Feature Extraction Approach in Hyperspectral Data

Munkh-Erdene Altangerel[1,*], Amarsaikhan Damdinsuren[1], Enkhjargal Damdinsuren[1], Odontuya Gendaram[2], Jargaldalai Enkhtuya[1]

[1]*Institute of Geography and Geoecology, Mongolian Academy of Sciences, Ulaanbaatar, Mongolia*
[2]*Mongolian University of Pharmaceutical Sciences, Ulaanbaatar, Mongolia*
[*]*Corresponding author. Email: munkherdenea@mas.ac.mn*

ABSTRACT

In general, feature extraction deals with the problem of finding the most informative, distinctive, and reduced set of features and improve the success of data processing. The features should contain information required to distinguish between classes, be insensitive to irrelevant variability in the input, and also be limited in number, to permit, efficient computation of the applied functions and to limit the amount of data required. In many cases, it is an important step in the solutions of many tasks aiming to extract the relevant information from the available large datasets. The aim of this study is to apply a feature extraction approach to a hyperspectral image and extract different features from the dataset and reduce its dimensionality into meaningful orthogonal features. The final analysis was performed in a test site situated in central Mongolia using 242 band Hyperion data. Overall, the study indicated that the Hyperion hyperspectral data could be effectively reduced into meaningful features through a feature extraction process.

Keywords: Feature extraction, hyperspectral data, principal component, meaningful features

1. INTRODUCTION

At present, feature extraction of hyperspectral data has a great interest for researchers dealing with remote sensing (RS) digital image processing [1]. In the field of image processing, it can be performed by extracting features for identification, classification, diagnosis, classification, clustering, recognition, and detection [2]. Feature extraction is a part of the dimensionality reduction process, in which, an initial set of the raw data is divided and reduced to more manageable groups. The most important characteristic of these large datasets is that they have a large number of variables. These variables require a lot of computing resources to process them. Therefore, feature extraction helps the experts in receiving the best feature from those large datasets to select and combine variables into features, thus, effectively reducing the amount of data. These features are easy to process, but still able to describe the actual data set with accuracy and originality. In the end, the reduction of the data helps to build the model with less machine's effort and also increases the speed of learning and generalization steps in the machine learning process [3].

The feature is defined as a function of the basic measurement variables or attributes that specify some quantifiable property of an object and are useful for pattern recognition. Obtaining a good data representation is a very domain-specific task and it is related to the available measurements [4]. Many consider the features as patterns of an object in an image that help to identify it. Features include properties like corners, edges, regions of interest points, ridges, and so on [5]. The feature plays a very important role in the area of image processing. Before getting features, various image preprocessing techniques like binarization, thresholding, resizing, and normalization can be applied to the sampled image. After that, feature extraction techniques are applied to get features that will be useful in classifying and recognition of images. Feature extraction techniques are helpful in various image processing applications, including character recognition. As features define the behavior of an image, they show its place in terms of storage taken, efficiency in classification, and obviously in time consumption also [6].

Generally, feature extraction is done after the preprocessing phase in the character recognition

system. The primary task of pattern recognition is to take an input pattern and correctly assign it as one of the possible output classes. This process can be divided into two general stages such as feature selection and classification. Feature selection is critical to the whole process since the classifier will not be able to recognize poorly selected features. The features should contain the information required to distinguish between classes, be insensitive to irrelevant variability in the input, and also be limited in number, to permit, efficient computation of discriminant functions and to limit the amount of data required. In many cases, it is an important step in the construction of any pattern classification and aiming to extract the relevant information that characterizes each class. In this process, relevant features are extracted from objects to form feature vectors. These feature vectors are then used by classifiers to recognize the input unit with target output unit [7].

Recent studies have found that feature extraction deals with the problem of finding the most informative, distinctive, and reduced set of features, to improve the success of data storage and processing. Important feature vectors remain the most common and suitable representation for classification problems. The current advances in both data analysis and machine learning fields made it possible to create a recognition system, which can achieve tasks that could not be accomplished in the past and feature extraction lies at the center of these advancements with applications in data analysis. In feature extraction, the researchers are concerned about finding a new set of dimensions, which are combinations of the original dimensions. The widely known and most commonly utilized feature extraction methods are principal component analysis (PCA) and linear discriminant analysis, unsupervised and supervised learning [8]. The PCA is considerably similar to two other unsupervised linear methods, factor analysis and multidimensional scaling [9].

