CEIG - Spanish Computer Graphics Conference (2025)
O. Argudo and O. Iparraguirre (Editors)

First steps in the dimensionality reduction of hyperspectral images of real-world scenarios based on curve representation

Luis René Domínguez Fuentes J. Roberto Jiménez-Pérez Juan M. Jurado David Jurado-Rodríguez

Department of Computer Science, University of Jaén, Spain
Center for Advanced Studies in Information and Communication Technologies, University of Jaén, Spain
Graphics and Geomatics Group of Jaen (GGGJ), University of Jaén, Spain

Abstract

Dimensionality reduction (DR) has been used in hyperspectral data mining for a variety of purposes. In particular, it has been used as a preprocessing technique to reduce a very high dimensional data space coupled with its characteristically large volume to a manageable low dimensional space in which data analysis can be performed more efficiently. This study focuses on developing a workflow for hyperspectral image processing employing feature extraction using vegetation indices. Combined with feature selection to develop a DR by band selection (DRBS) method that searches for a subset of bands representing the original data, so that the information of interest in the data can be retained in the selected subset of bands. It is proposed to employ curve simplification techniques such as Douglas-Peucker to achieve this end. To perform the experiments we have used a hyperspectral image (HSI) taken by a drone flight with 0.31 m resolution, in a scenario that presents both vegetation and other architectural elements, with the presence of illuminated areas and shadows. Our results show that our method manages to reduce the amount of information by 96.4% of the HSI. In this case, preserving the most important features with a minimum level of loss, in most cases with mean square error very close to 0, this allows subsequently recreating the original data with high reliability. (see https://www.acm.org/publications/class-2012)

CCS Concepts

• ***Computing methodologies*** → *Dimensionality reduction (DR); Hyperspectral Images (HSI); Feature Selection (FS);* • ***Hardware*** → *Sensors and UAV;*

1. Introduction

The use of remote sensing data sources, known as remote sensing (RS), has experienced growth in recent years. RS data have been widely used in studies in urban and rural areas, where buildings, industrial complexes and vegetation areas predominate.

It is in this field where the capture of HSIs stands out, as it has proven to be a remarkable advance in telematic data acquisition. Hyperspectral images contain dozens of narrow, contiguous spectral bands spanning the entire visible range of the electromagnetic spectrum (ES). In addition, the emitted or absorbed radiation captured by these images is stored in a hypercube format. Such data have three primary dimensions: two dimensions representing the spatial characteristics, which can be represented with coordinates (x, y), and a third dimension representing the reflectance in a particular band (λ).

Hyperspectral UAV systems cover smaller areas but capture much larger volumes of data due to their high spatial and spectral resolution. This results in data sets that are not only more detailed, but also data intensive, posing unique challenges and opportunities for analysis and processing. The main challenges facing HSI classification of UAVs are as follows: high computational expense, complex influences of shadows and terrain and extreme spatial-spectral heterogeneity.

The main contribution of this work is the proposal of a new workflow for the simplification and normalization of spectral signatures. This method aims to simplify the information by retaining only the most relevant and significant data from the original dataset. Its main strengths are low computational cost, high processing speed, and automatic band selection. In addition, it is designed to be used by non-specialized users, facilitating its application in other operating contexts. The description of the steps followed in this methodology is presented in section 4, which include: shadow filtering to reduce noise, vegetation detection using vegetation indices, DR by band selection and spectral signature similarity calculation to evaluate the method.

2. Previous work

HSI is an extension of multispectral imaging due to its high number of adjacent spectral bands along the ES, which allows a closer approximation to the actual spectral signature, leading to more com-

DOI: 10.2312/ceig.20251114

prehensive analysis of material properties. These images measure the intensity of energy emitted or absorbed through a group of contiguous spectral bands, where the energy intensity and wavelength depend directly on the characteristics of the material [ES10].

Unlike conventional multispectral remote sensing, hyperspectral remote sensing provides a higher degree of detail, enabling the detection of subtle biochemical changes and accurate representation of the spectral properties of various terrain features [LMVW14]. Its applications extend broadly to environmental monitoring [PMG*22], mineral identification [ZZH*23], oil spill detection [DKGL23] and precision agriculture [BPS*06].

