

# MULTISCALE STOCHASTIC WATERSHED FOR UNSUPERVISED HYPERSPECTRAL IMAGE SEGMENTATION

*J. Angulo, S. Velasco-Forero*

*J. Chanussot*

CMM-Centre de Morphologie Mathématique  
MINES Paristech, FRANCE

{jesus.angulo;santiago.velasco}@ensmp.fr

GIPSA-Lab  
Grenoble Institute of Technology, FRANCE

jocelyn.chanussot@gipsa-lab.inpg.fr

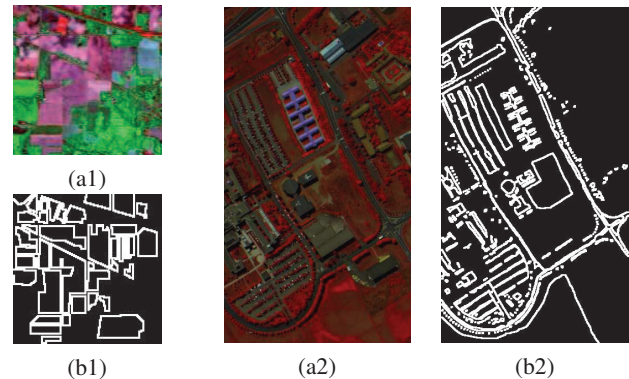
## ABSTRACT

This paper deals with unsupervised segmentation of hyperspectral images. It is based on the stochastic watershed, an approach to estimate a probability density function (pdf) of contours of an image using MonteCarlo simulations of watershed segmentations. In particular, it is introduced for the first time a multiscale framework for the computation of the pdf of contours using the stochastic watershed. Two multiscale approaches are considered: i) a linear scale-space using Gaussian filters, ii) a nonlinear morphological scale-space pyramid using levelings. In addition, a multiscale pyramid obtained by modifying the size of the random markers is also studied. Then, it is shown how the pdf of contours can finally be segmented using the non-parametric waterfalls algorithm. The performances of the proposed methods are compared using two examples of standard remote sensing hyperspectral images.

**Index Terms**— hyperspectral images, unsupervised segmentation, stochastic watershed, probabilistic hierarchical segmentation

## 1. INTRODUCTION

Watershed transformation is one of the most powerful tools for image segmentation. Starting from a gradient, the classical paradigm of watershed segmentation consists in determining markers for each region of interest. The markers avoid the over-segmentation (a region is associated to each minimum of the function) and moreover, the watershed is relatively robust to marker position [2]. The segmentation by watershed of hyperspectral images has shown to improve the results of classification in hyperspectral images [8]. To deal with some drawbacks of the classical deterministic watershed, the stochastic watershed [1] was recently introduced to detect and to regularize the contours which are robust with respect to variations in the segmentation conditions. The initial framework was then extended to multispectral images [6] and later, a specific methodology of classification-driven stochastic watershed for multivariate images such as multi/hyper-spectral images was also proposed [7].



**Fig. 1.** Two standard examples of remote sensing hyperspectral images: (1) “Indian Pines” and (2) “Pavia”. The false colour images (a) are obtained from three significant spectral bands. The images (b) represent the contours of the ground truth classification.

This paper deals with unsupervised segmentation of hyperspectral images. Fig. 1 depicts two standard examples of remote sensing hyperspectral images which will be used in the study. The first dataset is the hyperspectral image of the Indian Pines (200 spectral bands in the 400-2500 nm range, 145x145 pixels), obtained by the AVIRIS sensor. The other dataset is an airborne image from the ROSIS-3 optical sensor of the University of Pavia (103 spectral bands of 340x610 pixels). In the same figure are also given the contours associated to the ground truth. We observe, from both examples, that there are many spatial structures which clearly appear on the original images but which are not associated to contours of the spectral classes to be separated. In fact, the aim of unsupervised image segmentation is to obtain in a reliable way the contours of the most significant spatial/spectral structures. In high spatial resolution images, as the example “Pavia”, there are many complex structures at various scales and the definition of a single segmentation is a challenging difficult problem. In this paper an original probabilistic multiscale framework for the computation of the probability of contours using the stochastic watershed is introduced.

