Spectral Database Design and Development

Robert L. Fischer
Jeffrey G. Ruby
J. Ponder Henley

U.S. Army Topographic Engineering Center
7701 Telegraph Road
Alexandria, VA 22315-3864
fischer@tec.army.mil

Abstract

A spectral signatures database has been developed to support the use of spectral sensing for remote sensing applications. The database was developed to efficiently process and archive signatures and associated attribute information in support of data obtained from all regions of the electromagnetic spectrum including ultraviolet, visible, near-infrared, thermal infrared, and fluorescence. The database schema was designed to incorporate all sources of signature data including, laboratory equipment, field radiometers, and imaging spectrometers. Other associated metadata includes, project history, personnel information, collection location, date and time of collection, equipment parameters, reference documents, photos, and imagery. The database itself was developed with entirely off-the-shelve products and at completion will exist in both stand-alone and web-based versions. A full suite of metadata search and filtering utilities have been included to allow a user to quickly locate and extract signatures of interest. Application tools to perform numerical analysis are currently under development using distributed Matlab engine code. Currently available application tools include two- and three-dimensional visualization, signature statistics, and basic spectral band selection. Training tools and advanced signature processing are also under development.

Key Words

Remote Sensing, Spectral Signatures, Database, Reflectance, Thermal

Introduction

With the development of new multi- and hyperspectral remote sensing systems within the last decade, the need for high quality, well documented spectral signature libraries has been noted (Ben-Dor *et al*., 1999; Clark, 1999). Future imaging spectrometers will be designed to have the capability to obtain imagery, perform atmospheric correction, and compare pixels signatures with library signatures in realtime. It is obvious that systems, which perform spectral matching (both real and non-realtime) depend upon high quality spectral data. However, what can be just as important (and is often unavailable or discounted) is the descriptive information about the spectral data itself, also known as the metadata. Metadata would include information such as who collected the spectra, with what instrument, the date, time, collection protocol, etc. Additional information could and should describe the target object in some detail. For instance, if the feature of interest is vegetation, information such as *Genus*, *species*, height and phenology would be important. A key point to remember when deciding what metadata is pertinent is to consider the end-user requirements. If spectral data has been collected of multiple grass species in support of a effort to detect noxious weeds, to describe the spectra only as “grass” would be inadequate.

A benefit to developing a spectral database is the capability of linking spectral data from different sources. Over the past 20 years, the U.S. Army Topographic Engineering Center (TEC) has been collecting visible-near infrared, thermal and fluorescence signatures from a variety of laboratory and field instruments. This data was typically organized into project specific reports (Satterwhite & Henley, 1991) and existed on an assortment of media including lab books, floppy disks, zip disks, CD-ROMs, and internal hard disks. To perform what would seem to be a relatively simple task, such as extracting all signatures of deciduous vegetation could actually take several days due to searching through the storage media and reformatting data. With the implementation of the TEC spectral database, what may have required several days is now reduced to seconds. Of course, some effort is expended transferring the historical signatures into the database, but this time is usually well spent when considering the benefits of reviewing the data, and the functionality provided by the database. Additionally, new data are easily imported into the database by utilizing batch utilities developed for this purpose.

As a spectral library grows larger, the need to categorize and separate signatures becomes important. By collecting and utilizing signature metadata, spectral libraries (i.e. lists of spectra) are transformed into spectral databases. This is a critical capability not only for data access, organization and archiving, but also to assist in data dissemination between groups and agencies, which utilize spectral data. Ideally, a single/central signature database would be developed that could be accessed by users worldwide through the Internet. To accomplish this goal, groups that collect, process, and utilize spectral data would have to agree on standards for both signature content and the metadata information. The Jet Propulsion Laboratory (JPL) has implemented a version of an online spectral database with its ASTER spectral library (http://speclib.jpl.nasa.gov). The ASTER library is a composite of three separate spectral libraries from the U.S. Geological Survey, Johns Hopkins University, and JPL respectively. The ASTER library consists of roughly 2,000 signatures with the majority of rocks and minerals. Several agencies are pursuing the objective of formalizing metadata standards for spectral data. However, it is unlikely in the near future that a set of standards will be produced that all in the remote sensing community will support. In the intermediary, TEC has decided to develop its own version of a spectral database using metadata standards that TEC has incorporated over the years. The intent is not that the TEC format or software should become a standard, but that it be used as an example of a typical database, a platform in which to solicit comments and suggestions, and as a foundation on which to build.

