A TWO-LAYER CONDITIONAL RANDOM FIELD MODEL FOR SIMULTANEOUS CLASSIFICATION OF LAND COVER AND LAND USE

L. Albert*, F. Rottensteiner, C. Heipke

Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover - Germany
{albert, rottensteiner, heipke}@ipi.uni-hannover.de

Commission III, WG III/4

KEY WORDS: Contextual classification, Conditional Random Fields, multi-layer, land use classification

ABSTRACT:

This paper proposes a two-layer Conditional Random Field model for simultaneous classification of land cover and land use. Both classification tasks are integrated into a unified graphical model, which is reasonable due to the fact that land cover and land use exhibit strong contextual dependencies. In the CRF, we distinguish a *land cover layer* and a *land use layer*. Both layers differ with respect to the entities corresponding to the nodes and the classes to be distinguished. In the land cover layer, the nodes correspond to superpixels extracted from the image data, whereas in the land use layer the nodes correspond to objects of a geospatial land use database. Statistical dependencies between land cover and land use are explicitly modelled as pair-wise potentials. Thus, we obtain a consistent model, where the relations between land cover and land use are learned from representative training data. The approach is designed for input data based on aerial images. Experiments are performed on an urban test site. The experiments show the feasibility of the combination of both classification tasks into one overall approach and investigate the influence of the size of the superpixels on the classification result.

1. INTRODUCTION

1.1 Motivation

Land use classification is a standard task in remote sensing and forms the basis for the verification and update of geospatial land use databases. Land use can be obtained by an approach consisting of two steps, e.g. (Albert et al., 2014). In a first step, land cover information is classified using remote sensing data. The second step consists of a segment-based land use classification. Both classification tasks pursue different objectives. Land cover classification focuses on the assignment of land cover labels to small geometrical image sites having the same land cover. Land use reveals the socio-economic function of a piece of land, which is typically composed of different land cover segments. The goal of the land use classification is to assign a land use label to such pieces of land, which is based, amongst others, on features derived from the land cover classification results. For this purpose, a set of adequate features has to be derived from the land cover classification results in order to enable a correct inference of land use classes. The feature selection step requires a certain degree of knowledge about the overall characteristics of land use classes and their relations to land cover distributions within a land use object.

The two-step approach for land use classification can be interpreted as an image interpretation task at different interpretation levels. In contrast to land cover, land use cannot be derived directly from remote sensing data. Besides spectral characteristics, the composition of different land cover elements within a land use object is rather important to infer its socio-economic function. Thus, land use classification represents a higher level of image interpretation. In the two-step approach semantic relations describing the statistical dependencies between land cover and land use are indirectly introduced via additional features for the second classification derived from the results of the first step. Obviously, there also exist spatial dependencies between neighbouring image sites, such as pixels or segments in land cover as well as land use classification. In both cases, some classes are more likely to occur next to each other than others. In particular, land use classes typically occur in certain spatial configurations. It can be noted that land cover and land use classification exhibit strong contextual dependencies, where context incorporates spatial as well as semantic relations.

In this paper, we present an approach for land use classification of objects from a geospatial database, where the geometry of the objects is given and assumed to be correct. The rationale for this assumption is that our approach is the first step of a scheme for updating the given database. Rather than using a two-step procedure as outlined before, we determine land cover and land use simultaneously based on an explicit model of the statistical dependencies between land cover and land use. A simultaneous classification of land cover and land use is achieved by combining the two interpretation levels in a unified model, so that the information of the lower and the higher interpretation levels support each other and can help to improve the overall classification accuracy. Also, we avoid taking potentially wrong decisions, which cannot be reversed at later stages, too early (namely during land cover classification). Rather we try to determine the most probable label configuration at each layer and allow for the two layers to mutually influence each other. For this purpose, we use Conditional Random Fields (CRF) (Kumar & Hebert, 2006), which provide a flexible framework for contextual classification. The possibility of extending CRF to a multi-layer scheme allows modelling dependencies between labels of arbitrary image entities as well as semantic class structures and the data.

In the proposed a two-layer CRF, we distinguish a *land cover layer* and a *land use layer*. Both layers consist of nodes and

* Corresponding author.

intra-layer edges. The nodes at different layers are connected by inter-layer edges, which model the statistical dependencies between land cover and land use. Both layers differ with respect to the entities corresponding to the nodes and the classes to be distinguished, which is caused by the different nature of these classification tasks.

The approach is designed for input data based on aerial images. Experiments are performed on an urban test site and are carried out to evaluate the performance of the proposed method. The goal is to show the feasibility of the combination of both classification tasks into one overall approach. Furthermore, we investigate the influence of the size of the superpixels on the classification result.

1.2 Related Work

Several approaches for land use classification exist, which differ with respect to general processing strategy, extracted features, classifiers applied and input data. Some of the approaches apply a two-step processing strategy (Hermosilla et al., 2012; Helmholz, 2012). In a first step, a pixel- or segment-based land cover classification is performed. In a second step, the classification results are transferred to the land use objects of a geospatial database. In our previous work, we have presented a two-step land use classification approach using CRF (Albert et al., 2014). CRF are applied for land cover as well as land use classification, separately. Both CRFs model spatial dependencies between neighbouring sites, namely pixels in the case of land cover and segments in the case of land use classification. The benefit of the considering contextual knowledge in the classification process has already been identified, e.g. for the classification of urban structure types (Hermosilla et al., 2012). For this purpose, Hermosilla et al. (2012) incorporate contextual features in land use classification, which describe the relations of land cover areas within a land use object as well as relations between neighbouring land use objects. Instead of implicitly integrating context in the classification process by using contextual features, CRF offer the possibility to model relations between neighbouring land use objects as well as between land use and land cover objects directly, thus, explicitly considering context in the classification process.

