Saliency in Spectral Images

Steven Le Moan[1,2], Alamin Mansouri[1], Jon Hardeberg[2], and Yvon Voisin[1]

[1] Laboratoire Le2i, BP16, 89010 Auxerre Cédex, France
[2] Colorlab, Gjøvik University College, P.O. Box 191, N-2802 Gjøvik, Norway

Abstract. Even though the study of saliency for color images has been thoroughly investigated in the past, very little attention has been given to datasets that cannot be displayed on traditional computer screens such as spectral images. Nevertheless, more than a means to predict human gaze, the study of saliency primarily allows for measuring informative content. Thus, we propose a novel approach for the computation of saliency maps for spectral images. Based on the *Itti* model, it involves the extraction of both spatial and spectral features, suitable for high dimensionality images. As an application, we present a comparison framework to evaluate how dimensionality reduction techniques convey information from the initial image. Results on two datasets prove the efficiency and the relevance of the proposed approach.

1 Introduction

Visual attention modeling is the study of the human visual interpretation of a given scene. In other words, which objects/features will first draw attention and why ? This notion is closely linked to the analysis of saliency. Yet, the latter is a much broader concept in that it can be seen as a way of measuring informative content for any kind of data.

Following early influential work by Treisman *et al.* [1] and Koch & Ullman [2], Itti *et al.* [3] proposed a general visual attention model allowing for the computation of so-called saliency maps, which purpose is to predict human gaze given a certain scene. This model involves center-surround comparisons and combinations of three main feature channels, namely colors, intensity and orientations. More recent work involve for instance the use of graph theory [4], spectral residual [5], information theory [6], or face recognition [7].

Yet, only a few studies have extended the concept of saliency to objects which cannot be entirely displayed on traditional computer screens. Among them, it is worth mentioning the pioneer work by Lee *et al.* [8] on 3D mesh saliency. This paper does not tackle such objects but follows however the same idea of measuring prominent features (in a general way) for non-displayable objects, that is, multi- or hyperspectral images.

Spectral imaging consists of acquiring the same scene at several different ranges of wavelengths, usually several dozens. Since multispectral display devices are yet rare, most of today's popular display hardware is based on the three-stimulus paradigm [9]. Thus, in order to visualize spectral images, a dimensionality reduction step is required so that only three channels (Red, Green

and Blue) can contain most of the visual information. Such methods involve PCA [10,11], band selection [12], Color Matching Functions [13], and are based on a maximization of the informative content of the reduced dataset. However, assessing the quality of a dimensionality reduction technique is very challenging and application-dependent. When it comes to the task of visualization, one usually aims at displaying as much information as possible while easing interpretation by preserving natural colors and contrasts [14].

Even though visual attention has already been used in the context of spectral images for dimensionality reduction purposes [15], the computation of an actual saliency map from the entire high-dimensional image has not been tackled so far to our knowledge. In this paper, we propose a simple method to compute such a map, based on the extraction of both spectral and spatial features. It involves a spectrum segmentation for a local analysis of the reflectance curves. By extending the concept of saliency outside the scope of human visual attention and considering it as a measure of information, the proposed approach can be used for images ranging outside the visible wavelengths (400-700nm), especially since, in many cases, considering for instance the near Infra Red (nIR) allows to enhance the discrimination between materials and objects of a scene. Consequently, we have derived a simple technique to measure the efficiency of dimensionality reduction techniques to convey informative content from the initial spectral image.

In the following, we make a step-by-step description of the saliency map computation by explaining the feature extraction, the spectral center-surround comparisons and the creation of the final map. In a second section, results are shown on two spectral datasets and an evaluation framework involving one PCA-based band transformation approach and two band selection techniques is presented. Eventually, results are discussed and conclusions are drawn.

2   Saliency for Multispectral Images

We propose a simple and efficient method for saliency map computing for spectral images. It is based on the *Itti* model of which we have modified the feature maps computation in order to make it suitable for high dimensional reflectance vectors. Figure 1 gives the synopsis of the proposed approach.

