A New Approach to Segmentation of Multispectral Remote Sensing Images Based on MRF

Josef Baumgartner, Javier Gimenez, Marcelo Scavuzzo, and Julián Pucheta

***Abstract*—Segmentation of multispectral remote sensing images is a key competence for a great variety of applications. Many of the applied segmentation algorithms are generative models based on Markov random fields. These approaches are generally limited to multivariate probability densities such as the normal distribution. In addition, it is usually impossible to adjust the contextual parameters separately for each frequency band. In this letter, we present a new segmentation algorithm that avoids the aforementioned problems and allows the use of any univariate density function as emission probability in each band. The approach consists of three steps: first, calculate feature vectors for every frequency band; second, estimate contextual parameters for every band and apply local smoothing; and third, merge the feature vectors of the frequency bands to obtain final segmentation. This procedure can be iterated; however, experiments show that after the first iteration, most of the pixels are already in their final state. We call our approach successive band merging (SBM). To evaluate the performance of SBM, we segment a Landsat 8 and an AVIRIS image. In both cases, the $\widehat{\kappa}$ coefficients show that SBM outperforms the benchmark algorithms.**

***Index Terms*—Image segmentation, Markov random fields (MRFs), multispectral imaging, probability density function.**

I. Introduction

SEGMENTATION of remote sensing images is a key competence for a broad range of decision makers such as agricultural producers or local governments. In the case of agricultural producers, one can think of estimating crop parameters [1], whereas governments could be interested in wildfire management [2] or air quality measurements [3].

In the last decade, a huge number of image segmentation algorithms based on Markov random fields (MRFs) were proposed by researchers from different fields [4]–[6]. Most of these algorithms use multivariate probability functions such as the normal distribution to model multispectral images.

For many classes of images, the multivariate normal distribution might be a good choice, but in the case of remote sensing images, the gray values of the different frequency bands are often better described by univariate densities such as the Gamma distribution or Kernel density estimation. Still, many modern remote sensing algorithms are limited to the easy-to-handle normal distribution [7].

Another characteristic of remote sensing images is that the contrast of the gray values greatly varies from one band to another. In other words, it may be easy to distinguish two segments in one band but difficult in another. Therefore, a segmentation algorithm should be adoptable to the characteristics of each band when using contextual information. Nevertheless, most of the contextual segmentation algorithms require the same Markovian neighborhood in all bands [8], [9].

To overcome these two drawbacks of universal image segmentation methods, we propose a new approach for remote sensing images, which is similar to techniques such as Decision Templates or the Dempster–Shafer method [10]. The algorithm denominated *successive band merging* (SBM) has three parts: first, estimate the maximizer of the posterior marginals (MPM), then include contextual information in a nonparametric way, and finally assign a state to each pixel using a new method proposed in this work. If this procedure is iterated, it generally converges within few iterations to a final state map. Nevertheless, experiments show that after the first iteration, only few pixels are still switching states.

Note that SBM intentionally ignores the probabilistic relation between frequency bands in the first two steps. This enables us to extract hidden features of each band separately with an adequate univariate probability distribution. Only then are the feature vectors of all bands merged in the third step to obtain a segmented image. This contrasts segmentation algorithms that use multivariate distributions.

In addition, the described approach makes no assumptions about the used probability functions in each band. Suppose our image has K bands, and we want to distinguish L hidden states. Then, state one could be modeled by a Gamma distribution in band one and a Weibull distribution in band two and so forth. Moreover, our approach allows to set contextual parameters for each band according to their gray value characteristics. Hence, it is possible to work with neighborhoods of different sizes in different bands. Despite these useful features, the computational complexity of our approach is comparable to benchmark algorithms, particularly if the algorithm is not iterated.

This work is organized as follows. In Section II, we present the details of our segmentation algorithm and propose estimators for the parameters of SBM. Thereafter, we evaluate our method for two remote sensing images and compare the results to two benchmark algorithms in Section III. Finally, we outline the conclusions in Section IV.

Manuscript received November 10, 2014; revised March 6, 2015; accepted April 8, 2015.

AQ1

J. Baumgartner and J. Pucheta are with the Institute of Applied Math and Control, Universidad Nacional de Córdoba, 1611 Cordoba, Argentina (e-mail: jbaumgartner@efn.uncor.edu; jpucheta@efn.uncor.edu).

J. Gimenez is with the Institute of Automatic Control, National University of San Juan, J5402CWH San Juan, Argentina (e-mail: jgimenez@inaut.unsj.edu.ar).

M. Scavuzzo is with the Gulich Institute, Comisión Nacional de Actividades Espaciales, C1063ACH Buenos Aires, Argentina (e-mail: scavuzzo@conae.gov.ar).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2015.2421736

II. Methods

Here, the three parts of SBM are explained in detail. Suppose we have a multispectral remote sensing image I of size $M \times N$ with K frequency bands. We denote $I^{(k)}$ as the gray values of band $k \in 1, 2, \ldots, K$.

Our goal is to use the spectral information to find the most probable underlying state of every pixel of I. In other words, we are searching for the optimal state map $\mathbf{s}^*$, which is given by the maximum *a posteriori* probability, i.e.,

$$\mathbf{s}^* = \arg\max_{\mathbf{s}} P(\mathbf{s}|I, \theta). \tag{1}$$

In (1), θ's are the model parameters such as density functions and neighborhood parameters, and $\mathbf{s}$ is any admissible state map. Note that for L states, there are L^{MN} possible state maps. Even for small images, this huge number of state maps is prohibitive for the exact calculation of $\mathbf{s}^*$. Therefore, we try to approximate the optimal state map with the approach described in this section.

To start the SBM algorithm, we need an initial segmentation and the parameters of the probability functions. There are two ways to obtain the necessary data. The first option is to provide an initial guess of the density parameters. The initial segmentation can then be calculated by using maximum-likelihood classification. The second option is to run an unsupervised segmentation algorithm such as the expectation maximization (EM) algorithm [11] or k-means [12]. The resulting state map can then be used to estimate parameters of the density functions. In this letter, we use EM to initialize SBM as well as the benchmark algorithms.

A. MPM Criterion

The first step of the SBM algorithm consists in computing the marginal posterior probabilities for every pixel in every band. Therefore, let $s_{i,j}$ be the underlying state of pixel (i, j) with $i \in 1, 2, \ldots, M$ and $j \in 1, 2, \ldots, N$. Furthermore, we assume that the gray values of pixel (i, j) in the different bands depend only on $s_{i,j}$, which means $P(I_{i,j}|\mathbf{s}) = P(I_{i,j}|s_{i,j})$.

Hence, we can calculate the probability of pixel (i, j) being in state $l \in 1, 2, \ldots, L$ for the gray values of band k by using Bayes theorem, i.e.,

$$P\left(s_{i,j} = l|I_{i,j}^{(k)}\right) \propto P\left(I_{i,j}^{(k)}|s_{i,j} = l\right) P(s_{i,j} = l). \tag{2}$$

In this letter, we use noninformative priors, which means that $P(s_{i,j} = l) = 1/L$. Before we can go on with the next band, we have to normalize the posterior probabilities of all pixels such that

$$\sum_{l=1}^{L} P\left(s_{i,j} = l|I_{i,j}^{(k)}\right) = 1$$
$$\forall\, k \in 1, 2, \ldots, K; \quad i \in 1, 2, \ldots, M; \quad j \in 1, 2, \ldots, N.$$

Once we have calculated the marginal posteriors of all bands, we are done with the first step of SBM. This part of the algorithm is computationally extremely simple. Even for huge images, the computation and normalization of the marginal posteriors can be done by any average personal computer in less than a minute.

