

SALIENT OBJECT DETECTION IN HYPERSPECTRAL IMAGERY

Jie Liang¹ Jun Zhou² Xiao Bai³ Yuntao Qian⁴

¹ Research School of Computer Science, Australian National University, Canberra, Australia

² School of Information and Communication Technology, Griffith University, Nathan, Australia

³ Department of Computer Science and Engineering, Beihang university, Beijing, China

⁴ College of Computer Science, Zhejiang University, Hangzhou 310027, China

ABSTRACT

Object detection in hyperspectral images is an important task for many applications. While most traditional methods are pixel-based, many recent efforts have been put on extracting spatial-spectral features. In this paper, we introduce Itti's visual saliency model into the spectral domain for object detection. This enables the extraction of salient spectral features, which is related to the material property and spatial layout of objects, in the scale space. To our knowledge, this is the first attempt to combine hyperspectral data with salient object detection. Three methods have been implemented and compared to show how color component in the traditional saliency model can be replaced by spectral information. We have performed experiments on selected images from three online hyperspectral datasets, and show the effectiveness of the proposed methods.

Index Terms— Saliency detection, object detection, hyperspectral imaging

1. INTRODUCTION

A hyperspectral image consists of tens or hundreds of contiguous narrow spectral bands. Each pixel in a hyperspectral image is a vector of spectral responses across the electromagnetic spectrum (normally in the visible to the near-infrared range). Such spectral responses are related to the material of objects in a scene that has been imaged, which provides valuable information for automatic object detection.

Due to its high dimensionality, traditional pattern recognition and computer vision technology can not be directly applied to hyperspectral imagery. Most object detection methods are still pixel-based, i.e., performing pixel-wise detection and classification based on spectral signatures followed by post-processing to group pixels or to segment regions from an image [1, 2]. In this manner, feature extraction is only performed in the spectral domain, but the spatial distribution

of objects have not been fully explored. More recently, researchers have tried to use spectral-spatial structure modelling for hyperspectral image classification. Such efforts include Markov random field and conditional random field [3, 4], which introduce spatial information into classification steps using probabilistic discriminative function with contextual correlation. Furthermore, multi-scale time-frequency signal analysis methods based on 3D discrete wavelet transform have also been introduced for object detection and classification in remote sensing imagery [5].

Visual saliency is another type of approach to extract multi-scale image features. The concept of saliency is from human attention model, which detects objects or regions in a scene that stands out with respect to their neighborhood [6]. As a consequence, saliency detection models are normally established on the trichromatic or greyscale images, which are visible to human eyes. When used for object detection in computer vision and robotics applications, saliency map is often constructed in a bottom-up manner. For example, Itti et al computed multi-scale differences of intensity, color, and orientation features, and linearly combined them to form the final saliency map [7]. Liu et al formulated the saliency detection problem as a region of interest segmentation task [8]. Salient features were extracted at the local, regional and global levels, and were combined via learning with conditional random field. Similarly, many saliency detection methods try to detect image regions that are different from its neighborhood in the scale space, as reviewed in [6].

When applied to hyperspectral imagery, saliency model has been used for image visualization. Wilson et al employed contrast sensitivity saliency to fuse different bands of hyperspectral remote sensing images so that it can be used for visual analysis [9]. Itti's model [7] has been combined with dimensionality reduction method to covert a hyperspectral image to a trichromatic image that can be displayed on computer screen [10, 11]. Saliency has also been used to help edge detection and to predict eye fixation on hyperspectral images [12, 13].

Despite its success in object detection on RGB images, as far as we know, saliency model has not been used for ob-

This research is partly supported under the Australian Research Council's DECRA Projects funding scheme (project ID DE120102948) and by the National Natural Science Foundation of China projects No. 61171151 and No. 61105002.

ject detection in hyperspectral imagery. Therefore, the contribution of this paper is to explore how salient regions can be extracted from hyperspectral images, and then be used for object detection. Compared with the traditional pixel-level operations, this paper introduce a novel region-based approach for hyperspectral object detection. We propose three methods based on Itti's saliency detection model. The first method converts a hyperspectral image into an RGB image and apply Itti's model directly. The second method replaces the color double-opponent component with grouped band component. The last method directly use the raw spectral signature to replace the color component. We show in the experiments that all three methods have performed well.

2. ITTI'S SALIENCY MODEL

The saliency detection method proposed by Itti et al. mimics the behavior and structure of the early primate visual system [7]. It extracts three types of multiscale features, including intensity, color, and orientation, and then computes their center-surround differences. These differences are linearly combined to form the final saliency map.

