Monitoring Land-Cover Changes Using Satellite Imagery

Wayne Goosen
School of Mathematics, Statistics
and Computer Science
University of KwaZulu-Natal
Durban, South Africa
209515437@stu.ukzn.ac.za

Serestina Viriri
School of Mathematics, Statistics
and Computer Science
University of KwaZulu-Natal
Durban, South Africa
viriris@ukzn.ac.za

***Abstract*—Several regions around the World are currently undergoing rapid, wide-ranging changes in land cover due to human activities and natural events. These changes can have significant effect on regional and even global climate change. In this paper, focus is on urbanization in South Africa. A Post-Classification approach is employed to detect land cover changes on a specific area from a time series of satellite images. Variance in spatial resolution and radiometric resolution between images was taken into account using radiometric normalization and geometric registration. Maximum Likelihood, Neural Network, Decision tree, K-means clustering and Multi-Support Vector Machine methods were used to distinguish between the different land cover classes. All land cover class proportions are quantified in every image and this is used to monitor the change over time. For the classification using the Multi-Support Vector Machine method yielded the best classification success rate of 80.91%. The Maximum Likelihood and Neural Network methods produced 80.45% and 79.09%, respectively.**

I. Introduction

There are various types of imagery produced by a number of different satellites. SPOT, QuickBird, GeoEye-1, RapidEye and Landsat are the most commonly known satellites. Each of these satellites produce imagery that contain different image resolutions and auxiliary data. These images comprise of different bands, each on a set wavelength. In the classification process, different combinations of these bands are used to detect certain types of land cover classes. These satellite imagery all have one common aspect to them, the access to this imagery is fairly difficult from the relevant organizations that run these satellites. There exist various land cover classification programs, most commonly known is ArcGis, which uses Maximum likelihood, unsupervised Clustering and Principal Components as possible classification techniques. There has not been much research into land cover classification of True Color Satellite imagery with no auxiliary data. This process involves using only three bands of the visual spectrum comprised in the image, red, green and blue components, to classify the images through various techniques and detect the corresponding change.

This paper firstly provides a summary of the back ground and related work. The Methods and Techniques are described next, covering the preprocessing, classification and change detection methods. Finally the results are presented and discussed.

II. Related Works

There has been much research on the various change detection methods on different remotely sensed imagery. Selcuk Reis [12] Geo-referenced the images to the map of the research area , radiometric corrections were carried out by the dataset providers and the Supervised Maximum Likelihood method was used for classification. A 85.75% average classification success rate was achieved. This is higher than the result produced in this paper because of the larger testing set. The difficulties experienced stemmed from mountainous and sloping research area, noise induced by negative climate conditions and the complexity of the vegetation.

Jin Chen et al.[11] proposed a new method to improve Change Vector Analysis (CVA) in Change Detection. It attempts to determine the optimal thresholds for change detection which is the short fall of the CVA method. The process of empirically determining the threshold of change is replaced by the Double-Window Flexible Pace Search (DFPS) algorithm. The algorithm searches for the optimal threshold from training samples that leads to the maximum change detection accuracy.

Florian Sallabai [24] also utilized the Supervised Maximum Likelihood method to perform Post-Classification Change Detection and contained a number of preprocessing methods. Geometric registration was achieved by using the UTM projection. Normalization of different sun angle and solar radiance between images was achieved by calculating the Top of Atmosphere Reflectance. Spectral Radiance was used to place the images onto the same radiometric scale and cloud contaminated pixels were excluded. The results showed that the change detection suffered due to the research area being small-structured and heterogeneous.

Rupinder and Smriti Sehgal [19] compare the K-means and Back Propagation Neural Network (BPNN) methods on

ISBN: 978-1-61804-215-6

a Pixel Based Classification of remotely sensed images. The two methods were performed on the same remotely sensed image. The classification success rate clearly shows the BPNN out performing the K-means producing 80.5% and 63.3%, respectively.

