Abstract: Principal Component Analysis (PCA) technique is useful in reducing dimensionality of a data set in order to obtain a simple dataset where characteristics of the original dataset that contributes most to its variance are retained. Thus this method is to transform the original data set into a new dataset, which may better capture the essential information. Multi-spectral (or multi-band) images from orbiting satellites are gaining ground in recent years in inventory, mapping and monitoring of earth resources. These images are acquired in different wavelengths of the electromagnetic spectrum and therefore there exist correlation between the bands. The technique of PCA has been incorporated in digital image processing and found useful in removing or reducing such redundancy in the multi-spectral image data. In this paper the usefulness of PCA in processing of multispectral satellite images have been highlighted. It has been observed 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 quality of ground truth collection and in turn improving the image classification accuracy.

Index Terms: Principal Component Analysis (PCA), Multispectral Images, Satellites, Electromagnetic spectrum, Separability, Image classification.

INTRODUCTION

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 hyperspectral 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 thus retaining correlation amongst bands. The techniques of PCA have been incorporated as a special transformation in digital image processing of satellite images where a number 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 Principal Component Analysis in processing of satellite images on the basis of review of literature.

REVIEW OF PCA IN SATELLITE IMAGE ANALYSIS

The technique of Principal Component Analysis (PCA) has found wide use in digital processing of multispectral satellite images. PCA and image fusion techniques are often used to enhance an image particularly in the land-cover classification of satellite images, where such images are used for increasing the interpretability of human observers and for improving the accuracy of the classification. In the field of remote sensing, especially in hyper spectral imagery, reduction of the dimensionality is a key point for data analysis to prevent from Hughes phenomenon. A novel image fusion scheme is proposed for multispectral and panchromatic satellite images using Principal Component Analysis (PCA) and Non sub sampled Contour let Transform (NSCT) [1]. Experimental results show that the fusion scheme can effectively preserve spectral information while improving the spatial quality. Numerous satellites collect correlated data from earth surface with different spatial, spectral and temporal resolutions. Given the design constraints sensors with high spectral resolution, do not have an optimal spatial resolution, and vice versa. For optimal use of these data the fusion of multispectral and panchromatic images is becoming a promising technique to obtain images with high spatial and spectral resolution simultaneously.

Curvelet-based image fusion has been used to merge multispectral and panchromatic images of Landsat ETM where simultaneously rich information in the spatial and spectral domains is simultaneously obtained [2]. In the land-cover classification of satellite images, a well-fused image is a tool for increasing the interpretability of human observers and for improving the accuracy of the classification. Other methods include the Brovey method, the intensity–hue–saturation (IHS) color model, and principal component analysis (PCA). However, a limitation of these methods is that some distortion occurs in the spectral characteristics of the original multispectral images. Recently, developments in wavelet analysis have provided a potential solution to this problem where the fusion of a high-resolution
Pan image with a low-resolution multispectral image on the basis of an un-decimated à trous wavelet transform has been developed. Recently, other multi-resolution analyses have been developed, including ridgelets and curvelets. These approaches are very different from wavelet-like systems. Curvelets and ridgelets take the form of basic elements, which exhibit high directional sensitivity and are highly anisotropic.

A conventional principal component analysis is built on the basis of a statistical concept. The principal component analysis is to summarize data defined by many variants or various features into a few of principal components, so as that the meaning of the data can be lost as little as possible and interpreted as well. A fuzzy concept can be employed to construct a principal component model which can deal with fuzziness, vagueness or possibility of system considered, which is named a fuzzy principal component analysis for fuzzy data [3]. The fuzzy data is employed to deal with the possibility of the vague system, and to analyze a real state of the latent system under the data. As observed data should embody possibilities which the considered system has, the measured data can be interpreted using a possibility of the system. Therefore, the fuzzy principal component analysis for fuzzy data is built in terms of the possibility and evaluates all observed values as a possibility which the system should have.

