# SPATIAL/SPECTRAL INFORMATION TRADE-OFF IN HYPERSPECTRAL IMAGES

*Valero Laparra*[1] *and Raúl Santos-Rodríguez*[2]

[1] Image Processing Laboratoy (IPL), University of Valencia
[2] Dept. of Engineering Mathematics, University of Bristol, UK

## ABSTRACT

This paper shows an empirical analysis of the trade-off between the spectral and the spatial information content of hyperspectral images. The objective of this study is to provide some insights into how changes and variations of both resolutions may affect the information content of the resulting image. This is useful for different stages of hyperspectral image processing: from acquisition to final applications. We propose two alternative approaches to measure the information content of a hyperspectral image: first, a second order approximation where the data distribution is supposed to be Gaussian, and secondly a higher order approximation where no assumption about the data distribution is made.

**Keywords:** Information, hyperspectral images, dimensionality reduction

## 1. INTRODUCTION

Hyperspectral imaging is facing new methodological challenges. First of all, the availability of satellite images has increased exponentially over the last few years, imposing important constraints in order to effectively deal with the data. Secondly, much work has been devoted to develop optical sensors that provide unprecedented spectral and spatial resolution. Thanks to these advances, a wide variety of applications, ranging from monitoring the development and health of crops to surveillance, benefit from more accurate results and also new application domains have emerged [1]. However, even a single hyperspectral image has become a very complex high dimensional object that requires both adequate storage and efficient processing.

In order to alleviate this problem, usually some of the less informative data used to be discarded along the life cycle of a hyperspectral image, from the acquisition stage to the application-dependent post-processing. Even before that, the initial decision occurs during the design of the sensor, where the spatial and the spectral resolution are selected. Nevertheless, at acquisition time, there may still be room for maneuver. For example, NASA's AVIRIS has a spatial resolution of around 20 meters when flown above sea level at its typical altitude of 20 kilometers, but a 4-meter resolution when flown at an altitude of 4 kilometers over land. Once the image is captured, one has to decide among different compression algorithms, which will determine the actual image storage size. Finally, depending on the application, a variety of feature extraction algorithms and dimensionality reduction techniques (such as Principal Component Analysis) are applied. Through these subsequent steps, some of the data is kept or otherwise discarded according to different criteria. The choices are usually based on prior knowledge, experience o commonly used heuristics. Unfortunately, the theoretical justification of some of these criteria may be very convoluted.

The main contribution of this work is to measure the information trade-off between the spatial and the spectral resolution. To this end, we estimate the information content of a hyperspectral image in the terms of Shannon's entropy [2]. This concept provides a simple yet effective procedure to discard data. Previous works have suggested similar criteria to analyze hyperspectral images [3, 4, 5] but, up to our knowledge, a thorough analysis of this problem has not been presented before. Due to the high dimensional nature of hyperspectral images, we will use measures that effectively take into account these multidimensional relations. A interesting novelty of this work lies in the use of a non-parametric method to estimate the multidimensional entropy. In this way, we directly address the problem of spatial versus spectral information, providing a fair comparison for different configurations of spatial and spectral resolution.

The conclusions of our analysis can be applied to a wide variety of tasks. As we mentioned before, when designing an optical sensor a key decision is to select the spectral and the spatial resolution. Additionally, regarding the storage of huge amounts of gathered images, although all data may be necessary temporarily (e.g. to predict the weather), a lighter version of the images are kept as representative historical data (e.g. to study the climate over time), where the information criteria would used in order to decide which data we should persist. Lastly, how we handle the data is specially critical for statistical learning methods, both in terms of performance and scalability of the learning algorithms. For instance, applications such as weather forecasting or biological variable prediction are inherently high dimensional problems. In this sense, the information criteria would be interesting for feature selection, as an usual preprocessing step in all these models.

Supported by Vali+d project APOSTD/2014/095.

The remaining of this paper is organized as follows. In section 2 we describe the two different methods that we will use to empirically measure the information content. In section 3 we describe both the data and the methodology we follow. Finally, section 4 discusses the empirical results and the future lines.

## 2. MEASURING THE INFORMATION CONTENT

In order to measure the amount of information content for different spatio-spectral resolutions, we use the standard information definition by Shannon [2]. As a measure of information, we use the notion of entropy. Given a particular selection of spatial and spectral resolution, the entropy of an image gives the average amount of information that this configuration has for that scene.