Multi-layer CRF have been applied to several tasks in image analysis, e.g. hierarchical classification of building facades (Yang and Förstner, 2011). Yang and Förstner (2011) model spatial and multi-scale relationships between segments obtained by a multi-scale watershed segmentation of an image. The potential of multi-layer CRF has also been exploited for the classification of scenes with occlusions (Kosov et al., 2013) as well as the multi-temporal classification of remote sensing data (Hoberg et al. 2012). Kosov et al. (2013) propose a CRF which models the class labels of the occluded and the occluding object for each image site in two separate layers. Hoberg et al. (2012) model multi-temporal and multi-scale dependencies of remote sensing data at different epochs in a multi-layer CRF for land cover classification, where the underlying images of each layer can have different resolutions. In this approach, the spatial dependencies between image sites for one epoch are modelled within each layer. Temporal dependencies of image sites at different epochs are modelled by inter-layer edges connecting different layers. Most of these approaches aim to classify identical classes over time, i.e. change detection is not explicitly considered. This also applies to multi-scale approaches, although the class structure can slightly differ due to a different appearance of the entities to be classified across scale. For multi-scale approaches, the class structure complies with a part-based object model referring to object parts at finer scale and to compound objects at coarser scale. The multi-scale approaches simultaneously model local and global information in the CRF.

The statistical dependencies between sets of images sites can be captured by pair-wise or by higher order cliques. The approaches mentioned before only make use of pair-wise potentials. Higher order potentials have been exploited e.g. for image segmentation (Kohli et al., 2009). Kohli et al. (2009) present a class of higher order potentials, referred to as P^N-Potts model, which can be solved efficiently with move-making algorithms based on graph cuts. Higher order potentials allow modelling complex dependencies between random variables in a graphical model. However, inference on higher order potentials is challenging for generic formulations of the higher order potentials. This is why in this paper we restrict ourselves to a pair-wise model, too.

1.3 Contributions

To the best of our knowledge, this is the first approach making use of a multi-layer CRF for the classification of land use objects. The major contribution of this paper is the extension of the two-step approach using CRF by introducing statistical dependencies between land cover and land use in a multi-layer CRF. This paper focuses on the design of the graphical model. We explicitly model these dependencies in pair-wise cliques, i.e. as pair-wise potentials. Thus, training and inference are much easier than for higher order potentials, because algorithms can be simply adopted from the standard CRF. The integration of both classification tasks into one graphical model overcomes the challenge of an adequate feature selection for land use classification. Specific knowledge about the overall characteristics of certain land use classes and their relations to land cover distributions within a land use object is thus not required. It is rather important to consider the statistical relationships in order to define an abstract model, in which specific relations can be learned from representative training data. A main benefit of our model is that it tries to determine the most probable label configuration of the two layers simultaneously without taking early decisions (by fixing the land cover labels after the first classification stage). Explicitly modelling the contextual relations between the layers is supposed to lead to a better classification accuracy.

In contrast to our previous work (Albert et al., 2014), land cover classification is performed at the level of superpixels rather than pixels. For the classification of land use objects, the evidence and approximate arrangement of certain land cover classes within a land use object are more important than their precise pixel-wise distribution. The generalization to superpixels reduces the computational effort, while still being sufficiently detailed for the purpose of land use classification.

This paper is structured as follows. Section 2 introduces the standard CRF framework. Section 3 focuses on the methodology of the two-layer CRF model. Section 4 describes the experimental evaluation of the approach incl. the test setup and the feature extraction process. Finally, conclusions and an outlook on future work are given in section 5.

2. CONDITIONAL RANDOM FIELDS

Conditional Random Fields provide a flexible framework for contextual classification. They were introduced by Kumar and

Hebert (2006) for image classification. CRF are undirected graphical models, consisting of nodes n and edges e. The nodes represent the image sites, e.g. pixels or segments. The edges link adjacent nodes and model statistical dependencies between class labels and data at neighbouring image sites. The class labels of all image sites are combined in a label vector $\mathbf{y} = [y_1, \dots, y_i, \dots, y_n]$, where $i \in S$ is the index of an image site and S is the set of all image sites. The goal is to assign the most probable class labels $\mathbf{y}$ from a set of classes to all image sites simultaneously considering the data $\mathbf{x}$. CRF are discriminative classifiers, thus directly modelling the posterior probability $P(\mathbf{y}|\mathbf{x})$ of the label vector $\mathbf{y}$ given the observed data $\mathbf{x}$:

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \prod_{i \in S} \varphi_i(y_i, \mathbf{x}) \cdot \prod_{i \in S} \prod_{j \in N_i} \psi_{ij}(y_i, y_j, \mathbf{x})^{\omega} . \tag{1}$$

In equation (1), $\varphi_i(y_i, \mathbf{x})$ are the *association potentials* and $\psi_{ij}(y_i, y_j, \mathbf{x})$ are called the *interaction potentials*. The partition function $Z(\mathbf{x})$ acts as a normalization constant which transforms the potentials into probabilities, whereas N_i is the neighbourhood of image site i. The relative weight of the interaction potential compared to the association potential is modelled by the parameter ω. The association potential $\varphi_i(y_i, \mathbf{x})$ models the relations between class label y_i and the observations $\mathbf{x}$. The interaction potential $\psi_{ij}(y_i, y_j, \mathbf{x})$ models the relations between the labels y_i and y_j of adjacent nodes and the observations $\mathbf{x}$. CRF represent a general framework, which allows to introduce various functional models for both potentials (Kumar and Hebert, 2006). Thus, it is possible to choose any arbitrary discriminative classifier with a probabilistic output $P(y_i|\mathbf{x})$ for the association potential. This also applies for the interaction potential, where different models can be applied. Kumar and Hebert (2006) use a generalized linear model for the association potential, but several other classifiers have also proven to work well, for instance Random Forests (RF) (Breiman, 2001; Schindler, 2012). The models applied for the interaction potential are often more simple, favouring identical labels and penalising label changes (see Schindler, 2012 for a comparison). However, some approaches apply more complex models for the interaction potential in order to avoid over-smoothing (e.g. Niemeyer, 2014). CRF are a supervised classification technique, thus the parameters of the potentials are learned. In the inference step, the most probable label configuration of the graphical model is determined for all nodes simultaneously. This is based on maximizing the posterior probability $P(\mathbf{y}|\mathbf{x})$ of the labels given the data by an iterative optimization process.

