The SHGRWalk Cubes

Synthetic Hyperspectral Data

Josh Taylor

05-July-2020

Contents

Preface

The SHGRWalk Datasets (**S**ynthetic **H**yperspectral **G**aussian **R**andom **W**alk, or **SHGR** for short) are synthetically generated 100-dimensional hyperspectral image cubes. Each hyperspectral cube exhibits differing structural and probabilistic characteristics, as outlined below. They are made available as binary .rds files for R.

1   SHGR Class Maps

The synthetic spectra of each pixel of each cube were sampled from a Gaussian mixture distribution with either 8, 11, or 20 components (known as "8class", "11class" and "20class", respectively, from here on). By convention, each mixture component is labeled A - H (8class), A - K (11class) or A - T (20class). The samples from each class are arranged spatially within in a 128×128 pixel (so,

$N = 16,384$) image cube according to the class maps below, which were recyled from previous 6-dimensional hyperspectral cubes [1].
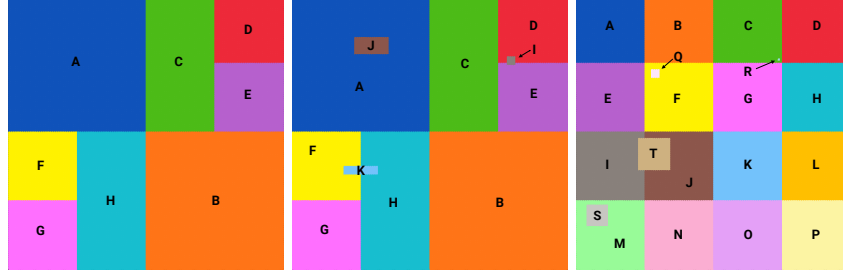
2 Class Means

The class-wise mean intensities were derived from Gaussian Random Walks (bounded within the range [5, 995]) across the synthetic spectral channels (which serve the role of time in a standard random walk). The starting point of each mean walk was sampled uniformly from [50, 950]. The change at each step was sampled $N(\mu, 5)$, where μ was sampled uniformly from $-8, -4, 0, 4, 8$. At any point where the random walk hit (surpassed) the above boundaries, a new μ with opposite sign was selected (uniformly, from the same set). Random change points were also selected where the mean trend was reversed in the same manner. To produce mean spectral intensities with more salient structure a LOWESS smoother [2] was applied across the dimensions (channels) to even out the mean walk trajectories.

