A metrological spectral difference space for the statistical modelling of hyperspectral images

Hilda Deborah*, Noël Richard†, Magnús Örn Úlfarsson‡, Jón Atli Benediktsson‡ and Jon Yngve Hardeberg*
*Department of Computer Science, Norwegian University of Science and Technology, Norway
†Laboratory XLIM, JRU CNRS 7252, University of Poitiers, France
‡Department of Electrical and Computer Engineering, University of Iceland, Iceland

***Abstract*—Answering to metrological constraints typically required in the context of industrial and medical applications, a spectral difference space is introduced in this work. In this space, an acquired hyperspectral data is treated as measurements. Then, modelling the spectral difference space as multivariate Normal laws, a Gaussian mixture model is used in a classification task of remote sensing images. An encouraging result is obtained, comparing the proposed space with a data-driven one. Moreover, it offers a starting point in developing a directly interpretable spectral analysis tools.**

***Index Terms*—metrology, image classification, hyperspectral imaging, Gaussian mixture model**

I. Introduction

Research and development activities and the use of hyperspectral imaging are mainly found in the remote sensing field. This is due to the cost of sensor and the complexity of its acquisition. Nevertheless, existing applications in, e.g., medicine [1], cultural heritage [2], and quality control [3], [4], show the interest of hyperspectral imaging and processing, especially since its acquisition provide high spectral accuracy.

In the recent years, the cost of HSI sensor has decreased quite significantly. There are ready solutions offered in the range of 10k Euro dedicated for industrial applications, which could also be suitable for medical and cultural heritage applications. However, these domains require *metrology* in order to manage bias and uncertainty necessary for quality control or diagnostic purposes. Metrology is the science of measurement. Several international recommendations express the different terms and approaches to assess accuracy, trueness, bias, etc. [5], [6]. These approaches are based on the assessment of the statistics of measurements relative to an average value, a theoretical value, etc. In other words, the assessment of metrological terms are expressed in the space of differences.

Another important point in understanding the need for metrology is in its ability to provide a generic solution of measurements. In metrology, measurements are dependent on the object to measure and not on the approach to measure it. As an illustration, the use of a measuring tape or a laser-based approach will not change the average measurement of an object. It will only change the accuracy of the measurement. This constraint of metrology strictly limits the use of a data-driven approach. By nature, the similarity between two given spectra x and y is always $d(x, y) = t$. However, in the context of data-driven approaches, the value of $d(x, y)$ depends on a given context and not on the accuracy of the sensor.

In this work, we are addressing the question of statistical modelling of spectral samples for quality control purposes, typically needed in industrial applications. Nevertheless, to compare the results with existing ones, the developed approach is applied for a classification task in remote sensing. The aim of this paper is to demonstrate that we can, at least, obtain similar results to that of a data-driven method, but with an adapted metrological solution. With this advance, autonomous and efficient measurement tools based on hyperspectral imaging can be devised in the future.

The article begins by Section II, which expresses the spectral difference space we will use. Histogram of spectral differences will be defined in this space, to further allow establishing *spectral statistics*. By construction, these histograms can be modelled using Gaussian mixture models (GMM). Its short presentation for a classification task will be recalled in Section III, before its use in the experimental study in Section IV. For ease of reading, mathematical notations frequently used throughout the article are provided in Table I.

II. Variability in the spectral difference space

For a set of spectra, there exists no way to directly obtain the statistical modelling or probability density function from its acquisition or reflectance spaces. Within the context of developing metrological processing, in this work, we translate the statistical modelling task into the domain of spectral

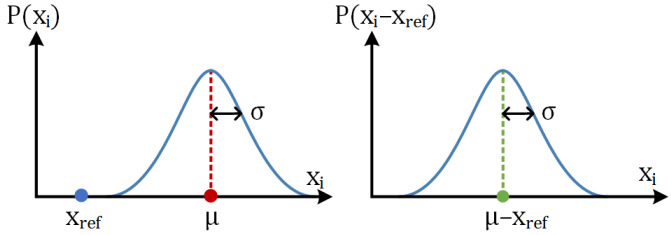


Fig. 1: Illustration of the translation of a statistical modelling of a one-dimensional probability density function $P(x_i)$ into a distance or difference space relative to a reference point x_{ref}.

