CLASSIFICATION OF MULTI-SENSOR REMOTE SENSING IMAGES USING AN ADAPTIVE HIERARCHICAL MARKOVIAN MODEL

Aurélie Voisin[1], Vladimir A. Krylov[1], Gabriele Moser[2], Sebastiano B. Serpico[2] and Josiane Zerubia[1]

[1]Ayin team, INRIA-SAM, 2004 route des Lucioles, BP93, F-06902 Sophia Antipolis, Cedex, France
{aurelie.voisin, vladimir.krylov, josiane.zerubia}@inria.fr
[2] Dept. of Biophys. and Elect. Eng. (DIBE), Univ. of Genoa, Via Opera Pia 11a, I-16145, Genoa, Italy
{gabriele.moser, sebastiano.serpico}@unige.it

ABSTRACT

In this paper, we propose a novel method for the classification of the multi-sensor remote sensing imagery, which represents a vital and fairly unexplored classification problem. The proposed classifier is based on an explicit hierarchical graph-based model sufficiently flexible to deal with multi-source coregistered datasets at each level of the graph. The suggested supervised method relies on a two-step technique. In the first step, a joint statistical model is developed for the input images that consists of the finite mixtures of automatically chosen parametric families for single images, and multivariate copulas to model joint class-conditional statistics at each resolution. As a second step, we plug the estimated joint probability density functions into a hierarchical Markovian model based on a quad-tree structure. Multi-scale features correspond to different resolution images or are extracted by discrete wavelet transforms. To obtain the classification map, we resort to an exact estimator of the marginal posterior mode.

Index Terms— Supervised classification, multi-sensor data, hierarchical Markov random fields, copulas, discrete wavelet transform

1. INTRODUCTION

Remote sensing image classification is of capital interest when investigating land-use, land-cover or areas damaged by natural disasters. The aim of this paper is to develop a novel classification method that can be applied to coregistered multi-sensor remote sensing images, an interesting typology which importance is underlined by the high diversity of the acquisition systems that are currently operating. Here, we focus on the relevant case of high resolution synthetic aperture radar amplitude (SAR) [1] and optical acquisitions.

A common way to classify multi-sensor data is to combine multiple classifiers, by using, for instance, boosting [2],

support vector machines [3], or Dempster-Shafer theory of evidence [4]. Posterior probabilities may also be combined [5]. But, most of these methods tend to degrade the acquisitions by requiring a re-sampling of the data given the fact that multi-sensor data intrinsically implies a multi-resolution aspect of such acquisitions. To our best knowledge, so far, no multi-sensor classifiers have overcome this limitation.

The Bayesian classification technique proposed in this paper is based on an explicit hierarchical Markovian context [6] and introduces a method to jointly model and classify the different data sources instead of fusing the final classifications or the input images. Moreover, at each level of the hierarchical graph, a flexible number of the input sources (images) is allowed, to adapt the graph to the available input dataset. The method consists of two steps. The first step is the statistical pixelwise modeling of the coregistered input images: for each class and each channel in the chosen stacked-vector input dataset, the marginal probability density functions (PDFs) are estimated by finite mixtures of automatically chosen parametric families [7], typically Gaussian mixtures for optical images or generalized Gamma mixtures for SAR imagery. Next, a d-variate copula [8] (d being the number of input channels) is applied to estimate multivariate joint class-conditional statistics, merging the marginal PDF estimates of the input channels. The best copula is selected among a predefined dictionary of copulas according to a chi-square goodness-of-fit test [7]. On the second step, we plug the estimated joint PDFs into a contextual model that uses a multi-scale approach via a hierarchical Markovian model [6] based on a quad-tree structure. The hierarchical structure is quite robust to noise and preserves spatial details of high resolution acquisitions. The quad-tree allows to integrate an exact estimator of the marginal posterior mode (MPM) [9] that aims to determine the unknown class labels. The prior probability is iteratively updated at each level of the tree, leading to an algorithm more robust with respect to speckle noise in SAR acquisitions [1] when compared to a non-updated prior [9, 10]. Multi-scale features are either the multi-resolution images, or the features extracted by discrete wavelet transforms (DWT) [11, 12]. A

similar model has already been introduced in [10] in the case of mono-resolution multi-band images, and is extended and validated here on multi-sensor and multi-resolution acquisitions that are directly integrated in the adaptive explicit hierarchical graph.

