



Dempster-Shafer Parzen-Rosenblatt Hidden Markov Fields for Multichannel Image Segmentation

Mohamed El Yazid Boudaren^(✉), Ali Hamache, Islam Debicha,
and Hamza Tarik Sadouk

Ecole Militaire Polytechnique, PO Box 17, 16111 Bordj El Bahri, Algiers, Algeria
boudaren@gmail.com

Abstract. Theory of evidence has been successfully used in many areas covering pattern recognition and image processing due to its effectiveness in both information fusion and reasoning under uncertainty. Such notoriety led to extension of many existing Bayesian tools such as hidden Markov models, extensively used for image segmentation. This paper falls under this category of frameworks and aims to propose a new hidden Markov field that better handles nonGaussian forms of noise, designed for multichannel image segmentation. To this end, we use a recent kernel smoothing- based noise density estimation combined with a genuine approach of mass determination from data. The proposed model is validated on sampled and real remote sensing images and the results obtained outperform those produced by conventional hidden Markov fields.

Keywords: Data classification · Dempster-Shafer theory · Hidden Markov Field · Multichannel image segmentation

1 Introduction

Multichannel image analysis and processing have gained more interest among the image and signal processing community following the development of computing technologies [3, 10, 12, 15]. The purpose of multichannel image classification, considered in this paper, is to produce a thematic map indicating the membership of each pixel in a specific class based on two sources: the spectral information and the spatial information. The first is represented by the different image channels. Each channel corresponds to an interval of the electromagnetic spectrum, where a dedicated sensor is used to measure the intensity of the spectrum received over this interval. The use of spectral information for image classification can be very effective especially in the supervised context because one has a knowledge base used at the learning stage. The interest of taking into consideration the second source of information, namely the contextual dependence, was quickly noticed. Image modeling through hidden Markov fields takes into account such dependencies which improves the classification performance [6, 9, 13, 14].

A more elaborated classification model should perform at both levels to produce significantly best class maps. For this purpose, we propose a new hidden Markov model that better handles general forms of noise, typically nonGaussian. More explicitly, we propose to adopt an evidential approach for estimating the noise parameters, which thus allows a better use of spectral information in a Markovian context towards a more effective multichannel image classification.

The remainder of this paper is organized as follows. Section 2 briefly recalls Dempster-Shafer theory, Parzen-Rosenblatt density estimation and hidden Markov fields. Section 3 describes the proposed approaches and related estimation tasks. Experimental results are presented and discussed in Sect. 4. Concluding remarks and future directions are given in Sect. 5.

2 Preliminaries

In this section, we briefly recall some basic notions of Dempster-Shafer theory and Hidden Markov Fields.

2.1 Dempster-Shafer Theory

Data fusion particularly enhances the quality of decision when more than one source of information are available. This is mainly due to the possibility of increasing the amount of relevant information by exploiting redundancy and complementariness among sources. One powerful and flexible mathematical tool that has shown its usefulness in this area is Dempster-Shafer theory (DST) [16], [17] that generalizes the Bayesian frame by allowing on one hand to reap a consensus decision from all information sources; and on the other hand, to handle information uncertainty within each information source. Hence, DST has been applied in many fields [4, 5, 7, 11]. In what follows, we give a quick overview about the DST concepts that will be needed for the sake of this paper.

Let $\Omega = \{\omega_1, \dots, \omega_K\}$, and let $\mathcal{P}(\Omega) = \{A_1, \dots, A_Q\}$ be its power set, with $Q = 2^K$. A function M defined from $\mathcal{P}(\Omega)$ to $[0, 1]$ is called a “basic belief assignment” (*bba*) if $M(\emptyset) = 0$ and $\sum_{A \in \mathcal{P}(\Omega)} M(A) = 1$. A *bba* M defines then a “plausibility” function Pl from $\mathcal{P}(\Omega)$ to $[0, 1]$ by $Pl(A) = \sum_{A \cap B \neq \emptyset} M(B)$, and a “credibility” function Cr from $\mathcal{P}(\Omega)$ to $[0, 1]$ by $Cr(A) = \sum_{B \subset A} M(B)$. Also, both aforementioned functions are linked by $Pl(A) + Cr(A^c) = 1$. Furthermore, a probability function p can be considered as a particular case for which $Pl = Cr = p$.