2.1   Channels Computation

Visual attention modeling usually involves the analysis of three main features, namely Color, Intensity and Orientation from which several sets of channels are created (e.g. two for color oppositions, one for intensity and four different orientations). Each set is in fact a gaussian pyramid of an initial channel at several different spatial scales (typically nine). The purpose of these channels is to allow for a straight-forward center-surround comparison by means of a simple across-scale difference. However, if computing color and intensity channels from RGB images is quite easy, it becomes much more challenging when it comes to create such channels from reflectance data. As stated in the introduction, many band
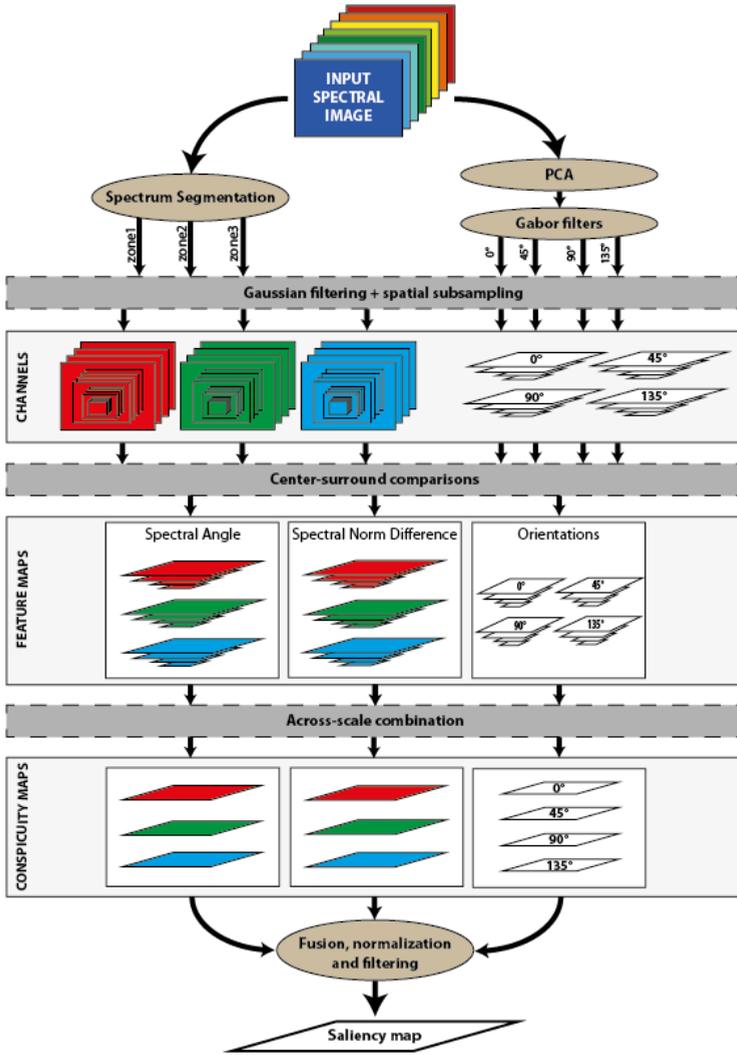
Fig. 1. Synopsis of the proposed approach

transformation and band selection methods exist at this aim, all of them producing different results, and each of them being suitable for a given task such as automatic classification, visualization, compression, noise reduction, etc. In this paper however, we aim at analyzing saliency in the high-dimensional space, therefore no dimensionality reduction must be involved at this point. Nevertheless, we propose to divide the spectrum into three regions, roughly corresponding to the blue, green and red wavelength ranges, in order to allow for a local analysis of the reflectance spectra. For images ranging outside the visible wavelengths (400-700nm), we suggest to add spectrum segments for the Ultra-Violet and near Infra-Red ranges. However, in the remaining of this paper, we will consider

only visible-ranging images. Consequently, three subset of the initial image are obtained and nine spatial scales are created for each of them (from $1:2^0$, the initial scale, to $1:2^8$, the coarser one). As for the creation of orientation channels, we have assumed that differences of orientations being relatively small from one spectral band to another, it would make sense to compute one orientation map for the whole set of spectral bands. Therefore, we have applied Principal Components Analysis (PCA) to the spectral dataset and applied four Gabor filters (at $0°$, $45°$, $90°$ and $135°$) to the first Principal Component (PC), which usually contains more than 95% of the data energy. Eventually, we obtain one set of three spectral cubes, each one of them at nine different spatial scales, as well as four sets of nine orientation channels, also at nine different spatial scales. From this point, the later four sets have been processed exactly the same way as in the *Itti* model.

