



SCYLLARUSTM
Beyond Vision

SLZ Data Structure and File Format

For SLZ Version 0.9

Publication Date 5 May 2014

The HSZ File Format, the Scyven Application, the Scyllarus system, the Scyllarus Matlab Toolbox and the Scyllarus C++ API are owned and managed by National ICT Australia Limited (NICTA). For more information go to www.scyllarus.com.

Table of Contents

Introduction	3
Document Scope	3
Intended Audience.....	3
Using this Guide	3
Reporting Problems	3
References	3
HSZ Format Overview	4
Using HDF5 Fields.....	4
String Fields and Subfields	4
Numerical and Binary Fields and Subfields	4
HDR Field	5
Endmembers Field.....	5
Document History	6
Appendix 1. ENVI Header Structure	7
General Rules	7
Required Header Entries.....	7
Optional Header Entries.....	8
Typical ENVI Library file header	9

Introduction

The Scyllarus Hyperspectral Zipped (HSZ) format is a compressed format for storing Hyperspectral images.

This format is used both in the HSZ file and the HSZ data structure in the Matlab environment.

Document Scope

This document describes the Hyperspectral Library Zipped (SLZ) format as used in both, the Scyllarus HSZ file and the Scyllarus Matlab toolbox.

Intended Audience

This document is intended for researchers and software engineers who wish to read and understand the Scyllarus SLZ file format of the Scyllarus Matlab Toolbox data structure. It assumes a workable knowledge of Hyperspectral imagery and the methods generally used to process it and store it.

Using this Guide

This guide is intended as a reference. Readers wishing to understand the Scyllarus processing pipeline and the theory behind it should refer to www.scyllarus.com

Reporting Problems

If you find a problem in this document, you can report it to NICTA using the Scyllarus web page. See <http://feedback.scyllarus.com>

Note that NICTA cannot commit to fixing all reported problems and this document is provided as is.

References

[1] <https://www.exelisvis.com/ProductsServices/ENVI/ENVI.aspx>

[2] <http://www.hdfgroup.org/HDF5/>

SLZ Format Overview

The SLZ Structure is used by Scyllarus in two ways.

- The Hyperspectral Library Zipped (HSZ) file format is used by Scyllarus to store compressed spectral libraries to disk using the HDF5 file format [2].
- SLZ is a data structure used in the Scyllarus Matlab toolbox.

The file format and the data structure have an identical hierarchical structure.

This has two advantages. Firstly, it allows for the Scyllarus Matlab toolbox, C++ API and Scyven to be compatible in the manner they handle hyperspectral data. Secondly, SLZ files can be opened and saved in a wide variety of platforms and operating systems using the libraries provided by the HDF group [2].

SLZ files have a structure based upon two fields. These are

- **HDR:** A field containing ENVI compatible header fields and other tags specific to the SLZ structure.
- **Endmembers:** An array containing the library spectra. For the sake of fidelity, this array is always stored as raw values.

Using HDF5 Fields

For the SLZ file format, the HDF5 file format is used to store the fields and the subfields that are described in this document.

String Fields and Subfields

If these fields or subfields are strings, they are stored using HDF5 attributes

Numerical and Binary Fields and Subfields

For numerical and binary subfields, they are stored as unsigned integers using HDF datasets.

Each of these subfields consist of DATA, MAX and MIN datasets such that the final value of the field is given by

$$x = \frac{DATA}{r} (MAX - MIN) + MIN$$

where r is the maximum value of the unsigned char type being used for saving the dataset. $DATA$ can be an array, whereas MIN and MAX are always single values.

For SLZ Version 0.9

5 May 2014

© National ICT Australia Limited (NICTA) 2014. All Rights Reserved.

When a sub-field has an array of dimensionality greater than 2, the last dimension is reserved for the non-spatial domain of the variable with respect to the image. That is, for the **HDR.wavelength** (the wavelength subfield of **HDR**), the second dimension corresponds to the wavelength domain.

HDR Field

The header (HDR) field in the SLZ format is based upon ENVI Standard tags. The ENVI Standard file header is compatible with several software vendors, such as ITT and ESRI. This also allows for the field to be populated in a straightforward manner from the input data for a flat file (FLA) as taken from its header file. For more information on the tags in ENVI Standard headers, see Appendix 1.