This method is processed as follows. An input image is firstly smoothed using low-pass filters so as to generate nine spatial scales. Three types of visual cues are then extracted from the intensity, color, and orientation features. The intensity feature is obtained by averaging the RGB channel-values at each pixel. By computing the differences between each pair of fine and coarse scales, 6 intensity channels are generated. The second set of features are computed from a set of color opponency between red, green and blue values against yellow value at each pixel. Center-surround differences for each pair of color opponent are then computed over three scales, which leads to 12 channels. The orientation features are computed using a set of even-symmetric Gabor filters. The dominant orientation at each pixel is recovered, whose center-surround differences are computed at six scales and four orientations. This leads to 24 orientation channels. Channels of each type are then linear combined to form three conspicuity maps. Finally, the mean of the conspicuity maps becomes the saliency map.

3. SALIENCY EXTRACTION ON HYPERSPECTRAL IMAGES

3.1. Hyperspectral to trichromatic conversion

What has hindered the adoption of saliency extraction by hyperspectral object detection is the large amount of bands in the spectral data. This makes the color component not be able to computed directly. Furthermore, effective computation of the intensity and texture saliency requires a grayscale image to be used. A direct solution to this problem is conversion of a hyperspectral image into a trichromatic image, which al-

lows traditional saliency model to be applicable. As pointed out in [11], this can be achieved by dimensionality reduction, band selection, or color matching functions.

In this research, we have followed the method of Foster [14]. This method first converts the hyperspectral image to a CIE XYZ image. Given a hyperspectral image $I(\lambda_i)$ for each of the bands λ_i , such conversion can be implemented by the following color matching function:

$$I_t = \sum_{i=1}^N I(\lambda_i) W_t(\lambda_i) \quad (1)$$

where N is the total number of bands, $t = \{X, Y, Z\}$ are the tristimulus component of the color space, and W_t comes from the spectral sensitivity curves of three linear light detectors that yield the CIE XYZ tristimulus values X , Y , and Z . This conversion is then followed by a further transform step to the sRGB color space [15], then Itti's method can be applied.



Fig. 1: Spectral band group.

3.2. Spectral band opponment

Although the above method is straightforward, it does not take advantage of the extra information provided by the hyperspectral image. Notice that the second conspicuity map in Itti's method is formed from RGB color channels, we shall be able to replace the color opponents with groups spectral bands that are approximately correspondence to these color channels. To do so, we divide the bands into four groups with each group occupying approximately the same width of visible spectrum, as shown in Figure 1. Then the original single value color component is replaced a vector, and the double opponency can be computed as follows

$$Opp_1(c, s) = |(G1(c) - G3(c)) \ominus (G3(s) - G1(s))|_1 \quad (2)$$

$$Opp_2(c, s) = |(G2(c) - G4(c)) \ominus (G4(s) - G2(s))|_1 \quad (3)$$

where G1 to G4 are vectors whose entries are extracted from the corresponding group of spectral bands, c and s are different scales for the across scale difference computation, and $|\cdot|_1$ is the 1-norm of a vector. \ominus is the cross-scale center-surround difference operator as defined in [7]. Then Opp_1 and Opp_2 can replace the red/green and blue/yellow opponency in [7]. In this method, the intensity and orientation maps can be extracted from the grayscale image converted from the trichromatic image generated using the color matching function as described in section 3.1.

3.3. Spectral Saliency with Euclidean Distance

Further extending the method in section 3.2 allows the using of whole spectral responses for saliency detection. When replacing the color saliency with spectral saliency, the rich information embedded in the spectral data can be fully explored. Following the general multi-scale operation, differences between the spectral responses and its neighborhood can be calculated. Both spectral angle distance (SAD) and Euclidean distance can be used to measure the similarity between two spectral vectors A_k and A_j , where the SAD is computed via

$$\text{SAD}_{kj} = \arccos \left(\frac{A_k^T A_j}{\|A_k\| \|A_j\|} \right) \quad (4)$$

This step leads to a set of center-surround spectral differences in the scale space. They can be combined into a spectral conspicuity map, which is used with the intensity and orientation conspicuity maps to form the final saliency map. The incorporation of spectral data suggests that not only visual clue has been extracted, i.e., from the color and orientation contrast, but also the intrinsic material property of objects. This has provided visual saliency model with additional information beyond the capability of human and traditional camera vision. Furthermore, the SAD and Euclidean measures provides two different spectral distance information that is useful for object detection.

4. OBJECT DETECTION

In the previous step, saliency map can be generated to highlight image regions that are different from their surrounding areas. To detect a region with a salient object, we binarize the saliency map using the optimal threshold recovery method [16]. This allows the pixels with low saliency values be removed. Then a set of morphological operations are used to fill the small holes in the connected components. Those small components that comes from noisy clutters are removed.

The object detection follows a winner-take-all strategy, i.e., assuming that there is only one salient object per image. The remaining image regions that contains the highest value in the saliency map is selected as the one that contains the target object. It should be noted that this method can be easily extended to detect more than one objects by sequentially selecting regions in order of their highest saliency values.

