Volume 5, Issue 5, May 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

PCA Classification Technique of Remote Sensing Analysis of Colour Composite Image of Chillika Lagoon, Odisha

S. .K Sahu
P.G. Dept. of Statistics,
Sambalpur University, Sambalpur
Odisha, India

M. B. N. V. Prasad*
Dean, V .I.T.A.M., Berhampur
Mantridi, Odisha
Odisha, India

B. K. Tripathy
Senior Professor. SCSE
VIT University
Tamil Nadu, India

Abstract— ***PCA is an efficient identifier in terms of time and provides a better accuracy in remote sensing image classification. A PCA based system provides the high speed processing with relatively better accuracy. PCA also easily handles a large number of image data due to its capability of reducing data dimensionality and complexity. PCA algorithm provides more accurate remote sensing image classification that infers better and concise results. This scheme is a practical and efficient method for dealing with hyper spectral remote sensing image data. It makes use of the block structure of the correlation matrix so that the PCA is conducted on data of smaller dimensionality. Therefore, computational load is reduced significantly. Here, we propose a fast alternative to iterative PCA that makes it suitable for remote sensing applications while ensuring its theoretical convergence illustrated in the challenging problem of monitoring. . As a novel approach for PCA-style analysis, the proposed PCA has proved to be both efficient and effective in feature extraction and data reduction as shown in the image of Chillika Lagoon, odisha. by using of ERDAS imagine-9.2 version software.***

Keywords— Classification, PCA, remote sensing, eigen value, image data.

I. INTRODUCTION

For feature extraction and dimension reduction in hypercubes, the most widely used approach is Principal Components Analysis (PCA) Dianat and Kasaei, (2010) and Ren et al., (2014) followed by several other approaches such as random projection He and Mei, (2010), singular value decomposition Phillips et al., (2009), maximum noise fraction Green et al., (1988) and singular spectral analysis Zabalza et al., (2014). In fact, PCA uses orthogonal transformation to convert high dimensional data into linearly uncorrelated variables, namely principal components. Often, the number of principal components is significantly reduced in comparison to the original feature dimension, here referring to the number of bands. Consequently, PCA is found to be a powerful tool in feature extraction and data reduction.

Principal Component Analysis (PCA) is a mathematical technique for reducing the dimensionality of a data set. Because digital remote sensing images are numeric, their dimensionality can be reduced using this technique. In multiband remote sensing images, the bands are the original variables. Some of the original bands may be highly correlated and to save on data storage space and computing time is less and also correlated Eigen images by PCA. In addition to its use in this way, PCA can be used as a change detection technique in remote sensing. Principally, there are two ways PCA can be used in change detection.

- Independent data transformation analysis - in which multi temporal remote sensing image data sets are spectrally, enhanced separately using PCA. Each image is separately classified for use in post classification change detection
- Merged data transformation - in which all the bands from the n - dimensional multi-temporal image data set are registered and treated as a single *N* - dimensional data set as input to the PCA.

Principal Component Analysis (PCA) is a statistical technique used to reduce a set of correlated multivariate measurements to a smaller set where the features are uncorrelated to each other. The advent of satellite remote sensing with multispectral and hyper spectral images in digital format has brought a new dimension in inventory, mapping and monitoring natural resources of the earth. The multispectral (or multi band) images have been acquired in different parts of the electromagnetic spectrum retaining correlation between the bands. The innovative techniques of PCA have been incorporated as a special transformation in digital image processing of satellite images where N numbers of correlated bands of the image data have been reduced to few uncorrelated bands. For example the LANDSAT satellite system provides seven band image data from which six bands are reduced to 3 bands using PCA and thereafter these three bands are used to create a false color composite where the visual interpretation for ground features is highly enhanced. The enhanced image is effectively used as a base for ground truth collection in supervised classification of land use and land cover, performing special tasks such as geologic interpretation etc. In this paper an attempt has been made to highlight the significance of PCA in processing of remote sensing images on the basis of evaluation process. However, in many remote sensing applications acquiring ground truth information for all classes is very difficult, especially complex and heterogeneous geographical areas are analyzed. Actually, many other applications have turned to recognize one specific land-cover class of interest and to discriminate it from the other classes present in the investigated area. This formulation of the problem relaxes the constraint of having an exhaustive training set, but requires the availability of representative training data for the analyzed class and if possible, some training samples representative of other classes considered. Recently, high interest has been played to this approach through the fields of:

- Anomaly Detection - where one tries to identify pixels differing significantly from the background.
- Target Detection - where the target spectral signature is assumed to be known (or available from spectral libraries) and the goal is to detect pixels that match the target.
- One-class or Multi-class Classification - where one tries to detect one class or extend to multi class and others.