This work is supported by FRIPRO FRINATEK Metrological texture analysis for hyperspectral images (projectnr. 274881) funded by the Research Council of Norway and French national projects ANR DigiPi and ERDF NUMERIC/e-Patrimoine.

TABLE I: Frequently used mathematical notations.

$S = \{s(\lambda)\}$	Radiance/ reflectance spectrum as a function of λ
λ	Wavelength, with the range of $\lambda \subset [\lambda_{\min}, \lambda_{\max}]$
S_{ref}	Reference spectrum used in the context of difference function
$\bar{S}$	Normalized spectrum S s.t. its sum is equal to 1
$\mathcal{S}$	An arbitrary set of spectra
$d(S_1, S_2)$	Difference function between arbitrary spectra S_1 and S_2
$\mathcal{D}_{\mathcal{S},S_{\text{ref}}}$	Spectral difference distribution of set $\mathcal{S}$ relative to S_{ref}
$\Gamma_{S_{\text{ref}}}$	Variance-covariance measure relative to S_{ref}
$d_M(S, \mathcal{S})$	Mahalanobis distance between a spectrum S and a set $\mathcal{S}$
$\mu_{\mathcal{S}}$, $\Gamma_{\mathcal{S}}$	Mean and covariance of $\mathcal{S}$, respectively

differences. This translation is illustrated in Fig. 1 for a one-dimensional case. Given a reference point x_{ref}, the probability density function $P(x_i)$ is now observed relative to x_{ref} s.t. the average location is changed from μ to $\mu - x_{\text{ref}}$. In the higher dimensional spaces, the probability density function can be extended simply by means of a difference to the reference point $P(d(x_i, x_{\text{ref}}))$. In the following, we will introduce the building blocks of a difference-based space for hyperspectral image processing. In a previous study, various difference functions have been compared and evaluated with metrological constraints [7], [8], and Kullback-Leibler pseudo-divergence (KLPD) [9] is currently the most suitable one.

A. Kullback-Leibler pseudo-divergence function

KLPD was specifically developed for metrological processing of hyperspectral data [9], assuming that the wavelengths are contiguously sampled over a certain range. It measures differences in terms of shape ΔG and intensity ΔW:

$$\begin{aligned} d_{\text{KLPD}}(S_1, S_2) &= \Delta G(S_1, S_2) + \Delta W(S_1, S_2) \\ \Delta G(S_1, S_2) &= k_1 \cdot d_{\text{KL}}(\bar{S}_1, \bar{S}_2) + k_2 \cdot d_{\text{KL}}(\bar{S}_2, \bar{S}_1) \\ \Delta W(S_1, S_2) &= (k_1 - k_2) \log\left(\frac{k_1}{k_2}\right) \end{aligned} \tag{1}$$

where the spectra are considered as probability density functions and normalized when computing $d_{\text{KL}}(\bar{S}_1, \bar{S}_2)$, and k_j is the normalization factor of each spectrum $\bar{S}_j$:

$$\begin{aligned} d_{\text{KL}}(\bar{S}_1, \bar{S}_2) &= \int_{\lambda_{\min}}^{\lambda_{\max}} \bar{S}_1(\lambda) \cdot \log \frac{\bar{S}_1(\lambda)}{\bar{S}_2(\lambda)} \, d\lambda \\ \bar{S}_j &= \left\{ \bar{s}_j(\lambda) = \frac{s_j(\lambda)}{k} \right\}, \; k_j = \int_{\lambda_{\min}}^{\lambda_{\max}} s_j(\lambda) \, d\lambda \end{aligned} \tag{2}$$

B. Histogram of spectral differences

Given a spectral set $\mathcal{S}$ and a reference spectrum S_{ref}, its distribution can be represented and visualized through a histogram of spectral differences (HSD), i.e., $\mathcal{D}_{\mathcal{S},S_{\text{ref}}} = \{d_{\text{KLPD}}(S_{\text{ref}}, S_i), \forall S_i \in \mathcal{S}\}$. As an example, the histogram of pixels belonging to classes Trees and Tiles from Pavia Center image is shown in Fig. 2. Here, S_{ref} was obtained by smoothing the marginal minimum spectrum $S_{\min} = \{\min(s_i(\lambda)), \forall S_i \in \mathcal{S}\}$ using Savitzky-Golay filter (SGF) [10].

