2017

**HyperDB: A Hyperspectral Land Class Database Designed for an Image Processing System**

Yizhou Fan  
*the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China.*

Ding Ni  
*the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China.*

Hongbing Ma  
*the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China.*

Follow this and additional works at: [https://tsinghuauniversitypress.researchcommons.org/tsinghua-science-and-technology](https://tsinghuauniversitypress.researchcommons.org/tsinghua-science-and-technology)

Part of the Computer Sciences Commons, and the Electrical and Computer Engineering Commons

**Recommended Citation**

Yizhou Fan, Ding Ni, Hongbing Ma. HyperDB: A Hyperspectral Land Class Database Designed for an Image Processing System. Tsinghua Science and Technology 2017, 22(1): 112-118.
HyperDB: A Hyperspectral Land Class Database Designed for an Image Processing System

Yizhou Fan*, Ding Ni, and Hongbing Ma

Abstract: Hyperspectral remote sensing is becoming more and more important amongst remote sensing techniques. In this paper, we present a hyperspectral database (HyperDB) designed to cooperate with an embedded hyperspectral image processing system developed by the authors. Hyperspectral data are recognized and categorized by their land coverage class and band information, and can be imported from various sources such as airborne and spaceborne sensors carried by airplanes or satellites, as well as handhold instruments based on in situ ground observations. Spectral library files can be easily stored, indexed, viewed, and exported. Since HyperDB follows standard design principles—indeedependence, data safety, and compatibility—it satisfies the practical demand for managing categorized hyperspectral data, and can be readily expanded to other peripheral applications.

Key words: hyperspectral database; spectral library; image processing

1 Introduction

Hyperspectral remote sensing is defined as the acquisition of images from hundreds of contiguous, registered, spectral bands such that for each pixel a radiance spectrum can be derived\textsuperscript{[1]}. In over three decades, hyperspectral remote sensing has played an increasingly important role in the remote sensing imaging field.

Compared to traditional multispectral remote sensing methods, hyperspectral imaging has several advantages. By creating a dense and continuous radiance spectrum, hyperspectral imaging can delineate the spectral characteristics of different genres of land class in detail, hence making precise spectral operations possible, such as identifying surface materials, removing atmospheric effects, correlating pixel spectra with spectral databases, etc.\textsuperscript{[2,3]}

The versatility of hyperspectral remote sensing has inspired its application in a wide range of areas. The retrievable and manipulatable data content types include plant and vegetation species\textsuperscript{[4]}, wildlife and livestock\textsuperscript{[5,6]}, soil and wetland\textsuperscript{[7]}, mineralogy products\textsuperscript{[8]}, fire monitoring\textsuperscript{[9]}, etc.

The importance of hyperspectral remote sensing yields an urgent requirement for a corresponding spectral library or database. To date, many databases that support hyperspectral remote sensing data have been established and developed. The Advanced Spaceborne Thermal Emission Radiometer (ASTER) spectral library is a collection of contributions of spectral data compiled by the Jet Propulsion Laboratory (JPL), Johns Hopkins University (JHU), and the United Stated Geological Survey (USGS)\textsuperscript{[10]}. The Spectral Input/Output (SPECCHIO) database is designed to manage heterogeneous data from different sources, with logical relations and consistency, intuitive interfaces, flexibility to changes in science context, file format independence, and data size scalability\textsuperscript{[11]}. Moreover, a redesign of SPECCHIO is undertaken in Ref.\textsuperscript{[12]} to strengthen the issues of user friendliness and inconsistencies in the data model. In contrast to these
integrated databases, the SpectraProc DB is designed exclusively for the rapid, precise, and repeatable operations of the Analytical System Device (ASD) FieldSpec Pro hyperspectral data format[13].

In this paper, we present a hyperspectral database, named HyperDB, which can store, index, and process hyperspectral data files, independent of the data content types. HyperDB has some distinguishing traits. It is inherently based on land-class indexing, which requires compulsory categorizing of the data files with the land-class tag. It allows data file input from multiple sources, including those retrieved from spaceborne and airborne sensors, as well as those provided by in situ ground observation apparatuses. HyperDB collaborates with an embedded hyperspectral image processing system developed by the authors, which can perform various image processing operations, while passing data and results to and from HyperDB using the unified data format, eliminating the gap between database and application.

