

LAND USE CLASSIFICATION USING CONDITIONAL RANDOM FIELDS FOR THE VERIFICATION OF GEOSPATIAL DATABASES

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

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

Commission IV, WG IV/1

KEY WORDS: Contextual classification, Conditional Random Fields, geospatial land use database

ABSTRACT:

Geospatial land use databases contain important information with high benefit for several users, especially when they provide a detailed description on parcel level. Due to many changes connected with a high effort of the update process, these large-scale land use maps become outdated quickly. This paper presents a two-step approach for the automatic verification of land use objects of a geospatial database using high-resolution aerial images. In the first step, a precise pixel-based land cover classification using spectral, textural and three-dimensional features is applied. In the second step, an object-based land use classification follows, which is based on features derived from the pixel-based land cover classification as well as geometrical, spectral and textural features. For both steps, the potential of the incorporation of contextual knowledge in the classification process is explored. For this purpose, we use Conditional Random Fields (CRF), which have proven to be a flexible, powerful framework for contextual classification in various applications in remote sensing. The results of the approach are evaluated on an urban test site and the influence of different features and models on the classification accuracy is analysed. It is shown that the use of CRF for the land cover classification yields an improved accuracy and smoother results compared to independent pixel-based approaches. The integration of contextual knowledge also has a remarkable positive effect on the results of the land use classification.

1. INTRODUCTION

1.1 Motivation

Geospatial land use databases contain important information with high benefit for several users. The number of possible applications of land use information increases with a higher level of detail, in terms of small geometrical entities as well as a high diversity of land use classes. The important drawback of these databases is the high effort required for the update process, which is necessary because of fast changes of the land use due to urban growth and land use conversion. As a consequence, these land use databases become outdated quickly. This observation motivates the development of an automatic update process for large-scale land use databases.

This paper presents a two-step approach for the automatic verification of land use objects of a geospatial database based on current high-resolution aerial images. In the first step, a precise pixel-based land cover classification using spectral, textural and three-dimensional features is applied, which assigns a land cover label to each pixel. The second step consists of an object-based land use classification, which is based on features derived from the pixel-based land cover classification as well as on geometrical, spectral and textural features. The objects underlying this second step are land use parcels obtained from the geospatial database to be verified. Finally, the verification of land use objects is done by a simple comparison of the old, possibly outdated land use class to the classification result, thus identifying contradictions between the geospatial database and current remote sensing data. This paper focuses on the two classification steps being prerequisite for the final verification. As there are naturally inherent relations between neighbouring pixels as well as between neighbouring land use objects, the

integration of contextual knowledge promises to form a considerable contribution to the classification process. The rather homogeneous appearance of pixels leads to this assumption for the land cover classification, but even for the land use objects, some land use classes are more likely to occur next to each other than others, and some are even more restricted by urban planning rules. For both steps, the potential of the incorporation of contextual knowledge in the classification process is explored in this paper. For this purpose, we use Conditional Random Fields (CRF) (Kumar & Hebert, 2006), which provide a flexible, powerful framework for contextual classification in various classification tasks in computer vision and remote sensing.

1.2 Related Work

Several different approaches for the verification of land use data bases exist. We can distinguish methods directly classifying objects from the database based on features extracted from image data within the object boundaries (Walter, 2004) from approaches which first carry out a pixel- or segment-based land cover classification and then transfer the classification results to the database objects (Helmholz, 2012; Hermosilla et al., 2012). In this paper, we will follow the latter strategy. Apart from the general strategy used for land use classification, approaches differ with respect to feature definition, classifiers applied and input data. Walter (2004) performs a Maximum Likelihood classification using spectral features derived from satellite images. Helmholz (2012) focuses on the verification of cropland and grassland objects. The discrimination of these classes results from a classification using Support Vector Machines based on spectral, textural and structural features derived from satellite images. Contextual relationships of land cover areas within a land use object can be analysed, like a

* Corresponding author.

