

Conditional random fields for land use/land cover classification and complex region detection

Gulcan Can^{1,*}, Orhan Firat^{1,*}, and Fatos T. Yarman Vural¹

¹ Department of Computer Science, Middle East Technical University, Ankara, Turkey
{gulcan, orhan.firat, vural}@ceng.metu.edu.tr

Abstract. Developing a complex region detection algorithm that is aware of its contextual relations with several classes necessitates statistical frameworks that can encode contextual relations rather than simple rule-based applications or heuristics. In this study, we present a conditional random field (CRF) model that is generated over the results of a robust local discriminative classifier in order to reveal contextual relations of complex objects and land use/land cover (LULC) classes. The proposed CRF model encodes the contextual relation between the LULC classes and complex regions (airfields) as well as updates labels of the discriminative classifier and labels the complex region in a unified framework. The significance of the developed model is that it does not need any explicit parameters and/or thresholds along with heuristics or expert rules.

Keywords: conditional random fields, land use/land cover, complex region detection, satellite imagery

1 INTRODUCTION

Visual patterns and object occurrences in remote sensing images exhibit high intra-class variance, meaning that two or more instances of the same object or object groups may look coercively different. For example, two airfields may have entirely different color structures, composing roads, shapes, sizes and configurations of their sub-parts (e.g. one may have just one, crossing, parallel runway(s) having hammer shaped, circular, polygonal dispersal areas and located in sandy, snowy, coastal or urban terrain). Occasionally these objects and object groups may even look more similar to instances within other classes than to instances within their own class, e.g. circular oil tanks of a refinery and circular dispersal area of a military airfield.

Contextual models are significantly useful in order to handle the huge variability within classes in the image because of their expressive representations. By forming a contextual framework, any object can be accurately classified not only by considering its low-level vision features but also its local context (spatial relations) over a probabilistic graphical model. Figure 1 illustrates some real world examples that exhibit high intra-class variance.

*corresponding author



Fig. 1. Examples of airfields from satellite images. Despite the high variance in color, texture, deployment terrain and composition, we can still recognize them as remotely sensed views of airfields.

The ability to recognize such complex objects comes from both appearance cues of the object itself and contextual relations of the complex objects with their surroundings. These contextual relations can be defined as co-occurrence frequency of other classes in a predefined neighborhood of the complex object. For instance, if the complex object in consideration is an airfield, we would expect urban, vegetation, water existence nearby to be less than a certain ratio. This information comes either from domain knowledge or by explicit observations. However, deciding this ratio by a static threshold is not desirable, since less likely configurations are not allowed at all. As in the case of urban areas in the surrounding of an airfield, we may set a 20% urban co-occurrence threshold in a 300-meters neighborhood by domain knowledge. Yet, Figure 1 demonstrates cases contradicting with such a threshold. Hence rather than determining crisp thresholds for the recognition task of a complex object, constructing a probabilistic model is much more flexible and suitable.

Probabilistic graphical models are the state-of-art approach for modeling contextual relations between semantic classes [1] and have many applications in remote sensing [2-3]. Since labels in spatial data are not independent as well as observations, assumptions on data being “independent and identically distributed” (i.i.d.) is violated by using traditional classifiers. Therefore such classifiers may produce undesirable results when applied to such data.

This problem motivates the use of Markov Random Fields (MRFs) and more recently Conditional Random Fields (CRFs) for spatial data. In the proposed approach, contextual relations between a complex object and its surroundings, which is characterized by LULC classes, are modeled within a CRF framework. The major contribution of the proposed model is that a random field is constructed over semantic classes rather than pixels or super-pixels as in the literature. Our model aims to correctly identify the complex object by recognizing the co-occurrence pattern of all other classes in its surrounding as well as updating previously assigned class labels which can be obtained by any kind of classification method.

The rest of this paper is organized as follows: In section 2, recent motivating studies in the field of probabilistic graphical models are stated, followed by section 3, the adopted methods are explained. In section 4, proposed algorithm is introduced and in section 5 dataset and experiments are described. Finally in section 6 some conclusion remarks and future work are given.