## 2. STANDARD STOCHASTIC WATERSHED FOR HYPER/MULTI-SPECTRAL IMAGES

Hyperspectral images are multivariate discrete functions with typically several tens or hundreds of spectral bands. Let  $\mathbf{f}_\lambda(\mathbf{x}) = \{f_{\lambda_j}(\mathbf{x})\}_{j=1}^L$  be a hyperspectral image, with  $\mathbf{f}_\lambda : E \rightarrow \mathcal{T}^L$ , where  $\mathbf{x} = (x, y)$  are the spatial coordinates of a vector pixel, such that  $\mathbf{x} \in E \subset \mathbb{Z}^2$ ; the space of spectral values is  $\mathcal{T}^L \subset \mathbb{R}^L$ . The scalar image  $f_{\lambda_j}(\mathbf{x})$  corresponds to the channel or band  $j$ ,  $j \in \{1, 2, \dots, L\}$ . Let  $g(\mathbf{x})$  and  $mrk(\mathbf{x})$  be a gradient image and a marker image, respectively. Intuitively, the associated watershed transformation,  $WS(g, mrk)$ , produces a binary image with the contours of regions ‘‘marked’’ by the image  $mrk$  according to the ‘‘energy of contour’’ given by the gradient image  $g$  [2]. In mathematical morphology, the gradient  $\varrho(f)(\mathbf{x})$  of an image  $f(\mathbf{x})$  is obtained as the pointwise difference between an unit dilation and an unit erosion. In addition, we consider that the values of the gradient image  $\varrho(f)$  are defined in the interval  $[0, 1]$ . The image  $mrk(\mathbf{x})$  is composed of binary connected components.

The classical paradigm of watershed segmentation lays on the appropriate choice of markers, which are the seeds to initiate the flooding procedure. In the stochastic watershed approach, an opposite direction is followed, by selecting random germs for markers on the watershed segmentation. This arbitrary choice will be balanced by the use of a given number  $M$  of realizations, in order to filter out non significant fluctuations. Each piece of contour may then be assigned the number of times it appears during the various simulations in order to estimate a probability density function (pdf) of contours. More precisely, given the scalar image of a spectral band,  $f_{\lambda_j}$ , the associated pdf of contours is obtained as follows. Let  $\{mrk_i(\mathbf{x})\}_{i=1}^M$  be a series of  $M$  realizations ( $M$  binary images) of  $N$  spatially uniformly distributed random markers. Each one of these binary images of points is considered as the markers for a watershed segmentation of gradient image  $\varrho(f_{\lambda_j})$  such that the obtained image,  $WS(\varrho(f_{\lambda_j}), mrk_i)(\mathbf{x})$ , is equal to 1 if  $\mathbf{x}$  belongs to the set of watershed lines and equal to 0 elsewhere. Consequently, a series of  $M$  segmentations is obtained, i.e.,  $\{WS(\varrho(f_{\lambda_j}), mrk_i)\}_{i=1}^M$ . Note that in each realization the number of points determines the number of regions obtained (i.e., essential property of watershed transformation). Then, the probability density function of contours is computed by the kernel density estimation method, or Parzen window method, as follows:  $pdf_{\lambda_j}(\mathbf{x}) = \frac{1}{M} \sum_{i=1}^M WS(\varrho(f_{\lambda_j}), mrk_i)(\mathbf{x}) * K(\mathbf{x}; \sigma_{spatial})$ . Typically, the kernel  $K(\mathbf{x}; \sigma_{spatial})$  is a spatial Gaussian function of width  $\sigma_{spatial}$ , which determines the smoothing effect. We took typically for our experiments  $\sigma_{spatial} = 5$ .