building coverage ratio (Van de Voorde et al., 2009). The incorporation of contextual features in land use classification has already been shown to improve the classification accuracy. Hermosilla et al. (2012), combining aerial images and LiDAR data, apply a contextual classification, which is realized by considering contextual features in the classification process. They define contextual features at two levels, referred to as internal and external context features. The internal context features describe the relations between different land cover elements within a land use object. The external context features describe each object with respect to the common properties of neighbouring objects. Thus, the context information is implicitly integrated in the classification process by the contextual features. On the other hand, CRF offer the possibility to model relations between objects directly, thus explicitly considering context in the classification process. CRF have already been applied for several tasks in photogrammetry and remote sensing, e.g. point cloud classification (Niemeyer et al., 2014), multi-temporal classification of optical satellite images (Hoberg et al., 2012) and land cover classification from aerial images (Schindler, 2012). Each of these papers highlights the improved classification performance of this context-based classifier. To the best of our knowledge no approach using CRF for land use classification exists.

1.3 Contribution

This paper focuses on a consistent statistical approach for land use verification by considering contextual knowledge in the classification process. We apply CRF for both, land cover and land use classification. Land cover works on the pixel level, whereas the second step is based on a classification of land use objects given the results of the first step. To our knowledge, this is the first approach making use of CRF for the classification of land use objects. The consideration of contextual knowledge is supposed to lead to improved classification accuracy and smoother results.

After introducing the CRF framework in section 2, the methodology used for both classification steps is presented in section 3. Section 4 describes the features used in both stages of our work flow. A thorough evaluation of both steps is presented in section 5. Finally, conclusions are given in section 6.

2. CONDITIONAL RANDOM FIELDS

Conditional Random Fields are 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 modeling the posterior probability $P(\mathbf{y}|\mathbf{x})$ of the label vector \mathbf{y} given the observed data \mathbf{x} :

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

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 indicates how likely a node i belongs to a class y_i given the observations \mathbf{x} . The interaction potential 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 classifiers 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 proven to work well, for instance a Random Forest (RF) classifier (Schindler, 2012). The models applied for the interaction potential are often more simple, favouring identical labels and penalising label changes. 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. CLASSIFICATION

The automatic verification of a large-scale geospatial land use database is achieved by a two-step approach. Both steps contain a supervised contextual classification using CRF. The workflow of each step is quite similar and can be subdivided into three stages. Firstly, we extract a suitable set of features for each specific task and image site. Secondly, the classifiers are learned based on representative training data. Thirdly, we build a graphical model for the test data, for instance an image, and determine the optimal label configuration in the inference step. The CRF differ for both classification tasks with respect to the graph structure, the parameter choice and partly concerning the models for the association and interaction potentials. Moreover, both classification tasks need different input data and features for the discrimination of different class structures.

3.1 Land Cover Classification

The goal of the first step is a pixel-based land cover classification of urban and rural scenes. Thus, the nodes of the graphical model correspond to the individual pixels of an image and the edges model spatial dependencies in a four-neighbourhood of each pixel.

3.1.1 Association potential: As stated above, the association potential indicates how likely a node n_i belongs to a class y_i given the observations \mathbf{x} . In this context, the observations are represented by the site-wise feature vectors $\mathbf{f}_i(\mathbf{x})$ (Kumar and Hebert, 2006), which may depend on all data. The association potential for node i is proportional to the probability of y_i given the site-wise feature vector $\mathbf{f}_i(\mathbf{x})$, i.e. $\varphi_i(y_i, \mathbf{x}) \propto P(y_i | \mathbf{x})$. It is possible to choose any arbitrary classifier with a probabilistic output for defining $P(y_i|\mathbf{x})$. Here, we apply the Random Forest

classifier for the association potential. This was first introduced by Breiman (2001) and has been shown to be a powerful classifier in remote sensing applications (e.g. Schindler, 2012). RF is a discriminative classification method, whose output can be easily converted into a probabilistic measure. RF generates an ensemble of randomized decision trees according to the bootstrap principle. In the classification, the features of an unknown sample of the dataset are presented to each tree, and each tree casts a vote for the most likely class. These votes are used to define a probability measure by dividing the sum of all votes for a class by the total number of trees. The main parameters that have to be adapted are the maximum number of samples used for training, the maximum depth and the number of trees in the forest (OpenCV, 2014).