2 RECENT STUDIES

Studies modeling spatial structures vary both in their representations used to encode the spatial information and their approaches for learning. Inference on the generated graphical models depends on the model selection in the studies and may be thought as a representation dependent step.

One of the pioneering studies employs both contextual and hierarchical representations with a *relationship learning* process and Bayesian inference algorithm is proposed by Porway et. al [4]. Their approach presents a grammar-based hierarchical and contextual model for object recognition. This grammar-based model combines a stochastic context free grammar (SCFG) [5] with a Markov Random Field (MRF) to capture both local and global context and combines bottom-up information with top-down knowledge. They represent the frequency of occurrence and type of object parts with a SCFG and model the spatial and appearance relationships between them using MRFs, thus create a constrained grammar that can represent a huge number of instances for a single category. Another contribution of this study is that, this contextual and hierarchical model learns statistical constraints on the appearances and relationships between different parts of the image classes with a minimax entropy framework [6]. This framework selects the set of contextual relationships necessary for modeling the object class; begins with a large set of relationships that could potentially exist between parts, then iteratively selects only those relationships that help the model best match true statistics for that image class. They separated hierarchy into two sets for objects and scene which enables to plug-in any object detection algorithm for bottom-up detection procedure. They employed compositional boosting [7] for some specific bottom-up proposals.

In [8], a region and object based model for object-detection is proposed through a hierarchy of CRFs. In the bottom level, a CRF is comprised of pixels as probabilistic graphical model nodes and features are extracted in pixel level accordingly, a unified energy function made it possible to incorporate bottom-middle and top level random fields. In the middle level, segments are formed as the model nodes and contextual

relations between segments are revealed with region statistics. Finally as the top-most level of the proposed hierarchical graphical model, segments and objects are connected to each other and contextual relations between objects are tried to be extracted from positional relations of the objects both considering segment level interactions at once. The model employed for this graphical model is a conditional MRF (CRF) that is trained by labeled images from both levels with logistic regression and inference is conducted by use of hill-climbing.

Jiang et.al. propose a *context based concept fusion* model for semantic concept detection [9]. In this study, posterior probabilities for several classifiers are fed to a CRF model for generating updated posterior probabilities through a fully-connected CRF where each node represents a concept. This corresponds to class labels in our case.

Lee et.al. propose a model, namely support vector random fields (SVRF), which combines the ability of CRFs to model different types of spatial dependencies and the appealing generalization properties of support vector machines (SVMs) [10]. Their approach employs an observation-matching potential by changing the association potential in CRF model. Therefore they combined the discriminative classification power of SVMs with spatial context encoding power of CRFs.

3 METHODS ADOPTED

SVM is a supervised classification approach which makes use of kernels (mapping functions) and sparsity [11]. By the help of kernels, samples in n -dimensional space are carried to a higher dimensional Hilbert space, therefore samples become linearly separable by a hyper plane. In this model a hyper plane that separates nearest samples from different classes with maximum margin is selected and named as support vectors. Tolerance and cost parameters can be used to allow or penalize outliers. SVM is also known as max-margin classifier.

As stated previously, spatial relations between neighboring pairs can be modeled by MRFs and CRFs. More specifically, the class labels can be assigned by maximum a posterior (MAP) estimation in image classification task as

$y_{MAP} = \operatorname{argmax}_y P(y|x)$. This can be interpreted as CRF framework that models directly the posterior probability of labels given the observed data. Consequently, besides the contextual in-formation in labels, the CRF framework has ability to capture the contextual infor-mation in observed data.

