IWSSIP 2010 - 17th International Conference on Systems, Signals and Image Processing

Improving Multispectral Image Classification by Using Maximum Pseudo-Likelihood Estimation and Higher-Order Markov Random Fields

Alexandre L. M. Levada[1], Alberto Tannús[3]
Physics Institute of São Carlos
University of São Paulo - USP
São Carlos/SP, Brazil
alexandreluis@ursa.ifsc.usp.br, goiano@ifsc.usp.br

Nelson D. A. Mascarenhas[2]
Computer Department
Federal University of São Carlos - UFSCar
São Carlos/SP, Brazil
nelson@dc.ufscar.br

Abstract—In this paper we address the multispectral image contextual classification problem following a Maximum *a Posteriori* (MAP) approach. The classification model is based on a Bayesian paradigm, with the definition of a *Gaussian Markov Random Field* model (GMRF) for the observed data and a Potts model for the *a priori* knowledge. The MAP estimator is approximated by the *Game Strategy Approach* (GSA) algorithm, a non-cooperative game theory based method. Maximum Pseudo-Likelihood is adopted for MRF model parameter (regularization parameter) estimation on higher-order neighborhood systems. To evaluate the proposed method, experiments using Nuclear Magnetic Resonance (NMR) images were proposed. Quantitative results obtained by using Cohen's Kappa coefficient for performance evaluation show significant improvement in contextual classification, indicating the effectiveness of our MAP-MRF approach.

Keywords-Markov Random Fields;Contextual classification; Maximum pseudo-likelihood; Game Strategy Approach.

I. INTRODUCTION

Undoubtedly, Markov Random Fields (MRF) define a powerful mathematical tool for contextual modeling of spatial data. With advances in probability and statistics, as the development of *Markov Chain Monte Carlo* (MCMC) simulation techniques and algorithms for combinatorial optimization, MRF's became a central topic in fields including image processing, computer vision and pattern recognition. In this paper, we are concerned with the multispectral image contextual classification problem using a MAP-MRF Bayesian framework that combines two MRF models: a Gaussian Markov Random Field (GMRF) for the observations (likelihood) and a Potts model as a smooth prior, acting as a regularization term in the presence of noisy data, using higher-order neighborhood systems.

In MAP-MRF approaches, the prior probability is given by a Markovian model. In this scenario, the MRF model parameter assumes the role of a regularization parameter, since it controls the tradeoff between the prior knowledge and the likelihood. However, in most contextual classification systems, MRF model parameters are still chosen by a trial-and-error procedure through simple manual adjustments [1]. In this work, we use novel pseudo-likelihood equations for the Potts model defined in both second and third order systems (8 and 12 neighbors) to estimate the MRF parameters [2,3].

For classification purposes, it is widely known that MAP-MRF based problems do not have closed-form solution, requiring the use of iterative combinatorial optimization algorithms. In this work, we use GSA [4] to approximate the MAP estimator, a stochastic algorithm based on non-cooperative game theory [5]. Several supervised pattern classifiers were used to provide different initializations to the contextual classifier. Statistical analysis of the obtained results indicated the effectiveness of the proposed method.

The remaining of the paper is organized as follows: Section 2 describes the proposed method, discussing the classification model, MPL estimation and the GSA algorithm for MRF-based contextual classification. Section 3 shows the experiments and the obtained results and Section 4 presents the conclusions and final remarks.

II. THE PROPOSED METHOD

A. MAP-MRF Contextual Classification

Let $\mathbf{x}_W^{(p)}$ be the label field at the *p-th* iteration, $\mathbf{y}$ the observed multispectral image, $\vec{\theta}$ the vector of GMRF hyperparameters, $\vec{\Phi}$ the vetor of class-dependent GMRF spectral parameters ($\vec{\mu}_m$ and $\mathbf{\Sigma}_m$) and β the Potts model hyperparameter. Then, combining the two models, according to the Bayes' rule, the current label of a pixel (i,j) can be updated by choosing the label that maximizes the functional given by [6]:

$$Q\left(x_{ij}=m \mid \mathbf{x}_W^{(p)},\mathbf{y},\vec{\theta},\vec{\Phi},\beta\right)=\frac{1}{2}\ln\left|\hat{\Sigma}_m\right|-$$
$$\frac{1}{2}\left\{\vec{y}_{ij}-\hat{\vec{\mu}}_m\left[\hat{\boldsymbol{\theta}}^T\vec{y}_{\eta_{ij}}-2\left(\sum_{ct}\hat{\boldsymbol{\theta}}^{ct}\right)\hat{\vec{\mu}}_m\right]\right\}^T\times\hat{\mathbf{\Sigma}}_m^{-1}\times \qquad (1)$$
$$\frac{1}{2}\left\{\vec{y}_{ij}-\hat{\vec{\mu}}_m\left[\hat{\boldsymbol{\theta}}^T\vec{y}_{\eta_{ij}}-2\left(\sum_{ct}\hat{\boldsymbol{\theta}}^{ct}\right)\hat{\vec{\mu}}_m\right]\right\}+\hat{\beta}U_{ij}(m)$$

where $\hat{\boldsymbol{\theta}}^{ct}$ is a diagonal matrix whose elements are the horizontal, vertical and diagonals hyperparameters (4×4), $ct=1,\ldots,K$, where K is the number of bands, $\hat{\boldsymbol{\theta}}^{T}$ is a $4\times4K$ matrix built by stacking the $\hat{\boldsymbol{\theta}}^{ct}$ matrices from each image band, that is, $\hat{\boldsymbol{\theta}}^{T}=[\hat{\boldsymbol{\theta}}^{ct1},\ldots,\hat{\boldsymbol{\theta}}^{ctK}]$ and $\vec{y}_{\eta_{ij}}$ is a $4K\times1$ vector whose elements are the sum of the two neighboring elements on each direction (horizontal, vertical and diagonals), for all bands.

Our main contribution here is the use of higher-order neighborhood systems in both MRF models. In the original work [6], all models were restricted to first-order systems (4 neighbors). In this work, we employ second and third orders systems, improving contextual modeling. We show that higher-order systems significantly improves the classification performance.

B. Maximum Pseudo-Likelihood Estimation

Now that contextual classification is stated as a MAP-MRF problem, the next step is the parameter estimation. In this work, we adopt the MPL approach for the MRF parameters. More precisely, we use MPL approach to estimate both GMRF and Potts model hyperparameters, $\vec{\theta}$ and β. The main motivation for MPL estimation is that the most traditional method, Maximum Likelihood (ML), is not computationally tractable. Besides, it has been shown that MPL estimators possess a series of desirable statistical properties such as consistency and asymptotic normality. The spectral GMRF model parameters are estimated by sample mean and covariance matrix.

1) GMRF Model: As there is no dependency between the hyperparameters from different image bands, it is quite reasonable to perform MPL estimation in each band independently. Assuming this hypothesis and considering a second-oder neighborhood system, the pseudo-likelihood function is given by:

$$\log PL\left(\vec{\theta},\mu,\sigma^2\right)=\sum_{(i,j)\in W}\left\{-\frac{1}{2}\log\left(2\pi\sigma^2\right)-\frac{1}{2}\left[y_{ij}-\vec{\theta}\vec{\Psi}_{ij}-\mu\left(1-2\vec{\theta}I\right)\right]^2\right\} \tag{2}$$

where W is an image band, $\vec{\theta}^{T}=[\theta_1,\theta_2,\theta_3,\theta_4]$ is the vector of spatial hyperparameters and $\vec{\Psi}_{ij}$ is a 4-D vector where each element is defined as the sum of the two neighboring pixels in the horizontal, vertical and both diagonal directions, respectively. The expression for the MPL estimator of $\vec{\theta}$ is given by [7]:

$$\hat{\vec{\theta}}_{MPL}=\left[\sum_{(i,j)\in W}\left(y_{ij}-\hat{\mu}\right)\tilde{\vec{\Psi}}_{ij}^{T}\right]\left[\sum_{(i,j)\in W}\tilde{\vec{\Psi}}_{ij}\tilde{\vec{\Psi}}_{ij}^{T}\right]^{-1} \tag{3}$$

where $\hat{\mu}$ is the sample mean of the image band and:

$$\tilde{\vec{\Psi}}_{ij}=\vec{\Psi}_{ij}-\frac{1}{N}\sum_{(i,j)\in W}\vec{\Psi}_{ij} \tag{4}$$

Although MPL estimation is widely used in MRF applications, little is known about its accuracy. In recent works, we have derived analytical expressions for the asymptotic variance of $\hat{\vec{\theta}}_{MPL}$, completely characterizing its behavior in the limiting case (when the number of samples grows infinitely) showing that this estimation method is reliable and accurate [8].