The rest of the paper is organized as follows. In Sec. 2, we develop the MPM model based on the quad-tree that we use to classify our data. In Sec. 3, we introduce the statistical multivariate copula-based model that combines the marginal PDF models of the input images at each resolution. In Sec. 4, we present an experimental validation of the developed model.

2. THE HIERARCHICAL CLASSIFIER MODEL

We aim to estimate a set of hidden labels X given a set of observations Y. X and Y are random processes, and X is a Markov random field (MRF) on the considered hierarchical graph. To address our classification problem, we use an explicit hierarchical graph-based model [6, 9] that has a specific quad-tree structure. Such a structure ensures the scale causality which allows a non-iterative algorithm to be used, thus implying a computational time decrease as compared to an iterative computation. Among the various algorithms proposed to estimate the labels on hierarchical graphs [6], we use an exact estimator of the marginal posterior mode (MPM). A classical MPM algorithm [9] is generally run in two passes, referred to as bottom-up ("forward") and top-down ("backward") passes, that respectively aim to estimate the partial posterior marginals $p(x_s|y_{d(s)})$, s representing a site and $d(s)$ its descendants, and then to estimate successively the labels at each level by maximizing the posterior marginals $p(x_s|y)$. To improve the robustness with respect to the speckle noise [10], we propose to truncate the top-down pass: we use the highest tree-level label estimates to update the prior, and we run a novel MPM algorithm on a smaller quad-tree (see Fig. 1).

The maximization of the posterior marginal $p(x_s|y)$ is done by employing a modified Metropolis Dynamics algorithm which has good properties for both its relative low computation time and the good precision of its results [13].

To estimate the posterior probability, we need information about the likelihood, the prior probability and the transition probability at each site s of the quad-tree. The likelihood term $p(y_s|x_s)$ is estimated for each stacked-vector input dataset at each tree level according to the model presented in Sec. 3. The transition probabilities between the scales, that determine the hierarchical MRF by representing the causality of the statistical interactions between the different levels of the tree, are expressed in the form introduced by Bouman et al. [14]. Specifically, for all sites $s \in S$ and all scales $n \in [0; R-1]$,

$$p(x_s^n = \omega_m | x_{s-}^n = \omega_k) = \begin{cases} \theta_n, & \text{if } \omega_m = \omega_k \\ \frac{1-\theta_n}{M-1}, & \text{otherwise} \end{cases}, \quad (1)$$

with the parameter $\theta_n > 1/M$ and where ω_m and ω_k represent the classes m and k, $m, k \in [1; M]$, respectively, and

M represents the number of considered classes for the final classification. This model favors an identical parent-child labeling.

Finally, the prior distributions at the current coarsest tree level are given by the estimated classification maps by resorting to a single-scale Markovian model [7]. The prior information at lower levels is determined with the knowledge of the transition probabilities. For a more comprehensive study, see [10].

3. JOINT PDF MODEL

At each tree level n and for each considered class m we perform the joint statistics modeling of the multi-sensor input images (e.g optical color bands and/or polarization SAR image channels), or their respective DWT features. The joint PDFs are independently modeled at each scale n, hence, in this section, we consider that n is fixed to simplify the notations. At each n, the following procedure is applied: the statistics of each input band are independently modeled, and are then used to generate a joint PDF by employing the statistical instrument of copulas [8].