We like to point out that, so far, we have made no assumptions about the probability density functions of the different states. All we need are the posterior probabilities of pixel (i, j), but this property can be calculated for any univariate probability function. Thus, we are free to use any combination of L density functions for frequency band k.

Note that, originally, every pixel was represented by a K-dimensional data vector containing information from the different bands. Now, we have projected the input data in KL-dimensional feature space. In the following sections, we show how to take advantage of this hyperspace to segment the pixels of I.

B. MPM Averaging

To incorporate contextual information in the segmentation process, we apply a nonparametric filter, namely, the bilateral filter (BF) [13]. However, instead of smoothing the gray values of the image, we propose to run the BF directly on the marginal posterior probabilities in the feature space to avoid blurring of the gray values over several iterations. Thereby, we make use of two fundamental characteristics of the BF: spatial averaging without smoothing edges [14], or in our notation: averaging of marginal posterior probabilities of similar pixels without blurring.

First of all, we denote $\mathbf{q}_{i,j}^{(k)}$ as the posterior probabilities of pixel (i, j) for band k as described in (2), i.e.,

$$\mathbf{q}_{i,j}^{(k)}(l) = P\left(s_{i,j} = l|I_{i,j}^{(k)}\right) \quad l = 1, 2, \ldots, L.$$

According to the BF framework, we can now calculate the smoothed feature vectors $\mathbf{q}_{i,j}^{(k)*}$ by using

$$\mathbf{q}_{i,j}^{(k)*} = \sum_{i',j' \in C} K_{i,j,i',j'}^{(k)}\, \mathbf{q}_{i',j'}^{(k)}. \tag{3}$$

In (3), C represents the user-defined neighborhood or clique for band k, and $K_{i,j,i',j'}$ is the kernel of pixels (i, j) and (i', j'). The neighborhood C consists typically of all pixels within a certain radius. In this letter, we choose a radius of three pixels for all experiments. For more information on neighborhoods in MRF, please refer to [4]. As a kernel function, we use the classical Gaussian kernel, which is defined by

$$K_{i,j,i',j'}^{(k)} = \exp\left(-\frac{\|(i,j) - (i',j')\|^2}{h_x^2} - \frac{\left\|I_{i,j}^{(k)} - I_{i',j'}^{(k)}\right\|^2}{h_y^2}\right). \tag{4}$$

Note in (4) that the kernel $K_{i,j,i',j'}$ depends on the Euclidean distance of pixels (i, j) and (i', j') as well as the gray values of the two pixels. Both components—the Euclidean distance and the gray value difference—are weighted by the kernel parameters h_x and h_y, respectively. If h_x is small, only pixels very close to (i, j) are taken into account, whereas for $h_x \to \infty$, all pixels in neighborhood C are equally weighted. The same is valid for the gray values. The smaller h_y, the more discriminative the kernel with respect to the gray values. Before we go on, keep in mind that $K_{i,j,i',j'}$ has to be normalized before applying it to the posterior probabilities in (3).

For a given neighborhood C, we propose methods to estimate h_x and h_y. Let us start with h_x. The idea is to put emphasis on pixels (i',j') close to the actual pixel (i,j), but on the other hand, we do not want the kernel weights for pixels on the boarder of C to be too small. Therefore, we calculate the maximum Euclidean distance d_c in neighborhood C and set

$$h_x = \frac{\sqrt{2}}{3} d_c. \tag{5}$$

The estimation of h_x by (5) has a geometrical interpretation, which is related to the spatial part of the kernel, i.e.,

$$K^{\text{spatial}}_{i,j,i',j'} = \exp\left(\frac{-\left\|(i,j)-(i',j')\right\|^2}{h_x^2}\right). \tag{6}$$

Note in (6) that the reflection point of the spatial kernel as a function of the spatial distance lies at exactly one third of d_c if we calculate h_x according to (5). Basic mathematics show that the kernel weights for the pixels on the boarder of C are more than 1% of the maximum kernel weight. This seems to be a reasonable value, particularly for huge neighborhoods.

The second parameter of the Gaussian kernel is h_y. This parameter weighs the photometric distance between two pixels. The goal is to set h_y such that the intensities of two pixels from the same class have a high kernel, while two pixels from different classes are discriminated by the BF. Thus, it is convenient to estimate h_y on the basis of the actual state map $\mathbf{s}$, which can be obtained from the marginal posterior probabilities of the previous section. Thereby, one has to keep in mind that each band of a remote sensing image can have gray values in different ranges. Hence, it is necessary to calculate $h_y^{(k)}$ for every band k.

At this point, the optimal way to estimate $h_y^{(k)}$ would be to look at every pixel of the image and analyze its neighborhood. With this information, one could calculate the optimal $h_y^{(k)}$ by maximizing the expected kernel of two neighboring pixels in the same state. Clearly, this procedure is very costly and unpractical. Therefore, we try to approximate a parameter $h_y^{(k)}(l)$ for every state $l \in 1,2,\ldots,L$ and then average the parameters $h_y^{(k)}(l)$ to obtain $h_y^{(k)}$.

All we have to do to simplify the estimation of $h_y^{(k)}(l)$ is to maximize the expected kernel of two arbitrary pixels from the same state instead of taking into account the neighborhood of every single pixel. This is equal to calculating the maximum likelihood of $h_y^{(k)}(l)$ for each state l. The formula for state l in band k is given by

$$h_y^{(k)}(l) = \sqrt{\frac{2\sum_{s_{i,j}=l}\left(I_{i,j}^{(k)} - \mu(l)\right)^2}{\sum_{s_{i,j}=l} 1}}. \tag{7}$$

Then, we calculate $h_y^{(k)}$ as the weighted average of all states, i.e.,

$$h_y^{(k)} = \frac{\sum_{l=1}^{L} h_y^{(k)}(l)}{L}.$$

Note that using a BF to incorporate contextual information is similar to running the iterated conditional modes (ICM) algorithm [15]. There are only two notable differences between BF and ICM in this context. First of all, BF assigns spatial weights to pixels, whereas ICM uses $h_x \to \infty$ in (4). Second, ICM updates the state of pixel (i,j) according to the states of its neighbors, whereas SBM takes into account the marginal posterior probabilities of the neighboring pixels.

In Section III, we study the differences of SBM and ICM in detail, but before that, we present a new method of merging multispectral data in the following section. Therefore, it is convenient to gather the posterior probabilities from (3) in a feature vector $\mathbf{x}^*_{i,j} \in \mathcal{R}^{KL}$, i.e.,

$$\mathbf{x}^*_{i,j} = \left[\mathbf{q}^{(1)*}_{i,j}, \mathbf{q}^{(2)*}_{i,j}, \ldots, \mathbf{q}^{(K)*}_{i,j}\right]. \tag{8}$$

C. Segmentation Step

The final step of SBM is to assign one of L states to the feature vectors $\mathbf{x}^*_{i,j}$ from (8). In other words, our goal is to find L basis vectors $b_1, b_2, \ldots, b_L \in \mathcal{R}^{KL}$, to segment the feature vectors according to their Euclidean distance to these basis vectors.

Keep in mind that the feature vector $\mathbf{x}^*_{i,j}$ is composed of the marginal posterior probabilities from K bands. Therefore, we can process each band successively, starting with the first band.

