Spectral signatures database for remote sensing applications

Jeffrey G. Ruby, Robert L. Fischer
U. S. Army Corps of Engineers/Engineer Research and Development Center

ABSTRACT

A spectral signatures database (SSDB) has been developed to support the use of spectral sensing for remote sensing applications. A broad range of data from multiple regions of the electromagnetic spectrum is supported, including ultraviolet, visible, near-infrared, thermal infrared, and fluorescence. Future plans include support of hemispherical reflectance data. A priority in the database development was schema flexibility. Data can be archived with minimum or detailed sets of attributes. Pre-defined, community-standard attributes are included, but custom attributes may be added to meet the specific project or data requirements. The database incorporates all sources of signature data, including laboratory equipment, field radiometers, and imaging spectrometers. The database can also incorporate and reference other metadata such as project history, personnel information, spatial information, temporal information, equipment parameters, reference documents, photos, and imagery. The database was developed with entirely off-the-shelf products and exists in both stand-alone and web-based versions. Searching and filtering utilities have been included to allow a user to quickly locate and extract signatures of interest. Currently available application tools include two- and three-dimensional visualization, signature statistics, and surface matching and comparison. Data exporting is also available, which includes the creation of commercial image processing spectral libraries.

Keywords: spectral signatures, remote sensing, database, VIS-NIR, reflectance, thermal, fluorescence

1. 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.[1,2] Imaging spectrometers soon to be available will have the capability to obtain imagery, perform atmospheric correction, and compare pixels signatures with library signatures in real-time. It is obvious that systems that perform spectral matching (both real and non-real-time) 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 were 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 of multiple grass species has been collected in support of an 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 25 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[3] 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 was expended transferring the historical signatures into the database, but this time was 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 critical. Collection and utilization of signature metadata transforms spectral libraries (i.e., lists of spectra) into spectral databases. This is a necessary capability not only for data access, organization and archiving, but also to assist in data dissemination between

Imaging Spectrometry VIII, Sylvia S. Shen, Editor, Proceedings of SPIE Vol. 48

groups and agencies that 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 Advanced Spaceborne Thermal Emission and Reflectance Radiometer (ASTER) spectral library (http://speclib.jpl.nasa.gov). TEC has recently implemented an online version of its spectral database. The functionality of the TEC web version is reduced compared to the PC-based stand-alone version. The web version currently contains approximately 1000 reflectance signatures.

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 everyone in the remote sensing community will support. In the intermediary, TEC has developed this 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, as a platform in which to solicit comments and suggestions, and as a foundation on which to build.

2. CONCEPTUAL OVERVIEW

The basic element within the database is an entry. Primarily, an entry will be a spectral signature, but other elements are also treated as entries, including (but not limited to) photographs, radiometers, and personnel. Each entry in the database has: 1) an entry name, 2) attributes associated with the entry, and 3) cross-references to other relevant entries in the database. Entries in the database are divided into classes. Current classes include spectrum, agency, project, personnel, photo, reference image, radiometer, imager and security. Each class has its own set of attributes, a few of which will be applicable to a given entry. Attributes are arranged in hierarchical fashion with sub-attributes arranged under more general attributes. This structure provides logical organization to the database and makes it simple for users to perform both broad and narrow searches.

An example entry from the database is shown in Figure 1. The entry selected is named "Aspen". As the name suggests, this is an entry for an Aspen tree. Attributes for this Aspen tree are shown on the left side of the figure. They include *Vegetation*, *Tree* (a sub-attribute of *Vegetation*), and *Broadleaf-Deciduous* (a sub-attribute of *Tree*). These attributes illustrate the hierarchical structure of the database.

The right half of Figure 1 contains a list of other relevant entries that are cross-referenced with this entry. This Aspen signature is cross-referenced with the following entries: GER 1022 (radiometer), Ponder Henley (personnel), Sep247r.003.jpg (photo), TEC (agency), and TEC Misc. Field Signatures (project). Each of these cross-referenced entries will have its own set of attributes. Entries may be cross-referenced with an unlimited number of other entries.