DR with the goal of identifying and eliminating statistical redundancies of hyperspectral data while keeping as much spectral information as possible. Relatively few bands can represent most of the information in HSIs, making DR very useful for: storage, transmission, classification, spectral unmixing, target detection, and visualization of remote-sensing data. Hyperspectral DR consists of both feature selection (FS) and feature extraction (FE) [GYL*17].

Unsupervised DR methods address cases where no labeled samples are available and aim to find an alternative representation of the data in a lower-dimensional space according to a given criterion. The objective of these methods is not to optimize the accuracy for a given classification task [Cha13]. For example, principal component analysis (PCA) reduces dimensionality by capturing the maximum variance in the data. Independent component analysis (ICA) finds the project matrix by maximizing the statistical independence. Minimum noise-fraction (MNF) transformation obtains the reduced features according to the image quality measured by the SNR, and local linear feature extraction (LLFE) methods seek a projection direction in which neighborhood relationships are preserved in the feature spaces. The nonlinear versions of these methods, such as kernel methods (e.g., kernel PCA, kernel ICA, and kernel MNF) have been widely used to detect higher-order statistical redundancies.

Hyperspectral classification typically consists of two steps:

1. DR, via either FE or FS [JKC13].
2. A training procedure for designing the classifier.

However, it is difficult to ensure that the best features from the first step will optimize the classification performance of the following one. Therefore, our proposal is to create a versatile DR method that allows selecting the bands that define the spectral behavior of each signature in a fast and efficient way. Later on, the original spectral signature can be recreated from the selected bands.

3. Materials and Datasets

The data used in these tests were provided by the Grupo de Gráficos y Geomática de Jaén (GGGJ). The dataset corresponds to an HSI captured by a flight conducted in 2023 over the campus of the **University of Trás-os-Montes e Alto Douro (UTAD)** in Vila Real, Portugal.

The flight was carried out using a drone equipped with a hyperspectral camera model **Nano Hyperspec VNIR1**, which provided images of **270 bands** between 398 nm and 1002 nm, and with a spatial resolution of **0.31 m**. The HSI used in this study is **25 GB** in size.

4. Our method

The stages of the proposed algorithm are shown in the figure 1. In the following sections the most important stages are explained which are: shadow filtering, vegetation detection and band selection.

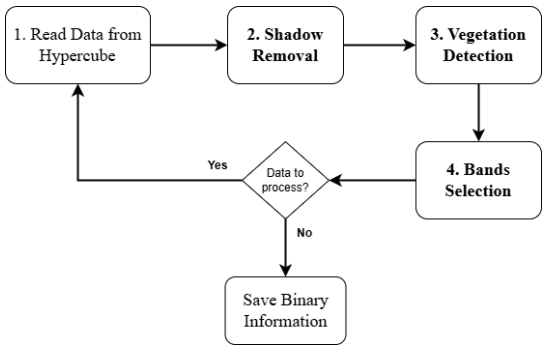

Figure 1: *Flowchart of the proposed algorithm.*

Due to the large size in gigabytes of the HSIs, the algorithm processes the image in parts in order not to occupy completely the memory of the computer where it is executed. Therefore, one of the parameters of the algorithm is the block size.

4.1. Shadow and noise removal

Before proceeding with the different phases of hyperspectral data processing, we proceed with shadow filtering. This consists of discarding all pixels that belong to a poorly illuminated area, this allows to remove noisy values. The main benefit of this section is to reduce the number of pixels to be processed in the following phases, which results in computational cost savings.

If the data is taken from an area with low illumination, it will be observed that the reflectance values are extremely low, not exceeding 0.1. Then, to perform this filtering, the mean of the reflectance values is calculated, and if it is below 0.1 it is discarded. This threshold can be parameterized and the filtering can be refined more or less depending on the case.

