

HYPERSPECTRAL PIGMENT DATASET

Hilda Deborah

Department of Computer Science
Norwegian University of Science and Technology (NTNU)
Gjøvik, Norway

ABSTRACT

Hyperspectral imaging has the potential of delivering highly accurate results due to its high spatial and spectral resolutions. However, to ensure relevant and highly accurate end results, the processing steps need to go through rigorous quality assessments. This article provides a generic hyperspectral dataset suitable for designing quality assessment protocols for spectral image processing algorithms. The dataset consists of hyperspectral images of 195 pigment patches and spectral libraries originating from 327 unique pigments. Additionally, two examples of how it can be used for the evaluation of distance functions are also provided.

Index Terms— hyperspectral imaging, dataset, quality assessment, pigment identification

1. INTRODUCTION

Acquiring and processing hyperspectral images are costly and complex. To ensure accurate measurements, a hyperspectral acquisition must be carried out following calibration and characterization protocols [1, 2]. Then, to fully exploit their accuracy, the data need to be considered as measurements of physical objects or surfaces. For image processing, this entails a need for metrology, where bias, uncertainty, trueness, and other metrological aspects are considered [3].

Embedding the notion of metrology in image processing can be done through the concept of image quality assessment (IQA) [4], especially its objective approaches. The basic idea is, *knowing what to expect* of the image processing results, we can devise a protocol to assess *the closeness of the results to the expectation*. An example is given in the quality assessment of spectral median filters [5]. The median filters are known for reducing noise and performing smoothing without blurring the edges. Their statistical and deterministic properties have also been thoroughly studied [6, 7], although only automatically applies in the scalar domain. As soon as the filters depart to the multivariate domain, e.g., color and spectral images, whether those properties still hold must be evaluated.

An integral part of designing an assessment protocol is the availability of data with characteristics appropriate for the properties to be evaluated. Using the example of median filters, we know how they should perform in reducing white noise with uniform, normal, and double-exponential distributions [6]. If that property is to be validated for spectral median filters, the test must use images with the appropriate noise characteristics. This can simply be done by modifying any arbitrary images with, e.g., uniformly distributed additive white noise. However, since separating the effects of signal and noise for nonlinear filters is complex [6], it is important to use constant signals or images with known characteristics.

There is a need for a generic purpose dataset with known characteristics to allow for the quality assessment of various spectral image processing algorithms. Available datasets are often for specific tasks, e.g., classification [8] and anomaly detection [9]. The dataset published in this work is aimed at filling the gap. An earlier version of the dataset has been used to assess the quality of various spectral processing algorithms [10, 11, 12]. In this article, two examples of its use in IQA protocols for distance functions will be demonstrated to inspire readers in designing their own protocols. In addition to IQA protocols, the dataset is also useful for applications in the cultural heritage domain, e.g., pigment identification.

2. DATA DESCRIPTION

This dataset consists of pigment patches and spectral libraries. They both come from the same image acquisition of Kremer color charts, thus having identical spectral dimensions, i.e., 186 bands at 3.26 nm intervals, captured between the range of 405.37 and 995.83 nm. The physical pigment patches themselves were flat objects, and the application of the colorants were done through screen-printing on white cards with printed grayscales on them [13].

2.1. Pigment patches

In total, there are 195 pigment patches belonging to 19 pigment groups. The groups roughly correspond to the Kremer color charts they belong to, e.g., red pigments. Table 1 pro-

This work was funded by the Research Council of Norway under FRIPRO Metrological texture analysis for hyperspectral images (projectnr. 274881).

Table 1: 19 pigment groups in the pigment patches sub-dataset and the number of pigments within each group.

Blue-1	11	Earth pigments	
Blue-2	12	Brown/ black	7
Cadmium	14	Green	12
Green-1	7	Red	13
Green-2	8	Umbræ	12
Iron oxides 1	12	Yellow	12
Iron oxides 4-5	8	Red-2	12
Organic pigments		Red-3	12
Blue/ violet	9	Yellow-3	11
Brown/ black, green	5	Yellow-4	7
Orange	11		

vides the list of pigment groups and their corresponding number of pigments. Examples are also provided in Fig. 1.

Each pigment patch has six associated files, i.e., a header file (*.hdr), a binary file containing the data (*.img), three color images (*.jpg, *_FIXED.png, *_CMF.png), and a ground truth image (*.cls.png). Together, the header and binary files form the hyperspectral image, and they must have the same file name. The rest are accompanying files, see Fig. 2, made for observation purposes and further uses of the hyperspectral image. Read below for more details of the files.

