

A Constrained Band Selection Method Based on Information Measures for Spectral Image Color Visualization

Steven Le Moan, Alamin Mansouri, Yvon Voisin, Jon Yngve Hardeberg

▶ **To cite this version:**

Steven Le Moan, Alamin Mansouri, Yvon Voisin, Jon Yngve Hardeberg. A Constrained Band Selection Method Based on Information Measures for Spectral Image Color Visualization. IEEE Transactions on Geoscience and Remote Sensing, 2011, PP (99), pp.1-12. 10.1109/TGRS.2011.2158319 . hal-00637700

HAL Id: hal-00637700

https://u-bourgogne.hal.science/hal-00637700

Submitted on 2 Nov 2011

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.

A Constrained Band Selection Method Based on Information Measures for Spectral Image Color Visualization

Steven Le Moan, Alamin Mansouri, Yvon Voisin, and Jon Y. Hardeberg

***Abstract*—We present a new method for the visualization of spectral images, based on a selection of three relevant spectral channels to build a Red-Green-Blue composite. Band selection is achieved by means of information measures at the first, second and third orders. Irrelevant channels are preliminarily removed by means of a center-surround entropy comparison. A visualization-oriented spectrum segmentation based on the use of color matching functions allows for computational ease and adjustment of the natural rendering. Results from the proposed method are presented and objectively compared to four other dimensionality reduction techniques in terms of naturalness and informative content.**

***Index Terms*—Multi/hyperspectral imagery, Visualization, Band Selection, Information Measures, Spectrum Segmentation, Color Matching Functions**

I. Introduction

Most of today's visualization devices are based on the paradigm that a combination of three primary colors (red, green and blue) is sufficient for the human eye to characterize any color [1]. However, in many applications such as remote sensing, medical or art imaging, measuring the electromagnetic properties of a scene has to be made with high spectral precision. Analogously to the need for a high spatial resolution for an enhanced separation of the different elements of a scene, a high spectral resolution allows for a better estimation of its reflectance, and thus, a better characterization of its "true" color, regardless of the conditions of acquisition (illumination, camera). Multispectral imaging consists of acquiring more than three spectral components from a scene, usually dozens, each one of them being captured in a small range of wavelengths, for a better spectral precision (analogously to a pixel covering a small area of the space). Usually, spectral images are acquired in the visible range of wavelengths [400..700]nm, but it is also very common to cover the infrared range (beyond 700nm). Images with a number of channels ranging from a hundred to a thousand are referred to as hyperspectral and higher dimensionality datasets are called ultraspectral.

High dimensionality often goes along with high computational burden, as well as some undesirable geometrical and statistical properties [2]. Moreover, it is well-known that neighboring spectral channels are highly correlated, hence a large amount of unnecessary redundancy. For these reasons, dimensionality reduction is a very common processing. It consists of reducing the number of bands so that only a few relevant ones are kept to represent the whole image. This step is often referred to as *feature extraction* and is achieved by combining the original channels according to an information-preservative criterion whose definition is usually application-related. If the combination is linear and made under the constraint that its weighting coefficients are in $\{0, 1\}$, one talks about *band selection*. In this particular case, the reduced set is a subset of the original one. *Band selection* is often prefered since it allows for a better interpretation of the dimensionality reduction by keeping the physical meaning of the spectral channels. In this paper, we consider the general application of the visualization of spectral images by means of a tri-stimulus-based display device such as a standard computer screen. Thus, our framework is restrained to an N-to-3 dimensionality reduction dedicated to the human perception.

We propose a new band selection method in three steps: first, irrelevant bands are excluded by comparing each channel's entropy with that of its close spectral neighborhood. Then, the image's spectrum is segmented by thresholding the CIE Standard Observer Color Matching Functions (CMF). Eventually, we use normalized information measures at second and third orders to select the triplet of bands with minimal redundancy and maximal informative content. The proposed technique is referred to as Constrained Band Selection for Visualization (CBSV$_\tau$), τ being a parameter that will be later explained. We focus our study on two aspects: the **perceptual appeal** (that we will refer to as **naturalness**) and the **informative content** of the rendered composites. We will demonstrate that, although being contradictory in many cases, these objectives can indeed be balanced, according to the user's need, either manually or automatically.

The remainder of this paper is thus organized as follows: first, we review some related works on the topic of dimensionality reduction for spectral image visualization. In section 3, we give a brief background on entropy, mutual information and review several N-order information measures. Then, we present all the steps of the band selection approach. Section 5 presents the experimental framework and gives objective results proving the efficiency of our approach. Several aspects of the methods are then discussed before conclusion.

A. Mansouri and Y. Voisin are with the Laboratoire d'Electronique Informatique et Image, Université de Bourgogne, Auxerre, France. J. Hardeberg is with the Norwegian Color Research Laboratory, Gjøvik University College, Gjøvik, Norway. S. Le Moan is affiliated to both laboratories. (emails: {steven.le-moan, alamin.mansouri, yvon.voisin}@u-bourgogne.fr, jon.hardeberg@hig.no)

Manuscript received July 13, 2010; revised May 25, 2011.

II. Related work

Tri-stimulus representation of multi/hyperspectral images for visualization is an active field of research that has been thoroughly investigated over the past decades. One of the most common approaches is probably the one referred to as *true color*. It can basically be achieved in two different ways: one consists of selecting the bands at 700nm, 546nm and 436nm (or the closest) and mapping them to the three primaries: R,G and B, respectively. The other one uses the CMF-based band transformation [3]. Even though it generally yields a very natural visual rendering, this approach does not take the data itself into account at all, and thus noise, redundancy, etc. are not accurately handled.

Another very common approach for dimensionality reduction is Principal Components Analysis (PCA), which has been extensively used for visualization purposes. Tyo *et al*. [4], investigated PCA for N-to-3 dimensionality reduction into the HSV color space. An automatic method to find the origin of the HSV cone is also introduced in order to enhance the final color representation. Later, Tsagaris *et al*. [5] proposed to use the fact that the red, green and blue channels, as they are interpreted by the human eye, contain some correlation, which is in contradiction to the underlying decorrelation engendered by PCA. For that reason, the authors proposed a constrained PCA-based technique in which the eigendecomposition of the correlation matrix is forced with non-zero elements in its non-diagonal elements. Several other PCA-based visualization techniques can be found in the literature [6], [7], [8].

In order to alleviate the computational burden of the traditional PCA, Jia et al. [9] proposed a correlation-based spectrum segmentation technique so that principal components are extracted from different segments and then used for visualization. Other segmented PCA approaches are investigated in [10] including equal subgroups, maximum energy and spectral-signature-based partitioning.

In [11], Du *et al*. compared seven feature extraction techniques in terms of class separability, including PCA, Independent Components Analysis (ICA) and Linear Discriminant Analysis (LDA). ICA has also been studied by Zhu *et al*. [12] for spectral image visualization. They used several spectrum segmentation techniques (equal subgroups, correlation coefficients and RGB-based) to extract the first IC in each segment. The use of different color spaces for mapping of PCs or ICs has been investigated by Zhang *et al*. [13].

In [14], [15], Jacobson *et al*. presented a band transformation method allowing the CMF to be extended to the whole image spectrum, and not only to the visible part. They proposed a series of criteria to assess the quality of a spectral image visualization. Later, Cui *et al*. [16] proposed to derive the dimensionality reduction problem into a simple convex optimization problem. In their paper, class separability is considered and manipulations on the HSV cone allow for color adjustments on the visualization. More recently, we have proposed a method based on class-separability in the CIELAB space for improved spectral image visualization [17].

All the previously presented approaches can be referred to as *band transformation* techniques inasmuch as they produce combinations of the original spectral channels to create an enhanced representative triplet. The often mentioned drawback of this kind of approach is the loss of physical meaning attached to a channel. That is, if, initially, a spectral band is implicitly linked to a range of wavelengths, what can be told about a combination of them? As previously explained, *band selection* approaches overcome this problem by preserving the underlying physical meaning of the spectral channels, thus allowing for an easier interpretation by the human end user.