III. Methodology and Design

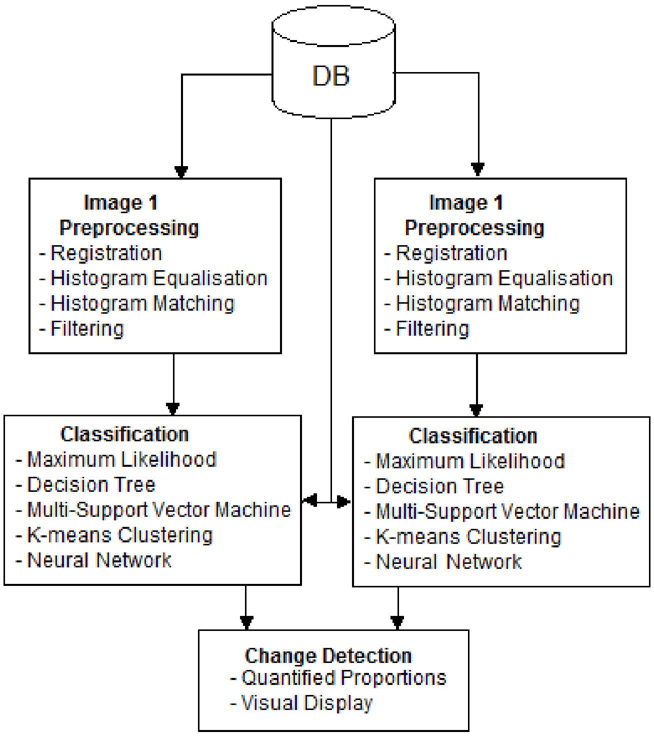


Fig. 1. Overview of the System

Figure 1 provides an overview of the system. The system is broken down into three steps, image preprocessing, classification and change detection. The image preprocessing and classification are implemented on both images. The preprocessing is divided into the following processes, Image-to-Image Registration, Histogram Equalization, Histogram Matching and Filtering. The classification consists of a number of methods, Maximum likelihood, Multi-Support Vector Machine, K-Means Clustering, Decision Tree and Neural Network methods. The change detection consists of simply quantifying the classification results.

A. Image Preprocessing

Image preprocessing consists of methods that will aid the classification process, removing any noise that will hamper the classification. Image-to-Image Registration must take place to ensure the pixels of both images correlate to one another. The three other methods are image enhancement and noise removal techniques.

1) Image-to-Image Registration: An Image-to-Image Registration is used whereby the one image is set as the reference image and the other input image is to be registered to the reference image. The method used to detect features was SURF (Speeded Up Robust Features) to detect scale and rotation-invariant points in both the images. The extracted feature vectors and relevant locations were stored. These extracted features were then matched between the two images. Using the locations of the feature vectors a geometric transformation matrix was produced between the matched features. This geometric transformation matrix is applied to the whole image and produces a image that is registered to the reference image.

2) Histogram Equalization: This method transforms the histogram of the image in an effort to enhance contrast. Once Histogram Equalization has taken place, the intensities will be better distributed on the histogram. A gray scale image G is represented by an MxN matrix of pixel intensity values. The intensity values ranging from 0 to L-1. Let p denote the normalized histogram of G as follows

$$P_n = \frac{\text{no. of pixels with intensity n}}{\text{total no. of pixels}} \quad \text{n=0,..,L-1} \tag{1}$$

The P_n is used to create a mapping function for the pixel intensities, T,

$$T(k) = (L-1)(\sum_{n=0}^{K} P_n) \quad \text{K=0,..,L-1} \tag{2}$$

The intensity value K will be mapped to T(K) in the outputted image. The dataset consists of images that contain three bands and thus this process will have to be performed on each band. An example can be seen in Figure 2.