This field also contains two Scyllarus specific tags. These are

MATx,	String attribute – For a library with N spectra, $x=\{1,2,3,\dots,N\}$. This subfield contains the names or identifiers of the end members in the library.
numEndmembers	Sub-field, numeric – Number of end members contained in the library. This is equal to the number of rows in the Endmember field array.

Endmembers Field

The **Endmembers** field is a numeric dataset which is given by a 2D array which contains a matrix of spectra where the second dimension corresponds to the wavelength. The first dimension is indexed to the **HDR.MATx** subfield. For instance, consider the attribute MAT27 containing the string 'Iron Ore'. This implies that the 27th row in the Endmembers field dataset corresponds to the spectra of iron ore.

Document History

Release	Date	Author	Description
0.1 draft		Antonio Robles-Kelly	First draft.
0.2 draft	1 May 2014	Antonio Robes-Kelly	Second draft

Appendix 1. ENVI Header Structure

General Rules

1. **Curly brace usage:** Used when a keyword contains more than one value.
2. **Entry format:** Entries have to use one of the two following formats (entry ENVI excepted):
 - a. Keyword = value, or
 - b. Keyword = {value1, value2, ..., value n}
3. **Nesting:** This is allowed for entries.
4. **Libraries:** Spectral library files in ENVI are widely used. These contain the additional keyword: spectra names. This additional keyword contains a comma-separated list of ASCII names enclosed in curly brackets.
5. **Additional keywords:** Additional keywords for specific required information. For example, classification results files will contain the additional keywords: classes, which defines the number of classes, including the unclassified, class lookup, which lists RGB color definitions for each respective class, and class names. These tags are to be used for the pipeline output header files.
6. **Coordinates:** In ENVI, pixel values always refer to the upper-left corner of the pixel. Map coordinates also typically refer to the upper-left corner of the pixel. However, if you entered "magic pixel" coordinates in the ENVI header, the map coordinates would refer to the x,y coordinates entered (e.g., x=1.5, y=1.5 would make the map coordinates refer to the center of the pixel).
7. **Comments:** Comments can be added to ENVI header files by inserting a line with a semicolon as the first character. These lines are ignored when the header file is parsed. Comments can appear anywhere within a header file but must be an entire line (and the semicolon must be the first character of that line). Comments cannot follow a keyword/value pair.

Required Header Entries

- **ENVI:** No value
- **description:** This is a character string describing the image or processing performed. It may contain the following tags
 - **time and date:** These are in the format hh:mm:ss, weekday, month dd, yyyy. The time is given in 24hrs format, with the month and the weekday denoted in full in words.
 - **gain:** Floating point value in the interval [0,1](if there's no information on the gain, it should be set to "1").
 - **exposure time:** This is a vector of exposure times per band in a vectorial form. The units are given in parenthesis next to the tag.
 - **exposure time units:** Text string indicating the unit of measure for the exposure time.
 - **GPS data:** Time, date, latitude, longitude, heading (or bearing), satellites in view and fix quality. This is for a mobile/vehicle mounted camera system.

For SLZ Version 0.9

5 May 2014

© National ICT Australia Limited (NICTA) 2014. All Rights Reserved.

- **headeroffset:** Number of bytes of imbedded header information present in the file (for example 128 bytes for ERDAS 7.5 .lan files). These bytes are skipped when the ENVI file is read.
- **samples:** Number of columns in the image.
- **lines:** Number of rows. The spatial resolution in the image is, hence, given by samples x lines
- **bands:** Number of wavelength indexed bands.
- **data type:** Data representation type, where 1=8 bit byte; 2=16-bit signed integer; 3=32-bit signed long integer; 4=32-bit floating point; 5=64-bit double precision floating point; 6=2x32-bit complex, real-imaginary pair of double precision; 9=2x64-bit double precision complex, real-imaginary pair of double precision; 12=16-bit unsigned integer; 13=32-bit unsigned long integer; 14=64-bit signed long integer; and 15=64-bit unsigned long integer.
- **interleave:** Denotes the interleaving on the data. These may be band sequential (bsq), band interleaved by pixel (bip), or band interleaved by line (bil)
- **sensor type:** Type of sensor used to acquire the data (if there's no information on the sensor, it should be "unknown").
- **byte order:** Describes the order of the bytes in integer, long integer, 64-bit integer, unsigned 64-bit integer, floating point, double precision, and complex data types; Byte order=0 is Least Significant Byte First (LSF) data (DEC and MS-DOS systems) and byte order=1 is Most Significant Byte First (MSF) data (all others - SUN, SGI, IBM, HP, DG).
- **wavelength:** Lists the center wavelength values of each band in an image. Units should be the same as those used for the fwhm parameter (see below) and set in the wavelength units parameter.