The discriminative CRF framework considers Markovian property of y conditioned on x and directly models the posterior as a Gibbs distribution with the following form:

$$P(y|x, \theta) = \frac{1}{Z} \exp \left\{ - \sum_{c \in C} \varphi_c (y_c, x, \theta) \right\} \quad (1)$$

where $Z = \sum_y \exp \{-\sum_{c \in C} \varphi_c(y_c, x, \theta)\}$ is partition function (normalization constant), φ_c is potential defined on clique c with parameters θ , C is set of cliques, and y_c is set of labels over clique c . Then the pair-wise CRF models can be written as

$$P(y|x, \theta) = \frac{1}{Z} \exp \left\{ -\sum_{i \in S} \varphi_i(y_i, x, w) - \sum_{i \in S} \sum_{j \in \eta_i} \varphi_{ij}(y_i, y_j, x, v) \right\} \quad (2)$$

where η_i is the set of neighbors of site i , φ_i and φ_{ij} are the unary and pair-wise clique potentials with parameters w and v , respectively, then θ denotes the parameter set $\theta = \{w, v\}$. The unary potential φ_i represents the association of a single site to semantic labels, whereas pair-wise potentials φ_{ij} can be seen as a measure of how the labels at neighboring sites at i and j should interact given image x . In fact, the unary and pair-wise potentials in CRF should be designed as discriminative as possible according to the domain it is applied to.

In our approach, we construct a fully connected random field over classes similar to concepts proposed by Jiang et.al. SVM is an intermediate step to assign class labels to segments. We could have applied other methods here, but preferred SVM due to its high performance. With this step, initial class map for LULC classes is obtained and they are then updated in the proposed CRF model according to context of the complex object.

4 PROPOSED ALGORITHM

In this study, *airports* are chosen as the complex regions and *water, forest, greenland, urban, concrete, soil* as LULC classes. The most significant cue of an airfield is the existence of runway(s), which consists of basically long straight parallel lines. We propose an algorithm for categorization of Parallel Line Bounded Regions (PLBR) which is stated as a strong indicator and invariant of airfields in highly variant contextual environments [12]. In the proposed algorithm, context information of airfields is formulated over LULC classes. Proposed model is also for updating labels assigned to LULC classes in the means of airport context.

Proposed model is a fully-connected conditional random field which can be seen in Figure 2. Fully-connected graphical model is selected in order to reveal contextual relations between all classes and their mutual influence.

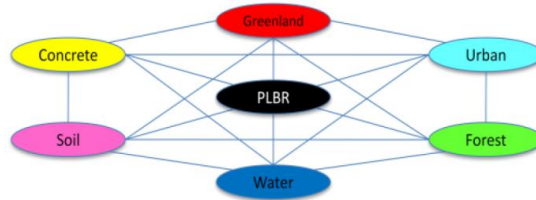


Fig. 2. Proposed CRF model, which is a fully connected graph aiming to capture all possible pairwise relations between semantic categories.

Figure 3 illustrates workflow of the proposed algorithm. In the proposed algorithm, the first step is the preprocessing step in which segmentation of the input image via the mean-shift algorithm [13] and PLBRs extraction. The PLBRs are extracted by finding the line segments [14] on the steerable-filter [15] response of the image first and then extracting the parallel ones. Note that PLBRs are treated as regular segments.

Pre-processing step is followed by a feature extraction step. For each segment, features are extracted as illustrated in Figure 4. For this purpose, fundamental maps are obtained first. These are spectral values (red, green, blue, near-infrared), DTED map, Gabor filter response, normalized difference water index (NDWI) map [16] and normalized difference vegetation index (NDVI) map calculated using (3) and (4) respectively.

$$NDWI = \frac{Green - NIR}{Green + NIR} \quad (3)$$

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (4)$$

Gabor response image is obtained by taking maximum response of a pixel at eight directions and one scale.

An initial labeling for LULC classes over segments is obtained with an SVM classifier by using features extracted. SVM is trained with a labeled dataset and the parameters are determined using cross-validation in validation set. Note that, the initial labeling in this step is conducted only for segments, rather than PLBRs, with considering only LULC classes (6 classes).

For each PLBR, a fully-connected CRF model as depicted in Figure 2 is constructed. No further operation is conducted over a PLBR before embedding them into this CRF model as a node. For each PLBR, the 300-meter neighborhood is analyzed in the initial class label map obtained by SVM classifier. All segments with the same class label are treated as one single node in the CRF model of the corresponding PLBR. As an example, in Figure 5, there are several segments initially labeled as urban class (cyan) by SVM. For the Urban node in the CRF model, we extract unary potentials considering all these urban segments.