3.1. Marginal PDF estimation

For each input image, we want to estimate the distributions of each class ω_m considered for the classification, $m \in [1; M]$, given a training set. The PDF $f_m^j(z^j)$ of the jth input band, $j \in [1; d]$, is then modeled via finite mixtures of K^j independent greylevel distributions:

$$f_m^j(z^j) = p_m^j(z^j|\omega_m) = \sum_{i=1}^{K^j} P_{mi}^j p_{mi}^j(z^j|\theta_{mi}^j), \quad (2)$$

where z^j is a greylevel, $z^j \in [0; Z-1]$. P_{mi}^j are the mixing proportions such that for a given m, $\sum_{i=1}^{K^j} P_{mi}^j = 1$ with $0 \leq P_{mi}^j \leq 1$. θ_{mi}^j is the set of parameters of the ith PDF mixture component of the mth class. The use of finite mixtures instead of single PDFs offers the possibility to consider heterogeneous PDFs, usually reflecting the contributions of the components present in each class, for instance, different kinds of crops for the vegetation class. Moreover, the use of finite mixtures can be seen as a generalization of the determination of a single PDF, and allows to estimate both the best finite mixture model and/or the best single PDF model.

When the jth input channel is an optical acquisition, we consider that the PDF $f_m^j(z^j)$ related to each class can be modeled by a finite mixture of Gaussian distributions. This choice is consistent with the fact that a Gaussian distribution is a usually accepted model for a homogeneous area in an optical remote sensing image. For each mixture component, the mean and the variance parameters are estimated by a stochastic expectation maximization (SEM) [15] algorithm.
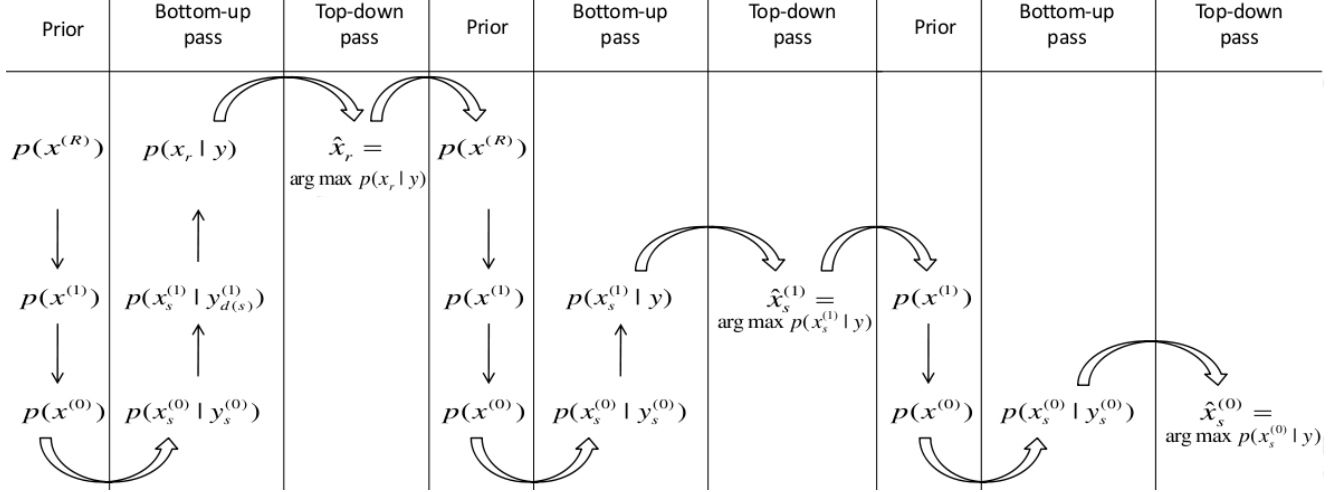
Fig. 1. Proposed MPM estimation on the quad-tree on which the set of sites is hierarchically partitioned as $S = S^0 \bigcup S^1 \bigcup ... \bigcup S^R$, where R corresponds to the coarsest resolution (the root), 0 corresponds to the reference level (finest resolution) and S^n is the collection of sites at the nth scale level ($n = 0, 1, ..., R$). In this figure, $R = 2$.

The SAR acquisitions are known to be affected by speckle noise [1], thus we propose to model each class marginal conditional PDF by a mixture of generalized Gamma distributions. When the generalized Gamma distribution can not be fitted, the PDF $p_{mi}^j(z^j|\theta_{mi}^j)$ is automatically chosen between log-normal, Weibull and Nakagami distributions, that are all commonly employed to model SAR imagery [7]. In that case, the best-fitting mixture model for each considered class is obtained by combining a density parameter estimation via the method of log-cumulants [16] and a SEM algorithm [7].

