# Significance of Eigen Matrix in Spectral Domain of Remote Sensing Images (RSI)

S.Murugan
Associate Professor of Computer Science, Nehru Memorial College, Puthanampatti, Tamil Nadu, India

Dr.C.Jothi Venkateswaran
Head, P.G.Department of Computer Science, Presidency College, Chennai, Tamil Nadu, India.

Dr.N.Radhakrishnan
Director, Geocare Research Foundation, Chennai, Tamil Nadu, India.

## ABSTRACT
Information extraction from RSI involves a significant level of testing and experimentation before arriving at an acceptable solution. It includes combination of techniques that hardly have clear cut rules except generating desired output with acceptable level of accuracy. This may contain many levels of mining techniques depending upon the level of information required, time and system efficiency. The first level of image mining may be involving some primitive operations to reduce noise, enhancement and filtering in RSI domain. Secondly, the process may involve image segmentation and recognition of features. Finally, the image mining could involve cognitive analysis and extraction of features from RSI. Another important factor about RSI is its multiband information about objects that require a more complicated procedure even at the preprocessing level. The multilayered RSI data may be reduced to a single band data without losing much information by using Eigen values. The output PCA image thus derived may help in identifying prominent features and encourage further extension towards cognitive information extraction process.

## Keywords
RSI, Eigen values, Eigen vectors image mining, PCA.

## 1. INTRODUCTION
Information extraction from Remote Sensing Image(RSI) data has become significant as it gives diverse varieties of information about various earth features and objects. RSI technology progresses from coarse level of data gathering to finer level, the dimension of information extraction has also become complex. Such complexities encouraged varied types of image mining techniques and their combination. Image mining concerns the extraction of implicit knowledge, image data relationship, or other patterns not explicitly stored in the images [1]. In the current information age there is an explosive expansion of digital data generated and stored in databases. Identifying potentially useful knowledge from such databases is not a trivial task and is resulting in the growing interest in data mining. As such it may be defined as an extension of data mining overlapping into image domain. Since RSI provides information about earth features such as land use, coast, forest, water bodies and mixing up of all these features, it requires an interdisciplinary endeavor drawing expertise in computing techniques, understanding the inherent qualities of an image [2] and data mining [3], apart from earth science domain expertise. RSI is employed in many applications such as environmental, resources inventorying, disaster monitoring and mitigation, urban planning and so on. Any such geospatial intelligence may require proper image retrieval, reconnaissance appraisal of environment and spatial relationship among objects to have significant understanding of relevant geospatial information. Many studies have been conducted in this sphere to establish certain methods to extract information about objects from RSI so that the study may be extended for a more thorough and sustained approach. Some studies as conducted by Datcu et al. [4] shows the results of breaking an image into Regions of Interest (ROI) for the purpose of classification, query-by-shape method as defined by Dell'Acqua [5] using point diffusion technique for efficient object comparisons in remote sensing images, Prasher and Zhou [6] explaining an efficient scheme for encoding the spatial relationships of objects, attempting to discuss techniques and intricacies within RSI towards relevant information extraction. It depends upon the use geographic properties to model global object dependence as shown by Bian and Xie [7]. However, it may be assumed that conceptual based image mining has to evolve further and the complexities are added by multi-dimensional nature of RSI and related information about objects. In this context present paper discusses reduction in the dimension of RSI by data compression technique using *eigen matrix*. It examines the specificity of image data mining and *eigen values*, its challenges, and implementation model.

## 2. BACKGROUND OF THE STUDY
The eigen value problem is a problem of considerable theoretical interest and wide-ranging application. *Eigen matrix* problem is crucial in solving systems of differential equations, analyzing population growth models, and calculating powers of matrices having wide applications such as physics, sociology, biology, economics and statistics have focused considerable attention on "eigen values" and "eigen vectors "their applications and their computations. To elaborate further, matrix $A$ is multiplied with column matrix $C$ ($AC$). Then det [P] from $P$ matrix formed by column matrices Cn is evaluated ($P^{-1} AP$) resulting in a diagonal matrix. In other words, let A be a square matrix and a non-zero vector C is called an *eigen vector* of A *if* and *only if* there exists a number (real or complex) $\lambda$ such that $AC = \lambda C$ and $\lambda$ is called the eigenvalue of A. The vector C (column matrix) is called eigenvector associated to the eigenvalue $\lambda$. Role of *eigen matrix* is significant in terms of Principal Component Transformation (PCT) reducing the

data redundancy, especially in image mining in RSI environment.

