Classification of Multitemporal Remote Sensing Data Using Conditional Random Fields

Thorsten Hoberg, Franz Rottensteiner, Christian Heipke
IPI, Institute of Photogrammetry and GeoInformation, Leibniz Universitaet Hannover, Germany
(hoberg, rottensteiner, heipke)@ipi.uni-hannover.de

Abstract

Land cover classification plays a key role for various geo-based applications. Many approaches for the classification of remote sensing data assume the features of neighboring image sites to be conditionally independent. However, using spatial and temporal context information may enhance classification accuracy. Conditional Random Fields (CRF) have the ability to model dependencies not only between the class labels of neighboring image sites, but also between the labels and the image features. In this work we present a novel approach for multitemporal classification in high resolution satellite imagery using CRF that is based on an extension of the CRF model by a time-dependant component. The potential of our approach is demonstrated using a set of two Ikonos and one RapidEye scenes of a rural area in Germany.

1. Introduction

With the increasing number of high resolution optical remote sensing satellites there is a higher availability of multitemporal image data. These data can be used for enhancing the classification accuracy and for analyzing land cover changes. However, up to now most approaches for multitemporal land cover analysis do not make use of temporal dependencies, but derive their results by some kind of difference measure between the monotemporal classification results of different epochs [1]. In contrast [2] models temporal dependencies by Markov chains for detecting land cover transitions in Landsat images, but spatial context is not taken into account. Bruzzone et al. [3] use a cascade of three multitemporal classifiers, one of them considering the k nearest neighbors of each pixel. In [4] the Markov Random Field (MRF) framework is extended by a temporal energy term based on a transition probability matrix in order to improve the classification results for two consecutive images. The interaction between neighboring image sites (pixels or segments) in MRFs is restricted to the class labels, whereas the features extracted from different sites are assumed to be conditionally independent.

This restriction is overcome by Conditional Random Fields (CRF) [5]. They provide a discriminative framework that can also model dependencies between the features from different image sites and interactions between the labels and the features. In remote sensing CRF have been used for the classification of settlement areas in high-resolution optical satellite images [6, 7] and for generating a digital terrain model from LiDAR [8].

In this work we present a novel approach for multitemporal classification of high resolution optical remote sensing data that is based on an extension of the CRF concept by an additional temporal interaction potential. Using this temporal model, we classify a set of M multispectral images simultaneously, which should increase the accuracy and reliability of the results while still allowing for changes in land cover between the individual epochs. No existing land cover map is required. In contrast to [6, 7] we consider a multi-class problem. The approach is demonstrated for a set of three multispectral images.

2. The Conditional Random Field model

In many classification algorithms the decision for a class at a certain image site is just based on information derived at the regarded site, where a site might be a pixel, a square block of pixels in a regular grid or a segment of arbitrary shape. In fact, the class labels and also the data of spatially and temporally neighboring sites are often similar or show characteristic patterns, which can be modeled using CRF. In monotemporal classification, we want to determine the vector of class labels **x** whose components x_i correspond to the classes of image site $i \in S$ from the given image data **y** by maximizing the posterior probability $P(\mathbf{x} \mid \mathbf{y})$ [5]:

$$P(\mathbf{x}|\mathbf{y}) = \frac{1}{Z} exp\left(\sum_{i\in S} A_i(x_i,\mathbf{y}) + \sum_{i\in S}\sum_{j\in N_i} I_{ij}\left(x_i,x_j,\mathbf{y}\right) \right) \qquad (1)$$

In Equation 1, N_i is the spatial neighborhood of image site i (thus, j is a spatial neighbor to i), and Z is a normalization constant. The *association potential* A_i links the class label x_i of image site i to the data **y**, whereas the term I_{ij}, called *interaction potential*, models the dependencies between the labels x_i and x_j of neighboring sites i and j and the data **y**. The model is very general in terms of the definition of the functional model for both A_i and I_{ij}; refer to [5] for details.