In the present study, we wanted to apply a feature extraction approach and extract different features from hyperspectral data and reduce its dimensionality into meaningful orthogonal features. As the feature extraction technique, the PCA has been selected. The final analysis was performed using 242 band Hyperion data of central Mongolia.

2. TEST SITE

In the present study, a site situated to the west of Ulaanbaatar city has been selected. The selected site is not very far from the western fringe of the capital city covers. It is extended from the west to the east

about 7.53 km and from the north to the south about 13.68 km. In the selected test area, such land cover classes as settlement, green vegetation, Salix, soil, bare land, the agricultural field, and water are available.

3. DATASETS AND METHODOLOGY

As RS data, the Hyperion dataset of 04 June 2014 has been used. The Hyperion is a high resolution hyperspectral imaging instrument that was carried out on the EO-1 mission. It is the first satellite in NASA's New Millennium Program Earth Observing series and can image the Earth's surface in 242 adjacent spectral bands with high radiometric accuracy, covering the region from 0.4 μm to 2.5 μm, with a spatial resolution of 30 m. Through this large number of spectral bands, different land and water surface features can be accurately imaged [10]. In addition, for identification of some ground truth information, a topographic map of scale 1:100,000 produced in 1984 for used. The selected test area shown in the Hyperion image is illustrated in Figure 1.
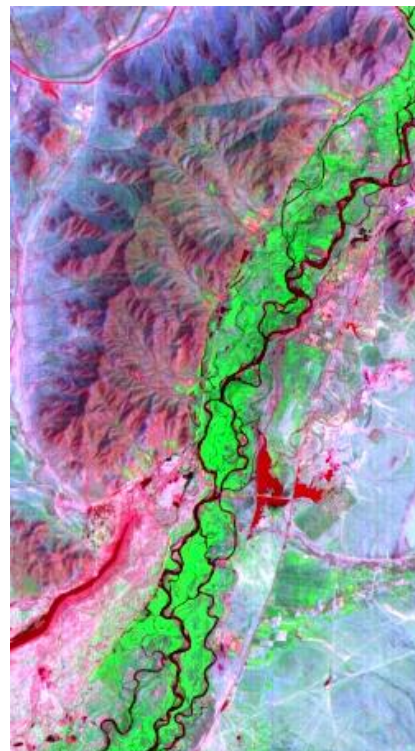

Figure 1. Hyperion image of the test area.

The most common understanding of the PCA is that it is a data compression technique used to reduce the dimensionality of the multidimensional datasets or bands. The bands of the PCA data are noncorrelated and are often more interpretable than the source data. The process is easily explained if we

consider a two dimensional histogram that forms an ellipse [11]. When the PCA is performed, the axes of the spectral space are rotated, changing the coordinates of each pixel in spectral space. The new axes are parallel to the axes of the ellipse. The length and direction of the widest transect of the ellipse are calculated using a matrix algebra. The transect, which corresponds to the major axis of the ellipse, is called the first principal component of the data [8]. The direction of the first principal component is the first eigenvector, and its length is the first eigenvalue. A new axis of the spectral space is defined by this first principal component. The second principal component is the widest transect of the ellipse that is perpendicular to the first principal component. As such, the second principal component describes the largest amount of variance in the data that is not already described by the first principal component. In a two-dimensional case, the second principal component corresponds to the minor axis of the ellipse [12].

In *n* dimensions, there are *n* principal components. Each successive principal component is the widest transect of the ellipse that is orthogonal to the previous components in the *n*-dimensional space and accounts for a decreasing amount of the variation in the data which is not already accounted for by previous principal components. Although there are *n* output bands in a PCA, the first few bands account for a high proportion of the variance in the data. Sometimes, useful information can be gathered from the principal component bands with the least variances and these bands can show subtle details in the image that were obscured by higher contrast in the original image [13].