3.1.2 Interaction potential: The interaction potential models the dependencies of the labels at adjacent nodes n_i and n_j , considering the data \mathbf{x} . In this case, the neighbourhood consists of four direct neighbours of each pixel in an image grid. The data are taken into account in the form of an interaction feature vector $\boldsymbol{\mu}_{ij}(\mathbf{x})$ for each edge. The simplest model for the interaction potential is the Potts model, where the smoothing effect only depends on the class labels of neighbouring image sites. This model favours identical labels and penalises label changes. More sophisticated models additionally take into consideration the data. They model the interaction potential based on the probability of both labels y_i and y_j being identical given $\boldsymbol{\mu}_{ij}(\mathbf{x})$, i.e. $\psi_{ij}(y_i, y_j, \mathbf{x}) \propto P(y_i = y_j | \boldsymbol{\mu}_{ij}(\mathbf{x}))$ (Kumar and Hebert, 2006). The contrast-sensitive Potts model belongs to this group of models (Kumar and Hebert, 2006). The interaction features define the degree of smoothing and consist of the Euclidian distance d_{ij} between the site-wise feature vectors $\mathbf{f}_i(\mathbf{x})$ and $\mathbf{f}_j(\mathbf{x})$ of two adjacent nodes n_i and n_j . We use an adapted version of the contrast-sensitive Potts model:

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

This model will result in a data-dependent smoothing of the resultant label image. The parameter $l_1 \in [0; 1]$ specifies the degree to which the smoothing effect will depend on the data. If l_1 is set to one, the model corresponds to a Potts model (Schindler, 2013) which will smooth the classification result independently from the image content. If l_1 equals zero the degree of smoothing is completely determined by the data-dependent term. The parameter σ^2 is the mean value of the squared distances d_{ij}^2 and is determined during training. The matrix $h(y_i, y_j)$ is based on a histogram of the co-occurrences of the class labels y_i and y_j at neighbouring image sites i and j , where the rows are scaled so that the largest value in a row is one in order to avoid a bias for classes covering a large area in the training data (Kosov et al., 2013). The contrast-sensitive Potts model has shown to produce good results (Schindler, 2012) and represents a good trade-off between accuracy and computation time, therefore we apply it for land cover classification.

3.2 Land Use Classification

The goal of the second step is to assign a land use class to each object obtained from a geospatial land use database, in order to identify contradictions between current remote sensing data and the outdated geospatial land use database. As in the first step, the graphical model consists of nodes and edges. However, in this step the nodes correspond to the land use objects and the edges model spatial relations between neighbouring land use

objects. The neighbourhood of an object is composed of its first order neighbours, i.e. all objects that share a common boundary with the given object. This classification is based on features derived from aerial images and a pixel-based land cover classification described in section 4.2.

3.2.1 Association potential: Again, we choose the RF classifier for the association potential of the land use classification. However, the parameter maximum number of samples used for training has to be adapted due to an overall smaller number of objects, and therefore less potential samples for training compared to the land cover classification.

3.2.2 Interaction potential: The incorporation of context into the land use classification does not intend to yield a pure smoothing effect, as it may be desirable for the land cover classification. Here, the interaction potentials should support more probable class relations, where the probability should result from real-world occurrences given the observations, learned from representative training data. Thus, it is reasonable to model the interaction potential as the joint posterior probability of both labels y_i and y_j given $\boldsymbol{\mu}_{ij}(\mathbf{x})$, i.e. $\psi_{ij}(y_i, y_j, \mathbf{x}) \propto P(y_i, y_j | \boldsymbol{\mu}_{ij}(\mathbf{x}))$. This group of models treats the estimation of the interaction potential as a standard classification task, where the relations are learned. Similarly to the association potential, it is possible to choose any classifier with a probabilistic output for this group of models. We choose the RF classifier for the interaction potential as well. In this context, each pair of land use classes is considered as a single class. The classifier should support more probable class relations given the data. The data are taken into account in the form of an interaction feature vector $\boldsymbol{\mu}_{ij}(\mathbf{x})$, which is defined as the concatenated site-wise feature vectors $\mathbf{f}_i(\mathbf{x})$ and $\mathbf{f}_j(\mathbf{x})$ of two adjacent nodes n_i and n_j . Other feature definitions are possible, such as the element-wise difference vector of the site-wise feature vectors or its absolute value. We use the concatenated feature vector, because in our previous work it has shown to give slightly better results.

