SPATIAL-TEMPORAL CONDITIONAL RANDOM FIELD BASED MODEL FOR CROP RECOGNITION IN TROPICAL REGIONS

P. Achanccaray[1], *R. Q. Feitosa* [1,2], *F. Rottensteiner*[3], *I. D. Sanches*[4], *C. Heipke*[3]

[1]Department of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro, Brazil
[2]Department of Computer Engineering, Rio de Janeiro State University, Brazil
[3]Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Germany
[4]Remote Sensing Division, National Institute for Space Research, Brazil

ABSTRACT

This work presents a spatio-temporal Conditional Random Field (CRF) based model for crop recognition from multi-temporal remote sensing image sequences. The association potential at each image site is based on the class posterior probabilities computed by a Random Forest (RF) classifier given the features at the corresponding site. A contrast-sensitive Potts model is used as a label smoothing method in the spatial domain, whereas the interactions in the temporal domain are modeled based on expert knowledge about the possible transitions between adjacent epochs. The CRF based model was tested for crop mapping in two subtropical areas based on a sequences of 9 Landsat and 14 Sentinel-1 images from Ipuã, São Paulo and Campo Verde, Mato Grosso, respectively, two municipalities in Brazil. The experiments showed significant improvements of the accumulated *F1 score* per class against a mono-temporal CRF approach of up to 50% and 75% for a total of 8 and 11 classes using Optical and SAR images respectively.

Index Terms— remote sensing, probabilistic graphical models, crop recognition, Landsat images, Sentinel-1

1. INTRODUCTION

Food security is a major concern worldwide. In order to assure that the food production meets the world population demands at all times, it is important to monitor agriculture activities at a regular basis. Such information can be derived from remote sensing data. With the use of multi-temporal image sequences, it is possible to deal with the data changes that occur as crops evolve through their phenological stages. Indeed, in recent years there has been an increasing interest in novel methods for agricultural land-cover mapping from multitemporal remote sensing image sequences (e.g., [1]).

Conditional Random Field (CRF) approaches are particularly attractive for crop recognition due to their ability to model contextual information, both in the spatial and in the temporal domains. In spite of these benefits, just a few CRF-based approaches have been proposed so far. Hoberg & Müller [2] used CRFs for spatio-temporal crop classification using site wise feature differences in two epochs to model temporal dynamics. In [3], a Dynamic Conditional Random Fields (DCRFs) approach is proposed to learn the phenological information from SAR images. However, all these works refer to agriculture in temperate regions. In the tropics, crop dynamics are more complicated: there are multiple agricultural practices (e.g. irrigation, non-tillage, crop rotation), multiple harvests per year are not uncommon, and phenological cycles of different crops are less coupled to each other than in temperate regions.

This work aims at filling this gap and proposes a CRF based approach for crop recognition in sub-tropical areas. The proposed model is validated upon a sequence of Landsat and SAR images. We analyze how the classification results vary depending on the number and acquisition dates of available images.

2. CONDITIONAL RANDOM FIELDS

2.1. CRF for multitemporal image sequences

CRF can be used to model a sequence of T georeferenced multitemporal images as a graph $G = \{V, E\}$ consisting of a set of nodes V and edges E. Nodes correspond to sites (pixels or segments) of the images. Nodes and their spatial and temporal neighborhoods are illustrated in Figure 1.

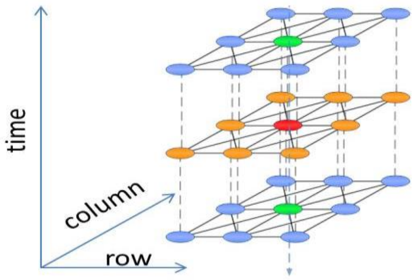


Figure 1. Graph structure comprising an image sequence of $T = 3$ epochs; a site (red) interacts with neighbors in the spatial domain (orange) and in the temporal domain (green).

Let $\boldsymbol{x} = \{\boldsymbol{x}_{i,t}\}_{(i,t)\in V}$ be the feature vectors extracted from the images, where $\boldsymbol{x}_{i,t}$ corresponds to the i-th geographical site in epoch t, for $t = 1, \ldots T$, and $i \in S$, where S is the set of imaged geographical sites. Thus, nodes in V are indexed by the geographical site i and by the time t. Their corresponding labels are given by $\boldsymbol{y} = \{y_{i,t}\}_{(i,t)\in V}$, where $\boldsymbol{y}$ is similarly indexed by the nodes of V and $y_{i,t}$ belongs to a set of classes $L = [l_1, \ldots, l_m]$.

