Detection and Analysis of Urban Change in Remotely Sensed Imagery by Principal Component Analysis of Image Data

Ramesh A¹, Rubina Parveen², Dr. Priya Narayanan³

Rubina Parveen, Assistant Director, Divisional Office, Employment and Training Department, Behind National Petrol bunk, MSK Mill Road, Gulbarga -585102, Karnataka, India

Abstract: This paper exhibits the time variant multispectral transformation of image data for detection and analysis of urban changes across two or more images over time. The technique covered, which appeal directly to the vector nature of the image, includes, the principal component analysis transformation (PCA). The usefulness of PCA in processing of multispectral satellite images has been highlighted. PCA reduces image dimensionality by defining new, uncorrelated bands composed of the principal components (PCs) of the input bands. Thus, this method transforms the original data set into a new dataset, which captures better the essential information. IRS IC LISS III images of 2002 and 2011 of Gulbarga area were geometrically co-registered on which PCA for the urban change Detection was undertaken. A colour composite of Eigen images from the resulting PCA was used for analysis of urban change. It has been experienced that PCA effectively summarize the dominant modes of spatial, spectral and temporal variation in data in terms of linear combinations of image frames. It provides maximum visual separability of image features thus improving the accuracy of urban change detection and analysis.

Keywords: Principal Component Analysis (PCA), Multispectral Images, Eigen values, eigenvector, Covariance Matrix, Correlation Matrix, Urban changes detection.

1. Introduction

Remote Sensing is a technique that uses sophisticated sensors to measure the amount of electromagnetic energy exiting in an object or geographic area from a distance. Hence, it helps to extract valuable information from the data using mathematically and statistically based algorithms (Basudeb 2010, Hardin PJ et al. 2007). Detection and analysis of regions of urban change in multiple images of the same scene taken at different times is of widespread interest due to a large number of applications in diverse disciplines, including surveillance, medical diagnosis and treatment, civil infrastructure, and underwater sensing (Richard J. et al. 2005). PCA is a linear projection method which, projects the data along the directions of maximal variance. The method consists of computing eigen values and eigen vectors of the covariance matrix of data. Each of the PC is otherwise the eigen vector. Using PCA change detection can be performed on urban images of two different time-periods (Basudeb 2010). This technique will result in low or null correlation for those pixels that underwent the change (Munyati C. 2004). According to this condition, high correlation will exist between unchanged corresponding pixels of two different time-period (Basudeb 2010). The most common method of utilizing PCA in urban change detection is to combine two images of different dates with n bands each into a single image with 2n bands, resulting in PC images. This transformation will produce 2n number of PC images (Basudeb 2010). The goals of PCA are to (a) extract the most important information from the data table, (b) compress the size of the data-set by keeping this important information only, (c) simplify the description of the data-set, (d) to analyze the structure of the observations and the variables with n bands. The goal is to identify the set of pixels that are "significantly different" between the last image of the sequence and the previous images; these pixels comprise the urban change mask (Richard J., Radke SA.,Omar Al-Kofahi. and Roysam. 2005). Multispectral Transformation of Image Data by Principal Component transformation uses image data statistics to define a rotation of the original image in such a way that the new axes are orthogonal to each other and point in the direction of decreasing order of variance (Eqlundh 1993). The transformed components are totally uncorrelated and are more interpretable than original data. Computationally, this can be done by 1.calculating of covariance or correlation matrix using input image data sets. 2. Calculation of eigen values and eigenvectors. 3. Calculation of principal components (Munyati 2004).

2. About PCA

Remote sensing urban change detection studies involve a series of sequential steps. Principal component analysis (Karhunen-Loeve or Hotelling transform) - PCA belongs to linear transforms based on the statistical techniques (Gonzalez R.C. and woods R.E. 2009). x1, x2xn are the values of a pixel in each of the n RGB component images, given by n-dimensional vector $x = [x1, x2, ...xn]^T$. This vector represents one common pixel in all n images as illustrated in Fig. 1. If the images are of size M*N, there will be total K=MN 3-D vectors. This is transformed into a vector y according to

 $y = A (x - m_x) \dots \dots \dots \dots \dots (1)$ The vector m_x in Eq. (1) is the vector of mean values of all input variables defined by relation

