Saliency for Spectral Image Analysis

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

Saliency for Spectral Image Analysis

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

***Abstract*—We introduce a new feature extraction model for purposes of image comparison, visualization and interpretation. We define the notion of spectral saliency, as the extent to which a certain group of pixels stands out in an image and in terms of reflectance, rather than in terms of colorimetric attributes as it is the case in traditional saliency studies. The model takes as an input a multi- or hyper-spectral image with any dimensionality, any range of wavelengths, and it uses a series of dedicated feature extractions to output a single saliency map. We also present a local analysis of the image spectrum allowing to produce such maps in color, thus depicting not only which objects are salients, but also in which range of wavelengths. A variety of applications can be derived from the resulting maps, particularly under the scope of visualization, such as the saliency-driven evaluation of dimensionality reduction techniques. Results show that spectral saliency provides valuable information, which do not correlate neither with visual saliency, second-order statistics nor with naturalness, but serve however well for visualization-related applications.**

***Index Terms*—Multi/hyperspectral imagery, Visualization, Saliency**

I. Introduction

When it comes to computer science, the notion of saliency is closely related to visual attention in color images: "from a given scene, which objects/features first draw attention and why ?". Following early influential work by Treisman *et al.* [1], Itti *et al.* [2] proposed a general visual attention model to design so-called saliency maps, in order to predict human gaze, given a certain digital picture. This model is based on the extraction of three different features: lightness, chroma (Red/Green and Blue/Yellow oppositions) and orientation (by means of Gabor filters in four different directions). Each feature is then derived into a Gaussian pyramid of so-called feature maps, where a pixel at a coarse scale thus represents the surrounding of the corresponding location in a finer scale. Through a series of dedicated combinations, blurring and normalizations, the model eventually outputs a single greyscale saliency map, where bright pixels represent the most salient locations. More recently, Harel *et al.* [3] proposed to model the feature spaces as Markov chains to achieve a better normalization of the feature maps through a so-called "activation". In [4], the authors suggested a method using the log-spectrum of the input image in which statistical singularities are assumed to be salient features (or so-called *proto objects*). An information-theoretic approach based on an extraction of independent components of the scene was proposed in [5]. As opposed to these techniques that seek biological plausibility to relate as much as possible to human vision, some methods use high-level features such as face recognition [6]. In [7], the authors proposed a supervised approach trained by a large database of eye-tracking data in a model based on low, mid and high-level image features. More recently, a signature-based approach was introduced by Hou *et al.* [8]. A very interesting review by Frintrop *et al.* [9] addresses thoroughly the cognitive foundations of many saliency detection techniques from the literature.

Yet, we would like to consider saliency as a much broader concept that goes beyond the simple scope of visual attention. Knowing the extent to which a certain object stands out in a certain image, w.r.t. a certain feature (such as pixel lightness), provides indeed with valuable information, which can be used for instance for image matching [10], segmentation [11] or artifact detection [12]. Saliency is thus not tied up to the notion of vision, and it can be seen as a way to measure informative content. Consequently, any kind of data can be considered, although only a few studies went outside the scope of two-dimensional trichromatic images. Among them, and though not directly related to this study, it is worth mentioning the pioneer work by Lee et al. [13] on 3D mesh saliency. As for high-dimensional images such as remotely sensed hyperspectral images, they have received very little attention when it comes to saliency. Particularly, there exists no model to detect salient objects in terms of spectral attributes, although the latter have evidently a variety of advantages over trichromacy: able to detect metameric matches, not bounded to visible wavelengths and device-independent. A strategy of that kind would moreover allow to compare images with different number of channels, a useful property to evaluate dimensionality reduction techniques for example.

In this paper, we introduce such a model for purposes of image comparison, visualization and interpretation. We define the notion of **spectral saliency**, as the extent to which a certain group of pixels stands out in an image, in terms of reflectance rather than in terms of colorimetric attributes. The model takes as an input a multi- or hyper-spectral image with any dimensionality, any range of wavelengths and it uses a series of dedicated feature extractions to output a single map, which can be further used for various applications. Note that we also present a local analysis of the image spectrum allowing to produce saliency maps in color, thus depicting not only which objects are salients, but also in which range of wavelengths. This is particularly useful for a fast and efficient scene interpretation.

The remainder of this document is organized as follows: first, we present the details of the architecture of the spectral saliency (**SS**) model before discussing a variety of results in Section III and eventually drawing our conclusions.

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)

II. Spectral saliency

A. Overall architecture

The **SS** model, whose general architecture is depicted in Figure 1, is based on the comparison of each pixel with its surroundings, in terms of different features and at several scales. The next sections give details on each steps of the model.

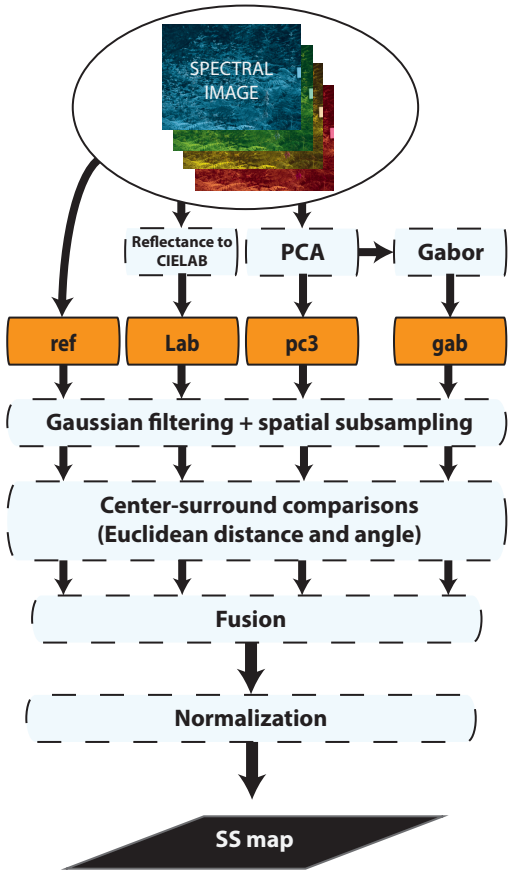

Figure 1. The **SS** model.