3.3 Training and Inference

Training and inference are similar for both classification tasks. The classifiers for the association and interaction potential are trained separately on representative training data. The edge weight ω and the parameter l_1 of the contrast-sensitive Potts model are defined by the user. If it is required, the parameters could be determined by cross-validation (Shotton et al., 2009). The RF classifiers as well as the parameter σ of the contrast-sensitive Potts model are learned from training data. The training of the classifier of the interaction potentials requires fully labelled training data in order to learn the relations between adjacent pixels or objects. As exact inference is computationally intractable (Kumar and Hebert, 2006), an approximate solution for the optimal label configuration is estimated. For this purpose, we apply the message passing algorithm Loopy Belief Propagation (Frey and MacKay, 1998).

4. FEATURE EXTRACTION

4.1 Land Cover Classification

An efficient classification requires an appropriate set of discriminative features. Our application is based on high-resolution aerial image data or derived products. These are digital orthophotos (DOP), digital surface models (DSM) and digital terrain models (DTM), which were derived from

multiple aerial images in a pre-processing step. The aerial images are assumed to have four channels, composed of one near-infrared channel besides three colour channels.

We use spectral, textural and three-dimensional features. Furthermore, some of the features are derived at different scales, so the feature set is complemented by multi-scale features. The first four spectral features are the original grey values of the image. Moreover, we use the normalized difference vegetation index (NDVI), derived from the near infrared and red band of the image, which is particularly suitable for the discrimination of vegetation from non-vegetation. Furthermore, the hue, saturation and intensity are used. All spectral features are computed for each single pixel, and in order to map the spectral characteristics of a local neighbourhood, we determine the mean value and the variance for the grey values, NDVI, hue, saturation and intensity from a local neighbourhood whose size depends on the image resolution. Here, we choose a local neighbourhood of 13×13 pixels. In addition, the magnitudes and orientations of the image gradients of the intensity image are estimated. For the textural features, we use features derived from the Grey Level Co-Occurrence Matrix (GLCM) proposed by Haralick (1973), namely the energy, contrast, homogeneity and entropy. The GLCM describes the spatial distribution of the intensity values in a local neighbourhood in a certain direction and distance. Here, we choose a size of 5×5 pixels. The three-dimensional features consist of the normalized digital surface model (nDSM), which is the difference between a DSM and a DTM, and derived features. The nDSM describes the height above ground, thus indicating whether the pixel belongs to an elevated object, such as a building or a tree, or is otherwise part of the ground surface. Furthermore, the mean and Gaussian curvatures of the nDSM as well as the magnitudes and orientations of its gradients are estimated. In addition, we determine multi-scale features filtering the original spectral features with Gaussian filters of several widths σ ; we use σ values of 2, 5 and 10. In total, we use 58 features for the classification, which are derived for each image site i and combined in the feature vector $\mathbf{f}_i(\mathbf{x})$ for each node n_i . The features are scaled to the interval $[0;1]$.

4.2 Land Use Classification

For this step, we additionally need the GIS-objects of the geospatial land use database, which represent the entities to be classified. The land use objects are assumed to have a polygonal representation. We extract an appropriate set of descriptive features, which characterizes single objects with respect to their spectral, textural, geometrical and three-dimensional properties. In addition, features are derived from the pixel-based land cover classification results. These features describe the internal context concerning the composition of different land cover elements within a land use object. Additionally, the number of neighbouring land use objects is estimated, which can be interpreted as an external context feature.

The spectral features describe the overall radiometric characteristics of each land use object. We compute the mean, standard deviation, minimum and maximum of the NDVI, hue, saturation and intensity values considering all pixels within an object. The textural features are also derived from the GLCM. We use the same features as for the land cover classification with the difference that the GLCM results from the intensity values of all pixels inside each object. The geometrical features characterize the geometric shape of the objects. We extract the features area, perimeter, compactness, shape index and fractal dimension (Krummel et al., 1987) from the polygon defining

the spatial extent of each object. In addition, the mean value, standard deviation, minimum and maximum values of the height above ground inside each object are determined, which are derived from the nDSM. The last group of features are inspired by Hermosilla (2012) and describe the area ratio of the different land cover segments to the total object area as well as the land cover areas in total. For instance, the ratio of the built-up area to the total object area, usually referred to as building coverage ratio (Van de Voorde et al., 2009), forms one feature. Such a ratio is also computed for the land cover classes *sealed area*, *bare soil*, *water*, *car* and *vegetation*, which includes the land cover classes *grass* and *tree*. Altogether, we combine 54 features in the feature vector $\mathbf{f}_i(\mathbf{x})$ for each node n_i . The features are scaled to the interval $[0;1]$.