Spectral Signatures Database Description

The current version of the TEC Spectral Signatures Database (SSDB) runs on Windows 95/98/NT systems. It requires two pieces of software: 1) Matlab 5.2 or higher and 2) the SSDB executable. The current version of the SSDB software has not implemented an install-shield because of the necessity of changing the path within Matlab, and the possible modification to the operating system's ODBC setup. The SSDB software provides the interface between the user, Matlab and a backend database (currently SQL Anywhere). The connection to the backend database and communication with Matlab are transparent to the user. The SSDB was developed using the PowerBuilder application software. PowerBuilder is an off-the-shelf development tool that allows the programmer to construct a user interface quickly and easily. PowerBuilder was chosen specifically for its ease of use, its ability to connect to all major commercially available databases, and for its ability to transition stand-alone applications to a web-based environment.

The SSDB software uses Sybase SQL Anywhere as the backend database. This database was selected because it was supplied with PowerBuilder, provides all required functionality, and because a runtime version can be distributed freely with PowerBuilder applications. PowerBuilder has the capability to access other common database packages, such as Microsoft Access and Oracle. However, unless the user plans to develop code directly accessing the database (i.e. bypassing the TEC interface), the backend database will be immaterial to the user.

All visualization and most application functions are performed by making runtime calls to Matlab. Matlab has sophisticated graphics and mathematical capabilities that are easily utilized. Stand-alone code could have been developed to perform these function independent of Matlab, but it was decided that the effort would be better spent further refining the user interface and database structure. While this current configuration requires that Matlab reside on the same PC as the database, we found that most agencies that were interested in using the database already owned Matlab, or would be willing to purchase a version to support the database. Future versions of the database will make calls to compiled Matlab code which will remove the requirement that the host PC have an installed version of Matlab.

The PowerBuilder-Matlab interface works as follows. The user starts the SSDB through the Windows Start button or through a shortcut as with any other Windows 95/98/NT application. The SSDB application then establishes a connection to the backend database. As the user selects a particular function, the SSDB software will query the database and pass parameters to Matlab for application or graphics support if required.

A large number of archiving, searching and visualization tools have been developed including:

- Adding Signatures Manually
- Adding Signatures in Batch Mode from ASCII Files
- Editing Signature Metadata
- Searching Signatures on Basis of Class or Attribute
- Viewing of Reflectance and/or Thermal Curve(s)
- Viewing an Average of Multiple Reflectance and/or Thermal Curves
- Viewing a 3-D Fluorescence Surface
- Viewing Slices of Fluorescence Surface(s)
- Calculation of Fluorescence Ratio(s)
- Exporting Selected Signatures to ENVI Spectral Library (.hdr/.sli) File
- Exporting Selected Signatures to ASCII File

Figure 1 below shows the SSDB software with its signature searching and viewing capabilities displayed. New tools are being developed on an as needed or requested basis.