20class
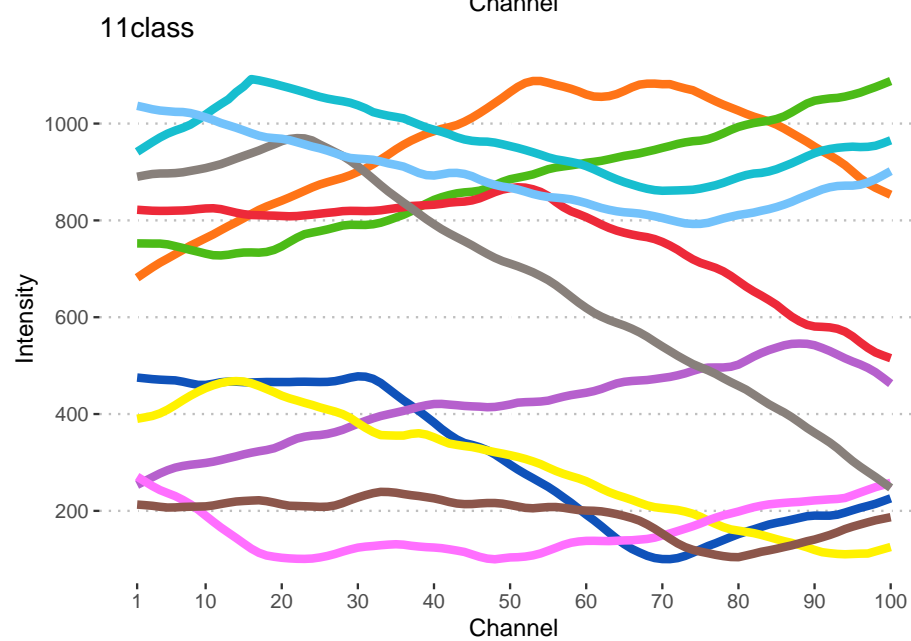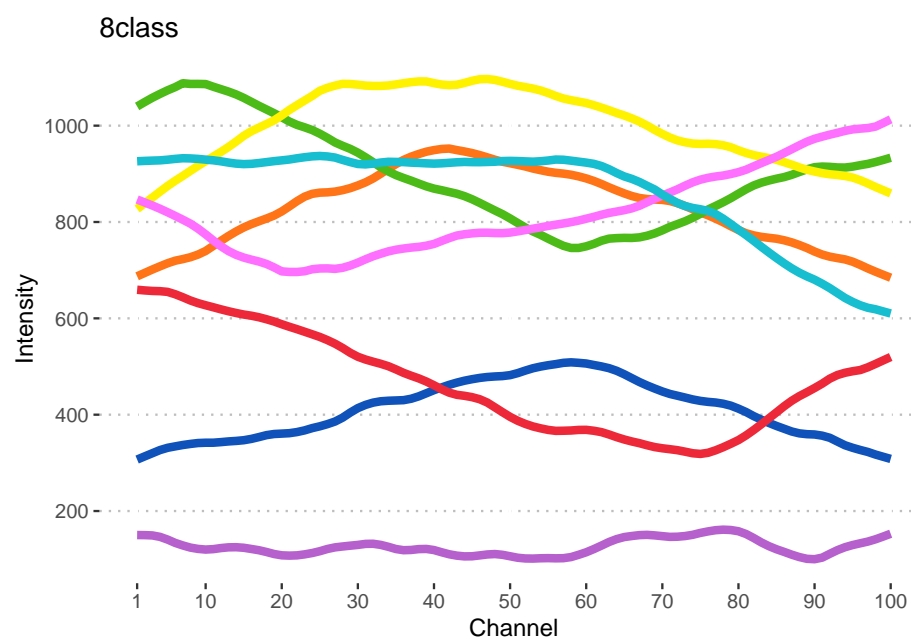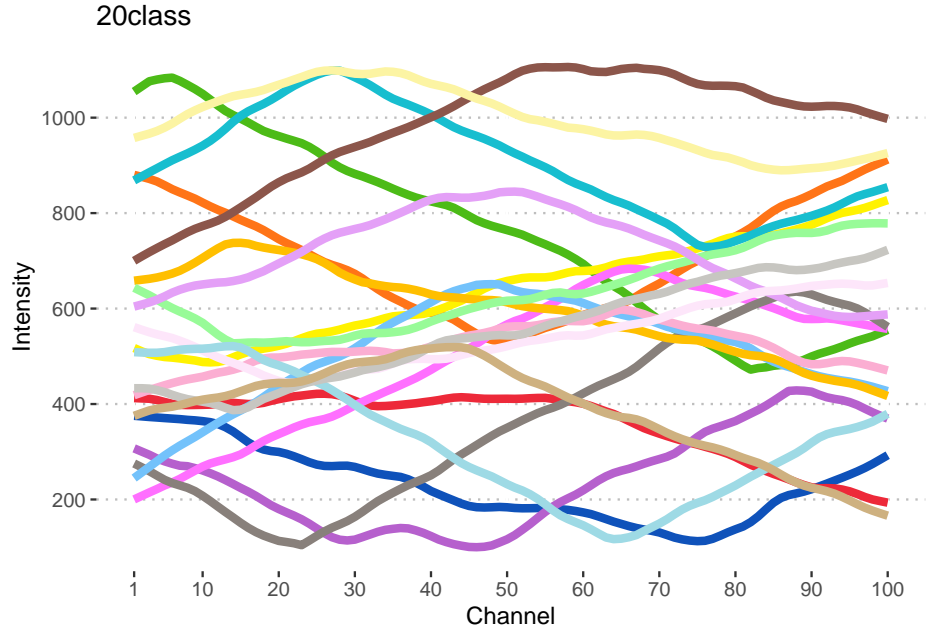
3 Class Covariance Structures