When two *bbas* M_1 and M_2 describe two pieces of evidence, we can fuse them using the so called “Dempster-Shafer fusion” (DS fusion), which gives $M = M_1 \oplus M_2$ defined by:

$$M(A) = (M_1 \oplus M_2)(A) \propto \sum_{B_1 \cap B_2 = A} M_1(B_1)M_2(B_2) \quad (1)$$

Finally, an evidential *bba* M can be transformed into a probabilistic one using Smets method, according to which each mass of belief $M(A)$ is equally

distributed among all elements of A , leading to the so called “pignistic probability”, Bet , given by:

$$Bet(\omega_i) = \sum_{\omega_i \in A \subseteq \Omega} \frac{M(A)}{|A|} \quad (2)$$

where $|A|$ is the number of elements of Ω in A .

2.2 Parzen-Rosenblatt Dempster-Shafer Classifier

In this subsection, we briefly recall the Parzen-Rosenblatt Dempster-Shafer (PRDS) Classifier proposed in [8]. To this end, let us assume we have a sample of N prelabeled multiattribute data (Z_1, \dots, Z_N) where each datum $Z_n = (X_n, Y_n)$ with $X_n \in \Omega = \{\omega_1, \dots, \omega_K\}$ being the label, and $Y_n = (Y_n^1, \dots, Y_n^P) \in \mathbb{R}^P$ being the P -attribute observation. The problem is then to estimate the label of any new observation $Y_{n'}$ that is optimal with respect to some criterion.

In what follows, we recall the training and classification procedures. According to the PRDS scheme, training consists in estimating for each class $\omega_k \in \Omega$ and for each attribute p ($1 < p < P$), the associated Parzen-Rosenblatt density \hat{f}_k^p . For further weighting sake, 5-fold cross-validation classification is achieved based on each attribute (taken alone) using the above Parzen-Rosenblatt PDFs according to maximum likelihood.

For a given new observation $Y_{n'}$, partial report about the identity of $X_{n'}$ can be made at each individual attribute level through a mass function M^p , on $P(\Omega)$, generated based on the Parzen-Rosenblatt PDF estimated at the training stage. Such reports are then combined, typically using DS- fusion to reap a consensus report M . Final decision is then be deduced through the Pignistic transform applied to M . In the following, we describe our approach step by step. For more details, the reader may refer to [8].

2.3 Hidden Markov Fields

Let S be a finite set, with $Card(S) = N$, and let $(Y_s)_{s \in S}$ and $(X_s)_{s \in S}$ be two collections of random variables, which will be called “random fields”. We assume that Y is observable with each Y_s taking its values in \mathbb{R} (or \mathbb{R}^m) whereas X is hidden with each X_s taking its values from a finite set of “classes” or “labels”. Such situation occurs in image segmentation problem, which will be used in this paper as illustrative frame. Realizations of such random fields will be denoted using lowercase letters. We deal with the problem of the estimation of $X = x$ from $Y = y$. Such estimation subsumes the distribution of (X, Y) to be beforehand defined.

In hidden Markov fields (HMFs) context, the field X is assumed Markovian with respect to a neighborhood system $\mathfrak{N} = (\mathfrak{N}_s)_{s \in S}$. X is then called a Markov random field (MRF) defined by

$$p(X_s = x_s | (X_t)_{t \in S, t \neq s}) = p(X_s = x_s | (X_t)_{t \in \mathfrak{N}_s}) \quad (3)$$

Under some conditions usually assumed in digital imagery, the Hammersley-Clifford theorem [2] establishes the equivalence between an MRF, defined with respect to the neighbourhood system \mathfrak{N} , and a Gibbs field with potentials associated with \mathfrak{N} . Such potentials, describing the elementary relationships within the neighbourhood, are computed with respect to the system of cliques C , where a clique $c \in C$ is a subset of S which is either a singleton or a set of pixels mutually neighbors with respect to \mathfrak{N} . Setting $x_c = (X_s)_{s \in c}$, $\phi_c(x_c)$ denotes the potential associated to the clique c .