2.2   Center-Surround Comparisons

While RGB-based saliency analysis involves the extraction of color and intensity channels prior to the center-surround differences computation, we propose to decouple color from intensity during this latter step, by means of relevant comparison metrics:

- Spectral Angle (SA) allows for an intensity-decoupled comparison of reflectance spectra. For a given couple of pixels with respective spectra being noted s_1 and s_2, the SA is given by the following formula:

$$SA(s_1, s_2) = \frac{s_1}{\|s_1\|} \cdot \frac{s_2}{\|s_2\|}$$

- Spectral Norm Difference (SND) depicts the difference, in terms of amount of reflected light, between two spectra. It is given by the following formula:

$$SND(s_1, s_2) = abs(\|s_1\| - \|s_2\|)$$

The substitution of across-scale difference by these metrics allows for the comparison of spectra without dimensionality reduction. Therefore, visual attention is analyzed all over the spectrum. In the common case of spectral images ranging outside the visible wavelengths (400-700nm), one can no longer talk about visual attention, but we believe this study to be relevant nonetheless, since its first aim is to provide a measure of informative content.

Figure 2 gives an example of three reflectance spectra extracted from the dataset presented in the next section. By applying the aforementioned metrics, we obtain the results given in the table in Figure 2. These results confirm that the SND allows for a discrimination between highly and poorly reflecting objects while the SA achieves a comparison in terms of the shape of the curves.

By means of these metrics, we have computed 6 maps by feature, corresponding to the comparison between scales $c \in \{2, 3, 4\}$ and $c + \delta$, $\delta \in \{3, 4\}$. An across-scale fusion at scale four (1:8) allows then for the creation of ten conspicuity maps
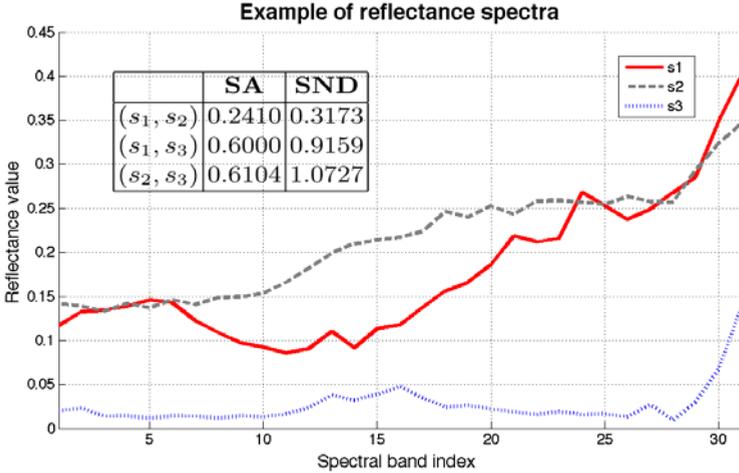
Fig. 2. Example of three different spectra s_1 (red plain), s_2 (grey dashed) and s_3 (blue dotted)

(3 SA, 3 SND and 4 Orientations) which are then respectively averaged in order to obtain one conspicuity map by mode. They are then fused into one single saliency map, here again averaging appears to be the best option considering the heterogeneity of the features. The map is then normalized (division by its global maximum) in order to reduce the number of salient locations and blurred (gaussian filtering) to avoid tiny salient spots and increasing global smoothness.

3   Experiments and Results

3.1   Data Sets

For our experiments, we used two calibrated multispectral datasets of 31 bands, ranging in the visible spectrum (400-700 nm):

- "Flowers" represents a natural scene with flowers, leaves and a background. It comes from a database presented in [16].
- "MacBeth" is the well-known MacBeth CC color calibration target.

Only raw reflectance has been used (no illuminant). Each dataset have been preprocessed so that bands with average reflectance value below 2% and those with low correlation (below 0.8) with their neighboring bands have been removed, as suggested in [17].