The estimation of K^j is integrated into the SEM procedure [7].

3.2. Combination of marginal distributions via multivariate copulas

A multivariate copula is a d-dimensional joint distribution defined on $[0, 1]^d$ such that marginal distributions are uniform on $[0, 1]$. According to Sklar's theorem [8], we can obtain a unique joint d-variate PDF for any continuous random variables $z^1, ..., z^d$ in $\mathbb{R}$:

$$h_m(z^1, ..., z^d) = \prod_{j=1}^{d} f_m^j(z^j) \times c_m(F_m^1(z^1), ..., F_m^d(z^d)),$$

(3)

where $\{f_m^1, ..., f_m^d\}$ are the marginal PDFs estimated for the jth input image and for the mth class, defined in Sec. 3.1, and $\{F_m^1, ..., F_m^d\}$ their corresponding cumulative distribution functions. c_m is the density of the copula C_m. For each m, the computation of the joint PDF h_m is boiled down to the determination of the copula family C_m according to Eq. (3).

The bivariate copulas have been studied extensively in the literature, which is not necessarily the case for multivariate copulas. We only focus in this paper on Archimedean copula families, which represent a good compromise between analytical tractability and their modeling capacity [8]. We consider a predefined dictionary of three single-parameter copulas: Clayton, Ali-Mikhail-Haq (AMH) and Gumbel copula families. For each class, we determine the best fitting copula as the one reporting the highest p-value in Pearson Chi-square (PCS) goodness-of-fit test, see [7]. The null hypothesis in PCS is that the sample frequencies $C_m(F_m^1(u^1), ..., F_m^d(u^d))$, where $(u^1, ..., u^d)$ represent the observed data, are consistent with the theoretical frequencies $C_m(v^1, ..., v^d)$ for each considered copula.

To estimate the copula parameters, we use the relationship between copulas and Kendall's τ [8], a concordance-discordance measure that can be empirically estimated by using the training sets.

4. EXPERIMENTAL RESULTS

We applied the developed classification approach to a multi-sensor dataset that consists of two images of the port area of Port-au-Prince (Haiti): an HH-polarized COSMO-SkyMed (CSK) SAR image (StripMap acquisition mode, 2.5 m pixel spacing, geocoded, single-look) of 320×400 pixels and a coregistered pan-sharpened (Pan.) GeoEye acquisition of 1280×1600 pixels, see Fig. 2(a),(b). Initially, the GeoEye image resolution is of 0.5 meters. To fit with the dyadic decomposition imposed by the quad-tree, we re-sampled the data to obtain the 0.625 m resolution (equal to 1/4 of the SAR image's resolution). To employ the multi-resolution information, and to avoid "empty" tree levels, we integrate
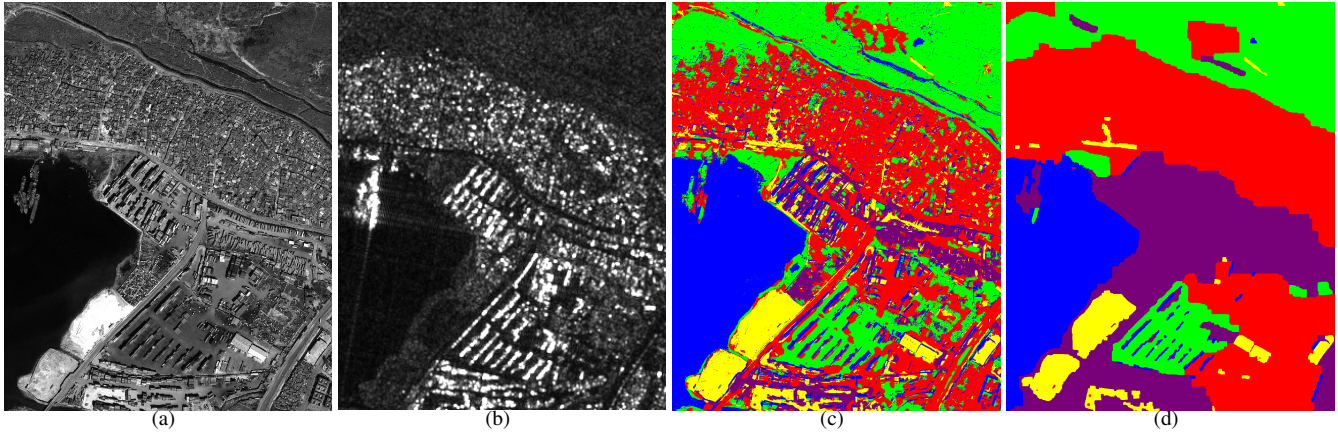
Fig. 2. (a) Optical (©GeoEye, 2010) and (b) SAR images (CSK, ©ASI, 2010), and different classification method results obtained with: (c) the proposed method applied to the multi-sensor dataset, and (d) a single-scale MRF applied to the highest resolution image. Legend: water (blue), urban areas (red), vegetation (green), sand (yellow) and containers (purple).