In [18], Bajcsy investigated several supervised and unsupervised criteria for band selection, including entropy, spectral derivatives, contrast, etc. Many signal processing techniques have been applied to band selection: Constrained Energy Minimization and Linear Constrained Minimum Variance [19], Orthogonal Subspace Projection (OSP) [20], [21] or the One-Bit Transform (1BT) [22]. Also information measures based on Shannon's theory of communication [23] have been proven to be very powerful in the identification of redundancy in high-dimensional datasets. Mutual Information (MI) was first used for band selection by Conese *et al*. [24]. In [25] and [26], two metrics based on MI are introduced in the context of image fusion evaluation. They measure how much information is shared by the original and the reduced datasets. In [27], MI is used to measure the similarity of each band with an estimated reference map. In [28], a normalized MI measure is used for hierarchical spectrum segmentation. More recently, Cariou *et al*. proposed an MI-based top-down band clustering technique [29].

To our knowledge, the use of third order information measures has not been investigated in the framework of band selection yet. Along with a visualization-oriented spectrum segmentation constraint, these are the main contributions of this paper.

III. Theoretical background

This section gives some background information about Shannon's mutual information and its generalizations to higher orders. When these measures are applied to a spectral image, it is generally considered that each channel is equivalent to a random variable X and all its pixels are events of X.

A. First and second orders

Entropy and mutual information were first introduced by Shannon [23]. In this part, we simply recall the formulas of (in the discrete case):

- the *entropy* of a random variable X:

$$H(X) = -\sum_{x \in \Xi(X)} p_X(x) \log_b(p_X(x)) \qquad (1)$$

  where x is an event of X, $\Xi(X)$ is the ensemble of possible values for X, $p_X(x)$ is the probability density of X and b is the order of the logarithm, usually set to 10.

- the *mutual information* between two random variables X and Y:

$$I(X;Y) = \sum_{\substack{x\in\Xi(X)\\ y\in\Xi(Y)}} p_{X,Y}(x,y)\log(\frac{p_{X,Y}(x,y)}{p_X(x).p_Y(y)}) = H(X) + H(Y) - H(X;Y) \quad (2)$$

where $p_{X,Y}(x,y)$ is the joint probability density of the couple (X,Y) and $H(X,Y)$ the joint entropy of the couple (X,Y).

- The *conditional mutual information* between two random variables X and Y, knowing a third one, Z:

$$I(X;Y|Z) = E(I(X;Y)|Z) \quad (3)$$

where $E(.)$ is the mathematical expectation.

- The *mutual information between two sets* of random variables:

$$\begin{aligned} I((X_1,X_2,X_3);(Y_1,Y_2)) =& H(X_1,X_2,X_3) \\ &+ H(Y_1,Y_2) \\ &- H((X_1,X_2,X_3);(Y_1,Y_2)) \end{aligned} \quad (4)$$

The well-known Venn diagram allows for a good understanding of these metrics, as shown in Figure 1.

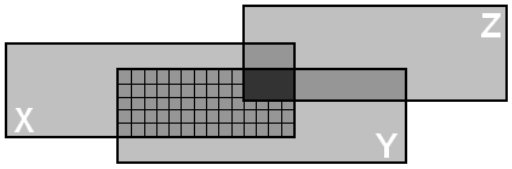


Figure 1. Venn representation of mutual information. The three rectangles depict the entropy of a random variable each. The overlappings represent the information shared by several variables. The squared area represents the conditional mutual information $I(X;Y|Z)$.

B. Third order

Several generalizations of mutual information have been proposed in the literature. In this section, for purposes of clarity, we will focus solely on the third order. However, extension to higher orders is quite straight forward.

Watanabe [30] introduced the *total correlation*, also known as *multivariate constraint* [31] or *multiinformation* [32]. It is defined as the difference between the sum of marginal entropies and the joint entropy of the set:

$$TC(X;Y;Z) = \sum_{i\in\{X;Y;Z\}} H(i) - H(X;Y;Z) \quad (5)$$

The main drawback of total correlation lays in the fact that it measures both second and third order, indiscriminately, while favoring the second order.

McGill [33] presented the *interaction information*, which is defined as follows:

$$\begin{aligned} A(X;Y;Z) &= I(X;Y|Z) - I(X;Y) \\ &= I(X;Z|Y) - I(X;Z) \\ &= I(Y;Z|X) - I(Y;Z) \end{aligned} \quad (6)$$

It can also be written as a sum of entropies at all orders:

$$\begin{aligned} A(X;Y;Z) = &- \sum_{i\in\{X;Y;Z\}} H(i) \\ &+ \sum_{\substack{i\in\{X;Y;Z\}\\ j\in\{X;Y;Z\}\setminus i}} H(i,j) \\ &- H(X;Y;Z) \end{aligned} \quad (7)$$

More recently, Bell [34] proposed the *co-information* which is identical to the interaction information, but with an opposite sign:

$$I(X;Y;Z) = -A(X;Y;Z) \quad (8)$$

A particularly interesting property of these measures is that they can take both positive and negative values. If we look at the Venn diagram in the last section, we naturally point out the center region as the third order redundancy. However, because of the very existence of the negative case, the Venn diagram representation is no longer relevant at orders higher than two. For the same reason, both aforementioned definitions are correct, nevertheless, we find Bell's definition more suitable and more intuitive for the following explanations.

In the case of positive co-information, we talk about redundancy, whereas in the case of negative values, we talk about synergy. Redundancies are foreseeable from lower orders while synergies only appear when the random variables are taken together. If we refer to equation 6, the synergy case appears when, for instance, $I(X;Y|Z) > I(X;Y)$ that is, when the knowledge of Z increases the dependency between X and Y. In order to explain this particular property, we consider a simple XOR cell with two binary inputs, X and Y and an output $Z = X \oplus Y$. If we consider the inputs as independent, the following stands true: $I(X;Y) = 0$. If we now introduce the knowledge of Z, we also introduce the underlying knowledge of the XOR relation linking the three variables. For instance, if we know that $Z = 0$, we can deduce that $X = Y$, and, by this, we increase the dependency between the inputs so that $I(X;Y|Z) > I(X;Y)$.

In the case of spectral images, this principle remains true. The knowledge of one channel can increase the mutual information between the two others and, in that case, the smaller the co-information, the higher the shared information. Therefore, in the context of minimizing the redundancy (generally speaking) inside a set of random variables, the absolute value of co-information must be minimal.

IV. Proposed method

In this section, we describe the different stages of the CBSV$_\tau$ method: exclusion of irrelevant channels, spectrum segmentation and band selection.

A. Exclusion of irrelevant bands

We propose to make a first coarse selection allowing for the removal of low informative channels. Multispectral images are known to contain high redundancy between neighboring bands and thus, channels that do not meet this definition are considered noisy or poorly informative [35]. Either way, these bands can be excluded. Measuring the similarity by means of second- order measures such as correlation or mutual information requires however a high computational burden, since all the pairs of bands must be considered. Another way of coarsely measuring similarity can be done by comparing the intrinsic informative content of a band with that of its neighbors. This can be achieved by means of a moving average as local threshold [36], [22]. We propose to use Shannon's entropy as a measure of informative content. It is computed for each channel, resulting in the solid curve in Figure 2 (for the "Jasper Ridge" image, see description in the next section).