For conciseness, we avoid describing CRF in its full generality and choose a formulation closer to our final model. The posterior probability $P(\boldsymbol{y}|\boldsymbol{x})$ of the labels given the observations is modelled by a CRF that takes the form:

$$P(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{Z}\left[exp\left(\sum_{t=1}^{T}\sum_{i\in S} A^t(y_{i,t}, \boldsymbol{x}_{i,t}) + \right.\right. \\ + \sum_{t=1}^{T}\sum_{i\in S}\sum_{j\in N_i} IS^t(y_{i,t}, y_{j,t}, \boldsymbol{x}_{i,t}, \boldsymbol{x}_{j,t}) \qquad \textbf{(1)} \\ \left.\left. + \sum_{t=1}^{T-1}\sum_{i\in S}\sum_{k\in C_t} IT^{tk}(y_{i,t}, y_{i,k}, \boldsymbol{x}_{i,t}, \boldsymbol{x}_{i,k})\right)\right]$$

where Z, is a normalizing constant called the partition function, $A^t(\cdot)$, $IS^t(\cdot)$ and $IT^{tk}(\cdot)$ are the association, spatial interaction and the temporal interaction potentials, respectively. The association potential $A^t(\cdot)$ measures how likely an image site $(i,t) \in V$ takes a label $y_{i,t}$ given the feature vector $\boldsymbol{x}_{i,t}$. The spatial interaction potential $IS^t(\cdot)$ determines how labels $y_{i,t}$ and $y_{j,t}$ at spatially neighboring sites i and j should interact given the features $\boldsymbol{x}_{i,t}$ and $\boldsymbol{x}_{j,t}$ at both sites in epoch t. N_i is the spatial neighborhood of site i. The temporal interaction potential $IT^{tk}(\cdot)$ models the interaction at one site i in two adjacent epochs, namely t and k. C_t denotes the set of epochs adjacent to t.

2.2. Models for the Potentials

The CRF model outlined in the previous section admits many variants depending on the functions chosen for the potentials. In the present work, the association potential is given by $A^t(y_{i,t}, \boldsymbol{x}_{i,t}) = logP(y_{i,t}|\boldsymbol{x}_{i,t})$, where $P(y_{i,t}|\boldsymbol{x}_{i,t})$ is a local class conditional probability at image site (i,t) given $\boldsymbol{x}_{i,t}$.

The spatial interaction potential $IS^t(\cdot)$, which measures the interaction between labels at spatially neighboring image sites, is modeled by the contrast-sensitive Potts model proposed in [4] and formulated in Equation 2.

$$IS^t(y_{i,t}, y_{j,t}, \boldsymbol{x}_{i,t}, \boldsymbol{x}_{j,t}) = \delta(y_i^t = y_j^t)\left[p + (1-p)e^{-\frac{d_{ij}^2}{2\sigma^2}}\right] \qquad \textbf{(2)}$$

It is based on the dissimilarity given by the Euclidian distance $d_{ij} = \|\boldsymbol{x}_{i,t} - \boldsymbol{x}_{j,t}\|$ of the feature vectors in the same epoch at spatially adjacent sites ($j \in N_i$), σ^2 refers to the mean value of squared feature distances d_{ij}^2 computed during training, and $\delta(\cdot)$ is an indicator function that returns 1 or 0 if its argument is true or false, respectively. The relative influence of the data-dependent and data-independent terms is controlled by parameter $p \in [0,1]$ in Equation 2.