In the multitemporal case, we have M co-registered images. The components of the image data vector **y** are site-wise data vectors $\mathbf{y}_i$ consisting of M components $\mathbf{y}_i^t$, where $\mathbf{y}_i^t$ is the vector of the observed pixel values at image site i at epoch $t \in T$ and $T = \{1,\dots M\}$. The components of **x** are vectors $\mathbf{x}_i = [x_i^1, \dots x_i^M]^T$, where x_i^t describes the class of image site i at epoch $t \in T$. For each image site we want to determine its class x_i^t for each time t from a set of pre-defined classes. In order to model the mutual dependency of the class labels at an image site at different epochs, the model for $P(\mathbf{x} \mid \mathbf{y})$ in Equation 1 has to be expanded:

$$\begin{aligned} P(\mathbf{x}|\mathbf{y}) = \ & \frac{1}{Z} exp\Bigg[\sum_{i\in S}\sum_{t\in T} A_i^t(x_i^t,\mathbf{y}^t) + \sum_{i\in S}\sum_{t\in T}\sum_{j\in N_i} IS_{ij}^t\left(x_i^t,x_j^t,\mathbf{y}^t\right) + \\ & + \sum_{i\in S}\sum_{t\in T}\sum_{k\in C_i} IT_i^{tk}\left(x_i^t,x_i^k,\mathbf{y}^t,\mathbf{y}^k\right) \Bigg] \end{aligned} \qquad (2)$$

In Equation 2, $\mathbf{y}^t$ and $\mathbf{y}^k$ are the images observed at epochs t and k, respectively. The association potential A_i^t is identical to A_i for epoch t in Equation 1. The second term in the exponent, IS_{ij}^t, is identical to I_{ij} for epoch t in Equation 1, but it is called *spatial interaction potential* in order to distinguish it from the third term in the exponent, the *temporal interaction potential* IT_i^{tk}. C_i is the *temporal neighborhood* of image site i at epoch t, thus k is the time index of an epoch that is a "temporal neighbor" of t. The temporal interaction potential models the dependency of class labels at consecutive epochs and the observed data. To our knowledge, such a time-dependant term has not yet been used with CRF. The image sites are chosen to be square blocks of pixels in a regular grid with side length s. We model the CRF to be isotropic and homogeneous, hence the functions used for A_i^t, IS_{ij}^t and IT_i^{tk} are independent of the location of image site i.

The association potential $A_i^t(x_i^t, \mathbf{y}^t)$ is related to the probability of label x_i^t given the image $\mathbf{y}^t$ at epoch t by $A_i^t(x_i^t, \mathbf{y}^t) = \log\{P[x_i^t \mid \mathbf{f}_i^t(\mathbf{y}^t)]\}$. The image data are represented by a site-wise feature vector $\mathbf{f}_i^t(\mathbf{y}^t)$ (section 3) that may depend on the whole image at epoch t, e.g. by using features at different scales in scale space [5]. We use a simple Gaussian model for $P[x_i^t \mid \mathbf{f}_i^t(\mathbf{y}^t)]$ [9]:

$$\begin{aligned} A_i\left(x_i^t,\mathbf{y}^t\right) = & -\frac{n}{2} log(2\pi) - \frac{1}{2} log\left[det\left(\mathbf{\Sigma}_{fc}\right)\right] - \\ & -\frac{1}{2}\left[\mathbf{f}_i^t\left(\mathbf{y}^t\right) - \mathbf{E}_{fc}^t\right]^T \cdot \mathbf{\Sigma}_{fc}^{-1} \cdot \left[\mathbf{f}_i^t\left(\mathbf{y}^t\right) - \mathbf{E}_{fc}^t\right] \end{aligned} \qquad (3)$$

In Equation 3, $\mathbf{E}_{fc}^t$ and $\mathbf{\Sigma}_{fc}$ are the mean and co-variance matrix of the features of class c, respectively. They are determined from the features $\mathbf{f}_i^t(\mathbf{y}^t)$ in training sites individually for each epoch t and each class c. If a semantic class corresponds to several clusters in feature space, we apply a Gaussian mixture model [9]. In this case, expectation maximization (EM) [9] is used to determine the individual clusters and their statistical parameters (mean and covariance matrix). The model in Equation 3 is used for all n individual clusters, and the sites assigned to one of these clusters in the classification stage are merged in post-processing.

