



HAL
open science

Quantization of hyperspectral image manifold using probabilistic distances

Gianni Franchi, Jesus Angulo

► **To cite this version:**

Gianni Franchi, Jesus Angulo. Quantization of hyperspectral image manifold using probabilistic distances. International Conference on Networked Geometric Science of Information, Oct 2015, Palaiseau, France. 10.1007/978-3-319-25040-3_44 . hal-01121104

HAL Id: hal-01121104

<https://hal.science/hal-01121104>

Submitted on 27 Feb 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Quantization of hyperspectral image manifold using probabilistic distances

Gianni Franchi and Jesús Angulo

MINES ParisTech, PSL-Research University,
CMM-Centre de Morphologie Mathématique; France
`gianni.franchi@mines-paristech.fr, jesus.angulo@mines-paristech.fr`

Abstract. A technique of spatial-spectral quantization of hyperspectral images is introduced. Thus a quantized hyperspectral image is just summarized by K spectra which represent the spatial and spectral structures of the image. The proposed technique is based on α -connected components on a region adjacency graph. The main ingredient is a dissimilarity metric. In order to choose the metric that best fit the hyperspectral data manifold, a comparison of different probabilistic dissimilarity measures is achieved.

Keywords: quantization, hyperspectral images, information geometry, probabilistic distances, mathematical morphology

1 Introduction

Gray-level images are pictures where each pixel is a scalar value which usually is quantized between 0 and 255. Color and hyperspectral images are images where each pixel can be considered as a vector, such that each coordinate of the vector corresponds to the intensity of the pixel at a certain wavelength. In addition, pixels values on spectral images are most of the time different. Quantization of images can be seen as a way to remove this excess of variability by reducing the number of (spectral) vectors and therefore to address the curse of dimensionality [3]. The problem can be addressed using dictionary learning techniques [2]. In these learning methods, a dictionary composed of k atoms is learned on a set of vectors. Then each vector is represented by a (low) number of atoms of the dictionary. Among these methods, VQ[6] is a well known technique that quantizes a set of vectors, since each vector is represented by just one atom of the dictionary. We consider in this paper a vector quantization method for hyperspectral images, halfway between manifold learning and dictionary learning methods, since we explore a quantization of the image manifold. First, we review existing similarity metrics on hyperspectral imaging and consider some probabilistic distance less known on this domain. Finally we quantize hyperspectral images using the best distances.

2 Quantization of hyperspectral images

Background on hyperspectral images. Let us consider a hyperspectral image where each pixel value v_i is a spectral vector of dimension $d \in \mathbb{N}$, such that $v_i \in \mathbb{R}^d$. Because of the curse of dimensionality [3], spectral variability is added to the vectors. First, let us consider a simple case where $v_i \in [0, 1]^d$ and $d = 1$; in this case it is easy to calculate that with 100 points, one get an interval between points of around 10^{-2} . However, if we consider a space of dimension $d = 10$ and if one wants the points to be separated by a ball of radius 10^{-2} , we can see that this time we need 10^{20} points. Indeed, in the case of hyperspectral images, d is usually about one or two hundred, which shows how sparse the sampled manifold is (or how empty the whole space is), and how crucial is the question about similarity between spectra. Moreover one can see a hyperspectral image as a set of spectral classes. If we consider for instance three classes: “road”, “water”, or “forest”, then each pixel belongs to one of these three classes. Due to the high dimensional space, each class has a high variability with respect to the Euclidean distance, and thus this distance will not be discriminative enough to separate objects from different classes. A solution could be to reduce the dimension of the manifold or to find another low-dimensional space to embed data, and then to use Euclidean distance in this space. Actually, we do not focus here on this kind of classical approach. We assess the interest of metrics on the original manifold on the d -dimensional space that are more invariant to spectral variability [1, 11, 14]. Thanks to this kind of similarities we can expect to improve hyperspectral image quantization.