The next sections are organized as follows. After briefly introducing the embedded image processing system in Section 2, the design principles and implementation details of HyperDB are depicted in Section 3. The main significant functions of HyperDB are elaborated in Section 4, and concluding remarks of our work are provided in Section 5.

2 Hyperspectral Image Processing System

As previously indicated, HyperDB is embedded with a hyperspectral image processing system developed by the authors. This image processing system, a very direct application of HyperDB per se, whose main interface is shown in Fig. 1, is designed to perform a series of operation on different spectral bands. The functions of the hyperspectral image processing system are briefly introduced below.

False-color composition and band selection. Amongst the bands whose number often easily transcends 1000, three bands are selected and normalized to represent the R, G, and B components of an image, to generate a visually-friendly false-color view. The three bands can either be selected manually by the user, or determined by an automatic band selection strategy based on Ref. [14] implemented in the system, which selects those bands with high signal-noise ratio and mutual information. Note that the bands selected by this strategy are not only optimal for the false-color view, but also ready for the classification function described below.

Image pre-processing. The system contains various filters to support image pre-processing operations such as noise elimination, topological and morphological transformation, etc.

Spectra importation and exportation. The system maintains an interactive relationship with HyperDB, which is distinct and crucial in this study. The system allows the user to select pixels in the false-color image, tag the pixel with its corresponding land-class category, and export its hyperspectral signatures into HyperDB. Conversely, the system also allows spectral data to be imported from HyperDB, serving as the reference and training set for the classification function.

Land cover classification. The system performs land cover classification using a support vector machine method[15]. This procedure can tag each pixel in the region of interest in the hyperspectral image with a land cover class in the known land class set. The classification can be executed using the training set from either the same series of the hyperspectral imagery, or from spectra imported from HyperDB. An example of the results obtained for land cover classification is shown in Fig. 2.
3 Design and Implementation of HyperDB

The structure of the hyperspectral database of spectral signatures should provide easy access for the users, including importing, exporting, viewing, comparing, and searching operations. Spectral data need to be organized and categorized by appropriate metadata to meet the research requirements.[16]

3.1 Design principles

To elaborate on the basic requirements of the hyperspectral database, three fundamental principles in designing HyperDB are proposed.

3.1.1 Independence

The independence principle of HyperDB can be explained in two aspects. Firstly, the spectral library files are independent from HyperDB, so that the data and software do not rely on each other. In this way, the data can be utilized by other softwares designed to use the data, and the updates and modifications of the softwares should not affect the data. Secondly, the compulsory tagging of every hyperspectral signature with a certain land-class guarantees the independence between data of different land-classes. This ensures that the HyperDB is indexed exclusively by the land-class tag and hence provides a convenient method for classification and other categorized operations.

3.1.2 Data safety

The intactness and security of the spectral library files must be ensured during any operation. The incidental data safety check must be conducted prior to any operation occurring in both read and write directions. Read transactions should check the validity of the format and content of the file to be read, as well as its homogeneity against the data that already exist in HyperDB. Write transactions must protect the target file before HyperDB performs an inspection of the data to be written.

3.1.3 Compatibility

Given the wide use of hyperspectral remote sensing and imagery, HyperDB should be compatible with commonly used software and operating system, and spectral library files need to be compatible with the universal standard format used in remote sensing area.

3.2 Data flow

HyperDB is designed to manage and view the spectral library files and provide data to the hyperspectral image processing system. This intent grants HyperDB a pivotal position in the data flow structure, as illustrated in Fig. 3.

From Fig. 3, HyperDB has complete access to the spectral library files, which includes spectral library data files and spectral library metadata files, depicted in detail in Section 3.3. HyperDB has permission to read and write spectral library files, enabling one to view and compare different spectra in its workspace, and modify existing spectral data upon safety inspection.