It is known that analysis of all multispectral information of an object in RSI is inefficient since the data is redundant. Such spectral redundancy may be reduced by a space transformation using linear transform type with an image oriented matrix called *eigen matrix*. In its simplest form this may be written as PC = $I_{PC}$ . DN. This transformation changes the covariance matrix as $C_{PC}=I_{PC}.C.I_{PC}^T$ and the PCT is optimized by choosing the particular $I_{PC}$ showing diagonal covariance matrix of the original multispectral RS image, which may be written as

$$C_{PC} = \begin{Bmatrix} \lambda_1 \ldots\ldots 0 \\ \\ 0 \ldots\ldots \lambda_k \end{Bmatrix}$$

showing the **K** roots of *eigen values* $\lambda_k$ forming [ C – λI ], where C is the original covariance matrix and I is the diagonal identity matrix. Each *eigen value* is equal to the variance of the respective transformation of PCT of image along a new coordinate axis. Also, the sum of all the *eigen values* is equal to the sum of all the band variances of the original image. In this way, the total variance in the image is preserved. Since the $C_{PC}$ is diagonal, the transformed principal component image is uncorrelated. The transformation order is determined by decreasing variance so that $PC_1$ has the largest variance and $PC_k$ has the smallest variance. The resultant transformed image reflects the compressed spectral information in *k* dimension to a single dimensional image apparently without losing out significant information. In this context, it is to be explained that the transformation axis is defined by the *eigen vector*s $e_k$ as obtained from the vector-matrix equation for *eigen values* $\lambda_k$ . This may be expressed as (C - $\lambda_k$ I) $e_k$ = 0 and form a row of transformation matrix $I_{PC}$.

$$I_{PC} = \begin{bmatrix} e^t{}_1 \\ \\ e^t{}_k \end{bmatrix} = \begin{bmatrix} e_{11} \ldots e_{1k} \\ \\ e_{k1} \ldots e_{kk} \end{bmatrix}$$

where $e_{ij}$ is the $j^{th}$ element of $i^{th}$ *eigen vector*. These *eigen vector* components are the direction cosines of the new axis to the relative original axis. The corresponding *eigen values* of the matrix gives an indication of amount of information the respective principal components represent.

## 3. DESIGN OF ALGORITHM

It involves finding *eigen values* and corresponding *eigen vector*s of the data set using covariance matrix. The methodology for calculating principal component is given by the following algorithm. Let $x_1$,……..$x_m$ are the layers of the image of dimension MN. The corresponding algorithm in its simplistic form is given below:

1. Computation of the global mean (x) from the selected image: $\bar{x} = 1 / M \Sigma x_i$
2. Computation of the sample covariance matrix (C) of dimension (N x N)
3. Computation of the *eigen values* of the covariance matrix. Computation of *eigen values* is performed by iteration method. C : $\lambda 1 > \lambda 2 > \ldots\ldots > \lambda_N$
4. Computation of the *eigen vector*s for the *eigen values*
5. Dimensionality reduction step and output image of RSI

The above said steps are needed to generate the principal components of the image. Corresponding *eigen vector*s are uncorrelated and have the greater variance. In order to avoid the components that have an undue influence on the analysis, the components are usually coded with mean as zero and variance as one. This standardization of the measurement ensures that they all have equal weight in the analysis.

## 4. SIGNIFICANCE OF PCT

The Principal Components Transformation (PCT) has been used in remote sensing image for different purposes. A comprehensive summary of different applications of PCA, including correlation analysis of RSI for effective feature recognition and change detection with multi-temporal images has been amply demonstrated [8]. PCA is a coordinate transformation typically used to remove the correlation contained within the multi-band imagery by creating a new set of components, which are often more interpretable than the original images. The images thus generated are uncorrelated and ordered by decreasing variance. The covariance matrix of the transformed data is a diagonal matrix of which the elements are composed of the eigen values.