1) Measures, features and center-surround comparisons: The key idea of saliency detection is the center-surround comparison, that is, a pixel or group of pixels is salient if it is somehow different from its neighborhood [14]. Thus, in order to achieve this detection, one needs to define these "centers" and "surrounds" in the spatial dimensions of the image, but also an efficient means to compare them. The model that we propose is inspired by the well-known Itti model [2], in which centers and surrounds are defined within an 8-level Gaussian pyramid. Let us consider for instance a three-dimensional representation of the image such as its first three principal components, that we will denote $\mathrm{PC}_{1,2,3}$. Then let us derive a Gaussian pyramid representation of $\mathrm{PC}_{1,2,3}$ by iteratively blurring and sub-sampling it so that each level of the pyramid depicts $\mathrm{PC}_{1,2,3}$ at a different spatial scale, ranging from number 1 (full resolution), to number 8 (lowest resolution). As suggested in [2], centers are then considered at scales $c \in \{2,3,4\}$ and surrounds at scales $s = c+\delta$, with $\delta \in \{3,4\}$. Thus, we obtain the straightforward correspondence of a pixel at a fine scale (center) and the corresponding location at a coarser scale (surround), as described by Figure 2.

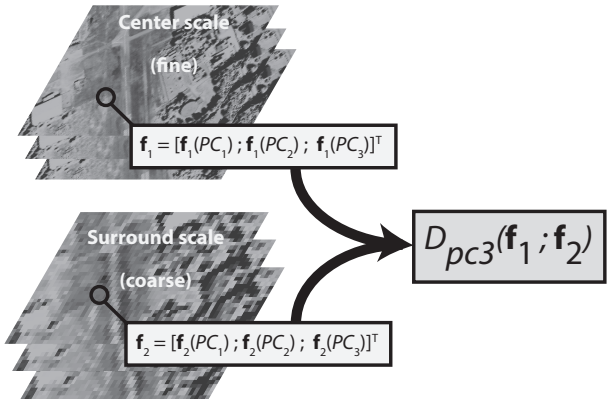

Figure 2. Description of the center-surround comparison for $D_{\mathrm{PC}_{1,2,3}}$, the Euclidean distance in the space spanned by the first three principal components of the image.

Now that we have defined these centers and surrounds, we need to find an efficient way to compare them. There exist many different approaches to compare high-dimensional pixels [15] that we can classify in two categories: the *perception-oriented* measures such as the CIE color difference metrics and the *spectral* measures such as the Euclidean distance between high-dimensional vectors. Measures in the first category usually to define particular viewing conditions (illuminant, standard observer) whereas spectral measures on the other hand, allow to compare full reflectance spectra in an unconstrained fashion. In this paper, we consider both perceptual and purely physical properties of the scene to detect salient objects. Consequently, we propose to use a set of 4 common spectra comparison measures for the **SS** model. In the following equations, the pair of pixels to compare are noted $\mathbf{f}_1$ and $\mathbf{f}_2$.

- D_{ref}, the **Euclidean distance** (L2-norm) between pixels, defined as high-dimensional vectors.

$$D_{ref}(\mathbf{f}_1;\mathbf{f}_2) = \sqrt{\sum_n \left(\mathbf{f}_1(\lambda_n) - \mathbf{f}_2(\lambda_n)\right)^2} \qquad (1)$$

  where $\mathbf{f}_1(\lambda_n)$ is the value of $\mathbf{f}_1$ in the band corresponding to wavelength λ_n (*idem* for $\mathbf{f}_2$). Although, using the full high-dimensional image for the multiscale analysis is very demanding computationally-wise, therefore we propose to use the same measure, but within the space spanned by the image's first three principal components.

$$D_{\mathrm{PC}_{1,2,3}}(\mathbf{f}_1;\mathbf{f}_2) = \sqrt{\sum_{n=1}^{3} \left(\mathbf{f}_1(PC_n) - \mathbf{f}_2(PC_n)\right)^2} \qquad (2)$$

  Based on results obtained over a population of 800 pixels randomly chosen from the 8-image Foster's 2002 database [16], we noted a correlation of 0.99 between D_{ref} and $D_{\mathrm{PC}_{1,2,3}}$, which leads us to choose the latter measure instead of the former.
- ΔE_{ab}, the Euclidean distance between pixels as **CIELAB trichromatic values**. The conversion from reflectance to

tristimuli values was made w.r.t. the CIED65 illuminant (daylight) as well as the CIE 10° standard observer data, or so-called Color Matching Functions [17]. These are indeed the standard assumptions when it comes to viewing conditions in the litterature. Tri-stimuli in XYZ were then converted to the perceptually pseudo-uniform CIELAB color space [18]. This feature includes both lightness and chromatic information.

$$\Delta E_{ab}(\mathbf{f}_1;\mathbf{f}_2) = \sqrt{(L_1^* - L_2^*)^2 + (a_1^* - a_2^*)^2 + (b_1^* - b_2^*)^2} \quad (3)$$

where (L_1^*, a_1^*, b_1^*) and (L_2^*, a_2^*, b_2^*) are respectively the coordinates of $\mathbf{f}_1$ and $\mathbf{f}_2$ in CIELAB.

- θ_{ref}, the **angle** between pixels.

$$\theta_{ref}(\mathbf{f}_1;\mathbf{f}_2) = cos^{-1}\left(\frac{\mathbf{f}_1 \cdot \mathbf{f}_2}{\|\mathbf{f}_1\| \, \|\mathbf{f}_2\|}\right) \quad (4)$$

where $\cdot$ is the dot product between two vectors. Note that this feature is independent of pixel lightness.

