

Acquisition and storage of multispectral material signatures – Workflow design and implementation

Stephan Irgenfried¹ and Jürgen Hock²

¹ Karlsruhe Institute of Technology
Institute for Anthropomatics and Robotics IAR
Intelligent Process Control and Robotics Lab IPR
Engler-Bunte-Ring 8, 76131 Karlsruhe

² Fraunhofer Institute of Optronics,
System Technologies and Image Exploitation
Fraunhoferstr. 1, 76131 Karlsruhe

Abstract This article describes design, hard- and software implementation of devices and a workflow for acquisition, processing, storage and analysis of multispectral material signatures. An approach for a consistent process from measurement to analysis is presented, utilizing standard formats for multi-dimensional image data. Finally applications are demonstrated that make use of the material signatures database.

1 Introduction

The “spectral fingerprint” of materials, which describes, how the material reflects or absorbs incident light at different wavelengths, is of use for many applications in science and industry, e.g. quality inspection, sorting, pharmacy or food freshness checks. All the applications have in common, that they need a base of material signatures data to be used for comparison, classification and analysis. Based on the ideas and the work presented in [1] we describe additional components of the workflow, the implementation of the whole tool chain, the data acquisition and storage infrastructure and the data analysis and access modules.

2 Related Work

In [2] Hüni et. al differentiate between file based spectral libraries and spectral databases which make use of a Database Management System (DBMS). Literature shows, that this assumption is not used consistently. There are spectral libraries with online access as well, e.g. [3]. In the view of Clark et. al [4], an ideal spectral library contains many materials and covers a large wavelength range together with metadata about the measurement. They created the USGS Digital Spectral Library³. In [5] Bojinski et. al introduced a change in infrastructure from pure file based libraries to using a relational database for their system called SPECCIO⁴. They describe the advantages of this architecture which have been picked up and extended by our work. Improvements to the SPECCIO database model were published by Hüni and Tuohy in [6]. In the same work they presented a database client software for data acquisition, processing and database access called SpectraProc. Ferwerda et. al implemented a similar architecture in [7], including a web-based interface and sample location view in Google Earth[®]. Their website⁵ was not accessible by Jan. 15th 2015. Some of the page can be accessed using way-back machine⁶. Compared to SPECCIO, SpectraProc and Ferwerda's work, our approach is a hybrid system which uses both, file based storage and a relational database. This is similar to the Common Hyperspectral Image Database Design CHIDB published by Tian et. al in [8]. Other databases accessible via a web-based interface are the ASU Spectral Library⁷, founded by Christensen et. al [9] and the ASTER spectral library⁸ by Baldrige et. al [3].

3 Workflow

Figure 10.1 shows the components of our workflow, starting with material measurement and ending with data analysis. For the repeatability of the measurements and to get results comparable to measurements

³ <http://speclab.cr.usgs.gov/spectral-lib.html>

⁴ <http://specchio.ch>

⁵ <http://www.hyperspectral.info>

⁶ <http://web.archive.org/web/20140103004905/http://hyperspectral.info/>

⁷ <http://speclib.asu.edu/>

⁸ <http://speclib.jpl.nasa.gov>



Figure 10.1: Workflow of the material signature acquisition and storage

taken by other institutions, we developed standard procedures to be executed for every measurement, e.g. white and black balance. With the help of a web-based application, the user is guided through the whole process reducing the risk of incomplete or inconsistent records in the database. To keep flexibility high in the whole process, we split it in several separate modules, which will be described in more detail later.

4 System Components

4.1 Data Acquisition / Material Measurement

Data acquisition is the fundamental part of the introduced workflow for multispectral data analysis and can be characterized by the following general requirements:

- capturing of samples for different kind of solid materials,
- repeatable capturing of broadband material spectra.

Considering these key points, a detailed material spectra analysis can be accomplished. Fulfilling the requirements, Fraunhofer IOSB Karlsruhe developed a data acquisition procedure and special technical infrastructure, shown in figure 10.2.