 $m_x = E(X)$(2) Matrix A in Eq. (1) is determined by the covariance matrix C_x . Rows in the A matrix are formed from the eigenvectors of C_x ordered according to corresponding eigenvalues in descending order. The evaluation of the matrix C_x of size n^*n , is possible according to relation

 $C_x = E_{i}\{(x - m_x)(x - m_x)^T\}.....(3)$ The elements C_x (i, j) lying in its main diagonal are the variances of **x** and the other values Cx (i, j) determine the covariance between input variables x_i , xj.

$$C_{\mathbf{x}}(i, i) = E_{\{(\mathbf{x}_{i} - m_{i})^{2}\}....(4)}$$

$$C_{\mathbf{x}}(i, j) = E_{\{(\mathbf{x}_{i} - m_{i})(\mathbf{x}_{i} - m_{i})\}....(5)}$$

The variances of the principal components (eigen values) contain useful information. To estimate the degree of interrelation between variables in a manner not influenced by measurement units, the *correlation coefficient*, r, is commonly used. The correlation between two bands of remotely sensed data, r_{kl} , is the ratio of their covariance (Cov_{kl}) to the product of their standard deviations ($S_k S_l$); thus:

$$v_{ki} = \frac{Cav_{ki}}{S_k S_i}$$

A correlation coefficient of +1 indicates a positive, perfect relationship between the brightness values of the two bands. A correlation coefficient of -1 indicates that the two bands are inversely related (Prieto 2000). A correlation coefficient of **zero** suggests that there is no linear relationship between the two bands of data.



Figure 1: Formation of a vector from corresponding pixels in eight bands of Multitemporal images of time T1 (2002) and T2 (2011).

Organizing the images as in Figure 1 leads to the formation of an eight element vector x from each set of the corresponding pixels in the images. The images are of size 200*200, resulting in total $(200)^2$ vectors. Covariance matrix, corresponding eigen values, and eigenvectors is computed. A set of eight principal component images were generated using the y vectors (Eq.(1)).

2.1 PCA for Urban Land Transformation

Looking at the frequency of occurrence of individual brightness values (or digital number-DN) in the image displayed in a histogram or computing multivariate statistics to determine the amount of between-band correlation (e.g., to identify redundancy), are the different ways to check the pixel values and statistics of remote sensing data(Prieto 2000).

A basic change detection algorithm takes the image sequence as input and generates a binary image [0,1] called

a *change mask* that identifies changed regions in the last image. According to the generic rule, if there is a significant change at pixel then generated image corresponding pixel will be 1 else 0 (Richard J. Radke 2005). PCA for Urban Land Transformation is of particular interest in multispectral image processing, because it can be used to de-correlate spectral correlation.

3. Study Area

Gulbarga city has a long history from the Bahamani Sultans who formed this city as their capital in 14th century, came into control of the sultanate of Delhi. From the year 1724 to the year 1948, Gulbarga was a part of Hyderabad state ruled by the famous Nizams (Naryanan 2012). Gulbarga was known as 'Kalburgi' means "rose petals" in poetic Persian. Gulbarga district is located in the Northern part of the state and lies between latitude 17010' and 17045' N and longitude 76 $^{\rm O}$ 10' and 77 $^{\rm O}$ 45'E. Gulbarga is the biggest district in Karnataka State covering 8.49% of the area and 5.9% of State's population (Naryanan 2012). It is bounded by Bijapur district (of Karnataka) and Sholapur district (of Maharashtra), in the west by Bidar district (of Karnataka) and Osmanabad district (of Maharashtra) on the north and Raichur district of Karnataka in the south. It is one of the three districts that were transferred from Hyderabad State to Karnataka state at the time of re-organization of the state in 1956. Gulbarga is an agriculture dominated District with crops such as Tur, Jowar, Bajra, Paddy, Sugarcane and Cotton.



Figure 2: Location Map of Gulbarga, Karnataka (India)

District receives and annual rainfall of 839 mm. Gulbarga area with 118.65 sq km region considered for the analysis. Gulbarga city has an area of 64.00 Sq Km with 55 Wards and a Population of 5.3lakhs (Census 2001) and is governed by Gulbarga Mahanagara Palike. Hence this study was undertaken to find the characters of urbanization in Gulbarga (Ramachandra T.V 2013).