In the case of multivariate images, such as hyperspectral images, a marginal pdf is built for each channel, according to the previous schema. Hence, the marginal probability density function of hyperspectral image  $\mathbf{f}_\lambda(\mathbf{x})$  is computed as

$pdf(\mathbf{x}) = \frac{1}{L} \sum_{j=1}^L pdf_{\lambda_j}(\mathbf{x}) * K(\lambda; \sigma_{spectral}, \omega_j)$ , where the kernel  $K(\lambda; \sigma_{spectral}, \omega_j)$  is the product of a spectral Gaussian function of width  $\sigma_{spectral}$  and a scalar weight  $\omega_j \geq 0$  for each spectral band  $j$ . The rationale that justified this kernel is the fact that contiguous spectral bands are similar and consequently, their marginal pdfs can be smoothed together. The set of values  $\{\omega_j\}_{j=1}^L$  can be used for instance to weight the pdf according to the entropy of the band, or to the noise of the band. For simplicity, we took  $\omega_j = 1, \forall j$  and  $\sigma_{spatial} = 3$ . We notice that  $pdf(\mathbf{x})$  gives the edge strength which can be interpreted as the probability of pixel  $\mathbf{x}$  to belong to the segmentation contours, i.e.,  $pdf(\mathbf{x}) \equiv \Pr(\mathbf{x} \in \text{segmentation})$ .

## 3. MULTISCALE EXTENSION OF STOCHASTIC WATERSHED FOR HYPER/MULTI-SPECTRAL IMAGES

A series of  $S$  scale images is associated to each spectral band, i.e.,  $f_{\lambda_j}(\mathbf{x}) \mapsto \{f_{\lambda_j}^{s_s}(\mathbf{x})\}_{s=0}^S$ . Typically,  $f_{\lambda_j}^{s_0}$  is the initial spectral band, and the image  $f_{\lambda_j}^{s_{s+1}}$  is ‘‘more simplified’’ than  $f_{\lambda_j}^{s_s}$  (i.e., the scale  $\varsigma_{s+1}$  is smoother than the scale  $\varsigma_s$ ). Before to study the alternative pyramids useful for multiscale segmentation, let us introduce the probabilistic strategy to calculate the pdf of contours of the hierarchy.

### 3.1. Probabilistic framework for multiscale segmentation

Let  $\Pr(j, \text{scale}_s)$  be the segmentation probability at scale  $\varsigma_s$  of spectral band  $j$ , i.e., which can be estimated as the marginal pdf using the corresponding scale image, denoted  $pdf_{\lambda_j}^{s_s}(\mathbf{x})$ . The aim is to calculate the probability of the pyramid:  $\Pr(j, \text{multiscale}) = \Pr(j, \text{scale}_0 \text{ or } \text{scale}_1 \text{ or } \dots \text{scale}_S)$ , or in more compact form  $\Pr(j, \text{multiscale}) = \Pr(j, \cup_{s=0}^S \text{scale}_s)$ . Considering for instance two scales, using the Poincaré formulae, we have:  $\Pr(j, \text{scale}_0 \cup \text{scale}_1) = \Pr(j, \text{scale}_0) + \Pr(j, \text{scale}_1) - \Pr(j, \text{scale}_0 \cap \text{scale}_1)$ , where the joint probability of scales  $\varsigma_0$  and  $\varsigma_1$  is  $\Pr(j, \text{scale}_0 \cap \text{scale}_1) = \Pr(j, \text{scale}_0 | \text{scale}_1) \Pr(j, \text{scale}_1)$ . The conditional probability  $\Pr(j, \text{scale}_0 | \text{scale}_1)$  is the probability of scale  $\varsigma_0$  given the probability of scale  $\varsigma_1$ . Now, the challenge consists in defining a way to introduce this conditioning for computing the corresponding joint probability of two scales using simulations. We propose a quite simple approach: the pdf of scale  $\varsigma_0$  conditioned to scale  $\varsigma_1$  is obtained as  $pdf_{\lambda_j}^{s_0|1}(\mathbf{x}) = \frac{1}{M} \sum_{i=1}^M WS(\tilde{\varrho}(f_{\lambda_j}^{s_0|1}), mrk_i)(\mathbf{x}) * K(\mathbf{x}; \sigma_{spatial})$ , where the conditioned gradient function used for the stochastic watershed simulations is the sum of the gradient (defined in the interval  $[0, 1]$ ) of scale  $\varsigma_0$  and the marginal pdf of scale  $\varsigma_1$ , i.e.,

$$\tilde{\varrho}(f_{\lambda_j}^{s_0|1})(\mathbf{x}) = \varrho(f_{\lambda_j}^{s_0})(\mathbf{x}) + pdf_{\lambda_j}^{\varsigma_1}(\mathbf{x}).$$