5. EXPERIMENTS

To compare the three salient object detection methods introduced in Section 3, we have performed experiments on ground-based hyperspectral images in three online datasets. The first two datasets were collected by Foster et al [17, 14]. They contain in total 55 hyperspectral images of natural

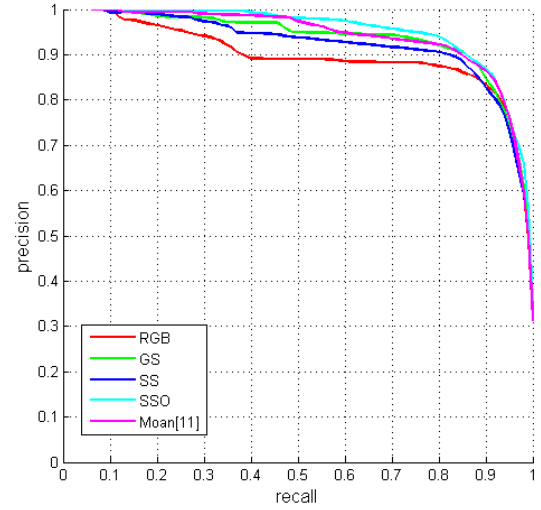


Fig. 2: Precision recall curve

scenes, with 16 images available online for free access. The second dataset consists of 50 hyperspectral images collected at Harvard University [18], which includes images captured in both indoor and outdoor setting.

It should be noted that these three datasets were not collected specifically for saliency object detection purpose. Therefore, in most of them, it is hard to find salient objects, or the scenes are cluttered with many objects. We have carefully selected image regions that contain salient objects in their surroundings from 13 images in these datasets for our experiments. To provide the ground truth, we have manually labeled the location of salient objects by bounding boxes. Because spectral saliency is not directly observable to human eyes, we have combined visual saliency on synthesized RGB images and domain knowledge on the object materials for judgement.

We have implemented each of the three method as proposed in section 3. The first method based on converted RGB images is named as *RGB*. The second method based on grouped spectral band opponent is named *GS*. We have implemented two versions of the third method, i.e., using only Euclidean distance of spectral response (*SS*), or use combined SAD and Euclidean distance (*SSO*). These methods are compared against the method from Moan et al [11]. This method firstly combined spectral channels into red, green, and blue groups, then computes the spectral differences on each of the color groups to get the spectral saliency. The orientation saliency is extracted on the first principle component generated by principle component analysis to reduce the dimensionality of the hyperspectral image, while the intensity saliency is not used.

To provide quantitative analysis to the saliency object detection methods, we have calculated the precision and recall

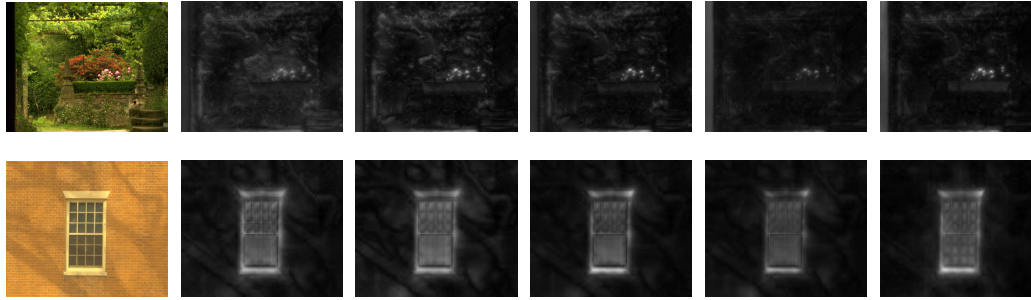


Fig. 3: Saliency map computed from different methods. From left to right: input image, hyperspectral to trichromatic conversion (RGB), spectral band opponent (GS), spectral saliency via Euclidean distance (SS), spectral saliency via combined SAD and Euclidean distances (SSO), and the method in [11].

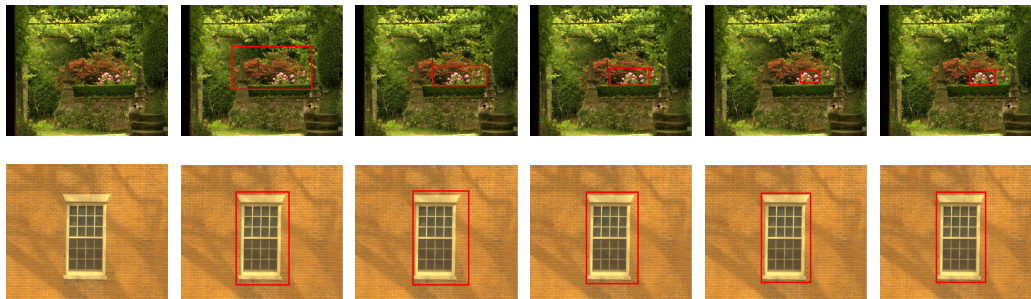


Fig. 4: Object detection results. From left to right: input image, hyperspectral to trichromatic conversion, spectral band opponent, spectral saliency via Euclidean distance, spectral saliency via combined SAD and Euclidean distances, and the method in [11].