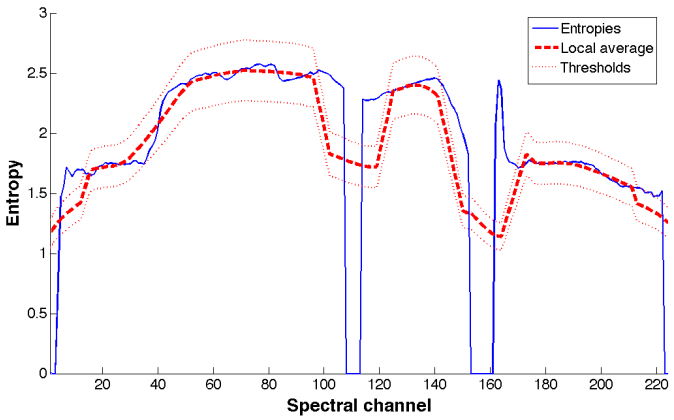


Figure 2. Exclusion of irrelevant bands: entropy, local average and thresholds for the "Jasper Ridge" image. Channels outside the rejection thresholds are excluded ($\delta = 10\%$ and $s = 11$).

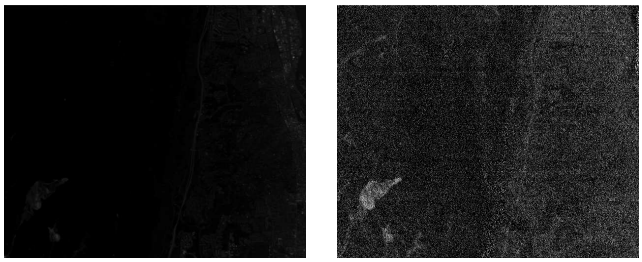
The dotted line represents the local average, which is defined for a band B_i as follows:

$$\bar{H}_s(B_i) = \frac{1}{s} \sum_{k=-\lfloor s/2 \rfloor}^{\lfloor s/2 \rfloor} H(B_{i+k}) \tag{9}$$

with s representing the size of the neighborhood. Bands whose entropy is higher (resp. lower) than their local average value moderated by the thresholding factor δ are then considered as irrelevant. Thus, if a band B_i reaches the condition in equation 10, it is excluded.

$$H(B_i) \notin [\bar{H}_s(B_i) * (100 - \delta); \bar{H}_s(B_i) * (100 + \delta)] \tag{10}$$

With δ, the thresholding parameter, given in percentage. Both the size of the window and the thresholds have to be set according to the smoothness of the entropy curve. Indeed, the smoother the latter, the lower the probability to have irrelevant bands and, accordingly, the smaller the amount of channels that fall outside the range of relevance. In that case, small value of δ and s are advised in order to gain in precision. On the other hand, a sharp-shaped curve implies strong differences between neighboring channels. In such a case, a large window size is prefered in order to reduce the influence of outliers. This yields a smoother local average curve, and also a less precise analysis. δ allows for adjusting how "strong" the exclusion will be (the lower, the stronger).



(a) (b)

Figure 3. Example of excluded bands: (a) num 5 (b) num 164. The first one falls below the relevance range because of its very low mean value. The second one is rejected for it contains to much spatial noise.

In the case of the example image "Jasper Ridge", one can notice steep curvature changes, hence our choice to use a large window size ($s = 11$). Moreover, since we first aim at a coarse band selection, we set the threshold to a moderate value ($\delta = 10\%$). Figure 3 shows examples of excluded channels for this image.

B. Spectrum segmentation

1) Principle: Spectrum segmentation aims at regrouping spectral channels so that bands of a same segment are considered similar in some way, in order to alleviate the computation of the feature extraction. Usually, similarity is measured in terms of shared information and groups are drawn contiguously among the spectrum. We propose to measure similarity in terms of human vision. At this aim, we propose to use the CIE 1964 Supplementary Standard Colorimetric Observer Color Matching Functions (CMF) [37] which are descriptors of the chromatic response of the human eye. The CMF are usually used to linearly combine spectral channels into a tristimulus (XYZ, RGB) representation roughly matching the human perception of Red, Green and Blue [38]. In other words, each wavelength is associated with three weighting coefficients corresponding to its contributions to the perception of three primary colors. We propose to interpret this statement as follows: the higher the weighting coefficient W_i^p of a channel $i \in [1..N]$ in the component $p \in \{R; G; B\}$, the higher the relevance for W_i^p to be a good representative of p. Consequently, we propose to cluster the CMF coefficients into two classes, by means of a binarizing threshold τ. Coefficients above τ depict the relevant wavelengths for band selection. We note the ensemble of the corresponding channels Seg_p^τ, $p \in \{R, G, B\}$.

Let us now consider the common case of spectral image ranging outside the visible range of wavelengths (400-700nm). Indeed, the CMF are designed only for this part of the electromagnetic spectrum. As a solution to this, Jacobson *et al.* [14] proposed to stretch the CMF so that it covers the entire image spectrum, no matter what wavelengths it ranges in. This stretched CMF principle is illustrated by Figure 4, for an image

covering the range [300..1000] nanometers. In the case of a non-constant spectral sampling step, that is, for example, when bands have been removed by the previous step, either the lacking channels must be replaced by interpolation methods, or the CMF coefficients must be adjusted. For computational ease, we recommend the latter solution.

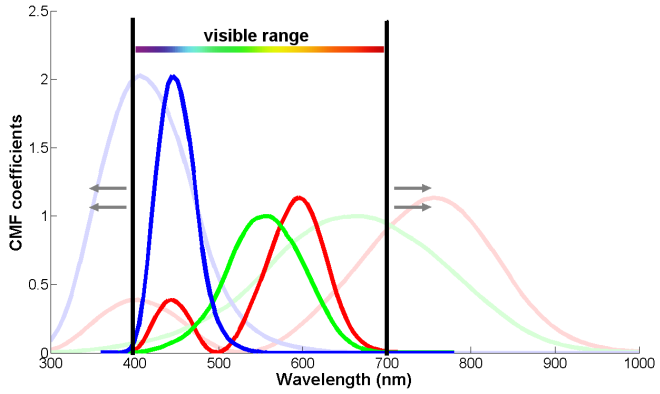


Figure 4. The stretched CMF principle: In strong colors, the original functions. In light colors, the same ones stretched to fit a larger range of wavelengths.

Eventually, three segments are obtained, depending on the binarization threshold: Seg_R^τ, Seg_G^τ and Seg_B^τ in which the band selection for the red, green and blue primitive channels will be performed, respectively. Consequently, for a growing value of τ, the size of segments gets smaller and:

$$\tau_1 > \tau_2 \rightarrow Seg_p^{\tau_2} \in Seg_p^{\tau_1}, \forall p \in \{R, G, B\} \qquad (11)$$

The spectrum segmentation is performed using normalized functions so that $\max_i(W_i^p) = 1, \forall P$. Figure 5 illustrates the technique as well as the role of τ.

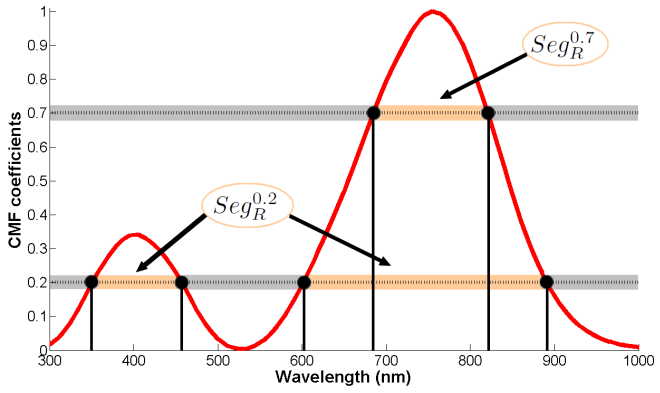


Figure 5. The spectrum segmentation. The curve represents the red CMF, normalized between 0 and 1 and stretched between 300 and 1000 nm. Extremum values have been removed for clarity. The horizontal lines represent two cases of spectrum segmentation, one for $\tau = 0.2$ and another for $\tau = 0.7$. In both cases, the grey segments highlight the removal areas while bands in the orange zones are kept.

We note two particular cases : if τ is set to 0, the hypothesis is rejected and band selection is totally unconstrained. On the contrary, if $\tau = 1$, the hypothesis is considered perfectly relevant and there is no need to proceed with band selection since, in that case, the size of each segment is reduced to 1.