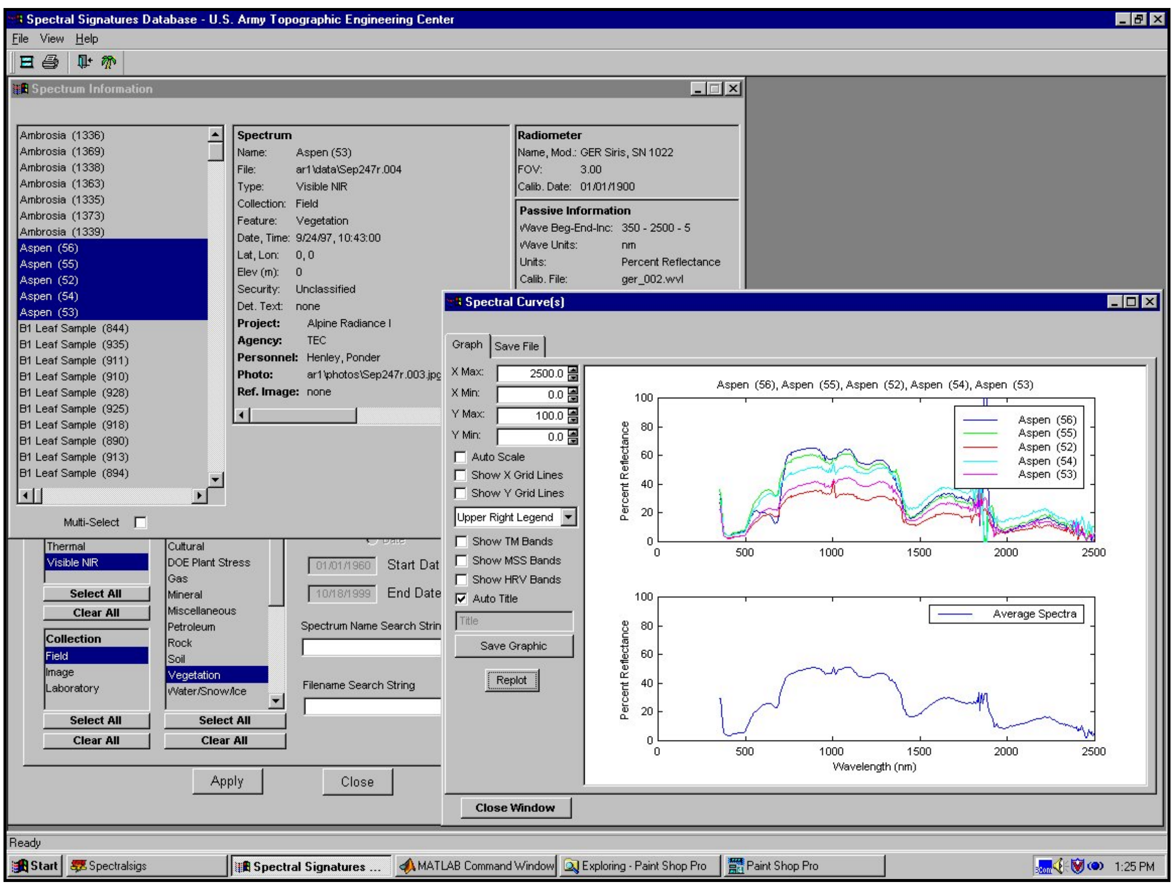

Figure 1. TEC Spectral Signatures Database (SSDB).

In the future, the database will be migrated to the World Wide Web. The Web version will have all the capabilities of the stand-alone version with the exception of addition and editing of signatures by client users. A central site will house the database where signatures can be added. The client user will need only a browser, such as Netscape or Internet Explorer to use the database on the web.

Schema

The organization of the metadata, sometimes referred to as the schema, is shown below in Figure 2. The larger font, bold names (such as Project), are referred to as classes, while the lower case fields (such as Name, Detailed Description) are referred to as attributes. All spectral signatures whether they are reflectance, thermal, fluorescence, field, laboratory, or image derived have Spectrum, Project, Personnel, and Agency class information. The signature will also have either a Radiometer or Imager table, and a Fluorescence Wavelength or Passive Wavelength table. All signatures will also be assigned to one of the eight feature classes on the left side of Figure 2. Photo and Reference Image classes are optional. For space reasons not all attributes are listed (especially attributes from the feature classes).