4.2. Vegetation detection

Vegetation, in general, has a very characteristic spectral signature. This is because the visible ES region (400 - 700 nm) of vegetation is controlled by pigments in the chloroplasts of green leaves. Specifically, chlorophyll is the main absorber of radiation in the visible region and its absorption is dominant in the visible red (600 - 700 nm), and a strong emitter in the NIR (700 - 850 nm). **Vegetation indices (VI)** exploit these spectral characteristics to identify vegetative cover efficiently, without the necessity of graphically representing the complete spectral signature.

During the development of this methodology, several VIs were

evaluated, and it was decided to use **NDVI** and **MSAVI**. Because in this study [ZSWZ15] , not all VIs respond to shadowed areas in the same manner. NDVI has been shown to be more tolerant to shadow effects, whereas MSAVI exhibits greater sensitivity to them. So including the choice of VI as a parameter in our methodology is essential.

4.3. Band Selection

This high level of granularity of the HSIs is one of the main factors responsible for the high memory consumption, commonly referred to as the curse of dimensionality. DR aims to identify and remove statistical redundancies in hyperspectral data while preserving as much spectral information as possible.

However, these techniques are presented in the state of the art as independent processes from the classification process, apart from the fact that they usually require complex statistical calculations and a deeper analysis. Our methodology proposes to perform a FS by means of **Douglas-Peucker** [Ram72]. As we have already seen, spectral signatures can be represented in a two-dimensional space, where x corresponds to the wavelength and y corresponds to the normalized reflectance value. So if we consider the spectral signature as a line described by numerous points in 2D, a subset of points describing the original curve can be found with a minimum loss of information. The main characteristic of this algorithm is its **adaptability to the complexity** of the data. It retains more representative points in those curves with more fluctuations, and selects fewer points in the simpler curves.

4.4. Calculation of spectral similarity

The **Spectral Angle Mapper (SAM)** [KLB*93] is a tool that allows to quickly calculate the spectral similarity by calculating the "angle" between the two spectra, treating them as vectors in a space with dimensionality equal to the number of bands (nb). A smaller angle indicates greater similarity between the signatures, being especially useful for evaluating materials that have a consistent spectral response but may differ in magnitude due to factors such as illumination. The SAM value ranges from **0º (perfect similarity)** to **90º (large difference)**, where a smaller angle indicates greater similarity. As explained before, Douglas-Peucker selects more or less representative bands depending on the fluctuations of the spectral signature. Therefore, before calculating SAM, a normalization of both vectors must be performed, so that they have the same dimensionality. This consists of performing an interpolation to find the reflectance values of those bands that were not selected.

5. Results

To test the validity of the methods described in the previous section, a series of experiments specifically designed to evaluate both the efficiency and effectiveness of the proposed techniques have been carried out. These experiments focus on the selection of bands by Douglas-Peucker and their corresponding evaluation with SAM.

5.1. Band selection and curve normalization

For these tests, the **threshold** (ε) of the Douglas-Peucker algorithm —which determines the number of points retained during the simplification process— was systematically varied. This threshold controls the algorithm's tolerance to fluctuations or variations within the spectral signatures: the lower the epsilon value, the more strictly the algorithm follows the original curve, resulting in a greater number of retained points. In this way, it is analyzed how the results vary as a function of this parameter. For this case the values must be very tight, because the distance between values of different bands are minimal, for this reason the values used were **0.01** and **0.015**.

For the realization of these tests, **250 samples** were randomly chosen along the image. This ensures that sufficient spectral signatures of the different materials found in the scene are obtained. Once the simplification is done, a reconstruction of each spectral signature is performed using **linear interpolation** between the selected points. In this way the original and the simplified spectral signature can be compared.

ε	**0.01**	**0.015**
Max points	119	78
Min points	12	2
MSE	0	0
Max. Accumulated error	0.807	1.617
Min. Accumulated error	0.393	0.647
Size (MB)	922	556

Table 1: *Table with the results of the different* ε

5.2. Checking the similarity measure

Another way to evaluate the proposed DR process is to calculate the similarity between the original spectral signatures and the synthesized signatures by first reconstructing them with linear interpolation. The similarity calculation is performed using **SAM**, from which the results in table 2 have been obtained.