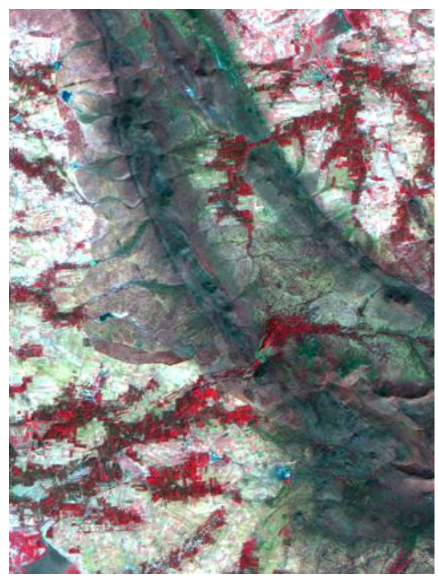

Figure 1. IRS-1C LISS III image

The significance of principal components is to produce uncorrelated output bands, to segregate noise components, and to reduce the dimensionality of data sets, a data compression [9, 10]. Because multispectral data bands are often highly correlated, the PCT is applied to produce uncorrelated output bands. This is done by finding a new set of orthogonal axes that have their origin at the data mean and that are rotated so the data variance is maximized. PC bands are linear combination of the original spectral bands. The first application of PCA consists in the image colour reduction, i.e., three colour components (RGB) of RSI are reduced into one containing a major part of information. The second use of PCA is to takes advantage of eigenvectors properties for determination of selected object orientation. In the present paper, the former aspect is applied on RSI and discussed in the following section.

## 5. PCA AND INFORMATION EXTRACTION FROM RSI

As stated in the above section, the concept behind PCA is to reduce the dimensionality and to perceive information from the selected RSI (Figure 1). It helps to reduce the size of the image to be processed and provide better run time extraction of relevant information. RSI showing a hilly terrain is selected for the present observation. Using ERDAS Imagine and Mat lab software, DN values of the pixels in each band of the RSI is extracted and processed for principal component transformation of the image to study the effect of reduction of band in PCA image and in turn the characteristic behavior of the statistical information of the image.

The selected RSI is an 8bit BIP image having grey values ranging from 0 to 255 and their pixel frequency and DN values are studied (Figure 2). While observing the histogram distribution of DN values of band that operates in *blue region* of the electromagnetic spectrum(ems), it shows that it has DN range from 0 to 178 with a mean of 94.96, median 91 and $\sigma$ =15.052. Second band in the *green* region shows that DN value has maximum value of 152 with a mean of 69.63, median value 65 and $\sigma$ =18.467. Third band in *red* region shows a maximum DN of 145, mean 87.73, median 86 and $\sigma$ =15.607. The value of DN in *infrared* region shows a maximum of 205, mean as 113.68, median 108 and $\sigma$ =29.43. An overview of statistical distribution of DN values increases the ability to appreciate the nature and occurrence of spectral reflectance of features. For example, mean values of all the four bands signify a higher frequency of DN values fall in the range values of vegetation that show a similar spectral reflectance. Distribution of DN values of the image (636 x 995) in all the selected four bands showed a distinct variation that could throw light on the density or frequency of the presence of objects or *features* within the selected RSI and knowledge on the significance of variance in DN values in extracting information about those objects (Figure 3). Such information on these objects could be extracted using PCA without losing much information. This would help in having a preliminary type of information extraction or reconnaissance information from the RSI. Higher DN value in the fourth band revealed regarding high reflectance objects mostly hills and vegetation and absence of waterbodies in the selected RSI.