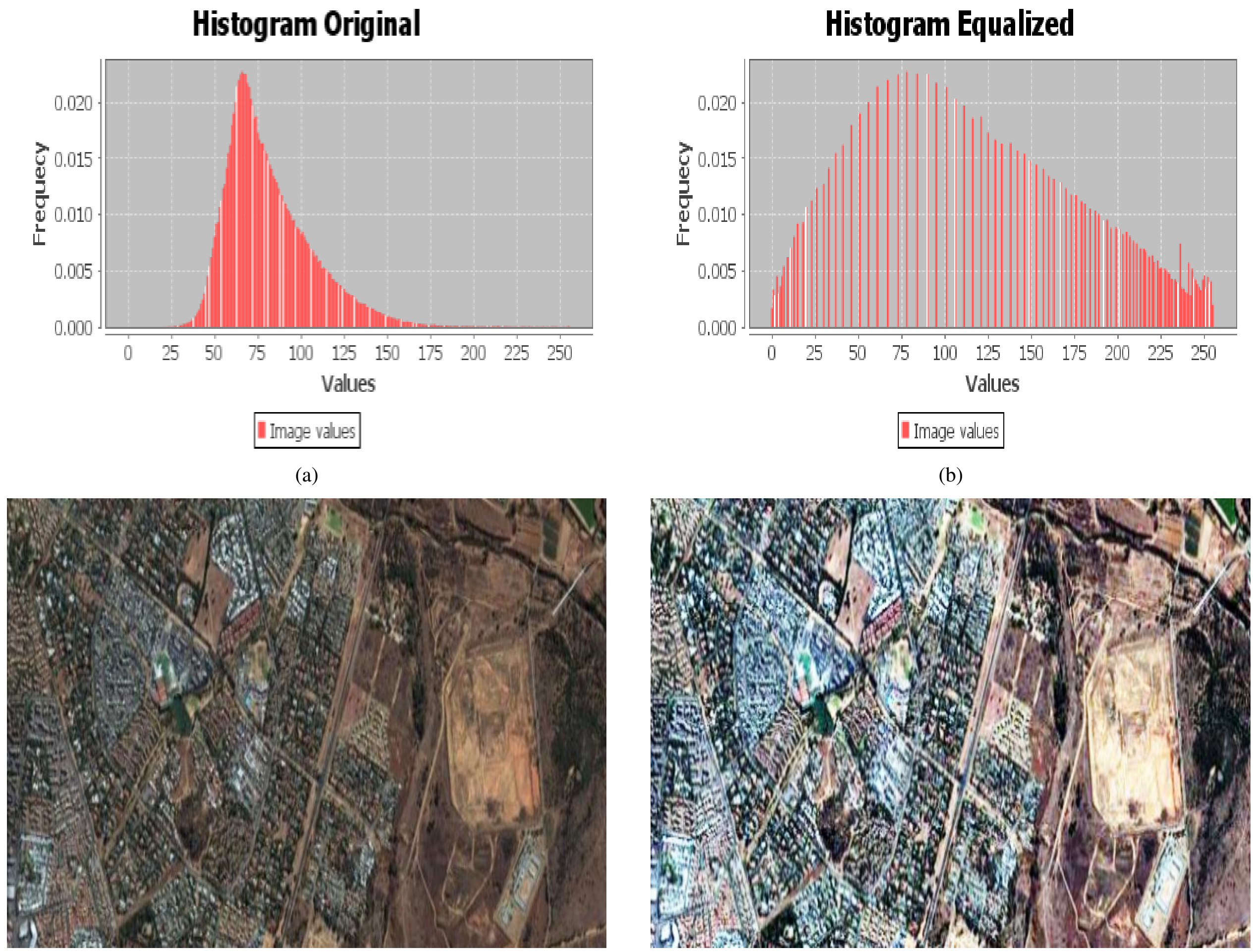
3) Histogram Matching: This method is the process of matching a histogram to a reference histogram. Using equations (1) and (2) to produce $P1_n$ and T1(k) for the original histogram, PR_n and TR(j) for the reference histogram. Suppose the mapping functions for both histograms produced arbitrary values a and b, T(k) = a and TR(k1) = b. The histogram matching utilizes a reverse mapping such that when a=b (or nearest value) the value k in the original histogram is mapped to k1. This process will be applied to all three bands of the image.

4) Filtering: Filtering is a method to reduce noise in an image. It uses a 2D matrix (known as a sliding window) that scans through the image performing different mathematical calculations on this window of values and replaces the center value in the image. There are different forms of this method but they all strive to accomplish a common goal, reducing the intensity variation amongst pixels. Median filter (median of the sliding window), Mean filter (mean of sliding window) and Gaussian filter (uses a Gaussian function to set values of the sliding window and performs the sum of products) are amongst the most common filters used.

B. Classification

Classification methods can be of two types, supervised and non-supervised. With supervised classification, training

ISBN: 978-1-61804-215-6



(a) (b)

(c) (d)

Fig. 2. (a) Histogram of Original image. (b) Transformed Histogram. (c) Original Image. (d) Resulting Image from Histogram Equalization of c.

data will have to be entered into the system to train the classifier. The number of classes and class names should be finalized before the classification process begins. The training data consists of the feature vectors of the image and the corresponding class labels. The non-supervised classifier just uses the feature vectors to classify the image into the different land cover classes. Once the pixel values are classified into the classes, a new image will be created displaying the classified pixels, each class will have a corresponding color. The images consist of a feature vector for every pixel as follows

$$F_v(i,j) = \{r,g,b\} \tag{3}$$

where i and j are co-ordinate values and r,g and b are the values from each corresponding band.