4. RESULTS AND DISCUSSION

In the beginning, the Hyperion bands have been analyzed for radiometric quality. From the analysis, it was seen that the water absorption bands and some other bands of the hyperspectral image had zero values. When these bands have been excluded, the original dataset was reduced from 242 bands to 169 bands.

After defining the final dataset, a PCA has been applied to the selected bands. From the final PCA, we have selected the first 10 PCs, because they contained a significantly high percentage of the overall variance. The result of the PCA is shown in Table 1 and PC1 is illustrated in Figure 2. As can be seen from Table 1, a PC1 includes 64% of the overall variance having the most compact SD values compared to the rest of PCs. Moreover, it is seen

from Figure 2 that, as it contained the majority of the variance, all available land cover classes are clearly seen.

As seen from Table 1, a PC2 contains 23% of the overall variance having the second most compact SD values compared to the other PCs. In addition, it is seen from Figure 3 that, it includes a high amount of the variance of the available land cover classes. Unlike the case of PC1, in PC2, Salix and green vegetation have very high negative loadings. Therefore, these 2 classes appear to vary dark in the PC2 image. Furthermore, it is seen from Table 1 that the PC2 has a higher minimum and maximum values compared to the PC1. It could be explained by the fact the PC2 image has a brighter appearance especially, in the mountainous areas, delineated in the upper left part of the image.

Figure 2. PC1 image of the test area.

A PC3 image obtained from the PCA is shown in Figure 4. Although the PC3 contains 5.4% of the overall variance, it still clearly delineates all available land cover classes such as water, barren soil, croplands, roads, willows, and grasslands. As seen from Table 1, it has more compact SD values compared to the other PCs and also higher minimum and maximum values in both PC1 and PC2. The mountainous areas in the PC3 have very high loading. As a result, these areas have a bright appearance in the PC3 image. Likewise, Salix and green vegetation have relatively brighter appearances having more higher loadings compared to the other PCs.

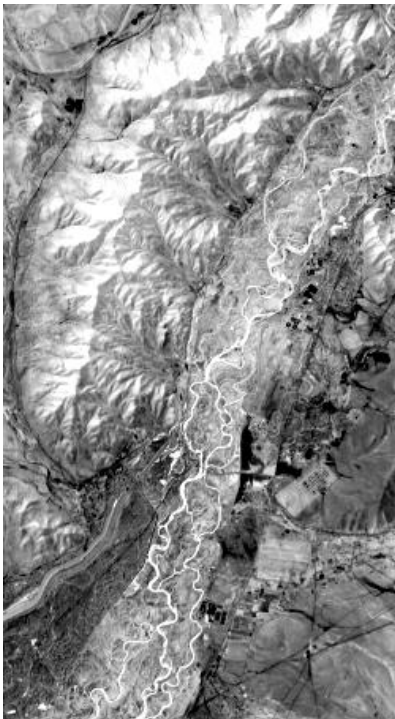**Figure 3.** PC2 image of the test area

Figure 4. PC3 image of the test area.

A PC4 image of the PCA is shown in Figure 5. As seen from Table 1, the PC4 contains 1.8% of the overall variance and has the third most compact SD values compared to the remaining PCs. Moreover, it is seen from Figure 3 that, although it includes an

insignificant amount of the variance, still contains some useful information about water and other classes. Furthermore, it is seen from Table 1, the PC4 has the lowest mean compared to the other PCs. Therefore, it has an overall darker appearance in the resulting image.