After obtaining the CRF model for a PLBR, node and edge features are extracted for all seven nodes. Node features for a node in the CRF model are extracted as in Figure 4, but this time; extraction is conducted over all segments of the corresponding 7 class label not separately but as a whole (e.g. mean and standard deviation of urban class in Figure 5, is computed over all cyan area and used as the node features of the Urban node).

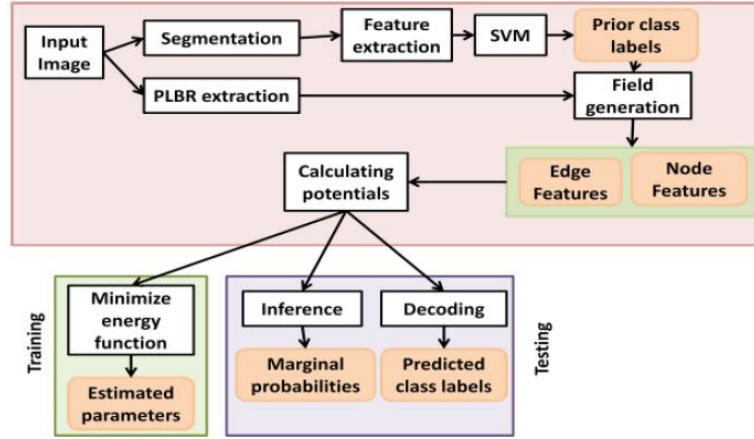


Fig. 3. Workflow of the algorithm

For the edge features, overlapping, adjacent and in-neighborhood class frequencies are used (see Figure 5). These are calculated based on pixel counts of each class in corresponding area.

For the CRF framework to be function reasonably, node features and edge features are converted into unary and pair-wise potentials using multi-class logistic function. Then potentials are fed to energy function and parameters are learnt for the model by minimizing this energy function (2). In test phase, CRF model for each PLBR is decoded according to trained parameters.

During parameter estimation in training phase, L-BFGS is used. It is a limited-memory quasi-Newton method for unconstrained optimization. Two distinct loss functions are applied during training, namely loopy belief propagation (LBP) and pseudo negative log-likelihood (NLL). As the decoding method, Iterated Conditional Modes (ICM) is used [17].

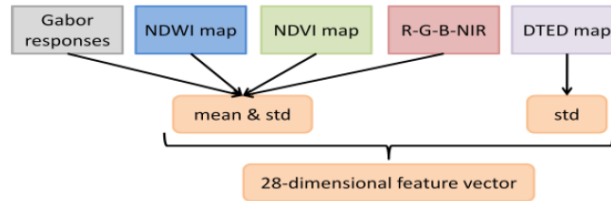


Fig. 4. Feature extraction

5 DATASET AND EXPERIMENTS

In this study, 4 GEOEYE multispectral images with size of $\sim 3800 \times 3800$ pixels are used. Each image contains between 23 and 89 PLBRs either corresponds to an air field or one of the LULC classes. 112 PLBRs in two images are used for training having 53 of them being actually airfield and 77 PLBRs in remaining two images are

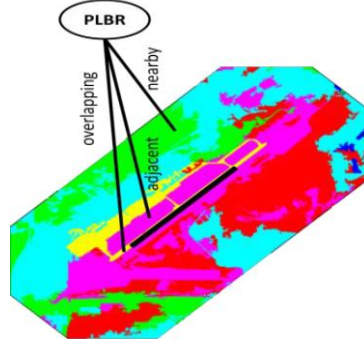


Fig. 5. Proposed edge features over SVM labels

used for testing having 49 of them actually airfield. Ground truths are prepared by labeling segments around each PLBR region.

There are seven classes in this study, namely *water*, *urban*, *forest*, *green land*, *soil*, *concrete*, and *airfield*. In our experiments, we used an eighth state, an extra state to enable the model to reject. Since we combine segments according to their prior labels, mixing cases may confuse the CRF model. Eighth state corresponds to "other" or "mixing" class.

As performance measures, recall and precision are used for airfield detection.