International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Impact Factor (2012): 3.358

3.2 Data and Software

A multi temporal, multispectral digital image data set i.e. LISS-III images of March 2002 and December 2011 of 23.5mts spatial resolution was used for change detection(Gomarasca 2009). The images underwent radiometric and geometric pre-processing before PCA. All work was undertaken using ENVI 4.7, GIS 10.1 and image processing software (ERDAS). It is important to use images from the same season to avoid the change detection error introduced by mere seasonal differences. Images acquired by the same sensor is used, i.e. LISS-III throughout the study. However, the need for high spatial resolution has limitation due to the high cost and the lack of suitable images in the archive (in terms of appropriate season and freedom from cloud cover). Geo rectified data is used for analysis. Image radiometric characteristics are shown in table 2 for LISS III multispectral images of the year 2002 and 2011 of Gulbarga area.

| Table 1: Bands and Wavelength of IRS LISS III Imag | ges |
|--|-----|
|--|-----|

| Band 1 | Green (0.52-0.59) |
|--------|-------------------------------|
| Band 2 | Red (0.62-0.68) |
| Band 3 | Near-Infrared (0.77-0.86) |
| Band 4 | Shortwave Infrared (1.55-1.7) |

4. Methodology

On the 2002 and 2011 images, bands MSS1 (green) and MSS2 (red) are highly correlated because they are both in the visible region of the electromagnetic spectrum (i.e. are visible bands). Table 2, below gives the range of pixel values of the images i.e. minimum, maximum and mean. Shortwave infrared band of both the images illustrates large deviation in pixel values. In 2002 image the pixel vales in SWIR band is less than 2012 image due to the seasonal variance. This seasonal variation is attributed due to the different months in which the imagery was obtained.

| Table 2: Ima | age data Statistics |
|--------------|---------------------|
| | |

| Image | Band | Minimum | Maximum | Mean | Median | Mode | Standard deviation |
|-------------------------------------|-----------------------|---------|---------|---------|--------|--------|--------------------|
| March 2002 IRS 1C LISS III 1(Green) | | 70 | 216 | 95.567 | 97 | 91 | 8.903 |
| | 2(Red) | 50 | 205 | 83.344 | 83 | 77 | 10.75 |
| | 3(Near-Infrared) | 32 | 147 | 76.25 | 77 | 87 | 10.591 |
| | 4(Shortwave Infrared) | 29 | 200 | 127.834 | 128 | 137 | 18.835 |
| December 2011 IRS 1C LISS III | 1(Green) | 67 | 192 | 90.332 | 90 | 91 | 8.809 |
| | 2(Red) | 47 | 206 | 87.555 | 89 | 94 | 15.762 |
| | 3(Near-Infrared) | 43 | 311 | 144.654 | 143.5 | 153.97 | 18.722 |
| | 4(Shortwave Infrared) | 83 | 417 | 154.799 | 154.76 | 154.76 | 22.1622 |

4.1 Flow Chart

The change detection procedure proposed in this study is carried out with following steps as shown in flow chart below (Figure 3): Spatially registered images from two different dates are taken, and then all eight bands from two different images of 2002 and 2011 are layer stacked. PCA transformation is applied to the multispectral images. The achieved results are used for accuracy analysis.



Figure 3: Flow chart of the proposed Method

4.2 Principal Component Analysis

The band correlation of individual image is shown in table 3 and 4. On the 2002 image, there is very high correlation between Band 1 and Band 2, as well as between Band 3 and Band 4. High interband correlation indicates that the bands contain nearly the same information (in terms of radiance or reflectance data depicted) and, therefore, using one of such bands instead of both may reduce the volume of data and save computation space and time.