Spectral/spatial image quantization. Quantization is the process which allows to approach a signal with large set of values by a signal on a smaller set. Images are signals on a spatial domain, so their quantization should takes into account the expected spatial coherence. To achieve this goal, we choose to use α -connected components representation [20, 22, 17], that produces an image partition into homogenous spatial classes. Two pixels belongs to the same α -connected component if there is a path linking these pixels such that the similarity between successive pixels of this path is lower than α . However, α -connected components algorithm often produces inadequate image partitions, since it fails to respect image contours. A solution is to first use an initial partition algorithm that would produce “superpixels” on our image and that must follows main image contours. Then, the superpixels are connected between then by their region adjacency graph (RAG). It is a graph where each node is a superpixel, and edges represent the dissimilarity between superpixels. Here edges are weighted by the dissimilarity between centroids of superpixels. In our case, the superpixels are obtained by computing the classical watershed on the image. Then the notion of α -connected components can be extended to RAG [17]. Moreover, by comparing nodes and thus regions, our quantization is more robust to noise. For this purpose on each superpixel SP_i the value of the original pixels is replaced by the value of its centroid C_i . Finally the choice of α is done in order to have a fixed number of centroids, and thus of different spectra.

Assessment of image quantization. Our quantization depends on the RAG and therefore, on the choice of distance. We focus on the potential interest of probabilistic distances on spectral pixels. In this context, our assessment is separated into two steps.

Step 1: Evaluation of the probabilistic distances.

We consider that a good probabilistic distance is a dissimilarity measure that has an invariant behaviour on each spectral class in a high dimension space. During the process of assessment, hyperspectral images with a ground truth spectral classification are used. Thus the class of each pixel is known. Moreover for each class the centroid for a dissimilarity measure can be computed: it corresponds to the vector which minimizes the cumulated distance to the other vectors of the class. The centroid represents the class. Figure 2 represents two scatter-plots from a three-classes hyperspectral image. These scatter-plots depict the distance of each spectrum to the centroids of each class. A good dissimilarity measure should concentrates the data of each class around its centroid. In this example, Kullback-Leibler divergence performs better than L_2 norm from this viewpoint. To quantitatively assess such property, the image is classified by calculating the distance of each image spectrum to the class centroids. Then, the class of the nearest centroid is given to each spectrum. After having classified the image, we evaluate the Overall Accuracy (OA), and the Average Accuracy (AA), that are two measures commonly used in hyperspectral imaging [9]. In addition, the rank of classification is also computed for each pixel. For example, let us consider that we have a spectrum X and three classes C_1, C_2, C_3 , and the spectrum X belongs to cluster C_1 . After computing the distance of X to each class centroid, denoted respectively D_1, D_2, D_3 , we obtain $D_1 > D_3 > D_2$. Obviously the classifier would consider that the class of X is C_2 . Since in this case D_1 is the highest distance, the classification mistake is very significant. The rank of good classification is in the present case 3. Hence, the rank of classification (Rank) is obtained as the mean of rank of good classification over all the pixels, divided by the number of classes. The more the rank of classification is near to one, the worst a dissimilarity measure is.

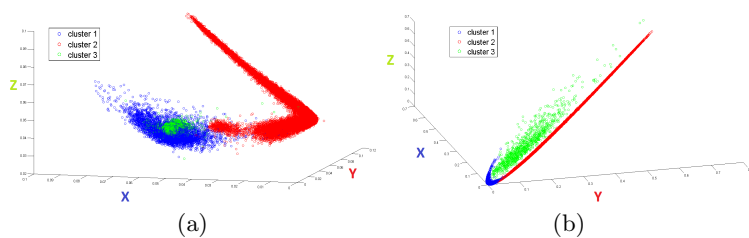


Fig. 1. Scatter plot of distances of pixels (of Pavia image) of cluster 1 (in blue) to the centroid of cluster 1 in X, to the centroid of cluster 2 in Y, to the centroid of cluster 3 in Z. Similarly, in red for cluster 2, and in green for cluster 3. In (a) we use the L_2 norm as dissimilarity measure, in (b) we used the Kullback-Leibler divergence.

Step 2: Evaluation of the quantization results. Quantization is a simplification operation that can destroy some relevant information. This introduces an error between the quantized signal and the original signal. This error is generally called quantization noise (or distortion). Let us consider an image f , and its quantized version \hat{f} . The distortion is measured as: $\|f - \hat{f}\|_2$. Thus this distortion is measured by the L_2 norm and this point can be problematic for use since we compare different measures of dissimilarity. To overcome this problem, we used as measure the SNR that is also used on signal quantization. Nevertheless, this metric might not be adapted to images since the spatial distortion is not taken into account. We propose to introduce an additional measure of distortion adapted to structured signals. The pattern spectrum (PS) [4] corresponds to the probability density function (pdf) of the granulometric decomposition, a multi-scale morphological image decomposition. We consider in fact the cumulative distribution function of the difference image $|f - \hat{f}|$. A good quantization schema filters out small spatial structures and keeps most of the large spatial objects of the image $|f - \hat{f}|$. A quantized image \hat{f}_A is better than a quantized image \hat{f}_B if the cumulative pattern spectrum of image $|f - \hat{f}_A|$ stochastically dominates [8] the one of $|f - \hat{f}_B|$.