Table 1. Results for the Port-au-Prince dataset: by class and overall accuracy.

	Port-au-Prince, Haiti					
	water	urban	vegetation	sand	containers	**overall**
Proposed classif.	100%	75.24%	87.16%	98.89%	49.31%	**82.12%**
Proposed classif. (Pan. only)	100%	67.12%	86.89%	98.83%	41.90%	**78.95%**
MRF-based classif.	100%	100%	81.42%	99.94%	59.62%	**88.20%**

additional information by employing a hierarchical decomposition of the original data. For our specific dataset, the optical image is integrated at the bottom of the quad-tree, and the SAR image, at level $R = 2$. We, thus, apply a 2-D DWT to decompose the finest resolution image along 2 levels or more, and integrate the obtained images in the tree. At the bottom and at the first stage of the graph, we only have one observation, and at higher tree levels, we have two observations, the SAR image and the corresponding decompositions of the optical one, that are combined by using the model presented in Sec.3. We can further increase the number of levels of the tree by introducing the SAR image decomposition.

Given the size of the input image, $R = 2$ appears to be a reasonable choice: a higher decomposition level might remove some relevant image details (e.g., roads). A wide choice of wavelet functions exists, and according to a preliminary comparative study [10], we chose Daubechies-10 wavelets and Haar wavelets to decompose the SAR and the optical images respectively. By a trial-and-error procedure, the single-scale MRF parameter (see Sec. 2) was set to 4.8 and the transition probability θ_n was set to 0.8 in (1), which means that a site s at scale n has a probability of 80% to belong to the same class as its ascendant s^-.

Manually annotated non-overlapping training and test sets were selected in homogeneous areas (no borders were taken into account). Five classes were chosen: water, urban areas, vegetation, sand and containers. We compare classification results obtained: 1) with the proposed approach on the multi-sensor SAR/optical dataset, 2) with the same approach applied only to the optical image and its decomposition, and 3) with a single-scale MRF-based approach [7] applied only to the optical image. The classification results are presented in Fig. 2(c),(d) and numerically assessed via accuracy estimates (Tab. 1). The classification map obtained when using a multi-sensor input shows satisfactory results because they lead to a detailed classification (Fig. 2(c)). Numerical results highlight the improvement related to the combination of the multi-sensor images as compared to single-sensor (only Pan.) classification performances, especially in the urban areas, in which the SAR acquisition seems helpful. The presented table (Tab. 1) indicates that the single-scale MRF-based method leads to a higher numerical accuracy. But we stress that the map obtained by using the MRF-based method is severely oversmoothed (Fig. 2(d)), and this affects only marginally the numerical accuracies due to the localization of the ground truth samples within homogeneous regions. Visually, the classification obtained with the hierarchical classifier is more detailed, which renders the proposed method preferable. Yet, it may be interesting to explore a prior despeckling of the SAR image [17] to decrease the noise effects on the final results.