3. A TWO-LAYER CONDITIONAL RANDOM FIELD MODEL

3.1 Mathematical description

In order to model the statistical dependencies between land cover and land use, we design a two-layer Conditional Random Field. The CRF is composed of a *land cover layer* and a *land use layer*. Figure 1 illustrates the design of the graphical model. Each layer consists of nodes and intra-layer edges. The nodes of both layers are connected by inter-layer edges. We aim to estimate the class labels for land cover y_i^c and land use y_i^u as random variables for each node i in the corresponding layer.

Both layers differ concerning the image entities represented by the nodes and the semantic classes to be distinguished. In the land cover layer, the nodes correspond to superpixels extracted from the image data, whereas in the land use layer the nodes correspond to land use objects. Examples for the shapes of both sites are shown in figure 2. Superpixels are small sets of pixels having similar characteristics. We use a method proposed by Achanta et al. (2012) for the generation of superpixels, called *Simple linear iterative clustering* (SLIC), which is based on an adapted version of k-means clustering. The size and compactness of the generated superpixels can be controlled by parameters in order to enable a certain adaptation to spectral boundaries in heterogeneous areas. In homogeneous areas, SLIC superpixels tend to have a compact shape. Land use objects are defined by land use parcels obtained from a geospatial database.

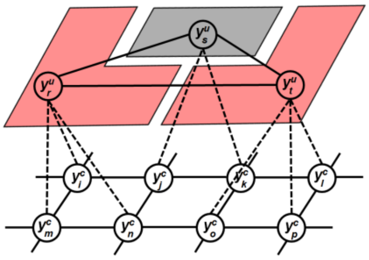

Figure 1: Graphical model consisting of two layers, land cover layer (c) and land use layer (u). Nodes are depicted as circles, intra-layer edges as solid and inter-layer edges as dashed lines.

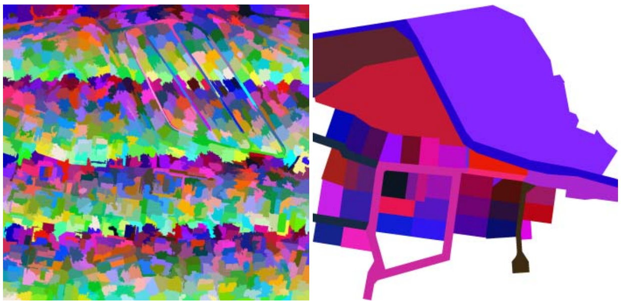

Figure 2: Region images representing superpixels (left) and land use objects (right). The colours bear no meaning.

The intra-layer edges model the spatial neighbourhood of adjacent nodes in the respective layer. The neighbourhood of a node i is composed of its first-order spatial neighbours, i.e. all sites that share a common boundary with the given site represented by node i. The inter-layer edges model the statistical dependencies between land cover and land use. Inter-layer edges form pair-wise cliques connecting two nodes, one in each layer. These edges connect spatially overlapping image sites, i.e. any node of the land use layer is connected with all nodes from the land cover layer having a spatial overlap with the object corresponding to the land use node.

The standard CRF notation given in section 2 is extended according to the two-layer graphical model described above:

$$\begin{aligned} & P(\mathbf{y^c}, \mathbf{y^u}|\mathbf{x}) \\ & = \frac{1}{Z(\mathbf{x})} \prod_{i \in S^c} \varphi^c\,(y_i^c, \mathbf{x})^{\omega_1} \cdot \prod_{k \in S^u} \varphi^u\,(y_k^u, \mathbf{x})^{\omega_2} \\ & \cdot \prod_{ij \in \mathrm{E_S^c}} \psi^c\,(y_i^c, y_j^c, \mathbf{x})^{\omega_3} \cdot \prod_{kl \in \mathrm{E_S^u}} \psi^u\,(y_k^u, y_l^u, \mathbf{x})^{\omega_4} \\ & \cdot \prod_{ik \in \mathrm{E}^{cu}} \psi^{cu}(y_i^c, y_k^u, \mathbf{x})^{\omega_5} . \end{aligned} \tag{2}$$

In equation (2), we have omitted the subscripts in the potential functions (e.g., φ^c) to indicate that the functional models are independent from the image location. S^c and S^u represent the sets of all nodes n_i^c and n_k^u in the land cover (*c*) and land use layers (*u*). E_s^c and E_s^c are the sets of all intra-layer edges connecting spatially neighboured sites in the respective layer. Inter-layer edges connecting nodes of both layers are collected in a set E^{cu}. The *association potentials* φ^c and φ^u model the relations between class labels y_i^c, y_i^u and the data **x**. ψ^c and ψ^u represent the *intra-layer interaction potentials*, which model the spatial dependencies between neighbouring sites within each layer in consideration of the data **x**. The *inter-layer interaction potential* ψ^{cu} models the statistical dependencies between land cover labels y_i^c and land use labels y_k^u at sites i and k being adjacent in-between both layers, also considering the data. The parameters $\Omega = (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5)$ determine the weights of each potential term relative to the first potential term (i.e. $\omega_1 \equiv 1$), which define the individual influence of the corresponding potentials in the classification process. Both layers differ partly with respect to the models for the association and interaction potentials and the chosen model parameters. Details are described in the subsequent sections.

