Principal Component Image Interpretation – A Logical and Statistical Approach

Md Shahid Latif
M.Tech Student, Department of Remote Sensing, Birla Institute of Technology, Mesra Ranchi, Jharkhand-835215

***Abstract* - Principal component analysis a multivariate statistical data analysis algorithm widely used as a dimensionality reduction algorithm in image processing task. In remote sensing data analysis, PCA used as a spectral enhancement pre-processing algorithm to reduce higher dimension space to lower dimension space with preservation of all the information in original variables. This paper provides a lucid approach to analyse and interpret PC images using statistical and logical approach. It also describe the dependency of the tonal variation of pixel vector of PCs image with magnitude and sign (negative or positive) of the coefficient of eigenvector and pixel value in original multispectral bands.**

***Keywords* - Principal Component Analysis, Eigenvalue, Eigenvector, Component Loading Factor**

I. INTRODUCTION

Principal component analysis is a multivariate statistical dimensionality reduction algorithm. Through linear orthogonal coordinate transformation it projects the sample dataset distributed in 'm' dimensional highly correlated space into uncorrelated orthogonal PC vector space with preservation of all the information of original variables. This algorithm tries to find out the direction of maximum variance on the basis of the distributed sample data in multivariate space towards which PCs vector exhibit maximum variability. Traditionally its application has been applied for image enhancement and channel reduction, also been effectively used in terrestrial change detection studies.

The transformed principal component vector exhibit orthogonal vector space in which every PC vector directed along the direction of maximum variance must orthogonal with other PC vectors. PC transformation does not leads any loss of information. Due to compression of information from high dimension to low dimension components it reduce computational complexity for interpreting the details of feature. Eigenvalue define the scene variance by each PCs and eigenvector define the direction of orientation of each PCs vector. Percentage of scene variance posed by PCs component decreases with increase in order of components. ERDAS IMAGIN, ENVI, MATLAB, be some image processing software for handling the remote sensing dataset provide the facility to perform principal component transformation (PCT) sorted the components in order of decrease in variance.

The purpose of the present article to explain the approach to interpret the principal component images in lucid manner through statistical and logical approach. We makes an assessment and evaluation of the effect of the tonal dependency of PC pixel with eigenvector (magnitude and sign (negative or positive eigenvectors (, i.e. loading))) and pixels value in original bands.

The paper has been organized into two sections as follow:

– Basic concept and terminology of PCA.
– Assessment and evaluation of the effect of combination of eigenvector with DN values of spectral bands(with the help of examples)

II. PRINCIPAL COMPONENT ANALYSIS- CONCEPTUAL UNDERSTANDING.

Principal component analysis is a lower order multivariate statistical data analysis algorithm used as a feature extraction process to reduce the dimensionality of the spectral band into few spectral component. It transform highly correlated bands into uncorrelated components through linear transformation orthogonal projection approaches. PCA preserve all the relevant information of ground features during transformation. PCA improve signal to noise ratio by discarding the extrinsic components (exhibits low variance) and remove inter-band correlation among the variables. In remote sensing, multispectral and hyper-spectral sensor observe or monitor the bio-physical characteristic of the ground features in multiple spectral channels (either continuous or discrete distribution). Hence these bands always undergoes different statistical error are as follows:

– Inter band correlation among the bands.
– Overlapping of spectral signature of multiple ground feature.
– Computational complexness.

PCA not only reduce the dimensionality of original bands but also provide a sharp spectral distinction among the multiple ground features posse spectral overlapping. Dimensionality of PC components must be less than equal to dimension of original bands.

Satellite remote sensing digital images are continuous numeric variable; therefore their dimensionality can be reduce using PCA algorithm. PCA reduces lower order dependency among the variables. PCA is a Gaussian based multivariate data analysis, which summarize the sample data using lower order statistics (1st or 2nd order). After performing PCA transformation we generate a new set of pixel array with pixel value is function of elements of eigenvector matrix and original pixel value.

Standardized and unstandardized PCA:
We perform PCA using two different statistical approaches;
Standardize PCA: standardized PC vector estimated using correlation matrix. Estimation of correlation matrix been made using the following mathematical relation;

$$V^{1/2} * \rho * V^{-1/2} = \sum \quad (1)$$

Or,

$$\rho = V^{-1/2} * \sum * V^{1/2} \quad (2)$$

Where; ρ = Correlation matrix.
$V^{1/2}$ = Diagonal Standard deviation matrix.
$\sum$ = Covariance matrix.

Hence from the above mathematical derivation equation-2 used to calculate correlation matrix from variance-covariance matrix after standardization of all the variables of covariance matrix. Standardized PCA always provide better signal to noise (SNR) than unstandardized PCA.

Unstandardized PCA: unstandardized PC vectors estimated using covariance (or $\sum$) matrix were the variables are not standardized. It poses always poor signal to noise ratio. Covariance matrix is a 2^{nd} order representative of sample data in multivariate feature space derive from mean vector measure the intersection in between bands or variables.
Mathematically:

$$\text{Cov (X Y)} = \frac{\sum_{i=1}^{n}(X-X')*(Y-Y')}{N} \quad (3)$$

Where: X and Y = observe random variable.
X' and Y' = mean of the random variable.
N= sample population of X and Y (assumed to be same).