The spatial interaction potential IS_{ij}^t is a measure for the influence of the data $\mathbf{y}^t$ and the neighboring labels x_j^t on the class x_i^t of site i at epoch t. The data are represented by site-wise vectors of *interaction features* $\mathbf{\mu}_{ij}^t(\mathbf{y}^t)$ [5]. We use the component-wise differences of the feature vectors $\mathbf{f}_i^t(\mathbf{y}^t)$, i.e. $\mathbf{\mu}_{ij}^t(\mathbf{y}^t) = [\mu_{ij1}^t, \dots \mu_{ijR}^t]^T$, where R is the dimension of the vectors $\mathbf{f}_i^t(\mathbf{y}^t)$, $\mu_{ijk}^t = |f_{ik}^t(\mathbf{y}^t) - f_{jk}^t(\mathbf{y}^t)|$, and $f_{ik}^t(\mathbf{y}^t)$ is the k^{th} component of $\mathbf{f}_i^t(\mathbf{y}^t)$. Introducing β as a weighting factor for the influence of the spatial interaction potential in the classification process, IS_{ij}^t is modeled as:

$$IS_{ij}^t\left(x_i^t,x_j^t,\mathbf{y}^t\right) = \begin{cases} \beta \cdot \exp\left[-\left\|\mathbf{\mu}_{ij}^t\left(\mathbf{y}^t\right)\right\|^2\right] & if \quad x_i^t = x_j^t \\ \beta \cdot \left\{1 - \exp\left[-\left\|\mathbf{\mu}_{ij}^t\left(\mathbf{y}^t\right)\right\|^2\right]\right\} & if \quad x_i^t \neq x_j^t \end{cases} \qquad (4)$$

In Equation 4, $\|\mathbf{\mu}_{ij}^t(\mathbf{y}^t)\|$ is the Euclidean norm of $\mathbf{\mu}_{ij}^t(\mathbf{y}^t)$. This is a very simple model that penalizes local changes of the class labels if the data are similar. The only model parameter is β. It could be determined from training data if fully labeled training images were available, but currently it is defined by the user. The neighborhood N_i of image site i over which IS_{ij}^t has to be summed in Equation 2 consists of the four neighboring image sites in a regular grid.

The temporal interaction potential IT_i^{tk} models the dependencies between the data **y** and the labels x_i^t and x_i^k of site i at epochs t and k. In general IT_i^{tk} can be modeled similarly to IS_i^k by penalizing temporal change of labels unless it is indicated by differences in the data. However, a more sophisticated functional model would be required due to different atmospheric and lighting conditions and due to seasonal effects on the vegetation. In this paper, we use a simple model that neglects the dependency of IT_i^{tk} from the data:

$$IT_i^{tk}\left(x_i^t, x_i^k, \mathbf{y}^t, \mathbf{y}^k\right) = \gamma \cdot \mathbf{TM}\left(x_i^t, x_i^k\right) \tag{5}$$

In Equation 5, γ is a weighting factor. **TM** is a temporal transition matrix similar to the transition probability matrix used in [3] with $\mathbf{TM}(c^t, c^k) = P(c^t \mid c^k)$ for $t \geq k$. That is, the elements of **TM** are the conditional probabilities of an image site belonging to class c^t at epoch t if it belonged to class c^k at epoch k. The temporal neighborhood C_i of x_i^t is chosen to contain the two elements x_i^{t-1} and x_i^{t+1}. The parameters of the model in Equation 5 are the weighting factor γ and the elements of **TM**. They could be estimated from training data, but this would require a large amount of multitemporal data, which is not at our disposal. Hence, both γ and **TM** are set by the user. **TM** is not symmetric, because some changes are more likely than others (e.g. farmland to settlement and vice versa). It also has to model the fact that change is not a very likely event. We apply a bidirectional transfer of temporal information instead of a "cascade" approach that hands information from one image to the next in a sequence [3].

Exact inference is computationally intractable for CRF [5]. In [10], several methods for parameter learning and inference are compared. In this paper we use Loopy-Belief-Propagation (LBP) [10], which is a standard technique for performing probability propagation in graphs with cycles.