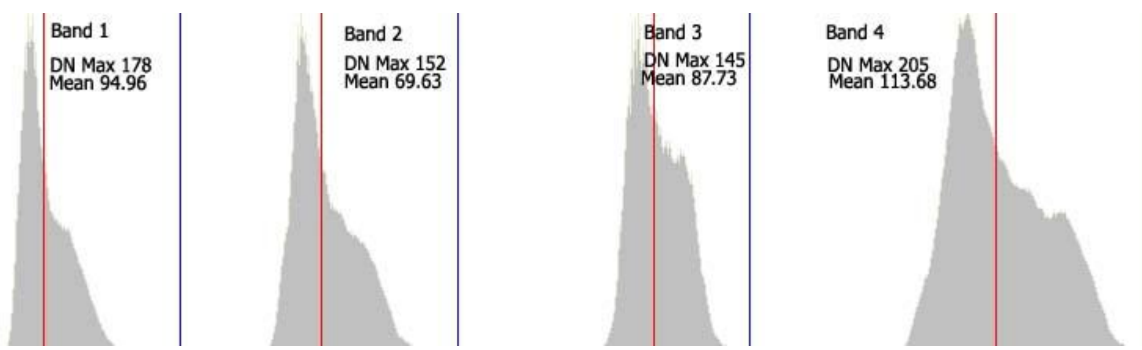


**Figure 2. Histogram showing pixel frequency and DN values of bands of RSI**

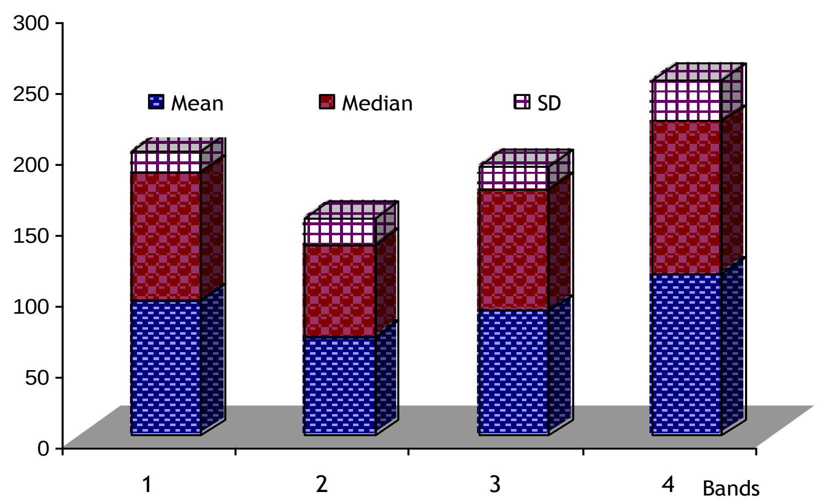


**Figure 3. Chart showing statistical parameters-mean, median and SD of RSI**

PCA on the selected RSI (Figure 4) has effected in compression of bands into a single layer with pixels having DN values of minimum 0 to maximum 326.5, mean 180, median 169.62 and $\sigma$ =39.575. When the same RSI is rotated for a second order PCT then the information present in the image is much sharper than the first order though some loss in information is evident. For such a rotation of RSI, the minimum value may go below 0, in this case-101.15 to maximum DN value of 1.8717 with a mean –35.97, median value -34.75 and $\sigma$ =8.772. Still the RSI may be rotated further to third order revealing much more information that is subtle and hazy for identification giving a reversal tonal value to the image. In this, the third order value of DN range from -104.57 to 11.693, an increase in DN value and a reduced mean and median values almost half to the mean and median values of the second order (-17.93 and -17.83 respectively) showing $\sigma$ =5.196 (Figure 5).