- D_{gab}, the difference between pixels in the **Gabor-filtered image of the first principal component** of the spectral image. Linear filter for edge detection according to a specified direction. Allows to compare pixels in terms of local orientations (at 0, 45, 90 and 135°, see [2]). Note that we make the assumption that noisy channels have been previously removed in order to minimize the noise in the first principal component.

These measures use features of various dimensionalities (reflectance curves, principal components and Gabor-filtered pixels) and which are in turn compared using the Euclidean distance or angle. Note that the dimensionality of each scale of the pyramids is the dimensionality of the corresponding feature. For example, for $D_{\mathrm{PC}_{1,2,3}}$, centers and surrounds are both 3-dimensional. The Euclidean distance and angle are then used to evaluate the discrepancies between these centers and surrounds, pixel by pixel, which results in intermediate maps (so-called *conspicuity maps* in the Itti model). Note that other measures like correlation, higher order statistics or information measures such as mutual information [19] are too demanding computationally-wise to be considered in this model.

2) Normalization: The intermediate maps are eventually fused (averaged, so as to treat the model's features equally) into a single saliency map, containing raw saliency information and which needs a specific normalization. The role of normalization is to concentrate lightness into a fewer key locations in order to increase the maps' interpretability [3]. We propose to do so by raising the maps to a power of τ, after scaling to the range $[0..1]$. In this study, we set this parameter to $\tau = 1.2$, in order to obtain the most representative results on the images used. Depending on the amount of information to be retained from the scene, we recommend to decrease this value in order to increase the number of detected salient pixels. Figure 3 shows an example of resulting map, before and after activation.

The final map is in turn normalized, but not blurred neither centrally biased, as it is traditionally done in visual saliency models. Indeed, the purpose of these processings pertains to visual attention modeling, which is not taken into account here.

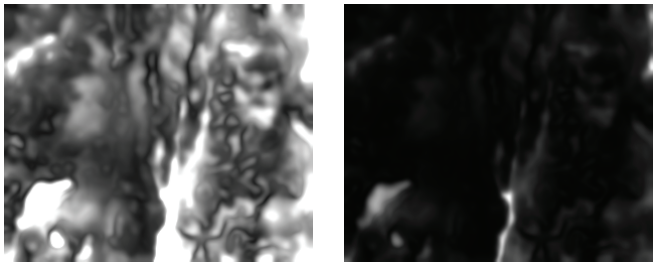

Figure 3. Example of normalization of an intermediate saliency map obtained from the Jasper Ridge scene ($\tau = 1.3$). Left: raw ; right: normalized. The brightness of a pixel represents its saliency (white pixel: very salient).

This final map will then be referred to as $\mathcal{M}_{\mathrm{SS}}$. Figure 4 shows an example of intermediate and final maps.

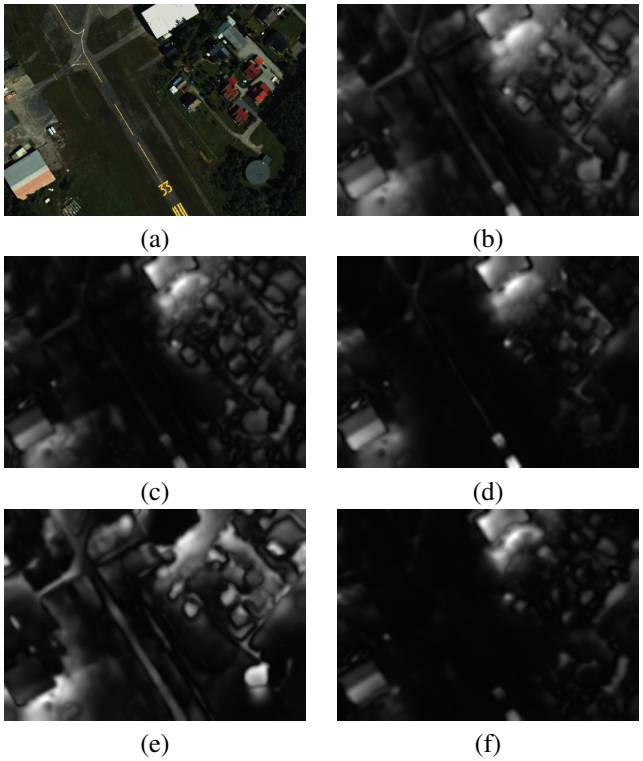

Figure 4. Example of final and intermediate saliency maps from the **SS** model. (a) *true color* composite, (b) final saliency map, (c) normalized $D_{\mathrm{PC}_{1,2,3}}$ map, (d) normalized ΔE_{ab} map, (e) normalized θ_{ref} map, (f) normalized D_{gab} map. We observe that all four intermediate maps emphasize slightly different regions, and especially the one corresponding to the angle feature (e), which is the only one detecting the middle road as well as the tower at the bottom right side of the scene (circle shape). The ΔE_{ab} map (d) detects especially the yellow painted sign on the road (bottom, middle), whereas all four features agree on the fact that the most salient area is the group of trees at the top of the image, next to the garage roof, which is very reflective all over the spectrum.

The output of the **SS** model thus represents those pixels whose reflectances stand out, in terms of the presented features.

3) Low-dimensional images: This model can be applied regardless of the dimensionality of the image. However for low-dimensional images (less than 4 channels), it is relevant to consider some modifications. For trichromatic images for instances, there is very little advantage in using $D_{\mathrm{PC}_{1,2,3}}$ instead of D_{ref}. Greyscale images require only D_{ref} (which

becomes a simple difference) and D_{gab} for 1-dimensional images.