5. EXPERIMENTS

5.1 Test Data and Test Setup

The performance of the presented approach is evaluated on an urban test site, which is located in the city of Hameln. The test area is mainly characterised by residential areas with detached houses as well as by densely built-up areas in the centre of the city. Besides, there are also industrial and rural areas and a river. The test site covers an area of $2 \text{ km} \times 6 \text{ km}$. The input data of the approach consist of a DOP, a DSM, a DTM and GIS-objects of the German geospatial land use database forming a part of the Authoritative Real Estate Cadastre Information System (ALKIS), corresponding to cadastral parcels. The DOP has a ground sampling distance of 20 cm and was acquired in spring, thus trees appear without leaves. The DSM and the DTM contain height information in a 0.5 m and a 5 m grid, respectively. The test data set for the land cover classification consists of 37 image tiles of $200 \text{ m} \times 200 \text{ m}$ each with ground truth obtained by manual annotation. These tiles are uniformly distributed over the test area and represent the main characteristics inherent in the data set. The test data for land use classification consist of a manually corrected version of the land use database for the whole test area, divided into 12 blocks of about $1000 \text{ m} \times 1000 \text{ m}$ each. In land cover classification, we distinguish the nine land cover classes *building (build.)*, *sealed area (seal.)*, *bare soil (soil)*, *grass*, *tree*, *water*, *rails*, *car* and *others*, which typically appear in urban and rural scenes. The definition of land use classes has to comply with the specifications of the German geospatial land use database. Hence, we distinguish the seven land use classes *residential (res.)*, *street*, *water*, *railway (rail.)*, *agriculture (agr.)*, *forest* and *others*, corresponding to the primary groups of this particular object type catalogue of land use objects. In both steps, we use 200 trees of a maximum depth of 25 for the RF classifier. The maximum number of samples to be used for training is set to 100,000 for the land cover classification and to 5,000 for the land use classification. The edge weight ω is set to 2 in both cases.

The evaluation of both steps is based on cross-validation. Concerning the land cover classification, the data are divided into seven groups, where each group represents all characteristics of the test site. For the land use classification, the training data are divided into 12 groups. In each test run, one group is used as test data for the quality evaluation and all others serve as training data. We repeat this procedure so that each group is used for testing (and, thus contributes to the evaluation) once. The comparison of the classification results of all test runs to ground truth data results in a confusion matrix and derived quality measures, more precisely overall accuracy,

kappa index, correctness and completeness (Rutzinger et al., 2009), also referred to as user's and producer's accuracy. Furthermore, the evaluation of the land use classification comprises an analysis of the influence of certain features on the classification accuracy.

5.2 Evaluation of Land Cover Classification

An example of the result of the land cover classification is shown in figure 1. A first visual evaluation shows that most of the buildings and streets in the scene are detected correctly. Problems occur at trees and at building boundaries, which seems to result from missing leaves and inaccuracies in the DSM, respectively. A visual comparison between the results of an independent RF-based classification of all pixels (using $\omega = 0$ in Eq. 1) and the CRF-classification shows that the result of the CRF approach is much smoother compared to the result of the independent classification.



Figure 1: Ground truth (left), result of a RF classification (centre) and of the CRF approach (right). Colours: *build.* (red), *seal.* (grey), *soil* (brown), *grass* (green), *tree* (dark green), *car* (red), *others* (pink).

The quantitative evaluation is based on the confusion matrix, shown in table 1. The mean overall accuracy is 81.3% and the mean kappa index is 76.2% for the CRF approach using the contrast-sensitive Potts model.