2) Automatic thresholding: Setting the parameter τ may be quite challenging, especially in the absence of a specific application. As will be seen and discussed in the results section, a manual setting allows to subjectively moderate the natural aspect of the result, but there is no guarantee to find a general optimum. What we suggest here is to empirically find a suboptimal solution based on a maximization of the number of discarded channels. Indeed, as explained in the previous section, not only does this step provide a relevant spectrum segmentation, it also allows to discard a certain amount of channels prior to band selection. Therefore, we propose to define the best thresholding as the one maximizing the amount of discarded information, under the constraint that the latter must always be inferior to the amount of preserved information. As a measure of informative content, we have used Shannon's entropy. Let S_τ^- be the set of channels removed by thresholding the CMF with τ and let S_τ^+ be its complementary set. Then this pseudo-optimal τ is given by the following equation:

$$\tau_{p-opt} = \arg\max_\tau Card(S_\tau^-) \text{ w.t.c. } H(S_\tau^-) < H(S_\tau^+)$$

We note two particular situations:

- If $H(S_\tau^-) > H(S_\tau^+)$ with $Card(S_\tau^-) = 1$, the default value is set to 0.
- If $H(S_\tau^-) < H(S_\tau^+)$ for every τ, then the default value is set to 1.

Let N be the dimensionality of the image and H_{total} its overall entropy. The curves on Figure 6 show the evolution of $\frac{H(S_\tau^-)}{H_{total}}$ for all the datasets described in section V.A.

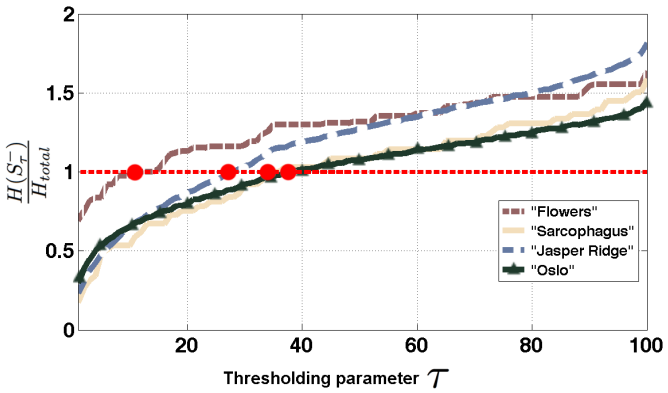


Figure 6. The proportion of discarded entropy in regards to τ

The crossing of the red dotted line indicates that the discarded information is higher than the remaining one, that is, the pseudo-optimal thresholding value according to the aforementioned criterion (depicted by red spots).

C. Band selection algorithm

The dimensionality reduction of the multispectral image must be carried out by optimizing two criteria:

- The **informative content** of the reduced dataset, which has to be maximal.
- Its **intrinsic redundancy**, which has to be minimal.

Hence the need for a metric measuring both quantities. In [28], the authors used a normalized version of mutual information that we propose to generalize for co-information and higher order information measures. We define the k-th order Normalized Information (NI) of the bandset $S = \{B_1, .., B_k\}$ as the following quantity:

$$NI_k(S) = \frac{k \times I(S)}{\sum_{i=1}^{k} H(B_i)} \tag{12}$$

Because a direct search for the best triplet of bands $B_{red}, B_{green}, B_{blue}$ would require a very high computational time, we propose to proceed iteratively, by using an algorithm similar to the one used in [21]. First, the Red and Blue primitive channels are sought as the most dissimilar couple of bands in the ensemble $Seg_R \times Seg_B$. This choice is guided by the fact that the couple R,B presents more orthogonality than any other couple in R,G,B, a property that is suggested by the very shape of the CMF and by the way they overlap along the spectrum. Then, the Green channel is selected as the one minimizing $|NI(B_{red}, B_k, B_{blue})|$. The procedure is detailed in algorithm 1.

Algorithm 1 Band selection

$i = 0$; $k = 1$;
randomly choose j so that $B_j \in Seg_R^\tau$;
iterations = 0;
while (i != k) and (iterations < 20) **do**
  if iterations is even **then**
    $\Xi = Seg_B^\tau$;
  else
    $\Xi = Seg_R^\tau$;
  end if
  find $temp = \arg\min_k [NI_2(B_j, B_k)]$ with $B_k \in \Xi$;
  $i \leftarrow j$; $j \leftarrow k$; $k \leftarrow temp$;
  iterations++;
end while
if iterations is even **then**
  $B_{red} = B_k$;
  $B_{blue} = B_j$;
else
  $B_{red} = B_j$;
  $B_{blue} = B_k$;
end if
find $k = \arg\min_k |NI_3(B_{red}, B_k, B_{blue})|$ with $B_k \in Seg_G^\tau$.
$B_{green} = B_k$;

V. Experiments and results

In this section, we first present the datasets, the metrics as well as the other techniques used for comparison. Computational consideration are then briefly tackled before presenting results both before and after spectrum segmentation, for a thorough comparison. The influence of τ is then discussed.

A. Data sets

For our experiments, we used four different reflectance datasets which are now presented:

- "Jasper Ridge" is a well-known 220 bands hyperspectral image from the AVIRIS sensor [39]. 5 classes were considered: water, road, urban area (houses), and two types of minerals.
- "Oslo" is a 160 bands remote sensing hyperspectral image, representing a urban area in the neighborhood of Oslo (Norway). It was acquired with the HySpex VNIR-1600 sensor, developed by the Norsk Elektro Optikk company. The sensor ranges from the early visible (400nm) to the near infrared (1000nm) with a spectral resolution of 3.7 nm. More information can be found on the constructor's website [40]. We considered 5 classes in this image: vegetation, road, roof tops (two kinds) and cars.
- "Sarcophagus" is a 35 bands multispectral image representing a portion of a 3^{rd} century sarcophagus from the St Matthias abbey in Trier, Germany [41]. It was acquired by means of a 8 channels filter wheel camera ranging only in the visible spectrum (400-740nm). A supervised neural-network-based reflectance estimation algorithm allowed for an enhancement of the sampling step to 10 nm. In this image, we considered solely the 24 patches of the MacBeth CC target for pixel classification.
- "Flowers" is a 31 bands multispectral image from the database used in [42]. Three classes are present on this image: flower, leaves and background.

Each reflectance dataset has been normalized so that it ranges in $[0..1]$ before dimensionality reduction and the representations are depicted here in their raw form in order to allow for a fair visual comparison. Only lightnesses have been equally increased in all the following figures in order to better the readableness of this document. Figure 7 depicts the *true color* representations of the datasets. They have been computed by means of the standard CMF-based transformation, after removal of irrelevant channels (see section IV.A) and without atmospheric correction.

B. Metrics

Two criteria have been used to assess the efficiency of the methods:

- the **natural rendering** of the composites. Subjectively measuring the naturalness of an image is really challenging since there is no exact definition for it, even though there have been some attempt to define it in a statistical manner [43]. In this study, we propose to make the assumption that the *true color* CMF-based transformation yields a reference image for naturalness. Consequently, the more natural an image is, the closer (in a perceptual fashion) it is to its *true color* representation. Thus, naturalness has been measured by computing the average euclidean perceptual distances of the resulting images with their respective CMF-based 'true color' representation. This metric is hereby referred to as **NR** (Natural Rendering).