The temporal interaction potential was modeled by Equation 3, where we dropped the dependency on the data by considering a transition matrix TM^{tk} which is constant between epochs t and k. This transition matrix could be estimated by training data; however, it would be necessary to have many samples of each possible transitions, which is not our case. Thus, TM^{tk} is based on expert knowledge containing "1" for possible class transitions between adjacent epochs and "0" otherwise.

$$IT^{tk}(y_{i,t}, y_{i,k}, \boldsymbol{x}_{i,t}, \boldsymbol{x}_{i,k}) = IT^{tk}(y_{i,t}, y_{i,k}) = TM^{tk} \qquad \textbf{(3)}$$

3. EXPERIMENTAL ANALYSIS

3.1. Datasets

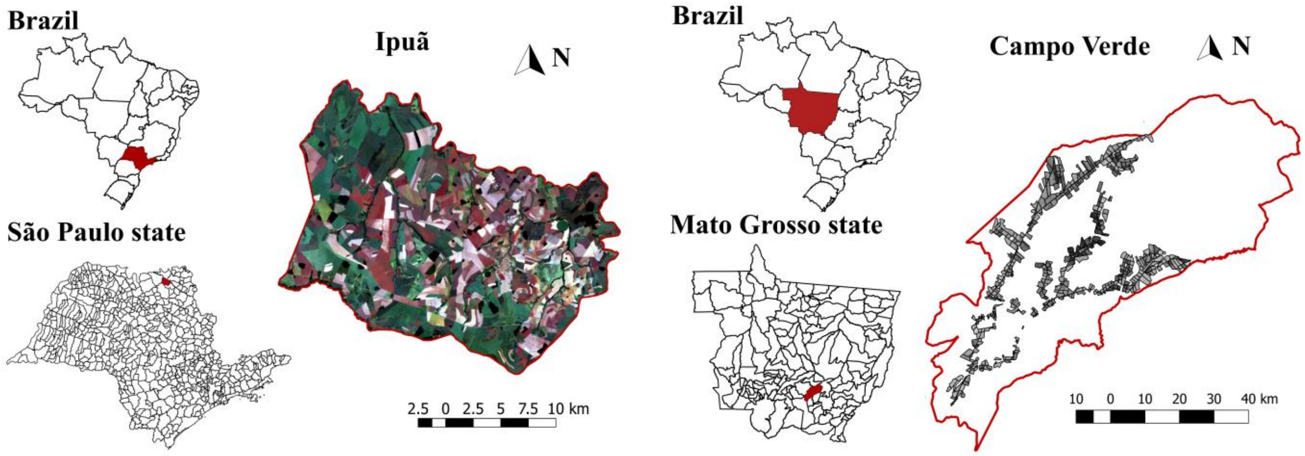
3.1.1. Ipuã

Ipuã municipality in the state of São Paulo, Brazil has an extension of 465 km^2 approximately (see Figure 2a). A sequence of 9 Landsat scenes, from August 2000 to July 2001, was taken, from either Landsat-5 (TM) or Landsat-7 (ETM+) with 30 m spatial resolution, each image having approximately 500K pixels. The reference for each epoch was produced manually by a human expert.

The most common crops are *Sugarcane*, *Soybeans* and *Maize*. In our study, we also included two classes related to no crops: *Prepared Soil*, which corresponds to ploughing and soil grooming phases, and *Post-Harvest*, characterized by vegetation residues lying on the ground. To complete the set of classes, *Pasture*, *Riparian Forest* and *Others* were also included in our model. The last one represents minor crops as well as rivers and urban areas. Distribution of samples per class could be found in Figure 3a.

3.1.2. Campo Verde

Campo Verde municipality in the state of Mato Grosso, Brazil has an extension of 4782 km^2 approximately (see Figure 2b). A total of 27 Level 1 Interferometric Wide Swath (IWS) mode Ground Range Detected (GRD) Sentinel-1 products in VV and VH polarizations were used to cover all the Campo Verde municipality from October 2015 to July 2016 resulting in a sequence of 14 images, with almost two images per month. These images were geometrically corrected using a Range Doppler terrain correction with a Digital Elevation Model from SRTM, radiometrically calibrated to a backscatter coefficient (sigma nought (σ^0) in our case), converted to *db*, co-registered using a RapidEye mosaic (5 m spatial resolution) and georeferenced to UTM projection Zone 21S and Datum WGS84. Distribution of samples per class could be found in Figure 3b.



(a) (b)

Figure 2. (a) Ipuã municipality in São Paulo state and (b) Campo Verde municipality in Mato Grosso state.