Finally, the distribution of X is given by

$$p(X = x) = \gamma \exp \left[- \sum_{c \in C} \phi_c(x_c) \right] \quad (4)$$

where γ is a normalizing constant which is impossible to compute in practice given the very high number of possible configurations K^N . The quantity $E(x) = \sum_{c \in C} \phi_c(x_c)$ is called “energy” and can also be expressed locally through $E_s(x_s) = \sum_{c \ni x_s} \phi_c(x_c)$. Hence, the local conditional probability of (3) becomes

$$p(X_s = x_s | (X_t)_{t \in S, t \neq s}) = \gamma_s \exp [-E_s(x_s)]$$

where γ_s is a computable normalizing constant.

To define the distribution of Y conditional on X , two assumptions are usually set:

- (i) the random variables $(Y_s)_{s \in S}$ are independent conditional on X ;
- (ii) the distribution of each Y_s conditional on X is equal to its distribution conditional on X_s .

When these two assumptions hold, the noise distribution is fully defined through K distributions $(f_i)_{1 \leq i \leq K}$ on \mathbb{R} where f_i denotes the density, with respect to the Lebesgue measure on \mathbb{R} , of the distribution of Y_s conditional on $X_s = \omega_i$: $p(Y_s = y_s | x_s = \omega_i) = f_i(y_s)$. Then we have

$$p(Y = y | X = x) = \prod_{s \in S} f_{x_s}(y_s) \quad (5)$$

that can equivalently be written as

$$p(Y = y | X = x) = \exp \left[\sum_{s \in S} \log f_{x_s}(y_s) \right] \quad (6)$$

Since $p(x, y) = p(x)p(y|x)$, we obtain

$$p(X = x, Y = y) = \gamma \exp - \left[\sum_{c \in C} \phi_c(x_c) - \sum_{s \in S} \log f_{x_s}(y_s) \right] \quad (7)$$

Hence, according to (7), the couple (X, Y) is a Markov field and also is the distribution of X conditional on $Y = y$. This allows to sample a realization of

X according to its posterior distribution $p(x|y)$ and hence, to apply Bayesian techniques like maximum posterior marginal (MPM) and maximum a posteriori (MAP).

The feasibility of the different estimations of interest in HMFs stems from the possibility of sampling realizations of the hidden process X from $Y = y$ according to the posterior distribution $p(x|y)$, and which is possible when this latter distribution is of Markov form. On the other hand, the Markovianity of this latter distribution relies itself on the assumption that the random variables $(Y_s)_{s \in S}$ are independent conditionally on X .

3 Proposed Approach

The problem considered in this paper is to derive a thematic map from a multichannel (typically remote sensing) image. As described in the previous section, the hidden Markov field (HMF) model allows to find a hidden field X representing in this case the thematic map, from an observed field Y representing the observed multichannel image. The novelty in this paper is to adopt the Pignistic probabilities provided by PRDS classifier [8] (after combination of different channel reports) instead of $f_{x_s}(y_s)$ in Eq. (5). Thus, one need to achieve a training process on the prelabeled set of pixels (typically a prelabeled image of subimage) to derive noise densities associated to different channels which will later produce parameters of spectral information. Such prelabeled data are not available however, given that we deal with unsupervised classification. Then, the estimation of both spectral and spatial parameters is achieved in an unsupervised iterative way. More explicitly, one starts by coarsely perform an initial clustering which will service as a basis for initial parameters estimation. Indeed, when an initial realization of X is available, one can perform training according to PRDS to derive noise densities and, at the same time, estimation of spatial parameters as in conventional HMF context.

In what follows, we describe PRDS training, parameter estimation and labeling procedures.