SHGR data was generated both with and without component-wise correlation. The correlation structures, when applied, were also randomly sampled from [0.4, 0.9] and structured according to the Toeplitz model of [3]. Mixtures with a component-wise correlation structure are denoted by "SHGR-cov", whereas mixtures with spherical component densities are denoted by "SHGR-var".Two variance levels have been added to each of the component-wise correlation matrices introduced above. We define "Low Noise" cases as those with components whose dimension-wise variances were sampled uniformly from [100 − 500] and "High Noise" cases with variances sampled from [500 − 4000].

4 Obtaining SHGR Image Cubes

SHGRWalk image cubes can be sourced in .rds format directly from git. The url for the 8class cube with uncorrellated low noise is below. Other cubes can be obtained by changing 8class to either 11class or 20class, var to cov, and 100-500 to 500-4000, as desired.

```
SHGR8.var.lo = readRDS(url("https://github.com/somdisco/SHGRWalk/raw/master/SHGR_100d_8class
```

Each SHGR `.rds` file returns a list with components

- **str_class** a string denoting whether the sourced file is 8class, 11class, or 20class

- **str_vcov** a string denoting the correlation structure and variance level defining the component-wise covariance matrices
- **X** a data matrix of the pixel spectra (nrows = 16,384, ncol = 100)
- **label** a vector of character labels of each pixel (length = 16,384)
- **mix_w** the component mixing weights underlying the true Gaussian mixture distribution
- **mix_mu** a matrix of component means, in rows
- **mix_sigma** an array of component covariance matrices, in slices
- **mix_label** a vector of character labels of each component

Example output from the 8class cube with uncorrellated low noise:

```
str(SHGR8.var.lo)
## List of 8
##  $ str_class: chr "8class"
##  $ str_vcov : chr "var100-500"
##  $ X        : num [1:16384, 1:100] 313 313 309 328 321 ...
##  $ label    : chr [1:16384] "A" "A" "A" "A" ...
##  $ mix_w    : num [1:8] 0.25 0.25 0.125 0.0625 0.0625 0.0625 0.0625 0.125
##  $ mix_mu   : num [1:8, 1:100] 307 687 1040 659 150 ...
##  $ mix_sigma: num [1:100, 1:100, 1:8] 206 0 0 0 0 ...
##  $ mix_label: chr [1:8] "A" "B" "C" "D" ...
```

5   Additional Functionality

Minor functionality for manipulating the cube structures is found in the associated SHGR_aux file:

```
load(url("https://github.com/somdisco/SHGRWalk/raw/master/SHGR_aux.RData"))
```

The above loads several items into the R workspace:

- **SHGRctab** is a data frame containing fields "label" and "color" which defines the color mapping of labels visible in the above class maps.

- `pixelmap_SHGR()` is a function returning a two-column matrix containing the (x, y) pixel coordinates of the spectra found in the rows of X.
- `convert_SHGR_mat2cube(datamat)` is a function to translate the SHGR data matrices (of dimension $N = 16,384 \times 100$) to data cubes (of dimension $128 \times 128 \times 100$).
- `convert_SHGR_cube2mat(datacube)` provides the reverse functionality

6   References

[1] E. Merényi, A. Jain, and T. Villmann, "Explicit magnification control of self-organizing maps for 'forbidden' data," *IEEE Transactions on Neural Networks*,

vol. 18, no. 3, pp. 786–797, 2007.

[2] W. S. Cleveland, "Robust locally weighted regression and smoothing scatterplots," *Journal of the American Statistical Association*, vol. 74, no. 368, pp. 829–836, 1979.

[3] J. Hardin, S. R. Garcia, and D. Golan, "A method for generating realistic correlation matrices," *The Annals of Applied Statistics*, vol. 7, no. 3, pp. 1733–1762, 2013.