Technical Report

Semantic Segmentation for Aerial Images using RF and a full-CRF

Nguyen Tien Quang[1], Nguyen Thi Thuy[2], Dinh Viet Sang[1] and Huynh Thi Thanh Binh[1]

[1]MSO-Lab, Ha Noi University of Science and Technology, Vietnam

[2]Dept. of Computer Science, Vietnam National University of Agriculture

05-2015

Introduction

This report presents a framework for semantic segmentation of aerial images. We propose an effective image segmentation system using some simple features, a random forest (RF) classifier and a full-connected Conditional Random Field (full-CRF) model. The system effectively exploits contextual information from color and position features in combination with unary potential built from the output of random forest classifier. The model is applied to the ISPRS 2D semantic labelling challenge dataset. We evaluate the results and show the competitive segmentation accuracy.

Method

The figure below summarizes the proposed framework

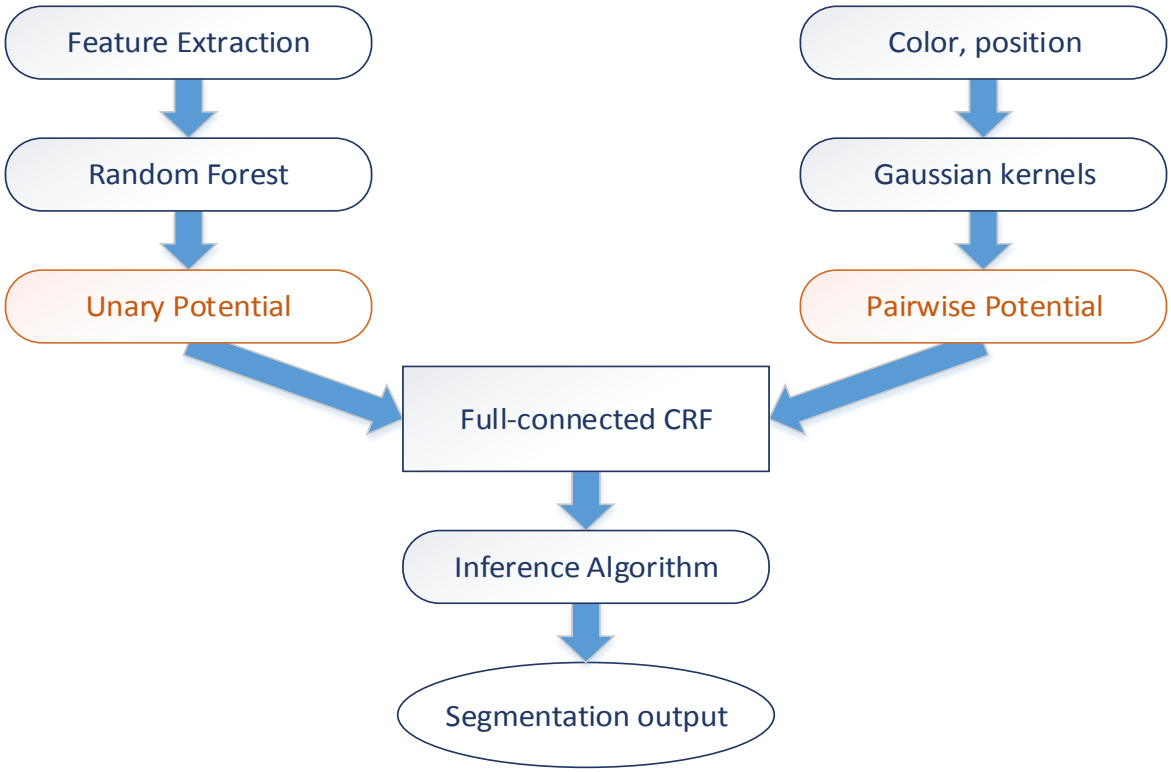


Figure 1. Semantic Segmentation full-CRF Model

1 - *Feature extraction:* We use the following features for the description of image data.