3.1 Training

As specified before, let us consider a set of prelabeled multichannel pixels. We recall that such data may be available through an initial coarse clustering providing a realization of X and then iteratively through successive updates of X during parameter estimation, as we are going to see later. Hence, training will be concerned exclusively with spectral information. More explicitly, we use PR-DS classifier to estimating for each class $\omega_k \in \Omega$ and for each channel p ($1 < p < P$), the Parzen-Rosenblatt density \hat{f}_k^p as described in the previous section. Let us now show how the estimated Parzen-Rosenblatt densities will produce spectral parameters. In other words, we demonstrate how one can replace the noise densities $f_{x_s}(y_s)$ in Eq. (5) for a given $Y_n \in \mathbb{R}^P$.

Step 1: Generation of Mass Functions. To define the mass associated to channel p , let us consider the rank function δ_p defined from $\{1, \dots, K\}$ to Ω such as $\delta_p(k)$ is the k -ranked element of Ω in terms of \hat{f}^p , i.e. $\hat{f}_{\delta_p(1)}^p(Y_n^p) \leq \hat{f}_{\delta_p(2)}^p(Y_n^p) \leq \dots \leq \hat{f}_{\delta_p(K)}^p(Y_n^p)$. Then, M^p is derived as follows:

$$\begin{cases} M^p(\Omega) \propto K \hat{f}_{\delta_p(1)}^p(Y_n^p) \\ M^p(\{\omega_{\delta_p(k)}, \dots, \omega_{\delta_p(K)}\}) \propto (K - k + 1) [\hat{f}_{\delta_p(k)}^p(Y_n^p) - \hat{f}_{\delta_p(k-1)}^p(Y_n^p)], \text{ for } k > 1 \end{cases} \quad (8)$$

Step 2: Combination of Mass Functions. Mass functions associated to different attributes are then combined into one collaborative mass $M = \bigoplus_{p=1}^P M^p$:

$$M(B) \propto \sum_{\bigcap_{p=1}^P B_p = B} \left[\prod_{p=1}^P M^p(B_p) \right], \text{ for } B, B_p \in \mathcal{P}(\Omega) \quad (9)$$

Step 3: Deriving Noise Density. Based on M , the noise density is then computed according to the Pignistic transform:

$$f_k(y_n) = \sum_{A \ni \omega_k} \frac{M(A)}{|A|} \quad (10)$$

3.2 Parameter Estimation

In this framework, we adopted ICE algorithm for parameter estimation. At each iteration i , a realization of X is simulated using Gibbs sampler. Then, we use the Derin and Elliott method for estimating spatial parameters ϕ_i ; and the PRDS method for spectral parameters η_i (which are noise densities as described in Step 3 above). The algorithm stops when an end criterion is reached. The parameter estimation procedure is illustrated through Fig. 1.

3.3 Labeling

Once parameter set θ is estimated by the ICE method [1] while the MPM estimator is used to infer X . Using the Gibbs sampler, T realization x^1, x^2, \dots, x^T of X are simulated according to $p(X|Y = y)$. Then, one estimates $\hat{p}(x_s = \omega | y)$ of each X_s from the realizations x^1, x^2, \dots, x^N . Finally, for each pixel x_s , one chooses the class whose number of appearances in the simulations is the highest.

4 Evaluation of the Proposed Approach

To validate our approach, we assess its performance in unsupervised segmentation of multichannel images against the conventional HMF model. To this end, we consider two series of experiments. The first series deal with synthetic images whereas the second series deals with a real multichannel remote sensing image.

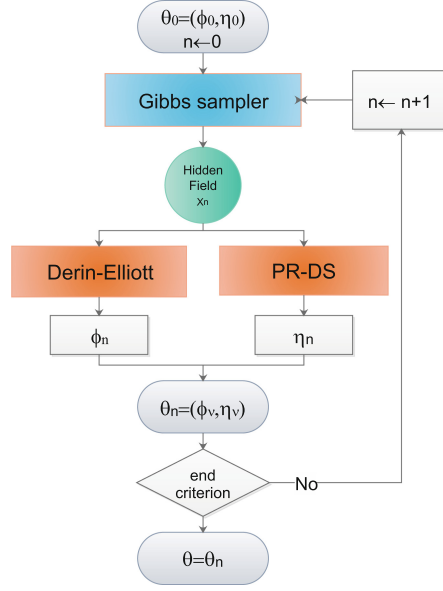


Fig. 1. Parameter estimation process.