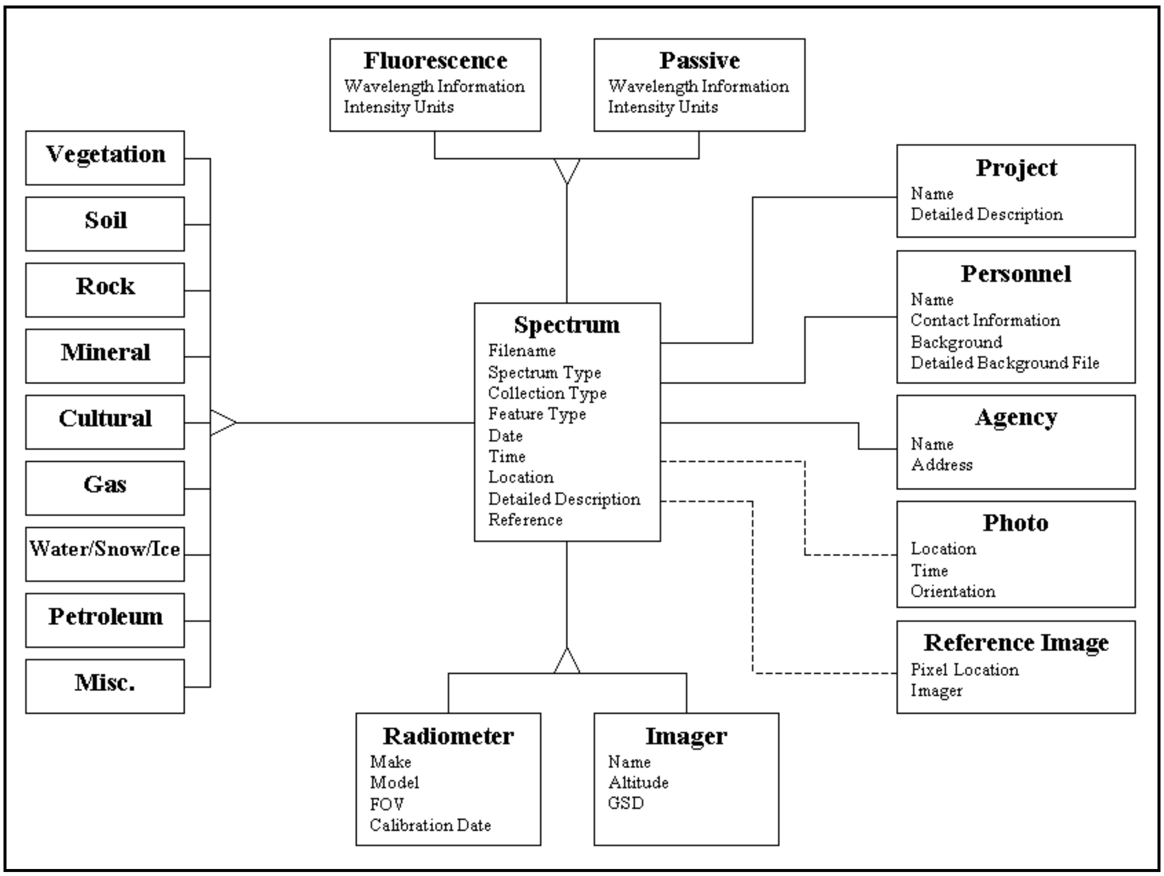

Figure 2. SSDB schema.

Data Sources

The database is designed to accommodate any type of spectral data from essentially any source. The current system is focused on three major spectral data types: (1) **Reflectance**, reflected energy over the visible-near infrared/short-wave infrared (V-NIR-SWIR) wavelengths, 0.4 to 2.5 microns, (2) **Thermal**, emitted and some reflected energy in the mid-wave (3-5 microns) and the long-wave (8-14 microns) infrared portions of the spectrum, and (3) **Fluorescence**, energy emitted from materials in the visible region (0.4-0.8 microns) when excited by energy of shorter wavelengths.

Reflectance

The sources of the spectral reflectance data are laboratory spectrophotometers, field-deployed spectroradiometers, and airborne or spacebased imaging spectrometers. Laboratory measurements have the advantage of a tightly controlled measurement environment, constant light source, and can have very high spectral resolution. Very precise spectral reflectance measurements can be obtained from the lab spectrophotometers. The disadvantage of the lab-based measurements is the typically very small sample area, and the inability to measure natural surfaces or materials *in-situ*.

A large portion of the data in the TEC SSDB was collected *in-situ* under solar illumination using field spectroradiometers. There are several common instruments in use for collecting field spectral reflectance data. The two most popular types are the ASD FieldSpec FR and the GER family of instruments. Field measurements of reflectance do not have the controlled environment of the lab instruments and must contend with varying illumination caused by the changing atmosphere and solar position. Spectra collected under these conditions will show the effects of atmospheric absorption in certain regions of the spectrum. Field measurement protocols must be developed and used to minimize the variation inherent in field collections. Obtaining consistent measurements of naturally variable materials in the field is difficult.

Another source of spectral reflectance signatures for the database is that collected by airborne imaging spectrometers. If the hyperspectral image data are calibrated and corrected to reflectance, the spectra at individual pixel locations can be extracted and stored in the database using any of several image data handling systems such as ENVI. Ground truth information to identify the material represented by the selected pixel location is essential. Spectra collected from an airborne sensor are most useful when developing detection and identification algorithms for that specific sensor.