Peripheral applications, such as the embedded hyperspectral image processing system shown in Fig. 3, have no direct access to the spectral library files. Instead, they can only retrieve desired data from HyperDB. The data retrieval process starts with the application sending an operation request to HyperDB, where the latter responds to the request and reads data from the spectral library files, then exports the required spectral data (tagged by the land-class) to the application, satisfying corresponding operation request.

The data sources of HyperDB are multiple, as previously outlined. In most cases, data retrieved by handheld instruments from in situ ground observations are often organized as spectral data of separate single pixels, while data retrieved by airborne and spaceborne sensors are often organized in a regional manner, as in hyperspectral imagery. The former case corresponds to the external source spectral data in Fig. 3, which can be directly imported to HyperDB after tagging with certain land-classes, and can then be viewed and written to the spectral library files. The latter case is recommended to be pre-processed in the hyperspectral image processing system before the spectral data of certain pixels are extracted and tagged, then imported into HyperDB.

Fig. 3 General data flow diagram with HyperDB in the pivotal position.
HyperDB clearly plays a mediator role in the data flow structure, separating the spectral library files from the peripheral applications as well as the external source spectral data. This structural characteristic abides by the independence design principle of HyperDB.

### 3.3 Spectral library files

Spectral library files are the objects that are managed and categorized by HyperDB. These files are exclusively recognized by the land class tag, and an ancillary band information label, so that data from different land class categories can be rigidly separated, following the independence design principle of HyperDB.

Two pairwise spectral library files are distributed to each land-class with specific band information. Figure 4 depicts the process of HyperDB responding to an operation request sent by the user or peripheral application.

**Spectral metadata file.** This file stores the mutual information of the land-class. It has the filename extension of `.hdr`, and contains the key words listed in Table 1. Spectral metadata files are stored in text format, and organized in the same way as the standardized format of ENVI, a commonly used geospatial analysis and spectral image processing software. Although some of the key-words do not already exist in the standardized format, it does not affect the compatibility design principle of HyperDB, since customized key-words will automatically be ignored when the file is read by ENVI. Actually, the key-words listed in Table 1 are mostly customized by HyperDB, and there are more areas of standardized key-words recognizable by ENVI and other popular softwares in spectral metadata files.

**Spectral data file.** This file stores every spectrum in the land class. It is assigned with the filename extension of `.sli`, and compressed into a binary file, designed to save memory space. Each spectrum in the land class, namely the reflectance value at each band arranged consecutively, is further placed next to each other.

### 3.4 File safety inspection

The file safety inspection procedure is highlighted with the dashed rectangle in Fig. 4. It contains two inspection steps, executed on different spectral library files.

**Validity inspection** is first performed on the spectral metadata file immediately after HyperDB reads the spectral library files of the requested land-class. It ensures the spectral metadata file is valid by checking every key-word area is filled with content of the correct format, and is also checked by some logical integrity verification. For example, the number of bands should be identical to the number of elements in the wavelength array.

**Homogeneity inspection** is next performed on both the spectral metadata and spectral data files. It mainly consists of a memory space verification between the two spectral library files, that is, ensuring the length of the binary spectral data file is identical to the product of the number of bands multiplied by the number of spectra in the spectral metadata file.

The validity inspection is performed prior to the homogeneity inspection, because the spectral metadata file is always significantly smaller than the spectral data file in size; hence, the calculation time cost of validity inspection is trivial compared to that of homogeneity inspection. HyperDB will assertively deny access to the spectral library files whose metadata file fails the validity inspection, with no further homogeneity inspection required.

If the requested spectral library files pass both inspections, they are permitted to be called into the HyperDB workspace, or alternatively exported to peripheral applications. However, if the operation ultimately cast a request to update the spectral...
library files, a safety re-inspection must be performed in a similar two-step manner, validating that the modification of the spectral library files will not violate the existing files.

4 Principal Functions of HyperDB

Based on its design and direct access to the spectral library files, HyperDB has several functions related to spectra processing and management, some of which are introduced below.

4.1 Viewing, comparing, and resampling spectra

Upon passing the file safety inspection, the spectral library files can be read into the HyperDB workspace. After analyzing the metadata files, their basic information is displayed in the upper left dialog area, as shown in Fig. 5 after analyzing the metadata files. When a land-class is selected, its spectra will appear in the lower left area. The spectra can be viewed in the upper right area, while accurate values at arbitrary band can be read by the tracking cursor, depicted as the white dashed crossed lines in Fig. 5.