II. GRAPHICAL REPRESENTATION OF PCA

This paper is organized in the following four distinct sectors. The background and Pre-processing sector introduces the methodological approach used and it also provides a description of data orthorectification. The sample data set sector points out the necessary procedures to obtain a sufficient number of training samples that are prerequisites for a successful classification. Finally accuracy assessment sector to evaluate the different PC bands of results. The graphical representation of PCA is shown in the figure.1.

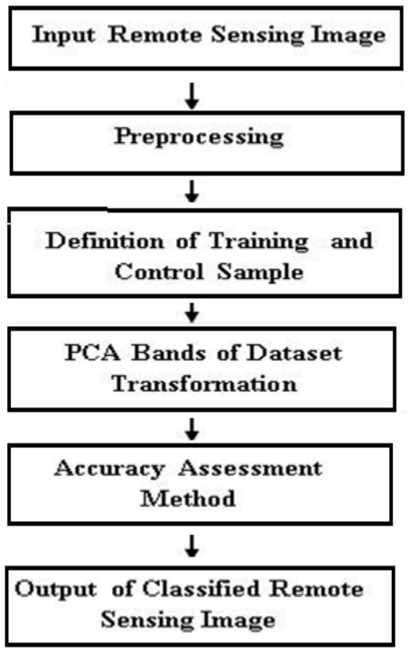

Figure .1**:** Graphical representation of the classification of remote sensing images using the PCA

Multispectral remote sensing PCA classification is a complex process of sorting pixels bands into a finite number of individual classes or categories of data. If a pixel band satisfies, a certain set of criteria, the pixel is assigned to the class that corresponds to those criteria. Thus an appropriate classification system and an adequate number of training samples are fundamentals for a successful classification. The schematic diagram of multistage PCA classification is shown in the figure 2:

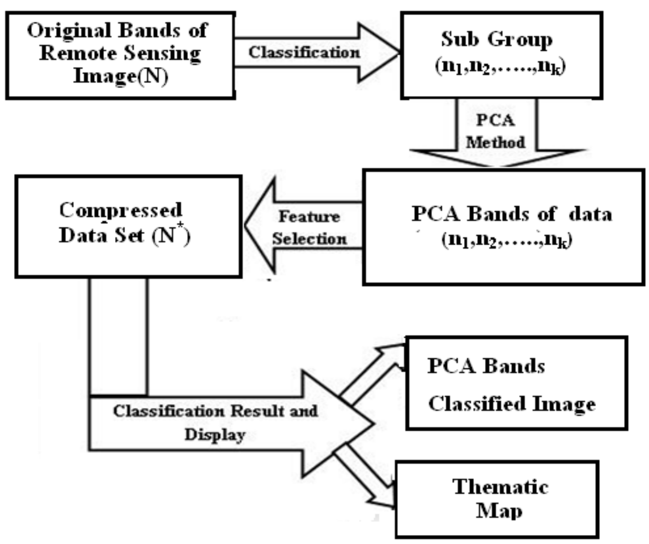

Figure .2**:** Schematic Diagram of Multistage PCA Classification

III. PRINCIPAL COMPONENT ANALYSIS (PCA) IN STATISTICAL CONCEPTS

The basic idea of the PCA was first proposed by Pearson . The concept of PCA is to convert the data sets of observation in which the variables may be correlated with each other into another sets of variables which are uncorrelated and called principal components. The mathematical procedure is finding an orthogonal matrix to transform the original data sets into another coordinate system where the principal components are located, such that the first principal component has the highest variance, and the subsequent components in turn have as high variance as possible. The principal components in the new coordinate system are theoretically uncorrelated with (orthogonal to) each other. The operation of PCA can reflect the internal structure of data sets in the way of most possible variability in the data. Therefore, it can provide us with a lower dimensional picture to characterize the multivariate data sets in a way of informative viewpoint when the data sets are observed in a higher dimensional space. For the purpose of simplifying the process of multivariate data analysis, only the first few principal components (more informative) are extracted for analysis without losing the major natures of the data sets, and thus the dimensionality of the data set is reduced.