The main crops found in this area are: *Soybean*, *Maize* and *Cotton*. Also, there are some minor crops as *Beans*, *Sorghum* and non-commercial crops (*NCC*) such as *Millet*, *Brachiaria* and *Crotalaria*. Other classes considered are *Pasture*, *Eucalyptus*, *Soil*, *Turfgrass* and *Cerrado*.

3.2. Experimental Protocol

The pixel-wise feature vectors $\boldsymbol{x}_{i,t}$ consisted of the spectral values directly observed at the image site $\boldsymbol{x}_{i,t}$ and the NDVI derived from the spectral values for the Landsat Images. For the Sentinel-1 images, texture attributes (mean, variance, correlation and homogeneity) were extracted from the Gray Level Co-occurrence Matrix (GLCM) in 3 × 3 windows in 4 directions (0, 45, 90 and 135 degrees).

The optimal label configuration $\boldsymbol{y}$, the one that maximizes the posterior probability in Equation 1, was computed using Loopy Belief Propagation (LBP) [5], which produces approximate solutions for graphs with cycles.

For each dataset, we used approximately 20% of the data for training and 80% for testing. Two sequences were extracted from each dataset. For Ipuã, a sequence from February to April 2001 and another one from August 2000 to July 2001 were analyzed due to the presence of *Maize* and *Sugarcane*, respectively. In Campo Verde, sequences from November 2015 to February 2016 and from March 2016 to July 2016 were considered to analyze the accuracy obtained for *Soybean* and *Maize & Cotton* respectively. In each sequence, we measured the accuracy on the last epoch of the sequence, which was increased by adding successively images of earlier epochs. For a sequence length equal to one, only the association and spatial interaction potentials are considered, which corresponds to a mono-temporal CRF. As accuracy metric, the *F1 score* was selected, which is the harmonic mean between *Precision* and *Recall*.

4. RESULTS AND DISCUSSION

The results obtained in our experiments are summarized in Figure 3, which shows the accumulated *F1 score* per class for each sequence in both datasets, Ipuã (Figure 3c and 3e) and Campo Verde (Figure 3d and 3f), always classifying the last image in the sequence.

In both sequences for Ipuã, there were improvements in the accumulated *F1 score*, approximately 30% in the first sequence and 50% in the second one, as more images were considered, especially for *Maize* in the first sequence and for *Sugarcane* in both sequences. *Sugarcane*´s *F1 score* increases regularly until a sequence length of 6. After that, there was no significant improvement. This is related to the gap in the acquisition dates in the sequence (from October 2000 to February 2001) and the decrease in the number of *Sugarcane* samples according to the class distribution per epoch (see Figure 3a).

Similarly, in both sequences for Campo Verde, there were improvements of 50% and 90% in the accumulated *F1 score* for the first and second sequence respectively. Some classes present a very low *F1 score*, mainly because same classes are under-presented in the dataset. Though over- and under-sampling has been applied to compensate for this problem, the accuracies for these classes remained low.

5. CONCLUSIONS

This work presented a spatio-temporal Conditional Random Field approach for crop recognition. The model was evaluated on two datasets comprising 9 Landsat images and 14 Sentinel-1 images of sub-tropical regions in Brazil. The inclusion of the temporal interaction led to an increase of up to 50% and 90% in accumulated *F1 score* compared to mono-temporal spatial context-based classification for Optical and SAR images respectively, demonstrating the effectiveness of considering the temporal context.

