

On the Influence of Markovian Models for Contextual-Based Optimum-Path Forest Classification

D. Osaku¹, A.L.M. Levada¹, and J.P. Papa²

¹ Federal University of São Carlos – UFSCar
Department of Computer Science, São Carlos, Brazil
danosaku@hotmail.com, alexandre@dc.ufscar.br

² São Paulo State University – UNESP
Department of Computing, Bauru, Brazil
papa@fc.unesp.br

Abstract. Contextual classification considers the information about a sample’s neighborhood to improve standard pixel-based classification approaches. In this work, we evaluated four different Markovian models for Optimum-Path Forest contextual classification considering land use recognition in remote sensing data. Some insights about the situations in which each of them should be applied are stated, as well as the idea behind them is explained.

Keywords: Optimum-Path Forest, Contextual Classification, Markov Random Fields.

1 Introduction

In pixel-based classification standards, we often classify samples according to their attributes (feature vector), but no information about the relationship between them is employed to the decision-making process. Besides, researchers have reported that such approaches frequently have salt-and-pepper and other undesirable artifacts associated to the classification results [2,16]. The main shortcoming is that we normally assume the samples are identically and independent distributed, such that their labels are not influenced by the neighborhood.

Several studies have proposed to employ contextual classification in order to reduce such effects, as well as to improve the classification results. The easiest way to perform contextual classification is to apply a mode filter over the label map generated by a pixel-based classifier as a post-processing step, for instance. A more robust manner to integrate contextual information is to use stochastic models based on the Markov Random Fields (MRF) theory, being a suitable representation of the prior knowledge by means of probability distributions. One of the most used Markov formulations is the well-known Potts Model [4], which can be defined by local conditional density functions computed over a finite neighborhood. In addition, several extensions of it have been proposed considering

a continuous distribution of the neighbouring pixels, such as the so-called Generalized Isotropic Multi-Level Logistic (GIMLL) [6], which incorporates pixel similarity in a softer way than Isotropic Multi-Level Logistic (IMLL) [8].

Later on, Nakamura et al. [9] proposed a contextual version of the Optimum-Path Forest [12,11] (OPF) classifier in order to improve naïve OPF results over grey matter and white matter classification in magnetic resonance images of the human brain. In that work, the authors have added contextual information extending the feature vector of each sample with a local energy function computed by Potts model to design the OPF-MRF. In the same direction, Osaku et al. [10] proposed an evolutionary-based framework to optimize a parameter that controls the amount of contextual information used by Potts model during the OPF-MRF classification process.

The OPF classifier has been widely used for several applications, and it has obtained very promising results, being comparable to the ones achieved by SVM, but much faster for training. Since OPF is a recently developed technique, there are just few works that address its formulation in contextual-based classification scenarios. As such, this paper proposes to extend the works of Nakamura et al. [9] and Osaku et al. [10] by providing a comparison between different MRF models for contextual classification, such as the Potts model, GIMLL with L1 (GIMLL-L1) and L2 (GIMLL-L2) norms, as well as the GMRF model. Therefore, we aim to provide a better understanding regarding MRF models applied for classification purposes, being the results validated over remote sensing image classification. The remainder of this paper is organized as follows: Section 2 presents a brief review about MRF models and the OPF-MRF formulation. Section 3 states the methodology employed in this work, and Section 4 presents the experiments. Finally, Section 5 discusses the conclusions and future works.

2 Theoretical Background

2.1 Markov Random Fields

Markov Random Fields are probabilistic models that can be used to integrate both spatial and contextual information in image classification problems. Such models are based on the idea that a pixel has a high probability to belong to the same class of its neighbors.

Let Ω be a bidimensional lattice that defines a set of image's pixels, \mathcal{M} be a Markov random field, and l_{ij} be the value of a possible occurrence of the random field \mathcal{L} , such that $l_{ij} \in \mathcal{L}$. We can also define a neighborhood system η_{ij} centred at pixel (i, j) as being the set of elements in which the distance from this central element is less than some $\sigma > 0$. This notion of neighborhood system is very important for MRF-based approaches, since a Markov random field defined over a lattice can be seen as a set of random variables such that the probability of one variable, constrained to all variables that fall in its field, is the same as the probability of the same variable restricted to a small subset of elements.