The stand-alone version was designed to allow creation of additional classes and attributes. This design allows users to customize their own stand-alone version of the database to fit their mission. The classes and attributes within the web-based version, which is hosted at TEC, cannot be modified by outside users.

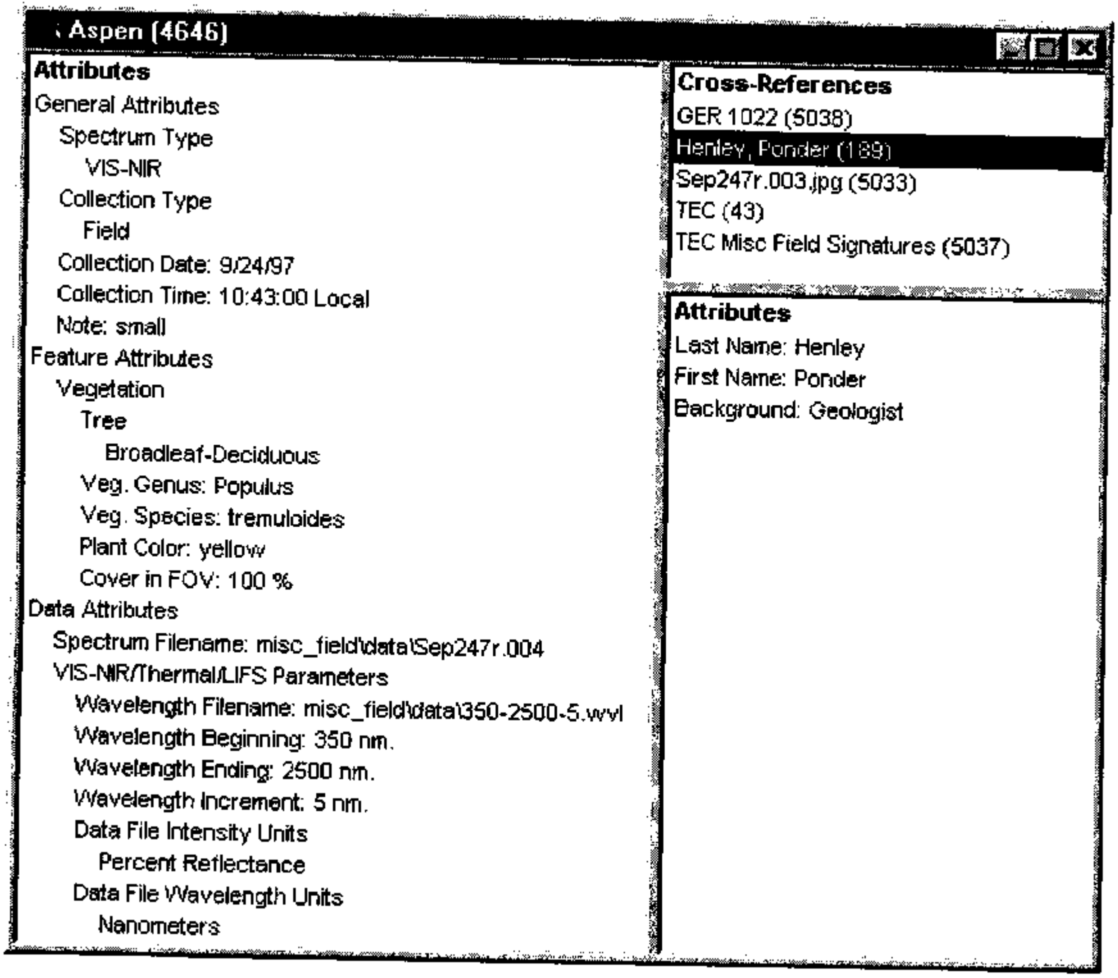

Fig. 1: Aspen tree metadata.