By means of this function, the contours that appear frequently in scale  $\varsigma_1$  will have a greater probability to appear that the

contours having the same gradient in  $\zeta_0$  but a lower probability in  $\zeta_1$ . There are of course other alternatives to introduce the conditioning between the scales which will be explored in ongoing work. By recurrence, the previous results can be generalised in order to have the probability of a multiscale approach for any number of scales:

$$pdf_{\lambda_j}^S(\mathbf{x}) \equiv \Pr(j, \text{multiscale}) = \Pi_0 - \Pi_1 + \dots + (-1)^{S-2} \Pi_{S-1},$$

$$\text{where } \Pi_0 = \sum_{s=0}^S \Pr(j, \zeta_s), \Pi_1 = \sum_{s_1 \leq s_2} \Pr(j, \zeta_{s_1} \cap \zeta_{s_2}), \Pi_2 = \sum_{s_1 \leq s_2 \leq s_3} \Pr(j, \zeta_{s_1} \cap \zeta_{s_2} \cap \zeta_{s_3}), \dots, \Pi_{S-1} = \Pr(j, \cap_{s=1}^S \zeta_s).$$

### 3.2. Image multiscale pyramids and multiscale pdf of contours

The first multiscale schema considered is the classical **linear scale-space Gaussian filters**. Let  $\{\mathcal{G}_{\sigma_s}\}_{s=0}^S$  be a family of Gaussian filters, each of size  $\sigma_s$ . This pyramid is applied to each band of the hyperspectral image in order to obtain the corresponding scale, i.e.,  $f_{\lambda_j}^{\zeta_s}(\mathbf{x}) = f_{\lambda_j}(\mathbf{x}) * \mathcal{G}_{\sigma_s}(\mathbf{x})$ . Using this linear scale-space the structures are simplified progressively and the low contrasted contours are blurred.

The second pyramid is based on a **morphological nonlinear scale-space using levelings** [5]. A leveling,  $\Lambda(m, r)(\mathbf{x})$ , is an operator which has two input images: the reference image  $r(\mathbf{x})$  and the marker image  $m(\mathbf{x})$  (the last is generally a rough simplification of the reference image), and it simplifies textures and eliminates small details of the reference image according to the marked structures, but preserving the contours of remaining objects. The leveling can be computed using an iterative algorithm of geodesic dilations and erosions [5]. In our framework, the pyramid of levelings is obtained by fixing as reference the image of the spectral band and by taking as markers the family of Gaussian filters, i.e.,  $\{f_{\lambda_j}^{\zeta_s}(\mathbf{x})\}_{s=0}^S = \{\Lambda(\mathcal{G}_{\sigma_s}, f_{\lambda_j})(\mathbf{x})\}_{s=0}^S$ .

Once the Gaussian pyramid or the leveling pyramid are calculated for each spectral band, the stochastic simulations allows to obtain the marginal scale pdf's as well as the conditioned pdf's, which are the required ingredients for the probabilistic computation of the corresponding multiscale pdf of spectral band  $j$ , denoted  $pdf_{\lambda_j}^S(\mathbf{x})$ . This framework is applied to each band and then, the multiscale pdf of the hyperpectral image is obtained as

$$f_{\lambda}(\mathbf{x}) \mapsto pdf^S(\mathbf{x}) = \frac{1}{L} \sum_{j=1}^L pdf_{\lambda_j}^S(\mathbf{x}) * K(\lambda; \sigma_{spectral}).$$

The last alternative considered to introduce a multiscale construction of the pdf of contours follows a very different approach. Instead of a pyramid of each spectral band; we consider a **multiscale model for the generation of random germs**. The value  $N$  of the number of random marker is always fixed, but the germs are now disks of random radius

following an uniform distribution:  $\mathcal{U}(1, r_{max}^s)$ , and the radial parameter  $r_{max}^s$  depends on the scale  $s$ . In summary, for a particular scale  $s$ , the  $M$  realizations of random markers  $\{mrk_i^s\}_{i=1}^M$  are obtained by the implantation of  $N$  random circular sets, with possible overlaps of neighbour sets which consequently become an unique marker. In addition, regions smaller than the connected components of random markers cannot appear in the segmentation. This kind of stochastic germs can be theoretically studied under the assumptions of the Boolean model [4]. Hence, we can calculate the  $pdf_{\lambda_j}^S(\mathbf{x})$  for each spectral band using also marginal and conditioned pdf's (here, the gradient image  $\varrho(f_{\lambda_j})(\mathbf{x})$  is fixed).