We have that $\mathcal{L} = \{l_{ij} | (i, j) \in \Omega\}$ is a MRF defined over Ω if \mathcal{L} holds the following properties for a given conditional probability p with respect to a certain pixel, as well as for the probability P for an image classification task:

- $p(l_{i,j} | \{l_{k,z}, (k, z) \in \Omega \setminus (i, j)\}) = p(l_{i,j} | \{l_{k,z}, (k, z) \in \eta_{ij}\}), \forall (i, j) \in \Omega$; and
 - $P(\mathcal{L}) > 0, \forall l \in \xi$, where ξ stands for the set of all possible realizations of \mathcal{L} .
- Therefore, this statement means all realizations of a given random field can be observed.

In other words, \mathcal{L} stands for the set of all possible labels for an image classification task.

One of the most widely used Markov models is the well-known Potts Model [4], which arose from the statistic physics to generalize the Ising model for multiple discrete energies [14,15]. In the context of image processing and pattern recognition, such model has been widely employed as a prior knowledge to hold the smoothness assumption, since the pixels that fall in the same neighborhood are likely to have the same label.

Given a neighborhood system η_{ij} , we can define the local conditional probability of Potts Model as follows:

$$p(l_{ij} = m | \eta_{ij}, \beta) = \frac{\exp\{\beta \mathcal{U}_{ij}(m)\}}{\sum_{k=1}^K \exp\{\beta \mathcal{U}_{ij}(k)\}}, \tag{1}$$

in which $\mathcal{U}_{ij}(k)$ stands for the number of pixels in η_{ij} that are from label k , K denotes the number of labels, β is a parameter that models the spatial dependence among neighbouring pixels, and $m \in \mathcal{M}$ means the observed label at central pixel (i, j) . It is important to shed light over that high values of β lead to a high spatial dependence among pixels, and $\beta = 0$ means we have no spatial dependence, i.e., we have a traditional classification.

The Generalized Isotropic Multi-Level Logistic is an extension of the Potts model, being defined over a continuous distribution of a given pixel's neighborhood. The idea is just to modify \mathcal{U}_{ij} from Equation 1 as follows:

$$\mathcal{U}_{ij}(m) = \sum_{t \in \eta_{ij}} [1 - 2e^{-(m-t)^2}]. \tag{2}$$

The reader can observe that the contribution of this energy function is given by the L2 norm. Thus, it is common to call this model as GIMLL-L2. We can still refer to another GIMLL version, which employs now a L1 norm, which is called GIMLL-L1:

$$\mathcal{U}_{ij}(m) = \sum_{t \in \eta_{ij}} [1 - 2e^{-|m-t|}]. \tag{3}$$

Finally, another model that has been widely used is the Gaussian MRF, whose local conditional probabilities are given by [6]:

$$p(l_{ij} = m | x_{ij}, \eta_{ij}, \beta, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2} \left[x_{ij} - \mu - \sum_{t \in \eta_{ij}} \beta(t - \mu) \right]^2}, \quad (4)$$

in which μ and σ^2 stand for the mean and variance estimated from the image label map, respectively, and x_{ij} stands for the sample at (i, j) position in the image.

2.2 Optimum-Path Forest-Based Pattern Classification

The OPF classifier models the problem of pattern recognition as a graph partition task into optimum-path trees (OPTs), which represent clusters in an unsupervised fashion, as well as classes in a supervised version. The partition process is ruled through a competition process between some key samples (prototypes), which are previously chosen [12,11]. According to a path-cost function defined by user, such prototypes compete among themselves in order to conquer the remaining samples, and thus generating the OPTs. This procedure is addressed in the training step, being the test phase in charge of adding each test sample to the generated optimum-path forest (collection of OPTS), and thus verifying the training sample that has conquered it using the same above idea. The label of the winning training sample is then assigned to that test node.

The works of Nakamura et al. [9] and Osaku et al. [10] proposed the OPF-MRF, which exploits the contextual information by means of a lattice-based neighborhood model. The proposed contextual-based OPF learning algorithm can be divided in three phases: (i) a pixel-wise classification is performed by traditional OPF to generate the initial label map (classified image), (ii) further, the local probability given by Markov model (Equation 1) is maximized over the previous label map using the well-known Iterated Conditional Modes (ICM) algorithm [1], and thus generating a new label map, which is finally (iii) classified using standard OPF again in order to generate an updated label map. The process is iterated from step (ii) to (iii) until a convergence criterion is reached. Basically, the idea of OPF-MRF is to iteratively maximize a local probability density function followed by the minimization of the path-cost function for all dataset samples.