3.2   Saliency Map

Figures 3 and 4 represent the final conspicuity maps for both datasets. The depicted maps are actually averages of the three conspicuity maps corresponding
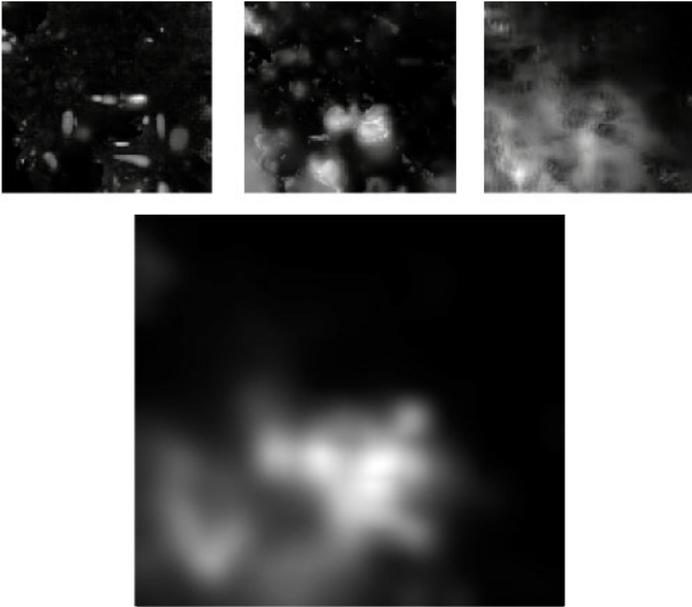
Fig. 3. Conspicuity and final saliency maps for the "Flowers" image. From left to right: SA, SND and Orientations.
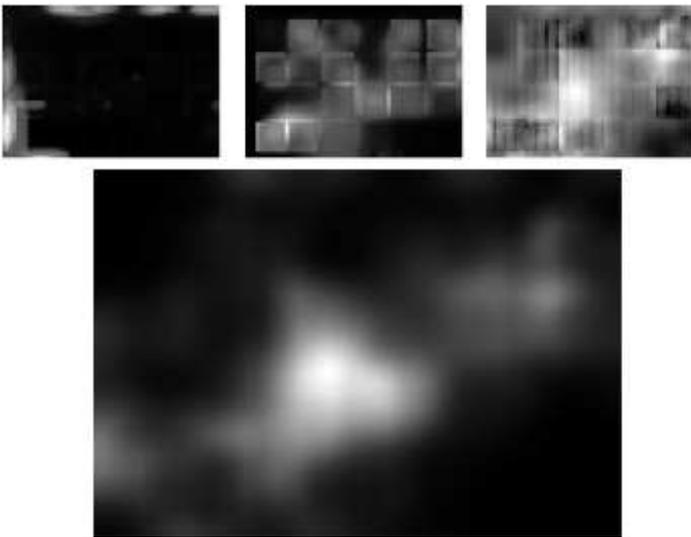
Fig. 4. Conspicuity and final saliency maps for the "Macbeth" image. From left to right: SA, SND and Orientations.

to each segment of the spectrum (for the SA and SND features) and of the four angles maps (for the orientation).

As explained before, these maps depict which locations are the most prominent in terms of SA, SND and orientations, *i.e.* the flowers in the first image and the white patch in the second one. The effect of normalization and filtering yields high contrast and smoothness, which allows for an easier interpretation by highlighting a restrained number of relatively large areas. On the "Flowers" image, SA highlights small areas of background through the leaves while SND mainly emphasizes the flowers. Both metrics find very different salient locations, hence their relative independence and the relevance to use both in the whole process. As for the "MacBeth" target, while the orientation-related saliency is meaningless due to the regular structure of the image, the orientation appears to be the most influential feature in the computation of the final map. This is due to the fact that, in this image, the orientation maps have the highest mean values. Moreover, this also induces that very few locations are salient in terms of colors, barely the white patch.

3.3   Evaluation of Dimensionality Reduction Techniques

We propose to illustrate one application of the spectral saliency map on the evaluation of dimensionality reduction techniques. At this aim, we have computed tri-stimulus representations of both datasets by means of one band transformation and two band selection methods:

- PCA_{rgb} is the traditional Principal Components Analysis of which components are mapped to the RGB color space ($PC1 \rightarrow R; PC2 \rightarrow G; PC3 \rightarrow B$).
- LP-based band selection has been proposed by Du *et al.* [12] and consists of progressively selecting bands by maximizing their respective orthogonality.
- Entropy-based band selection (ENT) is a naive method selecting the three bands with maximal entropy.