- NDVI: the normalized digital vegetation index, computed from the first (IR) and the second channels (R) of the CIR true ortho photo (TOP)

$$NDVI = \frac{IR - R}{IR + R}$$

The use of the NDVI is based on the fact that green vegetation has low reflectance in the red spectrum (R) due to chlorophyll and much higher reflectance in infrared spectrum (IR) due to its cell structure. Hence, this is a good feature to distinguish green vegetation from other classes.

- NDSM: the difference between the DSM and the derived DTM, which classifies pixel into ground and off-ground.

$$NDSM = DSM - DTM$$

This feature helps to distinguish the high object classes from the low object classes.

- Texton: Texton is a unit of texture, reflecting the human perception of textured images. It has been proven to be effective in image segmentation. Therefore, representing images in the form of texton, the pixels will contain more useful information than in the form of normal color [3].

- Color: In this work we use the CIELab color space. Unlike the RGB and CMYK color models, Lab color is designed to approximate human vision. It aspires to perceptual uniformity, and its L component closely matches human perception of lightness.
- Saturation of CIR image: some previous works have shown that the saturation is helpful to further support the separation of vegetation and impervious surfaces.
- Entropy gathered over a 9 × 9 neighborhood from the DSM to exploit spatial context information of a pixel (neighboring).

2 - *Random forest classifier:* With those extracted features, we used random forest classifier to train and build unary potential for CRF models. Random forest used in this work is Breiman's CART-RF [2], implemented parallel in R language.

The training algorithm for random forests applies the general technique of bootstrap aggregating (bagging), to tree learners. Given a training set $I = i_1, \dots, i_n$ where i_j is a feature vector at pixel j, with responses $X = x_1, \dots, x_n$ where $x_j \in L\{1 \dots l\}$, bagging repeatedly selects a random sample with replacement of the training set and fits trees to these samples:

for $b = 1, \dots, ntree$ **do**

Sample with replacement, n training examples from (I, X); call these (I_b, X_b).

Train a classification tree f_b on (I_b, X_b).

endfor

After training, predictions for unseen samples i' can be made by averaging the predictions from all the individual classification trees on i':

$$\hat{f} = \frac{1}{ntree} \sum_{b=1}^{ntree} f_b(i')$$

It means to take the majority votes in the case of classification trees.

The use of random forests has several advantages including: the computational efficiency in both training and classification, the probabilistic output, the seamless handling of a large variety of visual features and the inherent feature sharing of a multi-class classifier.

However, by using this technique the image pixels are labeled independently without regarding interrelations between them. Therefore, in the later process, we can further

improve the segmentation results by employing an efficient inference model (CRF) that can exploit the interrelations between image pixels.

3 - *Conditional random field model:* Following the standard definition of image labelling using CRFs, the energy function consists of unary and pairwise potential terms:

$$E(x) = \sum_{i \in V} \psi_u(x_i) + \sum_{(i,j) \in E} \psi_p(x_i, x_j)$$

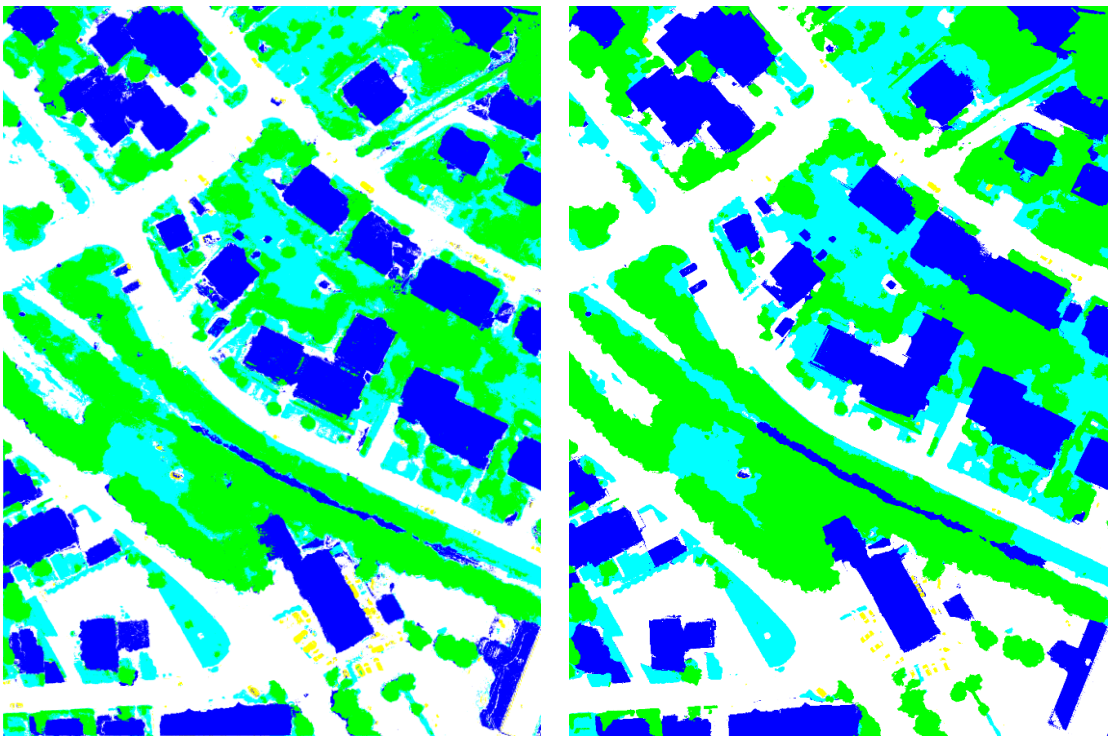
where V and E are nodes and edges of the CRF graph. We use a complete graph to exploit more effectively pairwise information of the pixels.

The unary potential $\psi_u(x_i)$ is computed by a random forest classifier as described above.

The pairwise potential $\psi_p(x_i, x_j)$ is built using a linear combination of Gaussian and sensitive Potts model over the information of color and position [1].

$$\psi_p(x_i, x_j) = \mu(x_i, x_j) \sum_m w^{(m)} k^{(m)}(f_i, f_j)$$

Due to the complexity of billions of edges in the graph, a Mean-Field approximate inference algorithm [1] is used to predict the labels of pixels. The Mean-Field approximation computes a distribution $Q(X)$ that minimizes the KL-divergence $\boldsymbol{D}(Q \parallel P)$ among all distributions Q that can be expressed as a product of independent marginals, $Q(X) = \prod_i Q_i(X_i)$. This approximation can be performed in $O(N)$ time. The inference takes about 30 to 40 seconds per image on a single CPU. The main improvements of the full-CRF are to change the label of regions with ambiguous probabilities based on the interrelations between image pixels, and to remove small noise regions. See the figure below for an illustration.



(a) Segment results given by RF (b) Improvement of (a) by a fullCRF

Figure 2. Improving the accuracy of full-CRF

Result

↓*Predict* *Reference* →	*Imp_surf*	*Building*	*Low_veg*	*Tree*	*Car*
Imp_surf	**93.1**	3.6	2.3	1.0	0.0
Building	7.0	**91.5**	0.5	0.8	0.1
Low_veg	10.1	2.9	**72.2**	14.8	0.0
Tree	2.1	0.8	8.0	**89.1**	0.0
Car	69.9	11.0	1.0	0.3	**17.8**
Precision	81.4	92.4	85.5	84.7	79.5
Recall	93.1	91.5	72.2	89.1	17.8
F1	**86.9**	**92.0**	**78.3**	**86.9**	**29.0**
Overall	***85.9***				

Reference

[1] P. Krahenbuhl, V. Koltun. Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. *NIPS, 2011*.

[2] L. Breiman. Random forests. *Machine Learning, 45(1): 5–32, 2001.*

[3] J. Winn, A. Criminisi, and T. Minka. Categorization by learned universal visual dictionary. *International Conference on Computer Vision, volume 2, pages 1800–1807, Beijing, China, October 2005.*