Table 3: Correlation Matrix of raw images of the year 2002

| Correlation | Band 1 | Band 2 | Band 3 | Band 4 |
|-------------|----------|----------|----------|----------|
| Band 1 | 1 | 0.959478 | 0.807876 | 0.879908 |
| Band 2 | 0.959478 | 1 | 0.738856 | 0.876853 |
| Band 3 | 0.807876 | 0.738856 | 1 | 0.798692 |
| Band 4 | 0.879908 | 0.876853 | 0.798692 | 1 |

Table 4: Correlation Matrix of raw images of the year 2011

| Correlation | Band 1 | Band 2 | Band 3 | Band 4 |
|-------------|----------|----------|----------|----------|
| Band 1 | 1 | 0.91375 | 0.35709 | 0.805246 |
| Band 2 | 0.91375 | 1 | 0.208116 | 0.866957 |
| Band 3 | 0.35709 | 0.208116 | 1 | 0.489504 |
| Band 4 | 0.805246 | 0.866957 | 0.489504 | 1 |

On the 2002 and 2011 images, bands MSS1 (green) and MSS2 (red) are highly correlated because they are both in the visible region of the electromagnetic spectrum (i.e. are

visible bands). Features on the urban zone which reflect highly in the green region (those that are cyan blue in colour), in the near infrared bands are independent of the visible bands(Jensen 2007), as shown by the low correlation coefficients in Tables 3 and Table 4. The near infrared bands are, however, less correlated, depicting decrease in vegetative area giving scope for increase in settlements. The marginal difference between band 3 and band 4 of correlation matrices can be attributed to the transformation that would have occurred from vegetation to built-up structures.

5. Results

The study area consists of large scale of an urban area and parts of their surrounding rural area. Number of change detection studies has been carried out for this area with Multi Scanner images of 2002 and 2011. For Change detection using PCA, the equivalent four bands from each temporal image were combined into an 8-band image. For change detection, PCA has been applied to a multi date multispectral dataset through layer stacking, with two date MSS images. Normally, the variance – covariance matrix obtained from the entire image in a systematic procedure is used to determine the PCA of the feature space. After applying the PCA to the multi-temporal MSS image, obtained a set of 8 PC Images and the 8 (figure 4).



Figure 4($\frac{ABC}{DCF}$): Eight PC Images Generated by PCA

PC2 and PC3 primarily include information about the change that can be visually detected. Change information dominant in the third principle component imagesPC3, PC1 does not include much change information. It would be desirable that change information be consistently preserved in PC3 for every change detection task.

When the images are stacked and analysed for change through PCA, if the two temporal images did not go undergo any change, nowhere the PCA will depict change. Whereas if it had undergone any change, it will be depicted in any part of the PCA as in band two and three in this case. The same has been illustrated by calculating the variancecovariance matrix mainly from some parts of changed area,

International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Impact Factor (2012): 3.358

later was applied to calculate the variance covariance matrix for the whole image.

PCA examines principal components eigenvector loadings to decide which of the PC images concentrate information

related directly to the theoretical spectral signatures of specific target materials. The technique can predict whether the target material is represented by bright or dark pixels in the relevant PC image according to the magnitude and sign of the eigenvectors (Xu 2007).

| Table 5: Image data univariate Statistics | | | | | | | | | |
|---|----------|----------|----------|----------------------|----------|----------|----------|----------|--|
| Correlation | Band 1 | Band 2 | Band 3 | Band 4 Band 5 Band 6 | | Band 7 | Band 8 | | |
| Band 1 | 1 | 0.961399 | 0.818633 | 0.885865 | 0.697533 | 0.589049 | 0.461203 | 0.586845 | |
| Band 2 | 0.961399 | 1 | 0.752298 | 0.882909 | 0.671483 | 0.617707 | 0.40877 | 0.598943 | |
| Band 3 | 0.818633 | 0.752298 | 1 | 0.808311 | 0.51045 | 0.360243 | 0.475696 | 0.429776 | |
| Band 4 | 0.885865 | 0.882909 | 0.808311 | 1 | 0.620515 | 0.546331 | 0.413439 | 0.57277 | |
| Band 5 | 0.697533 | 0.671483 | 0.51045 | 0.620515 | 1 | 0.922282 | 0.64212 | 0.879832 | |
| Band 6 | 0.589049 | 0.617707 | 0.360243 | 0.546331 | 0.922282 | 1 | 0.470255 | 0.905547 | |
| Band 7 | 0.461203 | 0.40877 | 0.475696 | 0.413439 | 0.64212 | 0.470255 | 1 | 0.682909 | |
| Band 8 | 0.586845 | 0.598943 | 0.429776 | 0.57277 | 0.879832 | 0.905547 | 0.682909 | 1 | |