Given an initial segmentation $\mathbf{s}$, we can set the basis vectors of the first band to the mean posterior probability, i.e.,

$$b_l^{(1)} = \frac{\sum_{s_{i,j}=1} \mathbf{x}^{(1)*}_{i,j}}{\sum_{s_{i,j}=l} 1} \qquad \forall\, l \in 1,\ldots,L. \tag{9}$$

In (9), $b_l^{(1)}$ stands for the basis vector of state l in band 1. The next step is to calculate the Euclidean distances of the feature vectors to the basis vectors from (9). Then, we resegment each pixel according to its distance to the basis vectors of the first band, i.e.,

$$\mathbf{s} = \left[\underset{l\in 1,2,\ldots,L}{\arg\min}\left(\mathbf{x}^{(1)*}_{i,j} - b_l^{(1)}\right)\right]_{i,j}. \tag{10}$$

Once we have finished the first step, we can sequentially add the remaining bands $2,\ldots,K$ to the segmentation process. We call $b_l^{(1:k)}$ the basis vector of state l for bands 1 to k and $\mathbf{x}^{(1:k)*}_{i,j}$ the feature vector for bands 1 to k. With this notation, we can extend (9) and (10) to

$$b_l^{(1:k)} = \frac{\sum_{s_{i,j}=1} \mathbf{x}^{(1:k)*}_{i,j}}{\sum_{s_{i,j}=l} 1} \qquad \forall\, l \in 1,\ldots,L \tag{11}$$

$$\mathbf{s} = \left[\underset{l\in 1,2,\ldots,L}{\arg\min}\left(\mathbf{x}^{(1:k)*}_{i,j} - b_l^{(1:k)}\right)\right]_{i,j}. \tag{12}$$

The idea behind this step of SBM is to update the hidden state map $\mathbf{s}$ according to (11) and (12) for $k=2$, then for $k=3$, and so on, until we reach $k=K$. After processing the last band K, we check for convergence of the state map $\mathbf{s}$. If the algorithm has not converged yet, we start again with band one. However, this time—as we have already completed one iteration—we use the state map obtained from the last iteration.

As a result, we obtain a hidden state map of a multispectral remote sensing image without using multidimensional probability density functions such as the multidimensional normal distribution. Note that once the algorithm converges, we found basis vectors $b_1, b_2, \ldots, b_L \in \mathcal{R}^{KL}$ that are valid for all bands. Hence, we achieved our main goal to incorporate the information of all bands.

In general, the presented algorithm converges within few iterations and leads to very promising results, as we will show in Section III. For a schematic description of the whole segmentation algorithm, please refer to Algorithm 1.

Algorithm 1: SBM Algorithm

1) Initialize parameters of probability distributions with a training set, k-means or GMM.
2) Calculate MPM of every pixel in every band using (2).
3) Apply BF as described by (3).
4) Segment image according to (11) and (12)
5) If no convergence of state map, go to step 2)

III. Experimental Results

Here, we use handmade ground truth and Cohen's $\widehat{\kappa}$ coefficient [16] to compare the performance of SBM with two benchmark algorithms, namely, Potts iterated conditional modes (ICM) and path constrained Viterbi training (PCVT). ICM goes back to a work of Geman and Geman [17], where they consolidated the use of Gibbs laws as prior evidence in the processing and analysis of images, whereas PCVT is based on 2-D hidden Markov models [18]. To estimate the β coefficient of ICM, we use the method proposed in [19].

All algorithms started from the same initial segmentation obtained from a GMM. Note that this might be a disadvantage for the distribution-independent SBM algorithm. Still, we chose this initialization method based on the normal distribution, because it is a widely accepted and applied algorithm.

The computational cost of SBM is approximately 10%–25% higher than PCVT and 50%–75% higher than ICM. Particularly for hyperspectral images, SBM demands more resources than the benchmark algorithms. On the other hand, the computational cost of SBM can be reduced by not iterating (9)–(12) until convergence.

A. AVIRIS Data

For the first experiment, we use Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data with 224 frequency bands. Because some bands contain negative values, we cannot use nonnegative probability functions such as the *Gamma* or the *Weibull* distribution.

The AVIRIS image with the identification number $f140528t01p00r10$ shows the Alameda Runway at N 37°47′10″, W 122°19′19″ with a pixel size of 16.40 m. In this image, we try to distinguish shallow water, deep water, sand, and the runways, as shown in Fig. 1. In the same figure, we use the bands 29, 20, and 12 to display the data as an RGB image. Moreover, we show some segmentation results. In Table I, the $\widehat{\kappa}$ coefficients of all algorithms are listed.

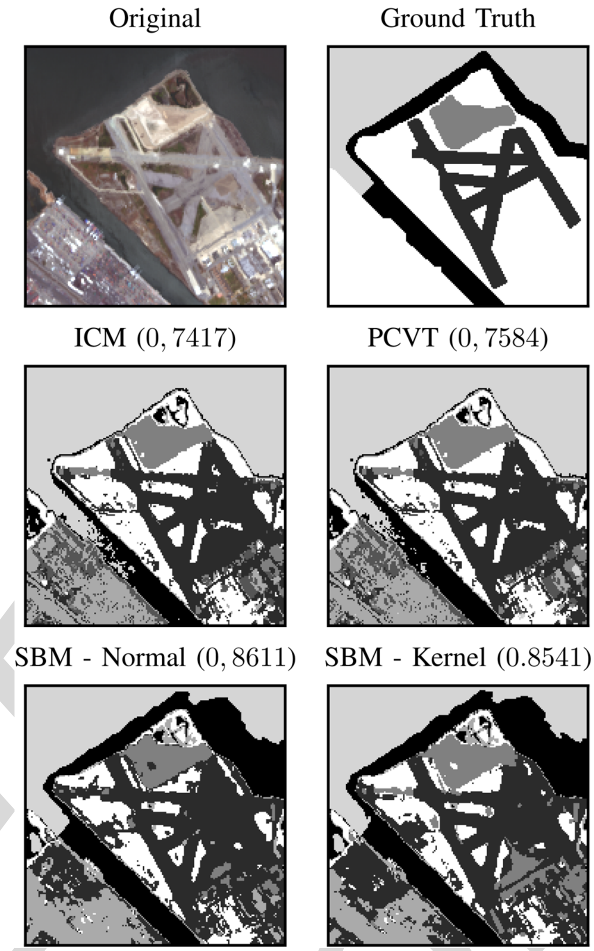

Fig. 1. Segmentation of the AVIRIS image. $\widehat{\kappa}$ values are shown in brackets.

TABLE I
Comparison of $\widehat{\kappa}$ Coefficients of the AVIRIS Image

Algorithm	$\widehat{\kappa}$	Algorithm	$\widehat{\kappa}$
ICM	0.7417	PCVT	0.7584
SBM Normal	0.8611	SBM Kernel	0.8541
SBM Gen. Extr. Value	0.8452	SBM Logistic	0.8468

B. Different Landscapes in a Landsat 8 Image

The second experiment is a multispectral Landsat 8 TM image of a mountainous region in the Humid Pampas of Argentina. It shows the San Roque lake with coordinates S 31°24′30″, W 64°29′45″, the city of Carlos Paz, agricultural fields of different sizes and orientations, and two areas that were burned by wildfires. The goal is to distinguish the following four ground-truth labels: wildfire, corn, fallow land, and water.

Some of the segmentation results for different emission probabilities are shown in Fig. 2. In Fig. 3, we compare the $\widehat{\kappa}$ coefficients of the benchmark functions and the SBM algorithm.