4.1 Unsupervised Segmentation of Sampled Multichannel Images

To sample a synthetic multichannel image, we use Gibbs sampler with the following parameters: $\beta = [1, 1, 1]$, $\alpha_h = \alpha_{vb} = 2I$ and $\alpha_d = 4I$. Then, the obtained image is noised considering multidimensional Gaussian mixture densities in two different ways, considering two different sets of noise parameters η :

- Image 1: noisy version of the reference image, with a two-dimensional mixture noise of four Gaussians and a variance-covariance matrix $\Sigma = \begin{bmatrix} 28 & 3 \\ 3 & 28 \end{bmatrix}$.
 - * For the first class $\mu_{11} = [10 \ 20]$, $\mu_{12} = [25 \ 20]$, $\mu_{13} = [40 \ 20]$, $\mu_{14} = [50 \ 20]$, $\Sigma_{11} = \Sigma_{12} = \Sigma_{13} = \Sigma_{14} = \Sigma$ and a proportion of mixture $p_1 = [0.25 \ 0.25 \ 0.25 \ 0.25]$;
 - * For the second class $\mu_{21} = [10 \ 30]$, $\mu_{22} = [25 \ 30]$, $\mu_{23} = [40 \ 30]$, $\mu_{24} = [50 \ 30]$, $\Sigma_{21} = \Sigma_{22} = \Sigma_{23} = \Sigma_{24} = \Sigma$ and a proportion of mixture $p_2 = [0.25 \ 0.25 \ 0.25 \ 0.25]$;
 - * For the third class $\mu_{31} = [10 \ 40]$, $\mu_{32} = [25 \ 40]$, $\mu_{33} = [40 \ 40]$, $\mu_{34} = [50 \ 40]$, $\Sigma_{31} = \Sigma_{32} = \Sigma_{33} = \Sigma_{34} = \Sigma$ and a proportion of mixture $p_3 = [0.25 \ 0.25 \ 0.25 \ 0.25]$;
- Image 2: noisy image with a three-dimensional mixture noise of five Gaussians and a variance-covariance matrix for each element of the mixture for each class c : $\Sigma_{c1} = \Sigma_{c2} = \Sigma_{c3} = \Sigma_{c4} = \Sigma_{c5} = \Sigma$ such as $c \in \{1, 2, 3\}$ and

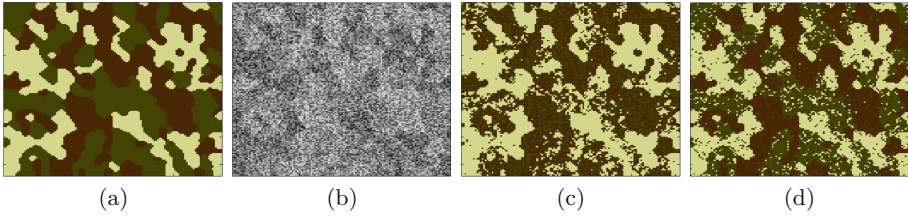


Fig. 2. Classification of synthetic data using HMF and PRDS-HMF. (a) Original class image. (b) Noisy image. (c) Classification of the image using HMF, success rate $\tau = 58.3740\%$. (d) Classification of the image using PRDS-HMF, success rate $\tau = 84.2285\%$.

$$\Sigma = \begin{bmatrix} 28 & 3 & 1 \\ 3 & 28 & 3 \\ 1 & 3 & 28 \end{bmatrix}.$$

* For the first class $\mu_{11} = [10 \ 20 \ 20]$, $\mu_{12} = [25 \ 20 \ 20]$, $\mu_{13} = [40 \ 20 \ 20]$, $\mu_{14} = [50 \ 20 \ 20]$, $\mu_{15} = [65 \ 20 \ 20]$ and a proportion of mixture $p_1 = [0.2 \ 0.2 \ 0.2 \ 0.2 \ 0.2]$;