| Table 0. mage data dilivaliate Statistics | | | | | | | | | |
|---|----------|---------|---------|--------|--------|--------|--------|--------|--|
| Eigenvector | Band 1 | Band 2 | Band 3 | Band 4 | Band 5 | Band 6 | Band 7 | Band 8 | |
| Band 1 | -0.245 | -0.243 | -0.195 | -0.394 | -0.280 | -0.339 | -0.412 | -0.570 | |
| Band 2 | -0.318 | -0.322 | -0.306 | -0.598 | 0.080 | 0.134 | 0.439 | 0.356 | |
| Band 3 | -0.054 | 0.016 | -0.243 | -0.079 | 0.152 | 0.472 | -0.754 | 0.343 | |
| Band 4 | -0.393 | -0.308 | -0.103 | 0.505 | -0.445 | -0.300 | -0.077 | 0.438 | |
| Band 5 | -0.089 | -0.294 | 0.876 | -0.267 | 0.004 | -0.019 | -0.179 | 0.187 | |
| Band 6 | -0.342 | -0.496 | 0.025 | 0.393 | 0.461 | 0.324 | 0.080 | -0.398 | |
| Band 7 | -0.381 | 0.322 | 0.157 | -0.007 | -0.558 | 0.592 | 0.148 | -0.206 | |
| Band 8 | -0.641 | 0.553 | 0.082 | -0.024 | 0.412 | -0.315 | -0.056 | 0.054 | |
| Eigenvalue | 2474.241 | 556.726 | 364.736 | 67.876 | 45.324 | 39.581 | 11.313 | 4.717 | |
| Percentage of data variance in component | 69.413 | 15.619 | 10.232 | 1.904 | 1.272 | 1.110 | 0.317 | 0.132 | |

Table 6: Image data univariate Statistics

A correlation matrix constructed for all eight different bands reveals the extent of change information redundancy among these difference images. An absolute correlation coefficient of 1 indicates that the change information contained in two band pairs different images is completely redundant. Correlation of 0 means there is no change information duplicating between two different images. Between the scales of 0 and 1 an absolute correlation close to 1 means that there is a high level of change information redundancy between two different images. Low absolute correlation indicates that there is a low or less significant change information redundancy .Statistics of correlation matrix (table 6) shows that change information varies from one band of change information redundancy.

According to the results obtained by this proposed method, covariance matrix is the base for identifying, PC with greatest loadings (values) represents the built-up land class, but that also has opposite Signs falling in the range of low negative to perfect negative correlations. The extraction of Built-up lands is from PC3 as four input bands have positive loadings. Multi temporal images have close positive loadings. Therefore, the built-up lands can only be mapped by PC3 based on a strong positive loading factor an opposite

sign mapped by PC1 indicates that the built-up lands presented by dark pixels. Accordingly, value (Table 5) was used to extract built-up features from the PC3 image is absolute 0 which shows perfect transformation.

To compare the extraction accuracy the original eight band images was classified using a supervised maximum fuzzy classification method with the same training regions as those used in previous classification. The area which transform from vegetation to built up accounted for 1.2 sqmtrs. The same was reflected in the PCA image with pixel value 0 which reflectance absolute transformation from vegetation to built up. According to the analysis through the proposed method, PCA image shows near about 70 pixels and the spatial resolution of the same pixels is 23.5mtrs.

Change = no of pixels multiplied by spatial resolution which is measured nearly 0.97 kms.

Hence it could be conformed apart from Spectral accuracy the PCA also proved to be spatially accurate by about 80%. Applying PCA to all the bands of multispectral images the data representation problems is automatically solved; this is an advantage of this method.