IV. Conclusion

In this letter, a new segmentation algorithm has been proposed and compared with two benchmark algorithms. For two test images, SBM showed good results and achieved higher $\widehat{\kappa}$ coefficients than the benchmark algorithms for most of the experiments. In the case of AVIRIS data with 224 frequency

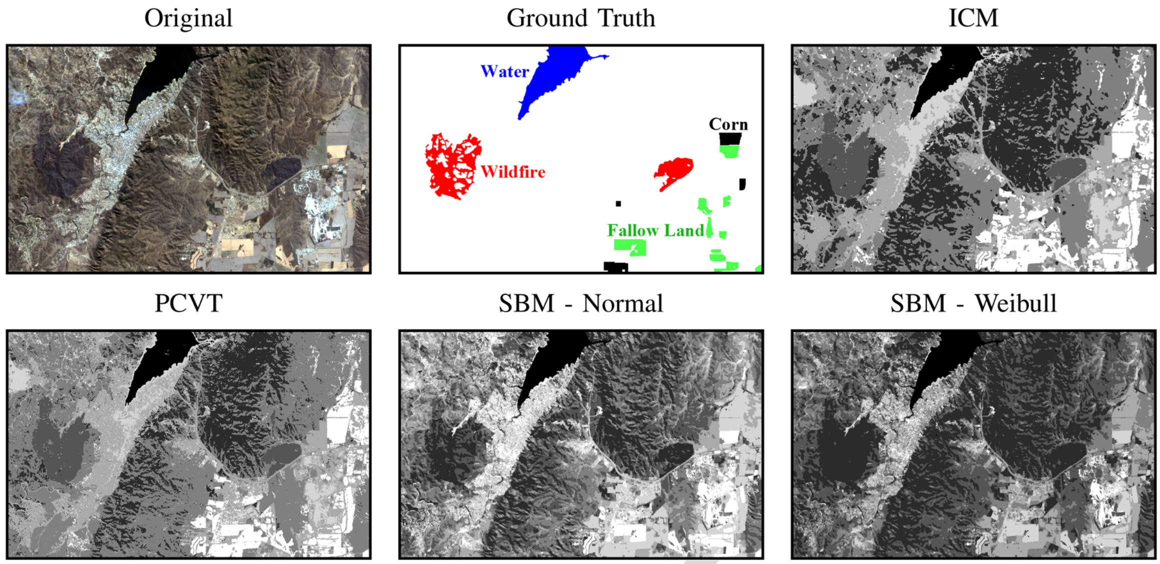

Fig. 2. Segmentations of a Landsat 8 image with seven hidden states. The four ground-truth labels are: wildfire, corn, fallow land, and water.

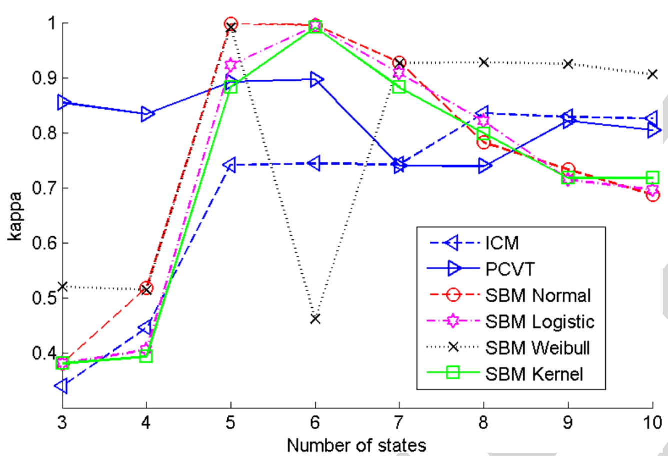

Fig. 3. Landsat 8 image: Comparison of $\widehat{\kappa}$ coefficients for different numbers of hidden states. For five to seven states, SBM clearly outperforms the benchmark algorithms for almost all probability functions. For more than seven states, SBM has the highest $\widehat{\kappa}$ values only when using the Weibull distribution.

bands, SBM was the only algorithm that distinguished shallow and deep water in a satisfactory way. In this experiment, the choice of the probability function had very little influence on the results. In the case of the Landsat 8 image, we found that the Weibull distribution is the best choice for SBM and that SBM tends to be relatively sensitive to the number of hidden states. The fact that the probability function can have great influence on the segmentation results encourages us to keep investigating algorithms that do not depend on a certain probability function.

Acknowledgment

The authors would like to thank CONICET and CONAE.

References

[1] K. Ushaan and B. Singhb, “Potential applications of remote sensing in horticulture,” *Scientia Horticulturae*, vol. 153, no. 4, pp. 71–83, Apr. 2013.

[2] S. W. Taylor, D. G. Woolford, C. B. Dean, and D. L. Martell, “Wildfire prediction to inform fire management: Statistical science challenges,” *Statistical Sci.*, vol. 28, no. 4, pp. 586–615, 2013.

[3] C. Manzo, R. Salvini, E. Guastaldi, V. Nicolardi, and G. Protano, “Reflectance spectral analyses for the assessment of environmental pollution in the geothermal site of Mt. Amiata (Italy),” *Atmospheric Environ.*, vol. 79, pp. 650–665, Nov. 2013.

[4] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. New York, NY, USA: Springer Science & Business Media, 2009.

[5] X.-Y. Fu, H.-J. You, and K. Fu, “Building segmentation from high-resolution SAR images based on improved Markov random field,” *Acta Electronica Sinica*, vol. 40, no. 6, pp. 1141–1147, 2012.

[6] J. Jing, Y. Li, P. Li, and Y. Jiao, “Textile printing pattern image segmentation based on algorithm of MRF,” *J. Inf. Comput. Sci.*, vol. 10, no. 13, pp. 4007–4015, 2013.

[7] P. Ruiz, J. Mateos, G. Camps-Valls, R. Molina, and A. K. Katsaggelos, “Bayesian active remote sensing image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2186–2196, Apr. 2014.

[8] A. L. M. Levada, N. D. A. Mascarenhas, and A. Tannús, “A novel MAP-MRF approach for multispectral image contextual classification using combination of suboptimal iterative algorithms,” *Pattern Recog. Lett.*, vol. 31, no. 13, pp. 1795–1808, Oct. 2010.

[9] F. Wang *et al.*, “Unsupervised SAR image segmentation using higher order neighborhood-based triplet Markov fields model,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5193–5205, Aug. 2014.

[10] L. Kuncheva, *Combining Pattern Classifiers*. Hoboken, NJ, USA: Wiley, 2004.

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *J. Royal Statistical Soc.*, vol. 39, no. 1, pp. 1–38, 1977.

[12] S. P. Lloyd, “Least squares quantization in PCM,” *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.

[13] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color images,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 1998, pp. 839–846.

[14] H. Peng, R. Rao, and S. A. Dianat, “Multispectral image denoising with optimized vector bilateral filter,” *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 264–273, 2014.

[15] J. E. Besag, “On the statistical analysis of dirty pictures,” *J. Royal Statistical Soc.*, vol. 48, no. 3, pp. 259–302, 1986.

[16] J. Cohen, “A coefficient of agreement for nominal scales,” *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960.

[17] S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.

[18] D. Joshi, J. Li, and J. Z. Wang, “A computationally efficient approach to the estimation of two- and three-dimensional hidden Markov models,” *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 1871–1886, Jul. 2006.

[19] J. Gimenez, A. C. Frery, and A. G. Flesia, “Inference strategies for the smoothness parameter in the Potts model,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2013, pp. 2539–2542.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

Please be aware that authors are required to pay overlength page charges ($200 per page) if the paper is longer than 3 pages. If you cannot pay any or all of these charges please let us know.

This pdf contains 2 proofs. The first half is the version that will appear on Xplore. The second half is the version that will appear in print. If you have any figures to print in color, they will be in color in both proofs.

The "Open Access" option for your paper expires when the paper is published on Xplore in an issue with page numbers. Papers in "Early Access" may be changed to Open Access.

AQ1 = Please check if the provided affiliations of all authors are correct. Otherwise, make the necessary corrections.

