Highlights

- To the best of our knowledge, this is the first time that salient objects are detected based on extracting explicit material property embedded in the spectral responses via retrieval of endmembers and estimating their abundance.
- The novelty also comes from adopting hyperspectral unmixing model as a preprocessing step for salient object detection. This allows the spatial distribution of endmembers be estimated, so the method is capable of dealing with mixed spectral responses in low spatial resolution hyperspectral images.
- Different from existing hyperspectral salient object detection methods, we developed a novel method to fuse both local and global features for hyperspectral salient object detection.
- We built a hyperspectral image dataset for salient detection, which contains mixed objects with similar color but different materials.

Material Based Salient Object Detection from Hyperspectral Images

Jie Liang^a, Jun Zhou^{b,*}, Lei Tong^c, Xiao Bai^d, Bin Wang^a

^a*Facility Design and Instrumentation Institute, China Aerodynamics Research and Development Center, Mianyang, Sichuan, China*

^b*School of Information and Communication Technology, Griffith University, Nathan, Australia*

^c*Faculty of Information Technology, Beijing University of Technology, Beijing, China*

^d*School of Computer Science and Engineer, Beihang University, Beijing, China*

Abstract

While salient object detection has been studied intensively by the computer vision and pattern recognition community, there are still great challenges in practical applications, especially when perceived objects have similar appearance such as intensity, color, and orientation, but different materials. Traditional methods do not provide good solution to this problem since they were mostly developed on color images and do not have the full capability in discriminating materials. More advanced technology and methodology are in demand to gain access to further information beyond human vision. In this paper, we extend the concept of salient object detection to material level based on hyperspectral imaging and present a material-based salient object detection method which can effectively distinguish objects with similar perceived color but different spectral responses. The proposed method first estimates the spatial distribution of different materials or endmembers using a hyperspectral unmixing approach. This step enables the calculation of a conspicuity map based on the global spatial variance of spectral responses. Then the multi-scale center-surround difference of local spectral features is calculated via spectral distance measures to generate local spectral conspicuity maps. These two types of conspicuity maps are fused for the final salient object detection. A new dataset of 45 hyperspectral images is introduced

*Corresponding author

Email address: jun.zhou@griffith.edu.au (Jun Zhou)

for experimental validation. The results show that our method outperforms several existing hyperspectral salient object detection approaches and the state-of-the-art methods proposed for RGB images.

Keywords: Salient object detection, Hyperspectral imaging, Material Composition, Hyperspectral unmixing, Spectral-spatial distribution

1. Introduction

Salient object detection is an important technique in computer vision and pattern recognition. The original concept of saliency comes from neuroscience, which refers to an object that stands out from its neighbors because of its state or quality [1, 2]. Human vision system perceives salient objects from their color, brightness, texture, movements, and other pertinent properties.

Based on the above assumption, existing salient object detection models extract different features on intensity, color, texture, orientation, and so on. However, these models often fail when the salient objects have similar color as their surrounding regions [3]. Furthermore, objects made of different materials may be perceived as the same color which are no long distinguishable for human vision. For example, in the natural environment, animals may present similar color with their living environment for hiding themselves, as shown in Fig. 1. In military, camouflage is widely used to disguise objects in the background. In daily life, it is not uncommon that an object of interest has similar color as its surrounding environment, for example, a pedestrian dressed in green standing beside a tree. Another example is an accident of Tesla self-driving car in May 2016 in which the driver was killed. One of the reasons for this accident is the failure of distinguishing a white side of the tractor trailer against a brightly lit sky¹. In such cases, objects consist of different materials may be perceived as the same color though their spectral power distributions are different. Such phenomenon is called metamerism in colorimetry, which is caused by the limitation of human vision system that can only perceive primary colors generated by three types of cone cells, each of which generates a receptor response by integrating light emitted, transmitted, or reflected in the visible light spectrum from objects in the scene.

¹<https://www.tesla.com/blog/tragic-loss>

Figure 1: Examples of metamerism. The upper row shows two examples of animals camouflage in similar color as their living environment. The bottom row shows differences of spectral responses of a real and a plastic plant with the similar color.

The same problem applies to widely used RGB cameras which produce red, green, and blue color channels.

Although objects hidden in the background are no longer salient in human vision, they still stand out from their surrounding environment due to their physical composition. The research in this area is still under-investigated, despite the wide success of salient object detection in RGB images. Therefore, we propose to tackle the problem of material based salient object detection. Just as traditional salient object detection which can be used for recognition, segmentation, image compression, and visual tracking, material based salient object detection can be used for similar applications. Particularly, it can be used in many cases in which traditional salient object detection cannot work well, such as pest control, anomaly detection and camouflage recognition.

An effective way of alleviating metamerism problem is to provide detailed description of the spectral responses perceived from the objects. Hyperspectral imaging is such a technology that generates fine wavelength indexed band images across the visible to infrared spectrum. A typical hyperspectral image contains tens or hundreds of continuous band images. The spectral response captured at a pixel forms a vector of intensity values which are

closely related to the material composition of objects under study, as shown in Fig. 1. Hyperspectral imaging has been widely adopted in remote sensing and has attracted increasing interests from the computer vision and pattern recognition community [4, 5, 6].

50 The main challenges of adopting hyperspectral image in salient object detection are from several aspects. First, there are very few open datasets available to support computer vision tasks due to the relatively high cost of hyperspectral imaging devices and image labelling process [7, 8, 9, 10]. Second, the quality of hyperspectral images is sometimes low since the im-
 55 ages are often deteriorated by noises and having low spatial resolution due to various constraints in the imaging process [11]. In many image analysis tasks, the quality of hyperspectral images has to be improved before undertaking image analysis [12, 13, 14]. Third, due to the high dimensionality of hyperspectral images, traditional pattern recognition and computer vision
 60 methods based on grayscale or color image analysis can not be directly applied to hyperspectral imagery. Effective methods have to be developed to explore both spectral and spatial property of images [15, 16, 17].

As a consequence, there are very few salient object detection models for hyperspectral images. Moan *et al.* [18] combined spectral channels into red,
 65 green, and blue groups then calculate the spectral differences on each of the color groups to get the spectral saliency. Zhang *et al.* [19] extended this work by detecting seed point from initially constructed saliency map and then segmenting a multi-scale image by region growing from the seed. Then the saliency map and segmentation map are combined to detect salient
 70 objects. Also based on segmentation strategy, Yan *et al.* [20] used spectral gradient instead of raw spectral vectors to calculate the salient region. In our previous work [21], we proposed several methods based on the visual saliency model from Itti *et al.* [1]. These included conversion a hyperspectral image into an RGB image, replacing the color double-opponent component with
 75 grouped band component, and directly using the raw spectral signature to replace the color component for cross-scale analysis. However, the problem of these methods is that they are all based on pixel-wise analysis and have not considered global distribution of spectral responses. None of these works
 80 extract the salient object from viewpoint of the material composition.

In this paper, we propose a method to detect salient objects in hyperspec-
 80 tral images. To detect the salient object based on their materials, we first adopt a hyperspectral unmixing method to decompose an image into two matrices. The first matrix contains spectral responses of endmembers which

correspond to distinctive materials in the image. The second matrix contains
 85 the abundances or percentage of different endmembers at each pixel, which
 also describes the spatial distribution of materials across the image. Then
 a conspicuity map is constructed according to the spatial variance of each
 material. When combined with two other conspicuity maps calculated from
 cross-scale center-surround differences of spectral responses in our previous
 90 work, the final saliency map can be generated so as to detect the salient
 object.