Table 1. Performance results of the proposed algorithm using ICM decoding

Loss Function	Pseudo Negative Log-likelihood	92	46.94
	Negative Log-likelihood with LBP	93.33	57.14
		Precision	Recall

6 CONCLUSION AND FUTURE WORK

The proposed model categorizes PLBRs and corrects/updates SVM results in the context of an airport-PLBR. The actual goal of this paper is focusing on detecting actual complex objects, airfields in our case, accurately. Recall values in Table 1 demonstrate that it is a hard task. PLBRs could be decoded as urban or forest-green land which may semantically correspond to roads and edges between green lands respectively. Other studies about airfield detection in the literature have specific datasets and employ generally heuristic or threshold-based approaches. This makes comparison with our approach inapplicable.

Instead of using SVM output of training images for training the CRF model, ground truth of training images can be directly employed. This would probably be better for CRF to learn true relationships among classes. However, in this case, train and test phases would have different steps, since SVM would not be involved in training phase to obtain areas under each class node.

The proposed CRF model is able to embed spatial information around complex objects in terms of LULC classes; however it is open to improvements. The model can

be designed as a star shape to make pair-wise relations more representative. For the CRF model corrects SVM results in segment level, not as a whole semantic class, segments layer can be added to the model which is connected to class nodes in the originally proposed model. This hierarchical model would make our model more relaxed and capable. However, computational tractability would be lost in this dense graph structure and only approximate methods could be applied.

7 REFERENCES

1. Galleguillos, C. & Belongie, S. Context based object categorization: A critical survey. *Computer Vision and Image Understanding* **114**, 712-722 (2010).
2. Benedek, C. & Sziranyi, T. Change Detection in Optical Aerial Images by a Multilayer Conditional Mixed Markov Model. *IEEE Transactions on Geoscience and Remote Sensing* **47**, 3416-3430 (2009).
3. Zhong, P. & Wang, R. Learning conditional random fields for classification of hyperspectral images. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society* **19**, 1890-907 (2010).
4. Porway, J., Wang, Q. & Zhu, S.C. A Hierarchical and Contextual Model for Aerial Image Parsing. *International Journal of Computer Vision* **88**, 254-283 (2009).
5. Lin, L., Wu, T., Porway, J. & Xu, Z. A stochastic graph grammar for compositional object representation and recognition. *Pattern Recognition* **42**, 1297-1307 (2009).
6. Zhu, S.C., Wu, Y.N. & Mumford, D. Minimax Entropy Principle and Its Application to Texture Modeling. (1997).
7. Wu, T.-fu, Xia, G.-song & Angeles, L. Compositional Boosting for Computing Hierarchical Image Structures. *Learning*
8. Gould, S. Probabilistic models for region-based scene understanding. (2010).
9. Jiang, W., Chang, S.-F. & Loui, A.C. Context-Based Concept Fusion with Boosted Conditional Random Fields. *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07* **1**, I-949-I-952 (2007).
10. Lee, C.-hoon, Greiner, R. & Schmidt, M. Support vector random fields for spatial classification. *IN PROC. OF PKDD 2005* 121 - 132 (2005).
11. Cortes, C. & Vapnik, V. Support-vector networks. *Machine Learning* **20**, 273-297 (1995).
12. Firat, O., Tursun, O.T. & Yarman Vural, F.T. Application of Context Invariants in Airport Region of Interest Detection for Multi-spectral Satellite Imagery. *IEEE 20th Conference on Signal Processing and Communications Applications (SIU)* (2012).
13. Comaniciu, D., Meer, P. & Member, S. Mean shift: A robust approach toward feature space analysis. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE* **24**, 603 - 619 (2002).
14. Grompone von Gioi, R., Jakubowicz, J., Morel, J.-M. & Randall, G. LSD: a fast line segment detector with a false detection control. *IEEE transactions on pattern analysis and machine intelligence* **32**, 722-32 (2010).
15. Freeman, W.T. & Adelson, E.H. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13**, 891-906 (1991).
16. McFEETERS, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing* **17**, 1425-1432 (1996).
17. Besag, J. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society. Series B (Methodological)* **36**, 192 - 236 (1974).