If multiple spectra are selected to be viewed, the user can compare them, either from an identical or different land-class. Each unique land-class will be assigned a corresponding color to distinguish different land classes. The tracking cursor always automatically locks to the nearest point in the monitoring spectra, displaying its coordinates in the text boxes in the lower right area.

HyperDB also supports the resampling of spectra, for those cases in which the band region or sampling rate of interest does not coordinate with the band information in the spectral library files. This resampling includes both up-sampling and down-sampling, and is implemented by a linear interpolation method. Resampled spectra can then be exported into other applications for further use.

4.2 Importing spectra and writing to the spectral library files

As mentioned in Section 1, HyperDB allows importing spectral data from multiple genres of sources. Irrespective of which source will be used, all spectra must be tagged with the land-class and ancillary band information while being imported, to ensure their validity and make them possible to pass the re-inspection before being written to the spectral library files. If the tagged land-class already exists, these spectra will be merged into corresponding existing spectral library files. If not, new spectral library files will be created using the information of these spectra.

Two types of data sources have been clarified in Section 3.2. In HyperDB, the single spectrum data file with standardized format can be directly imported, while the hyperspectral imagery data are recommended to be pre-processed in the hyperspectral image processing system. The pixels to be imported can be selected in the false-color image, displayed as red crosses in Fig. 6. After importing into HyperDB, these spectra can be viewed, compared, and resampled, as described in Section 4.1, and written to the spectral library files upon successfully passing the file safety re-inspection.

4.3 Exporting spectra

To expand the adaptability of HyperDB, the spectra should be able to be exported to other peripheral
applications to satisfy their requirement.

For example, in the land cover classification procedure in the hyperspectral image processing system, after deciding the list of possible land cover categories in the region of interest to be classified, the user should select these categories from HyperDB and export the spectra of these categories into the image processing system, which will perform the classification using these spectra as the training set and provide results in Fig. 2. The HyperDB interface for this operation is shown in Fig. 7.

Note that there exists a potential problem regarding band matching. Data exported from HyperDB are completely compatible with the peripheral applications if their band information is identical to those sent with the request. If not, the data will be resampled before exporting. However, only the coincident range of bands between the HyperDB spectral library files and the request from the peripheral applications is able to be resampled and exported, since HyperDB does not support extrapolation due to its unacceptable loss of accuracy.

5 Conclusion

A hyperspectral database that categorizes related spectral library files by land-class, named HyperDB, is designed and implemented. HyperDB attempts to provide a reliable and expandable database under the fundamental design principles of independence, data safety, and compatibility. HyperDB was originally designed to collaborate with an embedded hyperspectral image processing system, performing mainly land cover classification procedures. However, more peripheral applications in different platforms can definitely utilize HyperDB, provided they support the standardized spectral library data file format. This adaptability by other applications will enable the usage of HyperDB in such tasks as land object and land use classification, vegetation or mineral categorization, etc.

Since HyperDB supports numerous data sources, it can serve as an intermediary role between in situ ground observation, airborne and spaceborne sensors. Data from multiple sources can be integrated, compared, and merged if they share the same land-class; hence, providing better data abundance and heterogeneity, which is the ultimate target of building a database.