5. CONCLUSION

The method proposed in this paper allows to deal with multi-sensor acquisitions for land-cover mapping purposes. The main novelty lies in the direct integration of both multi-resolution and multi-sensor data in the hierarchical graph model without resorting to any external fusion as pre- or post-processing. The proposed methodological approach combines a joint copula-based statistical model for optical and SAR acquisitions, with a hierarchical MRF, leading to a statistical supervised classification approach. We have proposed a dictionary-based multivariate copulas model that enables to fuse multi-sensor acquisitions, and developed a novel MPM-based hierarchical Markov random field model that iteratively updates the prior probabilities and, thus, leads to the improved robustness of the classifier. The hierarchical MRF considered here has three advantages: it is quite robust to speckle noise, it does not require a time-expensive iterative optimization algorithm and it is sufficiently flexible to allow the integration of both multi-sensor and multi-resolution acquisitions. As a direction of future development, we intend to perform experiments with wider datasets including the novel high-resolution Pléiades (©CNES) optical imagery.

6. ACKNOWLEDGMENT

7. REFERENCES

[1] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar images*, SciTech Publishing, 2004.

[2] R. E. Schapire, "A brief introduction to boosting," in *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, 1999, vol. 16, pp. 1401–1406.

[3] B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensor data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3858–3866, 2007.

[4] A. Al-Ani and M. Deriche, "A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence," *Journal Of Artificial Intelligence Research*, vol. 17, pp. 333–361, 2011.

[5] J. A. Benediktsson and P. H. Swain, "Consensus theoretic classification methods," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 4, pp. 688–704, 1992.

[6] C. Graffigne, F. Heitz, P. Perez, F. Preteux, M. Sigelle, and J. Zerubia, "Hierarchical Markov random field models applied to image analysis: a review," in *Proc. SPIE Conf. on Neural, Morphological and Stochastic Methods in Image Proc.*, 1995, vol. 2568, pp. 2–17.

[7] V. Krylov, G. Moser, S. B. Serpico, and J. Zerubia, "Supervised high resolution dual polarization SAR image classification by finite mixtures and copulas," *IEEE Journal of Selected Topics in Signal Proc.*, vol. 5, no. 3, pp. 554–566, 2011.

[8] R. B. Nelsen, *An introduction to copulas*, Springer, New York, 2nd edition, 2006.

[9] J.-M. Laferte, P. Perez, and F. Heitz, "Discrete Markov modeling and inference on the quad-tree," *IEEE Trans. Image Process.*, vol. 9, no. 3, pp. 390–404, 2000.

[10] A. Voisin, V. Krylov, G. Moser, S. B. Serpico, and J. Zerubia, "Multichannel hierarchical image classification using multivariate copulas," in *Proc. of IS&T/SPIE Electronic Imaging 2012*, San Francisco, USA, Jan. 2012, vol. 8296, p. 82960K.

[11] S. G. Mallat, *A wavelet tour of signal processing*, Academic Press, 3rd edition, 2008.

[12] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 41, no. 7, pp. 909–996, 1988.

[13] M. Berthod, Z. Kato, S. Yu, and J. Zerubia, "Bayesian image classification using Markov random fields," *Image and Vision Computing*, vol. 14, no. 4, pp. 285–295, 1996.

[14] C. Bouman and M. Shapiro, "A multiscale random field model for Bayesian image segmentation," *IEEE Trans. Image Process.*, vol. 3, no. 2, pp. 162–177, 1994.

[15] G. Celeux, D. Chauveau, and J. Diebolt, "Stochastic versions of the EM algorithm: an experimental study in the mixture case," *Journal of Statistical Computation and Simulation*, vol. 55, no. 4, pp. 287–314, 1996.

[16] C. Tison, J.-M. Nicolas, F. Tupin, and H. Maitre, "A new statistical model for Markovian classification of urban areas in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2046–2057, 2004.

[17] S. Parrilli, M. Poderico, C. V. Angelino, and L. Verdoliva, "A nonlocal SAR image denoising algorithm based on LLMMSE wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 606–616, 2012.