Several measurement stations are equipped with different kind of spectrometer bundles, to cover the characteristic bandwidth ranges:

- UV 240-400nm
- VIS 400-1000nm
- SWIR 1000-2500nm

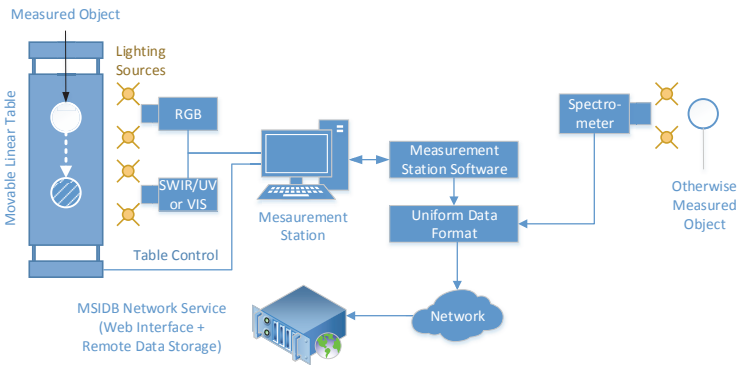


Figure 10.2: MSIDB technical environment infrastructure

The SWIR and UV spectrometer bundle includes a multispectral push broom scanner and an attached RGB scanner respectively. The attached RGB image is commonly used for image registration purposes. There-with material measurements from different measurement stations can be spatially registered with each other. The VIS measurement station covers already the RGB color space and requires no additional RGB image. After the data fusion from particular measurement stations a total bandwidth from 240nm up to 2500nm is reached.

The spectrometer bundles are mounted to the base frame of the measurement station, looking down on a linear movable table. Capturing of the material surface image is achieved by moving the table perpendicular to the line of sight of the cameras in synchronization with the camera clock. That requires fine mechanical characteristics and adjustment options, which can be achieved by using an electromechanical drive.

According to the physical characteristics, different kinds of light sources are installed: halogen lamps for SWIR, deuterium lamps for the UV measurement station. The VIS measurement station contains mixed fluorescent and halogen lamps with an additional Ulbricht sphere diffusor. The acquisition is accomplished by the measurement station software, which was developed to operate multiple spectrometer devices. Finally,

acquired material measurements are stored locally on the measurement station in the internal uniform data format, to be processed within the next workflow step.

4.2 Measurement Post-Processing

To reduce or preclude the influence of temperature or sensor aging, the base characteristics of the data acquisition devices are captured in separate steps.

As a first step, the dark current of the sensor is recorded with the sensor being protected from incoming light. This step is required to measure the actual dynamic range of the sensor. Together with the white balance step, both steps are also used to identify corrupt or defect pixels in the line sensor of the multispectral camera.

For the white balance, a two-substep approach is used. The white balance reference material Teflon is measured before every material sample to avoid sensor or environment drift between white balance and measurement. The white balance material itself is calibrated in regular intervals by comparing its reflectance properties with those of a SPECTRALON probe to compensate aging or usage caused degradation of white balance quality.

Data from black current and white balance measurements is used to correct the measured material image. The correction step corresponds to the flat-field correction technique described in [1]. Resulting image contains the sensor integration time independent reflectance of the captured material.

Finally, corrupted image pixels will be canceled out in the data. The position of this pixels is known by separate statistical analysis of specific camera sensor properties and is estimated from a large dark current and the white balance image sequence.

The measurement post-processing is being performed automatically during each data upload procedure.

4.3 Data Upload

The data upload procedure is performed through a web-based user interface. Multispectral image data captured by the measurement station

contains one or more objects of the material probe or even objects of different material types.

The images have to be annotated with metadata containing classification information for every pixel in the multispectral image data block. This separate step is performed using a semi-automatic, MATLAB[®]-based segmentation procedure. The user manually marks sample pixels belonging to the same class of material. Then the system automatically segments objects and background, based on an energy minimization function in the spectral domain.

Together with the corresponding mask file, the images are then uploaded to the database. During this step, the user also adds metadata like project number, environmental conditions or material description to the database record.