1) Maximum Likelihood: The Maximum Likelihood Classifier is a statistical based supervised classification method. It is derived from the Bayes Theorem and calculates a posteriori distribution known as P(k$|F_v$). This is the probability that a pixel with feature vector F_v belongs to class k, this calculation is given by:

$$P(k|F_v) = \frac{P(F_v|k)P(k)}{P(w)} \tag{4}$$

The likelihood function is P(F_v| k), P(k) is the probability that the class k occurs in the data used and P(F_v) is the probability that the feature vector has been observed which can be calculated using the following:

$$P(F_v) = \sum_{k=1}^{L} P(F_v|k)P(k) \tag{5}$$

L is the total number of classes. The whole image is traversed pixel by pixel and the pixel is assigned to class k by the following rule:

$$if P(k|F_v) > P(j|F_v) \text{ for all j} \neq k \tag{6}$$

For every pixel the probability will be calculated for each class and the one with the highest probability will be the class the pixel is assigned to.

2) Multi-Support Vector Machines: Support Vector machine (SVM) is a supervised learning model used in machine learning often to perform classification. The SVM classifies data into only two classes by finding the best hyperplane to separate them. An equation for a hyperplane is

$$< w, x > +b = 0 \tag{7}$$

ISBN: 978-1-61804-215-6

where $< w, x >$ is the inner product of w and x. w is an element of R^d where d is the dimension of the input vectors. Since the data can possibly be non separable, a soft margin is employed, the hyperplane will separate a good percentage of samples but not all. The soft margin implementation utilizes slack variables s_i and a penalty parameter C as follows

$$min_{\mathrm{w,b,s}}(\frac{1}{2} < w, w > +C\sum_i s_i) \tag{8}$$

Subject to

$$y_i(< w, x_i > +b) \geq 1 - s_i \tag{9}$$

$$s_i \geq 0 \tag{10}$$

This method is used for classification when only two classes are involved, in regards to land cover classes there are usually more than two classes. To use the SVM a simple transformation from the n dimensional class space to a two dimensional class space is performed. The method consists of making a SVM for each class versus the rest of the classes (combining the rest of the classes into one class to make the class space two dimensional), to produce the Multi-Support Vector Machine. The classification process consists of the feature vector of each sample (a pixel) being inputed into each SVM and which ever SVM returns the best margin from the hyperplane, the sample is assigned to that class.

3) Decision Tree: A decision tree is a method that consists of various ways of splitting a dataset into segments (branches). These branches form an inverted tree, with the root located at the top. This method can be used in classification and the leaf nodes of the decision tree will represent assignment to the land cover classes. A feature vector will start at the root and pass various decisions (branches) until it reaches a leaf node and it will be classified into that land cover class. This method also requires training data to learn how to split the data. An example is shown in Figure 3, five land cover classes were used and values at the leaf nodes will be rounded off to the nearest integer (1.364 = 1, This will represent assignment to class 1).

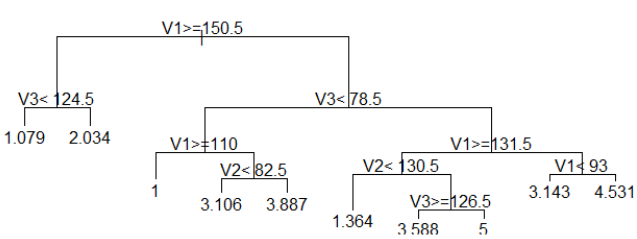


Fig. 3. Example of a Decision Tree, Input Feature Vector {V1 ,V2, V3}

4) K-means Clustering: This non-supervised classifier clusters the data by separating samples into K clusters to some objective function. This method requires a number of clusters to be specified which will be our land cover classes. The objective function used was a squared error function. For the dataset X, centroid C of a cluster is calculated as follows

$$J(X, C) = \sum_{i=0}^{n} \mathrm{min}_{\mu_j EC}(\| x_j - \mu_i \|^2) \tag{11}$$

Basic steps involved:

1) Initialize the centroids of all clusters.
2) Assign each sample to the cluster (class) that has the nearest centroid.
3) When all the samples have been assigned to a cluster (class), the centroids of the cluster need to be recalculated.
4) Repeat steps 2 and 3 until no change occurs in the centroid values. This creates clusters where by the metric to be minimized can be calculated.