In summary, the contribution of this paper includes the following four
 aspects:

- To the best of our knowledge, this is the first time that salient objects
 95 are detected based on extracting explicit material property embedded
 in the spectral responses via retrieval of endmembers and estimating
 their abundance.
- The novelty also comes from adopting hyperspectral unmixing model
 as a preprocessing step for salient object detection. This allows the
 100 spatial distribution of endmembers be estimated, so the method is ca-
 pable of dealing with mixed spectral responses in low spatial resolution
 hyperspectral images.
- Different from existing hyperspectral salient object detection methods,
 we developed a novel method to fuse both local and global features for
 105 hyperspectral salient object detection.
- We built a hyperspectral image dataset for material based salient object
 detection which contains mixed objects with similar color but different
 materials.

The rest of this paper is organized as follows. Section 2 introduces the
 110 related work including conversion of hyperspectral images to trichromatic
 images, several saliency detection methods developed for RGB images and
 hyperspectral unmixing. Section 3 describes the proposed method which
 consists of three parts: a graph regularized nonnegative matrix factorization
 115 based hyperspectral unmixing method, estimation of spectral-spatial distri-
 bution of materials, and conspicuity map fusion for salient object detection.
 Experimental results are presented in Section 4. Finally, conclusions are
 drawn in Section 5.

2. Related Works

In this section, we review methods on hyperspectral to color image conversion, hyperspectral saliency detection, and hyperspectral unmixing, which are closely related to key steps in our method.

2.1. Hyperspectral to Color Image Conversion

One reason that has hindered the adoption of saliency extraction in hyperspectral object detection is the large number of bands in the spectral data. This makes the color component not able to be calculated directly. Furthermore, effective extraction 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 allows traditional saliency model to be applicable. As pointed out in [18], this can be achieved by dimensionality reduction, band selection, or color matching functions.

Some methods can be used to convert the hyperspectral images into trichromatic images. The most straightforward way is to select three of the original spectral band images and treat them as red, green, and blue channels of a color image. More advanced methods use color matching function [7], principal component analysis [22], constrained band selection [23], or use additional color images to help the conversion [24]. A classic method was developed by Foster *et al.* [7] first converts the hyperspectral image to a CIE XYZ image. Given a hyperspectral image $\mathbf{H}(\lambda_i)$ for each of the bands λ_i , such conversion can be implemented by the following color matching function:

$$\mathbf{I}_t = \sum_{i=1}^B \mathbf{H}(\lambda_i) \mathbf{A}_t(\lambda_i) \quad (1)$$

where B is the total number of bands, $t = \{X, Y, Z\}$ are the tristimulus component of the color space, and $\mathbf{A}_t$ 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. In cases of hyperspectral images beyond visible range, it is not reasonable to disregard the extra bands that may be useful for the salient object detection. This can be solved by interpolating $\mathbf{A}_t$ to the required range to generate an extended color matching function. This approach is adopted in our experiments when a color image based method is applied.

2.2. Salient Object Detection in RGB images

After a trichromatic image is obtained, many methods developed for trichromatic images can be used. These methods are based on either the rarity/contrast of object or learning procedures. Reviewing all saliency detection models goes beyond the scope of this paper. We mainly review those methods that have motivated the idea in our method and those that are used for experimental comparison.

The most classic saliency detection method was proposed by Itti *et al.* [1]. It mimics the behavior and structure of the early primate visual system. The detailed step of 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 calculating the differences between each pair of fine and coarse scales, six intensity channels are derived. The second set of features are calculated 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 produced over three scales, which leads to 12 channels. The orientation features are generated using a set of even-symmetric Gabor filters. The dominant orientation at each pixel is recovered, whose center-surround differences are calculated 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.

Apart from the traditional methods based on the center-surround difference, many RGB-based saliency detection methods are built upon the learning procedure and other clues such as the center-surround histogram, color spatial distribution, and global contrast and spectral analysis [3, 25, 26, 27, 28, 29]. Particularly relevant to our method is the model from Liu *et al.* [27], which extracts a set of features to represent the salient object locally, regionally and globally. It adopts the conditional random field to learn the weight for each feature and then combines them into the saliency map. In this paper, it assumes that the color component of salient object tends to concentrate in the spatial domain but the background objects usually have color distributed across the image. Some methods have also been introduced in the compressed domain since most images over Internet are typically stored in such manner [30, 31, 32]. In [30], a novel saliency detection model was proposed to extract the intensity, color, and texture features of the images from

discrete cosine transform (DCT) coefficients in the JPEG bit-stream. This method was extended to video bitstreams, in which luminance, color, texture, and motion are extracted from the discrete cosine transform coefficients and motion vectors [31]. A benchmark database and computational model for stereoscopic video can be found in [32]. With flourishing of deep learning models, more recent works have focused on extracting the deep feature for salient object detection [33, 34, 35].

Similar to salient object detection in nature scene images, there are also a few works proposed for object detection in remote sensing images [36, 37, 38]. Han et al. converted the object detection to a weakly supervised learning (WSL) problem using high-level feature learning [36]. Cheng et al. proposed a novel feature representation for object detection by introducing and learning a new rotation-invariant layer on the basis of the existing CNN architectures [37].

However, as mentioned in previous section, the challenges faced by hyperspectral images have hindered the adoption of newly developed salient object detection for RGB images to hyperspectral images due to several reasons. First, the expense of hyperspectral image collection nowadays makes the dataset much smaller than the RGB images. It means that advanced learning methods can not be used for hyperspectral salient object detection due to lack of training samples. Second, the high-dimension of hyperspectral images is its advantage regarding the discriminative ability but a problem for those models proposed for RGB images. Band compression is a solution but also it may destroy the completeness of spectral features. Therefore, in our previous study[21], we borrowed the idea of multi-scale center-surrounding difference to extract the local salient features in hyperspectral images. The work from Liu et al. motivated our idea to calculate the distribution of material composition to extract the global salient features in hyperspectral images. In this work, we combine the global and local feature to develop a novel model for hyperspectral salient object detection.

2.3. Hyperspectral Unmixing

In the proposed method, hyperspectral unmixing is used to analyse the material composition in a hyperspectral image. Here we give a brief introduction to hyperspectral unmixing.

Due to low resolution of the imaging sensors, hyperspectral images usually contain mixed pixels. This phenomenon exists in most remote sensing

215 applications, but also often happens to close-range images. Hyperspectral un-
 mixing is an important technique to decompose mixed pixels to a collection
 of spectral signatures, or endmembers, and their corresponding proportions,
 i.e., abundances [39].

In recent years, many unmixing methods have been developed. They can
 220 be divided into two categories: geometry based and statistics based. Ge-
 ometrical based theory methods often assume that pure pixels exist in the
 image data and these pixels can be treated as endmembers. PPI [40, 41],
 N-FINDR [42] and VCA [43] belong to this category. The advantage of
 these methods is that they are fast and often generate good estimation of
 225 endmembers. When there is no pure pixels in the image data, endmembers
 can be treated as vertices in a minimum volume of simplex, which can be
 estimated using MVSA [44] and SISAL [45] approaches. However, when pix-
 els are highly mixed, geometrical based theory can hardly get good results.
 In this cases, statistics based methods are the solution. The representative
 230 methods are ICA [46], Bayesian approach [47] and nonnegative matrix fac-
 torization (NMF) methods [48, 49]. Some constraints are usually added to
 the unmixing model to enhance the results, such as spatial constraint [50],
 spectral constraint [51] and structure constraint [52].