Hyperspectral image Each image is of approximately 800-900 by 500-600 pixels. Then, each pixel corresponds to spectral reflectance values between 0-1, and represented in the 32-bit single precision floating point data type. Other metadata required to read the hyperspectral image, e.g., interleave, are present in the header file.

Accompanying color images They are shown in Fig. 2. The color images are generated using i) GLIMPS software, ii) FIXED: direct use of band numbers (75,46,19) as the RGB channels, and iii) CMF-D65: a color transformation simulating the normal vision human perception under daylight. The latter color transformation is a known transformation within color science [14], where the human visual system is represented by color matching functions. Here, the CIE Color Matching Function (CMF) 1931 2° standard observer and D65 standard illuminant are used. It will also be used throughout this work unless stated otherwise.

Ground truth/ reference image This image should not be treated as ground truth in the physical sense. They are made with digital image processing purposes in mind, e.g., image classification, where a clear separation between the color shades has to be made. In a physical reality, on the edge between two different color shades, there is practically no sharp change of colors or spectra but rather a gradual one.



Fig. 1: Example pigment patches from the dataset. The pigment groups are written in bold under each patch, followed by the corresponding pigment number and name.

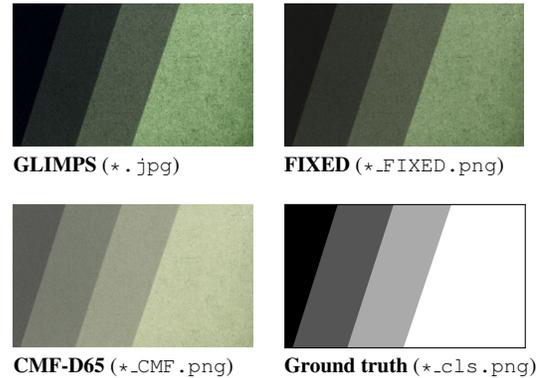


Fig. 2: Three color images and a ground truth image for Bohemian green earth pigment from the dataset.

2.2. Spectral libraries

Two spectral libraries are available, i.e., the average and median spectral libraries, obtained from 398 pigment patches. Note that this is significantly more pigments than what is made available in the form of hyperspectral images. Each pigment patch is represented by four signals in each library since they have four color shades, hence a total of 1592 reflectance spectra in each spectral library. Note also, that out of the 398 patches, only 327 are unique pigments due to several pigments belonging to different pigment groups.

The spectral libraries are stored as ENVI Spectral Library data, hence the header (*.hdr) and the spectral library data

(*.sli) files. Two additional metadata related to the pigment number and name are provided in the header file, i.e., `pnumber` and `pname`, respectively. It is important to note that the metadata `spectra` names are not the same as pigment names. For example, spectra shown in Fig. 3 are associated with `spectra` name `BL_4_1_p1`, `pnumber` 55500, and `pname` `studio pigment sky blue`. Average and median reflectance spectra originating from one pigment are shown in Fig. 3. Both types of spectra obtained from the same region of pixels. Users are encouraged to use the median spectral library in cases where it is important to measure metrological aspects, e.g., uncertainty, trueness, etc., to avoid the *false color* problem [15].

3. QUALITY ASSESSMENTS OF DISTANCE OR SIMILARITY FUNCTIONS

In the following are two examples of how the pigment dataset can be used to evaluate and finally choose suitable distance function and their combination for the task and data at hand. Due to space limitation, these examples will be presented concisely and many details, e.g., mathematical formula, will be left to the referred literature.

3.1. Sensitivity to intensity variation

Spectral capture of a physical surface will always give us different spectral values for every pixel. This results in spectral variability and is not only caused by noise but also, e.g., surface roughness and varying illumination condition. By observing Fig. 1, we can see how one pure pigment applied on a flat surface can already generate perceptual variability. Then, in Fig. 3, we can see how this perceived variability is also reflected in the spectral reflectance domain.

Spectral variability, especially one that is dominated by intensity variations as illustrated in Fig. 4, can be encountered in remote sensing applications as a result of shadows. Consequently, it is often important to suppress the impact of shadow by using distance functions that suppress intensity variations, e.g., spectral angle mapper (SAM) [16] and spectral information divergence (SID) [17]. However, for other purposes such as material discrimination, it is important to use a distance function that is sensitive to intensity variation, e.g., root mean square (RMS) or Chebyshev distances. There is a need to evaluate the appropriateness of various distance functions for the task at hand and this dataset allows such evaluation.