5) Neural Network: Neural Networks are models build based on the central nervous system (typically the brain) that are capable of machine learning. These models are comprised of a large number of interconnected neurons working together to solve specific problems. In this paper the multilayer perceptron (MPL) neural network model is used. This is a supervised model which requires a desired output to learn. The MPL consists of an input layer of nodes, one or more hidden layers of nodes and output layer of nodes. The input passes through all the hidden layers to produce an output. The calculation of a two layer MPL is as follows

$$x = f(s) = BH(As + a) + b \tag{12}$$

Where x is the vector output and s the vector of inputs. A and B are the matrix of weights for the first and second layer. a and b are the bias vector for first and second layer. H is the activation function. The data flows through the hidden layers where they get summed and processed by the activation function. With each transfer to a different layer the data gets multiplied by the interconnection weights. The learning process is where by data goes through the MLP, the output gets compared to the required output, an error calculation is performed and the weights are adjusted as seen in Figure 4.

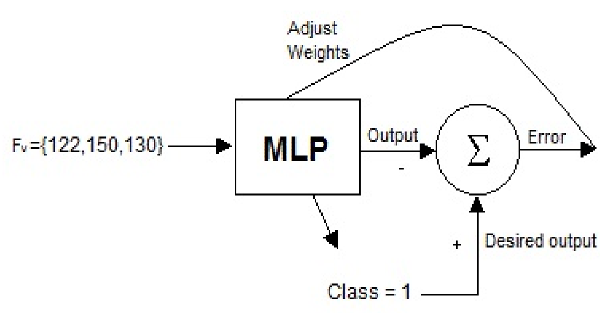


Fig. 4. Example of a MLP learning process

C. Change Detection

The Change Detection is the process of quantifying the results of the classification process. In each image the land cover proportions will be quantified, providing us with the

ISBN: 978-1-61804-215-6

amount (in percentage) of each class in that image. The percentages can be compared directly or graphed to detect the change happening over time in the land cover classes. The classified images are also displayed and the change can be detected visually.

IV. RESULTS AND DISCUSSION

The change detection accuracy is directly proportional to the classification accuracy rate. Due to this, the classification results are compared and a change detection example is examined thereafter.

A. Classification Rate

The images used to test these methods are of Amanzimtoti, Durban, South Africa. These images are from Google Earth and were acquired in 2008 and 2011. A testing set and training set were abstracted from the images with the relevant land cover class labels. For this specific image, five land cover classes were selected as follows

1) Bare soil
2) Urban area
3) Evergreen and forests
4) Grass land
5) Roads and railways

The following confusion matrices and classification rates were produced for the Maximum Likelihood (ML), Decision Tree (DT), Multi-Support Vector Machine (MSVM) and Neural Net (NN) methods. The results in the matrices are represented in percentages.

$$\mathrm{ML} = \begin{pmatrix} 21.36 & 2.27 & 0 & 0 & 0.45 \\ 1.8 & 13.18 & 0 & 1.36 & 1.8 \\ 0 & 0 & 23.63 & 2.27 & 1.36 \\ 0 & 2.27 & 0 & 11.81 & 2.7 \\ 0 & 3.18 & 0 & 0 & 10.45 \end{pmatrix}$$

Classification rate is: 80.45%

$$\mathrm{DT} = \begin{pmatrix} 18.18 & 0.91 & 0 & 0 & 5 \\ 3.18 & 8.63 & 0 & 4.09 & 2.27 \\ 0 & 0 & 19.09 & 8.18 & 0 \\ 0 & 0 & 0 & 4.09 & 12.72 \\ 0.45 & 0 & 0 & 8.18 & 5 \end{pmatrix}$$

Classification rate is: 55%

$$\mathrm{MSVM} = \begin{pmatrix} 21.82 & 0 & 0 & 0.45 & 1.8 \\ 2.7 & 8.18 & 0.91 & 0.45 & 5.91 \\ 0 & 0 & 19.09 & 3.18 & 0.45 \\ 0 & 0 & 0 & 15.45 & 1.36 \\ 0 & 0.91 & 0 & 0.91 & 11.81 \end{pmatrix}$$

Classification rate is: 80.91%

$$\mathrm{NN} = \begin{pmatrix} 22.27 & 0 & 0 & 0.45 & 1.36 \\ 4.09 & 8.64 & 0 & 0.91 & 4.54 \\ 0 & 0 & 21.82 & 2.7 & 2.7 \\ 0 & 0 & 0 & 12.72 & 4.09 \\ 0 & 0 & 0 & 0 & 13.63 \end{pmatrix}$$