2) Potts Model: Analyzing expression (1), we see that the decision rule incorporates both spectral and spatial information. Basically, it works as follows: in case of noisy situations no confidence is placed in the contextual information (since β is small) and the decision is made mostly based on the likelihood. On the other hand, when contextual information is informative, the Potts MRF model parameter controls the tradeoff between data fidelity and prior knowledge. Thus, a correct setting is crucial for the classification performance.

In order to perform contextual classification in higher-order systems, novel methods for Potts model parameter estimation are required. In this work, we use new methods for both second and third order systems recently proposed in the literature [2, 3]. Asymptotic evaluations and MCMC simulations have proved the accuracy of the derived equations [2, 3].

C. Game Strategy Approach for Image Labelling

It has been shown that MAP-MRF problems do not allow closed-form solutions. Thus, in order to approximate the MAP estimator, combinatorial optimization algorithms are applied to an initial solution provided by a classification technique (neural networks, clustering algorithms, statistical pattern classifiers), improving it iteratively. Some of the most popular optimization algorithms found in image processing literature are the *Iterated Conditional Modes* (ICM), *Maximizer of Posterior Marginals* (MPM), *Graduated Non Convexity* (GNC) and *Highest Confidence First* (HCF). In this work, we use an alternative algorithm named *Game Strategy Approach* (GSA) [4], based on non-cooperative game theory.

In a *n*-person game, $I=\{1,2,\ldots,n\}$ denotes the set of all players. Each player *i* has a set of pure strategies S_i (labels). The game process consists of, at a given instant, each player choosing a strategy $s_i\in S_i$. Hence, a play $s=\left(s_1,s_2,\ldots,s_n\right)$ is yielded, and a payoff $H_i\left(s\right)$ is assigned to each player. In GSA, two hypotheses are assumed: first, the payoff of a player depends only on its own strategy and the strategies of its neighbors; and second, it is supposed that each player knows all possible strategies and the payoff given by each one of them. In non cooperative games, each player selects independently his own strategy to maximize the local payoff. The solutions of such a game are the *Nash points*, a condition achieved when none of the players can improve his expected payoff by unilaterally changing his strategy. In mathematical terms, a play $t^*=\left(t_1^*,t_2^*,\ldots,t_n^*\right)$ satisfies the Nash Equilibrium if [5]:

$$\forall i : H_i\left(t^*\right) = \max_{s_i \in S_i} H_i\left(t^* \parallel t\right) \qquad (5)$$

where $t^* \parallel t$ is the play obtained by replacing t^* by t. It has been shown that Nash points always exist in non-cooperative *n*-person games with pure or mixed strategies [5]. The GSA fundamentals are based on two major results derived by [6]:

THEOREM 1: *The set of local maximum points of the posterior distribution in MRF image labeling problems is identical to the set of Nash points of the corresponding non-cooperative game.*

THEOREM 2: *The GSA algorithm converges to a Nash point in a finite number of steps.*

The GSA algorithm is given below.

ALGORITHM 1 – *GSA Algorithm*

INPUT: An initial label map $L^{(0)}$

OUTPUT: A contextual classification result

1. Initialize $L^{(0)}$ and set $k = 0$
2. At iteration $k \geq 0$, repeat for each pixel
   a. Choose the label $l_i' \neq l_i^{(k)}$ that maximizes the payoff, that is:
   $$H_i\left(l_i^{(k)} \parallel l_i'\right) = \max_{l_i} H_i\left(l_i^{(k)} \parallel l_i\right)$$
   b. If $H_i\left(l_i^{(k)} \parallel l_i'\right) \leq H_i\left(l_i^{(k)}\right)$ then
   $$l_i^{(k+1)} = l_i^{(k)}$$
   c. Otherwise,
   $$l_i^{(k+1)} = l_i', \quad \text{with probability } \alpha$$
   $$l_i^{(k+1)} = l_i^{(k)}, \quad \text{with probability } (1-\alpha)$$
   d. Let $L^{(k+1)} = \left(l_1^{(k+1)}, l_2^{(k+1)}, \ldots, l_n^{(k+1)}\right)$
3. If $L^{(k+1)}$ is a Nash point or $k \geq K$, stop. Otherwise, increment k and go back to (2)