3 Probabilistic distances on hyperspectral images

We focus on spectral metrics which model the data uncertainty as resulting from randomness. Thus we consider each pixel as a random variable with the probability distribution obtained by normalizing the vector. Given two vectors $X = (x_1, \dots, x_d) \in \mathbb{R}^d$, and $Y = (y_1, \dots, y_d) \in \mathbb{R}^d$, they are represented respectively by their probability distribution function (pdf) as $P_x = (\frac{x_1}{\sum_i x_i}, \dots, \frac{x_d}{\sum_i x_i}) \in \mathbb{R}^d$ and $P_y = (\frac{y_1}{\sum_i y_i}, \dots, \frac{y_d}{\sum_i y_i}) \in \mathbb{R}^d$. One can represent the data by their pdf or by their cumulative distribution function (cdf), or by their characteristic function. All these representations of the data and the different possible dissimilarities involves many ways to compare data.

L_p Minkowski norms. Given a real number $p \geq 1$, we can define the L_p norm on the pdf of vectors $(X, Y) \in \mathbb{R}^d$ as follows: $N_p(X, Y) = (\sum_i |X_i - Y_i|^p)^{\frac{1}{p}}$. Moreover the L_∞ norm is defined by: $N_\infty(X, Y) = \|X - Y\|_\infty = \max(|X_1 - Y_1|, \dots, |X_2 - Y_2|)$. One has to choose carefully the p -norm to be used. It is possible to notice that when the value of p increases, the distance gives more weight to the dimension of high dissimilarity. On the contrary, if one decrease the value of p , the distance tends to average dissimilarities. The p -norms are particularly used in hyperspectral imaging, because of their simplicity of use. However, the higher the dimension of the spectra, the more there is a data concentration effect. Thus it is impossible to discriminate data when the dimension increases. If we consider a set of data $\{X_i \in \mathbb{R}^d\}$ then:

$$\max_{i \neq j} (\|X_i - X_j\|_p) / \min_{i \neq j} (\|X_i - X_j\|_p) \rightarrow 1 \text{ when } d \rightarrow \infty$$

Spectral Angle Mapper (SAM). Let us consider a spectrum $X \in \mathbb{R}^d$ of an hyperspectral image. It can be considered as a set of n random tests, where $n = \sum_i X_i$, and each event X_i is independent and provides a binary outcome. Thus, each image spectrum can be seen as following a multinomial distribution of parameter P_x . It is then possible to define a Fisher-Rao distance between X and Y , represented by their pdf P_x, P_y . This notion corresponds to the spherical distance [15]:

$$D_{Spher}(X, Y) = 2 \arccos\left(\sum \sqrt{P_{x,i}P_{y,i}}\right). \quad (1)$$

It is called the spherical distance since it represents the geodesic distance between distributions that are embedded on a $(d-1)$ -unit sphere of equation: $\sum_{i=1}^d \sqrt{P_{x,i}}\sqrt{P_{y,i}} = 1$.

A distance classically used on hyperspectral images is the spectral angle mapper (SAM) [1], which for X, Y is defined between as:

$$D_{SAM}(X, Y) = \arccos\left(\frac{\sum X_i Y_i}{\sqrt{\sum X_i X_i} \sqrt{\sum Y_i Y_i}}\right) = \frac{1}{2} D_{Spherical}(X^2, Y^2). \quad (2)$$

This metric is invariant to spectral multiplication since $D_{SAM}(\alpha X, Y) = D_{SAM}(X, Y)$, $\forall \alpha \in \mathbb{R}^*$.

Hellinger distance. We considered above the geodesic distance between two vector of the $(d-1)$ - unit sphere of equation. Indde, one can consider on the unit sphere the chordal distance too, which is the distance between the points in the flat embedding space, and happens to be the Hellinger distance [15]:

$$D_{Helli}(X, Y) = (1/\sqrt{2}) \left(\sum_{i=1}^d (\sqrt{P_{x,i}} - \sqrt{P_{y,i}})^2 \right)^{1/2}, \quad (3)$$

which is commonly used to quantify the similarity between two probability distributions.