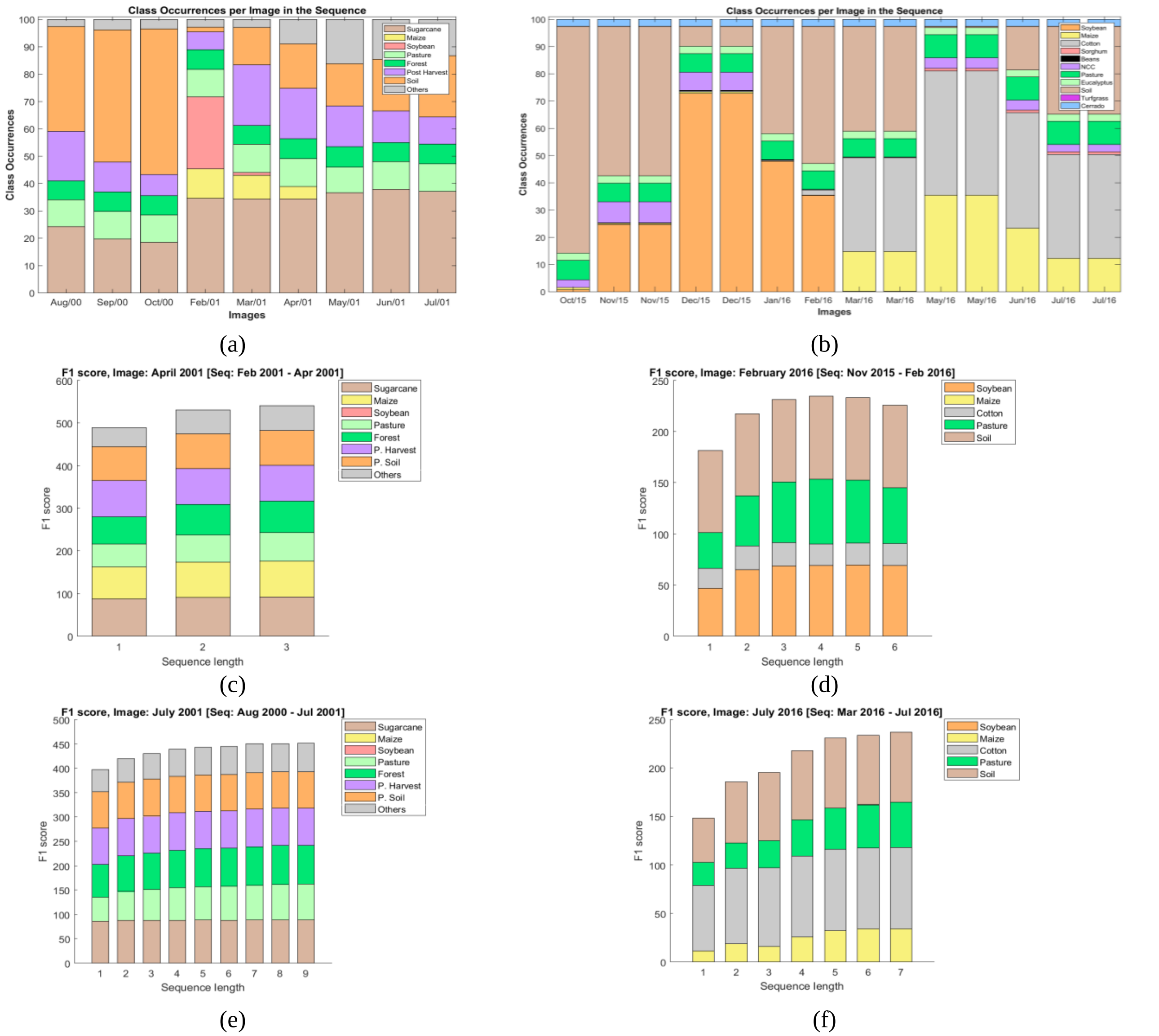


Figure 3. Distribution of samples per class per epoch for both datasets, Ipuã (a) and Campo Verde (b), *F1 score* accumulated per class obtained for sequences considered in Ipuã (c) and (e) and for Campo Verde (d) and (f).

6. REFERENCES

[1] R. Sonobe, H. Tani, X. Wang, N. Kobayashi and H. Shimamura, "Discrimination of crop types with TerraSARX-derived information," *Physics and Chemistry of the Earth, Parts A/B/C,* Vols. 83-84, pp. 2-13, 2015.

[2] T. Hoberg and S. Müller, "Multitemporal Crop Type Classification using Conditional Random FIelds and RapidEye data," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vols. XXXVIII-4, no. W19, pp. 115-121, 2011.

[3] B. K. Kenduiywo, D. Bargiel and U. Soergel, "Crop type mapping from a sequence of TerraSAR-X images with Dynamic Conditional Random Fields," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. III, no. 7, pp. 59-66, 2016.

[4] J. Shotton, J. Winn, C. Rother and A. Criminisi, "Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context," *International Journal of Computer Vision*, vol. 81, no. 1, pp. 2-23, 2009.

[5] B. J. Frey and D. J. c. MacKay, "A Revolution: Belief propagation in Graphs with Cycles," *Advances in neural information processing systems*, vol. 10, pp. 479-485, 1998.