3. Features

The elements of the site-wise feature vectors $\mathbf{f}_i^t(\mathbf{y}^t)$ used both for the association and the interaction potentials must be defined such that they can help to discriminate the different classes. We use two groups of features, namely gradient-based features and color-based features. The three gradient-based features are derived from a weighted histogram of the gradient orientations. We use the mean and the variance of these orientations along with the number of bins with values above the mean. These features allow a good distinction between textured and homogeneous areas. For the color-based features we carry out an IHS transformation. The hue channel turned out to be most suitable for our task. Following the strategy proposed in [9] for periodic variables, we introduced two features for the mean hue, namely $H_x=1/n\sum\cos(hue)$ and $H_y=1/n\sum\sin(hue)$. Moreover the variance of the hue was chosen. These altogether six features are computed at three different scales λ_1, λ_2 and λ_3. At scale λ_1, they are computed only from the pixels inside the image site i (a box of s x s pixels), whereas at scale λ_2 and λ_3 the pixels in a square of size $2 \cdot s$ and $3 \cdot s$, respectively, centered at the centre of image site i, are taken into account. In this way, dependencies between the image data of neighboring sites are modeled. Overall the site-wise feature vectors $\mathbf{f}_i^t(\mathbf{y}^t)$ thus consist of 18 elements. The values for each feature are normalized so that they are in the interval [0, 1].

4. Experiments

For our experiments we used the RGB bands of two multi-spectral Ikonos scenes from 2005 and 2007 with 4 m resolution and one RapidEye scene from 2009 with 5 m resolution of a rural region near Herne, Germany. Two areas having a similar type of land cover were cut out of the scenes. One of these areas, covering an area of 2.4 x 2.4 km^2, was used as training area whereas the other one of 4.1 x 3.4 km^2 served as test area. Ground truth was obtained by manually labeling these areas on a pixel-level. The classes to be distinguished are settlement (*set*), industrial areas (*ind*), forests (*for*), and cropland (*crp*). The size s of an image site was selected to be 6 pixels, so the test area consisted of 23.800 sites. An image site was labeled by a majority vote of its pixels. Whereas the association potential for the first three classes could be modeled by the simple Gaussian model, class *crp* showed different appearance for tilled and untilled areas. Thus, a Gaussian mixture model was applied to subdivide this class into two components using EM (cf. Section 2). The temporal transition matrix **TM** used in our experiments is shown in Table 1. The weighting factors were set to $\beta = 1.5$ and $\gamma = 15$. The image sites of the test area were classified by maximizing $P(\mathbf{x} \mid \mathbf{y})$ using LBP.

Table 1. Temporal transition matrix (**TM***5.05)

	$x_i^{t+1}=set$	$x_i^{t+1}=ind$	$x_i^{t+1}=for$	$x_i^{t+1}=crp$
$x_i^t=set$	1	0.05	0.05	0.05
$x_i^t=ind$	0.05	1	0.05	0.05
$x_i^t=for$	0.1	0.1	1	0.1
$x_i^t=crp$	0.2	0.2	0.05	1

To evaluate our multitemporal CRF classification results, all image sites were also classified applying

Maximum-Likelihood (ML)-Classification and monotemporal CRF classification. Applying $\mathrm{CRF}_{\mathrm{multi}}$ significantly increases the classification accuracy for all epochs, not only in comparison to ML but also to $\mathrm{CRF}_{\mathrm{mono}}$, especially for the 2009 RapidEye data (Table 2). Figure 1 shows one section of approx. 900 sites as an example. It becomes obvious, that the noisy appearance of the ML-classification is eliminated. Most of the misclassified sites (Table 3) are caused by the following reasons: The distinction between settlement and industrial areas is hard since both have a similar appearance. Farms that usually consist of some houses and trees were labeled as settlement, but in most cases are classified as forest. Main roads were labeled as settlement, but in many cases these elongated objects were oversmoothed by the spatial interaction potential. In the last case a more sophisticated model is needed.