ε	**0.01**	**0.015**
Mean	1.1045°	1.5764°
Median	1.0892°	1.5467°
Standard deviation	0.3625°	0.5117°
Minimum	0.3796°	0.5505°
Maximum	1.8511°	3.0298°
Coefficient of variation	0.3282	0.3246

Table 2: *Table of SAM results depending on the value* ε*. Where value ranges from 0º (perfect similarity) to 90º (large difference)*

Spectral signatures with different behaviors have also been selected to visually show how this RD method works with epsilon equal to 0.01. These plots can be seen in figure 2.

6. Discussion

This paper discusses methodologies for the analysis of spectral data. Initially, it is highlighted that the calculation of VI facilitates the identification of vegetation without a detailed analysis of

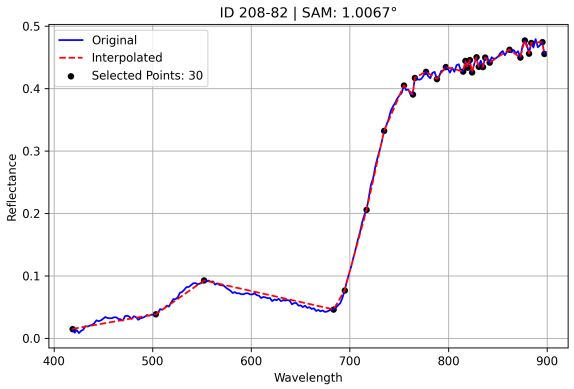

Figure 2: *An example of the comparison between original and simplified spectral signatures.*

spectral signatures. Subsequently, the Douglas-Peucker algorithm is introduced to synthesize the spectral signatures of vegetation and inanimate objects, noting an inversely proportional relationship between accuracy and data compression. A low ε value (0.01) is recommended to preserve data fidelity, while a higher value (0.015 or higher) results in further storage reduction with a slight loss of accuracy. The choice of the appropriate threshold will depend on the level of detail required and the objective of the data processing.

Finally, the SAM metric was used to evaluate the differentiation between the original spectral signatures and those recreated from the bands selected after application of the DP algorithm. The results indicate that an interpolation with a ε value of 0.01 generates significantly smaller angular errors and lower scatter of the SAM values compared to a value of 0.015. This suggests a higher accuracy of the method with a tighter tolerance, albeit with a lower efficiency in terms of storage size due to the selection of more bands. The selection of the ε value should consider the trade-off between storage efficiency and fidelity of the results.

7. Future works

Future lines of research focus on a **comprehensive comparison** of the Douglas-Peucker DR method with advanced methods such as PCA to improve computational and memory efficiency. It is also proposed to **integrate this methodology as an initial stage in classification** and clustering algorithms to facilitate the handling of large spectral data, **optimize processing** and increase the effectiveness of the analysis in systems **with limited resources**. Finally, it is planned to develop a diverse data set to simplify the categorization and validation of entities, which will allow a more accurate training of algorithms and extend the practical application of the methodology with greater computational efficiency.

8. Aknowledgements

This result has been partially supported through the research projects PID2022-137938OA-I00, TED2021-132120B-I00 and PID2021-126339OB-I00 by the Spanish Ministry of Science and Innovation and ERDF funds.

References

[BPS*06] BANNARI A., PACHECO A., STAENZ K., MCNAIRN H., OMARI K.: Estimating and mapping crop residues cover on agricultural lands using hyperspectral and ikonos data. *Remote Sensing of Environment 104*, 4 (2006), 447–459. URL: https://www.sciencedirect.com/science/article/pii/S0034425706002148, doi:https://doi.org/10.1016/j.rse.2006.05.018. 2

[Cha13] CHANG C.-I.: *Hyperspectral Data Processing: Algorithm Design and Analysis*. John Wiley & Sons, Feb. 2013. Google-Books-ID: uP1gRRfkMxgC. 2