The analysis of the PC5, PC6, PC7, PC8, and PC9 indicated that they contained either noise or insignificant information. However, the analysis of the PC10 (Figure 6) which contained only 0.26% of the overall variance indicated an interesting result. Unlike the PC4 image, it has less noise and much improved textural as well as land cover information. As seen from Table 1, the PC10 includes the most scattered minimum, maximum and SD values compared to the other PCs. Therefore, it should have a more scattered appearance in the resulting image.
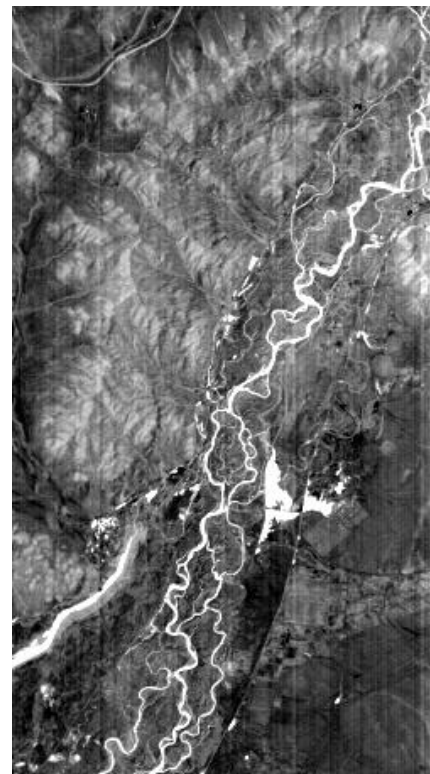

Figure 5. PC4 image of the test area.

As seen from the overall PCA study, it can be meaningfully applied to the dimensionality reduction process, in which, a multidimensional primary dataset is divided and reduced to more meaningful and manageable groups. As could be seen from the study, initially, we had 169 hyperspectral bands. However, after applying the PCA, we extracted 10 PCs. Of these PCs, the first four PCs contained over 95% of the overall variance. In most analysis cases, these PCs can represent all 169 bands. Nevertheless, a

thorough analysis of the PCA indicated that a PC that contained a negligible amount of the overall variance could still contain useful information.
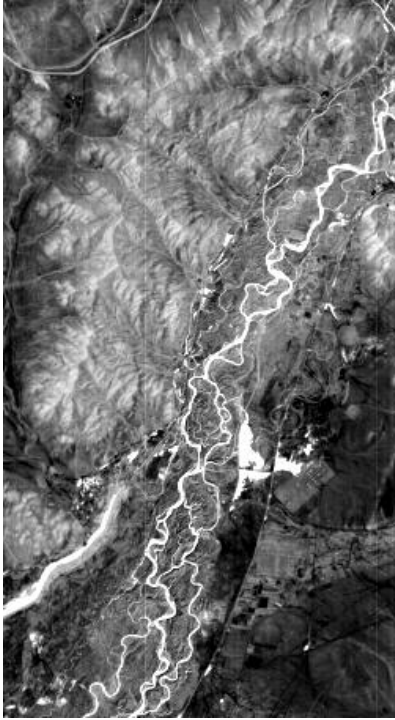

Figure 6. PC10 image of the test area.

Figure 7. PC1,2,3 image of the test area.

In our case, the PC10 contained some useful information regarding the land cover classes as well as the existing texture of the image.

For further analyses, the obtained PCs can be composed and different high quality color images could be created.

For example, a combination of the PC1, PC2, and PC3 can create a good color image for land cover classification illustrated in Figure 7. Likewise, by combining PC1, PC2, and PC10, we can create a color image shown in Figure 8.

Figure 8. PC1,2,10 image of the test area.

5. CONCLUSIONS

As known, PCA is one of the feature extraction methods based on a dimensionality reduction process, in which, an initial set of the raw data should be divided and reduced to more manageable groups. The most important characteristic of these large datasets is that they have a large number of variables and they require lots of computing resources to process them. Therefore, feature extraction helps the researchers in receiving the best feature from the large datasets to select and combine variables into features, thus, effectively reducing the amount of data. So, the aim of the study was to apply a feature extraction approach to a Hyperion image, reduce its

Table 1. Statistics of the PCA performed for the Hyperion image.