Typically a combination of attributes and/or cross-references is used to search for signatures. Figure 2 illustrates an example search. In this example, the user is interested in finding all the visible-near infrared signatures of trees from the project "TEC Misc. Field Signatures". This query retrieves entries with the *Tree* attribute that are cross-referenced with the "TEC Misc. Field Signatures" entry. The *OR* and *AND* toggle button can be used to search multiple attributes through Boolean logic. For example, if the user wanted to search for both trees and shrubs, an additional check mark would be placed in the *Shrub* box, and the toggle button would be left on *OR*. The right window is used to specify the search criteria and the left window contains a list of entries matching the query.

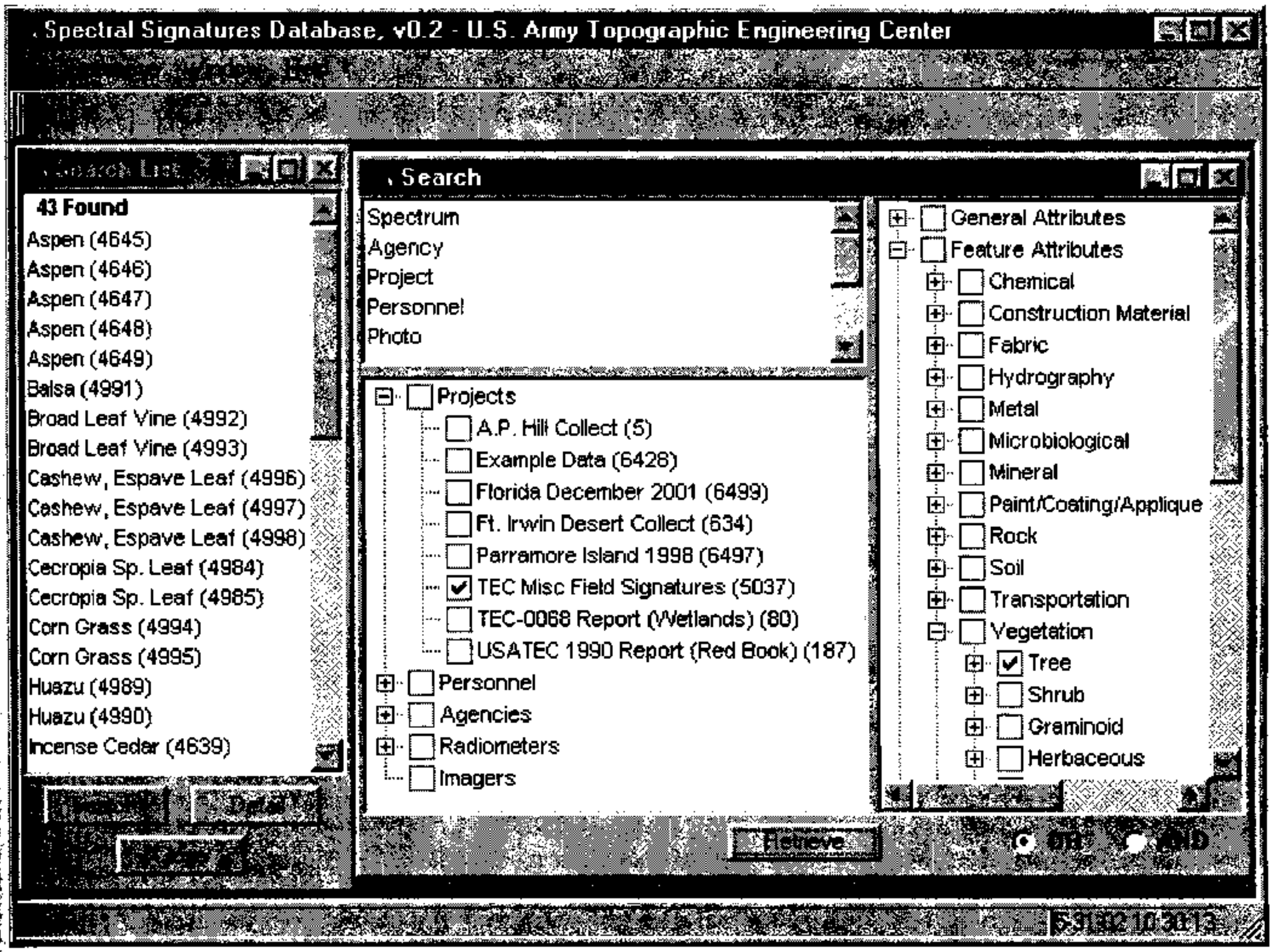

Fig 2: Example query.

3. VERSIONS

As stated previously, there are two versions of the database: a stand-alone version and a web-based version. The stand-alone version of the database runs on PCs with Microsoft Windows (98/NT/2000/XP) operating systems. SSDB application files and Microsoft Open Database Connectivity (ODBC) drivers are the only required software. ODBC drivers are included with Windows. The actual database is in Microsoft Access (.mdb) format, though Microsoft Access is not required. ODBC drivers interface to the database via a graphical user interface (GUI). The GUI was developed with PowerBuilder, Version 7.0.3, and PowerBuilder runtime files required by the database are included with the SSDB application files.

The software uses compiled Matlab executable modules for many of its graphical displays. Runtime Matlab libraries are the only required software, unlike previous versions, which required a full version of Matlab on the host computer. Runtime Matlab libraries are included with the application files. This arrangement relieves the requirement that each user have a valid version of Matlab on each machine hosting the stand-alone database.

A number of archiving, searching and visualization tools have been developed for the stand-alone version, including:

- Adding signatures manually

- Adding signatures in batch mode from ASCII files
- Editing signature metadata
- Searching signatures on basis of attribute and/or cross-reference
- Viewing of reflectance and/or thermal curve(s)
- Viewing an average of multiple reflectance and/or thermal curve(s)
- Viewing 3-D fluorescence surfaces
- Viewing slices of fluorescence surfaces
- Viewing signature data in ASCII format
- Calculation of fluorescence ratios
- Exporting selected signatures to ENVI Spectral Library (.hdr/.sli) files
- Exporting selected signatures to ASCII files

Figure 3 illustrates some graphical capabilities of the stand-alone version of the database. This figure shows the averaged response of four different Aspen signatures. The four component signatures and the averaged signature are displayed. Users may manipulate curve properties (line color, line weight, line style, curve name), change the plot title, modify legend placement, and toggle legend display on and off.