3.2 Association Potentials

The association potential predicts how likely node *i* belongs to a class y_i given the data **x**. The data are taken into account in the form of site-wise feature vectors $\mathbf{f}_i^c(\mathbf{x})$ and $\mathbf{f}_k^u(\mathbf{x})$ for the nodes n_i^c in the land cover layer and n_k^u in the land use layer (Kumar and Hebert, 2006), which may depend on all data. Both association potentials take a value proportional to the probability of y_i^c and y_k^u given the site-wise feature vectors $\mathbf{f}_i^c(\mathbf{x})$ and $\mathbf{f}_k^u(\mathbf{x})$, i.e. $\varphi^c\,(y_i^c, \mathbf{x}) \propto P\big(y_i^c \big| \mathbf{f}_i^c(\mathbf{x})\big)$ for the land cover layer and $\varphi^u\,(y_k^u, \mathbf{x}) \propto P\big(y_k^u \big| \mathbf{f}_k^u(\mathbf{x})\big)$ for the land use layer. We choose the Random Forest (RF) classifier, introduced by Breiman (2001), for determining the association potentials of both layers. However, each classification is based on different features. RF has proven to be an efficient classifier, also in remote sensing applications (e. g. Schindler, 2012). The RF classifier complies with the above mentioned requirements. In training, RF creates an ensemble of randomized decision trees. During classification, an unknown sample is classified by each tree based on the corresponding feature values, the tree thus casting a vote for the class it considers to be the most likely one. The sum of the votes for a class divided by the total number of trees defines the value of the association potential for that class. Some parameters of the RF classifier have to be set beforehand, these are, amongst others, the maximum number of samples used for training, the maximum depth and the number of trees in the forest. Due to considerable differences in the structure of both classification tasks, these parameters have to be selected individually.

3.3 Intra-layer Interaction Potentials

The intra-layer interaction potential models the dependencies of the labels of nodes n_i and n_j being adjacent within one layer, considering the data **x**. In both layers, the neighbourhood of node n_i consists of its first order neighbours in the respective layer. The data are taken into account in the form of an interaction feature vector $\boldsymbol{\mu}_{ij}(\mathbf{x})$ for each edge. We apply different models for the intra-layer interaction potentials of both layers.

For the interaction potential of the intra-layer edges in the land cover layer, we apply the *contrast-sensitive Potts model* (Boykov and Jolly, 2001). This model represents an extension of the *Potts model*, additionally taking into account the data (Kumar and Hebert, 2006). Here, the interaction potential is based on the probability of both labels y_i^c and y_j^c being identical given $\boldsymbol{\mu}_{ij}^c(\mathbf{x}) = d_{ij}$, i.e. $\psi^c\,(y_i^c, y_j^c, \mathbf{x}) \propto P\big(y_i^c = y_j^c \big| d_{ij}\big)$. Thus, the sole interaction feature is the Euclidian distance d_{ij} between the site-wise feature vectors $\mathbf{f}_i^c(\mathbf{x})$ and $\mathbf{f}_j^c(\mathbf{x})$ of two adjacent nodes n_i^c and n_j^c, and the degree of smoothing depends on the data. The contrast-sensitive Potts model defines the interaction potential:

$$\psi^c\,(y_i^c, y_j^c, \mathbf{x}) = \begin{cases} exp\left(l_1 + (1 - l_1) \cdot e^{-\frac{d_{\mathrm{ij}}{}^2}{2\sigma^2}}\right) & if\ y_i^c = y_j^c \\ 1 & otherwise \end{cases}. \quad (3)$$

In equation (3), the parameter $l_1 \in [0; 1]$ defines the relative weight between the data-dependent and data-independent smoothing term. If l_1 equals one, the model becomes a Potts model (Schindler, 2012). If l_1 is set to zero, the degree of smoothing depends completely on the data. The parameter σ^2 correspond to the mean value of the squared distances $d_{ij}{}^2$ and is learned during training. We choose the contrast-sensitive Potts model for the interaction potential of the land cover layer, because we want to achieve a smoothing effect, provided that the data support it. Furthermore, this kind of model has proven to achieve satisfactory results (Schindler, 2012) in reasonable computation time.

Land use classification has different demands for the modelling of context. A pure smoothing of the class labels of neighbouring land use objects is not desired. In contrast, more probable class configurations given the data should be favoured. How probable a class relation is, is to be learned from real-world occurrences in representative training data given the observations. Thus, the interaction potential is modelled as the joint posterior probability of both labels y_k^u and y_l^u given $\boldsymbol{\mu}_{kl}^u(\mathbf{x})$, i.e. $\psi^u\,(y_k^u, y_l^u, \mathbf{x}) \propto P\big(y_k^u, y_l^u \big| \boldsymbol{\mu}_{kl}^u(\mathbf{x})\big)$. This model formulation corresponds to a standard classification task, hence, it is possible to handle the interaction potential similar to the association potential by applying any discriminative classifier with a probabilistic output. The difference is that any pair of classes is considered as a single class. Again, and following (Niemeyer et al., 2014), we use the RF classifier due to its good classification performance. In this case, the interaction feature vector $\boldsymbol{\mu}_{kl}^u(\mathbf{x})$ corresponds to the concatenated site-wise feature vectors $\mathbf{f}_k^u(\mathbf{x})$ and $\mathbf{f}_l^u(\mathbf{x})$ of two adjacent nodes n_k^u and n_l^u.