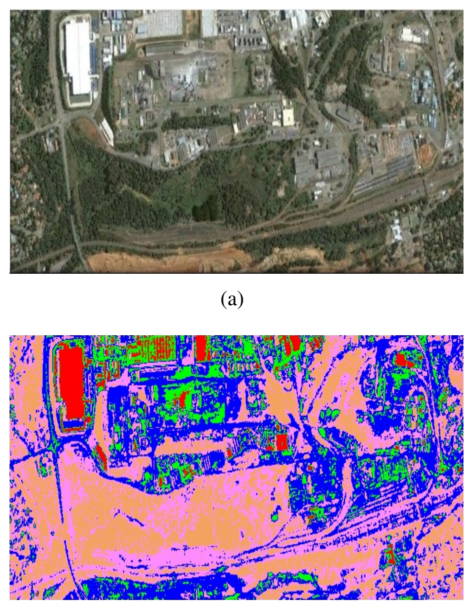
Classification rate is: 79.09%

Method	**%**
Maximum Likelihood	80.45
Decision Tree	55
Multi-Support Vector Machine	80.91
Neural Net	79.09

Table 1 Classification Success rate of methods.

Table 1 shows that the decision tree performs poorly and it suffers because the data can not be sequentially split up into the classes. The Maximum likelihood, Multi-Support Vector Machine and Neural Net methods perform well at an 80% average between them.

The sub images in Figure 5 show that even though the Maximum Likelihood method has given a classification rate of 80.45%, the results seem better in the Neural Net classified image which has a 79.09% rate. It is also so for the Multi-Support Vector Machine, visually the Neural Net out performs both the Maximum Likelihood and the Multi-Support Vector Machine methods. The K-means Clustering method is used to classify the same image, Figure 6 shows that the method performs poorly.



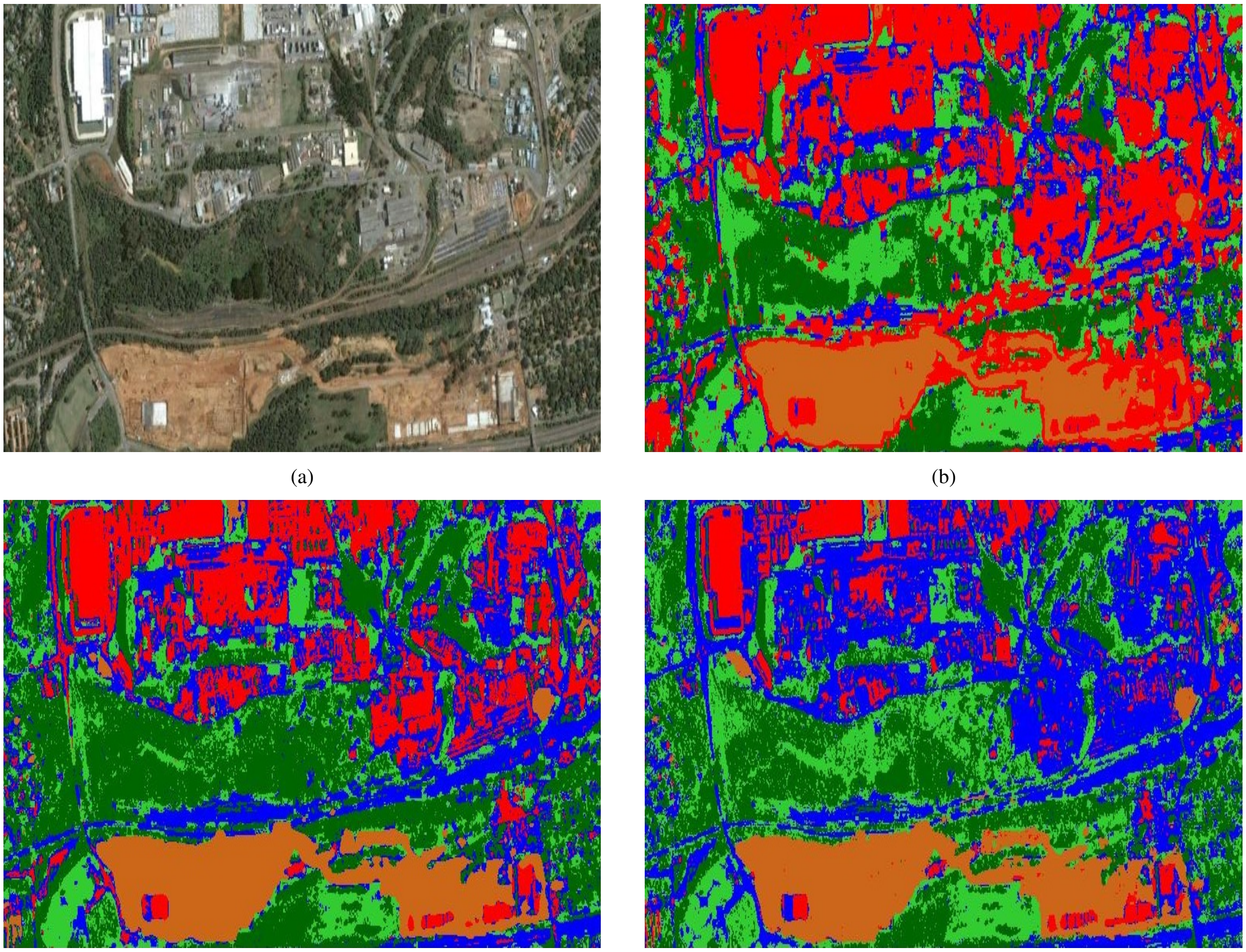
(a)