Since KLPD is composed of two components, each measuring shape and intensity differences, the one-dimensional histogram can be reconstructed as a two-dimensional one

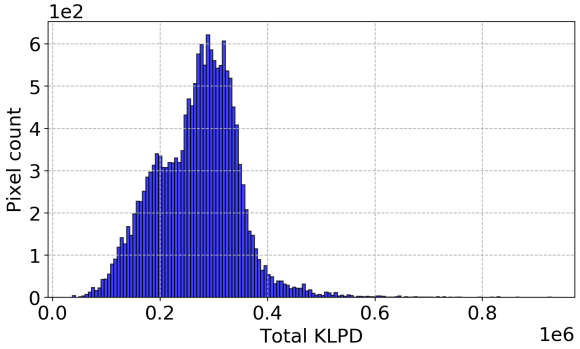


Fig. 2: Histogram of spectral differences obtained from the Trees and Tiles pixels of Pavia Center image.

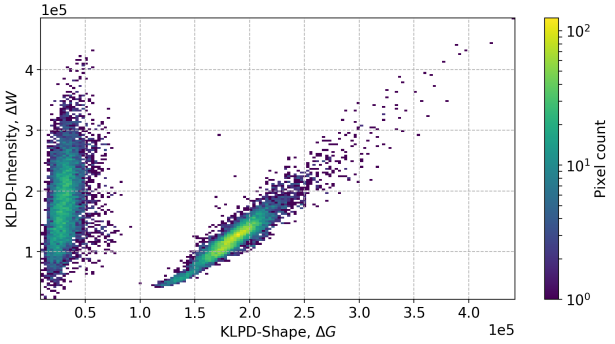


Fig. 3: Bidimensional histogram of spectral differences (BHSD) of two classes in Pavia Center, i.e., Trees and Tiles.

$\mathcal{D}_{\mathcal{S},S_{\text{ref}}} = \{(\Delta G(S_{\text{ref}}, S), \Delta W(S_{\text{ref}}, S)), \forall S \in \mathcal{S}\}$. Using the classes from Pavia Center image, the bidimensional histogram of spectral differences (BHSD) is provided in Fig. 3. Compared to its HSD, it can be observed that the two classes are significantly better separated. The class Trees is the ellipse lying closer to the origin, while Tiles is the one further away.

The representation of spectral distribution by means of histogram of spectral differences is not limited to two dimensions. This representation space can have more axes or dimensions by adding more spectral references. Specifically, the dimension is increased by two per every additional S_{ref}. In turn, this allows the construction of an n-dimensional feature space that can be used for more advanced image analysis and processing tasks.

C. Variance-covariance measure of a spectral difference set

The representation of a spectral set $\mathcal{S}$ through its spectral differences $\mathcal{D}_{\mathcal{S},S_{\text{ref}}}$ further allows computing the variance-covariance measure in (3). In this work, it is assumed that a homogeneous set of spectra would provide a multivariate normal distribution in the spectral difference domain, as illustrated for a one-dimensional case in Fig. 1. However, it has to be noted that the spectral difference values are always positive. Consequently, the $\widetilde{\Gamma}_{\mathcal{S},S_{\text{ref}}}$ will be biased when S_{ref} is chosen from within the set $\mathcal{S}$. It is for this reason that a smooth $S_{\min}$ was chosen as the reference for the case in Fig. 3.