3.4 Inter-layer Interaction Potential

The inter-layer interaction potential models the statistical dependencies of the land cover and land use labels at nodes n_i^c and n_k^u connected by an inter-layer edge, considering the data **x**. Similar to the interaction potential for adjacent nodes in the land use layer, we model this potential relating land cover and land use classes based on the joint posterior probability of both labels given the data, thus $\psi^{cu}(y_k^u, y_l^u, \mathbf{x}) \propto P\big(y_i^c, y_k^u \big| \boldsymbol{\mu}_{ik}^{cu}(\mathbf{x})\big)$. Again, we apply the RF classifier and define the interaction feature vector $\boldsymbol{\mu}_{ik}^{cu}(\mathbf{x})$ as the concatenated site-wise feature vectors $\mathbf{f}_i^c(\mathbf{x})$ and $\mathbf{f}_k^u(\mathbf{x})$ of two adjacent nodes n_i^c and n_k^u. The probability for the co-occurrence of land cover and land use classes given the observations is learned from training data.

3.5 Training and Inference

Basically, training and inference require an adequate set of features, which we extract beforehand for the nodes of both layers. In the training step, all individual potentials, association as well as interaction potentials, are trained separately on representative training data. This includes the generation of the randomized decision trees as well as the learning of parameter σ of the contrast-sensitive Potts model. Besides, the user has to define the potential weights Ω and the parameter l_1 of the contrast-sensitive Potts model. They could be determined by a procedure such as cross-validation (Shotton et al., 2009), but this has not been carried out yet. During the training of the intra-layer interaction potentials of the land use layer, the relations between adjacent nodes are learned. This requires fully-labelled training data for the corresponding layer. Moreover, the training of the inter-layer interaction potential requires training data where spatially overlapping image sites in both layers are labelled in order to learn the relations between land cover and land use appropriately.

In the inference step, we estimate an approximate solution of the optimal label configuration in the graphical model, because exact inference is computationally intractable (Kumar and Hebert, 2006). Here, we apply the max-sum version of the message passing algorithm Loopy Belief Propagation (Frey and MacKay, 1998).

4. EXPERIMENTS

4.1 Test Data and Test Setup

The experiments are carried out to evaluate the feasibility of the combination of both classification tasks into one overall approach. Furthermore, we investigate the influence of the size of the superpixels on the classification result. We perform our experiments on a test site containing the city of Hameln, Germany. The test area shows various urban as well as rural characteristics, such as residential areas with detached houses, densely built-up areas, industrial areas, a river, cropland and grassland. The test area has a size of 2 km x 6 km. The input data consist of an orthophoto, a digital terrain model (DTM) and a digital surface model (DSM) derived by image matching. The orthophoto has a ground sampling distance of 0.2 m and consist of four channels (one near-infrared channel, three colour channels). The DSM and DTM provide height information at a resolution of 0.5 m and 5 m, respectively. Furthermore, GIS-objects of the German geospatial land use database forming a part of the Authoritative Real Estate Cadastre Information System (ALKIS®) (AdV, 2008) are used to define the land use objects, which correspond to the nodes in the land use layer. The nodes of the land cover layer correspond to SLIC superpixels. The segmentation is performed on a three-channel image, where the channels correspond to the difference between the DSM and the DTM (normalised DSM or nDSM), i.e. the height above ground, the intensity and the normalized difference vegetation index (NDVI) extracted from the input data. The use of these three secondary channels instead of the original grey values enables a better adaptation to boundaries of certain land cover segments. Also, the current implementation is restricted to three channels only. We extract SLIC superpixels of size 50 x 50 and 20 x 20. Figure 3 shows examples for extracted SLIC superpixels in an urban and a rural scene.

For training and evaluation, reference data are available for both layers. The reference data for the land cover layer consist of pixel-wise reference labels for 37 image tiles, each of size 200 m x 200 m, obtained by manual annotation. The reference for each superpixel is assigned to the most frequent class label among its constituent pixels. The reference data for the land use layer consist of the geospatial land use database for the whole test area, divided into 12 blocks, each of size 1000 m x 1000 m.

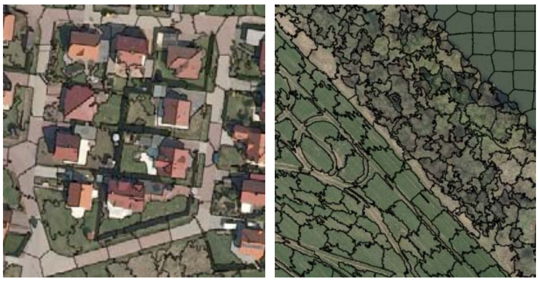

Figure 3: Extracted SLIC superpixels, each of size 50 x 50 pixels, superimposed to an orthophoto of an urban scene (left) and a rural scene (right). The compactness is set to 20 in a range of [1;100].

We distinguish the nine land cover classes *building* (*build.*), *sealed area* (*seal.*), *bare soil* (*soil*), *grass*, *tree*, *water*, *rails*, *car* and *others*, and the seven land use classes *residential* (*res.*), *street*, *water*, *railway* (*rail.*), *agriculture* (*agr.*), *forest* and *others*. The number of trees and the maximum depth of the RF classifier are set to 200 and 25 in each case this classifier is applied. The maximum number of training samples has to be adapted to the total number of samples available for training, which is much lower for the land use layer compared to the land cover layer. Thus, this parameter is set to 10,000 for the association potential in the land cover layer, 10,000 for the inter-layer interaction potential and 5,000 for the association potential and intra-layer interaction potential in the land use layer. The weights Ω for the potential terms are equally set to 1, thus all potentials have the same impact on the classification.