Thermal

Like reflectance, thermal signatures can be obtained in the laboratory, in the field under natural conditions, and from multi-band thermal imaging systems. Laboratory measurements are made in highly controlled environments and produce clean, high-resolution thermal reflectance data over the range (nominally) of 2 to 30 microns. Emissivity can be calculated from the reflectance data using additional information (i.e. temperature of object). The disadvantages of laboratory measurements are the same as with Vis-NIR-SWIR, including very small sample size, and the difficulty of measuring natural surfaces. Thermal signatures in the database originated from lab spectrophotometers and are usually presented in the ranges of the atmospheric windows as observed by ground and airborne sensors (3-5 microns and 8-14 microns)

Field measurements of thermal spectra also must contend with natural variation and are inherently more difficult to obtain than laboratory measurements. Most field measurement devices are large and may require vehicle mounts and auxiliary power. There are few truly portable field thermal instruments. The instrument used for collecting the field data in the database is a Designs & Prototypes FTIR thermal radiometer, which has a 4-6 wavenumber resolution. Field data contain atmospheric absorption's and irregularities that are minimized in laboratory data.

Thermal spectra can also be obtained from the current airborne multi- and hyperspectral thermal imagers. At the present time there is very little of this data in the library.

Fluorescence

The fluorescence spectra imported into the database is almost exclusively obtained from bench top spectrofluorometers. These instruments produce 3-dimensional data. The user controls both the excitation and emission wavelengths and measures the fluorescence intensity. However, the database has also been designed to accept 2-dimensional fluorescence spectra from sources, such as the Fraunhofer Line

Discriminator and the Laser Induced Fluorescence Spectrometer. Figure 3 below shows an example of a fluorescence spectrum from the TEC's SLM 8000C.