B. In color

In order to further explore this concept of spectral saliency, we propose to enhance the **SS** model so as to create saliency maps in color. We believe indeed that it is possible to display not only which pixels stand out, but also where they stand out, that is in which particular range of wavelengths. We propose to pre-define up to three ranges of interest by segmenting the image's spectrum in as many parts. The **SS** model is then applied locally, that is considering only the channels in each part. Three maps are thus obtained, which contain saliency information corresponding to each range of interest. Eventually, these maps are gathered by means of a convenient mapping to the Red-Green-Blue color space. Consequently, the regions which are salient for example in the second range of interest will appear green on the final map, therefore allowing for an easy and straightforward interpretation. We will refer to the resulting maps as $\mathcal{M}^c_{\text{SS}}$. Figure 5 illustrates this enhanced model, which will be referred to as SS-Color (**SS-C**), and Figure 6 gives an example of result, considering three equally-sized ranges of interest.

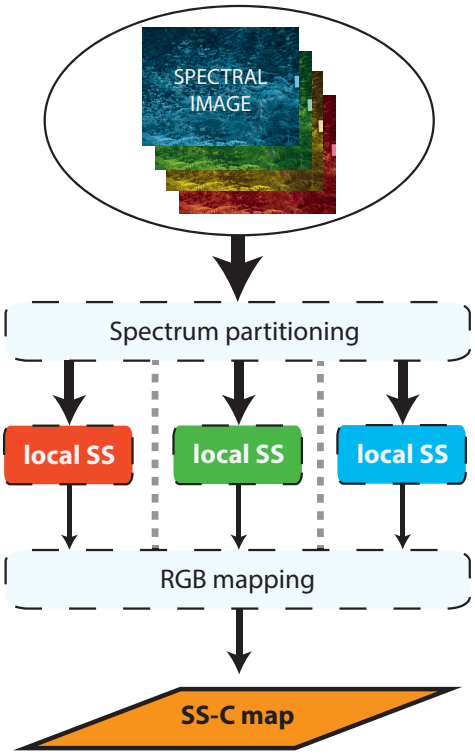

Figure 5. The **SS-C** model.

III. Experiments

A. Data

In this study, we used 2 hyperspectral images of remotely sensed natural scenes:

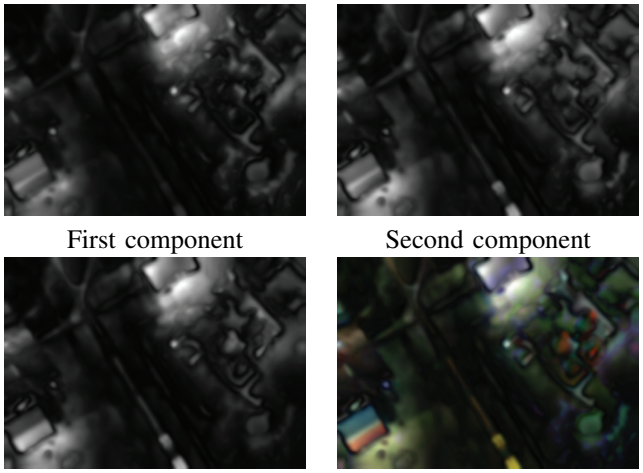

Figure 6. Example of enhanced saliency map obtained from a 160-band hyperspectral image, ranging from 400 to 1000nm. A uniform spectral partitioning was used, that is roughly: 400-600nm (first component, mapped to the Blue channel), 601-800nm (second component, mapped to the Green channel), 801-1000nm (third component, mapped to the Red channel). The variety of colors shows the sensitivity of the **SS** model to the range of wavelengths.

- The top part of the "Jasper Ridge" scene from the AVIRIS sensor[1]. It contains 224 channels ranging from 400 to 2500nm (sampling rate: 9.3nm).
- The "Norway" scene from the HySpex sensor[2]. It contains 160 channels ranging from 400 to 1000nm (sampling rate: 3.7nm).

Both images contain reflectance data, stored as double precision values (64 bits), and pre-processed as follows. For white point detection and coarse noise removal, all pixels with value greater than $m_R+4\sigma_R$ (where m_R and σ_R are respectively the average and standard deviation) of the whole raw image, were clipped to this boundary. Each image was then scaled so as to have a maximal value of 1, corresponding to a non-specular white. Moreover, bands with average reflectance values below 2% and those with low correlation (below 0.8) with their neighboring bands were removed, as suggested in [20].

B. Maps

In this section, we present the resulting maps obtained on the spectral images and we discuss the similarity of the **SS** with the Signature-based Visual Saliency model (**SVS**) presented in [8], which was applied on the *true color* composite (projection of the visible-ranged channels to a set of Color Matching Functions, as in [21]). Figures 7 and 9 show the results obtained on the two remotely sensed hyperspectral images. Note that, in this study, we considered only two partitioning approaches for the **HDS-C** model: uniform (three three equally-sized parts) and visible/near InfraRed (nIR). In the latter case, only two components of the RGB space were used: Red (for the nIR range) and Blue (for the visible range).

Overall, the **SS** and **SVS** models extract very different objects, mostly due to the fact that the latter is based on

[1] http://aviris.jpl.nasa.gov/html/aviris.freedata.html
[2] http://www.neo.no/hyspex/