In order to find the linear orthogonal transformation matrix, let the $n \times p$ matrix represent the n sets of data in which each data set consists of p variables,

$$X = \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ X_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & . & x_{1p} \\ x_{21} & x_{22} & . & x_{2p} \\ . & . & . & . \\ . & . & . & . \\ x_{n1} & x_{n2} & . & x_{np} \end{bmatrix}. \quad (1)$$

It would be difficult to observe and analyze the characteristics of the data sets if the variable number p is large and the p -dimensional space that the data sets form is complicated. Therefore, it is definitely beneficial if the complicated data sets are transformed into another space and thus they can reveal a few informative indicators through the linear combination.

Let $[x_1, x_2, \ldots\ldots, x_p]^T$ represent the data set of p variables, and Y is the vector of m principal components $(m \le p)$ through the linear transformation, that is,

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_m \end{bmatrix} = \begin{bmatrix} l_{11} & l_{12} & . & l_{1p} \\ l_{21} & l_{22} & . & l_{2p} \\ . & . & . & . \\ . & . & . & . \\ l_{n1} & l_{n2} & . & l_{np} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ x_p \end{bmatrix} = Lx, \quad (2)$$

where L is called projection matrix, and lmp is the loading of the principal component representing the weight of the $p-th$ variable projected onto the $m-th$ principal component. It is noted apparently that the values of the principal components are decided by the determination of the projection matrix L. In order to find the projection matrix, the correlation coefficient matrix of X (Eq. (2)) is first computed to measure the quantities and directions of correlation among the variables. The Pearson correlation coefficient is broadly utilized and formulated as

$$R = \begin{bmatrix} r_{11} & r_{12} & . & r_{1p} \\ r_{21} & r_{22} & . & r_{2p} \\ . & . & . & . \\ . & . & . & . \\ r_{n1} & r_{n2} & . & r_{pp} \end{bmatrix}, \quad (3)$$

where r_{ij} represents the correlation coefficient between x_i and x_j, and is formulated as

$$r_{ij} = \frac{\sum_{k=1}^{n}(x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{n}(x_{ki} - \bar{x}_i)^2 \sum_{k=1}^{n}(x_{kj} - \bar{x}_j)^2}}, \bar{x}_i = \frac{1}{n}\sum_{k=1}^{n} x_{ki}. \quad (4)$$

Conceptually, the eigen values of the correlation coefficient matrix R represent the variance of the variables. It is also known that the total quantity of variance is invariant for both the original observed data and the transformed variables in the new coordinate system. Based on this concept, the eigen values and their corresponding eigenvectors of the symmetric matrix R are calculated to describe the variance distribution of the variables (principal components) in the new coordinate system. According to the quantities, the eigen values of R can be sorted as $\lambda_1 \ge \lambda_2 \ge \ldots\ldots \ge \lambda_p \ge 0$, and $e_1, e_2, \ldots\ldots, e_p$ represent their corresponding eigenvectors. The individual contribution rate of the i-th principal component C_i is defined as

$$C_i = \frac{\lambda_i}{\sum_{k=1}^{p} \lambda_k}, \quad (5)$$

and the accumulative contribution rate of the first m principal components is formulated as

$$D_i = \frac{\sum_{k=1}^{m} \lambda_k}{\sum_{k=1}^{p} \lambda_k}, m = 1,2,...p. \quad (6)$$

The individual contribution rate of the i-th principal component represents its significance for characterizing the data sets. The accumulative contribution rate reveals the completeness of information that the m principal components can provide for characterizing the data sets. Therefore, the purpose of selecting the principal component number, m (usually $m < p$) is to reserve enough information representing the features of date sets, and to achieve the dimensionality reduction of the data sets concurrently.

The loading l_{jk} in Eq. (7.7) represents the contribution weight from the k-th variable in the original space of observation to the j-th principal component in the new coordinate system after the transformation. The loading l_{jk} is thus calculated in terms of its corresponding eigenvalue and eigenvector , that is,

$$l_{jk} = \sqrt{\lambda_j e_{jk}}, \quad (7)$$

where e_{jk} is the k-th component of eigenvector e_j. Once the number of principal components m is selected, the score of the k-th principal component in the j-th data set can be determined as

$$y_{jk} = l_{k1}x_{j1} + l_{k2}x_{j2} + ... + l_{kp}x_{jp}, \quad (8)$$

and the score matrix of principal components is formed to indicate the scores of m principal components in all the n sets of data, that is,

$$\bar{Y} == \begin{bmatrix} y_{11} & y_{12} & . & y_{1p} \\ y_{21} & y_{22} & . & y_{2p} \\ . & . & . & . \\ . & . & . & . \\ y_{n1} & y_{n2} & . & y_{nm} \end{bmatrix}, \quad (9)$$

Using the principal component scores of all data sets that are transformed from the features in the original measured data sets, the operation conditions of the gear transmission system can be diagnosed with reduced data dimensionality.