The quantitative and qualitative evaluation is based on cross-validation. For that purpose, the reference data are divided into 12 groups, where each group consist of one of the 1 km^2 blocks of land use reference data mentioned above combined with spatially overlapping land cover reference data. In each test run, we use one group for the evaluation and all others for training. In the 12 test runs, each group thus contributes to the evaluation once. We obtain a confusion matrix by site-wise comparison of the classification result to the reference for each layer separately; the comparison for the land cover layer is carried out on a per-superpixel-basis. The quantitative evaluation is based on the overall accuracy, kappa index, correctness and completeness values derived from the confusion matrix (Rutzinger et al., 2009).

4.2 Feature Extraction

We extract a set of features for the nodes of each layer. The feature extraction for both layers is based on the same image and height data, but differs concerning the type of features. Additionally, some of the land use features are derived from the polygonal representation of the land use objects obtained from a geospatial database.

4.2.1 Land Cover Layer: In the land cover layer, spectral, textural and three-dimensional features are extracted for superpixels. In addition, the feature set is complemented by multi-scale features. In a first step, the features are derived for each pixel and scaled to the interval [0;1]. In a second step, the pixel-wise feature values are averaged for each superpixel. The spectral features consist of the original grey values of the image, the NDVI, hue, saturation and intensity as well as their mean and variance values in a local neighbourhood (here, 13 x 13 pixels). Furthermore, we estimate the magnitudes of the image gradients of the intensity image. The textural features are energy, contrast and homogeneity derived from the Grey Level

Co-Occurrence Matrix (GLCM) proposed by Haralick (1973). The three-dimensional features are the normalized digital surface model (nDSM) and derived features, the mean and Gaussian curvatures as well as the gradient magnitude. The multi-scale features are estimated for linear scale space parameters σ taking the values of 2, 5 and 10. The set of features for the nodes of the land cover layer consist in total of 60 features, which are combined in the feature vector $\mathbf{f}_i^c(\mathbf{x})$ for each node n_i^c.

4.2.2 Land Use Layer: In the land use layer, features are extracted for land use objects, which are defined by the polygonal representation of the GIS-objects of a geospatial land use database. We determine spectral, textural, geometrical and three-dimensional features. Furthermore, the number of neighbouring land use objects is used as a feature. The spectral features consist of the mean, standard deviation, minimum and maximum of the NDVI, hue, saturation and intensity values, which are estimated from all pixels within an object. Again, we use the textural features energy, contrast and homogeneity derived from the GLCM, with the difference that the GLCM is now computed from the co-occurrences of the intensity values of all pixels within each object. The geometrical features are the area, perimeter and compactness, which can be determined from the polygonal representation of each object. The three-dimensional features consist of the mean value, standard deviation, minimum and maximum values of the height above ground within each object. In total, the feature set contains 36 features, which are combined in the feature vector $\mathbf{f}_k^u(\mathbf{x})$ for each node n_k^u.

4.3 Evaluation of Land Cover Classification

Figure 4 shows examples of the result for the land cover layer using the two-layer CRF approach for two different sizes of the superpixels. In both results, most of the *buildings* and *sealed areas* are classified correctly. The discrimination between *grass* and *tree* is in parts incorrect due to the fact that the trees did not carry leaves at image acquisition time. As a result the trees are not represented in the nDSM making the discrimination of both vegetation classes difficult. Furthermore, the boundaries of the superpixels frequently do not match the building boundaries. This is partly caused by inaccuracies in the DSM. Smaller superpixels represent the land cover segments more accurately in a geometric sense, and they can capture more detail such as cars or small trees. Larger superpixels partly cover different land cover classes which leads to inaccuracies.

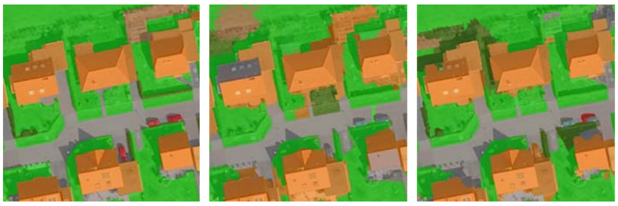

Figure 4: Pixel-wise ground truth (left), classification result of the two-layer CRF approach based on superpixels of size 50 x 50 (centre) and of size 20 x 20 (right). Colours: *build.* (red), *seal.* (grey), *soil* (brown), *grass* (green), *tree* (dark green), *car* (red), *others* (pink).

A quantitative evaluation of the comparison of the results obtained with the two-layer CRF approach for two different sizes of the superpixels is presented in table 1. The completeness and correctness values per class as well as the overall accuracy and kappa index are significantly improved for a smaller size of the superpixels. The two-layer CRF approach based on superpixels of size 20 x 20 yields a mean overall accuracy of about 80.8% and a mean kappa index of 75.3%, which is improved by 8.0% and 10.3% compared to the results obtained for superpixels of size 50 x 50. The completeness and correctness for the classes *building* and *sealed area* are improved by more than 10%. For the class *bare soil* the correctness increases by more than 15% and the completeness value by about 2.8%. The classes *grass* and *water* show a large increase in correctness, which goes along with a small decrease in completeness. For the class *tree*, the correctness value stays nearly the same, but the completeness increases by more than 10%. The class *car* is not detected when using large superpixels due to the lack of detail mentioned before. However, smaller superpixels achieve a completeness value of 31.5% and a correctness value of 55.9% for the class *car*. The results for the classes *rails* and *others* are based on a very small number of samples used both in training and for testing, so that these numbers are hardly representative.