PCA	Min	Max	Mean	SD	Eigenvalue	Percent
PC1	1577	14748	2214.190082	217.104567	16040395.034268	64
PC2	1599	15737	2250.994356	233.409985	4563359.220363	23
PC3	1634	17445	2350.800203	260.401749	884703.817610	5.4
PC4	1478	17045	2176.129832	258.198426	122692.817214	1.8
PC5	1413	17725	2173.387669	276.728774	56117.446186	0.85
PC6	1384	18462	2193.854975	293.507945	35880.155510	0.64
PC7	1333	17978	2110.521895	287.336463	23396.190179	0.51
PC8	1371	18971	2220.516574	298.004494	14952.870195	0.28
PC9	1400	19305	2252.358828	303.942349	13940.242730	0.27
PC10	1340	19641	2281.334364	317.871720	13128.532591	0.26

dimensionality and extract different meaningful features.

As seen from the PCA study, it might be profoundly applied to the dimensionality reduction process, where, multidimensional datasets could be divided and reduced to more meaningful and manageable groups. For example, we had 169 hyperspectral bands for the initial analysis and after applying the PCA, we extracted 10 PCs. These PCs contained more than 96%.

REFERENCES

[1] Kumar, K.K., Chaduvula, K. and Markapudi, B., 2020, A Detailed Survey On Feature Extraction Techniques In Image Processing For Medical Image Analysis, European Journal of Molecular & Clinical Medicine, Volume 07, Issue 10, 2020.

[2] Mutlag, W., Ali, S., and Taher, B., 2020, Feature Extraction Methods: A Review Journal of Physics: Conference Series 1591 (2020) 012028, IOP Publishing. DOI: https://doi.org/10.1088/1742-6596/1591/1/012028

[3] Chatterjee, S., 2020, What is Feature Extraction? Feature Extraction in Image Processing, Available at https://www.mygreatlearning.com/blog/feature-extraction-in-image-processing/.

[4] Elnemr, H.A., Zayed, N.M. and Fakhreldein, M.A., 2016, Feature Extraction Techniques: Fundamental Concepts and Survey, Handbook of Research on Emerging Perspectives in Intelligent Pattern Recognition, Analysis, and Image Processing, pp.31. DOI: https://doi.org/10.4018/978-1-4666-8654-0.ch013

[5] Krishnan, A. and Sreekumar, A., 2014, Survey on Image Segmentation and Feature Extraction Methods for Acute Myelogenous Leukemia Detection in Blood Microscopic Images, International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 7877-7879.

[6] Kumar, G. and Bhatia, P. K., 2014, A Detailed Review of Feature Extraction in Image Processing Systems," 2014 Fourth International Conference on Advanced Computing & Communication Technologies, pp. 5-12. DOI: https://doi.org/10.1109/ACCT.2014.74

[7] Suganya, S., 2015, Analysis of Feature Extraction of Optical Character detection in Image Processing Systems, International journal of engineering research & technology (ijert) ncrtet – 2015 (Volume 3 – Issue 04).

[8] D.Amarsaikhan and M. Saandar (June 24th 2011). Fusion of Multisource Images for Update of Urban GIS, Image Fusion and Its Applications, Yufeng Zheng, IntechOpen. DOI: https://doi.org/10.5772/16293

[9] Subashi, A., 2020, Chapter 2 - Data preprocessing, Practical Machine Learning for Data Analysis Using Python, pp.27-89. DOI: https://doi.org/10.1016/B978-0-12-821379-7.00002-3

[10] EO-1, 2017, EO1 Trajectory Details, National Space Science Data Center of NASA, Available at: https://nssdc.gsfc.nasa.gov/.

[11] Mather, P.M. and Koh, M., 2011, Computer Processing of Remotely-Sensed Images: an Introduction. Fourth edition (Wiley-Blackwell). DOI: https://doi.org/10.1002/9780470666517

[12] Amarsaikhan, D., Ganzorig, M., Saandar, M., Blotevogel, H.H., Egshiglen, E., Gantuya, R., Nergui, B. and Enkhjargal, D., 2012, Comparison of multisource image fusion methods and land cover classification, International Journal of Remote Sensing, Vol.33 (8), pp.2532-2550. DOI: https://doi.org/10.1080/01431161.2011.616552

[13] Gonzales, R. C. and Woods, R. E., 2002, Digital Image Processing, Second Edition. Upper Saddle River, (New Jersey: Prentice-Hall).