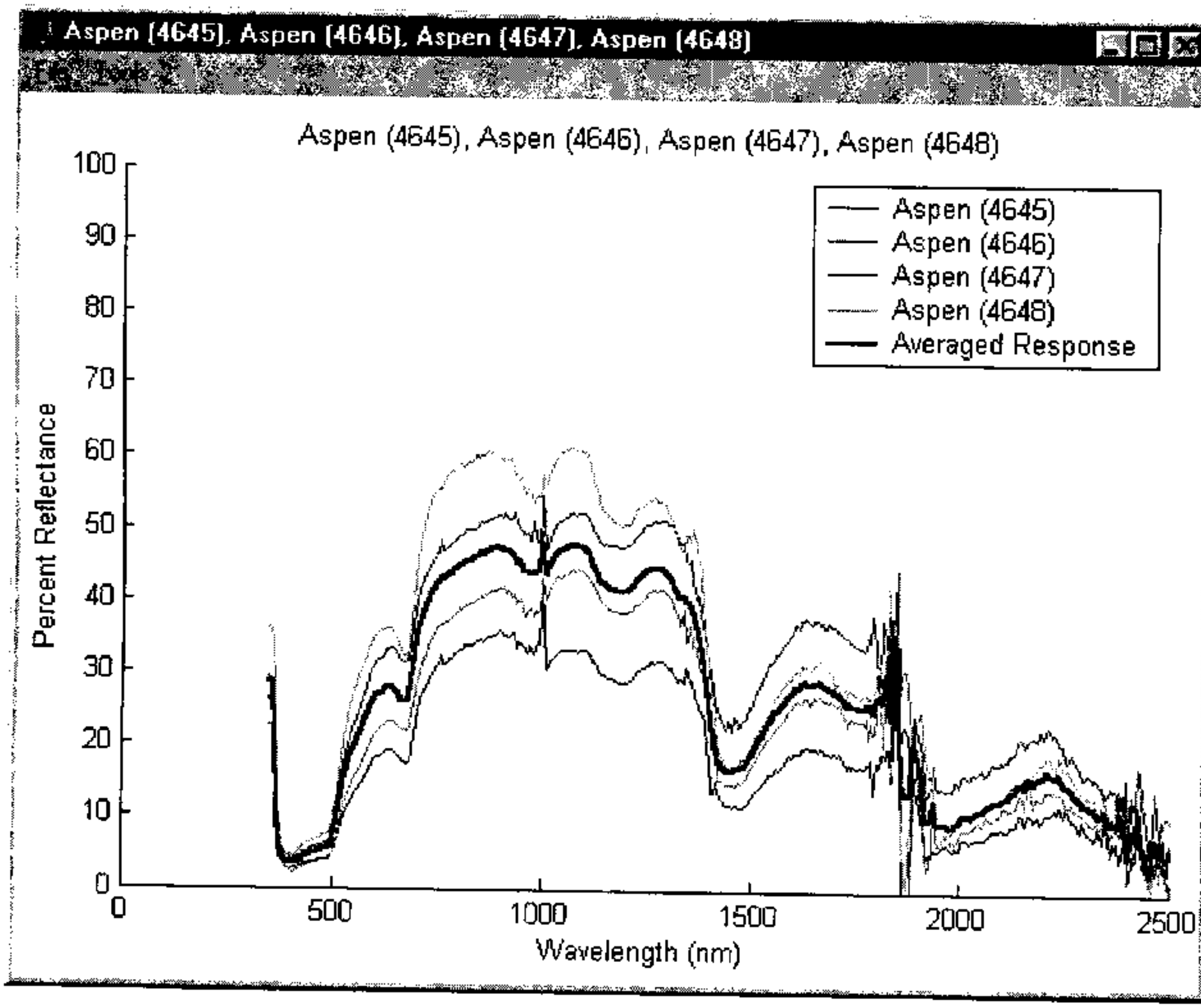

Fig 3: Graphical capabilities.

The web-based version requires only a browser such as Microsoft Internet Explorer or Netscape. This version provides access to data collected by the U. S. Army Topographic Engineering Center but does not allow users to add any of their own data. The web server uses a PowerBuilder application, Matlab and Matlab Webserver to serve and display data. The same Microsoft Access database used in the stand-alone version is used in the web version.

Tools included in the web-based version are:

- Searching signatures on basis of attribute and/or cross-reference
- Viewing of a reflectance or thermal curve
- Viewing a 3-D fluorescence surface
- Viewing of signature data in ASCII format

The web-based version is located at http://crunch.tec.army.mil/ssdb/index.html.

4. SCHEMA

Earlier work from a previous version of the database[4] revealed the importance of flexibility and simplicity in the database schema. It also revealed several problems with the existing database design. First, each class in the database had a predetermined, static set of attributes. When adding a new entry to the database, information had to be provided about specific attributes, even if that information was unavailable. Frequently, much of that information was unknown. This arrangement led entries such as "Not Applicable", "None", "Unknown", etc. in many of the fields. Entering unneeded information required unnecessary user effort and cluttered the screen during viewing.