		$\text{CRF}_{\text{multi, 50x50}}$		$\text{CRF}_{\text{multi, 20x20}}$	
		Comp. [%]	**Corr. [%]**	**Comp. [%]**	**Corr. [%]**
Land cover classes	*build.*	74.0	77.7	89.0	89.8
	seal.	73.2	58.6	83.2	70.3
	soil	28.5	65.9	31.3	85.0
	grass	79.0	76.3	78.6	85.6
	tree	77.2	77.3	89.7	77.1
	water	85.2	82.8	83.6	97.2
	rails	15.6	77.4	--	--
	car	--	--	31.5	55.9
	others	4.3	20.4	3.2	47.5
OA [%]		**72.8**		**80.8**	
Kappa [%]		**65.0**		**75.3**	

Table 1: Overall accuracy [%], kappa index [%], completeness (comp.) and correctness (corr.) values [%] for the land cover classes *build.*, *seal.*, *soil*, *grass*, *tree*, *water* and *car* obtained by classification using the two-layer CRF approach based on superpixels of size 50 x 50 ($\text{CRF}_{\text{multi, 50x50}}$) and of size 20 x 20 ($\text{CRF}_{\text{multi, 20x20}}$).

The best completeness and correctness values are achieved for the class *building*, but also the classes *grass*, *tree* and *water* achieve good completeness and correctness values. Lower completeness values for the class *bare soil* could be caused by an overall smaller number of training samples for this class, thus not sufficiently representing the whole range of characteristics of this class. The poor completeness and correctness values for the class *car* result from the size of the superpixels. Even for small superpixels, individual cars are not represented by a separate superpixel, but rather are merged into those mainly covering *sealed area*.

4.4 Evaluation of Land Use Classification

The confusion matrix obtained from the comparison of the classification results of the two-layer CRF approach based on superpixels of the size 20 x 20 with the ground truth is presented in table 2. The results for the class *residential* are quite good, with completeness and correctness values better than 85%. Lower completeness and correctness values for the class *street* mainly result from problems in the discrimination between *street* and *residential*. The classes *railway*, *water*, *forest*, *agriculture* and *others* are underrepresented in the training data, in fact only 5% of the objects in the training data belong to these classes. As a consequence, the discrimination of these classes is difficult, which may lead to lower correctness and especially completeness values. The class *agriculture* achieves a completeness and correctness value of better than 70%.

This contribution has been peer-reviewed.
doi:10.5194/isprsarchives-XL-3-17-2014

		Classification							Comp. [%]
		res.	*street*	*rail.*	*water*	*agr.*	*forest*	*others*	
Reference	*res.*	60.3	4.2	--	--	0.1	0.1	1.2	**91.6**
	street	3.3	14.7	--	--	0.3	0.1	0.6	**77.5**
	rail.	0.2	0.7	0.1	--	--	--	0.1	**11.4**
	water	0.1	0.2	--	0.3	--	--	0.1	**32.6**
	agr.	0.1	0.3	--	0.1	3.0	0.1	0.5	**73.2**
	forest	0.1	0.2	--	0.1	0.2	0.7	0.2	**49.7**
	others	3.2	0.9	--	--	0.4	0.1	3.2	**40.6**
Corr. [%]		**89.6**	**69.5**	**53.9**	**60.9**	**75.9**	**59.9**	**54.6**	

Table 2: Confusion matrix (in [%] of the total number of objects used for testing), completeness (comp.) and correctness (corr.) values [%] for the land use classes *res.*, *street*, *rail.*, *water*, *agr.*, *forest* and *others* obtained by classification applying the two-layer CRF approach based on superpixels of size 20 x 20.

A quantitative evaluation of the comparison of the results obtained by the two-layer CRF approach for two different sizes of the superpixels is presented in table 3. The two-layer CRF model based on superpixels of size 20 x 20 achieves a mean overall accuracy of 82.2% and a mean kappa index of 65.2%. In comparison to the results based on larger superpixels the overall accuracy and the kappa index are improved of more than 3.9% and 7.9%, respectively. The class *residential area* shows the lowest increase in completeness of about 1.3% and in correctness of about 2.2%. Large increases in completeness of more than 10% are achieved for the classes *street*, *water* and *agriculture*, which are accompanied by a smaller increase in correctness. The completeness and correctness values are also improved for the classes *forest* and *others*. For the class *railway* the completeness decreases of about 0.8%, but this goes along with a large improvement of the correctness of more than 20%.

		$\mathbf{CRF}_{\mathbf{multi,\ 50x50}}$		$\mathbf{CRF}_{\mathbf{multi,\ 20x20}}$	
		Comp. [%]	**Corr. [%]**	**Comp. [%]**	**Corr. [%]**
Land use class.	**res.**	90.3	87.4	91.6	89.6
	street	66.1	64.6	77.5	69.5
	rail.	12.2	26.8	11.4	53.9
	water	20.9	32.7	32.6	60.9
	agr.	63.2	64.3	73.2	75.9
	forest	44.0	40.9	49.7	59.9
	others	36.1	46.8	40.6	54.6
OA [%]		**78.3**		**82.2**	
Kappa [%]		**57.3**		**65.2**	

Table 3: Overall accuracy [%], kappa index [%], completeness (comp.) and correctness (corr.) values [%] for the land use classes *res.*, *street*, *rail.*, *water*, *agr.*, *forest* and *others* obtained by classification applying the two-layer CRF approach based on superpixels of size 50 x 50 ($\mathrm{CRF}_{\mathrm{multi,\ 50x50}}$) and of size 20 x 20 ($\mathrm{CRF}_{\mathrm{multi,\ 20x20}}$).

5. CONCLUSION

We propose a two-layer Conditional Random Field model for simultaneous classification of land cover and land use. The CRF represents a consistent model, where the statistical dependencies between land cover and land use are explicitly modelled and learned from training data. Preliminary results show that the presented approach yields good accuracies for the land use classes *residential area* and *street*. A lower accuracy is achieved for land use classes that occur less frequently in the test area. Furthermore, we have shown that reducing the size of the superpixels has a positive influence on the classification accuracy. Nevertheless, further enhancements are required in order to improve the classification result.