END OF ALL QUERIES

A New Approach to Segmentation of Multispectral Remote Sensing Images Based on MRF

Josef Baumgartner, Javier Gimenez, Marcelo Scavuzzo, and Julián Pucheta

***Abstract*—Segmentation of multispectral remote sensing images is a key competence for a great variety of applications. Many of the applied segmentation algorithms are generative models based on Markov random fields. These approaches are generally limited to multivariate probability densities such as the normal distribution. In addition, it is usually impossible to adjust the contextual parameters separately for each frequency band. In this letter, we present a new segmentation algorithm that avoids the aforementioned problems and allows the use of any univariate density function as emission probability in each band. The approach consists of three steps: first, calculate feature vectors for every frequency band; second, estimate contextual parameters for every band and apply local smoothing; and third, merge the feature vectors of the frequency bands to obtain final segmentation. This procedure can be iterated; however, experiments show that after the first iteration, most of the pixels are already in their final state. We call our approach successive band merging (SBM). To evaluate the performance of SBM, we segment a Landsat 8 and an AVIRIS image. In both cases, the $\widehat{\kappa}$ coefficients show that SBM outperforms the benchmark algorithms.**

***Index Terms*—Image segmentation, Markov random fields (MRFs), multispectral imaging, probability density function.**

I. Introduction

SEGMENTATION of remote sensing images is a key competence for a broad range of decision makers such as agricultural producers or local governments. In the case of agricultural producers, one can think of estimating crop parameters [1], whereas governments could be interested in wildfire management [2] or air quality measurements [3].

In the last decade, a huge number of image segmentation algorithms based on Markov random fields (MRFs) were proposed by researchers from different fields [4]–[6]. Most of these algorithms use multivariate probability functions such as the normal distribution to model multispectral images.

For many classes of images, the multivariate normal distribution might be a good choice, but in the case of remote sensing images, the gray values of the different frequency bands are often better described by univariate densities such as the Gamma distribution or Kernel density estimation. Still, many modern remote sensing algorithms are limited to the easy-to-handle normal distribution [7].

Another characteristic of remote sensing images is that the contrast of the gray values greatly varies from one band to another. In other words, it may be easy to distinguish two segments in one band but difficult in another. Therefore, a segmentation algorithm should be adoptable to the characteristics of each band when using contextual information. Nevertheless, most of the contextual segmentation algorithms require the same Markovian neighborhood in all bands [8], [9].

To overcome these two drawbacks of universal image segmentation methods, we propose a new approach for remote sensing images, which is similar to techniques such as Decision Templates or the Dempster–Shafer method [10]. The algorithm denominated *successive band merging* (SBM) has three parts: first, estimate the maximizer of the posterior marginals (MPM), then include contextual information in a nonparametric way, and finally assign a state to each pixel using a new method proposed in this work. If this procedure is iterated, it generally converges within few iterations to a final state map. Nevertheless, experiments show that after the first iteration, only few pixels are still switching states.

Note that SBM intentionally ignores the probabilistic relation between frequency bands in the first two steps. This enables us to extract hidden features of each band separately with an adequate univariate probability distribution. Only then are the feature vectors of all bands merged in the third step to obtain a segmented image. This contrasts segmentation algorithms that use multivariate distributions.

In addition, the described approach makes no assumptions about the used probability functions in each band. Suppose our image has K bands, and we want to distinguish L hidden states. Then, state one could be modeled by a Gamma distribution in band one and a Weibull distribution in band two and so forth. Moreover, our approach allows to set contextual parameters for each band according to their gray value characteristics. Hence, it is possible to work with neighborhoods of different sizes in different bands. Despite these useful features, the computational complexity of our approach is comparable to benchmark algorithms, particularly if the algorithm is not iterated.

This work is organized as follows. In Section II, we present the details of our segmentation algorithm and propose estimators for the parameters of SBM. Thereafter, we evaluate our method for two remote sensing images and compare the results to two benchmark algorithms in Section III. Finally, we outline the conclusions in Section IV.

Manuscript received November 10, 2014; revised March 6, 2015; accepted April 8, 2015.

AQ1 J. Baumgartner and J. Pucheta are with the Institute of Applied Math and Control, Universidad Nacional de Córdoba, 1611 Cordoba, Argentina (e-mail: jbaumgartner@efn.uncor.edu; jpucheta@efn.uncor.edu).

J. Gimenez is with the Institute of Automatic Control, National University of San Juan, J5402CWH San Juan, Argentina (e-mail: jgimenez@inaut.unsj.edu.ar).

M. Scavuzzo is with the Gulich Institute, Comisión Nacional de Actividades Espaciales, C1063ACH Buenos Aires, Argentina (e-mail: scavuzzo@conae.gov.ar).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2015.2421736

II. Methods

Here, the three parts of SBM are explained in detail. Suppose we have a multispectral remote sensing image I of size $M \times N$ with K frequency bands. We denote $I^{(k)}$ as the gray values of band $k \in 1, 2, \ldots, K$.

Our goal is to use the spectral information to find the most probable underlying state of every pixel of I. In other words, we are searching for the optimal state map $\mathbf{s}^*$, which is given by the maximum *a posteriori* probability, i.e.,

$$\mathbf{s}^* = \arg\max_{\mathbf{s}} P(\mathbf{s}|I, \theta). \tag{1}$$

In (1), θ's are the model parameters such as density functions and neighborhood parameters, and $\mathbf{s}$ is any admissible state map. Note that for L states, there are L^{MN} possible state maps. Even for small images, this huge number of state maps is prohibitive for the exact calculation of $\mathbf{s}^*$. Therefore, we try to approximate the optimal state map with the approach described in this section.

To start the SBM algorithm, we need an initial segmentation and the parameters of the probability functions. There are two ways to obtain the necessary data. The first option is to provide an initial guess of the density parameters. The initial segmentation can then be calculated by using maximum-likelihood classification. The second option is to run an unsupervised segmentation algorithm such as the expectation maximization (EM) algorithm [11] or k-means [12]. The resulting state map can then be used to estimate parameters of the density functions. In this letter, we use EM to initialize SBM as well as the benchmark algorithms.

A. MPM Criterion

The first step of the SBM algorithm consists in computing the marginal posterior probabilities for every pixel in every band. Therefore, let $s_{i,j}$ be the underlying state of pixel (i, j) with $i \in 1, 2, \ldots, M$ and $j \in 1, 2, \ldots, N$. Furthermore, we assume that the gray values of pixel (i, j) in the different bands depend only on $s_{i,j}$, which means $P(I_{i,j}|\mathbf{s}) = P(I_{i,j}|s_{i,j})$.

Hence, we can calculate the probability of pixel (i, j) being in state $l \in 1, 2, \ldots, L$ for the gray values of band k by using Bayes theorem, i.e.,

$$P\left(s_{i,j} = l|I_{i,j}^{(k)}\right) \propto P\left(I_{i,j}^{(k)}|s_{i,j} = l\right) P(s_{i,j} = l). \tag{2}$$

In this letter, we use noninformative priors, which means that $P\left(s_{i,j} = l\right) = 1/L$. Before we can go on with the next band, we have to normalize the posterior probabilities of all pixels such that

$$\sum_{l=1}^{L} P\left(s_{i,j} = l|I_{i,j}^{(k)}\right) = 1$$
$$\forall\, k \in 1, 2, \ldots, K; \quad i \in 1, 2, \ldots, M; \quad j \in 1, 2, \ldots, N.$$

Once we have calculated the marginal posteriors of all bands, we are done with the first step of SBM. This part of the algorithm is computationally extremely simple. Even for huge images, the computation and normalization of the marginal posteriors can be done by any average personal computer in less than a minute.