$$\Gamma_{S_{\text{ref}}} = \begin{pmatrix} \alpha_{GG,S_{\text{ref}}} & \alpha_{GW,S_{\text{ref}}} \\ \alpha_{GW,S_{\text{ref}}} & \alpha_{WW,S_{\text{ref}}} \end{pmatrix}, \text{ where} \tag{3}$$

$$\alpha_{GG,S_{\text{ref}}} = \sum_{S_i \in \mathcal{S}} \Big(\Delta G(S_i, S_{\text{ref}})\Big)^2 f(S_i)$$

$$\alpha_{WW,S_{\text{ref}}} = \sum_{S_i \in \mathcal{S}} \Big(\Delta W(S_i, S_{\text{ref}})\Big)^2 f(S_i)$$

$$\alpha_{GW,S_{\text{ref}}} = \sum_{S_i \in \mathcal{S}} \Big(\Delta G(S_i, S_{\text{ref}}) \cdot \Delta W(S_i, S_{\text{ref}})\Big) f(S_i)$$

III. Classification in Spectral Difference Space

A. Gaussian mixture model

Given a hyperspectral image with N different classes, each class $\mathcal{C}_j$ $(1 \leq j \leq N)$ will be a multivariate Gaussian probability density function in the BHSD space $h^j_{S_{ref}}$:

$$h^j_{S_{ref}}(\delta_i) = \frac{1}{(2\pi)^{\frac{r}{2}} |\Gamma_j|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}(\delta_i - \mu_j)^T \Gamma_j^{-1} (\delta_i - \mu_j)} \quad (4)$$

where δ_i is an observation, μ_j the average of $\mathcal{C}_j$ distribution, and Γ_j its associated variance defining the direction of the ellipsoid in the BHSD space. Despite hyperspectral images having typically hundreds of channels, dimensionality of the problem r is reduced to 1 or 2 when HSD or BHSD representations are used, respectively. Moreover, considering the mixed nature of endmembers in remote sensing applications, $\mathcal{C}_j$ can have M Gaussian components. Thus, it can be expressed as a linear mixture of these components [11], [12]. As an illustration, samples from class Trees of Pavia Center are plotted in Fig. 4. The contour lines show the variability within the class as explained by $M = 3$ Gaussian components.

B. Mahalanobis distance

Given any arbitrary spectral group $\mathcal{S}_j$ and a spectrum S, a Mahalanobis distance (MD) $d_M(S, \mathcal{S}_j)$ in the spectral difference space is defined as:

$$d_M(S, \mathcal{S}_j) = \sqrt{d_{\text{KLPD}}(S, \mu_j)^T \, \Gamma_j^{-1} \, d_{\text{KLPD}}(S, \mu_j)} \quad (5)$$

where the set $\mathcal{S}_j$ is modelled through its average μ_j and variance covariance Γ_j. This distance measure will be used in determining whether a spectrum belongs to a group in the context of classification. To obtain a certainty, the empirical rule in statistics will be imposed. This means that a spectrum is considered as belonging to a group only if $d_M(S, \mathcal{S}_j) \leq 3$.

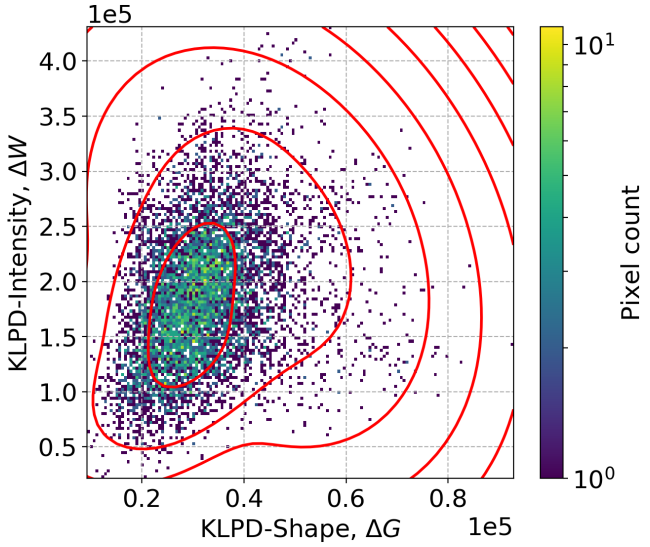


Fig. 4: BHSD showing the variability within class Trees of Pavia Center image, as explained by 3 Gaussian components.