3 Methodology

In this section, we described the methodology used to assess the OPF-MRF effectiveness using four different Markov models over the task of land use image classification in remote sensing data. We employed one image obtained from CBERS-2B satellite covering the area of Itatinga, SP-Brazil, and another image obtained from Ikonos-2 MS covering the area of Duque de Caxias, RJ-Brazil. Figure 1 displays these images, as well as their respective ground truth versions.

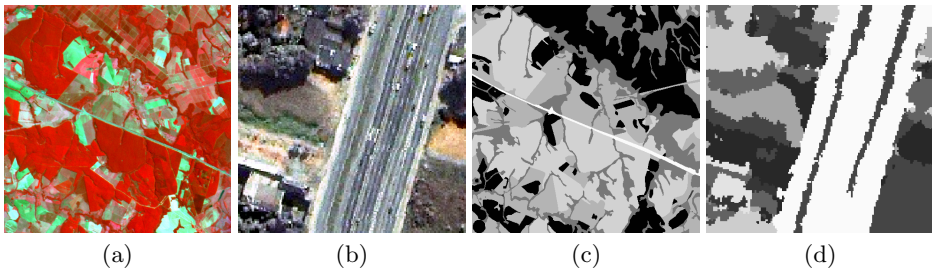


Fig. 1. Satellite images used in the experiments: covering the area of Itatinga, SP - Brazil by (a) CBERS-2B CCD (20m) sensor (R2G3B4) and (b) Ikonos-2 MS sensor (R4G3B2). The CBERS-2B and Ikonos-2 MS images have 526×492 and 258×250 pixels, respectively. Notice that Ikonos-2 MS image were obtained through a fusion process between the corresponding images from MS (4m) and PAN (1m) sensors using the pan-sharpening method. The final image has a spatial resolution of 1m. Labeled images used in the experiments: (c) and (d) refer to the images displayed in (a) and (b), respectively.

The experiments have been conducted as follows: for each satellite image, we have employed 5% of the samples (pixels) to compose the training set, and the remaining 95% samples to compose the test set¹. In addition, we estimated the β parameter (Equation 1) using an exhaustive search (BF - Brute Force) within the range $[0, \beta_{max}]$ with steps of 0.1, in which $\beta_{max} = \ln(1 + \sqrt{K})$ [15] (*critical value*). This approach has been employed to simulate an “optimal” reference (baseline) for the recognition rate using different β values. Thus, the β value that maximized the recognition rate over the test set has been used. This strategy was applied for Potts, GIMLL-L1, GIMLL-L2 and GMRF Markov models. As we have used 10 iterations for OPF-MRF, we computed the mean accuracy for each of them over a cross-validation procedure with 5 runnings using the β value found by BF.

4 Experiments

In this section, we discussed the experimental results using the methodology described in Section 3. In order to evaluate the behaviour of OPF-MRF under different Markov models, we compared four models in the context of land use image classification using two satellites with very different spatial resolutions. Figure 2 displays the classified images using OPF and OPF-MRF classifiers.

Clearly, it is possible to observe that OPF-MRF has been more accurate than naïve OPF for both images, since the latter results present salt-and-pepper effects in some regions of the images (bottom left and middle regions in CBERS-2B image, and up left and bottom left in Ikonos-2 image). Such regions have

¹ Notice that such values have been empirically set. In regard to OPF implementation, we have used the LibOPF [13] package.