χ^2 distance. The χ^2 distance is a similarity measure obtained from a statistical test that takes as input two distributions P_x, P_Y . This distance is defined as

$$D_{\chi^2}(X, Y) = \sum_{i=1}^d \frac{(P_{x,i} - m_i)^2}{m_i} \quad (4)$$

$$\text{with } m_i = \frac{P_{x,i} + P_{y,i}}{2}, \quad (5)$$

and measures the level of "adequacy" of pair P_x, P_Y [10]. More precisely, it corresponds to the probability that data of P_x follows the law P_y .

Kullback-Leibler divergence. The Kullback-Leibler divergence [15] is a distinguishability measure between two distributions P_x and P_y . Divergence

$S(P_x||P_y)$ tells us how much the expected lengths of a code change, when the coding is optimal but made under the assumption that X follows P_y . It is defined as:

$$S(P_x||P_y) = \sum_i P_{x,i} \log \frac{P_{x,i}}{P_{y,i}} \quad (6)$$

However, this Bregmann divergence [24] is not a metric distance since it is not symmetric. Moreover this dissimilarity is related to Fisher-Rao metric, since if we consider two distributions P_x and $P_x + dP_x$, that are close each other, their relative entropy is given by

$$S(P_x||P_x + dP_x) = \sum_i P_{x,i} \log \frac{P_{x,i}}{P_{x,i} + dP_{x,i}} \simeq (1/2) \sum_i \frac{dP_{x,i}dP_{x,i}}{P_{x,i}} \quad (7)$$

$$\simeq D_{Fisher-Rao}(P_x, P_x + dP_x). \quad (8)$$

The use of an asymmetric divergence may be problematic for some applications. This is why we try to symmetrize the Kullback-Leibler divergence. There are different ways to do it. The most commonly used in hyperspectral imaging is the Jeffreys approach [24], which leads to a divergence called in the information theory community the Jeffreys-Bregman Divergence, or in the hyperspectral community the Spectral Information Divergence (SID):

$$SID(P_x||P_y) = S(P_x||P_y) + S(P_y||P_x). \quad (9)$$

Rényi divergences. The Shannon entropy is a function that corresponds to the amount of information contained or sent by a source of information. Somehow, the more the source emits different information, the more the entropy is large. Let us consider a pdf P_x with d possible outputs. Then, Shannon defined that the amount of information produced by knowing that an event of probability $P_{x,i}$ took place can be approximate by [15, 18]: $I(P_{x,i}) = -\log(P_{x,i})$. The Shannon entropy is a mean of amount of information over P_x . It checks the postulate of additivity of information, which states that the information of two independent events is the sum of each information. However more general measures can be defined. Rényi [18] proved that to be able to verify the postulate of additivity of information, it is necessary to have entropy of the form:

$$H(P_x) = g^{-1} \left(\sum_{i=1}^d P_{x,i} g(I(P_{x,i})) \right), \quad (10)$$

with $g(x) = cx$ or $g(x) = c2^{(1-\alpha)x}$. The first case leads to the Shannon entropy, whereas the second leads to other functional class called the Rényi entropy, which is defined by:

$$H_\alpha(P_x) = \frac{1}{1-\alpha} \log \left(\sum_{i=1}^d P_{x,i}^\alpha \right), \quad (11)$$

Thus the Rényi entropy is a more flexible measure of uncertainty, where the parameter α allows different notions of information. The case $\alpha = 1$ leads to the Shannon entropy. Similarly to the Shannon entropy, it is possible to introduce divergence, that are called the Rényi divergence of order α , $\alpha > 0$ of a distribution P_x from a distribution P_y :

$$S_\alpha(P\|Q) = \frac{1}{\alpha - 1} \log \sum_i P_{x,i}^\alpha P_{y,i}^{1-\alpha}. \quad (12)$$

Parameter α involves a family divergences, where $S_{\alpha \rightarrow 1}(P_x\|P_y) = S(P_x\|P_y)$. There are two other special cases: $\alpha = 1/2$, which is: $S_{\alpha=1/2}(P_x\|P_y) = -2 \log(1 - D_{\text{Hellinger}}(X, Y)/2)$; and $\alpha = 2$ which leads to the quadratic Rényi divergence, that is a function mostly used on finance that check $S_{\alpha=2}(P_x\|P_y) = \log(1 + D_{\chi^2}(X, Y))$.