IV. Experiments

A. Setup

The performance of the proposed spectral difference space (SDiff) will be compared to a data-driven one, i.e., PCA. For each case, the target image will only be represented in four dimensions. This means SDiff will use two references and PCA will use four principal components (PCs). These features, of a training set, will be passed on to GMM, that will provide the statistical model of each class $\mathcal{C}_j$, i.e., μ_j and Γ_j $(1 \leq j \leq N)$. The labelling of each test sample is based on its minimum MD to the models of each class. It will be labelled as uncategorised if its minimum MD is above 3 standard unit.

B. Result and discussion

Pavia University image is the first target image. The two spectral references used to build the feature space are the marginal minimum $S_{\min} = \{\min(s_i(\lambda))\}$ and maximum $S_{\max} = \{\max(s_i(\lambda))\}, \forall S_i \in \mathcal{S}$, filtered by SGF to obtain smooth spectra. The obtained accuracy is provided in Table II, alongside PCA. Despite PCA outperforming SDiff in majority of the classes, note the significant difference of performance for class Bricks. SDiff is able to capture the variability within this class since the entire spectral bands are considered when computing spectral differences. For PCA, it is apparent that information regarding this class are lost when only 4 PCs are used. Increasing the number of PCs will increase the accuracy of Bricks, but at the expense of decreasing the performance in other classes due to the curse of dimensionality problem.

Table II also provides results for Pavia Center image. In this case, the references used are the marginal minimum of class Trees $S_{\text{Trees}} = \{\min(s_i(\lambda)), \forall S_i \in \mathcal{S}_{\text{Trees}}\}$ and global marginal

TABLE II: Classification accuracy (average of 5-fold cross validation) of GMM applied on the proposed spectral difference space (SDiff) and PCA.

Pavia University	Accuracy (%) SDiff	PCA	Pavia Center	Accuracy (%) SDiff	PCA
Asphalt	93.05	**97.41**	Water	96.64	**97.90**
Meadows	**71.16**	63.38	Trees	74.50	**92.62**
Bricks	**74.15**	12.22	Asphalt	**80.48**	62.27
Trees	89.82	**96.74**	Bricks	**76.39**	61.94
Metal sheets	94.57	**97.10**	Bare soil	95.60	**97.31**
Bare soil	54.25	**69.26**	Tiles	83.49	**93.53**
Shadows	96.73	**97.04**	Shadows	78.54	**89.31**
			Meadows	93.47	**97.88**
Overall accuracy (%)	**76.08**	69.97		92.13	**95.41**

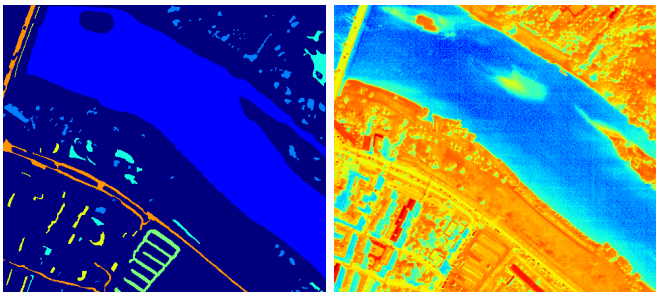
Fig. 5: A subset of Pavia Center. The color image is generated using CIE CMF 2°Standard Observer and D65 illuminant to simulate the human visual system.

maximum $S_{\max} = \{\min(s_i(\lambda)), \forall S_i \in \mathcal{S}\}$. The same way as for Pavia University, the two spectra are smoothed using SGF. In the table, it can be observed that PCA outperforms SDiff in majority of the classes. Nevertheless, their difference in the overall accuracy is only around 3%. It is important to note that SDiff is superior where PCA is challenged, i.e., for Bricks and Asphalts. This is because SDiff does not lose information the way PCA does. Rather, it captures the variability of the dataset from the point of view of several spectral references. Thus, the performance of SDiff is highly related to the choice of spectral references which, however, is out of the scope of this paper.