And,

$$\text{Corr (X Y)} = \rho_{xy} = \frac{\text{cov(XY)}}{\text{Var(X)}*\text{Var(Y)}} = \frac{\sum_{i=1}^{n}(X-X')*(Y-Y')}{\sqrt{\sum_{i=1}^{n}(x-x')}*\sqrt{\sum_{i=1}^{n}(y-y')}} \quad (4)$$

Equation-4 represents the statistical relation in between correlation matrix and covariance matrix. In multivariate data analysis we have multiple variable called bands as a vector measured the radiance flux from pixel vector and plot the data in accordance with their spectral value measured by each vector in 'n' dimensional space called feature space.

Mathematical procedure to estimate PC vector as follow:

- Estimation of mean vector from sample data or pixel vector distributed in 'n'(say) dimensional feature space using the following eq^n are as ;

$$\text{Mean vector} = m = \frac{\sum_{i=1}^{n}(Xk)}{N} = \begin{matrix} m1 \\ m2 \\ m3 \end{matrix} \; ; \quad \text{for k=3(dimensionality of bands)} \quad (5)$$

  Where: X_k = spectral value of n^{th} sample data measured by k^{th} vector.
- Estimation of covariance or correlation matrix using equation-3 and 4.
- Estimation of eigenvalue and eigenvector through Eigen matrix derive from correlation or covariance matrix .where eigenvalue measure the scene variance or variability of sample data of each PCs vector and eigenvector define the direction of orientation of each PCs vector(always along the direction of maximum variance).

  - **Eigenvalue estimation**
    Let A be covariance matrix derive using equation-3 of the sample data represented in 'n' dimensional feature space plot. Then using characteristic equation we got;

$$(A - \lambda * I) = 0 \quad (6)$$

    Where: I be the unit matrix and λ be the arbitrary value called root

    Solving equation- (6) we get the root of the characteristic equation called eigenvalue say, $\lambda 1 > \lambda 2 > \lambda 3 > \ldots\ldots\ldots \lambda n$ are arrange in decreasing order of variance and number of roots must equal with no of original bands.
    Statistically we must say $\lambda 1$ exhibit maximum variance or information than 2 , similarly $\lambda 2$ exhibit larger variance than $\lambda 3$ and so on. Statistically we must say that out of m^{th} principal components we maximize the variance of the first k^{th} effective principal components called "intrinsic dimensionality" and minimize the variance of 'm-k' principal components.

  - **Eigenvector estimation**:
    Let e1, e2……..en be the n^{th} set of eigenvector correspondence to eigenvalue 'λ' then,

$$(A - \lambda * I)[e_n] = 0 \quad (7)$$

    Solving equation- (7) for different value of λ give the values of Eigen vector also called the coefficient of Eigen matrix.

- Estimation of VAF (variance accounted for) for each components.
- Estimation of Loading Factor: measure Pearson correlation coefficient in between the pixel value of PC image and original band at same pixel location (i,j). Statistically we say it measure the importance or weightage of each spectral band to the particular PC axis.

Mathematically:

$$\rho_{pq} = \frac{eqp\sqrt{\lambda p}}{\sqrt{\sigma qq}} \quad \textbf{(8)}$$

Where: p = PC Components; q = original bands;

- Estimation of contribution: measure of fractional contribution of each bands for estimating PC component images.
  Mathematically:
  Contribution of Band ‘i’ to PC ‘k ‘= $\frac{(\text{eigen vector of band } 'i' \text{ for PC}'K')^{2}}{\sum_{t=i}^{n}(\text{eigen vector of each band to PC}'k')^{2}}$ **(9)**
- Finally estimation of PC images using the following linear transformation equation as follow:

PC_j = e1j*B1+e2j*B2+……………enj*Bn **(10)**

Where: PC_j = pixel value of jth PC at spatial location say (ij), e1j, e2j…….enj = eigenvector for PCj and B1, B2………Bn = pixel value at location (ij) recorded by ‘n’ spectral band.

III. Scope of Work

From the previous discussion we highlights the statistical approach to estimate PC images and also we analyse how the tonal value at each pixel location in PCs images is a linear function of eigenvector and DN values in original bands or variables. The purpose of this paper to highlight how the pixels value in PC images varies in accordance with spectral variation in original band with magnitude and sign (negative or positive) eigenvector through examples.

IV. Dependency of Eigen vector with DN values of Spectral bands on PC images

From the above discussion it has already been shown that,
PC image DN value = $\sum_{i=1}^{n}$ eigen vector for band i $*$ Pixel value of band i

Hence from the above equation we must say that larger numeric value eigenvector exhibit maximum contribution and significance for estimation of PCs than lower numeric value eigenvectors.

Now, our first attempt to analyse the above discussion by taking some examples. These examples are generated using a different subset data of LISS3 image and Landsat7 ETM image of Ranchi district.