IV. PCA ALGORITHM

Step 1: Get some data Let $X_i(j)$ represents the response. where i = (1, 2, …, m) where m is the number of experiments performed. j = (1, 2, …, n) where n is the number of quality characteristics.

Step 2: Normalization of the quality characteristics Equation (10) is used to normalize surface roughness as surface roughness is to be minimized. Equation (11) is used to normalize material removal rate as MRR is to be maximized.Lower-the-better.

$$x_{i,k}^* = \frac{\min x_i(k)}{x_i(k)} \quad (10)$$

Higher the better.

$$x_{i,k}^* = \frac{x_i(k)}{\max x_i(k)} \quad (11)$$

Where, $x_{i,k}^*$ represent the normalized data of the ith experiment and kth response.

Step 3: Calculate the covariance matrix

The normalized data is utilized to construct a variance covariance matrix R, which is illustrated as below:

$$R = \begin{bmatrix} R_{1,1} & R_{1,2} & . & R_{1,p} \\ R_{2,1} & R_{2,2} & . & R_{2,p} \\ . & . & . & . \\ . & . & . & . \\ R_{m,1} & R_{m,2} & . & R_{m,n} \end{bmatrix} \quad (12)$$

where $R_{ij} = Covx_i^*(j), x_i^*(k) / \sigma x_i^*(j)^* x_i^*(k)\theta$.

$Cov\left(x_i^*(j), x_i^*(k)\right)$ is the covariance of sequences $x_i^*(j)$ and $x_i^*(k)$;

$\sigma x_i^*(j)$ and $\sigma x_i^*(k)$ are the standard deviation of sequences $x_i^*(j)$ and $x_i^*(k)$.

Step 4: Calculate the eigen vectors and eigen value of the covariance matrix
Step 5: Form the feature vector: Choose the eigen vectors with the large eigen values. To form feature vector, arrange the eigen values from highest to lowest. Ignore the components of lowest significance. Final data set will have fewer dimensions. If we choose the first p eigen vectors and final data will have p dimensions.
Step 6: Derive the new data set.

PCA can be used to reduce the dimensionality of the data while minimizing loss of information. In addition, PCA images may be more easily interpreted than the conventional colour infrared composite. The PCA of the ETM+ datasets can distinguish the iron rich areas. In the present study, the PCA was followed for information in six bands viz., 1, 2, 3, 4, 5 and 6.

V. EXPERIMENTAL RESULTS

The study area (Geological setup of the area), covering the Chillika Lagoon and its surroundings in Odisha, India, and lying between latitudes 19° 0' 0''N and 20°N, and longitudes 84° 50' 0''E and 85°40'0''E has been extensively surveyed using ground-based geological techniques, principal component analysis. Algorithm which trains the dataset by using of ERDAS imagine-9.2 version software. False color composition image of Chillika Lagoon,Odisha and we get the PCA image in figure.4 and PCA classification in figure.5.The area in hectors classification shown in figure.6.