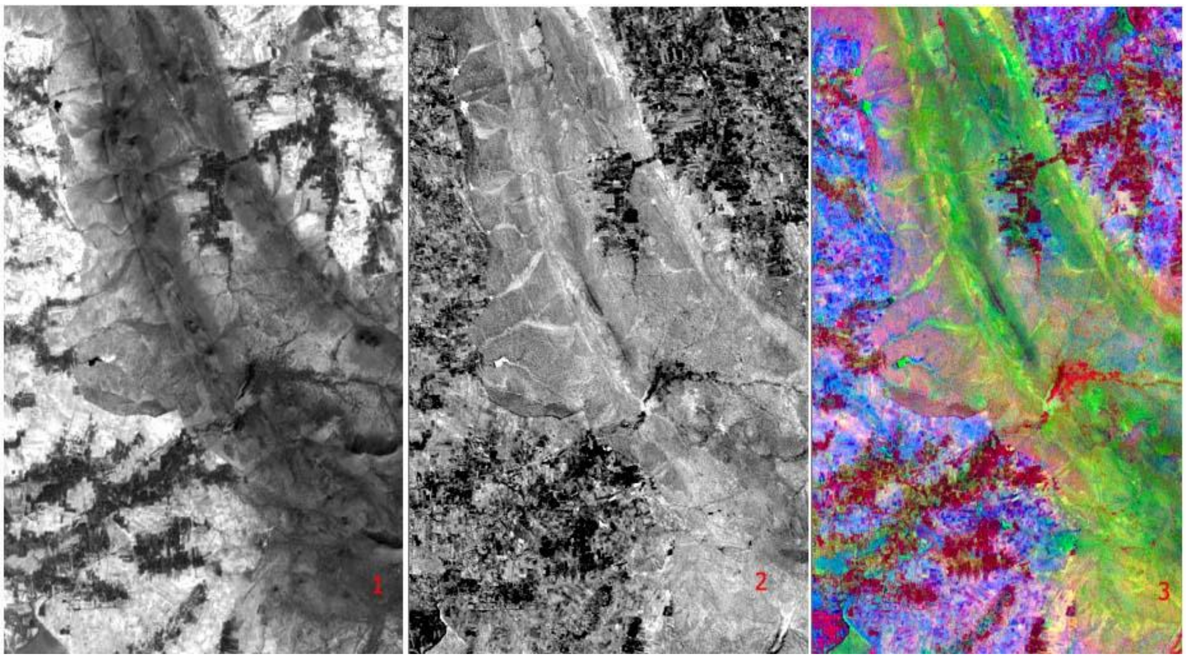


**Figure 4. PCA image of the selected RSI**

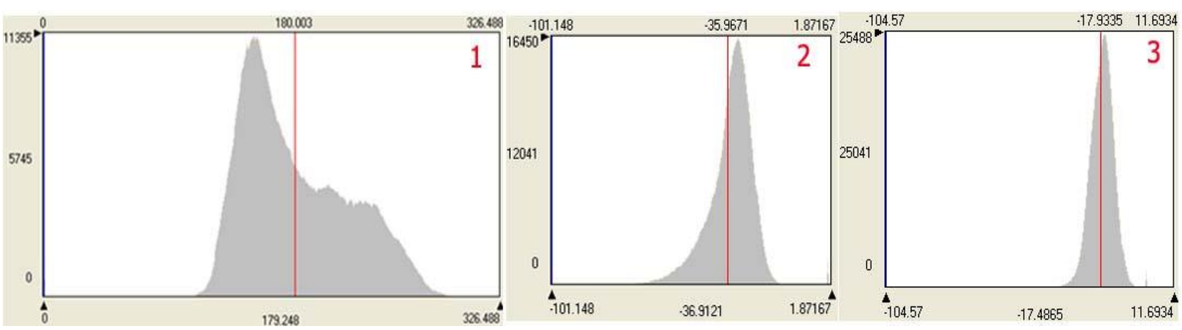


**Figure 5. Histogram showing statistical distribution of DN values in PCA image**

The above chart signifies the concentration of pixels and their respective DN values reaching peak values in one particular range and also in some way akin to k-means clustering, though this approach differs from the other in grouping the pixels. But, eventually such compression of band values with more rotation, evolve with more sharp objects, though with loss of some original information, in the RSI. This would be more towards separation of major objects especially linear objects and thereby assisting extraction of information about the *linear features* from the RSI. Some of the observations as perceived from such out are described below.