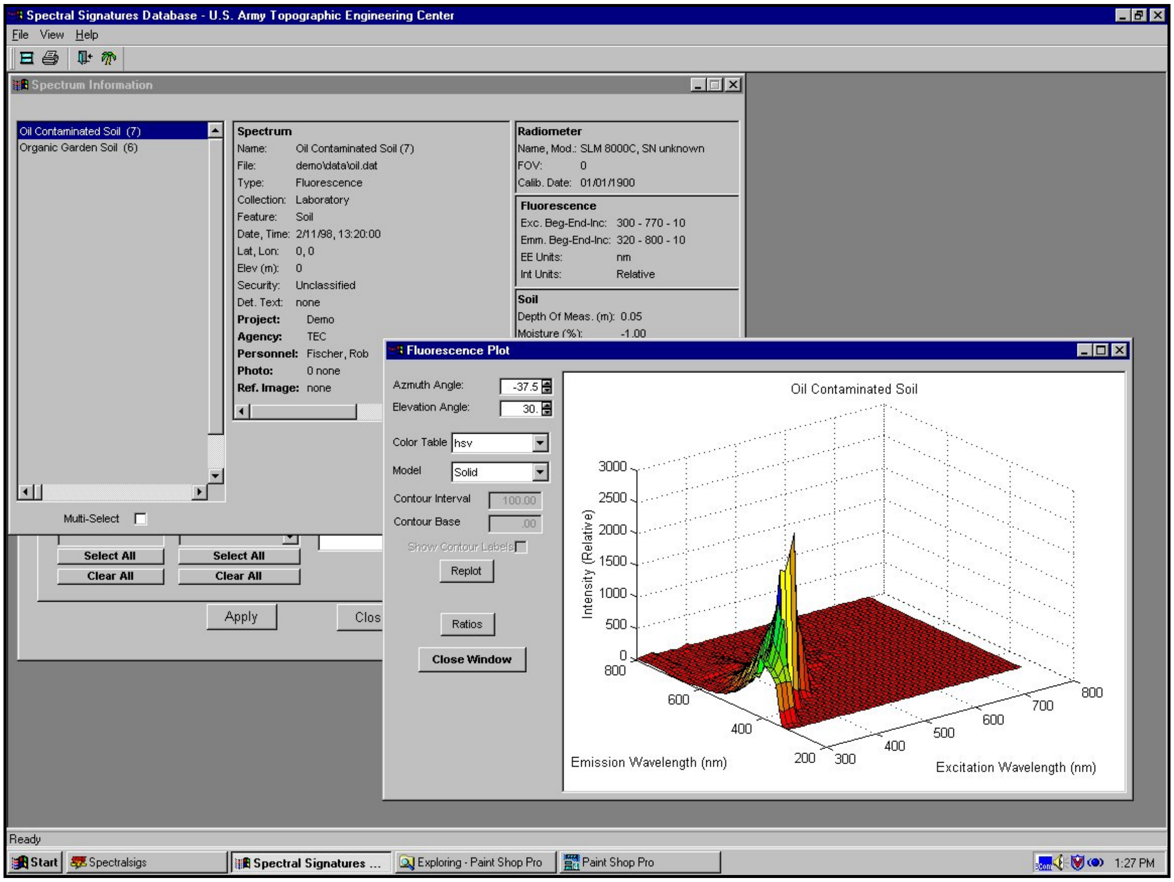

Figure 3. Fluorescence signature from the SSDB.

Outside Data Sources/Collaborators

As part of this effort, TEC has received substantial support from several sources. The Spectral Information Technology Application Center (SITAC) has provided guidance in the areas of metadata standardization, signature gaps, and schema design. Additionally, SITAC is actively coordinating with other producers of spectral data in an effort to expand the database outside the traditional remote sensing community.

The U.S. Air Force Warfighter1 Hyperspectral imaging satellite program (launch date planned for late 2000) is collaborating with TEC in the development and population of the database for use in their algorithm development and validation efforts. Warfighter1is planning to use the database to archive and access material target and natural background spectra to be collected at permanent test and validation sites located throughout the U.S.

Other groups and agencies have also contributed by either providing signatures or testing and evaluating various versions of the database. These collaborators include Virginia Commonwealth University, Virginia Institute of Marine Sciences, and Bechtel Special Technologies Laboratory. Of particular value have been the fluorescence signatures provided by Dr. Andrew Schuerger of Dynamac Corp., Kennedy Space Flight Center, Fla. These signatures were collected under the Department of Energy EM 50 Remote Sensing of Plant Stress Project, and have been used to develop and evaluate most of the fluorescence utilities within the database.

Discussion

Further effort is still needed in many areas. Arguably, the most important is metadata standardization. Several agencies (most notably SITAC and the National Imagery & Mapping Agency) are addressing this problem within the remote sensing community, but presently there is no consensus on metadata content. Normally, defacto metadata standards are developed within particular disciplines that utilize similar equipment. Trying to develop a single standard that applies to all disciplines which utilize remote sensing spectra will be difficult at best.

A second issue is the collection of the metadata itself. Obtaining detailed metadata for field spectra can be extremely time-consuming and expensive. Often, this requires that experts in a particular discipline be present during spectral collection. For instance, a botanist would need to be included on a team collecting vegetation spectra, while a geologist would be necessary when collecting mineral spectra.

Another concern is the assignment of feature class. For example, the spectral signature of "gravel" could have several meanings depending upon the discipline of the user. Within the current database structure, the "gravel" signature could be classified as a "rock", a "building material", or a "road "(and probably others). Unfortunately, a scientist interested in rock signatures may not search the database under building materials and thus, might miss signatures that may be of value. TEC's current policy is to classify each spectral signature as directed by the individual who collected the data. It may be desirable that a single signature be assigned to more than one feature class. In essence, the numerical spectral data would then have two or more sets of metadata. This would require considerable effort in reviewing possible candidates and may only be feasible (cost effective) for large, universally accessed databases.

References

Ben-Dor, E., Irons, J. R. & Epema, G. F., 1999. Soil Reflectance. In: *Remote Sensing for the Earth Sciences* (ed Rencz, A. N.), pp. 111-188., John Wiley & Sons, Inc., New York.

Clark, R. N., 1999. Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy. In: *Remote Sensing for the Earth Sciences* (ed Rencz, A. N.), pp. 3-58, John Wiley & Sons, Inc., New York.

Satterwhite, M. B. & Henley, J. P., 1991. Hyperspectral Signatures (400-2,500 nm) of Vegetation, Minerals, Soils, Rocks and Cultural Features: Laboratory and Field Measurements, U.S. Army Corps of Engineers, Fort Belvoir, VA.