The problem of statistics based method is that the optimization pro-
 235 cess can be time-consuming when the initial estimation of endmember is not
 good. For this reason, some approaches use a geometry based method to
 give initialize endmembers, and then adopt statistics based method to refine
 the unmixing results [53, 54]. In our work, we adopted this strategy and
 combined VCA and a graph regularized NMF method for endmember and
 240 abundance estimation in the scene.

3. Proposed Method

To detect the salient object with material information, we firstly analyse
 the material composition and its proportion in the hyperspectral images us-
 ing a hyperspectral unmixing method based on graph regularized nonnegative
 245 matrix factorization. Inspired by the color spatial distribution method [27],
 we calculate the spectral-spatial distribution using the abundance map from
 the unmixing results such that the saliency is measured based on the variance
 of spectral distribution in the spatial domain. After that, the derived con-
 spicuity map is combined with two conspicuity maps generated from multi-
 250 scale center-surround difference in the spectral spatial domain. The final

Figure 2: Flowchart of the proposed method.

salient object is labelled by binarizing the saliency map using an optimal threshold. The flowchart of the proposed method is summarized in Fig. 2.

3.1. Spectral Abundance Estimation via Hyperspectral Unmixing

Due to low the spatial resolution of hyperspectral images, a pixel in the image may contain mixed responses from several endmembers which correspond to distinctive materials in the scene. Furthermore, hyperspectral images are prone to be corrupted by noises due to low signal to noise ratio caused by narrow spectral band, short exposure, and intrinsic limitation of hyperspectral sensors. Hyperspectral unmixing is a method to deal with such problems. It decomposes an image into an endmember matrix and an abundance matrix with additive noises. The endmember matrix contains the spectral signatures of distinctive materials in the image. The abundance map shows the percentage of each endmember at each pixel.

To help the understanding of hyperspectral unmixing, we show the unmixing results of a hyperspectral scene in Fig. 3. The original hyperspectral image is converted to a color image as shown in Fig. 4. The abundance map in the second row of Fig. 3 associates image intensity with proportion of end-

Figure 3: Unmixing results for a hyperspectral scene. The first row shows spectral responses of endmembers and second row illustrates the corresponding abundance maps. The last row show the distribution of color components extracted using the method in [27].

members, the higher the intensity, the higher abundance of an endmember. It can be observed that the abundance maps give clear representation of the distribution of endmembers which correspond to different material components, including pink petals, white petals, leaves, stamen, and so on. At the same time, we also draw the distribution of color components based on Liu's method [27]. It clearly shows the advantage of hyperspectral unmixing regarding decomposing a scene based on its material composition.

Many methods have been proposed to solve hyperspectral unmixing problem [55]. To this end, we propose to build a linear mixture model so that the spectral responses measured by the imaging sensor can be represented by a linear combination of spectral signature of the endmembers and their corresponding fractional abundance. Our unmixing method contains two steps. In the first step, vertex component analysis (VCA) [43] method is used to give an initial estimation of the endmembers. Then in the second step, a

graph regularized nonnegative matrix factorization (NMF) method is used to refine the endmembers and estimate the abundance.

3.1.1. Vertex Component Analysis

285 VCA method [43] is a fast and effective endmember estimation method. It works particularly well when there are pure materials presented in the scene. This approach estimates endmembers one by one by projecting the data to orthogonal subspaces. A new endmember is considered as the extreme of the projection. The algorithm is shown in the following steps

- 290 1. Projecting data onto a vector orthogonal to the subspace spanned by the endmembers already determined.
2. Finding pixel with extreme projection.
3. Incorporating this pixel into endmember.
4. Generating orthogonal vector and projecting data to new vector.

295 The algorithm repeats from steps 1 to 4 until all endmembers are exhausted.

3.1.2. Graph Regularized NMF

In order to refine the endmember estimation results from the VCA method so as to cope with the sparsity distribution of endmembers and their structural relationship, we propose to adopt a graph regularized nonnegative matrix factorization method [52].

300 In linear mixture model, each pixel can be considered as the linear combination of several endmembers. Given a hyperspectral image $\mathbf{H}$ of $M \times H \times B$ in size, where M , H , and B are the width, height, and number of wavelength-indexed bands of this image, respectively. A pixel $\mathbf{h}$ is an $B \times 1$ column vector whose entries correspond to the reflectance of object in different bands. Let $\mathbf{M}$ be an $B \times P$ endmember matrix $(\mathbf{m}_1, \dots, \mathbf{m}_j, \dots, \mathbf{m}_P)$, where $\mathbf{m}_j$ is an $B \times 1$ column vector representing the spectral signature of the j^{th} endmember, and P is the total number of endmembers. Then $\mathbf{h}$ can be approximated by a linear combination of endmembers

$$\mathbf{h} = \mathbf{M}\mathbf{r} + \mathbf{e} \quad (2)$$

where $\mathbf{r}$ is a $P \times 1$ column vector for endmember abundances, and $\mathbf{e}$ is the additive Gaussian white noise.

Let N be the number of pixels in image $\mathbf{H}$, the matrix form of the linear mixture model can be defined as

$$\mathbf{H} = \mathbf{M}\mathbf{R} + \mathbf{E} \quad (3)$$

where $\mathbf{R}$ is a $P \times N$ abundance matrix which contains the proportion of each
 305 endmember at each pixel. $\mathbf{E}$ is a $B \times N$ matrix representing the additive
 noise.

The purpose of unmixing is to estimate the signature matrix $\mathbf{M}$ and
 abundance matrix $\mathbf{R}$. NMF is a nature method to solve this problem, which
 treats the unmixing as a blind source separation problem, and estimates both
 310 endmembers and abundances simultaneously. However, a disadvantage of the
 NMF method is that the cost function is non-convex, so it does not lead to
 a unique solution. This requires a constraint to be applied to the unmixing
 model. Sparsity is one of the most useful constraints, which forces matrix $\mathbf{R}$
 to be sparse in most circumstances. This is consistent with the assumption
 that most materials in a scene shall appear in only limited locations.

315 However, sparsity constrained NMF only considers the sparse characteris-
 tic and Euclidean structure of the hyperspectral data. It does not reflect the
 relationship of spectral responses of endmembers in the high dimensional data
 space. Mikhail and Partha [56] propose a geometrically motivated algorithm
 for constructing a representation for data sampled from a low dimensional
 320 manifold embedded in a higher dimensional space. This idea was adopted by
 Lu et. al. [52] who proposed that a graph regularized sparse constraint can
 be further combined with the $L_{1/2}$ sparsity constraint.