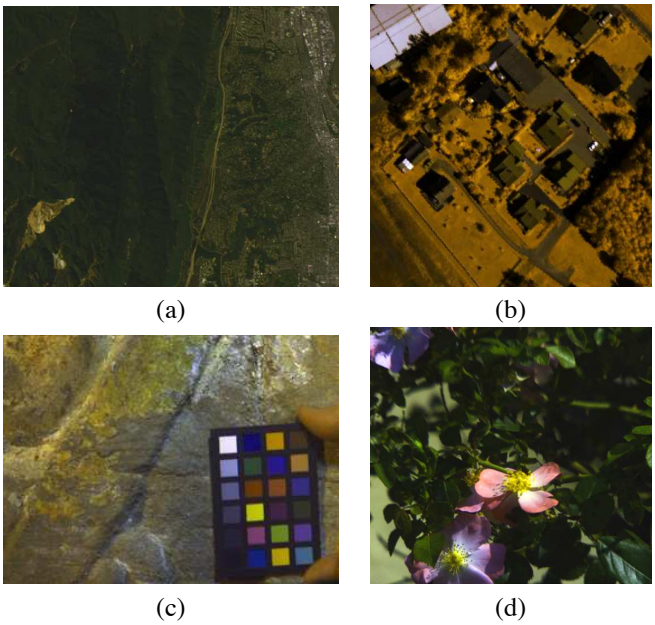


Figure 7. "True color" representations of the datasets. (a) "Jasper Ridge" (b) "Oslo" (c) "Sarcophagus" (d) "Flowers"

- The **visual informative content**, which has been measured by the average inter-class perceptual distance, as suggested by Du *et al*. [11]. At this aim, we have manually selected 20 pixels by class, in each image. Then, each class centroid has been computed and projected in a perceptually uniform color space. The euclidean distances between each couple of centroids have then been averaged. This metric will be referred to as **ICPD** (Inter-Class Perceptual Distance).

For both metrics, we have projected the RGB data into the L*a*b* color space, by assuming a D65 illuminant. We used the following ranges: L* $\in [0..100]$, a* $\in [-110..110]$ and b* $\in [-110..110]$.

C. Benchmarking methods

We propose to compare our approach with four other methods.

- The OSP-based band selection (OSP) [21] which consists of progressively selecting the channel which is the most orthogonal to the group of bands already selected. Due to the high memory requirements of this method, a spatial subsampling is necessary. According to Du *et al*., the subsampling rate can be chosen as high as 1:100 (only 1% of the pixels are kept) without affecting the results. This rate has been applied in this study.
- The 1BT-based band selection (1BT) [22]. The one-bit-transform (1BT) allows for a measure of the informative content of each channel by analyzing its structure and thus presents itself as an alternative to entropy.
- The "True Color" band transformation (TC) which consists of transformation from reflectance to XYZ then to RGB by means of the CIE standard observer CMF.
- The Principal Components Analysis (PCA) used with the HSV color space. PC_i being the i-th principal component, we have used the following mapping: $PC_1 \rightarrow V$, $PC_2 \rightarrow H$, $PC_3 \rightarrow S$.

D. Computational considerations

The histogram estimation involved in the entropy computation can be optimized by selecting the adequate number of bins. If a channel's pixels are represented by one byte each, it is pointless to take more than 256 bins, however, we can choose a lower number to improve the computational efficiency without changing the accuracy of the measure. In this paper, all entropy computations have been achieved with 32 bins with no significative difference on the results.

E. Comparison prior to spectrum segmentation

In this section, we focus on the evaluation of dimensionality reduction techniques applied on raw data, that is, prior to spectrum segmentation.

Figure 8 shows the obtained composites for each dataset by means of PCA, OSP, 1BT and CBSV_0. In order to obtain a fair comparison with the OSP method, each dataset has been spatially subsampled by a ratio of 1:100 before dimensionality reduction. Tables I and II give the corresponding values of both **NR** and **ICPD**.

	PCA	**OSP**	**1BT**	**CBSV_0**
"Jasper Ridge"	60.21	64.32	54.01	**44.23**
"Oslo"	67.23	57.00	50.33	**47.41**
"Sarcophagus"	31.11	**24.32**	37.76	26.14
"Flowers"	38.22	31.40	32.12	**20.39**

Table I
NATURAL RENDERINGS

	TC	**PCA**	**OSP**	**1BT**	**CBSV_0**
"Jasper Ridge"	17.74	26.18	26.33	30.67	**40.44**
"Oslo"	19.62	**25.22**	23.02	23.63	23.77
"Sarcophagus"	39.91	35.09	**47.27**	28.39	41.70
"Flowers"	104.95	**137.00**	87.78	88.29	118.00

Table II
INTER-CLASS PERCEPTUAL DISTANCES

High color differences are observed between the results, and particularly, hue differences. For instance, on the "Oslo" image, one can notice blue-, red-, yellow- and even orange-tinted renderings. This shows how numerous are the possibilities when it comes to visualizing spectral images, and illustrates the need for a common reference, namely the naturalness.

On the first two datasets, the 1BT-based band selection composites contain more spatial noise than the others (see red pixels on the AVIRIS image). It seems fair to assume that the information measurement on which is based this technique tends to be mislead by spatial noise. Hence the usefulness of the prior removal of irrelevant channels. In the case of the two multispectral datasets ("Sarcophagus" and "Flowers"), the amount of spatial noise along the channels being less important, this effect is not as pronounced. However, on these latter, one can observe a very poor contrast in terms of colors (no colors at all for the "Sarcophagus" image and high correlation of blue and green for the "Flowers" one), resulting from the redundancy between spectral channels. Indeed, the

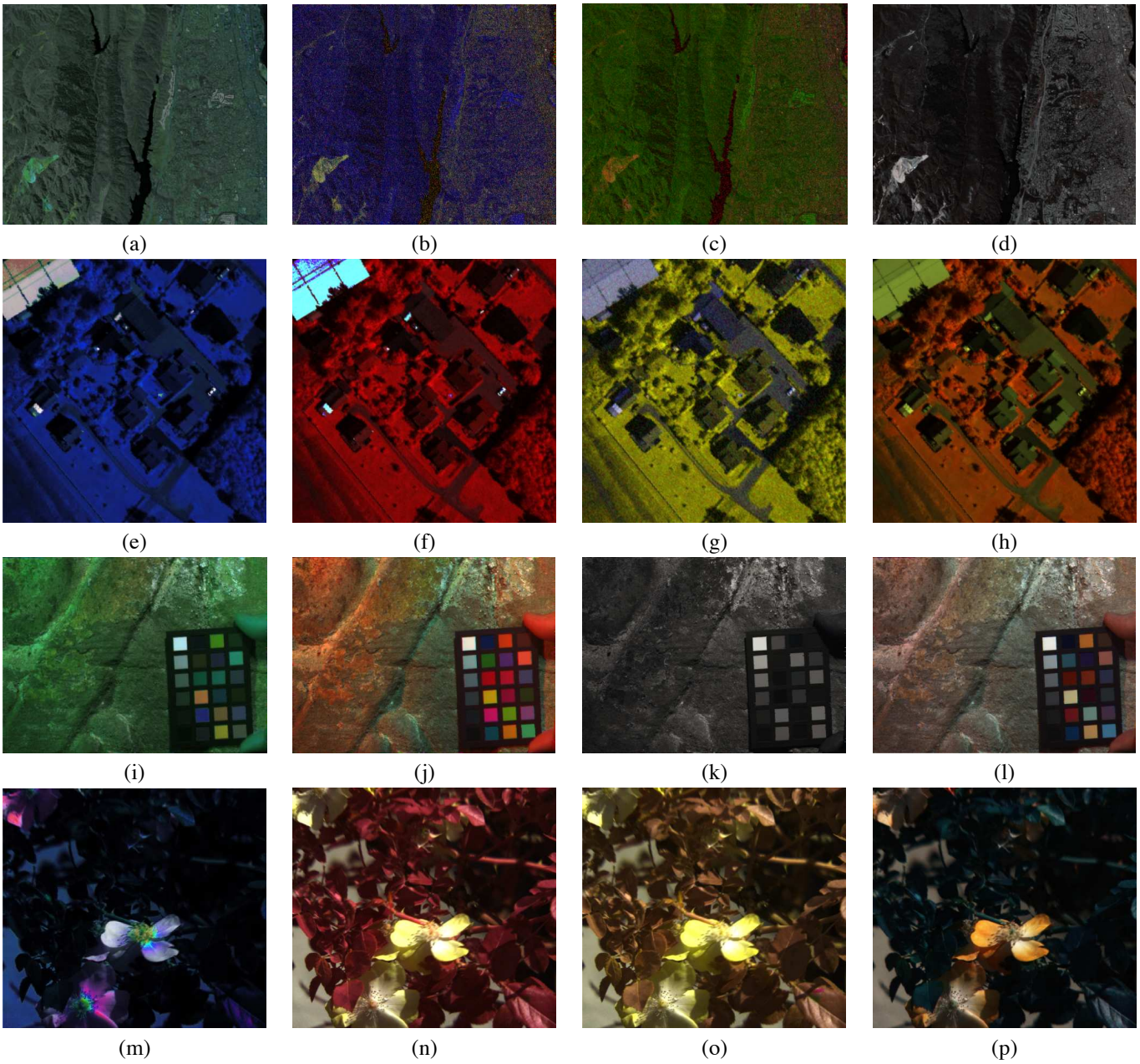


Figure 8. Rendered composites before spectrum segmentation. Column-wise: PCA, OSP, 1BT and $CBSV_0$.