Another interest of the use of SDiff lies in the fact that it provides results that are directly related to physical measurements, thus easily interpretable. A subset of Pavia Center is shown in Fig. 5. The color image is generated using CIE CMF 2°Standard Observer and D65 illuminant. This color image is simulating how a human eye would see the scene under daylight. In the ground truth image (Fig. 6a), the large bright blue area is tagged as class Water. However, observing again the color image, we can see that the area pointed by the red arrow should not be tagged as Water. The variability within this Water-tagged area is captured by SDiff in Fig. 6b. It shows in a continuous manner that the area pointed by the red arrow has relatively big differences to Water, which has more bluish colors in this map. This limitation of the ground truth has an impact on the numbers we see on Table II. Nevertheless, GMM is able to lessen the impact since it allows explaining the class Water in two or more Gaussian components.

V. Conclusion

This work takes place in the quest for metrological solutions to the analysis of hyperspectral images. Such solutions are typically needed for industrial or medical purposes, where the high spectral sampling of hyperspectral sensors are exploited



(a) Ground truth (b) SDiff map

Fig. 6: Ground truth and spectral difference map of a Pavia Center subset (Fig. 5). SDiff map is computed relative to a marginal minimum of class Water, shown in a log colormap.

to improve the accuracy of measurements. To preserve the metrological aspects of the acquired images, we propose to analyze spectral image content in a spectral difference space. The proposed space allows us to tackle the dimensionality issue by translating the n-dimensional problem to a two-dimensional one without losing the acquired spectral accuracy.

Compared to PCA, the results are encouraging. When PCA outperforms the proposed SDiff, the difference in overall accuracy is relatively small. When PCA is challenged, SDiff significantly outperforms it. These results are obtained with a constraint of preserving the metrological content, as opposed to reducing the data volume. The obtained modelling also offers a very good starting point to directly explain and interpret physical variations captured by a set of spectra, a behavior of spectral analysis tools expected in the industrial and medical context.

Acknowledgment

Authors would like to thank Prof. Paolo Gamba for providing the Pavia images used in this work.

References

[1] G. Lu and B. Fei, "Medical hyperspectral imaging: A review," *J Biomed Opt*, vol. 19, no. 1, pp. 1–23, 2014.

[2] H. Deborah, S. George, and J. Y. Hardeberg, "Spectral-divergence based pigment discrimination and mapping: A case study on The Scream (1893) by Edvard Munch," *Journal of the American Institute for Conservation*, 2019. In press.

[3] G. Mirschel, O. Daikos, T. Scherzer, and C. Steckert, "Near-infrared chemical imaging used for in-line analysis of functional finishes on textiles," *Talanta*, vol. 188, pp. 91–98, 2018.

[4] S. Serranti, A. Gargiulo, and G. Bonifazi, "Hyperspectral imaging for process and quality control in recycling plants of polyolefin flakes," *J Near Infrared Spec*, vol. 20, no. 5, pp. 573–581, 2012.

[5] The Joint Committee for Guides in Metrology (JCGM), "International vocabulary of metrology – Basic and general concepts and associated terms," tech. rep., BIPM, IEC and IFCC, ILAC and IUPAC, IUPAP and ISO, OIML, 2012.

[6] The Joint Committee for Guides in Metrology (JCGM), "Evaluation of measurement data – Guide to the expression of uncertainty in measurment (GUM) - JCGM100: 2008," tech. rep., BIPM, IEC and IFCC, ILAC and IUPAC, IUPAP and ISO, OIML, 2008.

[7] H. Deborah, N. Richard, and J. Y. Hardeberg, "A comprehensive evaluation on spectral distance functions and metrics for hyperspectral image processing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3224–3234, 2015.

[8] H. Deborah, *Towards Spectral Mathematical Morphology*. phdthesis, Norwegian University of Science & Technology, University of Poitiers, Dec 2016.

[9] N. Richard, D. Helbert, C. Olivier, and M. Tamisier, "Pseudo-divergence and bidimensional histogram of spectral differences for hyperspectral image processing," *J Imaging Sci Technol*, vol. 60, no. 5, pp. 1–13, 2016.

[10] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, no. 8, p. 1, 1964.

[11] Y. Zhou, A. Rangarajan, and P. D. Gader, "A Gaussian Mixture Model representation of endmember variability in hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 27, pp. 2242–2256, May 2018.

[12] W. Li, S. Prasad, and J. E. Fowler, "Hyperspectral image classification using Gaussian Mixture Models and Markov Random Fields," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, pp. 153–157, Jan 2014.