III. EXPERIMENTS AND RESULTS

In order to test and evaluate the contextual classification methods previously described, we show some experiments in NMR images of marmocets brains, a monkey species often used in medical experiments. These images were acquired by the CInAPCe project, an abbreviation for the Portuguese expression "Inter-Institutional Cooperation to Support Brain Research", a Brazilian research project that has as main purpose the establishment of a scientific network seeking the development of neuroscience research through multidisciplinary approaches. In this sense, pattern recognition can contribute to the development of new methods and tools for processing and analyzing magnetic resonance imaging and its integration with other methodologies in the investigation of epilepsies. Figure 9 shows the bands PD, T1 and T2 of a marmocet NMR multispectral brain image.

Figure 1.PD, T1 and T2 NMR image bands of a marmoset brain

The contextual classification of the multispectral image was performed by applying the GSA algorithm using second and third order neighborhood systems based on several initializations provided by seven different pattern classifiers: Quadratic Bayesian Classifier (*QDC*) and Linear Bayesian Classifier (*LDC*) under Gaussian hypothesis, Parzen-Windows Classifier (*PARZENC*), K-Nearest Neighbour Classifier (*KNNC*), Logistic Classifier (*LOGLC*), Nearest Mean Classifier (*NMC*) and Decision Tree Classifier (*TREEC*). We used 300 training samples for each one of the 3 classes: white matter, gray matter and background. The classification errors and confusion matrix are estimated by the *leave-one-out cross validation* method. Convergence was considered by achieving one of two conditions: less than 0.1% of the pixels are updated in the current iteration, or the maximum of 5 iterations is reached.

Quantitative analysis is performed by calculating the Cohen's *Kappa* coefficient from the confusion matrix. To analyze the obtained results we use both Z and T statistics, to compare local and average performances, respectively. The Z statistic is normally distributed and it is given by:

$$Z = \frac{\left|k_1 - k_2\right|}{\sqrt{\sigma_{k_1}^2 + \sigma_{k_2}^2}} \qquad (6)$$

where k_1 and k_2 are the Kappa coefficients obtained by each classifier, and $\sigma_{k_1}^2$ and $\sigma_{k_1}^2$ are their variances. Considering a significance level $\alpha = 0.05$, if $Z > 1.96$ the difference beetwen two *Kappas* is considered significant. For a global analysis, two average performances can be compared by calculating the following statistic:

$$T = \frac{\left|\overline{k_1} - \overline{k_2}\right|}{\sigma_d / \sqrt{n}} \qquad (7)$$

where $\overline{k_1}$ and $\overline{k_2}$ denotes the average *Kappas* obtained by each technique, σ_d is the standard deviation of the punctual differences between the respective *Kappas* belonging to each group and *n* is the number of elements of the groups. The higher the difference between the two means (numerator), the higher the chance of having two distinct groups. But, the higher the variability of the observed results (denominator), the harder the discrimination in two distinct groups. Considering a significance level $\alpha = 0.05$, the two means are statistically different if $T > 1.943$ (T is distributed according a t_{n-1} law). Table 1 shows the

MPL estimators of the Potts MRF model parameters (*regularization parameters*), obtained by using the equations proposed in [2, 3]. Tables 2 and 3 show the contextual classification performances, in terms of *Kappa*, obtained by applying the GSA algorithm in each one of the seven initializations previously described considering second and third order Potts MRF models.

TABLE I. MPL ESTIMATORS OF POTTS MRF MODEL PARAMETERS ON SECOND AND THIRD ORDER SYSTEMS.