3.1 Example 1: (Figure 1)

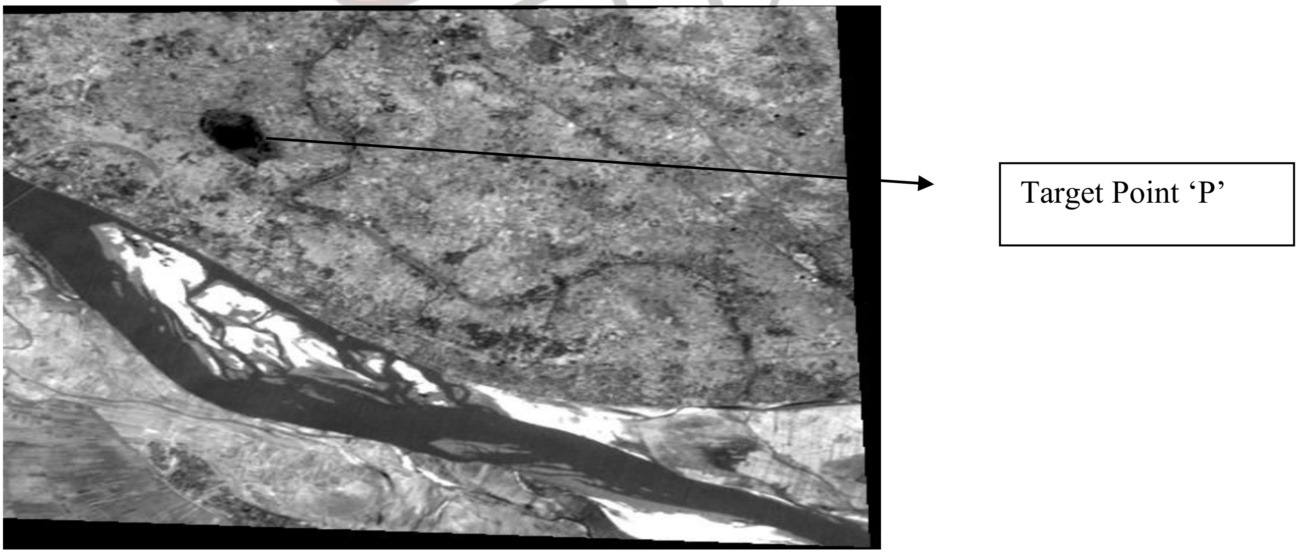

Fig-1(a): PC1 Image of Ranchi district derived from LISS3 image

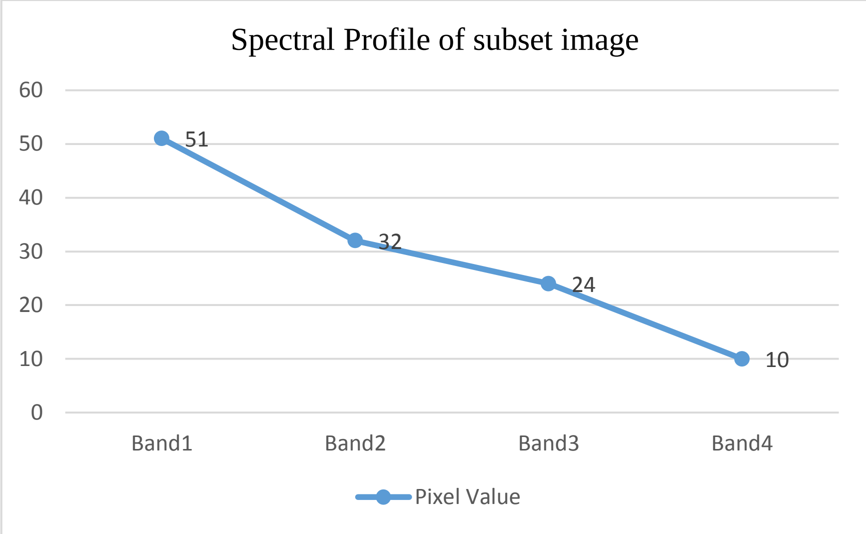

Fig-1(b): Spectral Profile curve of LISS3 subset image

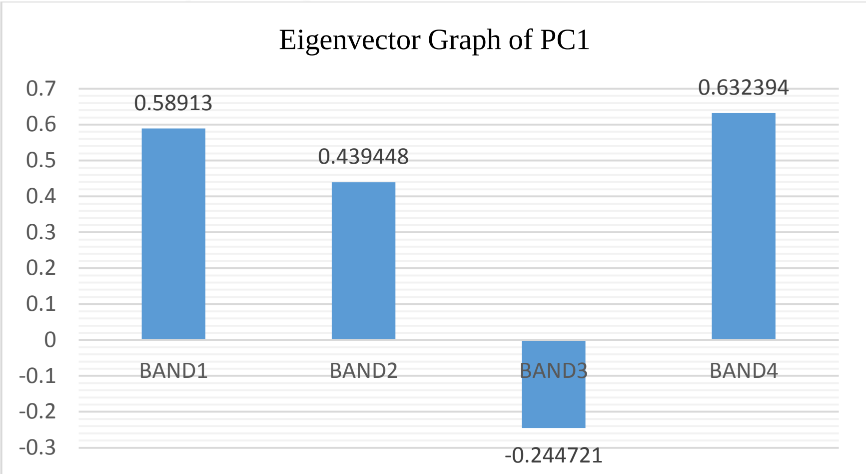

Fig-1(c): Eigenvector graph of PC1

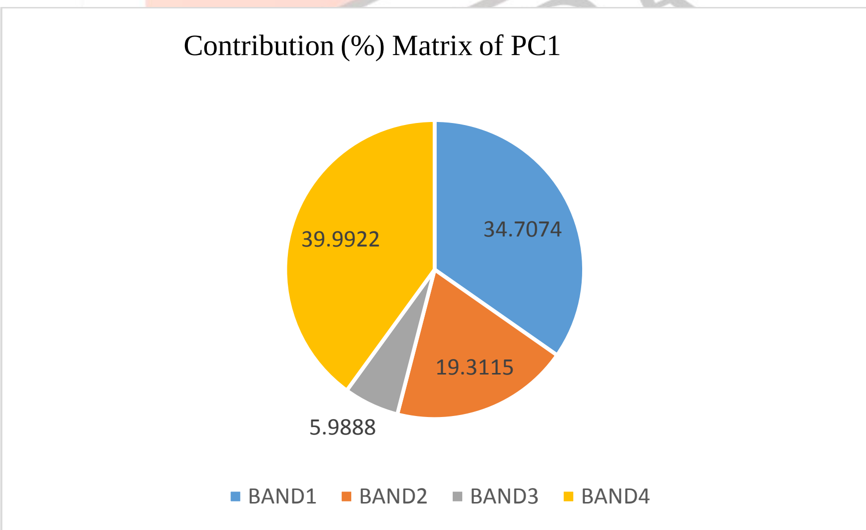

Fig-1(d): Contribution Percent Matrix of PC1

From the above Fig-1(a) shows the PC1 image of subset area of Ranchi district comprising of VNIR-SWIR bands, Fig-1(b) shows the spectral profile curve at point 'P'(as mention in fig1(a))in subset image and Fig-1(c) shows eigenvector plot for PC1 to all the bands. From Fig-1(a) we found that spectral profile curve continuously decreasing from band1 to band4 with maximum value at band1 for point 'P' in Fig-1(a) shows in band1 point 'P' are more reflective but as we move towards band4 its reflectivity goes decreasing but simultaneously from Fig- 1(c) we observe band4 exhibit maximum contribution with positive value to PC1 but pixel value at this band is very low, also spectral value at band3 is quite better than band4 but negative eigenvector makes the contribution of band3 at pixel location 'P' be low or insignificant. From contribution graph for PC1 we found that band4 possess maximum contribution or more weightage than band1.Hence from above discussion we conclude that pixel location 'P' exhibit low DN value in PC1 image or appears to darker signature.