Equation 2 tells that $\mathbf{r}$ is the representation of $\mathbf{h}$ in a P -dimensional
 space and matrix $\mathbf{M}$ is the basis matrix of the new space. It means that
 the abundance vector $\mathbf{r}$ can be considered as the projection of $\mathbf{h}$ from an
 B dimensional space to P dimensional space, with $B \gg P$. In this way, it
 establishes a link between the original image $\mathbf{H}$ and the abundance map $\mathbf{R}$.
 Based on the manifold learning theory [56], the structure of the data manifold
 shall be preserved before and after projection from a high dimensional space
 to a low dimensional space. That is to say, a good unmixing model shall
 preserve the manifold structure between the original spectral responses and
 their mapped abundances, i.e., if a B -dimensional data point $\mathbf{h}_i$ is close to
 $\mathbf{h}_j$, then the corresponding abundance $\mathbf{r}_i$ is also close to $\mathbf{r}_j$. Based on this
 assumption, manifold constraint can be implemented by building a nearest
 neighbor graph from high dimensional data with the data points as vertices.
 In this graph, $\mathbf{h}_i$ and $\mathbf{h}_j$ are adjacent to each other if $\mathbf{h}_i$ is one of the nearest
 neighbors of $\mathbf{h}_j$. Given two adjacent pixels $\mathbf{h}_i$ and $\mathbf{h}_j$, the corresponding
 entry of the adjacency matrix $\mathbf{W}$ is calculated using a heat kernel

$$\mathbf{W}_{ij} = \exp - \frac{\|\mathbf{h}_i - \mathbf{h}_j\|^2}{\sigma} \quad (4)$$

According to the analysis above, when $\mathbf{h}_i$ and $\mathbf{h}_j$ are close, $\mathbf{r}_i$ and $\mathbf{r}_j$ in the new space should also be close. We implement this with the following equation:

$$\begin{aligned} \frac{1}{2} \sum_{i,j=1}^N \|\mathbf{r}_i - \mathbf{r}_j\| \mathbf{W}_{ij} &= \sum_{i=1}^N \mathbf{r}_i^T \mathbf{r}_i \mathbf{D}_{ii} - \sum_{i,j=1}^N \mathbf{r}_i^T \mathbf{r}_j \mathbf{W}_{ij} \\ &= Tr(\mathbf{RDR}^T) - Tr(\mathbf{RWR}^T) \\ &= Tr(\mathbf{RLR}^T) \end{aligned} \quad (5)$$

where $Tr()$ is the trace of the matrix. $\mathbf{D}$ is a diagonal matrix which is can be calculated as follows

$$\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij} \quad (6)$$

$\mathbf{L}$ is a Laplacian matrix used as the manifold constraint. It can be calculated as the difference of the degree matrix $\mathbf{D}$ and the adjacency matrix $\mathbf{W}$ of the graph

$$\mathbf{L} = \mathbf{D} - \mathbf{W} \quad (7)$$

Then, the objective function of the graph regularized sparse NMF can be defined as

$$obj(\mathbf{M}, \mathbf{R}) = \frac{1}{2} \|\mathbf{H} - \mathbf{MR}\|_F^2 + \gamma \|\mathbf{R}\|_{1/2} + \frac{\mu}{2} Tr(\mathbf{RLR}^T) \quad (8)$$

where $\|\mathbf{R}\|_{1/2}$ is a sparsity regularization term whose importance is controlled by γ . μ is a scalar which weights the contribution of manifold constraint defined in $Tr()$ that represents the trace of a matrix. The first two terms on the right-hand side of the equation forms the $L_{1/2}$ -NMF method [50]. The third term is the manifold term. According to [52], the solution of the objective function is as follows:

$$\mathbf{M} \leftarrow \mathbf{M} * \mathbf{HR}^T ./ \mathbf{MRR}^T \quad (9)$$

$$\mathbf{R} \leftarrow \mathbf{R} * (\mathbf{M}^T \mathbf{H} + \mu \mathbf{RW}) ./ (\mathbf{M}^T \mathbf{MR} + \frac{\gamma}{2} \mathbf{R}^{-\frac{1}{2}} + \mu \mathbf{RD}) \quad (10)$$

Figure 4: Conspicuity map built from spectral-spatial distribution of a hyperspectral scene.

3.2. Saliency Extraction Based on Spatial Distribution of Materials

Many RGB-based saliency detection methods calculate the spatial distribution of color components. For example, the model from Liu *et al.* [27] assumes that the color component of salient object tends to concentrate in spatial domain but the background objects usually have color distributed across the image. Consequently, the variance of global distribution of color components can be used to indicate the extent of saliency of an object. This characteristic of saliency suits hyperspectral images very well considering its highly discriminative ability based on the intrinsic reflectance property of objects. Instead of estimating the color spatial distribution, we propose to use the spectral spatial distribution of materials to calculate the saliency.

In the unmixing step, the abundance map $\mathbf{R}$ can be used to estimate the probability of each endmember so as to calculate the spectral spatial distribution. The probability of an endmember $\mathbf{m}_c$ at a pixel u can be calculated as follows

$$p(\mathbf{m}_c | (u_x, u_y)) = \mathbf{R}(\mathbf{m}_c, u_x + u_y * M) \quad (11)$$

where (u_x, u_y) are the coordinates of pixel u . $u_x + u_y * M$ represents the index of a pixel after the all pixels in the image are rearranged sequentially based on their spatial location in the unmixing step.

The variance of the spatial distribution of an endmember $\mathbf{m}_c$ is the com-

bination of variance along x and y dimensions, which are calculated as:

$$\begin{aligned}\sigma_x(\mathbf{m}_c) &= \frac{1}{Z} \sum_{i=1}^M \sum_{j=1}^H p(\mathbf{m}_c | (u_i, u_j)) \cdot |u_i - \bar{x}_c|^2 \\ \sigma_y(\mathbf{m}_c) &= \frac{1}{Z} \sum_{j=1}^H \sum_{i=1}^M p(\mathbf{m}_c | (u_i, u_j)) \cdot |u_j - \bar{y}_c|^2 \\ \sigma(\mathbf{m}_c) &= \sigma_x(\mathbf{m}_c) + \sigma_y(\mathbf{m}_c)\end{aligned}\tag{12}$$

where $\mathcal{N}$ is normalizing factor calculated as follows:

$$Z = \sum_{i=1}^W \sum_{j=1}^H p(\mathbf{m}_c | (u_i, u_j))\tag{13}$$

$\bar{x}_c$ is the average x coordinate for component c and can be calculated as

$$\bar{x}_c = \frac{1}{Z} \sum_{i=1}^W \sum_{j=1}^H p(\mathbf{m}_c | (u_i, u_j)) \cdot u_i\tag{14}$$

$\bar{y}_c$ can be calculated in a similar way.

With $\sigma(\mathbf{m}_c)$ being ready, we can calculate the conspicuity map from the spectral spatial distribution. The conspicuity value at each pixel u is calculated according to the probability of u belonging to endmember c and the spatial variance of endmember c . To reduce the effect from the uncommon components of small objects near boundaries, we also weight the conspicuity value based on the distance between a pixel and the image center d_u . As a result, the conspicuity map based on spectral spatial distribution is formulated as:

$$f(u) = \sum_c p(\mathbf{m}_c | u) (1 - \sigma(\mathbf{m}_c)) (1 - D(\mathbf{m}_c))\tag{15}$$

where

$$D(c) = \sum_u p(\mathbf{m}_c | u) d_u\tag{16}$$

which assigns less importance to the components near boundaries. Both $\sigma(\mathbf{m}_c)$ and $D(\mathbf{m}_c)$ are normalized to $[0, 1]$.