### 3.3. Significant contours of multiscale pdf using waterfalls algorithm

The  $pdf^S(\mathbf{x})$  could be directly thresholded in order to obtain the most prominent contours, however the results are only pieces of contours (not enclosing regions). In addition, we have studied the histograms for several examples and there is not an optimal threshold to separate the classes of contours. An alternative technique to segment the pdf of the hyperspectral image is to use the method known as waterfalls: a hierarchical non parametric generalization of the watershed [3], which proceeds iteratively using geodesic reconstruction.

### 3.4. Experiments and Results

We have compared the multiscale results with the standard stochastic watershed. For the experiments, we fixed  $M = 50$ ,  $N = 15$  for "Indian Pines" and  $N = 100$  for "Pavia" (previous works have shown that the stochastic watershed is robust to these parameters [1]). Four levels (i.e.,  $S=3$ ) have been used for the Gaussian pyramid  $\sigma_1 = 10$ ,  $\sigma_2 = 20$  and  $\sigma_3 = 40$  and the leveling pyramid. For the model of random disks we fixed  $r_{max}^1 = 5$ ,  $r_{max}^2 = 10$  and  $r_{max}^3 = 20$ . A comparison of the various alternative multiscale stochastic watershed pdf's are given in Fig 2. In Fig. 2 are also given the hierarchical waterfall segmentation for two pdf. The first three levels of the waterfall pyramid are represented (in yellow the 1st level, in red the contours of the 2nd and in black the 3rd level). By selecting a particular level of the waterfall hierarchy, contours of the regions having a higher probability to belong to the final segmentation than the regions appearing in lower levels are obtained. Using the ground truth segmentations, we have compared quantitatively the performances with three parameters: i) the average probability of  $pdf^S$  on the ground truth contours  $\mu(\Pr)$ , ii) the detection percentage  $DP = 100TP/(TP + FN)$ , iii) the quality percentage  $QP = 100TP/(TP + FP + FN)$  (where  $TP$ ,  $FP$ ,  $FN$  are defined in terms of pixels of the contours). The values are given in the table of Fig 2. As we can observe, the three multiscale frameworks improve the standard one. We have also compared, using the leveling pyramid, the interest of the appropriate computation of the multiscale pdf including the

conditional probabilities (M-Lev) with respect to a simpler marginal calculation (M-Lev\*).