To assess how sensitive various distance functions are to intensity variations, extract one row of pixels from a pigment patch such that we obtain spectra from all four color shades. An example is given for pigment 45110 ultramarine violet (reddish), see their spectra in Fig. 4. The spectrum in green will be used as the reference spectrum S_r in the distance computation $d(S_r, S_i)$, where S_i is the spectrum associated to pixel i , $i \in [1, n]$. The obtained distance values $d(S_r, S_i)$

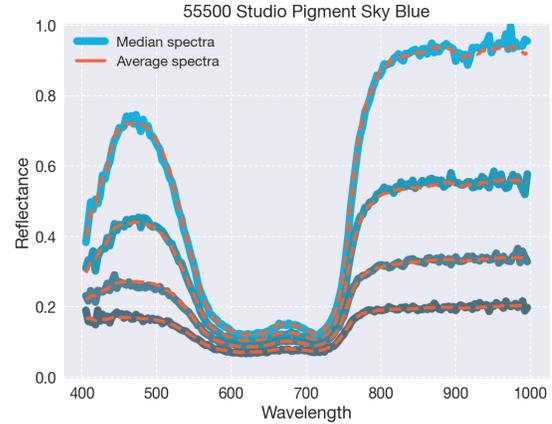


Fig. 3: Median and average reflectance spectra entries for *studio pigment sky blue*, available from their corresponding spectral libraries. Four spectra are available from each library since every pigment patch has four color shades.

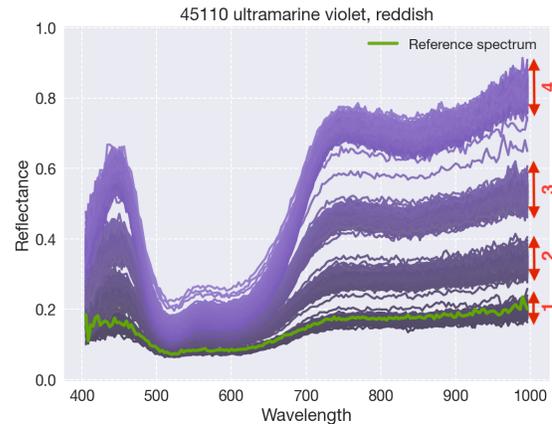


Fig. 4: Spectral reflectance obtained from the pigment patch 45110 *ultramarine violet, reddish* in Fig. 1. The reference spectrum is used in the calculation of distance values. The number 1 to 4 roughly corresponds to the color shades, from darkest to lightest.

for all pixels from location 1 to n should then resemble a step function with four steps, each corresponds to a color shade.

Distance values have been obtained for various distance and similarity functions and the results are plotted in Fig. 5. The expected step functions are indeed obtained from the distance functions shown in the left plot, i.e., RMS, Manhattan, Chebyshev, and Canberra functions. Note that they are shown in normalized values to reflect their sensitivity to intensity variation, despite the actual ranges of the obtained values. In this plot, we can also see that there are fluctuations within each step as expected due to variations in the spectra. The distance values obtained for SAM, SID, GFC [18], and Smith functions are shown in their real values to compare their performance in suppressing intensity variation. A better perfor-

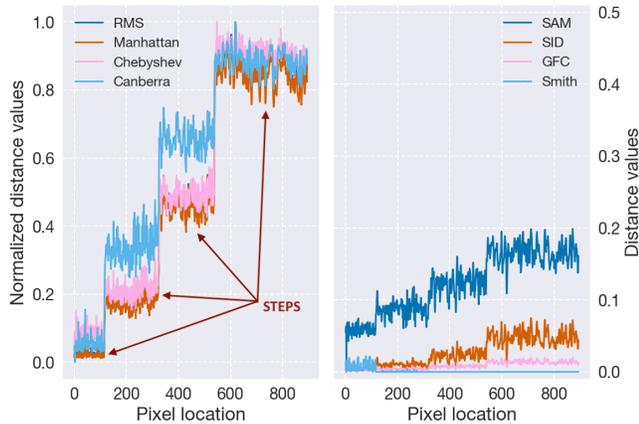


Fig. 5: Distance values for various distance functions computed from spectra shown in Fig. 4. Distance values in the left plot are shown in normalized values, while those in the right plot are in their real values. The expected steps corresponding to each color shade are indicated by the red arrows.

mance is indicated by a function that has values closer to zero and where the steps demonstrated by, e.g., RMS, are not as prominent.

