Journal of Computational Science 26 (2018) 13–21

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs

Interactive visual analysis and classification of hyperspectral imaging data

Anatoliy Antonov[a], Lars Linsen[a,b,*]

[a] Jacobs University, Bremen, Germany
[b] Westfälische Wilhelms-Universität Münster, Germany

ARTICLE INFO

Article history:
Received 14 September 2017
Received in revised form 6 February 2018
Accepted 4 March 2018
Available online 6 March 2018

Keywords:
Interactive visual analysis
Hyperspectral imagery
Data reduction
Clustering

ABSTRACT

Hyperspectral imaging is a widely used remote sensing technique in planetary sciences. Captured data consist of arrays of images of the same scene taken at a high number of sensor wavelengths. Studying these data, scientists search to understand, for example, the mineral composition of the surface, or types and kinds of vegetation present in the region. Different mineral or vegetation classes induce spectral features that allow for their identification and mapping. The two main tasks are selecting a subset of bands that best capture the relevant spectral features and separating different mixtures from summary products in these bands. It is common practice that the subsets of bands used in the analysis are predefined based on experience, i.e., without taking into account actual data. Then, classification of the regions is performed by analyzing the summary products. We, instead, propose an approach that allows for data-driven selection of subsets of bands and for a more accurate separation of different mixtures. Our approach relies on an interactive visual analysis using suitable visual encodings and interaction mechanisms. In a first step, we produce a similarity plot of the bands of the hyperspectral imaging data by employing a projection-based dimensionality reduction technique. The similarity plot allows for the selection of most informative bands. In a second step, we apply an automatic hierarchical density-based clustering approach to the pixels of the selected bands. The resulting cluster hierarchy is interactively explored and adjusted using coordinated views of a cluster tree visualization, a parallel coordinate plot of the bands, and a spatial data visualization. Brushing and linking in the coordinated views allows for an intuitive interactive analysis of the bands leading to the desired mineral mapping result. The effectiveness of our approach is demonstrated by applying it to mineral mapping of Compact Reconnaissance Imaging Spectrometer for Mars data and vegetation classification of Airborne Visible/Infrared Imaging Spectrometer data.

1. Hyperspectrology

Hyperspectral imaging is a remote sensing technique with various applications in planetary sciences. Captured data consist of arrays of images of the same scene taken at a high number of sensor wavelengths. Studying these data, scientists search to understand, for example, the mineral composition of the surface, or types and kinds of vegetation present in the region.

As the technology progresses, spectral and spatial resolution of the instruments continue to increase. Nowadays, hyperspectral data collection generates hundreds of high-resolution 2D images. The volume presents a need for integrated automated components and an intuitive interactive workflow.

Common approaches include per-pixel manual analysis of spectral signatures and machine learning approaches. Each 1D spectrum consists of values of a point across all images. Other approaches include generating summary products and studying them or their combinations visually, combining several in red, green, and blue color channels.

The mineral mapping scenario relies on the knowledge of the spectral features of the mineral classes for their separation and identification. Such features are discovered in laboratory spectra of clean minerals, afterwards they are captured in summary products. The summary products use values, band ratios, band depths, and spectral slopes around a few selected wavelengths to diagnose the likelihood of the features to be present in any measured spec-

* Corresponding author at: Westfälische Wilhelms-Universität Münster, Germany.
E-mail address: l.linsen@jacobs-university.de (L. Linsen).

https://doi.org/10.1016/j.jocs.2018.03.003
1877-7503/

trum. Thus, they also can indicate the likelihood that a mineral of interest is present in a scanned mineral mixture.

However, multiple factors influence the spectral shape in the remote sensing measurements and complicate the analysis by diverting the spectral shape from the laboratory results. For example, even in high-resolution images, each pixel contains information about hundreds of square meters of the surface, essentially creating a mineral mixture signal even when there might be relatively large spots of clean minerals. Additionally, the properties of the surface like purity, abundance, and grain sizes modify the resulting spectral shape. Finally, scanning artifacts and noise bring distortions into the signal.

As the data are perturbed by these factors, the indicators selected in the laboratory and used in the summary products may become less usable for the characterization. However, it is common practice to rely on the identified summary products without adjusting them for the actual measured data.

Another difficulty of using summary products is that their output is given in the form of an image with continuous distribution, which requires additional efforts to define thresholds and separate different mineral mixtures.