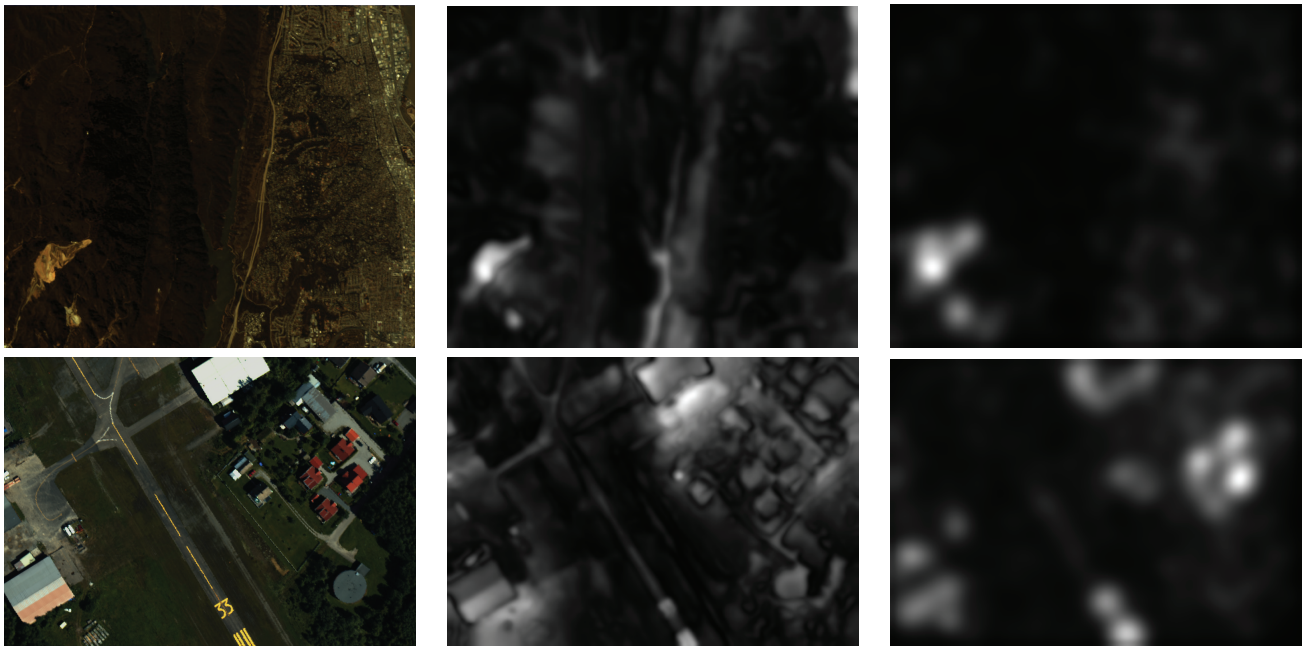

Figure 7. Results obtained on 2 images. Column-wise: *true color* composite, $\mathcal{M}_{SS}$, **SVS** map [8]. Row-wise: "Jasper Ridge" and "Norway" scenes.

visual attention and perceptual attributes only, which is not the case of our method. Note also that the **SVS** method uses a final blur, which induces a more "diffuse" and less precise rendering. On the "Jasper Ridge" scene, the **SVS** model fails to detect the river in the middle of the scene, due to its lack of conspicuity in the *true color* composite. It is however retained by our model, and particularly salient in the nIR (red in the visible/nIR partitioning-based **HDS-C** map). Figure 8 shows samples of pixels from the river-area and its surroundings, which suggest indeed that this river can only be detected in non-visible wavelengths.

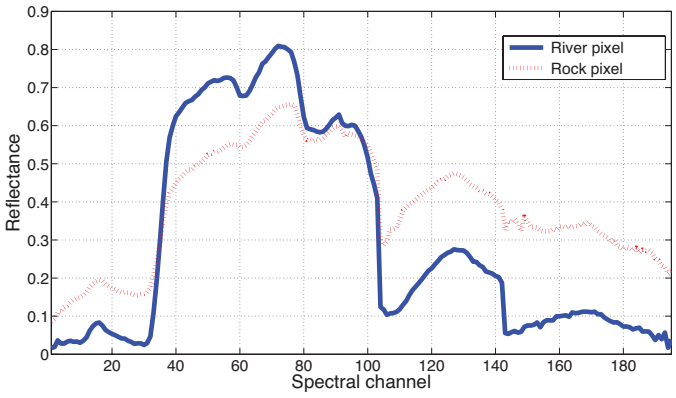

Figure 8. Sample pixels from the "Jasper Ridge" scene: river area (continuous line) and its surroundings (dashed). Clearly, perceptual attributes are not relevant to segregate these reflectance curves as they are both very little reflective in the visible range of wavelengths (channels 1 to 33). Among the features of the **SS** model, it is likely that the angle measures the lowest discrepancy as these curves look indeed very correlated. On the other hand, the Euclidean distance and difference of standard deviations are most definitely the discriminant attributes here.

The area containing a bright sand-like mineral at the bottom left side of the scene is apparently its prominent feature, as all models detect it. Furthermore, the $\mathcal{M}^c_{SS}$ maps suggest that it is uniformly salient all over the spectrum, because it is represented in white (uniform partitioning) and purple (visible/nIR partitioning). Also the urban area at the right side of the scene is detected as salient in the nIR.

On the "Norway" scene, the **SS** models globally detect either portions of grass surrounded with buildings and/or asphalt, or the other way around, because of the contrast created in the nIR range, where lifeforms are particularly reflective, unlike cold materials such as concrete or asphalt. Aside from that, the prominent features appear to be the two farm roofs and their close surroundings, although the **SVS** model detects principally the red house roofs and the yellow painting on the road.

The color maps thus allow to thoroughly interpret the scene in terms of saliency and throughout the spectral dimension, so as to understand which locations stand out and in which ranges of wavelengths.