3.2. Separability in the two-dimensional distance space

For discrimination tasks, e.g., classification and clustering, often one feature is insufficient. When it comes to spectral data, the available information is abundant from, typically, the hundreds of wavelengths. However, using all the wavelengths as features will cause the curse of dimensionality, hence the need for feature selection. There are multitudes of ways to select the feature for a spectral classification task, e.g., dimensionality reduction or band selection. For the purpose of demonstrating the use of this dataset, we will use combinations of distance functions as features.

Instead of using only one row of pixels, all pixels in a pigment patch shall be used. In the following, the same pigment 45110 ultramarine violet (reddish) is used for demonstration. The reference spectrum used to compute the distance values is also the same as shown in Fig. 4.

The results of three combinations of distance functions can be seen in Fig. 6, provided in terms of two-dimensional histograms of distance functions. Since the pigment patch we use consists of four color shades, we expect to see four main clusters or distributions of pixels that are well-separated. The first histogram in Fig. 6 is a combination of RMS and SAM, and is called the Normalized Spectral Similarity Score (NS^3) [19]. It has been shown to be superior than the maximum likelihood classifier (MLC), and SAM and SID for a spectral library search task. Since SID is also commonly used in remote sensing, a result is also provided for RMS-SID combination in Fig. 6b. Upon observing Fig. 5, Chebyshev shows

the largest inter-class separation especially for the two last steps. Therefore, a combination of SAM and Chebyshev is also experimented and the result can be seen in Fig. 6c. Comparing the three histograms, it is immediately evident that the SAM-Chebyshev combination generates more well-separated clusters. The four distributions also better resembles a normal distribution and, therefore, a better choice for discrimination purposes. For NS^3 , since it was developed for library search task, this result is reasonable. The reference spectrum used is coming from the cluster closest to 0 in the y-axis. Therefore, other spectra from the same cluster would be densely distributed around 0 to the extent that it is almost imperceptible in this plot. In SID-RMS combination, the obtained distribution indicates that there is a nonlinear relationship between SID-RMS, which mainly is due to the logarithmic operation embedded in the calculation of SID.

4. CONCLUSION

In this work, a hyperspectral pigment dataset has been presented with its main use as a generic purpose dataset for the quality assessment of spectral image processing algorithms. Two examples of such use was given for assessing spectral distance functions in its sensitivity to spectral variation and its discriminatory power when used in a two-dimensional distance space. In both examples, the protocol can be taken further by measuring the intra- and inter-class separations of the obtained distance values using, e.g., an extension of Jeffries-Matusita measure for a multiclass problem. Besides uses in quality assessments, especially the spectral libraries will be useful for applications in the cultural heritage domain, e.g., pigment identification. The pigment dataset can be accessed by the following DOI: 10.5281/zenodo.5592484.

5. REFERENCES

- [1] G. Polder and G. W. A. M. van der Heijden, "Calibration and characterization of spectral imaging systems," in *Multispectral and Hyperspectral Image Acquisition and Processing*, 2001, vol. 4548, pp. 10–17.
- [2] R. Pillay, J. Y. Hardeberg, and S. George, "Hyperspectral imaging of art: Acquisition and calibration workflows," *J. Am. Inst. Conserv.*, vol. 58, no. 1-2, pp. 3–15, 2019.
- [3] BIPM, IEC, IFCC, ILAC, IUPAC, IUPAP, ISO, OIML, "The international vocabulary of metrology-basic and general concepts and associated terms, 3rd ed.," 2012.
- [4] D. M. Chandler, "Seven challenges in image quality assessment: Past, present, and future research," *ISRN Signal Processing*, vol. 2013, pp. 1–53, 2013.
- [5] H. Deborah, N. Richard, and J. Y. Hardeberg, "Spectral ordering assessment using spectral median filters,"