		Classification							Comp. [%]
		build.	seal.	soil	grass	tree	water	car	
Reference	build.	22.5	1.2	--	0.2	0.3	--	0.3	90.1
	seal.	1.5	17.6	0.1	0.6	0.4	--	1.3	79.2
	soil	--	0.2	1.9	0.4	0.2	--	--	68.7
	grass	0.4	1.2	0.6	22.7	2.9	0.1	0.1	79.7
	tree	0.2	0.6	0.2	2.2	13.2	0.1	--	79.3
	water	--	0.1	--	0.1	0.1	1.8	--	85.9
	car	0.1	0.2	--	--	--	--	1.2	76.0
Corr. [%]		90.7	82.3	66.9	86.4	76.8	89.5	38.7	

Table 1: Confusion matrix obtained by classification using the CRF approach with correctness (corr.) and completeness (comp.) values [%] for the land cover classes *build.*, *seal.*, *soil*, *grass*, *tree*, *water* and *car*. The class *rails* is omitted.

The best completeness and correctness values are achieved for the class *building*, but also the classes *sealed area*, *grass*, *tree* and *water* achieve good completeness and correctness values. Lower values for the class *bare soil* are caused by an overall smaller number of training samples for this class, thus not sufficiently representing the whole range of characteristics of this class. A problem turns out to be the discrimination of cars from sealed area. Although most of the car pixels in the ground truth are found, as confirmed by a good completeness, most of the classified car pixels actually do not correspond to a car, reflected in a low correctness value. This is amongst other factors caused by the fact that rows of individual cars are merged, thus falsely including sealed area in between.

Table 2 shows a comparison of the results achieved by different models for the interaction potential. The approach referred to as *RF* is an independent pixel-based classification without

considering interactions, where the classification is only based on the association potentials obtained by a RF classifier. Furthermore, the results of two models for the interaction potential described in section 3.1.2 are listed. All approaches apply a RF classifier with identical parameters for the association potential. The best result is achieved by the use of the contrast-sensitive Potts model, which increases the overall accuracy by 1.1% compared to an increase of 0.8% by the use of the Potts model for the interaction potential. In fact, the improvement become more obvious when analysing the completeness and correctness of classes covering smaller areas in the training data. The correctness of *bare soil* and *water* are improved by 5.7% and 2.2%, respectively, while maintaining completeness; for *car*, the correctness increases by 5%, but this is accompanied by a decrease in completeness of 3.3%.

	RF	CRF-Potts	CRF-CS Potts
Overall accuracy [%]	80.2	81.0	81.3
Kappa index [%]	74.9	75.8	76.2

Table 2: Comparison of the overall accuracy and kappa index of results obtained by different models for the interaction potential: independent RF classification (*RF*), Potts model (*CRF-Potts*) and contrast-sensitive Potts model (*CRF-CS Potts*).

5.3 Evaluation of Land Use Classification

The confusion matrix of the result obtained by the land use classification is presented in table 3. The CRF approach achieves a mean overall accuracy of 85.5% and a mean kappa index of 73.4%. The results for the class *residential* are quite good, with completeness and correctness values better than 85%. The completeness value of the class *street* also reaches this value, but only about 70% of these objects are correct. This is caused by 4.2% of objects classified as *street* while actually corresponding to class *residential*. Lower correctness and especially completeness values can be explained by the fact that the amount of training data is not sufficient for an adequate discrimination of the corresponding classes. In total, only about 5% of the objects in the training data belong to the classes *railway*, *water*, *agriculture* and *forest*.

		Classification						Comp. [%]
		res.	street	rail.	water	agr.	forest	
Reference	res.	58.8	4.2	--	--	0.1	--	89.2
	street	1.6	16.3	0.1	--	0.2	--	85.6
	rail.	0.1	0.5	0.2	--	--	--	21.1
	water	0.1	0.1	--	0.4	--	--	51.2
	agr.	--	0.3	--	0.1	3.0	--	75.2
	forest	--	0.2	--	--	0.2	0.6	38.4
Corr. [%]		96.1	73.3	65.0	74.6	81.4	68.5	

Table 3: Confusion matrix obtained by classification using the CRF approach with correctness (corr.) and completeness (comp.) values [%] for the land use classes *res.*, *street*, *rail.*, *water*, *agr.* and *forest*.