A secondary and more pervasive problem was lack of sufficient attributes to satisfy user requirements. Adding attributes to the database was a difficult and cumbersome task. A third problem involved assigning attributes from different features to an entry. For example, a gravel road could be considered a transportation feature, a rock, or soil depending on circumstance. The existing version of the database did not have the capability to include attributes for all three features. Including information about all three features required three separate entries in the database. This procedure was unacceptable. A fourth problem was the complexity of the database structure. The previous structure had almost 100 tables. Maintaining tables and code to access data in these tables was complex and difficult.

After working with several users, it was obvious that no "standard" set of attributes would meet all user needs. Their needs were too varying, complex and in some cases too specific to allow for a "standard" set of attributes. The solution to the above problems was to redesign the database with a simpler, yet more flexible, structure. From a user standpoint, there was minimal change, but the few changes that were made were very important. These changes included:

- Allowing a variable number of attributes for each entry
- The capability to add new classes
- The capability to add extend attributes sets for existing and new classes
- Allowing an unlimited number of cross-references

From a programming standpoint, the database is vastly different. The new schema has only eight tables. The relationships between tables, combined with the fewer number of tables, makes it easier to maintain the database and easier to program software to access information in the database. This improved structure led to the inclusion of the important features listed above while allowing it to remain conceptually the same.

The entire database schema will not be presented here; only a brief outline of the three major tables in the database will be included. These tables are shown in Figure 4. They are: 1) the entry table, 2) the attribute table, and 3) the entry-attribute table. This figure also illustrates the relationship between these tables. The entry table maintains a list of all entries in the database and a unique identification number for each entry. The attribute table maintains the set of all attributes in the database and a unique identification number for each attribute. The entry-attribute table links attributes with specific entries by referencing entry identification numbers and attribute identification numbers.

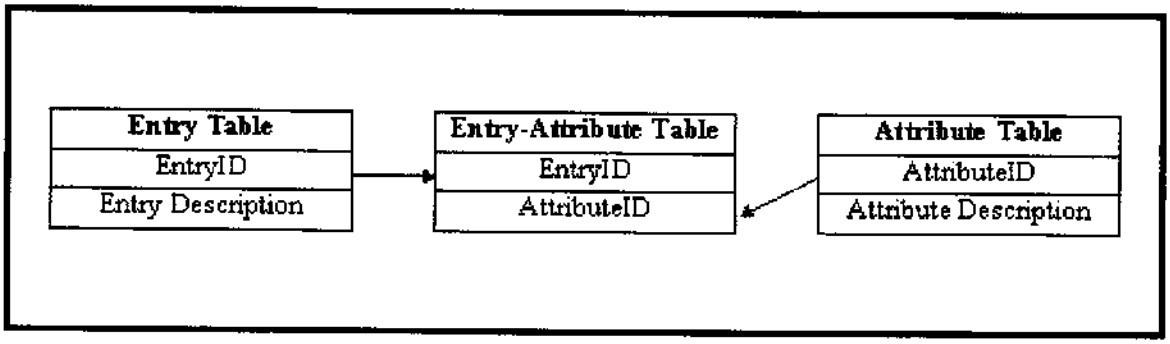

Fig. 4: Entry, attribute, and entry-attribute tables.

An example will better illustrate the relations between these tables. Figure 5 shows a simplified database; one with only three entries and seven attributes. This figure illustrates how entries and attributes are linked. "Aspen" has *Vegetation*, *Tree*, and *Spectrum Filename* attributes. "Paver Brick" has *Construction Material*, *Brick*, and *Spectrum Filename* attributes. "Currant" has *Vegetation*, *Shrub*, and *Spectrum Filename* attributes.