C. Application to composite evaluation

Despite the wide variety of existing dimensionality reduction methods, their automatic quality assessment for visualization remains a very challenging and application-dependent task. Jacobson *et al.* [21] introduced 9 general criteria (so-called *design goals*) to do so: *consistent rendering*, *edge preservation*, *computational ease*, *color symbolism*, *equal energy white point*, *wavelength shift invariance* (all wavelengths are given an equal weight), *smallest effective difference* (visual distinctions are no larger than needed to effectively show relative differences), *appropriate pre-attentive features* and *natural palette*. Aside from the computational ease (which was used for instance in [22], [23]), these design goals can roughly be categorized according to three general criteria: the *conveyed information* from the raw, unreduced image,

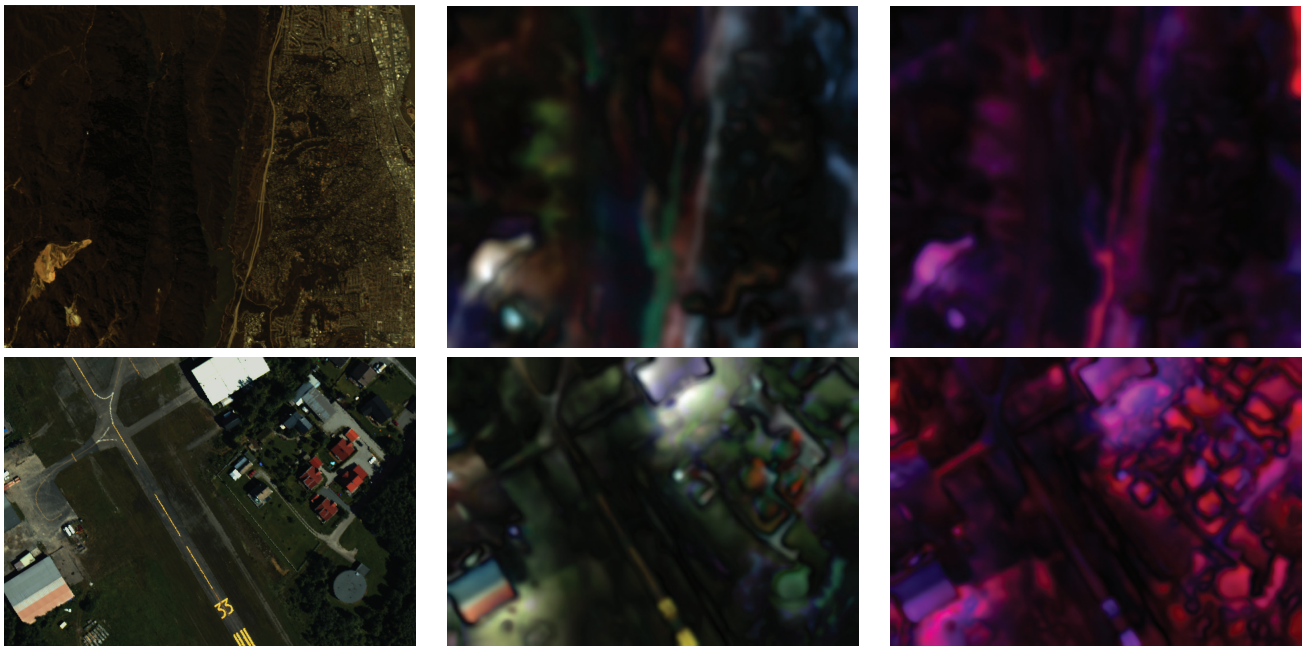

Figure 9. Example of enhanced saliency maps with two different partitioning: uniform (second column) and visible/nIR (third column). They show very valuable information such as the river in the first image (surrounded by minerals) and the urban area in the second one (surrounded by vegetation) that stand out mostly in the infrared.

the *intrinsic information* contained by the composite, and finally the *appeal*, which is related to naturalness, intuitiveness and ease of interpretation. Whereas the two first categories are easily derived into objective and quantitative metrics and measures, the appeal of an image pertains rather to a subjective evaluation.

Since the **SS** model can be applied to any dimensionality, it is possible to compare two images with different number of channels. Consequently, it can be used to assess how accurately a dimensionality reduction technique conveys information into a trichromatic composite. Because it uses both perceptual and physical features, spectral saliency can be used in regard to both the *conveyed information* and *appeal* criteria.

In order to compare the maps produced by the spectral image and its corresponding composite, we propose to use Shannon's mutual information: $I(X;Y)$, which measures the information shared by a couple of random variables X and Y, with regard to the notion of entropy. Because it is usually expressed in units such as *bits* or *nats*, which are not always a intuitive to evaluate redundancy, we propose to normalize it so as to represent a percentage of shared information, to obtain what we refer to as the **true mutual information**. We know indeed that $I(X;Y)$ is bounded between zero and the minimal entropy $\min[H(X), H(Y)]$, as the highest possible value for $I(X;Y)$ is when X "contains" (information-wise) Y or vice-versa [24]. Thus, we propose a measure that will be referred to as the Mutual Saliency (*MS*) of a couple of images $\mathbf{B}_1$ and $\mathbf{B}_2$ based on the true mutual information between the maps in high- and low-dimensionality:

$$MS(\mathbf{B}_1;\mathbf{B}_2)|_{\mathcal{M}_1,\mathcal{M}_2} = \frac{I(\mathcal{M}_1;\mathcal{M}_2)}{\min[H(\mathcal{M}_1);H(\mathcal{M}_2)]} \qquad (5)$$

where $\mathcal{M}_1$ and $\mathcal{M}_2$ are two saliency maps of images with the same spatial dimensions but a potentially different number of spectral channels.

Figures 10 and 11 depict 6 different composites each, obtained from the Jasper Ridge and Norway scenes, respectively. Here is a brief overview of the dimensionality reduction techniques used to obtain these composites:

- The *true color* approach is designed so as to imitate the human visual system. It consists of a set of three functions to project reflectance data to a trichromatic color space. Although these functions are bounded to the visible range of wavelengths, they can be stretched [17] so as to fit any range of wavelengths, in order to obtain what we refer to as a pseudo-*true color* composite.
- Principal Component Analysis (PCA) is a pure data-analysis approach based on second order statistics and of which 3 first components can be conveniently mapped to, respectively the L*, a* and b* components of the pseudo-uniform color space CIELAB [25], it can also be derived into a local approach referred to as segmented PCA [26].
- Band selection consists of preserving the bands during the dimensionality reduction as the final composite is a subset of the original set of bands. We used two approaches: one based on orthogonality, the linear prediction-based band selection (LPBS) [27] as well as one based on band-pass filtering for coarse texture analysis, the one-bit-transform-based band selection (1BTBS) [28].