* For the second class $\mu_{21} = [10 \ 30 \ 25]$, $\mu_{22} = [25 \ 30 \ 25]$, $\mu_{23} = [40 \ 30 \ 25]$, $\mu_{24} = [50 \ 30 \ 25]$, $\mu_{25} = [65 \ 30 \ 25]$ and a proportion of mixture $p_2 = [0.2 \ 0.2 \ 0.2 \ 0.2 \ 0.2]$;

* For the third class $\mu_{31} = [10 \ 40 \ 30]$, $\mu_{32} = [25 \ 40 \ 30]$, $\mu_{33} = [40 \ 40 \ 30]$, $\mu_{34} = [50 \ 40 \ 30]$, $\mu_{35} = [65 \ 40 \ 30]$ and a proportion of mixture $p_3 = [0.2 \ 0.2 \ 0.2 \ 0.2 \ 0.2]$;

Then, unsupervised segmentation is performed using conventional HMFs; and the proposed PRDS- HMF. The results obtained are illustrated in Fig. 2 and Fig. 3 where the noisy multichannel images are depicted in monochannel gray level by averaging the channels' intensities for illustrative purpose. The segmentation accuracy rates obtained confirm the interest of the proposed model with respect to the classic HMF. The supremacy of the proposed model is mainly due to the possibility of considering more general forms of noise by the PRDS-

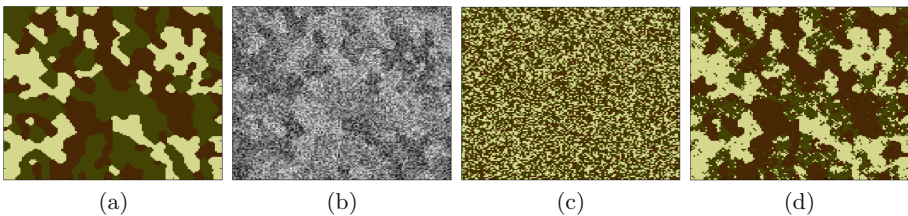


Fig. 3. Classification of synthetic data using HMF and PRDS-HM. (a) Original class image. (b) Noisy image. (c) Classification of the image using HMF, success rate $\tau = 32.7332\%$. (d) Classification of the image using PRDS-HMF, success rate $\tau = 80.8350\%$.

HMF thanks to the kernel smoothing technique that makes it possible to fit any form of noise instead of assuming it Gaussian as in HMF context.

4.2 Unsupervised Segmentation of Multichannel Remote Sensing Image

In this series of experiments, we consider a multichannel image of the Landsat-7 satellite ETM+ sensor acquired on June 03, 2001. The acquisition was made in an area around the city of Algiers. The image used has a resolution of $30\text{ m} \times 30\text{ m}$ and a size of 256×256 . The area of study includes 4 classes: (i) Urban Dense (UD); (ii) Urban Less Dense (ULD); (iii) Barren Land (BL); and (iv) Vegetation (V).

Figure 4 represents the 6 bands at the gray scale image.

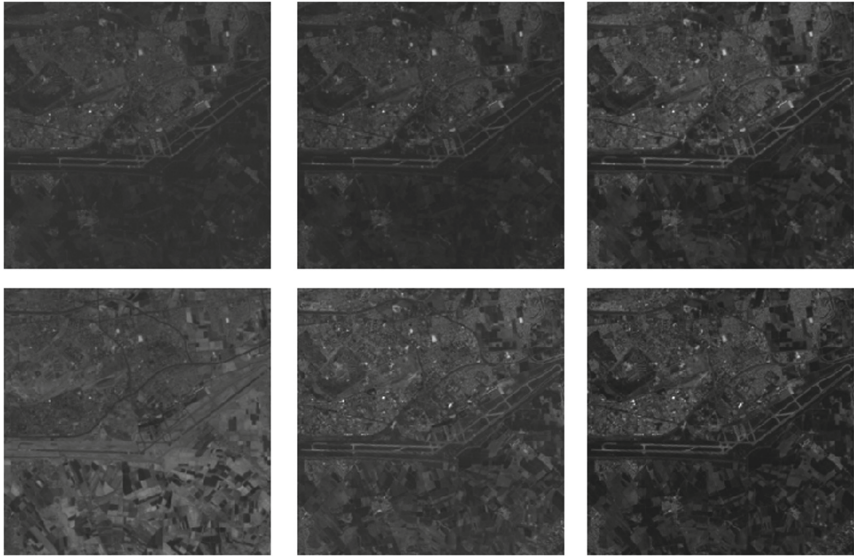


Fig. 4. Different channel observations of the studied image.