(b)

Fig. 6. (a) Original Image. (b) K-means method classified image a.

The graphical visualization of the classification in Figure 7 shows the limitations of the K-means method. The five clusters can be seen quite distinctively responding to the five classes in Figure 7 (a). They are grouped down the diagonal because of the distance measure used to create the clusters. The graph in Figure 7 (b) corresponds to the Maximum Likelihood method and the benefits are evident. In the RGB feature space the method allows classification of two classes to be parallel

ISBN: 978-1-61804-215-6



(a) (b) (c) (d)

Fig. 5. (a) Original Image. (b) Maximum Likelihood method. (c) Multi-Support Vector Machine method. (d) Neural Net method.

on the diagonal. It allows a sample to classified to a class without necessarily being in the neighborhood of the rest of the samples of that class.

B. Change Detection in Amanzimtoti

The area in Amanzimtoti was chosen because of recent developments in the small town. A golf course and surrounding areas were cleared to introduce two shopping malls and also the business park near by where development is ongoing. Evaluation of the systems potential to detect the change that is occurring in this area is performed. The resulting classification process using a Neural Network Model can be seen in Figure 8.

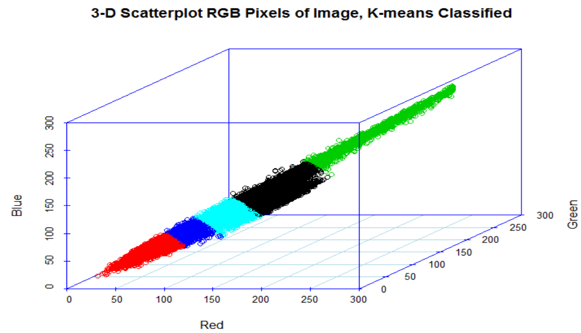
There is quite a lot of change happening and the system detects the overall change quite well as seen in Table 2. The shopping centers and adjacent roads can be seen visually occupying the bare soil over time. This is present in the system with bare soil decreasing from 10% to 1%, Urban/Residential increasing from 8% to 10% and Roads & Railways increasing from 35% to 41% in the period from 2008 to 2011.

	Year	**Year**
	2008	2011
Land Cover Class	%	%
Bare Soil	10	1
Urban/Residential	8	10
Evergreen/Forestry	27	14
Grass Land	18	32
Roads and Railways	35	41

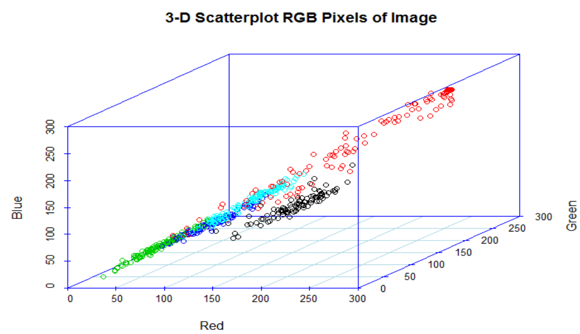
Table 2 Comparison of 2008 to 2011 Amanzimtoti Quantified Land Cover Classes.

V. CONCLUSION

This paper has compared a number of classification methods and produced competitive results compared to ones found in literature. Most notable result was a 80.91% classification success achieved by the Multi-Support Vector Machine method. The results vary compared to literature due to each change detection system using a different dataset. It can be concluded that on the dataset used, this system using a choice of preprocessing and classification methods at the users will can provide the relevant change over time from the dataset.