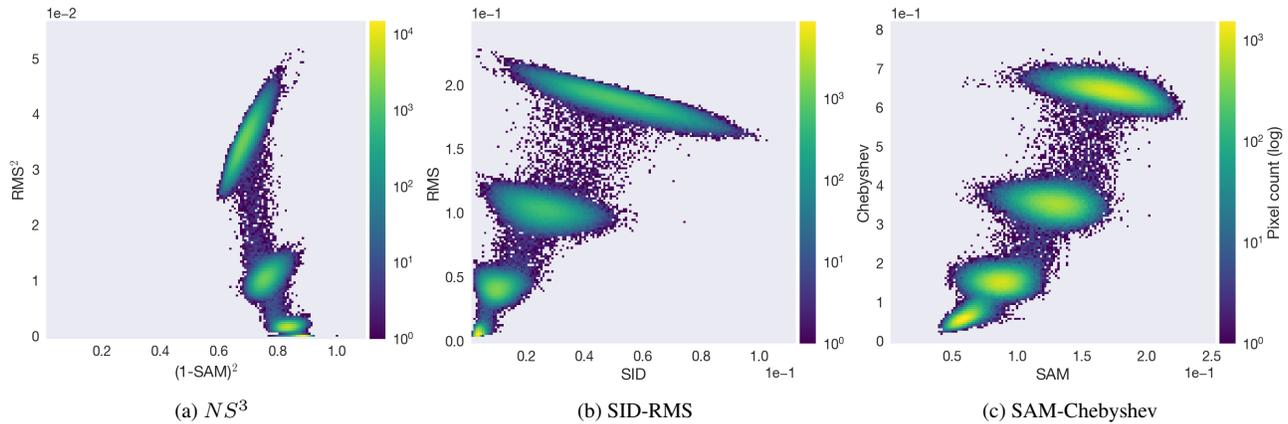


Fig. 6: Two-dimensional histogram of spectral distance values computed for pigment pigment 45110 ultramarine violet (reddish) by using various combination of distance functions. A good combination will give us four clusters that are well-separated, each corresponding to a color shade in the pigment patch.

in *International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing (ISMM)*, 2015, vol. 9082, pp. 387–397.

- [6] B. I. Justusson, “Median filtering: Statistical properties,” in *Two-Dimensional Digital Signal Processing II: Transforms and Median Filters*, Topics in Applied Physics, pp. 161–196. Springer, 1981.
- [7] S. G. Tyan, “Median filtering: Deterministic properties,” in *Two-Dimensional Digital Signal Processing II: Transforms and Median Filters*, Topics in Applied Physics, pp. 197–217. Springer, 1981.
- [8] M. Baumgardner, L. Biehl, and D. Landgrebe, “220 band AVIRIS hyperspectral image data set: June 12, 1992 Indian Pine Test Site 3,” 2015.
- [9] X. Kang, X. Zhang, S. Li, et al., “Hyperspectral Anomaly Detection with Attribute and Edge-Preserving Filters,” *IEEE T. Geosci. Remote*, vol. 55, no. 10, pp. 5600–5611, 2017.
- [10] H. Deborah, N. Richard, and J. Y. Hardeberg, “A comprehensive evaluation of spectral distance functions and metrics for hyperspectral image processing,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 8, no. 6, pp. 3224–3234, 2015.
- [11] H. Deborah, N. Richard, J. Y. Hardeberg, and C. Fernandez-Maloigne, “Evaluation of gradient operators for hyperspectral image processing,” in *Color and Imaging Conference*, 2017, pp. 182–187.
- [12] H. Deborah, N. Richard, J. Y. Hardeberg, and C. Fernandez-Maloigne, “Assessment protocols and comparison of ordering relations for spectral image processing,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 4, pp. 1253–1265, 2018.
- [13] Kremer Pigmente GmbH & Co, “Books and Color Charts,” <https://www.kremer-pigmente.com/en/shop/books-color-charts/>.
- [14] G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, John Wiley & Sons, 2000.
- [15] J. Serra, “The ”False Colour” Problem,” in *Mathematical Morphology and Its Application to Signal and Image Processing*, 2009, pp. 13–23.
- [16] F. A. Kruse, A. B. Lefkoff, J. W. Boardman, et al., “The spectral image processing system (SIPS) – interactive visualization and analysis of imaging spectrometer data,” *Remote Sens. Environ.*, vol. 44, no. 2-3, pp. 145–163, 1993.
- [17] C.-I. Chang, “Spectral information divergence for hyperspectral image analysis,” in *Geoscience and Remote Sensing Symposium (IGARSS), 1999 IEEE International Conference on*, 1999, vol. 1, pp. 509–511.
- [18] J. Hernández-Andrés, J. Romero, A. García-Beltrán, and J. L. Nieves, “Testing linear models on spectral daylight measurements,” *Appl. Opt.*, vol. 37, no. 6, pp. 971–977, 1998.
- [19] R. R. Nidamanuri and B. Zbell, “Normalized spectral similarity score (NS^3) as an efficient spectral library searching method for hyperspectral image classification,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 4, no. 1, pp. 226–240, 2011.