Mahalanobis distance. We have considered above that each spectrum follows a multinomial distribution. It might be possible to consider alternatively that the spectra follow normal distributions. This model is often used on hyperspectral imaging. In our case, we assume that each spectrum X follows a multivariate normal of mean itself and with a fixed covariance for all the spectra. Hence we have that $X \sim \mathcal{N}(X, \Sigma)$. It turns out that the Fisher-Rao distance between X, Y corresponds to the Mahalanobis distance [12, 13] defined as:

$$D_{\text{Mahal}}(X, Y) = (X - Y)^T \Sigma^{-1} (X - Y). \quad (13)$$

Endowed with this distance, the hyperspectral space is a submanifold of the manifold of multivariate normal distributions[13]. Moreover this kind of metric depends on the estimation of the covariance matrix. We have considered two ways to estimate the covariance matrix. The first one is the biased empirical covariance estimator:

$$\Sigma = \frac{1}{d} \sum_{i=1}^d (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T, \quad (14)$$

where $\bar{\mathbf{X}} = \frac{1}{d} \sum_{i=1}^d \mathbf{X}_i$ is the empirical mean. The second one is inspired from the work [17, 19, 21], where thanks to random projections they succeed to have a good estimator of the Mahalanobis distance. We will denote respectively the corresponding distance D_{Mahal1} and D_{Mahal2} .

Kolmogorov distance. Spectra X of a hyperspectral image can be represented by a pdf, but they can also be represented by their cumulative distribution function (cdf), denoted C_X , where $C_{x,i} = \sum_{k=1}^i P_{x,k}$. This function is smoother and less subject to high noise variation. Then, one can define the Kolmogorov-Smirnov distance [10] for two spectra X and Y as the maximal difference between their cumulative distribution functions:

$$D_{\text{Kolmo}} = \max(|P_{x,1} - P_{y,1}|, \dots, |P_{x,2} - P_{y,2}|). \quad (15)$$

This distance happens to be the L_∞ norm applied to the cdf's.

Earth Mover's distance. Let us consider two spectra X and Y represented by their respective pdf P_x and P_y . Their Earth Mover's Distance [10] can be defined as

$$D_{\text{EMD}} = \min_{\alpha_{i,j} \in \mathcal{M}} \left(\sum_{i=1}^d \sum_{j=1}^d \alpha_{i,j} C(i,j) \right), \quad (16)$$

where $\mathcal{M} = \{\alpha_{i,j} \geq 0; \sum_{i=1}^d \alpha_{i,j} = P_{y,j}; \sum_{j=1}^d \alpha_{i,j} = P_{x,i}\}$ and C is the cost function. Different choices of cost functions have been considered. We adopt here two different cost functions. The first one can be defined as:

$$C_1(i,j) = \frac{1}{d} |i - j|. \quad (17)$$

In such a case, the Earth Mover's Distance is given by [16]:

$$D_{\text{EMD1}} = \frac{1}{d} \|C_x - C_y\|_1 \quad (18)$$

One can also use a thresholded version of the cost function [16] to increase the computation speed:

$$C_2(i,j) = \begin{cases} |i - j| & \text{if } |i - j| \leq s \\ s & \text{otherwise} \end{cases} \quad (19)$$

where s is the value of the threshold. We will write this distance D_{EMD2} .

4 Results on hyperspectral images

The question of the evaluation of the metrics is crucial. We first use two images conventionally used in hyperspectral image processing: i) the Pavia image, which represents the campus of Pavia university (urban scene), of size 610×340 pixels and $d = 103$ spectral bands, and is composed of 9 classes; ii) the Indian Pines image, test site in North-western Indiana composed for two thirds of agriculture, and one-third of forest, of 145×145 pixels and $d = 224$ spectral bands, and is composed of 16 classes. On these images, we calculate on Table 1, the evaluation measures introduced in section 2. We also plotted the cumulative pattern spectra, on Fig. 2, and example of quantized Indian Pines image on Fig. 3. From this study we can deduce that Mahalanobis distance may have good results, however the quality of the results of Mahalanobis distance depends on the estimation of the covariance matrix. We can also deduce that χ^2 distance can have excellent results, and seems to be quite robust to the curse of dimensionality, using this distance in dimensionality reduction [7] might improve the extracted feature.