(a) K-means Clustering method



(b) Maximum Likelihood method

Fig. 7. 3D graph of the Classification.

Further research would be to investigate the Change Detection accuracy of these methods on images that contain more than just the three visual bands.

Acknowledgment

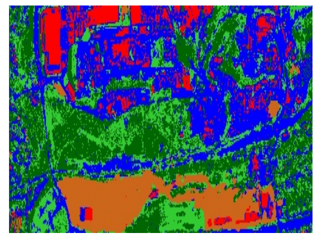
The authors would like to thank the National Research Foundation (NRF) and the South Africa National Space Agency (SANSA) for all forms of assistance during this research work.

References

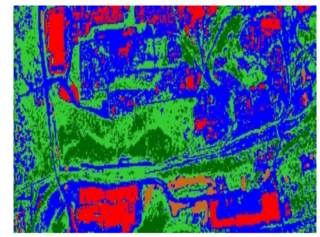
[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.
[2] Penglin Zhang, Wenzhong Shi, Man Sing Wong and Jiangping Chen. *A Reliability-Based Multi-Algorithm Fusion Technique in Detecting Changes in Land Cover*. 2013
[3] J. F Mas. *Monitoring Land-Cover changes: a comparison of change detection techniques*. 20 May, 1998
[4] D. Lu, E. Moran, S. Hetrick and G. Li. *Land-Use and Land-Cover Change Detection* 2010
[5] Wafa Nori, El Nour Elsiddig and Irmgard Niemeyer. *Land-Use and Land-Cover Change Detection* 2008
[6] Andrew A. Millward and Joseph M. Piwowar. *Exploring ordination a a method for normalizing disparate datasets: implications for digital change detection* 2002
[7] Limin Yang, George Xian, Jacequelin M. Klaver and Brian Deal. *Urban Land-Cover Change Detection through sub-pixel imperviousness Mapping using remotely sensed data* 2003
[8] Compton J. Tucker, John R. G. Townshend and Thomas E. Gof. *AfricanLand-Cover Classification Using satellite Data* 25 January, 1985

(a)

(b)



(c)



(d)

Fig. 8. (a) Amanzimtoti 2008. (b) Amanzimtoti 2011. (c) Classified image a. (d) Classified image b.

[9] Maryne Rymasheuskaya. *Land Cover change detection in Northern Belarus* 2007
[10] I Gatsis, S. Kouskoulas, A. Vafeidis, E. Lagoudakis and K Gkoltsiou. *Monitoring and Mapping of Land cover/use changes in an agriculture and natural environment, using multitemporal satellite data and GIS(Lesvos Island, Greece*
[11] Jin Chen, Peng Gong, Chunyang He, Ruiliang Pu and Peijun Shi. *Land-Use/Land-Cover change detection using improved Change Vector Analysis* 2003
[12] Selcuk Reis. *Analyzing Land Use/Land Cover Changes using Remote Sensing and GIS in Rize, North-East Turkey* 2008.
[13] Mario Caetano. *ESA advanced training course on land remote sensing, Image Classification* 30 June, 2009.
[14] J. M. Bardsley, Marylesa Wilde, Chris Gotschalk, M. S. Lorang *MATLAB Software for Supervised Classification in Remote Sensing and Image Processing*
[15] John R. Jenson *Introductory Digital Image Processing, A Remote Sensing Perspective* Third Edition
[16] Paul M. Mather and Magaly Koch *Computer Processing of Remotely-Sensed Images, An Introduction* Fourth Edition
[17] Gang Hong, Yun Zhang *Radiometric Normalization of Ikonos image using QuickBird Image for Urban Area Change Detection*
[18] Barbara Zitova, Jan Flusser *Image Registration methods: a survey*
[19] Rupinder Kaur, Smriti Sehgal *Pixel Based Classification of Remotely Sensed Image using K-means and BPNN*
[20] Anders Karlson *Classification of high resolution satellite images* August 2003
[21] Asmala Ahmad and Shaun Quegan *Analysis of Maximum Likelihood Classification on Multi-spectral Data*
[22] T. Hastie, R. Tibshirani and J. H. Friedman *Elements of Statistical Learning* Second Edition
[23] Florian Sallabai 2009 *Potential of a Post-Classification Change Detection Analysis to Identify Land Use and Land Cover Changes.*
[24] Offer Rozenstein, and Arnon Karnieli *Comparison of Methods for Land-Use Classification incorporating Remote Sensing and GIS Input* 2011