We like to point out that, so far, we have made no assumptions about the probability density functions of the different states. All we need are the posterior probabilities of pixel (i, j), but this property can be calculated for any univariate probability function. Thus, we are free to use any combination of L density functions for frequency band k.

Note that, originally, every pixel was represented by a K-dimensional data vector containing information from the different bands. Now, we have projected the input data in KL-dimensional feature space. In the following sections, we show how to take advantage of this hyperspace to segment the pixels of I.

B. MPM Averaging

To incorporate contextual information in the segmentation process, we apply a nonparametric filter, namely, the bilateral filter (BF) [13]. However, instead of smoothing the gray values of the image, we propose to run the BF directly on the marginal posterior probabilities in the feature space to avoid blurring of the gray values over several iterations. Thereby, we make use of two fundamental characteristics of the BF: spatial averaging without smoothing edges [14], or in our notation: averaging of marginal posterior probabilities of similar pixels without blurring.

First of all, we denote $\mathbf{q}_{i,j}^{(k)}$ as the posterior probabilities of pixel (i, j) for band k as described in (2), i.e.,

$$\mathbf{q}_{i,j}^{(k)}(l) = P\left(s_{i,j} = l|I_{i,j}^{(k)}\right) \quad l = 1, 2, \ldots, L.$$

According to the BF framework, we can now calculate the smoothed feature vectors $\mathbf{q}_{i,j}^{(k)*}$ by using

$$\mathbf{q}_{i,j}^{(k)*} = \sum_{i',j' \in C} K_{i,j,i',j'}^{(k)}\, \mathbf{q}_{i',j'}^{(k)}. \tag{3}$$

In (3), C represents the user-defined neighborhood or clique for band k, and $K_{i,j,i',j'}$ is the kernel of pixels (i, j) and (i', j'). The neighborhood C consists typically of all pixels within a certain radius. In this letter, we choose a radius of three pixels for all experiments. For more information on neighborhoods in MRF, please refer to [4]. As a kernel function, we use the classical Gaussian kernel, which is defined by

$$K_{i,j,i',j'}^{(k)} = \exp\left(-\frac{\|(i,j)-(i',j')\|^2}{h_x^2} - \frac{\left\|I_{i,j}^{(k)} - I_{i',j'}^{(k)}\right\|^2}{h_y^2}\right). \tag{4}$$

Note in (4) that the kernel $K_{i,j,i',j'}$ depends on the Euclidean distance of pixels (i, j) and (i', j') as well as the gray values of the two pixels. Both components—the Euclidean distance and the gray value difference—are weighted by the kernel parameters h_x and h_y, respectively. If h_x is small, only pixels very close to (i, j) are taken into account, whereas for $h_x \rightarrow \infty$, all pixels in neighborhood C are equally weighted. The same is valid for the gray values. The smaller h_y, the more discriminative the kernel with respect to the gray values. Before we go on, keep in mind that $K_{i,j,i',j'}$ has to be normalized before applying it to the posterior probabilities in (3).

For a given neighborhood C, we propose methods to estimate h_x and h_y. Let us start with h_x. The idea is to put emphasis on pixels (i',j') close to the actual pixel (i,j), but on the other hand, we do not want the kernel weights for pixels on the boarder of C to be too small. Therefore, we calculate the maximum Euclidean distance d_c in neighborhood C and set

$$h_x = \frac{\sqrt{2}}{3} d_c. \tag{5}$$

The estimation of h_x by (5) has a geometrical interpretation, which is related to the spatial part of the kernel, i.e.,

$$K^{\text{spatial}}_{i,j,i',j'} = \exp\left(\frac{-\left\|(i,j)-(i',j')\right\|^2}{h_x^2}\right). \tag{6}$$

Note in (6) that the reflection point of the spatial kernel as a function of the spatial distance lies at exactly one third of d_c if we calculate h_x according to (5). Basic mathematics show that the kernel weights for the pixels on the boarder of C are more than 1% of the maximum kernel weight. This seems to be a reasonable value, particularly for huge neighborhoods.

The second parameter of the Gaussian kernel is h_y. This parameter weighs the photometric distance between two pixels. The goal is to set h_y such that the intensities of two pixels from the same class have a high kernel, while two pixels from different classes are discriminated by the BF. Thus, it is convenient to estimate h_y on the basis of the actual state map $\mathbf{s}$, which can be obtained from the marginal posterior probabilities of the previous section. Thereby, one has to keep in mind that each band of a remote sensing image can have gray values in different ranges. Hence, it is necessary to calculate $h_y^{(k)}$ for every band k.

At this point, the optimal way to estimate $h_y^{(k)}$ would be to look at every pixel of the image and analyze its neighborhood. With this information, one could calculate the optimal $h_y^{(k)}$ by maximizing the expected kernel of two neighboring pixels in the same state. Clearly, this procedure is very costly and unpractical. Therefore, we try to approximate a parameter $h_y^{(k)}(l)$ for every state $l \in 1,2,\ldots,L$ and then average the parameters $h_y^{(k)}(l)$ to obtain $h_y^{(k)}$.

All we have to do to simplify the estimation of $h_y^{(k)}(l)$ is to maximize the expected kernel of two arbitrary pixels from the same state instead of taking into account the neighborhood of every single pixel. This is equal to calculating the maximum likelihood of $h_y^{(k)}(l)$ for each state l. The formula for state l in band k is given by

$$h_y^{(k)}(l) = \sqrt{\frac{2\sum_{s_{i,j}=l}\left(I_{i,j}^{(k)} - \mu(l)\right)^2}{\sum_{s_{i,j}=l} 1}}. \tag{7}$$

Then, we calculate $h_y^{(k)}$ as the weighted average of all states, i.e.,

$$h_y^{(k)} = \frac{\sum_{l=1}^{L} h_y^{(k)}(l)}{L}.$$

Note that using a BF to incorporate contextual information is similar to running the iterated conditional modes (ICM) algorithm [15]. There are only two notable differences between BF and ICM in this context. First of all, BF assigns spatial weights to pixels, whereas ICM uses $h_x \rightarrow \infty$ in (4). Second, ICM updates the state of pixel (i,j) according to the states of its neighbors, whereas SBM takes into account the marginal posterior probabilities of the neighboring pixels.

In Section III, we study the differences of SBM and ICM in detail, but before that, we present a new method of merging multispectral data in the following section. Therefore, it is convenient to gather the posterior probabilities from (3) in a feature vector $\mathbf{x}^*_{i,j} \in \mathcal{R}^{KL}$, i.e.,

$$\mathbf{x}^*_{i,j} = \left[\mathbf{q}^{(1)*}_{i,j}, \mathbf{q}^{(2)*}_{i,j}, \ldots, \mathbf{q}^{(K)*}_{i,j}\right]. \tag{8}$$

C. Segmentation Step

The final step of SBM is to assign one of L states to the feature vectors $\mathbf{x}^*_{i,j}$ from (8). In other words, our goal is to find L basis vectors $b_1, b_2, \ldots, b_L \in \mathcal{R}^{KL}$, to segment the feature vectors according to their Euclidean distance to these basis vectors.

Keep in mind that the feature vector $\mathbf{x}^*_{i,j}$ is composed of the marginal posterior probabilities from K bands. Therefore, we can process each band successively, starting with the first band.