To quantitatively assess the performance of the proposed approach against the conventional HMF model, we have a partial ground truth (see Fig. 4).

Qualitative assessment of the results obtained shows that the thematic map provided by the proposed approach contains less salt and pepper effect. This is confirmed by the quantitative assessment in terms of overall accuracy and kappa metrics. Indeed, the PRDS- HMF yields an accuracy rate of 79% (resp. a kappa of 0.70) against an accuracy of 68% (resp. a kappa of 0.5) by the conventional HMF model (Figs. 5 and 6).

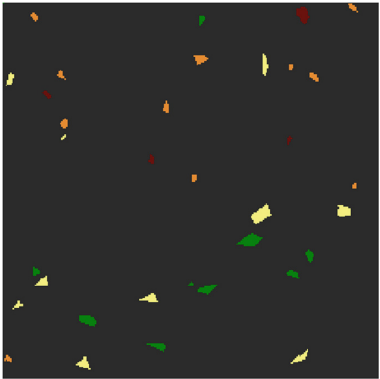


Fig. 5. Partial ground truth of the studied area

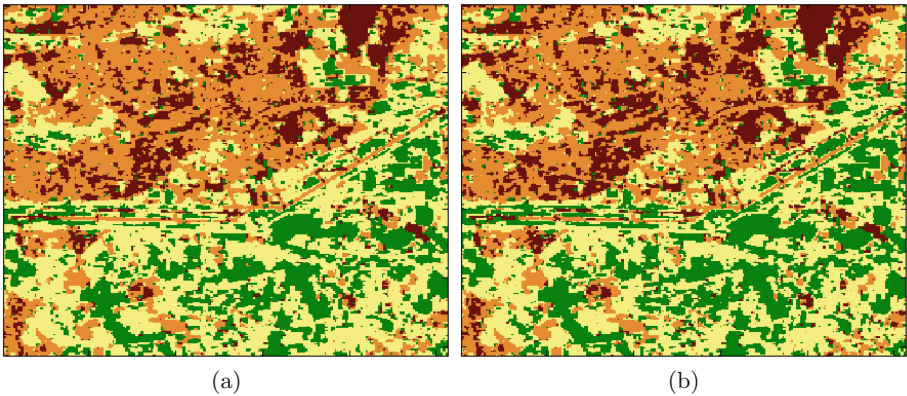


Fig. 6. Classification of a multichannel image using HMF and PRDS-HMF: Urban Dense (red), Urban Less Dense (orange), Barren Land (Yellow) and Vegetation (Green). (a) Thematic map obtained by PRDS-HMF: accuracy= 79.42%, Kappa= 0.70 (b) Thematic map obtained by HMF: accuracy= 68.23%, Kappa= 0.55. (Color figure online)

Table 1. Confusion matrix obtained by PRDS-HMF.

	(UD)	(ULD)	(BL)	(V)	Truth	Recall
(UD)	41	0	0	1	42	97.62%
(ULD)	19	84	7	1	111	75.68%
(BL)	0	2	192	31	225	85.33%
(V)	4	22	27	123	176	69.89%
Classification	64	108	226	156	554	
Precision	64.06%	77.78%	84.96%	78.85%		

Table 2. Confusion matrix obtained by HMF.

	(UD)	(ULD)	(BL)	(V)	Truth	Recall
(UD)	41	0	0	1	42	97.62%
(ULD)	43	60	7	1	111	54.05%
(BL)	1	1	164	59	225	72.89%
(V)	7	16	40	113	176	64.20%
Classification	92	77	211	174	554	
Precision	44.57%	77.92%	77.73%	64.94%		

Confusing matrices obtained by PRDS-HMF and HMF classifications are also given in Tables 1 and 2. We can confirm that the proposed HMF outperforms the plain one in terms of both precision and recall per each class. It is worth mentioning that a better modeling of noise allows also to a better estimation of spatial parameters. In fact, parameter estimation is an iterative process in which a good perception of noise leads to a better parameter estimation of spatial features.