1BT-based band selection does not allow for an efficient decorrelation of the selected bands.

The OSP-based approach, on the other hand, seems to perform very well on the "Sarcophagus" image as it looks quite appealing, with a very high contrast of colors, quite natural as well (the MacBeth patches look accurately depicted). On the other images, though relatively good contrasts are obtained, the technique doesn't show good performances in terms of natural rendering. Moreover, the same remark as before applies for the "Jasper Ridge" image, of which composite contains spatial noise.

The $CBSV_\tau$ technique manages to avoid the selection of noisy channels in the case of the first image, but yields a poorly colorful composite, though highly contrasted. Overall, table I indicates that the most appealing results are given by our approach, except for the "Sarcophagus" image, which is the only one whose number of channels has been artificially increased (see data description). On the first dataset, it gives however a poorly colorful result, but natural-looking nonetheless, and also with the highest ICPD, due to high luminance contrasts. Regarding class-separability, if we exclude again the "Sarcophagus" image, $CBSV_\tau$ outperforms the two other band selection methods according to table II. This indicates that the information-measures-based criterion is very efficient in conveying energy from the original high-dimensional image.

Overall, it is fair to conclude that $CBSV_0$ allows for the best tradeoff between both metrics.

F. Comparison post-spectrum segmentation

In this section, we compare only the band selection criteria, as used on the segmented data, in order to demonstrate the actual contribution of the spectrum segmentation to band selection. Moreover, we have also applied the "irrelevant channels removal" step described in section IV.A, so as to make the CBSV$_\tau$ method complete. We have then computed the pseudo-optimal thresholding values following the procedure described in section IV.B.2. (see table III) and applied these values for the results given in tables IV and V.

	Pseudo-optimal τ value
"Jasper Ridge"	0.28
"Oslo"	0.37
"Sarcophagus"	0.34
"Flowers"	0.11

Table III
PSEUDO-OPTIMAL VALUES OF τ

	OSP	**1BT**	**CBSV$_{p-opt}$**
"Jasper Ridge"	28.88	26.01	**24.32**
"Oslo"	37.00	45.78	**33.67**
"Sarcophagus"	31.11	35.05	**23.21**
"Flowers"	27.80	24.11	**11.85**

Table IV
NATURAL RENDERINGS

	OSP	**1BT**	**CBSV$_{p-opt}$**
"Jasper Ridge"	21.33	25.67	**37.21**
"Oslo"	26.02	26.78	**28.03**
"Sarcophagus"	**47.27**	28.37	43.09
"Flowers"	72.89	89.25	**134.61**

Table V
INTER-CLASS PERCEPTUAL DISTANCES

We can observe that the final composites look now much more natural and that a smaller variety of hues is rendered. This is due to the constraint spawned by the spectrum segmentation, forcing the Red, Green and Blue channels to be sought in the relevant regions of the spectrum.

We also notice no more noisy rendering, as noisy channels have been successfully removed by the pre-processing step.

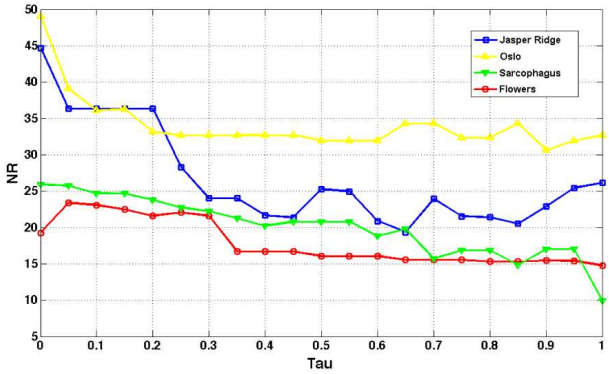
Table IV allows us to remark that the inclusion of the spectrum segmentation yields an overall increase of naturalness. CBSV$_\tau$ once again outperforms the other technique, but this time on each dataset.

In terms of class-separability, as expected, values are overall lower than without constraining the dimensionality reduction. Only one configuration gives favor to the OSP-based band selection: the composite from the "Sarcophagus" image, which is the same as the one obtained previously. However, in this case, our method gives a fairly good result and ranks second. On the same image, the 1BT-based band selection is barely more colorful than without spectrum segmentation, but we can still observe a greenish hue. Moreover, one of the main advantages of CBSV$_\tau$ over OSP band selection is the computational burden. Indeed, we have measured computing times twice to three times lower with our method, with the same configuration (same data, same subsampling).

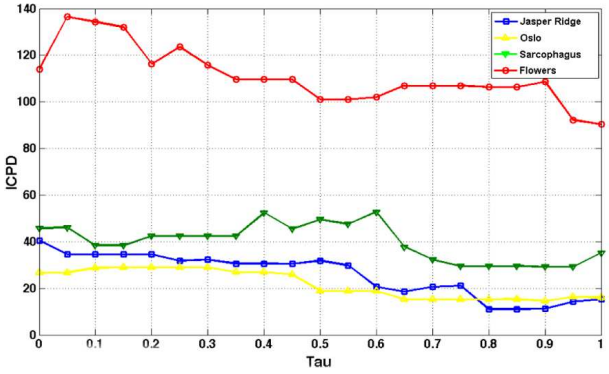
Therefore, we can conclude that the addition of both pre-processing steps allows for an overall increase of the naturalness, the best results being obtained by CBSV$_{p-opt}$, while moderately decreasing the intrinsic class-separability and reducing noise.

G. On the influence of τ

In this section, we show how the parameter τ affects the results. Figures 10a-b depict the evolution of, respectively, the **NR** and **ICPD** metrics versus τ.



(a)



(b)

Figure 10. Influence of τ on: (a) NR (b) ICPD.

As shown on figures 10a-b, τ has no direct influence on the metrics considered. Nevertheless, one can notice some overall behaviors: as τ is increased, naturalness is increased and informative content is decreased. This tells us that, by constraining too much the band selection, less energy is conveyed from the initial dataset, while the spectrum segmentation leads progressively towards a "true color" composite with maximal naturalness.