For the two latter techniques, resulting channels are mapped to RGB by descending wavelengths. In order to compare the different tri-stimulus composites, we have considered two simple metrics that we will refer to as Saliency Discrepancy number 1 and 2, respectively ($SD1$ and $SD2$). The first one is based on a point-by-point difference, summed along both spatial dimensions and divided by the total number of pixels:

$$SD1(im_1, im_2) = \frac{\sum\limits_{i=1:sizeX} \sum\limits_{j=1:sizeY} smap(im_1(i,j)) - smap(im_2(i,j))}{sizeX * sizeY}$$

with im_1 and im_2 any two images of same size but potentially different number of spectral channels and $smap(.)$ an operator computing the saliency map of its input. The second metric is the inverse of *Shannon*'s mutual information, normalized by the sum of entropies:

$$SD2(im_1, im_2) = \frac{H(smap(im_1)) + smap(im_2))}{MI(smap(im_1)); smap(im_2))}$$
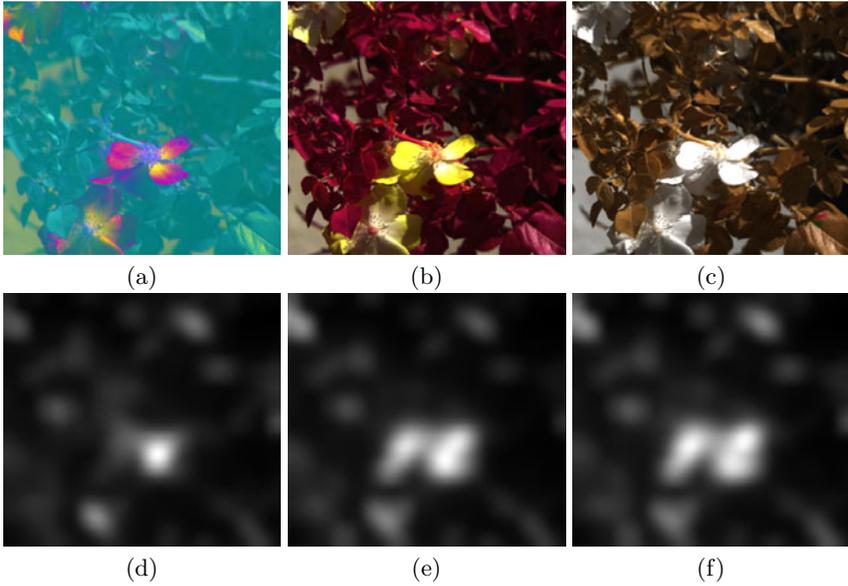
Fig. 5. Results from the dimensionality reduction techniques: (a) PCA (b) LP (c) Entropy and their associated saliency maps (d-f)
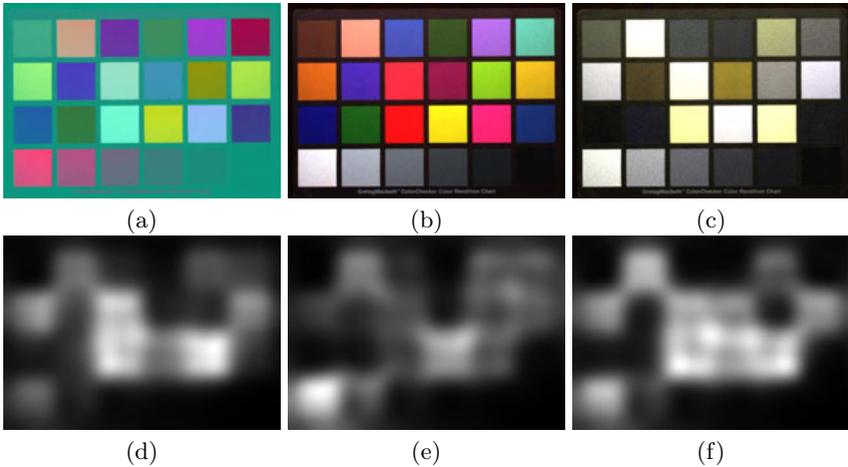
Fig. 6. Results from the dimensionality reduction techniques: (a) PCA (b) LP (c) Entropy and their associated saliency maps (d-f)

with $H(.)$ and $MI(.;.)$ being respectively the entropy and mutual information operators. Figures 5 and 6 depict the resulting tri-stimulus representations by means of the aforementioned techniques as well as their associated saliency maps.