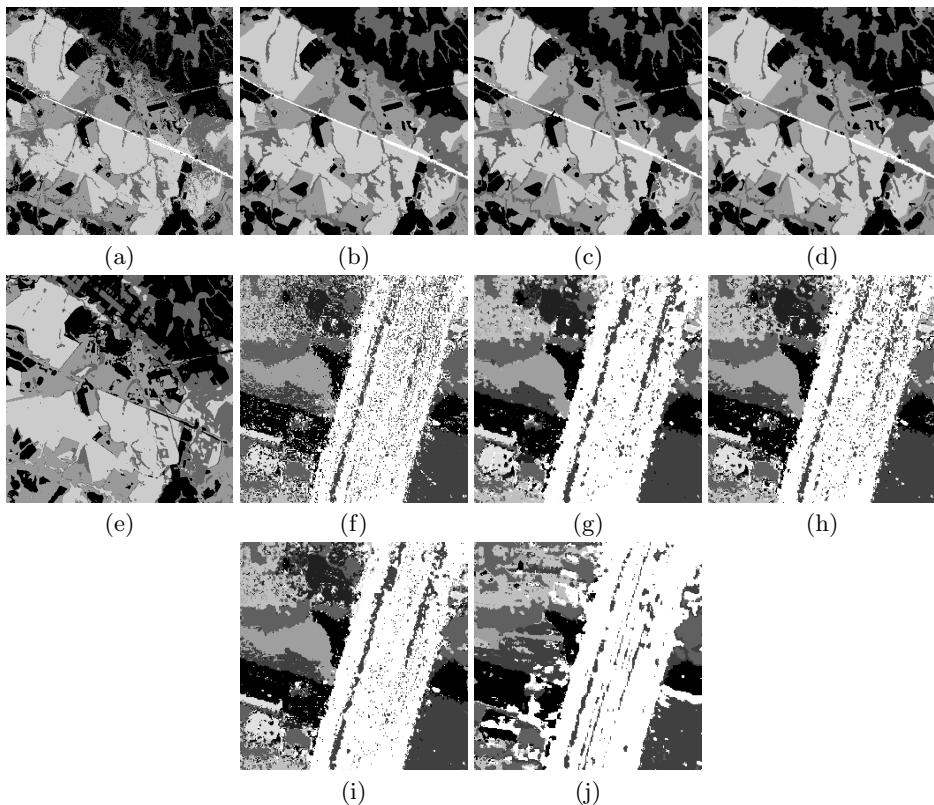


Fig. 2. CBERS-2B classified image considering traditional OPF in (a), and OPF-MRF with GIMLL-L1 (b), GIMLL-L2 (c), GMRF (d) and Potts (e) models. Ikonos-2 classified image considering traditional OPF in (f), and OPF-MRF with GIMLL-L1 (g), GIMLL-L2 (h), GMRF (i) and Potts (j) models.

been smoothed by all Markov models considering OPF-MRF, since contextual information has been used. However, it seems difficult to realize whether there are differences between the four Markov Models. In order to answer that question, Figure 3 displays the mean accuracy curve over the iterations for OPF-MRF. As aforementioned, we used the best results obtained from the β values that maximized the OPF-MRF recognition rate for Potts, GIMLL-L1, GIMLL-L2 and GMRF models. Notice the accuracy stated in iteration #0 stands for the original OPF classification, i.e., the one that does not consider the contextual information.

From Figure 3, it is possible to observe three important information: (i) OPF-MRF using all models clearly improved OPF for both images (the recognition rates at iteration #1 are better than the recognition rates at iteration #0), (ii) OPF-MRF has converged over the iterations for all models, and (iii) Potts model has outperformed GIMLL-L1, GIMLL-L2 and GMRF models for both images. The first two statements have been also observed by Nakamura et al. [9]

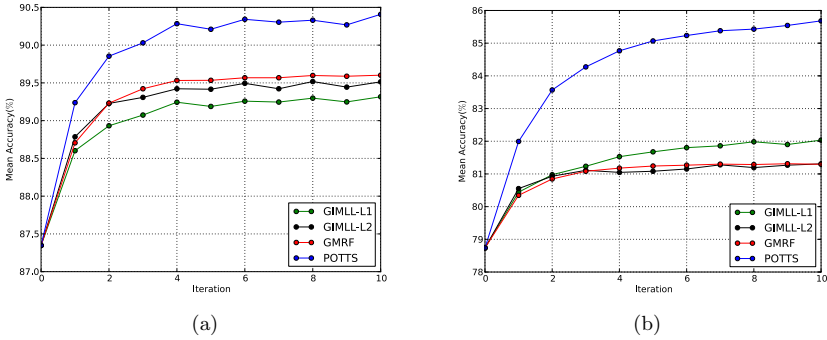


Fig. 3. Mean accuracy curves obtained over the iterations through cross-validation for (a) CBERS-2B and (b) Ikonos-2 MS images

and Osaku et al. [10], being the main contribution of this paper to evaluate whether there is a significant improvement regarding other Markovian models for contextual-based OPF classification.

