Vegetation	80.70% (14.29)	81.30% (3.65)
Settlement	77.47% (10.19)	81.17% (2.76)
Overall	79.20%	81.24%

3.2. Change detection - Simulated land cover change

In section 2.3 two testing data subsets were introduced which produced four possible outcomes of the land cover change detection analysis. Only the *true positive* and *true negative* cases were reported, as the other two cases were simply the

Table 3. Matching matrix representing the land cover change detection accuracy on the simulated data set.

	Ward algorithm	K-means
True positive	71.49%	70.52%
True negative	75.81%	75.20%

inverse. The outcome of the change detection simulations is summarised in the matching matrix shown in Table 3. The land cover change detection accuracy differs by less than 1% between the two clustering algorithms (Table 3). The K-means was considered the better option, due to the lower standard deviation reported in the average clustering accuracy in the *no change* time series (Table 2).

3.3. Change detection - Real land cover change

Figure 1 illustrates SPOT images of real land cover change from natural vegetation (2 May 2000) to a new human settlement (10 May 2006) in the Limpopo province. This shows a new settlement had been established in the last six years. The clustering algorithms were tested on all the known new settlements developed on previously natural vegetated areas, which amounted to 21 MODIS pixels in the Limpopo province (Table 4). Even though the accuracy of 76.12% reported in Table 4 were exactly the same for all the unsupervised clustering techniques, different areas were detected by different algorithms.

4. CONCLUSIONS

In this paper, a method for unsupervised land cover change detection incorporating a temporal sliding window, operating on MODIS time series data was demonstrated. The unsupervised approaches reported *true positive* measurements of higher than 70.5% on all simulated land cover change using cross validation. The results for the detection of simulated land cover change was compared to real mapped settlement development and a *true positive* accuracy of 76.12% was achieved. The difference in change detection accuracy between the real and simulated land cover change were still acceptably close in these experiments, even though only a limited number of real land conversion examples were available.

Since the MODIS time series has a very strong periodic component due to seasonal vegetation growth, it provides the remote sensing community with a special type of data which, if processed correctly, is immune to the limitation pointed out by Keogh and Lin [4]. This is mainly due to the extraction process which produced a short-term FFT that fixed the features' positions, which allows the features to be analyzed and permits the temporal sliding window to be moved at any time increment. This should rekindle the remote sensing community's quest for automated change detection using time series as it allows them to use many different types of algorithms and methodologies on sequential time series extracted from satellite data.

Table 4. Change detection accuracy on new settlement development.

Ward algorithm	K-means
76.12%	76.12%

5. REFERENCES

[1] R. S. DeFries, L. Bounoua, and G. J. Collatz, "Human modification of the landscape and surface climate in the next fifty years," *Global Change Biology*, vol. 8, no. 5, pp. 438–458, May 2002.

[2] J. R. G. Townshend and C. O. Justice, "Selecting the spatial resolution of satellite sensors required for global monitoring of land transformations," *International Journal of Remote Sensing*, vol. 9, no. 2, pp. 187–236, February 1988.

[3] T. Loveland *et al.*, "A strategy for estimating the rates of recent united states land-cover changes," *Photogrammetric Engineering and Remote Sensing*, vol. 68, no. 10, pp. 1091–1099, October 2002.

[4] E. Keogh and J. Lin, "Clustering of time-series subsequences is meaningless: implications for previous and future research," *Knowledge and Information systems*, vol. 8, no. 2, pp. 154–177, August 2005.

[5] C. Schaaf *et al.*, "First operational BRDF, albedo nadir reflectance products from MODIS," *Journal of Remote Sensing of Environment*, vol. 83, no. 1-2, pp. 135–148, November 2002.

[6] S. Lhermitte *et al.*, "Hierarchical image segmentation based on similarity of NDVI time series," *Remote Sensing of Environment*, vol. 112, no. 2, pp. 506–521, February 2008.

[7] R. S. Lunetta *et al.*, "Land-cover change detection using multi-temporal MODIS NDVI data," *Remote Sensing of Environment*, vol. 105, no. 2, pp. 142–154, November 2006.

[8] R. S. Lunetta, D. Johnson, J. Lyon, and J. Crotwell, "Impacts of imagery temporal frequency on land-cover change detection monitoring," *Remote Sensing of Environment*, vol. 89, no. 4, pp. 444–454, February 2004.

[9] J. Ward, "Hierarchical grouping to optimize an objective function," *Journal of American Statistical Association*, vol. 58, no. 301, pp. 236–244, March 1963.

[10] A. Jain and R. Dubes, *Algorithms for clustering data*. Prentice Hall, 1988.