Tables I and II give the results obtained with the proposed measure as well as two others, for comparison: ρ_θ, the correlation of pairwise angles as in [21] and ν, which measures naturalness as the average Euclidean distance over all pixels in CIELAB between the considered composite and the *true color* one, which serves as a ground truth for naturalness, as

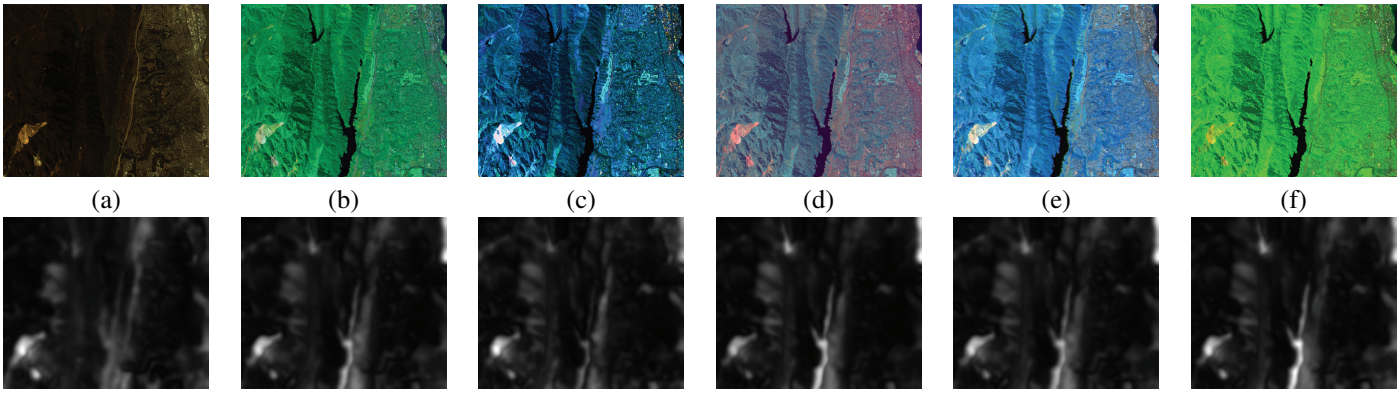

Figure 10. Example of composites from the Jasper Ridge scene. First row: composites (a) *true color*, (b) pseudo-*true color*, (c) PCA to CIELAB, (d) segmented PCA, (e) LPBS and (f) 1BTBS. Second row: corresponding saliency maps. If we focus on the central river (dark shape starting from the bottom middle of the scene), we observe indeed various rendering of this object compared to its surroundings, and therefore different ways to convey saliency through dimensionality reduction.

Table I
SALIENCY-BASED EVALUATION OF DIMENSIONALITY REDUCTION TECHNIQUES: JASPER RIDGE SCENE COMPOSITES.

	true color	pseudo-*true color*	PCA to CIELAB	segmented PCA	LPBS	1BTBS
MS	0.21	**0.54**	0.34	0.41	0.50	0.42
ρ_θ	0.20	**0.93**	0.64	0.74	0.50	0.81
ν	0.00	59.35	58.14	**50.02**	60.71	80.79

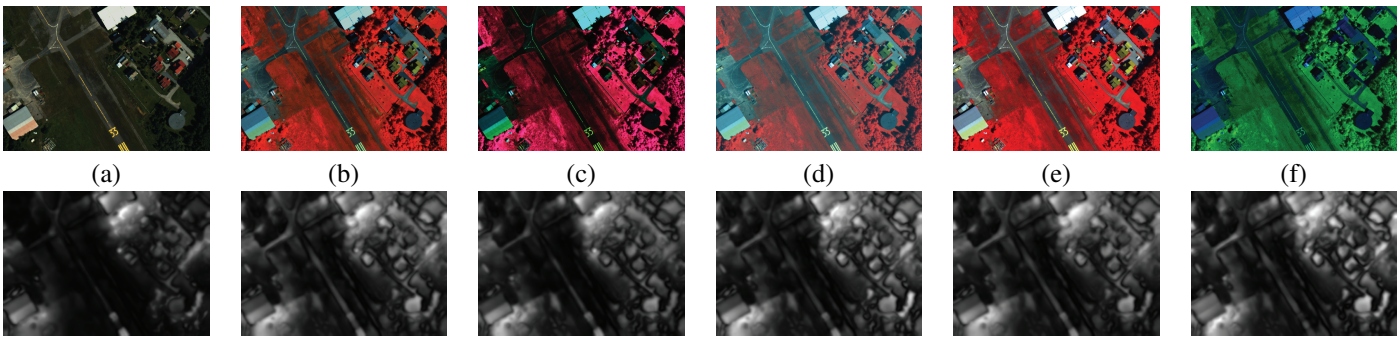

Figure 11. Example of composites from the Norway scene. First row: composites (a) *true color*, (b) pseudo-*true color*, (c) PCA to CIELAB, (d) segmented PCA, (e) LPBS and (f) 1BTBS. Second row: corresponding saliency maps from the **SS** model.

Table II
SALIENCY-BASED EVALUATION OF DIMENSIONALITY REDUCTION TECHNIQUES: NORWAY SCENE COMPOSITES.

	true color	pseudo-*true color*	PCA to CIELAB	segmented PCA	LPBS	1BTBS
MS	0.34	**0.47**	0.33	0.35	0.46	0.35
ρ_θ	0.27	0.83	0.72	0.78	**0.87**	0.68
ν	0	35.19	45.80	**30.65**	42.39	40.53

in [29].