In order to assess the influence of the incorporation of contextual knowledge in the classification process, we compare the results of the CRF approach with a classification based exclusively on association potentials. In fact, the quantitative evaluation yields only a small improvement of the overall accuracy of about 0.5% and of the kappa index of 1.2%. The benefits of including context become more obvious when analysing the completeness and correctness of some of the classes only having a small number of instances. In particular, the completeness and correctness of *water* are improved by 10.5% and 3.6%, respectively; for *agriculture*, the improvement is about 3% both in completeness and correctness. However, this is contrasted by a decrease in completeness of 13.2% for

forest, partly compensated by an increase in correctness of 9.1% for that class. Figure 2 shows two instances where context helps to improve the classification results.

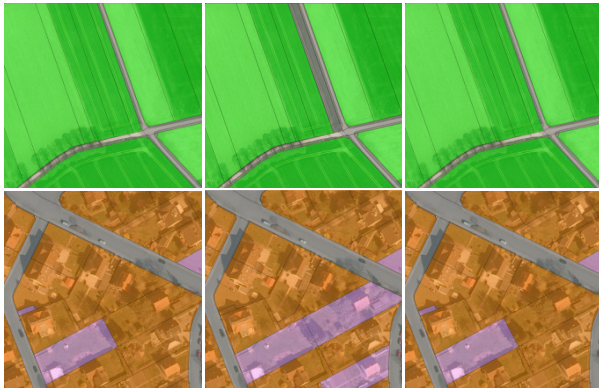


Figure 2: Ground truth (left column), classification results exclusively based on association potentials (centre column), results of the CRF approach (right column). Colours: *res.* (red), *street* (grey), *agr.* (green), *forest* (dark green), *others* (pink).

The relevance of features results from a permutation importance measure (Breiman, 2001), which can be obtained from the RF classifier. An analysis of the overall importance values per feature shows a large influence of the contextual features on the classification accuracy, as the two most relevant features belong to this group, namely the ratio of the areas covered by the land cover classes *building* and *vegetation* to the total area. Figure 3 shows the predicted overall accuracy achieved by progressively including features in the classification according to their importance value, starting with the most relevant one. The predicted overall accuracy converges to a value of approximately 85%, which is already achieved with the use of the 20 most relevant features.

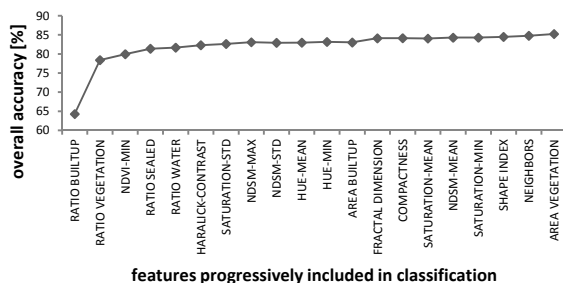


Figure 3: Overall accuracy as a function of the features used for classification, progressively included in classification according to their importance value.

6. CONCLUSION

We present a two-step land use verification approach using Conditional Random Fields. The evaluation shows an improved overall accuracy and in case of the land cover classification smoother results. We identify a potential of the use of CRF as context-based classifiers for the task of land use classification, but further enhancements are required. In future work, we will examine if both classification tasks can be integrated into one graphical model. Furthermore, we will verify our approach on more test areas with different characteristics and more training data, especially for currently underrepresented classes. Moreover, we plan a step-wise refinement of the land use classes in order to investigate the maximum level of semantic resolution which still delivers acceptable results.

ACKNOWLEDGEMENT

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

REFERENCES

- 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.
- 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.
- Krummel, J.R., Gardner, R.H., Sugihara, G., O'Neill, V., Coleman, P.R., 1987. Landscape patterns in a disturbed environment. *OIKOS* 48(3), pp.321-324.
- 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.
- OpenCV, 2014. Random Trees. OpenCV Reference Manual. Available online (accessed 15/01/2014): http://docs.opencv.org/modules/ml/doc/random_trees.html
- 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.

Van de Voorde, T., Van de Kwast, J., Engelen, G., Binard, M., Cornet, Y., Canters, F., 2009. Quantifying intra-urban morphology of the Greater Dublin area with spatial metrics derived from medium resolution remote sensing data. In: *Proc. of the 7th joint urban remote sensing event*, Shanghai, China.

Walter, V., 2004. Object-based classification of remote sensing data for change detection. *ISPRS Journal of Photogrammetry and Remote Sensing* 58(3-4), pp.225-238.