Given an initial segmentation $\mathbf{s}$, we can set the basis vectors of the first band to the mean posterior probability, i.e.,

$$b_l^{(1)} = \frac{\sum_{s_{i,j}=1} \mathbf{x}^{(1)*}_{i,j}}{\sum_{s_{i,j}=l} 1} \qquad \forall\, l \in 1,\ldots,L. \tag{9}$$

In (9), $b_l^{(1)}$ stands for the basis vector of state l in band 1. The next step is to calculate the Euclidean distances of the feature vectors to the basis vectors from (9). Then, we resegment each pixel according to its distance to the basis vectors of the first band, i.e.,

$$\mathbf{s} = \left[\underset{l\in 1,2,\ldots,L}{\arg\min}\left(\mathbf{x}^{(1)*}_{i,j} - b_l^{(1)}\right)\right]_{i,j}. \tag{10}$$

Once we have finished the first step, we can sequentially add the remaining bands $2,\ldots,K$ to the segmentation process. We call $b_l^{(1:k)}$ the basis vector of state l for bands 1 to k and $\mathbf{x}^{(1:k)*}_{i,j}$ the feature vector for bands 1 to k. With this notation, we can extend (9) and (10) to

$$b_l^{(1:k)} = \frac{\sum_{s_{i,j}=1} \mathbf{x}^{(1:k)*}_{i,j}}{\sum_{s_{i,j}=l} 1} \qquad \forall\, l \in 1,\ldots,L \tag{11}$$

$$\mathbf{s} = \left[\underset{l\in 1,2,\ldots,L}{\arg\min}\left(\mathbf{x}^{(1:k)*}_{i,j} - b_l^{(1:k)}\right)\right]_{i,j}. \tag{12}$$

The idea behind this step of SBM is to update the hidden state map $\mathbf{s}$ according to (11) and (12) for $k=2$, then for $k=3$, and so on, until we reach $k=K$. After processing the last band K, we check for convergence of the state map $\mathbf{s}$. If the algorithm has not converged yet, we start again with band one. However, this time—as we have already completed one iteration—we use the state map obtained from the last iteration.

As a result, we obtain a hidden state map of a multispectral remote sensing image without using multidimensional probability density functions such as the multidimensional normal distribution. Note that once the algorithm converges, we found basis vectors $b_1, b_2, \ldots, b_L \in \mathcal{R}^{KL}$ that are valid for all bands. Hence, we achieved our main goal to incorporate the information of all bands.

In general, the presented algorithm converges within few iterations and leads to very promising results, as we will show in Section III. For a schematic description of the whole segmentation algorithm, please refer to Algorithm 1.

Algorithm 1: SBM Algorithm

1) Initialize parameters of probability distributions with a training set, k-means or GMM.
2) Calculate MPM of every pixel in every band using (2).
3) Apply BF as described by (3).
4) Segment image according to (11) and (12)
5) If no convergence of state map, go to step 2)

III. Experimental Results

Here, we use handmade ground truth and Cohen's $\widehat{\kappa}$ coefficient [16] to compare the performance of SBM with two benchmark algorithms, namely, Potts iterated conditional modes (ICM) and path constrained Viterbi training (PCVT). ICM goes back to a work of Geman and Geman [17], where they consolidated the use of Gibbs laws as prior evidence in the processing and analysis of images, whereas PCVT is based on 2-D hidden Markov models [18]. To estimate the β coefficient of ICM, we use the method proposed in [19].

All algorithms started from the same initial segmentation obtained from a GMM. Note that this might be a disadvantage for the distribution-independent SBM algorithm. Still, we chose this initialization method based on the normal distribution, because it is a widely accepted and applied algorithm.

The computational cost of SBM is approximately 10%–25% higher than PCVT and 50%–75% higher than ICM. Particularly for hyperspectral images, SBM demands more resources than the benchmark algorithms. On the other hand, the computational cost of SBM can be reduced by not iterating (9)–(12) until convergence.

A. AVIRIS Data

For the first experiment, we use Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data with 224 frequency bands. Because some bands contain negative values, we cannot use nonnegative probability functions such as the *Gamma* or the *Weibull* distribution.

The AVIRIS image with the identification number $f140528t01p00r10$ shows the Alameda Runway at N $37^\circ 47' 10''$, W $122^\circ 19' 19''$ with a pixel size of 16.40 m. In this image, we try to distinguish shallow water, deep water, sand, and the runways, as shown in Fig. 1. In the same figure, we use the bands 29, 20, and 12 to display the data as an RGB image. Moreover, we show some segmentation results. In Table I, the $\widehat{\kappa}$ coefficients of all algorithms are listed.

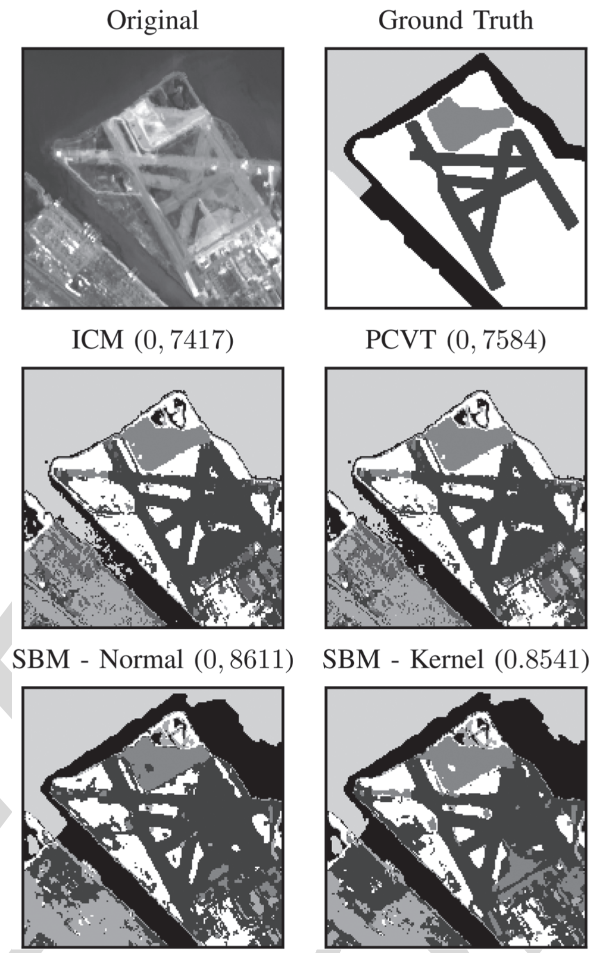

Fig. 1. Segmentation of the AVIRIS image. $\widehat{\kappa}$ values are shown in brackets.

TABLE I
Comparison of $\widehat{\kappa}$ Coefficients of the AVIRIS Image

Algorithm	$\widehat{\kappa}$	Algorithm	$\widehat{\kappa}$
ICM	0.7417	PCVT	0.7584
SBM Normal	0.8611	SBM Kernel	0.8541
SBM Gen. Extr. Value	0.8452	SBM Logistic	0.8468

B. Different Landscapes in a Landsat 8 Image

The second experiment is a multispectral Landsat 8 TM image of a mountainous region in the Humid Pampas of Argentina. It shows the San Roque lake with coordinates S $31^\circ 24' 30''$, W $64^\circ 29' 45''$, the city of Carlos Paz, agricultural fields of different sizes and orientations, and two areas that were burned by wildfires. The goal is to distinguish the following four ground-truth labels: wildfire, corn, fallow land, and water.

Some of the segmentation results for different emission probabilities are shown in Fig. 2. In Fig. 3, we compare the $\widehat{\kappa}$ coefficients of the benchmark functions and the SBM algorithm.