[DKGL23] DUAN P., KANG X., GHAMISI P., LI S.: Hyperspectral remote sensing benchmark database for oil spill detection with an isolation forest-guided unsupervised detector. *IEEE Transactions on Geoscience and Remote Sensing 61* (2023), 1–11. doi:10.1109/TGRS.2023.3268944. 2

[ES10] ELMASRY G., SUN D.-W.: CHAPTER 1 - Principles of Hyperspectral Imaging Technology. In *Hyperspectral Imaging for Food Quality Analysis and Control*, Sun D.-W., (Ed.). Academic Press, San Diego, Jan. 2010, pp. 3–43. URL: https://www.sciencedirect.com/science/article/pii/B9780123747532100012, doi:10.1016/B978-0-12-374753-2.10001-2. 2

[GYL*17] GHAMISI P., YOKOYA N., LI J., LIAO W., LIU S., PLAZA J., RASTI B., PLAZA A.: Advances in Hyperspectral Image and Signal Processing: A Comprehensive Overview of the State of the Art. *IEEE Geoscience and Remote Sensing Magazine 5*, 4 (Dec. 2017), 37–78. Conference Name: IEEE Geoscience and Remote Sensing Magazine. URL: https://ieeexplore.ieee.org/abstract/document/8113122, doi:10.1109/MGRS.2017.2762087. 2

[JKC13] JIA X., KUO B.-C., CRAWFORD M. M.: Feature Mining for Hyperspectral Image Classification. *Proceedings of the IEEE 101*, 3 (Mar. 2013), 676–697. Conference Name: Proceedings of the IEEE. URL: https://ieeexplore.ieee.org/abstract/document/6450025, doi:10.1109/JPROC.2012.2229082. 2

[KLB*93] KRUSE F. A., LEFKOFF A. B., BOARDMAN J. W., HEIDEBRECHT K. B., SHAPIRO A. T., BARLOON P. J., GOETZ A. F. H.: The spectral image processing system (SIPS) - interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment 44*, 2 (May 1993), 145–163. URL: https://www.sciencedirect.com/science/article/pii/003442579390013N, doi:10.1016/0034-4257(93)90013-N. 3

[LMVW14] LUCIEER A., MALENOVSKÝ Z., VENESS T., WALLACE L.: HyperUAS—Imaging Spectroscopy from a Multirotor Unmanned Aircraft System. *Journal of Field Robotics 31*, 4 (2014), 571–590. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/rob.21508. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21508, doi:10.1002/rob.21508. 2

[PMG*22] PIARULLI S., MALEGORI C., GRASSELLI F., AIROLDI L., PRATI S., MAZZEO R., SCIUTTO G., OLIVERI P.: An effective strategy for the monitoring of microplastics in complex aquatic matrices: Exploiting the potential of near infrared hyperspectral imaging (NIR-HSI). *Chemosphere 286* (Jan. 2022), 131861. URL: https://www.sciencedirect.com/science/article/pii/S004565352102333X, doi:10.1016/j.chemosphere.2021.131861. 2

[Ram72] RAMER U.: An iterative procedure for the polygonal approximation of plane curves. *Computer Graphics and Image Processing 1*, 3 (Nov. 1972), 244–256. URL: https://www.sciencedirect.com/science/article/pii/S0146664X72800170, doi:10.1016/S0146-664X(72)80017-0. 3

[ZSWZ15] ZHANG L., SUN X., WU T., ZHANG H.: An Analysis of Shadow Effects on Spectral Vegetation Indexes Using a Ground-Based Imaging Spectrometer. *IEEE Geoscience and Remote Sensing Letters 12*, 11 (Nov. 2015), 2188–2192. Conference Name: IEEE Geoscience and Remote Sensing Letters. URL: https://ieeexplore.ieee.org/abstract/document/7180331, doi:10.1109/LGRS.2015.2450218. 3

[ZZH*23] ZHANG L., ZHANG M., HUANG J., ZHANG C., YE F., PAN W.: A new approach for mineral mapping using drill-core hyperspectral image. *IEEE Geoscience and Remote Sensing Letters 20* (2023), 1–5. doi:10.1109/LGRS.2023.3328139. 2