In this paper, we make an attempt to overcome the shortcomings of traditional methods and assess the possibilities of integration of our approach into the scientific workflow. For the mineral mapping, we suggest a two-step approach based on interactive visual analysis that, first, allows for data-driven adjustment of the bands selected to represent the spectral features and, second, for a user-guided extraction of regions with distinct mineral mixtures from the data. To document that our approach is not limited to the mineral mapping scenario, the same approach is applied to the vegetation classification context, and is able to provide a first classification estimate from a highly reduced number of bands and in the absence of a priori knowledge of number of classes or their properties.

Our contributions can be summarized as follows:

- Application of a dimension-based projection technique to hyperspectral bands to produce similarity plots and interactively select most informative bands for classification.
- Application of multidimensional clustering approach to pixels of hyperspectral images to extract regions with distinct spectral signatures (mineral mixtures or vegetation).
- User-guided classification by investigating clustering results using multiple coordinated views.

2. Hyperspectral data analysis and mineral mapping

The raw hyperspectral data recorded by the detector undergo a number of processing steps before being investigated. Processing includes calibration and noise reduction, photometric corrections (reducing effects of the observation angle), and atmospheric corrections (removing atmospheric absorption bands from the spectra). For a comprehensive list of typically applied processing steps we refer to the literature [1]. The outcome of this pre-processing pipeline is a registered set of cleaned and properly transformed 2D images, one for each of the hyperspectral bands.

Decisions about the mineral composition of the surface are made by the characteristic absorption features of minerals in the spectral signal. In other words, the information that is required for making the decision about a mineral being present or absent at a certain location is not inherent in the absolute reflectance values of this point at the certain wavelengths, but stored in the relationships between values at many wavelengths. Approaches to the analysis of the hyperspectral data extract this information with the help of summary products or by the spectral unmixing family of approaches.

Summary products are designed to capture the particular spectral features. Each of them is aimed at diagnostics of a particular feature, e.g., BD2100 computes the band depth at 2100 nm wavelength. The (relative) band depth is defined by the difference of the amplitude of the measured signal to a fitted baseline. Hence, summary products try to capture how likely it is that a specific feature is present in the data coming from the observed point. By combining several features, the summary products encode how likely it is that a certain mineral or a mineral group is included in the signal at the given location. For a comprehensive overview of summary products used in the analysis of CRISM data we refer to the literature [2].

Spectral unmixing approaches use libraries of laboratory-taken spectra of clean minerals. Approaches of this type attempt to combine the library spectra in a way that would model the data best. As features in measured spectra are influenced not only by the mixture but also by many other factors, the results are only estimates of the mineral content. Hence, it is desirable to investigate the images for the bands chosen for the mineral mapping for the given data set to check their suitability. However, due to the lack of effective and efficient analysis tools for this task, this is typically not done. We aim at providing an interactive visual analysis approach to this task in the first step of our analysis.

Summary products are typically examined using color-coded images. Visual inspection of certain combinations of these summary products in RGB images allow scientists to identify mineral classes present in the data. However, the combination of summary products of multiple minerals is challenging and quantitative analysis of the automatically computed summary products is not provided. Our second analysis step allows for an intuitive interactive analysis and adjustment of an automatically generated result.

An description of analysis methods for hyperspectral imaging data is provided by Landgrebe [3]. The described approaches as well as follow-up work in literature is concerned with the classification of the pixels using various automatic approaches [4–6]. We instead provide an interactive analysis process, where the user can bring in his/her expertise into the analysis process by interacting with visual representations of automatically extracted information in an iterative work process.

3. Data

The specific hyperspectral data analyzed in the paper were the following:

- CRISM instrument: up to 438 hyperspectral bands, wavelengths are from 362 nm to 3920 nm at 6.55 nm/channel, spatial extent varies by dataset: 640×450 pixels (Gale Crater), 800×739 pixels (Nili Fossae), with a resolution of about 20 m/pixel.
- AVIRIS instrument: up to 224 bands, wavelengths are from 400 nm to 2500 nm, spatial extent varies by dataset: 217×512 pixels (Salinas), 145×145 pixels (Pines), 86×83 pixels (Salinas-A spatial subset), with a resolution of about 30 m/pixel.

In both cases we are dealing with an array of 2D images of the same spatial region, where each image corresponds to a certain sensor wavelength, and represents a micro-band.

4. Similarity plot of hyperspectral bands

Our visual analysis approach starts with the construction of a similarity plot for the different spectral bands of the hyperspectral dataset. The idea behind the plot is that it can be used for