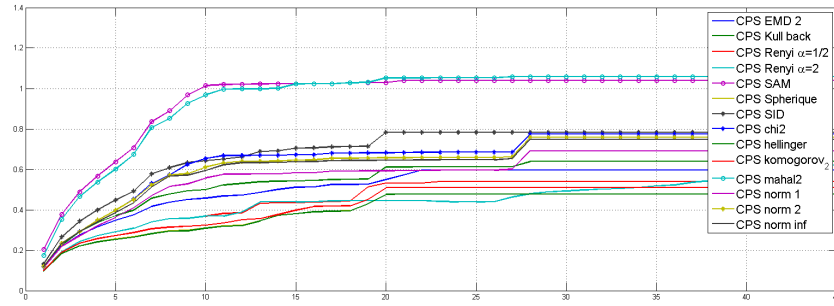


Fig. 2. Representation of the cumulative Pattern Spectrum of Pavia images for different similarity measures.

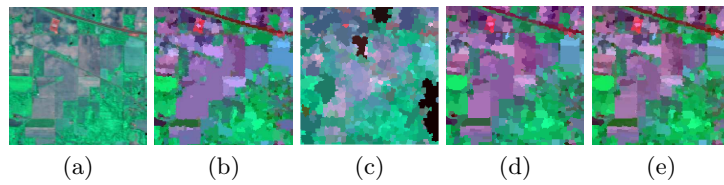


Fig. 3. (a) False RGB color image (using three spectral bands) of Indian Pines hyperspectral image. False RGB color image of the quantized hyperspectral image thanks to in (b) the norm 2, (c) the SAM, (d) the χ^2 distance, (e) the EMD.

Results on "Pavia" image																
	L_1	L_2	L_∞	D_{Spher}	D_{SAM}	D_{Helli}	D_{χ^2}	S	$S_{\alpha=1/2}$	$S_{\alpha=2}$	SID	D_{Mahal1}	D_{Mahal2}	D_{Kolmo}	D_{EMD1}	D_{EMD2}
OA	0.001	0.001	0.003	0.012	0.056	0.001	0.023	0.51	0.50	0.50	0.012	0.009	0.057	0.006	0.008	0.006
AA	0.001	0.001	0.013	0.085	0.12	0.001	0.11	0.25	0.22	0.22	0.089	0.003	0.22	0.09	0.1	0.08
Rank	0.93	0.93	0.91	0.81	0.62	0.46	0.33	0.21	0.47	0.22	0.36	0.58	0.37	0.90	0.90	0.41
SNR	22.82	22.96	22.89	22.72	14.90	21.92	23.90	21.88	21.97	21.20	21.26		21.4	22.0	22.47	
Results on "Indian Pines" image																
	L_1	L_2	L_∞	D_{Spher}	D_{SAM}	D_{Helli}	D_{χ^2}	S	$S_{\alpha=1/2}$	$S_{\alpha=2}$	SID	D_{Mahal1}	D_{Mahal2}	D_{Kolmo}	D_{EMD1}	D_{EMD2}
OA	0.016	0.016	0.011	0.016	0.09	0.016	0.30	0.010	0.010	0.010	0.010	0.019	0.068	0.0162	0.029	0.0162
AA	0.012	0.0022	0.014	0.016	0.13	0.016	0.24	0.09	0.09	0.09	0.09	0.016	0.13	0.0019	0.069	0.016
Rank	0.44	0.45	0.45	0.45	0.22	0.46	0.18	0.46	0.44	0.45	0.46	0.39	0.34	0.45	0.45	0.42
SNR	12.86	12.74	12.73	12.73	5.59	12.69	14.01	12.88	12.81	12.86	12.73		12.74	11.01	12.55	

Table 1. Comparison of probabilistic distances on hyperspectral images.