3.2 Example 2 (Figure 2)

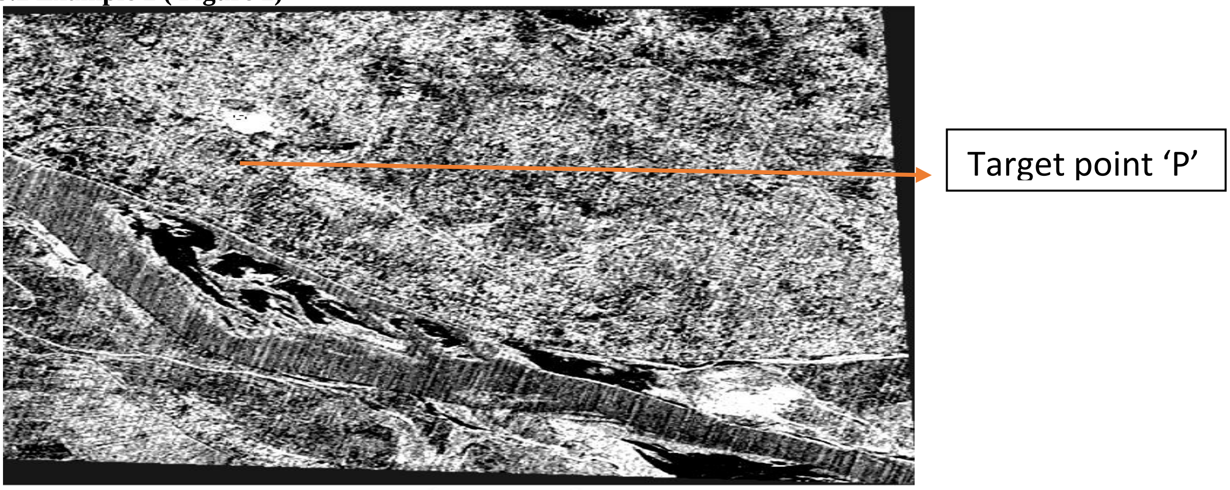

Fig-2(a): PC4 subset images of Ranchi district from LISSS3 image

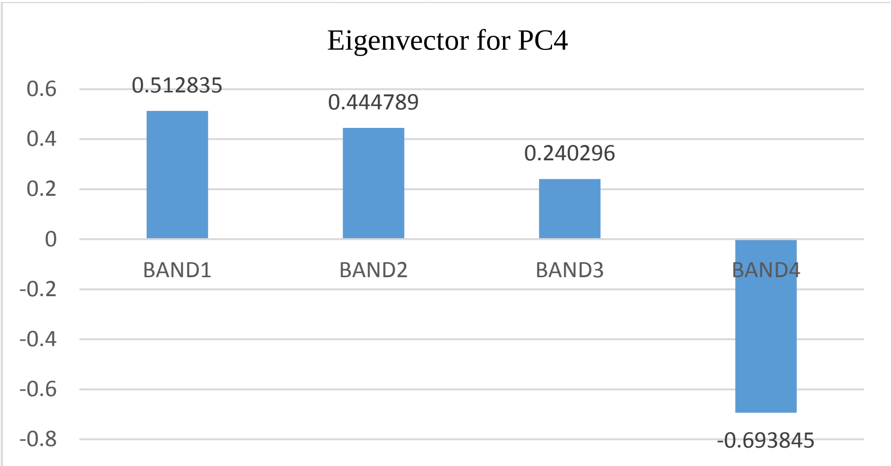

Fig-2(b): Eigenvector Graph of PC4 layer

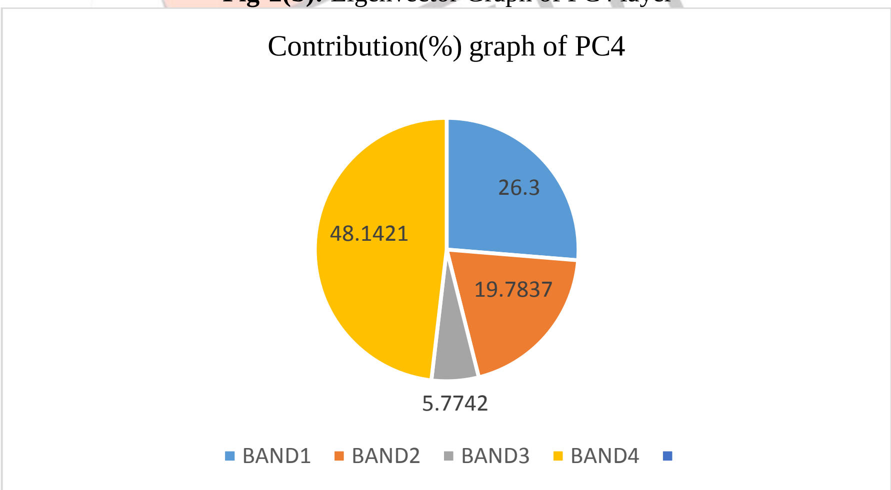

Fig-2(c): Contribution Percent Graph of PC4

Fig-2(a) show the subset image of the same area as in Fig-1(a) in PC4 component**.** Target point 'P' appears to brighter signature which are irrespective of Fig-1(a) were it appear to darker signature. From statistics graph in Fig-2(b) we observe that all the bands except band4 exhibit positive eigenvector with maximum contribution of 48.1241% to estimate PC4. Spectral profile curve Fig-1(b) shows the darkest spectral signature at point 'P'. Similarly eigenvector at band1, 2&3 exhibit positive value. From above discussion we got two result:

- Combination of negative maximum eigenvector at band4 with low spectral DN value in band4 satisfied brighter spectral signature in PC4 image at target location 'P'.

- Combination of positive moderate eigenvector at band1,2,&3 with high spectral DN value in band1,2&3 also satisfied brighter spectral signature at pixel location 'P' in PC4 images.

Hence on combination of the above two statement we got brighter DN value at this location in PC4 image.