References

[1] J. Solomon and B. Rock, Imaging spectrometry for earth remote sensing, *Science*, vol. 228, no. 4704, pp. 1147–1152, 1985.
[2] A. F. Goetz, Three decades of hyperspectral remote sensing of the earth: A personal view, *Remote Sensing of Environment*, vol. 113, pp. S5–S16, 2009.
[3] F. D. Van der Meer, H. M. Van derWerff, F. J. van Ruitenbeek, C. A. Hecker, W. H. Bakker, M. F. Noomen, M. van der Meijde, E. J. M. Carranza, J. B. de Smeth, and T. Woldai, Multi-and hyperspectral geologic remote sensing: A review, *International Journal of Applied Earth Observation and Geoinformation*, vol. 14, no. 1, pp. 112–128, 2012.
[4] K. Manjunath, A. Kumar, M. Meenakshi, R. Renu, S. Uniyal, R. Singh, P. Abuja, S. Ray, and S. Panigrahy, Developing spectral library of major plant species of western himalayas using ground observations, *Journal of the Indian Society of Remote Sensing*, vol. 42, no. 1, pp. 201–216, 2014.
[5] P. Hyde, R. Dubayah, W. Walker, J. B. Blair, M. Hofton, and C. Hunsaker, Mapping forest structure for wildlife habitat analysis using multi-sensor (lidar, sar/insar, etm+, quickbird) synergy, *Remote Sensing of Environment*, vol. 102, no. 1, pp. 63–73, 2006.
[6] R.-G. Zhu, B.-X. Ma, Z.-J. Gao, and J.-B. Ge, Research progress in nondestructive detection of livestock product quality based on hyperspectral imaging, *Laser & Infrared*, vol. 10, p. 003, 2011.
[7] R. Zomer, A. Trabucco, and S. Ustin, Building spectral libraries for wetlands land cover classification and hyperspectral remote sensing, *Journal of Environmental Management*, vol. 90, no. 7, pp. 2170–2177, 2009.
[8] A. T. DePersia, A. P. Bowman, P. G. Lucey, and E. M. Winter, Phenomenology considerations for hyperspectral mine detection, in *SPIE’s 1995 Symposium on OE/Aerospace Sensing and Dual Use Photonics*, International Society for Optics and Photonics, 1995, pp. 159–167.
[9] Y. Fan and H. Ma, Forest-fire smoke detection method based on video signal, (in Chinese), *Journal of Tsinghua University (Science and Technology)*, vol. 55, no. 2, pp. 243–250, 2015.
[10] A. Baldridge, S. Hook, C. Grove, and G. Rivera, The aster spectral library version 2.0, *Remote Sensing of Environment*, vol. 113, no. 4, pp. 711–715, 2009.
[11] S. Bojinski, M. Schaepman, D. Schläpfer, and K. Itten, Specchio: A spectrum database for remote sensing applications, *Computers & Geosciences*, vol. 29, no. 1, pp. 27–38, 2003.

[12] A. Hühni, J. Nieke, J. Schopfer, M. Kneubühler, and K. Itten, 2nd generation of RSL’s spectrum database “SPECCHIO”, presented at 10th Intl. Symposium on Physical Measurements and Spectral Signatures in Remote Sensing, Davos, Switzerland, 2007.

[13] A. Hueni and M. Tuohy, Spectroradiometer data structuring, pre-processing and analysis—An IT based approach, *Journal of Spatial Science*, vol. 51, no. 2, pp. 93–102, 2006.

[14] B. Guo, S. R. Gunn, R. Damper, and J. Nelson, Band selection for hyperspectral image classification using mutual information, *Geoscience and Remote Sensing Letters, IEEE*, vol. 3, no. 4, pp. 522–526, 2006.

[15] T. Kavzoglu and I. Colkesen, A kernel functions analysis for support vector machines for land cover classification, *International Journal of Applied Earth Observation and Geoinformation*, vol. 11, no. 5, pp. 352–359, 2009.

[16] A. Hueni, J. Nieke, J. Schopfer, M. Kneubühler, and K. I. Itten, The spectral database specchio for improved long-term usability and data sharing, *Computers & Geosciences*, vol. 35, no. 3, pp. 557–565, 2009.

Yizhou Fan received the BEng degree from Tsinghua University, China, in 2013. He is currently a PhD candidate at the Department of Electronic Engineering, Tsinghua University, China. His research interest includes image processing, remote sensing, and machine learning.

Ding Ni received the BS degree from Huazhong University of Science and Technology, China, in 2012. He is currently a PhD candidate at the Department of Electronic Engineering, Tsinghua University, China. His research interest covers remote sensing image processing, hyperspectral data classification, pattern recognition, and machine learning.

Hongbing Ma received the PhD degree from Peking University, China in 1999. He is currently an associate professor with the Department of Electronic Engineering, Tsinghua University, China. His research interests include image processing, pattern recognition, and spatial information processing and application.