In future work, we will further investigate the influence of the size of the superpixels on the classification result in order to determine a level of detail, which represents a good trade-off between accuracy and computation time. In this context, we will also address further problems associated with the superpixels, for instance the congruence of superpixels with a land use object as well as inaccuracies in the training data. Currently, a land use object is connected with all spatially overlapping superpixels not taken into account the degree of overlap. Thus, some superpixels are assigned to land use objects, even though they cover mostly neighbouring land use objects. This may have a misleading effect on the inference, which, for instance, can be reduced by defining a threshold for the spatial overlap, or even avoided by restricting the superpixels to coincide with land use object boundaries. Another obvious test is a comparison of the new two-layer CRF approach with a two-step processing strategy presented in our previous work (Albert et al., 2014).

Furthermore, the “winner-takes-all”-strategy for the assignment of the ground truth label to each superpixel leads to inaccuracies in the training data. In future work, we will consider uncertain superpixels, for instance by eliminating uncertain training samples or considering the uncertainty of training data in the classification approach.

Remaining problems may also result from the fact that for some relations we currently have only a low number of training samples, thus not all statistical dependencies are properly represented in the training data. Therefore, we want to apply our approach on more test areas with different characteristics and more training data, especially for currently underrepresented class relations. Moreover, complex dependencies, like the composition of several land cover classes within a land use object, cannot be modelled explicitly using pair-wise potentials. In this context, we aim to investigate whether inter-layer interaction potentials can be modelled more appropriately with higher order potentials using a suitable model, which on the one hand, can capture the complex dependencies between land cover and land use, and on the other hand, allows efficient inference.

Currently we only differentiate a very small number of different land use classes, corresponding to the coarsest semantic level of the geospatial data base. In our future work, we also aim to determine the maximum level of semantic resolution of the land use classes which still delivers acceptable results.

Finally, we strive to embed the presented method into the overall task of updating the given geospatial database, In this context we will also work on automatically inferring changes to the geometric delineation of the geospatial objects assumed to be correct at this stage, e. g. by deforming the current object outlines and by splitting and merging objects.

ACKNOWLEDGEMENT

We gratefully thank the Landesamt für Geoinformation und Landesvermessung Niedersachsen (LGLN) and the Landesamt für Vermessung und Geoinformation Schleswig Holstein (LVermGeo) for providing data and support of this project.

REFERENCES

Albert, L., Rottensteiner, F., Heipke, C., 2014. Land Use Classification using Conditional Random Fields for the Verification of Geospatial Databases. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. II-4, pp. 1-7.

Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV), 2008. ALKIS®-Objektartenkatalog 6.0. Available online (accessed 16/07/2014): http://www.adv-online.de/AAA-Modell/Dokumente-der-GeoInfoDok/

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P. & Susstrunk, S., 2012. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34(11), pp. 2274-2282.

Boykov, Y. and Jolly, M.-P., 2001. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In: *Proc. of International Conference on Computer Vision,* vol.1, pp. 105-112.

Breiman, L., 2001. Random Forests. *Machine Learning* 45, pp.5-32.

Frey, B. and MacKay, D., 1998. A revolution: Belief propagation in graphs with cycles. In: *Advances in Neural Information Processing Systems,* vol. 10, pp. 479-485.

Haralick, R. M., Shanmugan, K., Dinstein, I., 1973. Texture features for image classification. *IEEE Transactions on Systems, Man and Cybernetics* 3, pp. 610-621.

Helmholz, P. 2012. Verifikation von Ackerland- und Grünlandobjekten eines topographischen Datensatzes mit monotemporalen Bildern. Phd Thesis, *Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover*, Nr. 299.

Hermosilla, T., Ruiz, L. A., Recio, J. A., Cambra-López, 2012. Assessing contextual descriptive features for plot-based classification of urban areas. *Landscape and Urban Planning* 106(1), pp. 124-137.

Hoberg, T., Rottensteiner, F., Heipke, C., 2012. Context models for CRF-based classification of multitemporal remote sensing data. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. I-7, pp. 128-134.

Kohli, P., Ladicky, L., Torr, P., 2009. Robust Higher Order Potentials for Enforcing Label Consistency. *International Journal of Computer Vision* 82(3), pp. 302-324.

Kosov, S., Rottensteiner, F., Heipke, C., 2013. Sequential Gaussian Mixture Models for two-level Conditional Random Fields. In: *Proc. of the 35th German Conference on Pattern Recognition (GCPR),* LNCS 8142, Springer, Heidelberg, pp. 153-163.

Kumar, S., Hebert, M., 2006. Discriminative Random Fields. *International Journal of Computer Vision* 68(2), pp. 179–201.

Niemeyer, J., Rottensteiner, F., Sörgel, U., 2014. Contextual classification of lidar data and building object detection in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing* 87, pp.152-165.

Rutzinger, M., Rottensteiner, F., Pfeifer, N., 2009. A comparison of evaluation techniques for building extraction from airborne laser scanning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 2(1), pp. 11-20.

Schindler, K., 2012. An overview and comparison of smooth labeling methods for land-cover classification. *Transactions on Geoscience and Remote Sensing* 50, pp. 4534 – 4545.

Shotton, J., Winn, J., Rother, C., Criminisi, A., 2009. TextonBoost for image understanding: multi-class object recognition and segmentation by jointly modelling texture, layout, and context. *International Journal of Computer Vision* 81(1), pp. 2-23.

Yang, M. Y., Förstner, W., 2011. A hierarchical conditional random field model for labeling and classifying images of man-made scenes. *ICCV Workshop on Computer Vision for Remote Sensing of the Environment 2011*, IEEE, pp. 196-203.

This contribution has been peer-reviewed.
doi:10.5194/isprsarchives-XL-3-17-2014