A prominent hill structure at the middle of the RSI is shown very clearly in the three PCA output image. The linearity of the structural hill is lucidly displayed sharply against vegetation covering the hill and is aligned continuously. In the present image, vegetation appears grey in PCA1, dark grey in PCA2 with definite shape and pinkish red color in PCA3. Waterbodies that appear dark blue in color in RSI could be clearly identified on the PCA output as grey mass of pixels. The density of grey color depends upon the depth of these waterbodies. It has helped to extract some specific information about the size, shape and volume of these waterbodies. Apart from these features, small river course from the hills that are linear in nature are also brought out. Coarseness of dark tone helped to extract information on the density of vegetation. Thus, PCA has helped to extract certain information about various types of *features* from the selected RSI. Not only this, it has indirectly aided to infer the inherent nature of the terrain *features* such as predominance of hilly terrain with vegetative cover, which may probably be a forest cover, intermittently covered with other vegetation mostly agricultural plantation, beside waterbodies. The size and shape of the waterbodies indicate the storage capacity of the same.

Much of the information that could be extracted from PCA output would require multiband analysis of RSI to identify the spectral pattern and subsequently to categorize *features* of interest. But, grey levels in the output image, because of compression of multiband data of RSI depicted as a single band PCA output, attribute its significance in extracting information of such *features* and lead to the specific inference about the terrain.

## 6. CONCLUSION

In image mining, the data are stored as pictures or images that are used in many applications, especially using remote sensing image (RSI). It is also pretty well understood that image contain many layered *information* about a specific object in its *spectral domain* and *changes* in the specifically selected object in *temporal domain* necessitate handling a large volume of data to extract information of features of a terrain. It involved tedious process of segmenting the image, generating micro-array, indexing and extracting information. An attempt has been made in the present paper depicting the significance of eigen matrix in compressing multi-layered RSI into a single band PCA output. The output thus derived from the selected RSI helped in identifying and extracting information about *features* in the RSI. It also assisted in deriving significant information of such features and an instantaneous appraisal of their pattern and in turn, nature of the terrain. The aggregation of grey levels of pixels, their distribution pattern and frequency helped to extract information about features such as hills, ridges, barren land, vegetation and waterbodies. Present work may further be extended and improved by applying various algorithms such as filtering, edge detection and classification.

## 7. REFERENCE

[1] Zhang, J., Hsu, W., andLee, M.L, “Image mining: Issues, frameworks and techniques,” in *Proceedings of the 2nd International Workshop Multimedia Data Mining*, 2001, pp. 13-20.

[2] Shi-Fei Ding, Zhong –Zhi Shi,Yong Liang , Feng-Xiang Jin, “Information Feature Analysis and Improved Algorithm of PCA,” *Proc. of the 4th International Conference on Machine Learning and Cybernetics*, Guangzhou, Aug 2005, pp 1756-1761 , 18-21.

[3] Burl, M.C. “Mining large image collections,” in R. Grossman, C. Kamath, V.Kumar, and R. Namburu, eds., *Data Mining for Scientific and Engineering Applications*, Kluwer Academic Publishers, New York, 2001, pp. 63-84.

[4] Datcu, M., et al.: Information Mining in Remote Sensing Image Archives: System Concepts. *IEEE Trans. on Geosci. Remote Scensing*, 41 Dec 2003, pp. 2923–2936

[5] Dell’Acqua, F., Gamba, P.: Query-by-Shape in Meteorological Image Archives Using the Point Diffusion Technique. *IEEE Trans. on Geosci. Remote Sensing*, 39 Sep 2001, pp. 1834–1843.

[6] Rasher, S., Zhou, X.: Efficient Update and Retrieval of objects in a multiresolution geospatial database. *Proc. 15th Int. Conf. on scientific and Statistical Database Management*, July 2003, pp.193–201.

[7] Bian, L., Xie, Z.: A spatial dependence approach to retrieving industrial complexes from digital images. *The Professional Geographer*, V56, No.3, 2004, pp. 381–393.

[8] Jensen, J.R, “*Introductory Digital Image Processing: A Remote Sensing Perspective*”, 1996, Prentice-Hall, NJ.

[9] Subramanya A, “Image Compression Technique”, Potentials IEEE, Vol.20, Issue 1, Mar2001, pp 19-23.

[10] Ming Yang and Nikolaos Bourbakis, “An Overview of Lossless Digital Image Compression Techniques,” Circuits and Systems, 2005, 48th Midwest Symposium, Vol. 2 IEEE, pp 1099-1102.