Furthermore the uploaded data can be accessed manually by the web-based user interface or from within applications by using the data access interface through the network.

4.4 Database

Database infrastructure

All acquired measurements are stored in the database together with the measurement metadata. As it contains a large amount of raw and pre-processed files, it is important to ensure fast access to the data, enough storage volume and metadata management flexibility. Therefore the database is splitted into two divisions – the SQL database and the blob storage, as shown in the figure 10.3. The SQL database is used to store the common metadata about projects, material lists, pixel classification annotations and measurement relationships. The binary data of the measurements is stored in the separate blob division. Each measurement blob contains a binary file and corresponding metadata file, which describes the binary file format.

The separation in data storage divisions allows fine-grained utilisation of hardware components, like mixing of solid-state and hard drives. An usefull option for fast and transparent caching of non-persistent data is utilisation of the virtual drive assigned to a physical RAM sector. This technique is commonly used for the internal data preparation routines during the upload. The database server itself utilizes the Intel Xeon[®]

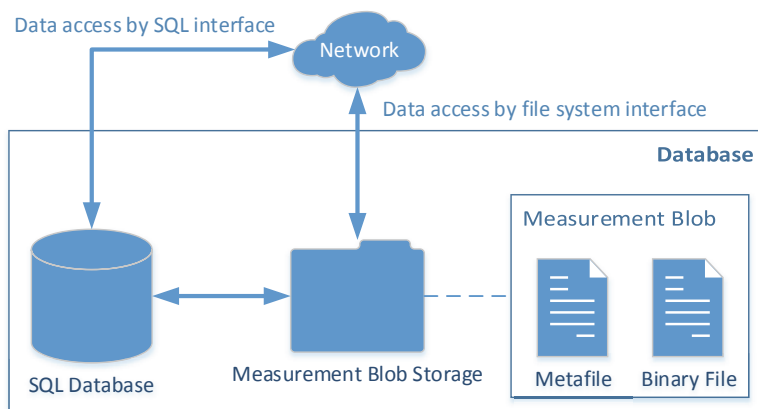


Figure 10.3: Database and data access overview

64-bit multi-core processor architecture. This allows a high degree of computing task parallelization for online data processing and analysis.

Data Model

In the relational database we keep the metadata about the measurements. In addition, project, campaign, devices and user management is implemented. For a maximum of flexibility, it is possible to add arbitrary key-value-pairs of information to each data record. A set of metadata about a measurement consists of

- Material information
Which material was measured and how was its' status, e.g. grain size, temperature, humidity
- Measurement information
Where and when did the measurement take place? Which device was used and who operated it? Which files belong to the measurement?

- Campaign information
To which campaign and project does the measurement belong?
Who is allowed to access the data?

A vast amount of metadata is stored to provide detailed information on how a measurement was carried out and which processing steps were applied to the data. We store file results of every processing step in the long term storage to provide access, analysis and possible error correction at every step.

4.5 Data Access

The measurement database provides a bi-directional interface to import and export data via network. Depending on the application needs there are two separate data access methods, as shown in figure 10.3.

The high-level data access is provided through the SQL-Server standard connection mechanisms, e.g. SQL-Server, ODBC or ADO. The flexibility of SQL allows to send complex queries from within applications like MATLAB[®] or programming languages like C#. One of the possible use case scenarios is precise searching of measurements by specific properties, like camera kind or acquisition location in a given time period.

The preprocessing and sharing of the binary blob data over the SQL interface is achieved by the Microsoft SQL CLR⁹ technique. This allows custom C# classes to be run inside the SQL database environment. The classes are called from within a regular SQL query, have read-write permissions on the physical hard-drive and return a SQL conform query response. Calls to SQL CLR applications are therefore transparent for the end user. One of the possible use case scenarios is computation of statistical metrics or comparison of different measured images.

The low-level data access is implemented as a file sharing interface. Using this interface requires knowledge of the internally used data format, but also allows additional functionalities like mass data modifications or data sharing via network.