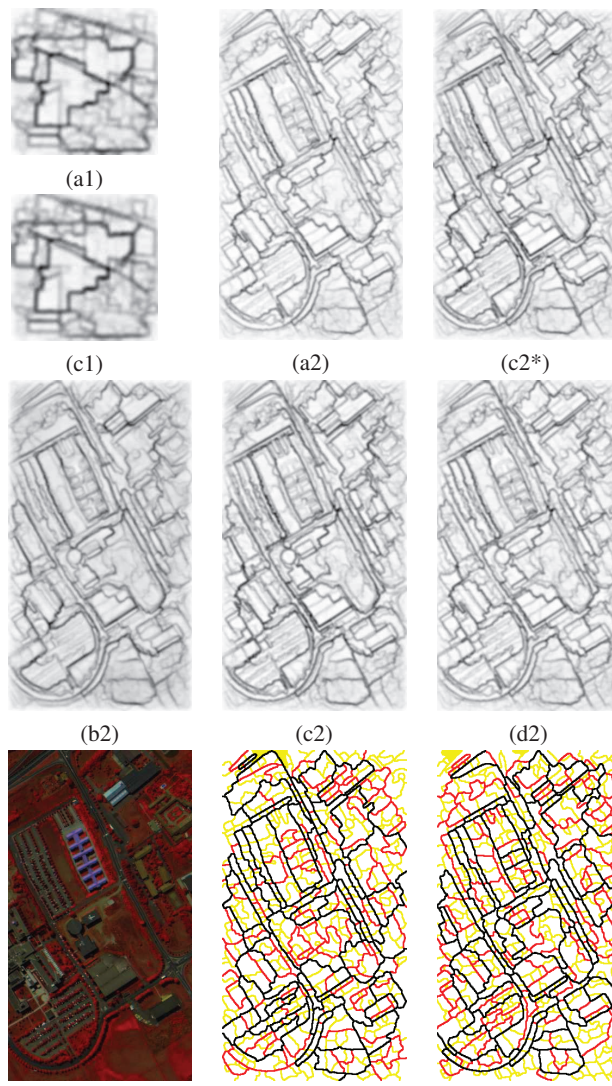
#### 4. CONCLUSIONS AND PERSPECTIVES

In this paper a novel probabilistic framework for multiscale segmentation of high spatial resolution hyperspectral images using stochastic watershed has been introduced. Work remains to be done in the integration of the unsupervised spatial information, associated to the probability of a pixel to belong to a contour, with the spectral information of the pixel used for classification. A promising direction is to use graph-based semi-supervised image segmentation techniques integrating spatial-spectral kernels [9].

**Acknowledgment.** The authors would like to thank the IAPR-TC7 for providing the data, and Prof. P. Gamba and Prof. F. Dell’Acqua of the University of Pavia, Italy, for providing reference data.

#### 5. REFERENCES

- [1] J. Angulo, D. Jeulin. Stochastic watershed segmentation. In *Proc. Int. Symp. Mathematical Morphology ISMM’07*, Rio, Brazil, INPE, Eds. Banon, G. et al., 265–276, 2007.
- [2] S. Beucher, F. Meyer. The Morphological Approach to Segmentation: The Watershed Transformation. In *Mathematical Morphology in Image Processing*, Marcel Dekker, Eds. E. Dougherty, 1992, 433–481, 1992.
- [3] S. Beucher. Watershed, hierarchical segmentation and waterfall algorithm. In *Mathematical Morphology and its Applications to Image and Signal Processing, Proc. ISMM’94* Kluwer, 69–76, 1994.
- [4] G. Matheron. *Random sets and integral geometry*, Wiley, New York, 1975.
- [5] F. Meyer. Levelings, Image Simplification Filters for Segmentation. *Journal of Mathematical Imaging and Vision*, 20: 59–72, 2004.
- [6] G. Noyel, J. Angulo, D. Jeulin. Random Germs and Stochastic Watershed for Unsupervised Multispectral Image Segmentation. In *Proc. KES 2007/ WIRN 2007*, Part III, LNAI 4694, Springer, 17–24, 2007.
- [7] G. Noyel, J. Angulo, D. Jeulin. Classification-driven stochastic watershed. Application to multispectral segmentation. In *Proc. of the IS&T’s Fourth European Conference on Color on Graphics, Imaging and Vision (CGIV’2008)*, Terrasa-Barcelona, Spain, 471–476, 2008.
- [8] Y. Tarabalka, J. Chanussot, J.-A. Benediktsson, J. Angulo, M. Fauvel. Segmentation and classification of hyperspectral data using watershed. In *Proc. of IEEE IGARSS’08*, Boston-Massachusetts, USA, July 2008.
- [9] S. Velasco-Forero, V. Manian. Improving Hyperspectral Image Classification Using Spatial Preprocessing. *IEEE Geoscience and Remote Sensing Letters*, 6(2): 297–301, 2009.



	Wfall(c2*)			Wfall(c2)	
	Standard	M-Gauss	M-Lev	M-Lev*	M-Disks
$\mu(\text{Pr})$ (1)	0.31	0.34	0.33	0.30	0.35
DP (1)	83.19	87.57	91.01	86.19	87.74
QP (1)	22.31	23.14	24.07	23.14	23.20
$\mu(\text{Pr})$ (2)	0.14	0.23	0.20	0.18	0.17
DP (2)	81.27	89.92	90.93	87.62	88.89
QP (2)	16.87	19.52	19.49	19.50	19.89

**Fig. 2.** Results using images (1) “Indian Pines” and (2) “Pavia”. Two first rows, examples of pdf of contours: (a) standard, (b) multiscale Gaussian, (c\*) multiscale leveling using only marginal computation, (c) multiscale leveling, (d) multiscale random disks. Third row, original false colour image and three levels of waterfall of pdf. Fourth row, table of quantitative results. See the text for details.