curve when different binarization thresholds are used. The results are shown in Figure 2. The precision is computed as the percentage of true object pixels out of all detected pixels. The recall calculates the percentage of true object pixels that have been detected. It can be seen that the spectral saliency methods are clearly better than the method based on RGB images. The performance of the spectral-based solutions are very close to each other, with the *SSO* option slightly outperforming the other options. This shows the advantage of combining different spectral distance measures for saliency detection.

Figure 3 shows the saliency map on two sample images generated by each method under comparison. The object detection results are shown in Figure 4. It can be observed that each method can generate very good saliency feature. When it comes to object detection, the spectral-based methods can detect the truth object more accurately. The method based on RGB image, however, has included large amount of background regions into the results.

6. CONCLUSION

We have extended Itti’s visual attention model to generate saliency map from hyperspectral imagery for object detection. Such extension is mainly based on replacing the color com-

ponent with spectral saliency, which can be implemented by dividing visual spectrum into groups, or use the whole spectral responses. These methods allow extra information from spectral data to contribute to the traditional visual attention model. Experiments have shown the effectiveness of the proposed methods in salient object detection. In the future, we will apply the method to more hyperspectral data, for example, on remote sensing images. We will also incorporate other saliency models into the hyperspectral object detection tasks.

7. REFERENCES

- [1] Z. Fu, A. Robles-Kelly, and J. Zhou, “MILIS: Multiple instance learning with instance selection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 958–977, 2011.
- [2] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, “Multiple spectral-spatial classification approach for hyperspectral data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4122–4132, Nov. 2010.
- [3] P. Zhong and R. Wang, “Learning conditional random fields for classification of hyperspectral images,” *IEEE*

Transactions on Image Processing, vol. 19, no. 7, pp. 1890–1907, 2010.

- [4] J. Li, J. Bioucas-Dias, and A. Plaza, “Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 3, pp. 809–823, 2012.
- [5] Yuntao Qian, Minchao Ye, , and Jun Zhou, “Hyperspectral image classification based on structured sparse logistic regression and 3d wavelet texture features,” *IEEE Transactions on Geoscience and Remote Sensing*, 2013.
- [6] A. Borji and L. Itti, “State-of-the-art in visual attention modeling,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185–207, 2013.
- [7] 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.
- [8] Tie Liu, Jian Sun, Nanning Zheng, Xiaou Tang, and Heung-Yeung Shum, “Learning to detect a salient object,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 353–367, 2007.
- [9] Terry A. Wilson, Steven K. Rogers, and Matthew Kabrisky, “Perceptual-based image fusion for hyperspectral data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 4, pp. 1007–1017, 1997.
- [10] Hongqin Zhang, Honghong Peng, Mark D. Fairchild, and Ethan D. Montag, “Hyperspectral image visualization based on a human visual model,” in *Proceedings of SPIE*, 2008, vol. 6806.
- [11] Steven Le Moan, Alamin Mansouri, Jon Hardeberg, and Yvon Voisin, “Saliency in spectral images,” in *Proceedings of the 17th Scandinavian conference on Image analysis*, 2011, pp. 114–123.
- [12] Cuong V. Dinh, Raimund Leitner, Pavel Paclik, Marco Loog, and Robert Duin, “SEDMI: Saliency based edge detection in multispectral images,” *Image and Vision Computing*, vol. 29, no. 8, pp. 546–556, 2011.
- [13] Anton Garcia-Diaz, Victor Leboran, Xose R. Fdez-Vidal, and Xose M. Pardo, “On the relationship between optical variability, visual saliency, and eye fixation: A computational approach,” *Journal of Vision*, vol. 12, no. 6, pp. 1–22, 2002.
- [14] D. Foster, S. Nascimento, and K. Amano, “Information limits on neural identification of colored surfaces in natural scenes,” *Visual Neuroscience*, vol. 21, pp. 331–336, 2004.
- [15] Matthew Anderson, Ricardo Motta, Srinivasan Chandrasekar, and Michael Stokes, “Proposal for a standard default color space for the internet: srgb,” in *Proceedings of Fourth Color Imaging Conference: Color Science, Systems, and Applications*.
- [16] N. Otsu, “A thresholding selection method from gray-level histograms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, pp. 62–66, 1979.
- [17] Sérgio M. C. Nascimento, Flávio P. Ferreira, and David 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.
- [18] A. Chakrabarti and T. Zickler, “Statistics of real-world hyperspectral images,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 193–200.