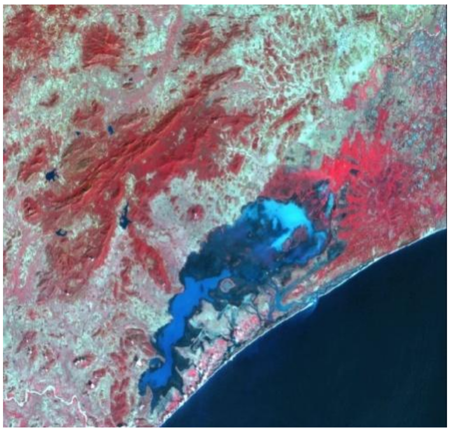

Figure.3. False color composition image of Chillika Lagoon, odisha

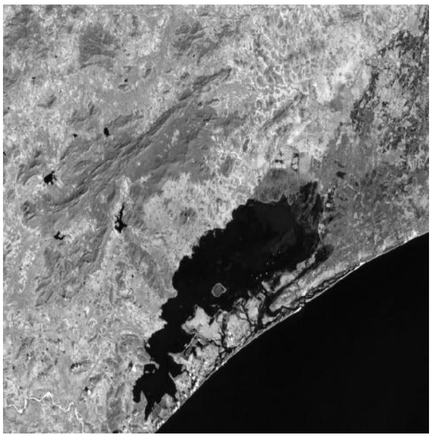

Figure.4 PCA Image

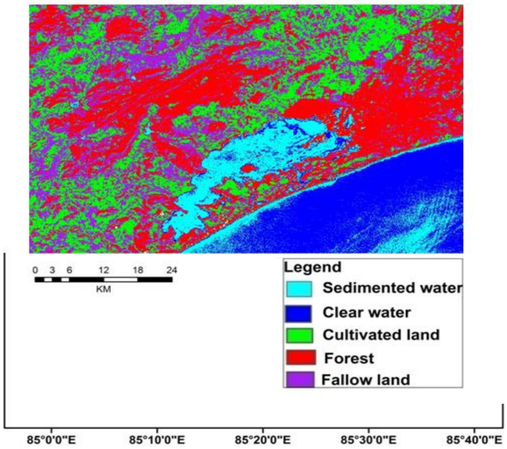

Figure 5. PCA Classification

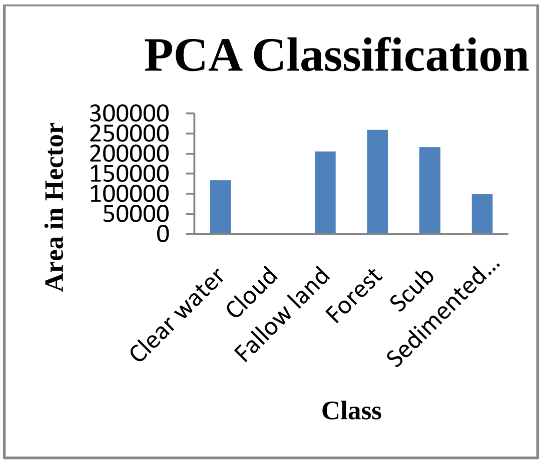

Figure .6. distribution of PCA Classification

VI. CONCLUSION

PCA is an efficient identifier in terms of time and provides a better accuracy in remote sensing image classification. A PCA based system provides the high speed processing with relatively better accuracy. PCA also easily handles a large number of image data due to its capability of reducing data dimensionality and complexity. PCA algorithm provides more accurate remote sensing image classification that infers better and concise results. This scheme is a practical and efficient method for dealing with hyper spectral remote sensing image data. It makes use of the block structure of the correlation matrix so that the PCA is conducted on data of smaller dimensionality. Therefore, computational load is reduced significantly. As a novel approach for PCA-style analysis, the proposed PCA has proved to be both efficient and effective in feature extraction and data reduction as shown in fig .4 of image of Chillika Lagoon, odisha.

REFERENCES

[1] Dianat, R., Kasaei, S., (2010) Dimension reduction of optical remote sensing images via minimum change rate deviation method,IEEE Trans. Geosci. Remote Sens., 48 (1) (198–206

[2] Ren, J., Zabalza, J., Marshall, S., Zheng, J., (2014)Effective feature extraction and data reduction in remote sensing using hyperspectral imaging,IEEE Signal Proc. Mag., 31 (7)

[3] He, M., Mei, S., (2010). Dimension reduction by random projection for endmember extraction. In: Proc. IEEE Conf. Ind. Electron. Appl., pp. 2323–2327

[4] Phillips, R.D., Watson, W.T., Wynne, R.H., Blinn, C.E., (2009),Feature reduction using a singular value decomposition for the iterative guided spectral class rejection hybrid classifier.,ISPRS J. Photogramm. Remote Sensing, 64 (1) 107–116.

[5] Green, A., Berman, M., Switzer, P., Craig, M.,(1988).A transformation for ordering multispectral data in terms of image quality with implications for noise removal. IEEE Trans. Geosci. Remote Sens., 26 (1) . 65–74

[6] Zabalza, J., Clemente, C., Di Caterina, G., Ren, J., Soraghan, J.J., Marshall, S., (2014). Robust PCA micro-doppler classification using svm on embedded systems. IEEE Trans. Aerospace Electr. Syst.