Despite the fact that we proposed an automatic method for the computation of an optimal τ, one may want to adjust it manually. The following remarks aim at helping the user in doing so. Due to the diversity of the datasets, and because the CMF-based spectrum segmentation presented in this paper is data-independent, the influence of τ is very challenging to properly understand. Indeed, none of the curves depicted above is perfectly monotonous, and no generic behavior can be drawn from them. One also have to consider the fact that both extremities ($\tau = 0$ and $\tau = 1$) represent particular cases and thus potential outliers. However, the whole purpose of the constraint is to allow for balancing between naturalness and

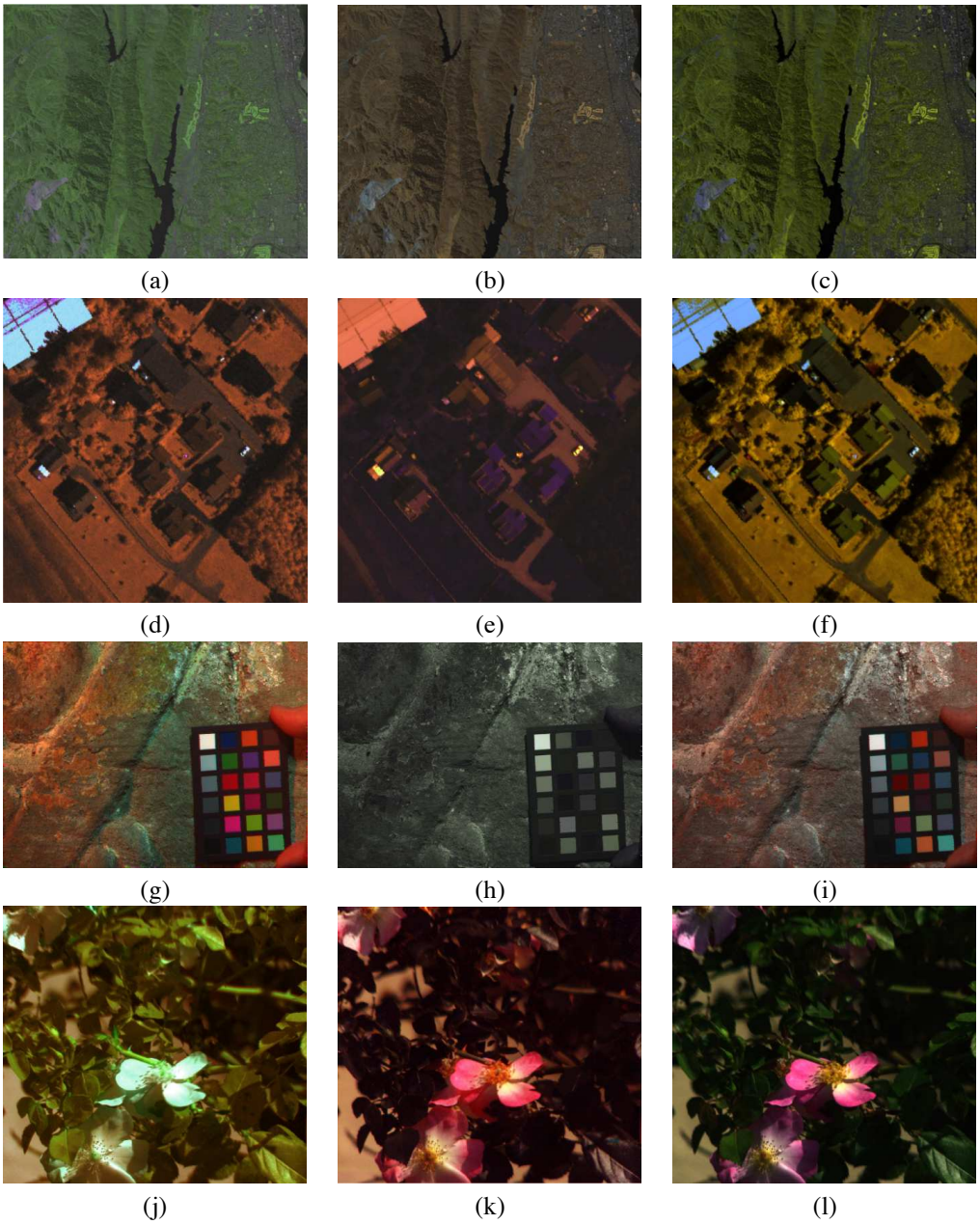


Figure 9. Rendered composites after spectrum segmentation. Column-wise: OSP, 1BT and CBSV$_{p-opt}$.

visual informative content, while also reducing computational burden when increased. In the results, even if the CBSV$_\tau$ method can be outperformed in certain configurations, it gives the best tradeoff between those characteristics.

VI. Discussions

A. Selection versus transformation

Results from the Principal Components Analysis are generally highly contrasted but poorly natural. On the other hand, results from the TC band transformation are, by definition, perfectly natural but relatively poor when it comes to informative content. From these remarks, we can assume that band selection can outperform band transformation in finding the best compromise between both characteristics. However, we believe that whether or not band transformation can surpass band selection is not the point. These are two different philosophies and band selection has an important advantage over transformation. Since the underlying task of visualization is the interpretation of the image, we believe that preserving the physical meaning of spectral channels allows for a better understanding, and also an easier interaction for the end user.

B. Limitations of the probabilistic information measures

One major drawback of the information measures presented in this paper is that they do not take into account the order in which the values appear. That is, in the case of images, the spatial information. For this reason, two different images with the same histogram cannot be distinguished by their respective entropies and the same problem appears for mutual information and its generalizations at the N-th order. Several

attempts have been made to enhance the standard entropy measure by including spatial information. In [44], the authors have proposed to no longer consider a random variable for each channel, but rather couples of neighboring pixels. Even if this improvement appears to be useful in the context of comparing low-correlated and highly textured data, it is not relevant in the general case of spectral images. During our experiments, we did not notice any major influence on the results when enhanced entropy was used instead of the standard one. On the contrary, computational burden was considerably increased.

C. Extension to general band selection

By removing the spectrum segmentation step, our approach can easily be extended to the selection of N bands for other purposes than visualization (classification, compression). In this case, some precautions have to be taken when computing high-dimensional entropies. Indeed, this requires the manipulation of N-dimensional matrices, and the computational requirement of our method would then be increasingly unfeasible. Methods consisting of approximating joint probability density functions (PDF) from lower order PDFs [45] allow for the breaking up of such complexity.

VII. Conclusions

A new method for RGB-based visualization of spectral images called CBSV$_\tau$ has been proposed. Information measures at three different orders are used to compare spectral channels with each other and select the best triplet, minimizing redundancy and maximizing informative content. A visualization-oriented spectrum segmentation approach has been presented and the role of its parameter has been discussed. Objective results have been presented and compared to four other techniques, assessing the efficiency of the proposed approach in the rendering of visual information.

VIII. Acknowledgements

The authors would like to thank the Regional Council of Burgundy for supporting this work, the Norsk Elektro Optikk in Oslo as well as the i3Mainz lab in Mainz and the Institut Für Steinkonservierung for providing useful data.

References

[1] H. Grassmann, "On the theory of compound colors," *Phil. Mag*, vol. 7, pp. 254–64, 1854.

[2] LO Jimenez and DA Landgrebe, "Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data," *IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 28, no. 1, pp. 39–54, 1998.

[3] G. Poldera and G.W.A.M. van der Heijden, "Visualization of spectral images," in *Proc. SPIE*, 2001, vol. 4553, p. 133.

[4] J.S. Tyo, A. Konsolakis, D.I. Diersen, and R.C. Olsen, "Principal-components-based display strategy for spectral imagery," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 41, no. 3, pp. 708–718, 2003.

[5] V. Tsagaris and V. Anastassopoulos, "Multispectral image fusion for improved rgb representation based on perceptual attributes," *International Journal of Remote Sensing*, vol. 26, no. 15, pp. 3241–3254, 2005.

[6] JM Durand and YH Kerr, "An improved decorrelation method for the efficient display of multispectral data," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 27, no. 5, pp. 611–619, 1989.

[7] P. Scheunders, "Multispectral image fusion using local mapping techniques," in *International conference on pattern recognition*, 2000, vol. 15, pp. 311–314.

[8] C. Yang, L. Lu, H. Lin, R. Guan, X. Shi, and Y. Liang, "A fuzzy-statistics-based principal component analysis (fs-pca) method for multispectral image enhancement and display," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 46, no. 11, pp. 3937–3947, 2008.

[9] X. Jia and J. Richards, "Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 37, no. 1, pp. 538–542, 1999.