Entropy is usually defined in the context of a probabilistic model. However, as the underlying probability distribution of hyperspectral images is unknown, we can not compute entropy using a closed-form expression and we face a multidimensional estimation problem. To measure entropy we propose two methods. The first one assumes that images follow a Gaussian distribution. This method is easy to implement while keeping a small estimation error. This approximation was originally presented in [3]. However, although it is a reasonable first choice, the distribution of the hyperspectral images is known to be different from a Gaussian [6]. The second approach is a non-parametric method (i.e. it does not assume any particular distribution of the data). Although it is a much more realistic model, measures are noisier. In section 3 we will analyze the different results obtained by both alternatives.

### 2.1. Gaussian distribution assumption

The first approximation is based on the assumption that the data is distributed as a multidimensional Gaussian. Assuming a functional model of the data distribution always implies an error in the results. Mainly we are taking into account just second order relationships, and therefore discarding any higher order terms. Since multidimensional probability estimation is a difficult problem, by assuming this model we greatly simplify the computation. Additionally, although images are not distributed as multidimensional Gaussians, the assumption is not completely unrealistic [6]. Actually, the error caused by ignoring higher order relations in natural images is usually not very high [7]. Given a hyperspectral image in matrix form $\mathbf{X}$, the expression of the entropy of a multidimensional Gaussian distribution is as follows:

$$h(\mathbf{X}) = \frac{d}{2}(1 + \ln(2\pi))) + \frac{1}{2}\ln|\mathbf{\Sigma}| \quad (1)$$

where $d$ is he number of dimensions, and $|\mathbf{\Sigma}|$ is the determinant of the covariance matrix. Note that the error in this measure will be determined by the error in the estimation of the covariance matrix, which is not critical when a large number of samples are available. In order to reduce the dimensionality, using the entropy is an alternative (although similar) criteria to just selecting the number of retained features given by Principal Component Analysis.

### 2.2. Non-parametric approach

Rotation based iterative Gaussianization (RBIG) [8] is a non-parametric method that has been successfully applied to estimate different multidimensional information-theoretic measures. RBIG is especially interesting for our problem because it works particularly well in high dimensional scenarios while keeping a low computational cost. In [8] its ability for measuring multi-information was shown, and more recently, in [9], RBIG was used to estimate the Kullback-Leibler divergence between multidimensional probability distributions. As opposed to the previous method, RBIG does not assume a functional form for the data distribution. Due to its formulation its accuracy depends mainly on marginal entropy estimations. In particular we can take advantage of its ability for measuring multi-information and apply the following formula to then estimate the multidimensional entropy:

$$h(\mathbf{X}) = \sum_{i=1}^{d} h(\mathbf{X}_i) - I(\mathbf{X}) \quad (2)$$

where $h$ represents the entropy (both for multidimensional data, $h(\mathbf{X})$, and for unidimensional data, $h(\mathbf{X}_i)$), and $I$ is the multi-information. Note that marginal entropies are easy to estimate and multi-information can be estimated using RBIG.

## 3. EXPERIMENTS

We aim to empirically determine the trade-off between spatial information and spectral information, comparing the two aforementioned approaches. The underlying objective is to analyze images with a collection of spatial and spectral resolutions, and show the amount of information that is kept when using limited spatial or spectral resolutions and how this value varies for each configuration.

### 3.1. Experimental setup

Ideally, we would want to analyze different types of hyperspectral images with a variety of spatial and spectral resolution configurations and similar content. Up to our knowledge, a dataset with these properties does not exist, so we synthetically reproduce the effects of different resolutions using an image with medium spatial and spectral resolution. For these experiments, the selected image was captured using NASA AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) instrument in the Salinas Valley in California in

1998 [1]. The image has 224 bands of 10 nm width with center wavelengths from 400 - 2500 nm. The KSC data, acquired from an altitude of approximately 20 km, have a spatial resolution of 18 m. The data generation process takes the selected image as the starting point, and iteratively constructs images with lower spatial and spectral resolution to explore different resolution configurations.