³⁴⁰ We draw the conspicuity map extracted from the spectral spatial distribution of a hyperspectral scene in Fig. 4. It can be observed from Figs. 3-4

that petals and stamen of flowers stand out in the conspicuity map as their spectral responses concentrate in the spatial domain. On the contrary, leave responses are suppressed due to the wide spreading of their corresponding spectral responses. Essentially, the spectral spatial distribution reflects the material composition in a scene. The less frequent a material appears across the image and more concentrated its spatial distribution is, it is more likely that this material belongs to a salient object. On one hand, the assumption of this method is consistent with our instinct. On the other hand, since this method takes the advantage of spectral information in hyperspectral images, it enables detection of salient objects beyond the human vision capability. Therefore, we expect that this method will give more accurate estimation of the saliency map compared to the traditional methods based on color information.

3.3. Salient Object Detection

Besides the spectral spatial conspicuity map generated above, we also follow the method in our previous work [21] to generate two more maps based on cross-scale center-surround spectral differences. This allows local spectral differences of salient objects being detected. These two maps are calculated based on the vector distances between the spectral responses and its neighborhood, followed by calculating the differences between each pair of fine and coarse scales in hyperspectral images. Two vector distances, i.e., Euclidean distance and spectral angle distance are used to measure the similarity between two spectral vectors $\mathbf{H}_i$ and $\mathbf{H}_j$:

$$d_{\text{EUD}}(\mathbf{H}_i, \mathbf{H}_j) = \sqrt{\sum_{k=1}^B (\mathbf{H}_i^k - \mathbf{H}_j^k)^2} \quad (17)$$

where B denotes the dimension of the vector, and

$$d_{\text{SAD}}(\mathbf{H}_i, \mathbf{H}_j) = \arccos \left(\frac{\mathbf{H}_i^T \mathbf{H}_j}{\|\mathbf{H}_i\| \|\mathbf{H}_j\|} \right) \quad (18)$$

In total, three conspicuity maps are generated and combined to form a saliency map. To detect a region with a salient object, we binarize the saliency map using the optimal threshold recovery method [57]. This allows pixels with low saliency values be removed. The object detection follows a winner-take-all strategy so only one salient object is detected per image.

Figure 5: Examples from the collected hyperspectral saliency dataset: the first and third rows show the images captured by a Nikon DSLR camera; the second and fourth rows show the false color images converted from hyperspectral images.

4. Experiments

4.1. Hyperspectral Dataset

In this section, we present the experimental results on the proposed method using a hyperspectral salient object detection dataset of 32 images generated in a close-range imaging setting. These images were captured using hyperspectral cameras with Acousto-optical tunable filters (AOTF). Each image has 61 bands from 400nm to 1000nm or from 900nm to 1700nm, depending on the type of objects being captured. The corresponding color images were also captured by a Nikon DSLR camera. In this dataset, the scenes are specially designed and consist of daily objects, such as fruits, toys, pens, water stains etc. Each image contains mixed objects with similar color but different materials, for example, plastic fruits among real fruits and real

Figure 6: Examples from online datasets[58, 8]. Since they are in the visible range, the false color images converted from hyperspectral images are the same as RGB ones.

pens in the background of printed pens on a paper. Fig. 5 shows some examples from the captured dataset. It can be observed that the single object made of different materials from other objects is almost impossible to detect in normal RGB images but can be distinguished via conversion from hyperspectral images to false color images. This is because hyperspectral imaging has higher distinguishability than color images thanks to the fine spectral information embedded in the images. Although this dataset is quite different from scenes that can be captured in some real applications, such as military camouflage or insect detection in the soil, it is still a good way to test the effectiveness of the proposed method. The corresponding ground truth are manually labelled using a rectangle covering the single object with distinctive materials.

To verify the effectiveness of proposed method on natural hyperspectral images, we also ran experiments on images retrieved from two online datasets. The first dataset is provided by the University of Manchester [58, 7] and contains 16 images with 33 bands from 400nm to 720nm. The second dataset consists of 50 hyperspectral images collected at Harvard University [8], which includes both indoor and outdoor scenes captured from 420nm to 720nm with 10nm interval. The images in these two datasets were not collected for salient object detection purpose, so we only selected 13 images with salient objects in their surrounding environment for the experiments. For each dataset, we used bounding boxes to label the locations of salient objects. Fig. 6 shows some examples from online datasets.

Figure 7: Material decomposition maps calculated by unmixing and color clustering. From left to right: input image, abundance 1, abundance 2, abundance 3, abundance 4, color map 1, color map 2, color map 3, color map 4.

4.2. Material Analysis Using Hyperspectral Unmixing

The advantage of the proposed method is to employ a hyperspectral unmixing method to analyse the material composition so that the salient object with rare materials can be detected. Therefore, in the first experiment, we compare the results of unmixing on hyperspectral images and color clustering on hyperspectral images. Due to the high dimension of hyperspectral images, the color clustering were conducted on the false color images converted from

hyperspectral images rather than the original RGB images for fair comparison. The hyperspectral unmixing used the graph regularized nonnegative matrix factorization method as introduced in Section 3. The color clustering was based on the method in [27]. The number of endmembers and color components were both set to four. Fig. 7 shows the comparison results.

It can be observed that unmixing method dealt effectively with noises and mixed spectral responses in the images, and provided more smooth component maps than the color clustering method. Take the apples in the fourth row as an example. The unmixing step generated four abundance maps, which correspond to the background, specularity, real apples, and plastic apple. The brighter a pixel on the abundance map is, the higher proportion the corresponding material is at the location. In the unmixing model, we increased the contribution from the abundance sparsity term, so that the generated results have higher contrast values in each abundance map. For pixels made of similar material, e.g., plastic apple, the adopted unmixing method tries to assign similar abundance to endmembers with similar spectral responses. The decomposition results based on color information could not highlight the plastic apple as a unique cluster. When the color of material based salient object is close to surrounding objects, it is very difficult to distinguish them based on color information. On the contrary, unmixing results distinguished different materials and provided accurate description on spatial distributions. The experiment demonstrates the advantage of hyperspectral unmixing compared to traditional color based method regarding material decomposition and analysis, which can be used for the subsequent material based salient object detection.

4.3. Conspicuity Maps Built From Different Clues

We then compare the conspicuity maps calculated from trichromatic images and hyperspectral images. Following the cross-scale center-surround difference concept in Itti's model [1], we calculated conspicuity maps using different features for all 45 scenes. These included intensity, orientation, trichromatic color, spectral Euclidean distance, spectral angle distance, and proposed spectral spatial distribution. Five examples of extracted conspicuity maps are shown in Fig. 8. From this figure, we have several observations. Firstly, the conspicuity maps from spectral bands are generally better than those from trichromatic color. The color feature based method completely failed to highlight the objects in the third and fourth scenes. Furthermore, results from spectral Euclidean distance and spectral angle distance were

Figure 8: Conspicuity maps generated from different features. From left to right: input image, conspicuity maps from intensity, orientation, trichromatic color, spectral Euclidean distance, spectral angle distance, and spectral spatial distribution.

complementary to each other. They were able to emphasize different parts of the object in the scene. It is probably because they made use of different characteristics of distance measurements. At last, the proposed spectral spatial distribution method achieved excellent results. For example, it successfully extracted the whole bottle in the second scene and highlighted the plastic pear out of real ones. Overall, this experiment indicates that spectral information is more powerful than trichromatic colors in distinguishing material differences.