Principal component analysis (PCA) is deployed in JPEG2000 to provide spectral de-correlation as well as spectral dimensionality reduction[4]. This scheme is evaluated in terms of rate-distortion performance as well as in terms of information preservation in an anomaly-detection task. Additionally, this scheme is also compared to the common approach of JPEG2000 coupled with a wavelet transform for spectral de-correlation. The results reveal that, not only does the proposed PCA-based coder yield rate distortion and information preservation performance superior to that of the wavelet-based coder, the best PCA performance occurs when a reduced number of PCs are retained and coded. A linear model to estimate the optimal number of PCs to use in such dimensionality reduction.

Remote sensing produces large amounts of digital data which are collected into databases. Since a variety of applications utilize the multispectral data, the data cannot be compressed with lossy methods. The combination of two reversible methods for the lossless compression of the multispectral images: first, the principal component analysis is applied to the spectra of the image and then, the integer wavelet transform is applied to the residual image to further concentrate the energy and reduce the entropy [5]. The coding quality of the method is measured with the the zero-order entropy, and it is clearly lower with this method than with the other methods. Depending on the AVIRIS image, the entropies varied from 5.6 to 5.9 bits per pixel. With the same images, the actual compression ratios, calculated from the files sizes, were in the range from 2.8 to 2.9.

The study reveals that the lossless compression techniques that capture the energy of the image to a low number of coefficients results to a low entropy. The computation of the residual image through PCA is much faster than that through the vector quantization, which is typically very time consuming. The wavelet transform and the arithmetic coding were more time consuming, around 100 secs for each.

Feature extraction of hyper-spectral remote sensing data can be investigated using Principal component analysis (PCA) which has shown to be a good unsupervised feature extraction. On the other hand, this method only focuses on second orders statistics. By mapping the data onto another feature space and using nonlinear function, Kernel PCA (KPCA) can extract higher order statistics [6]. Using kernel methods, all computation are done in the original space, thus saving computing time. KPCA is used for the preprocessing step to extract relevant feature for classification and to prevent from the Hughes phenomenon & then the classification is done with a back-propagation neural network on real hyper-spectral ROSIS data from urban area. An unsupervised nonlinear feature extraction method was investigated. Based on kernel methods, linear PCA was turn to nonlinear KPCA. This method was used to extract features that are uncorrelated in some feature space. Basically KPCA is used as feature extraction on hyper-spectral data, which performed well in terms of accuracy.

A nonlinear approach based on a combination of the fuzzy -means clustering (FCMC) [7], feature vector selection and principal component analysis (PCA) to extract features of multispectral images when a very large number of samples need to be processed. A preprocessing method for classifying these images with higher accuracy compared to the single PCA and kernel PCA has been found. A nonlinear approach based on a combination of three key methods (i.e., the fuzzy -means clustering, feature vector selection, and principal component analysis) to extract the features in the nonlinear space for classifying multispectral images. Also, a good strategy is given for selecting the input samples for the FVS so that the samples chosen can better represent the cluster that they belong to. The results
not only improve the classification accuracy greatly but also significantly reduce the computational time compared to the single KPCA when a large number of samples are considered.

Texture information offers an extensive solution for image classification by providing better accuracy of image information. However, huge amounts of improper additional texture information may result in a chaotic state, and this leads to uncertainty in the classification process. The PCA method has flaws in the area of influenced and non-influenced attributes. On the whole, whether PCA provides an effective solution to determine the value of knowledge rule in image information still remains a question. Here an innovative method, called Discrete Rough Set method [8], is used as a tool for image classification. This method focuses on two crucial issues:

1. The core attributes of the target categories in image classification are systematically analyzed while eliminating surplus attributes rationally;

2. The unique point of each attribute, which influenced the target categories, is successfully found. This is a crucial aspect that is very helpful for the construction of decision rule.