Optional Header Entries

- **wavelength units:** Units for the wavelength indexed band centres.
- **projection info:** This keyword can be nested and is added to the ENVI header file if a user-defined projection is used instead of a standard projection.
- **default bands:** If set, indicates which band numbers to automatically use to construct an RGB pseudocolour image. If only 1 band number is used, then a grayscale image is constructed.
- **reflectance scale factor:** The value that scales the image cube to be between zero and unity.
- **file type:** ENVI defined file type. The file type ASCII string must match verbatim, including case. This file is an ASCII file that lists the specific file types that are used in the ENVI header to identify special files. These file types include ENVI specific files such as meta files, classification results, virtual mosaics, spectral libraries, and fft results. The file types also include data specific formats such as ADRG and AVHRR. Files such as TIFF and BMP are also recognized as well as ERDAS 8.x and PCI files. The file type field allows these files to have an ENVI header but still exist in their native formats. These types can be edited or even be user defined.
- **map info:** Time, date, latitude, longitude, heading (or bearing), satellites in view and fix quality. This is for a mobile/vehicle mounted camera system.
- **x start** and **y start:** Image coordinates for the upper-left hand pixel in the image. The values in the header file are specified in "file coordinates," which is a zero-based number.
- **fwhm:** Full-width-half-max values of each band in an image. Units should be the same as those used for wavelength and set in the wavelength units parameter.
- **z plot range :**Default minimum and maximum values for Z plots.
- **z plot average :** Number of pixels in the X and Y directions to average for Z plots.
- **z plot titles:** X and Y axis titles for Z plots.
- **pixel size:** Indicates width and length pixel size in meters for non-georeferenced files
- **default stretch:** Determines what type of stretch (% linear, linear range, gaussian, equalize, square root) is used when the image displays
- **band names:**List of names for each band.
- **bbi:**List of the bad band multiplier values for each band in an image, typically zero for bad bands and one for good bands.

- **data gain values:**List of individual gain values per band. This has to be multiplied by the value in the file description gain parameter.
- **data offset values:** List of individual offset values per band. These are the number of bytes to be skipped when the ENVI FLA file is read.

Typical ENVI Library file header

This is the typical header for an ENVI standard library file. The library corresponds to three materials (samples) acquired with a StellarNet spectrometer over 151 bands.

```
ENVI
description = {
gain = 1.000 }
wavelength units = nm
wavelength = {420, 422, 424, 426, 428, 430, 432, 434, 436, 438, 440, 442, 444, 446,
448, 450, 452, 454, 456, 458, 460, 462, 464, 466, 468, 470, 472, 474, 476, 478,
480, 482, 484, 486, 488, 490, 492, 494, 496, 498, 500, 502, 504, 506, 508, 510,
512, 514, 516, 518, 520, 522, 524, 526, 528, 530, 532, 534, 536, 538, 540, 542,
544, 546, 548, 550, 552, 554, 556, 558, 560, 562, 564, 566, 568, 570, 572, 574,
576, 578, 580, 582, 584, 586, 588, 590, 592, 594, 596, 598, 600, 602, 604, 606,
608, 610, 612, 614, 616, 618, 620, 622, 624, 626, 628, 630, 632, 634, 636, 638,
640, 642, 644, 646, 648, 650, 652, 654, 656, 658, 660, 662, 664, 666, 668, 670,
672, 674, 676, 678, 680, 682, 684, 686, 688, 690, 692, 694, 696, 698, 700, 702,
704, 706, 708, 710, 712, 714, 716, 718, 720}
offset = 0
samples = 3
lines = 1
bands = 151
header offset = 0
filetype = ENVI Standard
data type = 12
interleave = bsq
sensor type = StellarNetBLUE-WaveSpectrometer
byte order = 0
spectra names = {
Acmite NMNH133746 Pyroxene, Actinolite HS22, Anhydrite GDS42}
```