4.4. Hyperspectral Salient Object Detection

To evaluate the performance of the proposed method, we compared it with three representative methods developed for RGB images and other three methods for hyperspectral images. The model from Itti *et al.* [1] can be treated as a baseline for salient object detection. The method from Liu *et al.* [27] uses a learning procedure to combine three saliency maps. Due to limited number of images in our dataset, we set the weight for each map as reported in the original paper instead of learning them via the conditional random field method. As a newly developed method, BMS [59] has been tested on several datasets and achieved the state-of-the-art performance. We adopted this method for comparison due to its simplicity and advantage in saliency object detection. For fair evaluation, all these methods were tested on the false color images converted from hyperspectral images instead of the original RGB images as shown in Fig. 5. The reason is that many salient objects in our dataset have similar color, texture, and shape from its surrounding environment, and the saliency comes mainly from differences in materials, so it is too difficult for these color image based methods to work. We adopted a hyperspectral visualisation tool "Scyven" [60] to perform the hyperspectral to false color image conversion. Regarding salient object detection method proposed for hyperspectral images, we implemented the method from Moan *et al.* [18] and Zhang *et al.* [19]. Moreover, we also included the results from our previous method [21] which combines spectral Euclidean map, spectral angle map and opponent band group map for salient object detection. Overall, we compared our method against six other methods. The results include saliency maps and bounding boxes for salient object detection for qualitative evaluation, and precision-recall curve and the F-measure for quantitative evaluation.

Fig. 9 displays the saliency maps generated by all seven methods under comparison. All three color image based methods were able to detect some

Figure 9: Saliency maps generated by different methods. From left to right: input image, Itti *et al.* [1], Liu *et al.* [27], BMS [59], Moan *et al.* [18], Liang *et al.* [21], Zhang *et al.* [19], and the proposed method.

Figure 10: Salient object detection results. From left to right: manually labelled image, results generated by Itti *et al.* [1], Liu *et al.* [27], BMS [59], Moan *et al.* [18], Liang *et al.* [21], Zhang *et al.* [19], and the proposed method.

material information from false color images. As a baseline method, Itti *et al.*'s method [1] could only detect part of the saliency but failed to highlight the complete salient objects in the saliency maps. This is mainly due to the similar color represented in the images. Although using three kinds of features, without the help of learning procedure and distinct colors, Liu *et al.*'s method [27] mostly distinguished foreground and background, and generated large salient regions in the center of the images. BMS method [59] achieved much better results. The drawback is that in some results, it highlighted parts of or too many objects. In comparison, methods based on spectral information showed overall better saliency maps than those based on color information.

For further comparison, Fig. 10 shows the corresponding object detection results. It can be seen that the methods in [1] and [27] either detected part of the salient objects or included large amount of background regions into the results. As expected, BMS gave the best results among the color based methods and it successfully covered the salient objects in all scenes except the rubic cubes. However, it only showed part of the bottle and was misled to nearby similar objects in several other scenes. The method from Moan *et al.* and Liang *et al.*, although detected the right objects for the second and last scenes, failed to detection the right objects in other images. The reason is that these two methods are based on the local center-surrounding features, and they are vulnerable to the local distinct regions, such as the gap among pears. By combining more features, the multiscale target extraction method [19] achieved better results, but was less accurate than the proposed method due to lack of global information. Overall, the proposed method, using both the local and global information, achieved the best results and successfully detected the salient objects in five out of eight scenes.

To provide quantitative analysis to the salient object detection methods, we calculated the precision and recall curves by changing the binarization thresholds from 0 to 1 with an interval of 0.01 for cropping the salient objects. The results are shown in Fig. 11. It can be seen that BMS and proposed method are clearly better than the rest of six methods. The color based method BMS even achieved higher precision than three hyperspectral based methods when the recall is larger than 0.4, which verifies the superiority and robust of topological structure of the scene captured by BMS. However, the proposed method still outperformed BMS to a significant extent.

Furthermore, we calculated the F-measure on each scene, for which the optimal threshold recovery method [57] was used to determine the binariza-

Figure 11: Precision-recall curves generated by different salient object detection methods.

tion threshold for each image adaptively. Then the final F-measure was calculated as the average F-measure value over the whole dataset. The results are shown in Fig. 12. Overall, this result is consistent with our previous observations. There are slight different in the ranking of BMS, Zhang and Liang's methods. This is mainly because the optimal threshold method [57] automatically found the threshold for each scene but the precision recall curve were calculated using the same threshold. Despite that, the proposed method achieved significantly better result than all other methods. These experimental results also show the advantage of combining the local and global spectral spatial features for saliency detection.

At last, we report the average computation time to generate a saliency map on a PC (Inter Core i5-4300 CPU 2.5 GHz, RAM 4GB, Windows 10). The results are shown in Fig. 13. The traditional Itti's method and BMS are the most efficient ones. In contrast, all spectral information based methods need at least seven seconds to process a scene. The proposed method spent the longest time to compute the saliency map. Therefore, efficiency is a drawback of the proposed method. This is mainly due to the time-consuming unmixing step which takes long time to optimize and converge. We aim to solve this problem in our future work.

Figure 12: F-Measure generated by different salient object detection methods.

Figure 13: The average computation time to generate a saliency map using different salient object detection methods.

5. Conclusion

535

We have presented a novel method to detect salient object detection in hyperspectral images. This method takes advantage of the detailed spectral information in hyperspectral images, and therefore enables detecting salient objects that have similar colors but different materials with its neighborhood.

To cope with noises and low spatial resolution problems, hyperspectral un-
 540 mixing approach is adopted to decompose an image into an endmember ma-
 trix and an abundance matrix. The abundance map is then used to estimate
 the global spatial distribution of materials in the scene so as to calculate a
 conspicuity map. This map is then combined with local spectral spatial dis-
 tribution map for final salient object detection. Experimental results show
 545 that this method has outperformed several baseline approaches. To the best
 of our knowledge, this is the first time that hyperspectral unmixing is intro-
 duced into salient object detection framework and is verified to be effective
 for this task.

Material based saliency detection is a challenging task. In the proposed
 550 method, we have used a basic salient object detection method to prove the
 concept. More powerful models shall be explored in the future for more ro-
 bust saliency detection. Additionally, since our method measure the saliency
 based on the variance of spectral-spatial distribution in the spatial domain,
 it can be influenced by small regions in the background. A filtering proce-
 555 dure has to be developed to solve this problem. More importantly, there is
 huge potential to apply the material based salient object detection to appli-
 cations such as agriculture and medical image processing. We will collect
 more hyperspectral saliency datasets to support this research.

6. Acknowledgements

560 This work is supported by the NSFC projects No.11602292, No.61370123,
 No.61772057, No.61701009 and BNSF project No.4162037.

References

- [1] L. Itti, C. Koch, E. Niebur, A model of saliency based visual attention
 for rapid scene analysis, IEEE Transactions on Pattern Analysis and
 565 Machine Intelligence 20 (11) (1998) 1254–1259.
- [2] A. Borji, L. Itti, State-of-the-art in visual attention modeling, IEEE
 Transactions on Pattern Analysis and Machine Intelligence 35 (1) (2013)
 185 –207.
- [3] A. Borji, M. Cheng, H. Jiang, J. Li, Salient object detection: a bench-
 570 mark, IEEE Transactions on Image Processing 24 (12) (2015) 5706–5722.