### 3.2. Results

Starting from the AVIRIS image, which has moderate spatial and spectral resolution, we synthetically simulate images with different spatial and spectral resolutions. Then, we compute the entropy of each of these new images - with a fixed number samples (20.000) in each estimation - as a measurement of the amount of information captured by each spatial/spectral configuration. The results are reported in Figure 1 for the Gaussian distribution assumption (Gaussian) and Figure 2 for the non-parametric approach (RBIG).

Figures 1(a) and 2(a) show a matrix where each element represents the estimated entropy for a specific spatial and spectral resolution. In this case, although the values differ slightly, both methods present a similar behavior: as expected, the entropy grows as the spatial and spectral resolutions increase. Figures 1(b) and 2(b), illustrate the comparison of the information content for different configuration with fixed number of coefficients. Using either of the methods, combining spectral and spatial data clearly provides more information when using the same number of coefficients. However, determining the exact trade-off between spectral and spatial information requires a more detailed analysis. In this case, the accuracy of RBIG becomes very relevant. Although it is computationally cheaper, the Gaussian assumption is not entirely realistic for hyperspectral images. RBIG does not assume an underlying model achieving a better performance specially in high dimensional problems [8]. While the Gaussian approach suggests that a higher spatial resolution results in a higher information content, RBIG encourages a slightly higher spectral resolution.

## 4. CONCLUSION

This paper reports an empirical analysis of the trade-off between the spectral and the spatial information content of hyperspectral images. We illustrate how variations of both resolutions determine the entropy of the corresponding image. In order to estimate the entropy of a hyperspectral image, we try two different methods: first, a second order approximation where the data distribution is assumed to be Gaussian, and secondly a non-parametric model (RBIG). While the first approach is easily computed (just a covariance matrix eigenvalues computation is required), it assumes a Gaussian model. The second approach is more accurate in this sense since it is adapted to the real model of the hyperspectral images. However it is more expensive computationally. Results show that the combining spectral and spatial data provides clearly more information when using the same number of dimensions. Interestingly, the results show that, for a constant number of coefficients, a higher spectral resolution provides a higher information content. Although these are preliminary results obtained with just one image, multiple images with different content will be considered in future works.

[1] Authors want to acknowledge Dr. A. Gualtieri, for providing the image

## 5. REFERENCES

[1] G. Camps-Valls, D. Tuia, L. Bruzzone, and J.A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, 2014.

[2] Claude E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July, October 1948.

[3] Clive D. Rodgers, *Inverse methods for atmospheric sounding : theory and practice*, vol. 2 of *Series on atmospheric, oceanic and planetary physics*, World Scientific, 2000.

[4] B. Aiazzi, S. Baronti, L. Santurri, M. Selva, and L. Alparone, "Information-theoretic assessment of multidimensional signals," *Signal Processing*, vol. 85, no. 5, pp. 903 – 916, 2005, Information Theoretic Signal Processing.

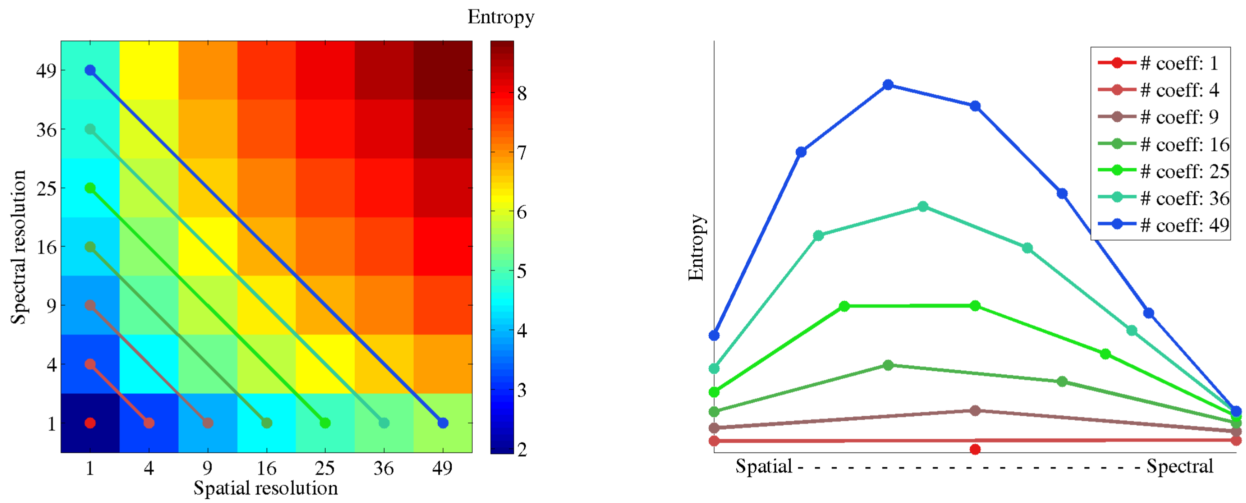
[5] N. Longbotham, F. Pacifici, B. Baugh, and G. Camps-Valls, "Prelaunch assessment of worldview-3 information content," in *IEEE Workshop on Hyperspectral Image and Signal Processing, Whispers*, 2014.