3.3 Example 3(Figure 3)

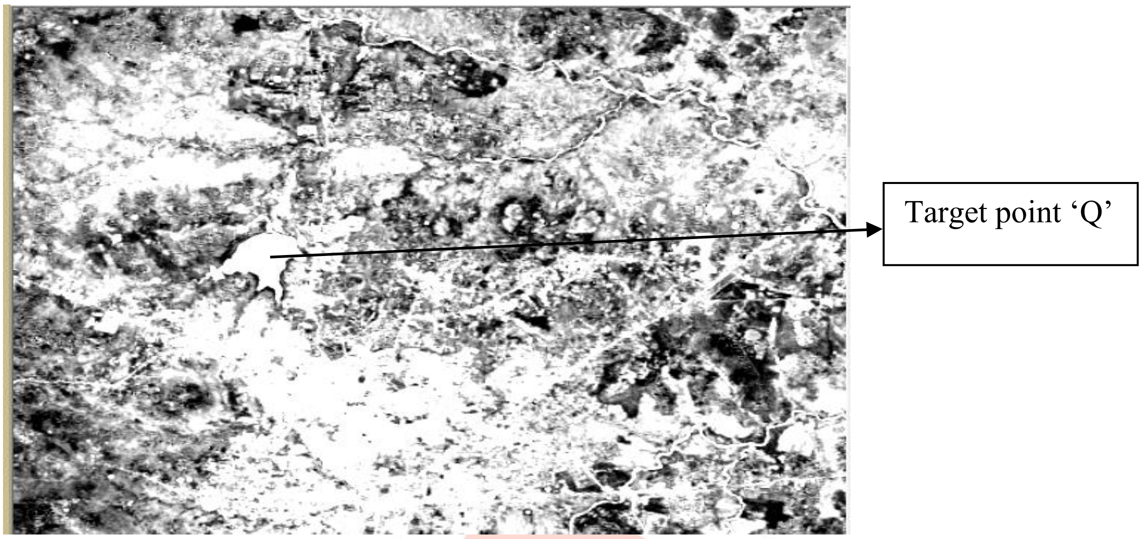

Fig-3(a): Subset Image of PC-3 of Ranchi district derived from Landsat 7 image

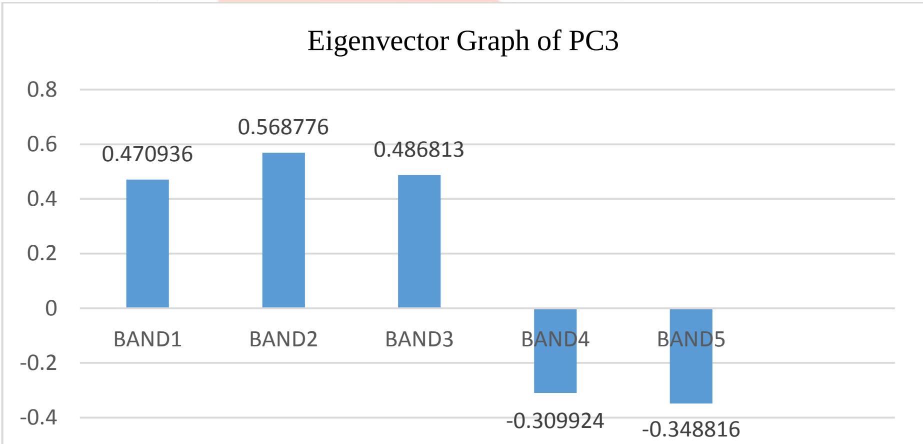

Fig-3(b): Eigenvector Graph of PC3

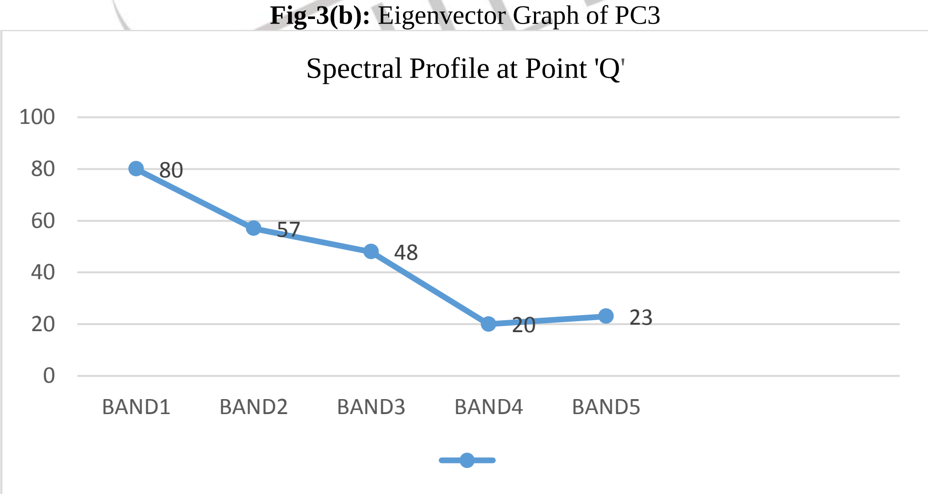

Fig-3(c): Spectral Profile at Point 'Q'

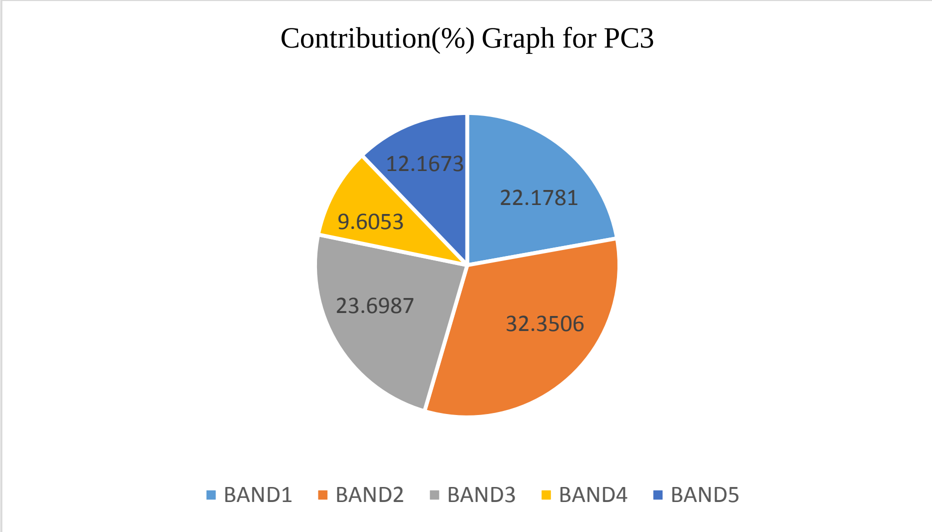

Fig-3(d): Contribution Graph of PC3

Fig-3(a) shows the PC3 subset image of Ranchi district derived from Landsat7 ETM sensor comprising VNIR-SWIR bands. Fig-3(b) shows that band2 exhibit maximum contribution of eigenvector for estimating the PC3 image with positive value. Band 4&5 exhibit low eigenvector contribution for PC3 image possess less significance than band 1, 2, &3. Spectral profile graph in Fig-3(c) shows that maximum spectral DN value at location ‘Q’ is achieved by band1 and as we move towards other bands pixel value linearly decreases. From contribution graph we observe that band1 & band2 in combination provide 54.5287 percent of the total contribution for estimation of PC3 than other bands from Fig-3(d). Band 4 &5 exhibit negative and low eigenvector with low percentage of contribution for PC3 and poor spectral value at target point ‘P’ from Fig-3(c). Statistically we get following conclusion:

- Positive and high eigenvector exhibit band 1, 2&3 with high spectral DN value enforce the DN value at target point in PC3 to be higher. Hence it emphasize brighter pixel value
- Low and negative eigenvector consist band 4&5 with low spectral DN value due to counter effect enforce the output DN value in PC image at target point to be positive or brighter spectral signature.
- Actually in both case; whether high negative eigenvector with low spectral DN value bands or high positive eigenvector with high spectral DN value bands, we got brighter pixel vector in PC images at that location. Similarly, whether high negative eigenvector with high spectral DN value bands or high positive eigenvector with low spectral DN value bands, we must got dark pixel vector in PCs image at that location.

3.4 Example 4(Figure 4)

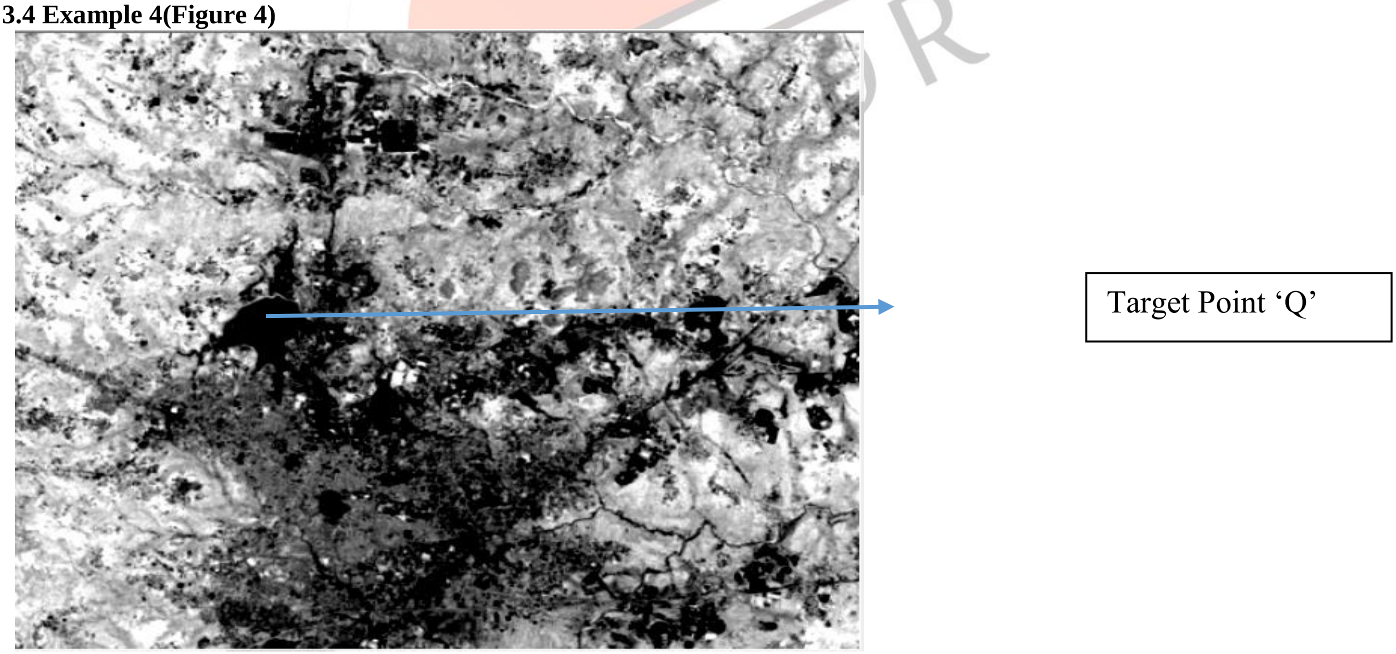

Fig-4(a): Subset of PC-2 layer of Ranchi district derived from Landsat 7

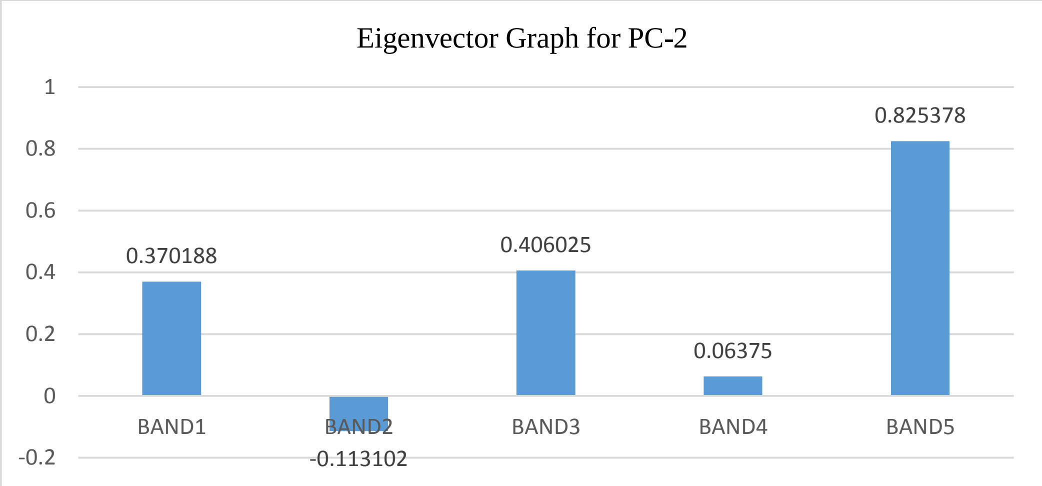

Fig-4(b): Eigenvector Graph of PC-2

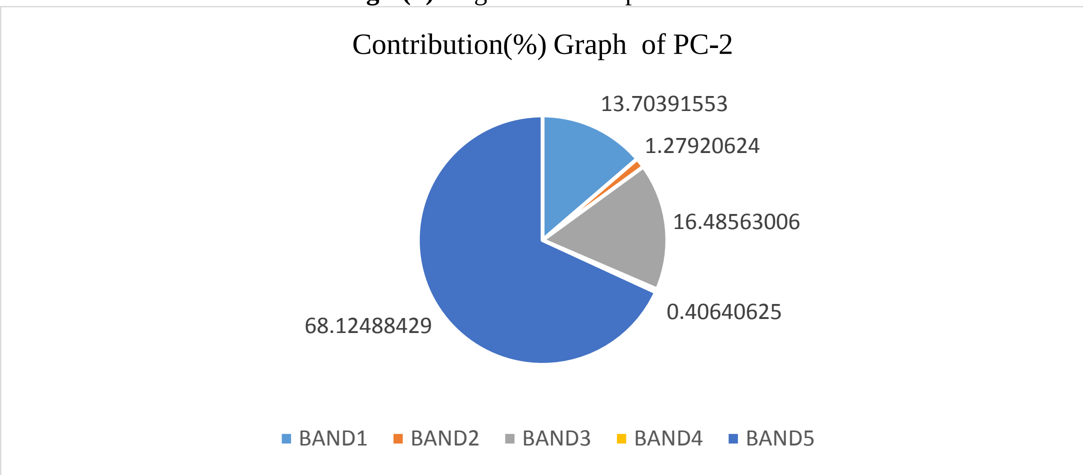

Fig-4(c): Contribution Percent Matrix of PC-2

From Fig-4(a) we observe that the brightness condition at pixel location 'Q' in PC-2 is dark which are opposite w.r.t to Fig-3(a). From the above statistical calculation and graph associated with PC2 image and original bands, we found that band 5 exhibit maximum positive eigenvector from Fig-4(b) with maximum percentage of contribution to estimate the PC-2 of 68.12% from Fig-4(c) and band 1& 3 both exhibit low contribution of 13.07% and 16.4856% to estimate PC-2 image. But spectral graph of Fig-3(c) shows band 5 exhibit very low spectral value than band 1, 2&3. From above discussion we conclude the following facts:

- Low spectral DN value with high eigenvector contribution by band5 has negative impact on PC-2 at point 'Q' makes the low DN value.
- High spectral DN value with low positive eigenvector contribution by band 1&3 has negative effect on PC-2 image. Hence pixel value must be dark spectral signature at point 'Q'.

V. CONCLUSION

From the above experimental analysis of different multispectral dataset we observe the dependency of PCs image pixel value at any spatial location as a linear function of magnitude and sign of eigenvector and spectral profile of original multispectral bands or layer for same pixel location, i.e. brightness value in PC images at each pixel vector location depend upon spectral profile of the pixel vector at same spatial location and magnitude and sign of Eigenvectors.

REFERENCES

[1] Hair, Andreson, Tatham and Black, "Multivariate Data Analysis," Low Price Edition, 2001
[2] Johnson and Wichern, "Applied Multivariate Statistical Analysis," Prentice Hall, New Jersey, 2001.
[3] P.M. Mather, "Computer Processing of Remote Sensed Images, "John Wiley & Sons Ltd, West Sussex, 2004.
[4] Lillesand , Keifer, Chipman ," Remote Sensing and Image Interpretation," John Wiley & Sons Ltd, 2008.
[5]
[6] A. Singh and A. Harrison," Standardized Principal Component Analysis," *International Journal of Remote Sensing*, Vol.6, no.6, 1985, pp.883-96
[7] Sabins, F.F., Jr., Remote Sensing –Principal and interpretation, 3rd ed., W.H.Freeman, New York, 1997.
[8] Ravi P.Gupta, Reet K. Tiwari, Varinder Saini and Neeraj Srivastava," A Simplified Approaches For Interpreting Principal Component Images," *Advance in Remote Sensing*, 2013, 2, 111-119