Here the overall accuracy of Discrete Rough Set (96.67%) exceeds that of the conventional PCA (86.00%) of paddy rice area evaluation from Quickbird image. This shows that the appropriate classification knowledge can be presented by Discrete Rough Set, and the information can effectively improve the accuracy of image classification. The researchers generally encountered difficulties in finding a solution for effectively enhancing the image classification accuracy. Most of the researchers would like to append texture information as the preliminary process of the improvements. Unfortunately, the add-in texture information is not always necessary for classification. To develop a systematic analysis of evaluating the add-in texture information, this method provides an innovative solution of selecting appropriate terms of texture information. Once the suitable add-in texture information is selected, the accuracy of target categories is enhanced significantly. To display the advantage of Discrete Rough Sets, the image classifications of paddy rice are used as an example for demonstration. Furthermore, the well-known method of conventional PCA is also used for comparison purposes. The overall accuracy of Discrete Rough Set + RSKC (96.67%) is better than the conventional PCA + RSKC (86.00%). The Rough Set method offers a great improvement by reducing the commission errors from 13.67% to 1.67%. In addition, two main contributions are also obtained from the Discrete Rough Set method. That is, the separate points and core factors are also found for the target categories. It is believed that the concept of separate points and core factors is very crucial for the related research.

Remote sensing data have become very widespread in recent years, and the exploitation of this technology has gone from developments mainly conducted by government intelligence agencies to those carried out by general users and companies. There is a great deal more to remote sensing data than meets the eye, and extracting that information turns out to be a major computational challenge. For this purpose, high performance computing (HPC) infrastructure such as clusters, distributed networks or specialized hardware devices provide important architectural developments to accelerate the computations related with information extraction in remote sensing. Here the recent advances in HPC applied to remote sensing problems; in particular, the HPC-based paradigms included comprises of multiprocessor systems, large-scale and heterogeneous networks of computers, grid and cloud computing environments, and hardware systems such as field programmable gate arrays (FPGAs) and graphics processing units (GPUs) [9]. Combining these parts deliver a snapshot of the state-of-the-art and most recent developments in those areas and also offer a thoughtful perspective of the potential and emerging challenges.

Fig. 1. The GEO grid field observation network virtual organization.

GEO Grid: The goal of GEO Grid is to provide a disaster assessment capability, and could be considered a prototype for an operational disaster monitoring system. GEO Grid integrates grid technology to securely manage federated resources with standard geospatial tools for a variety of
applications focusing on utilization of various remote sensing data sources. GEO Grid is an on-going project funded by the Japanese government and built by the Grid Technology Research Center at the National Institute of Advanced Information Science Technology. Japan GEO Gridingests both ASTER and MODIS data, storing this data using the Gfarm data grid middleware to achieve the desired scalability and distribution. Like many other systems, GEO Grid can be accessed through a portal. However, GEO Grid offers both a portal development kit (PDK) and a service development kit (SDK). The GEO Grid PDK enables users to build customized portals from a library of portlets that include workflow engines, data access tools, and OGC web services. The GEO Grid SDK enables users to create their own services that can be registered and shared with other users and sites.

Here the review of the state-of-the-art in the application of HPC techniques and application of this HPC technique on remote sensing problems is being discussed. Techniques discussed include specialized hardware devices, multi-processor systems and distributed networks, which provide important architectural developments to accelerate the computations related with information extraction in remote sensing. The study of this paper reveals that specialized hardware systems can satisfy the time-critical constraints introduced by several remote sensing applications, while the computational power offered by clusters and distributed networks is ready to introduce substantial benefits from the viewpoint of integrating available computing resources and exploiting large volumes of remotely sensed data.

**CONCLUSION**

Application of PCA on image time series can effectively filter compositing noise when data are reduced in a small number of components. It is an important tool for multi-temporal analysis. PCA of a high temporal resolution image series can attenuate temporal autocorrelation, thereby increasing the suitability of the data for image segmentation and classification procedures. Moreover, principal components are intrinsic interest because they effectively summarize the dominant modes of spatial, spectral & temporal variation in data in terms of linear combinations of image frames. PCA has been widely used in pattern recognition and remote sensing application, mathematically establishes a new set of variables, which describe the variance in the original data set. The principal components is useful in providing maximum visual separability of image feature. Therefore, principal component analysis can be used in image classification to improve the accuracy.

**REFERENCES**