5 Conclusion

In this paper, we proposed a new hidden Markov field model designed for unsupervised segmentation of multichannel images. The main novelty of the proposed model relies in the use of Dempster-Shafer theory and Parzen-Rosenblatt window for noise density estimation which makes it possible to model general forms of multidimensional noise. To assess the performance of the proposed PRDS-HMF, experiments were conducted on both synthetic and real multichannel images. The results obtained confirmed its interest with respect to the conventional HMF model. A possible future direction of this approach would be to consider more general Markov models with the same extension.

References

1. Benboudjema, D., Pieczynski, W.: Unsupervised image segmentation using triplet Markov fields. *Comput. Vis. Image Underst.* **99**(3), 476–498 (2005)
2. Besag, J.: Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Stat. Soc. Ser. B* **6**, 192–236 (1974)
3. Borges, J.S., Bioucas-Dias, J.M., Marcal, A.R.: Bayesian hyperspectral image segmentation with discriminative class learning. *IEEE Trans. Geosci. Remote Sens.* **49**(6), 2151–2164 (2011)
4. Boudaren, M.E.Y., An, L., Pieczynski, W.: Dempster-Shafer fusion of evidential pairwise Markov fields. *Int. J. Approximate Reason.* **74**, 13–29 (2016)
5. Deneux, T.: 40 years of Dempster-Shafer theory. *Int. J. Approximate Reason.* **79**, 1–6 (2016)

6. Ghamisi, P., Benediktsson, J.A., Ulfarsson, M.O.: Spectral-spatial classification of hyperspectral images based on hidden Markov random fields. *IEEE Trans. Geosci. Remote Sens.* **52**(5), 2565–2574 (2014)
7. Guo, H., Shi, W., Deng, Y.: Evaluating sensor reliability in classification problems based on evidence theory. *IEEE Trans. Syst. Man Cybern. Part B (Cybern.)* **36**(5), 970–981 (2006)
8. Hamache, A., et al.: Uncertainty-aware Parzen-Rosenblatt classifier for multi-attribute data. In: Destercke, S., Denoeux, T., Cuzzolin, F., Martin, A. (eds.) *BELIEF 2018. LNCS (LNAI)*, vol. 11069, pp. 103–111. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99383-6_14
9. Jackson, Q., Landgrebe, D.A.: Adaptive Bayesian contextual classification based on Markov random fields. *IEEE Trans. Geosci. Remote Sens.* **40**(11), 2454–2463 (2002)
10. Jia, X., Richards, J.A.: Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification. *IEEE Trans. Geosci. Remote Sens.* **37**(1), 538–542 (1999)
11. Jones, R.W., Lowe, A., Harrison, M.J.: A framework for intelligent medical diagnosis using the theory of evidence. *Knowl.-Based Syst.* **15**(1), 77–84 (2002)
12. Li, J., Bioucas-Dias, J.M., Plaza, A.: Hyperspectral image segmentation using a new Bayesian approach with active learning. *IEEE Trans. Geosci. Remote Sens.* **49**(10), 3947–3960 (2011)
13. Li, J., Bioucas-Dias, J.M., Plaza, A.: Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields. *IEEE Trans. Geosci. Remote Sens.* **50**(3), 809–823 (2012)
14. Mercier, G., Derrode, S., Lennon, M.: Hyperspectral image segmentation with Markov chain model. In: 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No. 03CH37477), IGARSS 2003, vol. 6, pp. 3766–3768. IEEE (2003)
15. Plaza, A., et al.: Recent advances in techniques for hyperspectral image processing. *Remote Sens. Environ.* **113**, S110–S122 (2009)
16. Shafer, G.: *A Mathematical Theory of Evidence*, vol. 1. Princeton University Press, Princeton (1976)
17. Shafer, G.: A mathematical theory of evidence turns 40. *Int. J. Approximate Reason.* **79**, 7–25 (2016)