First of all, we observe drastic differences in terms of rendering among the composites. Not only do they differ by their overall naturalness, contrast and variety of colors, they also have different pre-attentive properties. In particular, we note that the *true color* and PCA to CIELAB composites seem to contain slightly less salient objects than the other four, despite the important contrasts yielded by PCA. Indeed, the corresponding maps indicate fewer and more compact salient regions in these composites. In particular, in the "Jasper Ridge" scene, these composites are the only ones failing to properly depict the central river, thus giving them the lowest *MS* scores (34% and 33%). The same rankings are obtained in the "Norway" scene, although in that one, the saliency discrepancies are slightly more complex to evaluate. These two composites also give the worst ρ scores, indicating a bad preservation of pairwise pixel angles. These results can be explained by the fact that the PCA-based approach drastically stretches the data into orthogonal components and may thus emphasizes contrasts and distorts saliency, whereas the *true color* approach discards all the channels in the non-visible range of wavelengths and therefore fails to properly retain the full range of information contained in the scene. This also suggests that perceptual attributes are not sufficient to fully describe spectral saliency as we define it. On the other hand, the best results in terms of *MS*are obtained by the pseudo-*true color* strategy, with 54% for the first scene and 47% for the second one. These composites are neither the most natural (ν

scores of 59.35 and 35.19) nor the most contrasted, but they do preserve the spectral saliency of the high-dimensional image. Furthermore, we measured on these results a correlation as low as 0.65 between *MS* and ρ_θ and 0.60 between *MS* and ν, which implies that our measure conveys a different kind of information, although slightly correlated with both informative content (ρ_θ) and naturalness (ν). Note finally that despite the discrepancies between all the false-color composites, which induce unnatural and misleading renderings, the saliency maps contain consistent information which can serve for interpretation and rapid scene understanding.

IV. Conclusions

We introduced the concept of spectral saliency for multi/hyperspectral images analysis. Unlike traditional research on saliency, this new concept pertains not to the focus of attention, but should rather be seen as a novel approach to feature extraction, taking into account both visual perception and physical properties of the scene within a same framework. In that, it presents all the advantages of spectral imaging over trichromacy: able to detect metameric matches, not bounded to visible wavelengths and device-independent. A "spectrally salient" object stands out from its neighborhood, not only perceptually, but more generally in terms of reflectance, according to the four simple features that were presented. A new model was introduced to compute saliency maps from images of any dimensionality. Particularly, we proposed to partition the image spectrum so as to analyze spectral saliency locally and eventually gather local results within a single highly-information map in color, in order to depict not only which objects stand out, but also in which range of wavelengths. The resulting maps allow for an easy interpretation of the scene, unlike false-color composites that can induce unnatural and misleading renderings. Moreover, we tackled the evaluation of trichromatic composites as a possible application of this model. Results show that spectral saliency provides information which are reliable for interpretation, scene understanding and also image comparison. This information does not correlate neither with visual saliency, second-order statistics nor with naturalness. We are confident that many other applications can be derived from the presented framework, and particularly under the scope of visualization, but also for image registration, segmentation and target detection.

References

[1] A.M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive psychology*, vol. 12, no. 1, pp. 97–136, 1980.

[2] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[3] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Advances in neural information processing systems*, vol. 19, pp. 545, 2007.

[4] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Conference on Computer Vision and Pattern Recognition (CVPR07)*. 2007, IEEE.

[5] N.D.B. Bruce and J.K. Tsotsos, "Saliency, attention, and visual search: An information theoretic approach," *Journal of Vision*, vol. 9, no. 3, 2009.

[6] P. Sharma, F. Alaya Cheikh, and J.Y. Hardeberg, "Saliency map for human gaze prediction in images," in *Sixteenth Color Imaging Conference*, Portland, Oregon, USA, Nov 2008.

[7] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *International Conference on Computer Vision*. 2009, pp. 2106–2113, IEEE.

[8] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011.

[9] S. Frintrop, E. Rome, and H.I. Christensen, "Computational visual attention systems and their cognitive foundations: A survey," *ACM IEEE Transactions on Applied Perception*, vol. 7, no. 1, pp. 6, 2010.

[10] A. Toshev, J. Shi, and K. Daniilidis, "Image matching via saliency region correspondences," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[11] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2009, pp. 1597–1604, IEEE.

[12] G. Cao, M. Pedersen, and Z. Baranczuk, "Saliency models as gamut-mapping artifact detectors," in *5th European Conference on Colour in Graphics, Imaging, and Vision (CGIV)*. 2010, pp. 437–443, IS&T.

[13] C.H. Lee, A. Varshney, and D.W. Jacobs, "Mesh saliency," in *ACM SIGGRAPH 2005 Papers*. 2005, p. 666, ACM.

[14] D. Gao, V. Mahadevan, and N. Vasconcelos, "On the plausibility of the discriminant center-surround hypothesis for visual saliency," *Journal of Vision*, vol. 8, no. 7, 2008.

[15] F.H. Imai, M.R. Rosen, and R.S. Berns, "Comparative study of metrics for spectral match quality," in *Proceedings of the First European Conference on Colour in Graphics, Imaging and Vision*, 2002, pp. 492–496.

[16] 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.

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

[18] P. Green and L. MacDonald, *Colour engineering: achieving device independent colour*, John Wiley and Sons Ltd, 2002.

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

[20] S. Cai, Q. Du, and RJ Moorhead, "Hyperspectral imagery visualization using double layers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3028–3036, 2007.

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

[22] S. Kaewpijit, J. Le Moigne, and T. El-Ghazawi, "Automatic reduction of hyperspectral imagery using wavelet spectral analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 4, pp. 863–871, 2003.

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

[24] T.O. Kvalseth, "Entropy and correlation: some comments," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 17, no. 3, pp. 517–519, 1987.

[25] 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, vol. 6233, p. 31, SPIE.

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

[27] 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.

[28] 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 Transactions on Geoscience and Remote Sensing*, vol. 47, no. 1 Part 1, pp. 97–105, 2009.

[29] S. Le Moan, A. Mansouri, Y. Voisin, and J.Y. Hardeberg, "A constrained band selection method based on information measures for spectral image color visualization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 12, pp. 5104–5115, 2011.