[6] S. Jimenez and J. Malo, "The role of spatial information in disentangling the irradiance-reflectance-transmittance ambiguity," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 8, pp. 4881–4894, Aug 2014.

[7] M. Bethge, "Factorial coding of natural images: How effective are linear model in removing higher-order dependencies?," *Journal of the Optical Society of America A*, Jun 2006.

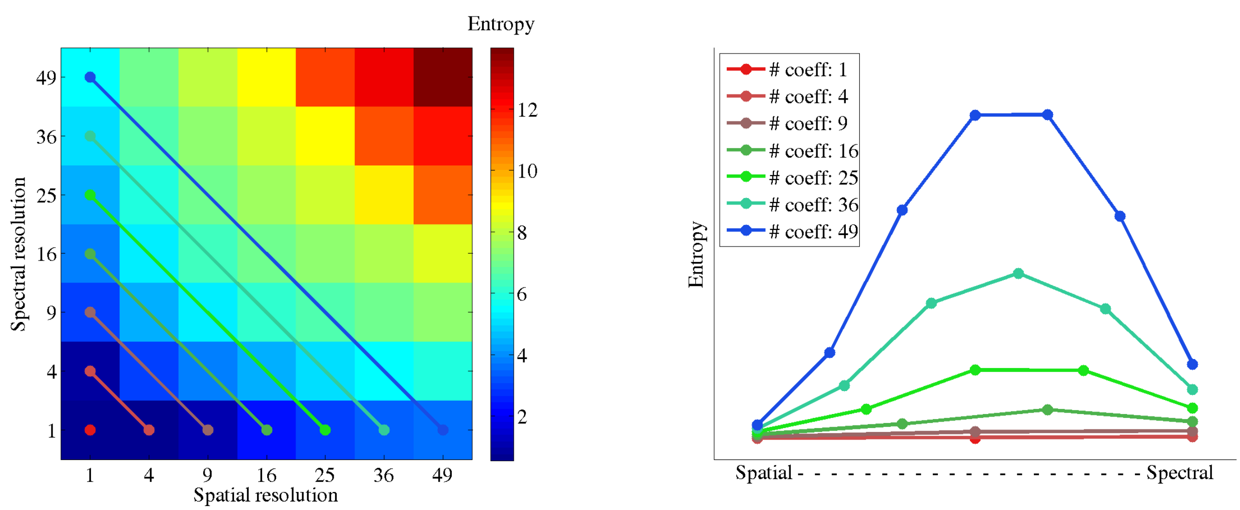
[8] V. Laparra, G. Camps-Valls, and J. Malo, "Iterative gaussianization: From ica to random rotations," *Trans. Neur. Netw.*, vol. 22, no. 4, pp. 537–549, Apr. 2011.

[9] V. Laparra, J. Malo, G. Camps-Valls, and R. Santos-Rodriguez, "Kullback-leibler divergence estimation based on iterative gaussianization," in *IEEE-EURASIP Seminar on Signal Processing, Communication and Information Theory*.




(a) (b)

**Fig. 1**. Entropy estimation in AVIRIS using the Gaussian model: (a) entropy for different configurations of spatial and spectral resolution. Entropy values are in logarithmic scale for visualization pourposes. (b) Entropy for a fixed number of coefficients (selected diagonals of the entropy matrix).




(a) (b)

**Fig. 2**. Entropy estimation in AVIRIS using RBIG: (a) entropy for different configurations of spatial and spectral resolution. Entropy values are in logarithmic scale for visualization pourposes. (b) Entropy for a fixed number of coefficients (selected diagonals of the entropy matrix).