One reason behind the best results obtained by Potts model concerns with the fact that it has been proposed to model situations when the particles of a given random field assume discrete values. In the OPF-MRF formulation, each element (sample) assumes a discrete state represented by its label (a natural number within $[1, 2, \dots, K]$), thus been favoured by this model. The remaining models, i.e., GIMLL-L1, GIMLL-L2 and GMRF, are more adequate to handle situations in which the variables assume continuous values.

5 Conclusions

In this paper, we have addressed the influence of different Markovian models for contextual-based classification by means of Optimum-Path Forest in the task of automatic land use image classification in two satellite images. We have observed that Potts model works better for both images, since it has been designed to handle situations in which the variables assume only discrete values, instead of GIMLL-L1, GIMLL-L2 and GMRF models, that assume a continuous distributions of the variables. For future works, we intend to employ automatic approaches to find out β values for GIMLL-L1 and GIMLL-L2 approaches.

Acknowledgments. The authors are grateful to FAPESP grants #2012/06472-9 and #2013/20387-7, as well as to CNPq grants #303182/2011-3 and #470571/2013-6.

References

1. Besag, J.: On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B (Methodological)* 48(3), 259–302 (1986)
2. Bischof, H., Schneider, W., Pinz, A.J.: Multispectral classification of landsat-images using neural networks. *IEEE Transactions on Geoscience and Remote Sensing* 30(3), 482–490 (1992)
3. Chellappa, R., Chatterjee, S.: Classification of textures using gaussian markov random fields. *IEEE Transactions on Acoustics, Speech and Signal Processing* 33(4), 959–963 (1985)
4. Geman, S., Geman, D.: Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6(6), 721–741 (1984)
5. Levada, A.L.M., Mascarenhas, N.D.A., Tannús, A.: A novel MAP-MRF approach for multispectral image contextual classification using combination of suboptimal iterative algorithms. *Pattern Recognition Letters* 31(13), 1795–1808 (2010)
6. Li, S.Z.: *Markov Random Field Modeling in Image Analysis*, 3rd edn. Springer Publishing Company, Incorporated (2009)
7. Manjunath, B., Chellappa, R.: Unsupervised texture segmentation using markov random field models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(5), 478–482 (1991)
8. Martins, A.L.D., Homem, M.R.P., Mascarenhas, N.D.A.: Super-resolution image reconstruction using the generalized isotropic multi level logistic model. In: *Proceedings of the ACM Symposium on Applied Computing*, pp. 934–938. ACM, New York (2009)
9. Nakamura, R., Osaku, D., Levada, A., Cappabianco, F., Falcão, A., Papa, J.: OPF-MRF: Optimum-path forest and markov random fields for contextual-based image classification. In: Wilson, R., Hancock, E., Bors, A., Smith, W. (eds.) *CAIP 2013, Part II. LNCS*, vol. 8048, pp. 233–240. Springer, Heidelberg (2013)
10. Osaku, D., Nakamura, R., Papa, J., Levada, A., Cappabianco, F., Falcão, A.: Optimizing contextual-based optimum-forest classification through swarm intelligence. In: Blanc-Talon, J., Kasinski, A., Philips, W., Popescu, D., Scheunders, P. (eds.) *ACIVS 2013. LNCS*, vol. 8192, pp. 203–214. Springer, Heidelberg (2013)
11. Papa, J.P., Falcão, A.X., Albuquerque, V.H.C., Tavares, J.M.R.S.: Efficient supervised optimum-path forest classification for large datasets. *Pattern Recognition* 45(1), 512–520 (2012)
12. Papa, J.P., Falcão, A.X., Suzuki, C.T.N.: Supervised pattern classification based on optimum-path forest. *International Journal of Imaging Systems and Technology* 19(2), 120–131 (2009)
13. Papa, J.P., Falcão, S.C.T.N., LibOPF, A.X.: *LibOPF: A library for the design of optimum-path forest classifiers* (2009), available at <http://www.ic.unicamp.br/~afalcao/LibOPF>
14. Potts, R.: Some generalized order-disorder transformations. *Mathematical Proceedings of the Cambridge Philosophical Society* 48(1), 106–109 (1952)
15. Wu, F.: The potts model. *Reviews of Modern Physics* 54, 235–268 (1982)
16. Zhang, B., Li, S., Jia, X., Gao, L., Peng, M.: Adaptive markov random field approach for classification of hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters* 8(5), 973–977 (2011)