Table 2. Accuracy / kappa coefficients

	ML	$\mathrm{CRF}_{\mathrm{mono}}$	$\mathrm{CRF}_{\mathrm{multi}}$
t_1 (2005)	73.6 % / 0.61	80.8 % / 0.71	82.7 % / 0.73
t_2 (2007)	75.8 % / 0.64	82.2 % / 0.72	83.2 % / 0.74
t_3 (2009)	67.4 % / 0.52	75.0 % / 0.62	79.6 % / 0.68

Table 3. Confusion matrix for t_2 (2007)

	$x_i^{lab} = set$	$x_i^{lab} = ind$	$x_i^{lab} = for$	$x_i^{lab} = crp$
$x_i = set$	9003	309	271	119
$x_i = ind$	574	666	84	79
$x_i = for$	954	75	1863	565
$x_i = crp$	329	12	627	8270

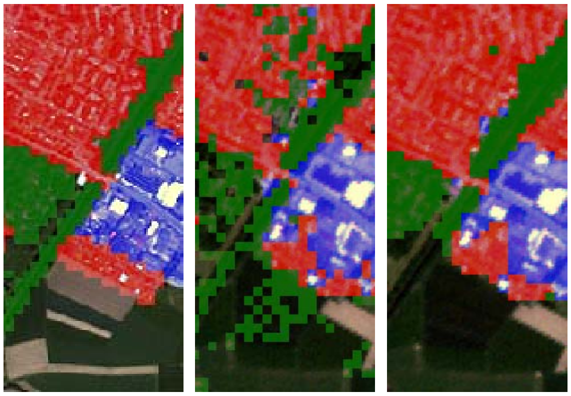

Figure 1. Ground truth overlayed to 2005 image (left), results of ML (mid.) and $\mathrm{CRF}_{\mathrm{multi}}$ (right) for 2009; Red/blue/green/transp.: *set/ind/for/crp.*

5. Conclusion and Outlook

We have presented a new multitemporal CRF-based approach for land cover classification in high resolution optical satellite imagery. The framework allows incorporating spatial and temporal context information. Our results are quite promising, even more so because they were achieved using only a small set of features and relatively simple models for the interaction potentials. Further research will concentrate on the applicability of the model on data of different scales, which requires an appropriate selection of different features for the epochs and a more advanced temporal model, and on testing non-Gaussian models. Moreover the CRF framework should use the results of a preliminary segmentation in order to obtain a more precise determination of the class boundaries.

Acknowledgement

This study was funded by the German Science Foundation (Deutsche Forschungsgemeinschaft) under grant HE 1822/22-1. Our implementation is based on the CRF Toolbox by K. Murphy & M. Schmidt [10]: www.cs.ubc.ca/~murphyk/Software/CRF/crf.html.

References

[1] D. Lu, P. Mausel, E. Brondizio and E. Moran. Change detection techniques. *Int. J. of Remote Sensing,* 25(12):2365-2401, 2004.

[2] R. Q. Feitosa, G. A. O. P. Costa, G. L. A. Mota, K. Pakzad and M. C. O. Costa. Cascade multitemporal classification based on fuzzy Markov chains. *ISPRS J. Photogrammetry Remote Sens.* 64(2):159-170, 2009.

[3] L. Bruzzone, R. Cossu and G. Vernazza. Detection of land-cover transitions by combining multidate classifiers. *Pattern Recognition Letters,* 25(13):1491-1500, 2004.

[4] F. Melgani and S. B. Serpico. A Markov Random Field approach to spatio-temporal contextual image classification. *IEEE-TGARS,* 41(11):2478-2487, 2003.

[5] S. Kumar and M. Hebert. Discriminative Random Fields. *Int. J. Computer Vision*, 68(2):179-201, 2006.

[6] P. Zhong and R. Wang. A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images. *IEEE-TGARS,* 45(12):3978-3988, 2007.

[7] T. Hoberg and F. Rottensteiner. Classification of settlement areas in remote sensing imagery using Conditional Random Fields. *Int. Arch. Photogrammetry, Remote Sens.*, SIS XXXVIII (7), 2010.

[8] W.-L. Lu, K. P. Murphy, J. J. Little, A. Sheffer, and H. Fu. A hybrid conditional random field for estimating the underlying ground surface from airborne Lidar data. *IEEE-TGARS,* 47(8/2):2913-2922, 2009.

[9] C. M. Bishop. *Pattern recognition and machine learning.* 1st edition, Springer New York, 2006.

[10] S. Vishwanathan, N. N. Schraudolph, M. W. Schmidt, K. P. Murphy. Accelerated training of conditional random fields with stochastic gradient methods. *23rd Int. Conf. on Machine Learning*:969-976, 2006.