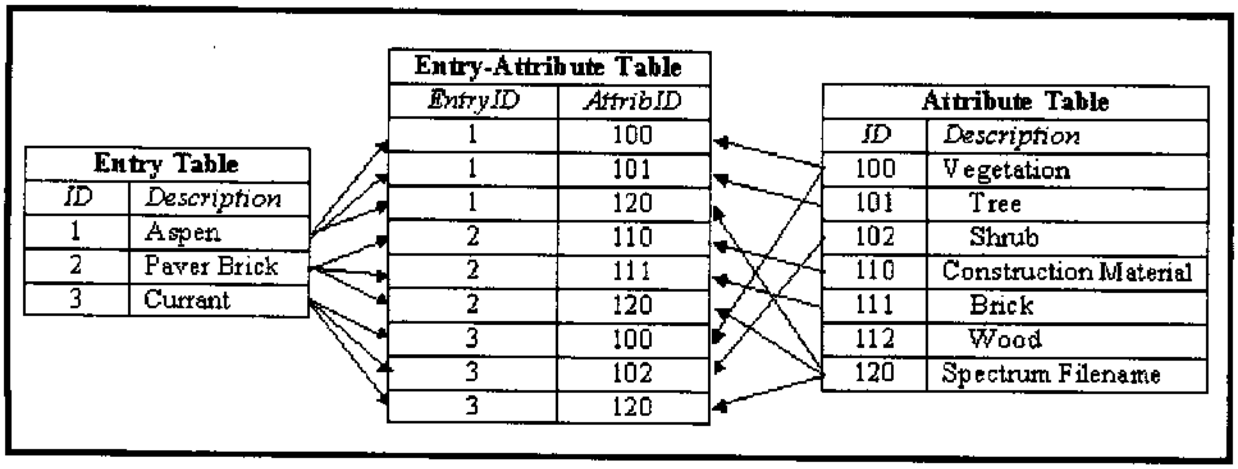

Fig. 5: Simple database.

Figure 5 is a simplified database. Each table in the real database contains more information than shown in Figure 5. Each record in the Entry table contains information about the class of the entry. Each record in the attribute table contains information about display order, parent attribute, and class to which attribute is relevant. Each record in the Entry-Attribute table may contain further attribute information as it applies to a specific entry. For example, the *Spectrum Filename* attribute is associated with all three entries below. The actual filename for each entry would be included in the Entry-Attribute table.

5. FURTHER WORK

Further work will focus on development of the web-based version of the database. Ideally, all functionality now found within the stand-alone version (with the exception of schema modification) would be supported by the web-based version. As an example, web-based exporting tools such as creating local ENVI .hdr/.sli files would be useful. For those interested in using the stand-alone version, automated tools that modify systems parameters during installation would ease this sometimes cumbersome process.

Application software under development includes the processing and visualization of hemispherical data (i.e., bi-directional reflectance), two- and three-dimensional signature matching, and more robust reference image manipulations.

Large amounts of multispectral, hyperspectral, and fluorescence imagery are currently being collected in conjunction with signatures. The capability to efficiently organize and access this data within the database structure is critical.

A final, and by no means trivial, effort is to link the TEC Spectral Signatures database with both internal and external data sources. As noted, several other groups and agencies are producing spectral databases and libraries in various forms. Ideally, a "standard" format could be settled upon that will be used to represent spectral signature data, but at a minimum, it is critical that each group publish its schema formats to allow easy interpretation by others.

REFERENCES

1. E. Ben-Dor, J. R. Irons, G. F. Epema, "Soil Reflectance", *Remote Sensing for the Earth Sciences*, A. N. Rencz (ed.), pp. 11-188, John Wiley & Sons, Inc., New York, 1999.
2. R. N. Clark, "Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy", *Remote Sensing for the Earth Sciences*, A. N. Rencz (ed.), pp. 3-58, John Wiley & Sons, Inc., New York, 1999.
3. M. B. Satterwhite, J. P. Henley, *Hyperspectral Signatures (400-2,500 nm) of Vegetation, Minerals, Soils, Rocks and Cultural Features: Laboratory and Field Measurements*, U. S. Army Corps of Engineers, Ft. Belvoir, VA, 1991.
4. R. L. Fischer, J. G. Ruby, J. P. Henley, "Spectral Database Design and Development", *1999 International Symposium on Spectral Sensing Research*, Las Vegas, NV, 1999.