Figure 5(ABC): The Highest Contrast PC`s

Fig. 5A was created from first component of the y vectors (Eqn 1); fig. 5B was formed from second component of these vectors, and so on, so these images are of the same size as the original images. The most obvious feature in the principal component images is that a significant portion of the contrast detail is contained in the first three images and decreases rapidly from there (woods 2009.). Table 6 shows, the first three eigenvalues, are much larger than the others. Because the eigenvalues are the variances of the elements of the small y vector and variance is a measure of intensity contrast. Images resulting from vector components corresponding to the largest eigenvalues would exhibit the highest contrast. First three PC's accounts for about 95.26 % of the total variance (Figure 3). The other six PC's have low contrast detail because they account for only the remaining 4.58% of the total variance. Variances in multi temporal images represent different degree of change in different feature. In band two and band three of the PCA, most of the change information are visualized. The first three images (figure 5) from PCA image visualizes the changed area.

6. Conclusion

The choice of urban change detection algorithm is pragmatic rather than scientifically based, in other words, the selection was driven more by the application itself, than by the main issues of urban change monitoring in general. This makes it very difficult to draw comparative statements. The Proposed method detects the Urban Sprawl and analysis's the statistics of given imagery by applying PCA. In this approach change information is always accumulated in the first few PCA images and thus makes it a trivial task for an image analysis to determine which PCA image contain change information. With this procedure illustrated in this paper, it could be conformed apart from Spectral accuracy the PCA also proved to be spatially accurate by about 80%. Applying proposed method Urban Sprawl can be automatically identified, this is an advantage of this method. Further, change information of various land cover types can be eventually preserved and stored in one image channel, with high compression ratio. Despite the substantial amount of work in the field, change detection is still an active and interesting area of research. We expect future algorithm developments to be fuelled by increasingly more integrated approaches combining elaborate models of change, implicit pre and post-processing. Finally, and, perhaps, most importantly, we expect a continued growth of societal applications of change detection and interpretation in key area like geospatial intelligence.

References

- [1] Basudeb B. 2010. Analysis of Urban Growth and sprawl from remote sensing data Advances in Geographic Information Science.
- [2] Eqlundh A. 1993. A Comparative analysis of standardized and unstandardised Principal Component Analysis in remote sensing. International Journal Remote Sensing. Epub july;14.Epub 1370.
- [3] Gomarasca MA. 2009. Basics of Geoinformatics Italy: Springer.
- [4] Gomez DD, C. Butakoff, B. Ersboll et al. 2007. Automatic change detection and. 2007. Automatic change detection and quantification of dermatological diseases with an application to psoriasis images.Pattern Recognition Letters 28. Epub 2007.
- [5] Gonzalez R.C., woods R.E. 2009. Digital Image Processing Singapore: Pearson Education.
- [6] Hardin PJ, Jackson M.W., Otterstrom SM editors. 2007. Mapping, measuring, and Modelling Urban growth. Berlin: Springer.
- [7] Jensen JRaI, J. . 2007. Remote Sensing Change Detection in Urban Environment Verlag, Heidelberg: Springer.
- [8] Munyati C. 2004. Use of Principal Component Analysis (PCA) of Remote Sensing Images in Wetland Change Detection on the Kafue Flats, Zambia. Geocarto International. September 2004;19.
- [9] Prieto LBDF. 2000. Automatic Analysis of the Difference Image for Unsupervised Urban change Detection. IEEE Transaction on Geoscience and Remote Sensing.38.
- [10] Priya N. 2012. Ananlysing the growth of Gulbarga City by Semi-Automated extraction of Urban forms International Journal of the Union Geographic Information Technologists 1:7. Epub December 2012.
- [11] Ramachandra T.V BHA. 2013. Understanding urban sprawl dynamics of Gulbarga - Tier II city in Karnataka through spatio-temporal data and spatial metrics.

International journal of geomatics and geosciences 3:14. Epub March 2013.

- [12] Richard J. Radke SA, Omar Al-Kofahi, and Badrinath Roysam. 2005. Image Urban change Detection Algorithms: A Systematic Survey. IEEE Transactions on image processing. Mach;13.
- [13] Xu H. 2007. Extarction of Urban Built-up Land Features from Landsat Imagery Using aThematic oriented Index Combination Technique. Photogrammetric Engineering & Remote Sensing. December 2007; 73:10.