IV. Conclusion

In this letter, a new segmentation algorithm has been proposed and compared with two benchmark algorithms. For two test images, SBM showed good results and achieved higher $\widehat{\kappa}$ coefficients than the benchmark algorithms for most of the experiments. In the case of AVIRIS data with 224 frequency

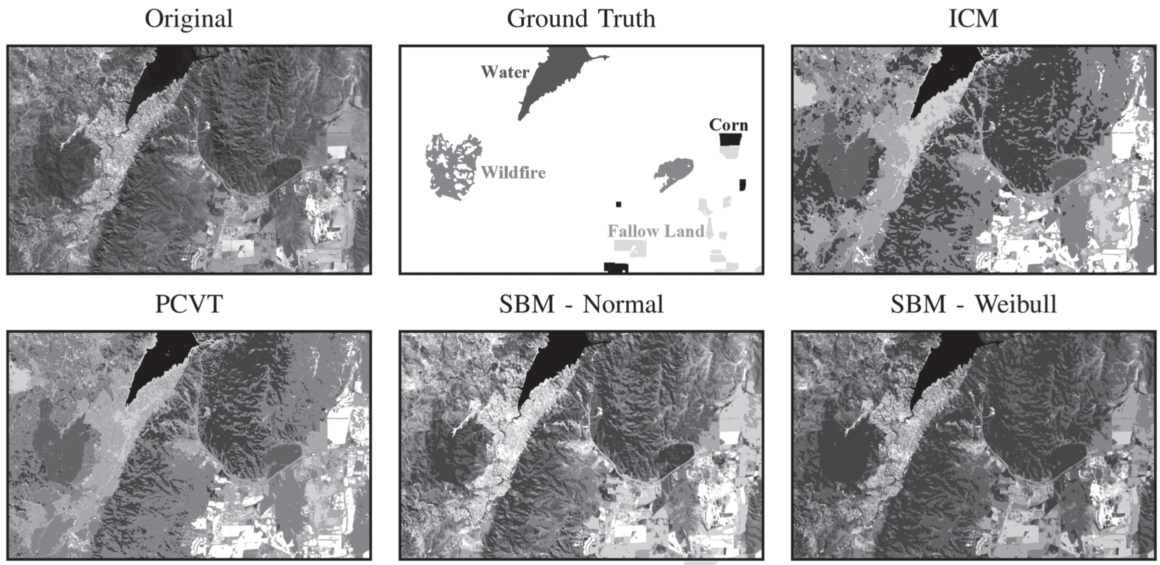

Fig. 2. Segmentations of a Landsat 8 image with seven hidden states. The four ground-truth labels are: wildfire, corn, fallow land, and water.

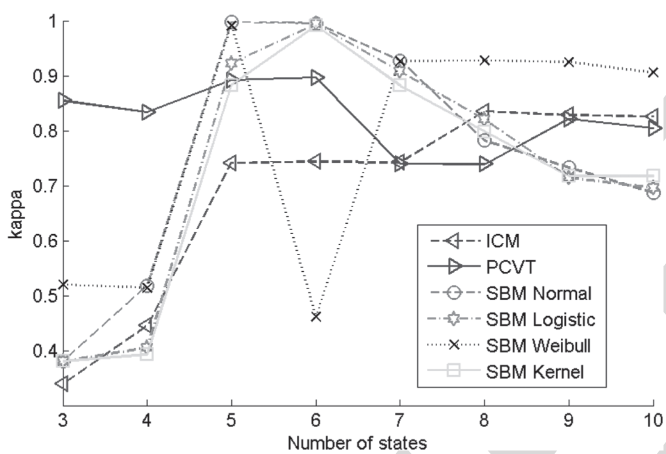

Fig. 3. Landsat 8 image: Comparison of $\widehat{\kappa}$ coefficients for different numbers of hidden states. For five to seven states, SBM clearly outperforms the benchmark algorithms for almost all probability functions. For more than seven states, SBM has the highest $\widehat{\kappa}$ values only when using the Weibull distribution.

bands, SBM was the only algorithm that distinguished shallow and deep water in a satisfactory way. In this experiment, the choice of the probability function had very little influence on the results. In the case of the Landsat 8 image, we found that the Weibull distribution is the best choice for SBM and that SBM tends to be relatively sensitive to the number of hidden states. The fact that the probability function can have great influence on the segmentation results encourages us to keep investigating algorithms that do not depend on a certain probability function.

ACKNOWLEDGMENT

The authors would like to thank CONICET and CONAE.

REFERENCES

[1] K. Ushaan and B. Singhb, "Potential applications of remote sensing in horticulture," *Scientia Horticulturae*, vol. 153, no. 4, pp. 71–83, Apr. 2013.

[2] S. W. Taylor, D. G. Woolford, C. B. Dean, and D. L. Martell, "Wildfire prediction to inform fire management: Statistical science challenges," *Statistical Sci.*, vol. 28, no. 4, pp. 586–615, 2013.

[3] C. Manzo, R. Salvini, E. Guastaldi, V. Nicolardi, and G. Protano, "Reflectance spectral analyses for the assessment of environmental pollution in the geothermal site of Mt. Amiata (Italy)," *Atmospheric Environ.*, vol. 79, pp. 650–665, Nov. 2013.

[4] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. New York, NY, USA: Springer Science & Business Media, 2009.

[5] X.-Y. Fu, H.-J. You, and K. Fu, "Building segmentation from high-resolution SAR images based on improved Markov random field," *Acta Electronica Sinica*, vol. 40, no. 6, pp. 1141–1147, 2012.

[6] J. Jing, Y. Li, P. Li, and Y. Jiao, "Textile printing pattern image segmentation based on algorithm of MRF," *J. Inf. Comput. Sci.*, vol. 10, no. 13, pp. 4007–4015, 2013.

[7] P. Ruiz, J. Mateos, G. Camps-Valls, R. Molina, and A. K. Katsaggelos, "Bayesian active remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2186–2196, Apr. 2014.

[8] A. L. M. Levada, N. D. A. Mascarenhas, and A. Tannús, "A novel MAP-MRF approach for multispectral image contextual classification using combination of suboptimal iterative algorithms," *Pattern Recog. Lett.*, vol. 31, no. 13, pp. 1795–1808, Oct. 2010.

[9] F. Wang *et al.*, "Unsupervised SAR image segmentation using higher order neighborhood-based triplet Markov fields model," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5193–5205, Aug. 2014.

[10] L. Kuncheva, *Combining Pattern Classifiers*. Hoboken, NJ, USA: Wiley, 2004.

[11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Statistical Soc.*, vol. 39, no. 1, pp. 1–38, 1977.

[12] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.

[13] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE Int. Conf. Comput. Vis.*, 1998, pp. 839–846.

[14] H. Peng, R. Rao, and S. A. Dianat, "Multispectral image denoising with optimized vector bilateral filter," *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 264–273, 2014.

[15] J. E. Besag, "On the statistical analysis of dirty pictures," *J. Royal Statistical Soc.*, vol. 48, no. 3, pp. 259–302, 1986.

[16] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960.

[17] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984.

[18] D. Joshi, J. Li, and J. Z. Wang, "A computationally efficient approach to the estimation of two- and three-dimensional hidden Markov models," *IEEE Trans. Image Process.*, vol. 15, no. 7, pp. 1871–1886, Jul. 2006.

[19] J. Gimenez, A. C. Frery, and A. G. Flesia, "Inference strategies for the smoothness parameter in the Potts model," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2013, pp. 2539–2542.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

Please be aware that authors are required to pay overlength page charges ($200 per page) if the paper is longer than 3 pages. If you cannot pay any or all of these charges please let us know.

This pdf contains 2 proofs. The first half is the version that will appear on Xplore. The second half is the version that will appear in print. If you have any figures to print in color, they will be in color in both proofs.

The "Open Access" option for your paper expires when the paper is published on Xplore in an issue with page numbers. Papers in "Early Access" may be changed to Open Access.

AQ1 = Please check if the provided affiliations of all authors are correct. Otherwise, make the necessary corrections.

END OF ALL QUERIES