Initialization	***2ª order***	***3ª order***
PARZENC	0.5949	0.4024
KNNC	0.5954	0.4020
LOGLC	0.7799	0.5157
LDC	0.6250	0.4195
QDC	0.5995	0.4040
NMC	0.6029	0.4053
TREEC	0.5790	0.3901

TABLE II. KAPPA COEFFICIENTS AND THEIR VARIANCES FOR CONTEXTUAL CLASSIFICATION ON SECOND ORDER SYSTEMS

Initialization	***Kappa***	σ_k^2
PARZENC	0.9067	0.0001453
KNNC	0.8817	0.0001809
LOGLC	0.7850	0.0002909
LDC	0.9367	0.0001083
QDC	0.9317	0.0001083
NMC	0.8967	0.0001599
TREEC	0.9067	0.0001453

TABLE III. KAPPA COEFFICIENTS AND THEIR VARIANCES FOR CONTEXTUAL CLASSIFICATION ON THIRD ORDER SYSTEMS

Initialization	***Kappa***	σ_k^2
PARZENC	0.9833	0.00002746
KNNC	0.9567	0.00007006
LOGLC	0.8167	0.00002577
LDC	0.9917	0.00001381
QDC	0.9933	0.00001106
NMC	0.9750	0.00004096
TREEC	0.9783	0.00003558

Comparing the best performances on second and third order systems (LDC and QDC initializations), we have $Z = 5.1807$, leading to a *p-value* smaller than 0.0001, which provides strong evidences in favor of the proposed method. Figure 2 compares the label maps obtained by the best performances on second and third order systems.

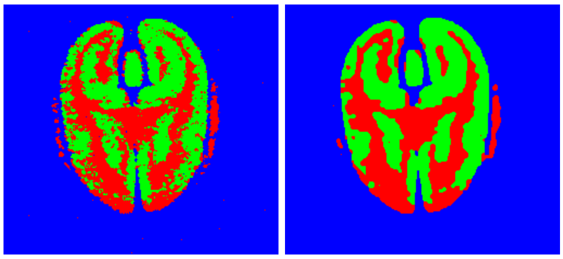

Figure 2.Best classification results on second (LDC) and third (QDC) order neighborhood systems.

Finally, a comparison between the average *Kappas* also indicates a clear improvement on classification performance when using higher-order systems. From the data in Tables 1 and 2, we have $T = 4.9655$, leading to a *p-value* smaller than 0.0005, which reinforces the strength of the obtained results.

IV. CONCLUDING REMARKS

This paper addressed the problem of multispectral image contextual classification following a MAP-MRF approach on higher-order systems. Parameter estimation was performed by a Maximum Pseudo-Likelihood approach. Quantitative analysis using the *Kappa* coefficient indicated that the use of higher-order systems significantly improves classification performance. Future works include the study of other optimization algorithms, the use of multiple initial conditions and classifier combination rules to further improve classification performance.

ACKNOWLEDGMENTS

We would like to thank FAPESP for the financial support. We also would like to thank Hilde Buzzá for the NMR images acquisition.

REFERENCES

[1] J. Wu, A. C. S. Chung, "A segmentation model using compound Markov Random Fields based on a boundary model", IEEE Trans. on Image Processing, v. 16, n. 1, 241-252, 2007.

[2] A. L. M. Levada, N. D. A. Mascarenhas, A. Tannús, "A novel pseudo-likelihood equation for Potts MRF model parameter estimation in image analysis", In: Procs. of the 15th International Conference on Image Processing (ICIP), San Diego, pp.1828–1831, 2008.

[3] A. L. M. Levada, N. D. A. Mascarenhas, A. Tannús, "Improving Potts MRF model parameter estimation using higher-order neighborhood systems on stochastic image modeling", In: Procs. of the 15th International Conference on Systems, Signals and Image Processing (IWSSIP), Bratislava, pp.385-388, 2008.

[4] S. Yu, M. Berthod, "A game strategy approach for image labelling", Computer Vision and Image Understanding, v. 61, n. 1, pp. 32-37, 1995.

[5] J. F. Nash, "Equilibrium points in n-person games", Procs. of the National Academy of Sciences, v. 36, pp. 48-49, 1950.

[6] T. Yamazaki, D. Gingras, "A contextual classification system for remote sensing using a multivariate gaussian MRF model", In:Procs. of the 9th International Symposium on Circuits and Systems (ISCAS), Atlanta, pp. 648-651, 1996.

[7] C. S. Won, R. M. Gray, "Stochastic Image Processing", Kluwer Academic/Plenum Publishers, 2004.

[8] A. L. M. Levada, N. D. A. Mascarenhas, A. Tannús, "On the asymptotic variances of Gaussian Markov random field model hyperparameters in stochastic image modeling", In: Procs. of the 19th International Conference on Pattern Recognition (ICPR), Tampa, pp.1-4, 2008.