Internally, an ENVI (ENvironment for Visualizing Images)-¹⁰ based data format is used to store measurement blobs, see also figure 10.3.

⁹ <http://msdn.microsoft.com/en-us/library/ms254498.aspx>

¹⁰ <http://www.exelisvis.com/docs>

Every measurement blob contains a metafile. It stores information on measurement properties, e.g.:

- Image size (height, width, spectral band count)
- Image value type (16bit, 32bit, signed, unsigned, complex)
- Integration or exposure time
- Short description and acquisition timestamp

The metafile allows quick reconstruction of image acquisition conditions, which is essential for measurement data import and export routines.

The binary file contains only the image data, according to the specified binary format. ENVI is a simple data format, which can hold common data types, real and complex values.

This uniform measurement data format rounds up the database access interface and is the basis for the multispectral data processing. On the one hand, it makes it possible to import measurements from custom spectrometer devices or other databases, as previously shown in figure 10.2. Using custom conversion tools, measurements from external data sources like the ASTER spectral library [4] can be imported into the database. On the other hand due to simplicity and flexibility of the ENVI data format complex data analysis, like spectral filter design or classifier design, can be easily realized in common programming environments.

4.6 Applications

- Visualization of Multispectral Data (ViMuDat)¹¹
This web-based application allows online access to the data in the multispectral database and data analysis. Materials can be searched in the database by name, project, customer, material and measurement properties. Spectral signatures of the materials found can be compared and visualized, several basic statistics algorithms, e.g. normalization and averaging, can be applied on the

¹¹ <http://msidb.iosb.fraunhofer.de/>

data. Also the expected spectral signature after applying an optical filter can be previewed, which gives the user a first impression whether different materials can be separated by using their spectral signatures. Figure 10.4 shows the user interface of ViMuDat.

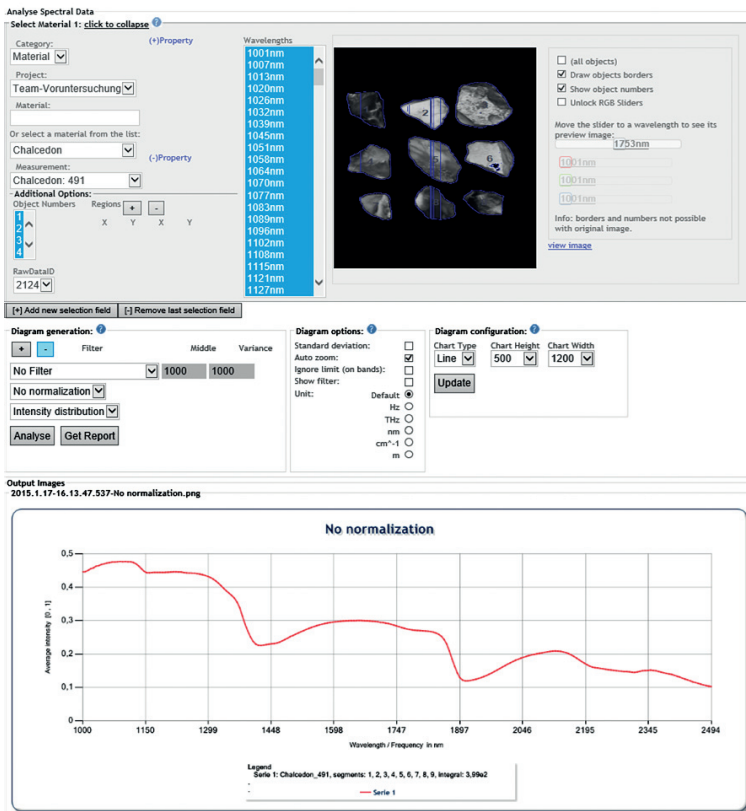


Figure 10.4: Visualization of Multispectral Data (ViMuDat)

The material selection area is located on the left side of the picture. On the right side, the image for the selected wavelength is shown. Page bottom shows the spectral signature of the selected material.