- [4] S. Kim, F. Deng, M. Brown, Visual enhancement of old documents with hyperspectral imaging, *Pattern Recognition* 44 (7) (2012) 1461–1469.
- [5] Y. Fu, Y. Zheng, I. Sato, Y. Sato, Exploiting spectral-spatial correlation for coded hyperspectral image restoration, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3727–3736.
- [6] M. Uzair, A. Mahmood, A. Mian, Hyperpsectral face recognition with spatio-spectral information fusion and PLS regression, *IEEE Transactions on Image Processing* 24 (3) (2015) 1127–1137.
- [7] D. Foster, S. Nascimento, K. Amano, Information limits on neural identification of colored surfaces in natural scenes, *Visual Neuroscience* 21 (2004) 331–336.
- [8] A. Chakrabarti, T. Zickler, Statistics of real-world hyperspectral images, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 193–200.
- [9] S. Oh, M. Brown, M. Pollefeys, S. Kim, Do it yourself hyperspectral imaging with everyday digital cameras, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2461–2469.
- [10] L. Wang, Z. Xiong, D. Gao, G. Shi, W. Zeng, F. Wu, High-speed hyperspectral video acquisition with a dual-camera architecture, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4942–4950.
- [11] H. Kwon, Y. Tai, RGB-guided hyperspectral image upsampling, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 307–315.
- [12] C. Lanaras, E. Baltsavias, K. Schindler, Hyperspectral super-resolution by coupled spectral unmixing, in: *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3586–3594.
- [13] N. Akhtar, F. Shafait, A. Mian, Bayesian sparse representation for hyperspectral image super resolution, in: *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference*, 2015, pp. 3631–3640.

- [14] M. Ye, Y. Qian, J. Zhou, Multi-task sparse nonnegative matrix factorization for joint spectral-spatial hyperspectral imagery denoising, *IEEE Transactions on Geoscience and Remote Sensing* 53 (5) (2015) 2621–2639.
- [15] Z. Fu, A. Robles-Kelly, J. Zhou, MILIS: Multiple instance learning with instance selection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33 (5) (2006) 958–977.
- [16] J. Li, J. Bioucas-Dias, A. Plaza, Spectralspatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields, *IEEE Transactions on Geoscience and Remote Sensing* 50 (3) (2012) 809–823.
- [17] Y. Qian, M. Ye, J. Zhou, Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features, *IEEE Transactions on Geoscience and Remote Sensing* 51 (4) (2013) 2276–2291.
- [18] S. Le Moan, A. Mansouri, J. Hardeberg, Y. Voisin, Saliency in spectral images, in: *Proceedings of the 17th Scandinavian Conference on Image Analysis*, 2011, pp. 114–123.
- [19] J. Zhang, W. Geng, L. Zhuo, Q. Tian, Y. Cao, Multiscale target extraction using a spectral saliency map for a hyperspectral image, *Applied Optics* 55 (28) (2016) 8089–8100.
- [20] H. Yan, Y. Zhang, W. Wei, L. Zhang, Y. Li, Salient object detection in hyperspectral imagery using spectral gradient contrast, in: *IEEE International Geoscience and Remote Sensing Symposium*, 2016, pp. 1560–1563.
- [21] J. Liang, J. Zhou, X. Bai, Y. Qian, Salient object detection in hyperspectral imagery, in: *Proceedings of the IEEE International Conference on Image Processing*, IEEE, 2013, pp. 2393–2397.
- [22] Q. Du, N. Raksuntorn, S. Cai, R. Moorhead, Color display for hyperspectral imagery, *Image and Vision Computing* 46 (6) (2008) 18581866.
- [23] S. Le Moan, A. Mansouri, Y. Voisin, J. Hardeberg, A constrained band selection method based on information measures for spectral image color

- 635 visualization, *IEEE Transactions on Geoscience and Remote Sensing* 49 (12) (2011) 5104–5115.
- [24] D. Liao, Y. Qian, J. Zhou, Y. Tang, A manifold alignment approach for hyperspectral image visualization with natural color, *IEEE Transactions on Geoscience and Remote Sensing* 54 (6) (2016) 3151–3162.
- 640 [25] P. Rosin, A simple method for detecting salient regions, *Pattern Recognition* 42 (11) (2009) 2363–2371.
- [26] T. Vikram, M. Tscherepanow, B. Wrede, A saliency map based on sampling an image into random rectangular regions of interest, *Pattern Recognition* 45 (9) (2012) 3114–3124.
- 645 [27] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, H. Shum, Learning to detect a salient object, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33 (2) (2011) 353–367.
- [28] M. Cheng, N. Mitra, X. Huang, P. Torr, S. Hu, G. Zhang, N. Mitra, X. Huang, S. Hu, Global contrast based salient region detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37 (3) (2011) 409–416.
- 650 [29] X. Hou, L. Zhang, Saliency detection: A spectral residual approach, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.
- [30] Y. Fang, Z. Chen, W. Lin, C. Lin, Saliency detection in the compressed domain for adaptive image retargeting, *IEEE Transactions on Image Processing* 21 (9) (2012) 3888–3901.
- 655 [31] Y. Fang, W. Lin, Z. Chen, C. Tsai, C. Lin, A video saliency detection model in compressed domain, *IEEE Transactions on Circuits and Systems for Video Technology* 24 (1) (2014) 27–38.
- 660 [32] Y. Fang, C. Zhang, J. Li, J. Lei, M. Perreira Da Silva, P. Le Callet, Visual attention modeling for stereoscopic video: a benchmark and computational model, *IEEE Transactions on Image Processing* 26 (10) (2017) 4684–4696.

- 665 [33] D. Zhang, J. Han, C. Li, J. Wang, X. Li, Detection of co-salient objects by looking deep and wide, *International Journal of Computer Vision* 120 (2) (2016) 215–232.
- [34] J. Han, D. Zhang, X. Hu, L. Guo, J. Ren, F. Wu, Background prior-based salient object detection via deep reconstruction residual, *IEEE Transactions on Circuits and Systems for Video Technology* 25 (8) (2015) 1309–1321.
- 670 [35] R. Zhao, W. Ouyang, H. Li, X. Wang, Saliency detection by multi-context deep learning, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1265–1274.
- [36] J. Han, D. Zhang, G. Cheng, L. Guo, J. Ren, Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning, *IEEE Transactions on Geoscience and Remote Sensing* 53 (6) (2015) 3325–3337.
- 675 [37] G. Cheng, P. Zhou, J. Han, Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images, *IEEE Transactions on Geoscience and Remote Sensing* 54 (12) (2016) 7405–7415.
- [38] X. Yao, J. Han, L. Guo, S. Bu, Z. Liu, A coarse-to-fine model for airport detection from remote sensing images using target-oriented visual saliency and CRF, *Neurocomputing* 164 (2015) 162–172.
- 685 [39] N. Keshava, J. Mustard, Spectral unmixing, *IEEE Signal Processing Magazine* 19 (1) (2002) 44–57.
- [40] J. Boardman, Automating spectral unmixing of AVIRIS data using convex geometry concepts, in: *Annual JPL Airborne Geoscience Workshop*, Vol. 1, 1993, pp. 11–14.
- 690 [41] J. Boardman, F. Kruse, R. Green, Mapping target signatures via partial unmixing of AVIRIS data 1 (1995) 23–26.
- [42] M. Winter, N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data, in: *SPIE International Symposium on Optical Science, Engineering, and Instrumentation*, 1999, pp. 266–275.
- 695