Table 1. Saliency Discrepancies between the spectral images and their respective tri-stimulus composites

		PCA_{rgb}	LP	ENT
SD1	"Flowers"	**13.20**	18.73	19.72
	"MacBeth"	23.95	**16.07**	33.74
SD2	"Flowers"	19.41	**10.78**	12.31
	"MacBeth"	12.01	**10.61**	10.90

Table 1 gives the results of the SD metrics. One must notice that the resulting tri-stimulus composites are very different from each other in terms of color and consequently in terms of saliency. Different regions are highlighted and this reflects the variety of manners to conveys information from the spectral image. Consequently, we observe large fluctuations of SD between the initial datasets and their tri-stimulus composites. As expected, we observe that the SD is independent from the overall visual appeal of the image. Indeed, images with very different hues such as the ones from PCA compared to the results from LP in the case of the "Flowers" image are close nonetheless (13.20 and 18.73). Both metrics are in accordance to elect LP as the most suitable method for the visualization of the "MacBeth" target. However, $SD1$ ranks ENT as the worst method for both images while $SD2$ gives PCA_{rgb} as the method conveying the less saliency. This divergence is due to the fact that Figures 5d and 6d have respectively lower entropies than Figures 5f and 6f, a property that is conveyed by the normalization in $SD2$. For this, we believe that $SD2$ is more efficient and accurate than $SD1$ in measuring discrepancies. *In fine*, PCA is globally outperformed by the two band selection techniques in retaining saliency from the input dataset and the LP is the method giving best results.

4   Conclusion

We have presented a new method to create saliency maps for spectral images. It is based on the extraction of both spectral and spatial features and involves a spectrum segmentation for a better handling of local variations of the reflectance curves. Further than the single scope of visual attention, it is meant to be a way of measuring informative content. An example of application has been given on the comparison of dimensionality reduction techniques for visualization. Further study will investigate the use and influence of different spectrum segmentation techniques as well as means to use saliency as a criterion for dimensionality reduction, in order to control the visual features of a tri-stimulus composite.

Acknowledgements

References

1. Treisman, A., Gelade, G.: A feature-integration theory of attention. Cognitive Psychology 12, 97–136 (1980)
2. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. Hum. Neurobiol. 4, 219–227 (1985)
3. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence 20, 1254–1259 (1998)
4. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. Advances in Neural Information Processing Systems 19, 545 (2007)
5. Hou, X., Zhang, L.: Saliency detection: A spectral residual approach. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2007, pp. 1–8. IEEE, Los Alamitos (2007)
6. Bruce, N., Tsotsos, J.: Saliency, attention, and visual search: An information theoretic approach. Journal of Vision 9 (2009)
7. Cerf, M., Harel, J., Einhäuser, W., Koch, C.: Predicting human gaze using low-level saliency combined with face detection. Advances in Neural Information Processing Systems 20, 241–248 (2008)
8. Lee, C., Varshney, A., Jacobs, D.: Mesh saliency. In: ACM SIGGRAPH 2005 Papers, p. 666. ACM, New York (2005)
9. Grassmann, H.: On the theory of compound colors. Phil. Mag. 7, 254–264 (1854)
10. Jia, X., Richards, J.: Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification. IEEE Trans. on Geoscience and Remote Sensing 37, 538–542 (1999)
11. Tyo, J., Konsolakis, A., Diersen, D., Olsen, R.: Principal-components-based display strategy for spectral imagery. IEEE Trans. on Geoscience and Remote Sensing 41, 708–718 (2003)
12. Du, Q., Yang, H.: Similarity-based unsupervised band selection for hyperspectral image analysis. IEEE Geoscience and Remote Sensing Letters 5, 564–568 (2008)
13. Poldera, G., van der Heijdena, G.: Visualization of spectral images. In: Proc. SPIE., vol. 4553, p. 133 (2001)
14. Jacobson, N., Gupta, M.: Design goals and solutions for display of hyperspectral images. IEEE Trans. on Geoscience and Remote Sensing 43, 2684–2692 (2005)
15. Zhang, H., Peng, H., Fairchild, M., Montag, E.: Hyperspectral image visualization based on a human visual model. In: Proceedings of SPIE, vol. 6806, p. 68060N (2008)
16. Nascimento, S., Ferreira, F., Foster, D.: Statistics of spatial cone-excitation ratios in natural scenes. Journal of the Optical Society of America A 19, 1484–1490 (2002)
17. Cai, S., Du, Q., Moorhead, R.: Hyperspectral imagery visualization using double layers. IEEE Trans. on Geoscience and Remote Sensing 45, 3028–3036 (2007)