- **Spectral Unmixing**
For applications like demonstrated in [10] it is required to have a broad basis of material measurements to improve success rates for material classification using spectral unmixing. Our database supports these applications by providing direct access to the data for bulk queries.
- **QueryMe**
Implemented as web- and standalone-applications, QueryMe allows real-time classification of materials, e.g. with a handheld spectrometer, based on classifiers trained with the data from our multispectral database.

5 Summary

We presented a detailed view inside our infrastructure for acquisition, processing, long-term storage and analysis of multispectral material signatures. Material samples are measured in a wide spectral range, normalized by using white- and black-balance of the measurement devices, uploaded to the hybrid material signatures database and analyzed using online, web-based analysis or offline tools like ENVI[®] or MATLAB[®]. Our online data analysis includes spectral signature view as well as standard mathematical operations and optical filter preview for fast testing on discriminability of materials. User rights management ensures configurable access to data and analysis functions.

6 Outlook

Currently our infrastructure is extended on both ends. Mobile measurement devices are included in the workflow for easy sampling of materials and data upload to the database. Material classification and spectral unmixing are set up to create mobile applications for material identification, especially focused on handheld measurement devices. Our material signature fund is also continuously growing by numerous measurements of food samples, minerals, miscellaneous organic and inorganic materials.

7 Acknowledgements

We especially thank the students that spend a lot of creativity and effort in numerous projects over the last years in setting up and improving this infrastructure and applications as part of the projects *ViMuDat*, *Up-ToDatE*, *Bring2Light* and *QueryMe*.

References

1. S. Irgenfried and C. Negara, "A framework for storage, visualization and analysis of multispectral data," in *OCM 2013 - Optical Characterization of Materials - conference proceedings*, J. Beyerer, F. Puente León, and T. Längle, Eds. Karlsruhe: KIT Scientific Publishing, 2013, pp. 203–214.
2. A. Hüni, 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.
3. 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.
4. R. N. Clark, G. A. Swayze, A. J. Gallagher, T. V. V. King, and Calvinm W. M., "The u. s. geological survey, digital spectral library: Version 1: 0.2 to 3.0 microns: U.s. geological survey open file report 93- 592," 1993. [Online]. Available: <http://speclab.cr.usgs.gov/spectral.lib04/clark1993/spectral.lib.html#Abs>
5. S. Bojinski, M. Schaepman, D. Schläpfer, and K. Itten, "Specchio: a web-accessible database for the administration and storage of heterogeneous spectral data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 57, no. 3, pp. 204–211, 2002.
6. A. Hüni and M. Tuohy, "Spectroradiometer data structuring, pre-processing and analysis – an it based approach," *Spacial Sciene*, vol. 52, no. 2, pp. 93–102, 2006.
7. J. G. Ferwerda, S. D. Jones, and P.-J. Du, "A web-based open-source database for the distribution of hyperspectral signatures," in *Geoinformatics 2006: Geospatial Information Technology*, ser. Proceedings of SPIE, H. Wu and Q. Zhu, Eds. Seattle, WA and USA: SPIE, 2006, vol. 6421. [Online]. Available: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1295832>

8. L. Tian, N. Liao, and A. Chai, "Common hyperspectral image database design," in *SPIE Proceedings Vol 7513*, ser. SPIE Proceedings, SPIE, Ed., 2009, vol. 7513, pp. 75 132Y–75 132Y–9. [Online]. Available: [http%3A//proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=774930](http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=774930)
9. P. R. Christensen, J. L. Bandfield, V. E. Hamilton, D. A. Howard, M. D. Lane, J. L. Piatek, S. W. Ruff, and W. L. Stefanov, "A thermal emission spectral library of rock-forming minerals," *Journal of Geophysical Research*, vol. 105, no. E4, p. 9735, 2000.
10. M. Michelsburg and F. Puente Leon, "Spectral and spatial unmixing for material recognition in sorting plants," in *OCM 2013 - Optical Characterization of Materials - conference proceedings*, J. Beyerer, F. Puente León, and T. Längle, Eds. Karlsruhe: KIT Scientific Publishing, 2013, pp. 179–190.