5 Conclusion

We have done a systematic study to compare and assess different probabilistic distances in the context of hyperspectral image quantization. Our results are consistent to those previously published [1, 11, 14] for some of the distances. We infer from our study the importance of using appropriate distances to address the curse of dimensionality in hyperspectral imaging. A distance that seems to be rather efficient is the χ^2 distance. However, each dissimilarity measure has its disadvantages and benefits. A potentially interesting approach would be to take a dissimilarity measure as a linear combination of dissimilarity measures, taking advantage of the discriminatory power of each of them. This kind of approach is link with the multiple kernel learning [23]. For the particular problem of image quantization, which involves computation of centroids, advanced methods to compute centroid from divergences [5] can improve the results.

References

1. Chang, C. I. (Ed.). (2003). Hyperspectral imaging: techniques for spectral detection and classification (Vol. 1). Springer Science & Business Media.
2. Tomic, I., & Frossard, P. (2011). Dictionary learning. Signal Processing Magazine, IEEE, 28(2), 27-38.
3. R. E. BELLMAN, Adaptive control processes. 1961.
4. Maragos, P. (1989). Pattern spectrum and multiscale shape representation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 11(7), 701-716.
5. Nielsen, F., & Nock, R. (2009). Sided and symmetrized Bregman centroids. Information Theory, IEEE Transactions on, 55(6), 2882-2904.
6. Schmid-Saugeon, P., & Zakhor, A. (2004). Dictionary design for matching pursuit and application to motion-compensated video coding. Circuits and Systems for Video Technology, IEEE Transactions on, 14(6), 880-886.
7. Noyel, G., Angulo, J., & Jeulin, D. (2007). Morphological segmentation of hyperspectral images. Image Anal. Stereol, 26(3), 101-109.
8. Saporta, G. (2011). Probabilits, analyse des donnes et statistique. Editions Technip.
9. Fauvel, M. (2007). Spectral and spatial methods for the classification of urban remote sensing data (Doctoral dissertation, Institut National Polytechnique de Grenoble-INPG; Universit d'Islande).
10. Rubner, Y., Tomasi, C., & Guibas, L. J. (2000). The earth mover's distance as a metric for image retrieval. International journal of computer vision, 40(2), 99-121.

11. Robila, S. (2005, March). An investigation of spectral metrics in hyperspectral image preprocessing for classification. In Geospatial goes global: from your neighborhood to the whole planet. ASPRS Annual Conference, Baltimore, Maryland.
12. Mahalanobis, P. C. (1936). On the generalized distance in statistics. Proceedings of the National Institute of Sciences (Calcutta), 2, 49-55.
13. Strapasson, J. E., Porto, J. P., & Costa, S. I. (2015, January). On bounds for the Fisher-Rao distance between multivariate normal distributions. MAXENT 2014 (Vol. 1641, pp. 313-320). AIP Publishing.
14. Paclik, P., & Duin, R. P. (2003). Dissimilarity-based classification of spectra: computational issues. Real-Time Imaging, 9(4), 237-244.
15. Bengtsson, I., & Zyczkowski, K. (2006). Geometry of quantum states: an introduction to quantum entanglement. Cambridge University Press.
16. Pele, O., & Werman, M. (2009, September). Fast and robust earth mover's distances. In Computer vision, 2009 IEEE 12th international conference.
17. Gueguen, L., Velasco-Forero, S., & Soille, P. (2014). Local mutual information for dissimilarity-based image segmentation. Journal of mathematical imaging and vision, 48(3), 625-644.
18. Renyi, A. (1961). On measures of entropy and information. In Fourth Berkeley symposium on mathematical statistics and probability (Vol. 1, pp. 547-561).
19. Velasco-Forero, S., & Angulo, J. (2012). Random projection depth for multivariate mathematical morphology. Selected Topics in Signal Processing, IEEE Journal of, 6(7), 753-763.
20. Meyer, F., & Maragos, P. (2000). Nonlinear scale-space representation with morphological levelings. Journal of Visual Communication and Image Representation, 11(2), 245-265.
21. Zuo, Y. (2003). Projection-based depth functions and associated medians. Annals of Statistics, 1460-1490.
22. Soille, P. (2008). Constrained connectivity for hierarchical image partitioning and simplification. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 30(7), 1132-1145.
23. Bach, F. R., Lanckriet, G. R., & Jordan, M. I. (2004, July). Multiple kernel learning, conic duality, and the SMO algorithm. In Proceedings of the twenty-first international conference on Machine learning (p. 6). ACM.
24. Schwander, O. (2013). Methodes de gomtrie de l'information pour les modes de mlange (Doctoral dissertation, Ecole Polytechnique X).