- [43] J. Nascimento, J. Bioucas-Dias, Vertex component analysis: A fast algorithm to unmix hyperspectral data, *IEEE Transactions on Geoscience and Remote Sensing* 43 (4) (2005) 898–910.
- 700 [44] J. Li, A. Agathos, D. Zaharie, J. Bioucas-Dias, A. Plaza, X. Li, Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing, *IEEE Transactions on Geoscience and Remote Sensing* 53 (9) (2015) 5067–5082.
- 705 [45] J. Bioucas-Dias, A variable splitting augmented Lagrangian approach to linear spectral unmixing, in: *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2009, pp. 1–4.
- [46] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, *Neural Networks* 13 (4) (2000) 411–430.
- 710 [47] M. Arngren, M. Schmidt, J. Larsen, Unmixing of hyperspectral images using Bayesian non-negative matrix factorization with volume prior, *Journal of Signal Processing Systems* 65 (3) (2011) 479–496.
- [48] D. Lee, H. Seung, Learning the parts of objects by non-negative matrix factorization, *Nature* 401 (6755) (1999) 788–791.
- 715 [49] D. Lee, H. Seung, Algorithms for non-negative matrix factorization, in: *Advances in Neural Information Processing Systems*, 2001, pp. 556–562.
- [50] Y. Qian, S. Jia, J. Zhou, A. Robles-Kelly, Hyperspectral unmixing via $L_{1/2}$ sparsity-constrained nonnegative matrix factorization, *IEEE Transactions on Geoscience and Remote Sensing* 49 (11) (2011) 4282–4297.
- 720 [51] N. Wang, B. Du, L. Zhang, An endmember dissimilarity constrained non-negative matrix factorization method for hyperspectral unmixing, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6 (2) (2013) 554–569.
- 725 [52] X. Lu, H. Wu, Y. Yuan, P. Yan, X. Li, Manifold regularized sparse NMF for hyperspectral unmixing, *IEEE Transactions on Geoscience and Remote Sensing* 51 (5) (2013) 2815–2826.

- [53] L. Tong, J. Zhou, Y. Qian, X. Bai, Y. Gao, Nonnegative-matrix-factorization-based hyperspectral unmixing with partially known end-members, *IEEE Transactions on Geoscience and Remote Sensing* 54 (11) (2016) 6531–6544.
- 730 [54] L. Tong, J. Zhou, X. Li, Y. Qian, Y. Gao, Region-based structure preserving nonnegative matrix factorization for hyperspectral unmixing, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10 (4) (2017) 1575–1588.
- 735 [55] J. Bioucas-Dias, A. Plaza, N. Dobigeon, Q. Du, P. Gader, J. Chanussot, Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5 (2) (2012) 354–379.
- 740 [56] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, in: *Proceedings of the 14th International Conference on Neural Information Processing Systems*, Vol. 14, 2001, pp. 585–591.
- [57] N. Otsu, A thresholding selection method from gray-level histograms, *IEEE Transactions on Systems, Man, and Cybernetics* 9 (1979) 62–66.
- 745 [58] S. Nascimento, F. Ferreira, D. Foster, Statistics of spatial cone-excitation ratios in natural scenes, *Journal of the Optical Society of America A* 19 (8) (2002) 1484–1490.
- [59] J. Zhang, S. Sclaroff, Exploiting surroundedness for saliency detection: A boolean map approach, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38 (5) (2016) 889–902.
- 750 [60] N. Habili, J. Oorloff, Scyllarus: from research to commercial software, *Proceedings of the 24th Australasian Software Engineering Conference II* (2015) 1–4.

Jie Liang received the B.E. degree in automatic control from National University of Defense Technology, Changsha, China, in 2011. He received the Ph.D. degree in computer vision and pattern recognition at Australian National University, Canberra, Australia, in 2017.

He is now working at Facility Design and Instrumentation of China Aerodynamics Research and Development Center, Mianyang, Sichuan, China. His research interest includes hyperspectral imaging, image measurement techniques and robotics.

Jun Zhou received the B.S. degree in computer science and the B.E. degree in international business from Nanjing University of Science and Technology, Nanjing, China, in 1996 and 1998, respectively. He received the M.S. degree in computer science from Concordia University, Montreal, Canada, in 2002, and the Ph.D. degree from the University of Alberta, Edmonton, Canada, in 2006.

He is a senior lecturer in the School of Information and Communication Technology at Griffith University, Nathan, Australia. Previously, he had been a research fellow in the Research School of Computer Science at the Australian National University, Canberra, Australia, and a researcher in the Canberra Research Laboratory, NICTA, Australia. His research interests include pattern recognition, computer vision and spectral imaging with their applications to remote sensing and environmental informatics.

Lei Tong received the B.E. degree in measurement and control technology and instrumentation and the M.E. degree in measurement technology and automation devices from Beijing Jiaotong University, Beijing, China, in 2010 and 2012, respectively, and the Ph.D. degree in engineering from Griffith University, Brisbane, Australia in 2016.

He is now working as a lecture at Faculty of Information Technology, Beijing University of Technology, Beijing, China. His current research interests include signal and image processing, pattern recognition, and remote sensing

Xiao Bai received the B.Eng. degree in computer science from Beihang University of China, Beijing, China, in 2001, and the Ph.D. degree from the University of York, York, U.K., in 2006. He was a Research Officer (Fellow, Scientist) in the Computer Science Department, University of Bath, until 2008. He is currently a Full Professor in the School of Computer Science and Engineering, Beihang University. He has published more than eighty papers in high quality international journals and refereed conferences. His current research interests include pattern recognition, image processing and remote sensing image analysis. He has been awarded New Century Excellent Talents

in University by Ministry of Education of China in 2012 and Excellency Scientist by China Association for Science and Technology in 2016.

Bin Wang received the the M.S. degree from National University of Defense Technology, Changsha, China, in 2008 and 2011.

795

He is now working at Facility Design and Instrumentation of China Aerodynamics Research and Development Center, Mianyang, Sichuan, China. His research interest includes image measurement and computer vision.

Accepted Manuscript

Material Based Salient Object Detection from Hyperspectral Images

Jie Liang, Jun Zhou, Lei Tong, Xiao Bai, Bin Wang

PII: S0031-3203(17)30480-6
DOI: [10.1016/j.patcog.2017.11.024](https://doi.org/10.1016/j.patcog.2017.11.024)
Reference: PR 6377

To appear in: *Pattern Recognition*

Received date: 15 February 2017
Revised date: 22 October 2017
Accepted date: 19 November 2017

Please cite this article as: Jie Liang, Jun Zhou, Lei Tong, Xiao Bai, Bin Wang, Material Based Salient Object Detection from Hyperspectral Images, *Pattern Recognition* (2017), doi: [10.1016/j.patcog.2017.11.024](https://doi.org/10.1016/j.patcog.2017.11.024)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.