[10] V. Tsagaris, V. Anastassopoulos, and GA Lampropoulos, "Fusion of hyperspectral data using segmented pct for color representation and classification," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 43, no. 10, pp. 2365–2375, 2005.

[11] Q. Du, N. Raksuntorn, S. Cai, and R.J. Moorhead, "Color display for hyperspectral imagery," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 46, pp. 1858–1866, 2008.

[12] Y. Zhu, P.K. Varshney, and H. Chen, "Evaluation of ica based fusion of hyperspectral images for color display," in *Information Fusion, 2007 10th International Conference on*, 2007, pp. 1–7.

[13] H. Zhang, D.W. Messinger, and E.D. Montag, "Perceptual display strategies of hyperspectral imagery based on pca and ica," in *Proceedings of SPIE*, 2006.

[14] N.P. Jacobson and M.R. Gupta, "Design goals and solutions for display of hyperspectral images," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 43, no. 11, pp. 2684–2692, 2005.

[15] N.P. Jacobson, M.R. Gupta, and J.B. Cole, "Linear fusion of image sets for display," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3277–3288, 2007.

[16] M. Cui, A. Razdan, J. Hu, and P. Wonka, "Interactive hyperspectral image visualization using convex optimization," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 47, no. 6, pp. 1673, 2009.

[17] S. Le Moan, A. Mansouri, J.Y. Hardeberg, and Y. Voisin, "A class-separability-based method for multi/hyperspectral image color visualization," in *IEEE International Conference on Image Processing*, 2010.

[18] P. Bajcsy and P. Groves, "Methodology for hyperspectral band selection," *Photogrammetric engineering and remote sensing*, vol. 70, pp. 793–802, 2004.

[19] C.I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 44, no. 6, pp. 1575–1585, 2006.

[20] C.I. Chang, Q. Du, T.L. Sun, and M.L.G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 37, no. 6, pp. 2631–2641, 1999.

[21] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 4, pp. 564–568, 2008.

[22] B. Demir, A. Celebi, and S. Erturk, "A low-complexity approach for the color display of hyperspectral remote-sensing images using one-bit-transform-based band selection," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 47, no. 1 Part 1, pp. 97–105, 2009.

[23] C.E. Shannon and W. Weaver, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[24] C. Conese and F. Maselli, "Selection of optimum bands from tm scenes through mutual information analysis," *ISPRS journal of photogrammetry and remote sensing*, vol. 48, no. 3, pp. 2–11, 1993.

[25] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electronics Letters*, vol. 38, pp. 313, 2002.

[26] V. Tsagaris and V. Anastassopoulos, "An information measure for assessing pixel-level fusion methods," *Image and Signal Processing for Remote Sensing*, 2004.

[27] B. Guo, S.R. Gunn, RI Damper, and JDB Nelson, "Band selection for hyperspectral image classification using mutual information," *Geoscience and Remote Sensing Letters, IEEE*, vol. 3, no. 4, pp. 522–526, 2006.

[28] A. Martinez-Uso, F. Pla, J.M. Sotoca, and P. Garcia-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 4158–4171, 2007.

[29] C. Cariou, K. Chehdi, and S. Le Moan, "Bandclust: An unsupervised band reduction method for hyperspectral remote sensing," *Geoscience and Remote Sensing Letters, IEEE*, , no. 99, pp. 564–568, 2010.

[30] S. Watanabe, "Information theoretical analysis of multivariate correlation," *IBM Journal of research and development*, vol. 4, no. 1, pp. 66–82, 1960.

[31] W.R. Garner, *Uncertainty and structure as psychological concepts*, Wiley New York, 1962.
[32] M. Studeny and J. Vejnarova, "The multiinformation function as a tool for measuring stochastic dependence," *Learning in graphical models*, vol. 261, 1998.
[33] W.J. McGill, "Multivariate information transmission," *Psychometrika*, vol. 19, no. 2, pp. 97–116, 1954.
[34] A.J. Bell, "The co-information lattice," in *Proceedings of the Fifth International Workshop on Independent Component Analysis and Blind Signal Separation: ICA 2003*, 2003.
[35] S. Cai, Q. Du, and RJ Moorhead, "Hyperspectral imagery visualization using double layers," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3028–3036, 2007.
[36] R.C. Gonzalez, R.E. Woods, and S.L. Eddins, *Digital image processing using MATLAB*, Prentice Hall Upper Saddle River, NJ, 2004.
[37] [online], "http://www.cie.co.at/main/freepubs.html," last check: May. 11, 2011.
[38] HJ Trussell, "Color and multispectral image representation and display," *Handbook of image and video processing*, p. 411, 2005.
[39] [online], "http://aviris.jpl.nasa.gov/html/aviris.freedata.html," last check: May. 11, 2011.
[40] [online], "http://www.neo.no/hyspex/," last check: May. 11, 2011.
[41] C. Simon, U. Huxhagen, A. Mansouri, A. Heritage, F. Boochs, and F.S. Marzani, "Integration of high-resolution spatial and spectral data acquisition systems to provide complementary datasets for cultural heritage applications," in *Proceedings of SPIE*, 2010, vol. 7531, p. 75310L.
[42] S.M.C. Nascimento, F.P. Ferreira, and D.H. Foster, "Statistics of spatial cone-excitation ratios in natural scenes," *Journal of the Optical Society of America A*, vol. 19, no. 8, pp. 1484–1490, 2002.
[43] D.L. Ruderman, "The statistics of natural images," *Network: computation in neural systems*, vol. 5, no. 4, pp. 517–548, 1994.
[44] N.R. Pal and S.K. Pal, "Entropy: a new definition and its applications," *IEEE Trans. on Systems Man and Cybernetics*, vol. 21, no. 5, pp. 1260–1270, 1991.
[45] S. Somani, B.J. Killian, and M.K. Gilson, "Sampling conformations in high dimensions using low-dimensional distribution functions," *The Journal of chemical physics*, vol. 130, pp. 134102, 2009.



Steven Le Moan received his MSc degree in signal processing and embedded systems from the Université de Rennes 1, France, in 2009. He is currently working towards the PhD degree at the Laboratoire d'Electronique, Informatique et Image (Université de Bourgogne, France) and in collaboration with the Norwegian Color Research Laboratory (Gjøvik University College, Norway). His research interests include multispectral color imaging, information theory and 3D reconstruction.



Alamin Mansouri received his PhD from the university of Bourgogne, France in 2005. Afterwards, he was postdoctoral fellow at the ColorLab laboratory in Norway. Since 2006 he is associate professor at the same university and member of the Laboratory LE2I (Laboratoire d'Electronique, Informatique et Image) where he is leading the project of 3D multispectral scanning. His current research is focused on multispectral imaging (both sides acquisition and processing/analysis) and its interactions with cultural heritage and media technology.



Yvon Voisin received his PhD degree in 1993. He is currently a full professor at the University of Burgundy, France. He is also a member of the Image Processing group of the Le2i - Laboratoire d'Electronique, Informatique et Image. His research interests are 3D reconstruction and metrology by artificial vision. He is also working on the application of artificial vision in medicine.



Jon Y. Hardeberg received his Sivilingeniør (MSc) degree in signal processing from the Norwegian Institute of Technology in Trondheim, Norway in 1995, and his his PhD from Ecole Nationale Supérieure des Télécommunications in Paris, France in 1999. After a short but extremely valuable industry career near Seattle, Washington, where he designed, implemented, and evaluated color imaging system solutions for multifunction peripherals and other imaging devices and systems, he joined Gjøvik University College (GUC) in Gjøvik, Norway, in 2001. He is currently Professor of Color Imaging at GUC's Faculty of Computer Science and Media Technology, and director of the Norwegian Color Research Laboratory. His current research interests include multispectral color imaging, print and image quality, colorimetric device characterization, and color management, and he has co-authored more than 150 publications within the field. His professional memberships include IS&T, SPIE, and ISCC, he is GUC's representative in iarigai, and the Norwegian delegate to CIE Division 8.