CLASSIFICATION OF SETTLEMENT AREAS IN REMOTE SENSING IMAGERY USING CONDITIONAL RANDOM FIELDS

T. Hoberg*, F. Rottensteiner

IPI, Institute of Photogrammetry and GeoInformation, Leibniz Universitaet Hannover, Germany
(hoberg, rottensteiner)@ipi.uni-hannover.de

Commission VII, WG VII/4

KEY WORDS: Conditional Random Fields, contextual information, classification, satellite imagery, urban area

ABSTRACT:

Land cover classification plays a key role for various geo-based applications. Numerous approaches for the classification of settlements in remote sensing imagery have been developed. Most of them assume the features of neighbouring image sites to be conditionally independent. Using spatial context information may enhance classification accuracy, because dependencies of neighbouring areas are taken into account. Conditional Random Fields (CRF) have become popular in the field of pattern recognition for incorporating contextual information because of their ability to model dependencies not only between the class labels of neighbouring image sites, but also between the labels and the image features. In this work we investigate the potential of CRF for the classification of settlements in high resolution satellite imagery. To highlight the power of CRF, tests were carried out using only a minimum set of features and a simple model of context. Experiments were performed on an Ikonos scene of a rural area in Germany. In our experiments, completeness and correctness values of 90% and better could be achieved, the CRF approach was clearly outperforming a standard Maximum-Likelihood-classification based on the same set of features.

1. INTRODUCTION

1.1 Motivation

The detection of settlement areas in satellite imagery is the basis for many applications, e.g. regional planning, the observation of urban expansion, or disaster prevention and management. In optical remote sensing images settlement areas have a heterogeneous appearance because they consist of a large number of different objects such as buildings, trees, and roads. The variety of these objects results in specific local patterns in the images. Whereas these patterns make a spectral classification of such areas very difficult, they can at the same time be exploited to improve the classification result if they are properly modelled. It is the main goal of this paper to model the contextual information contained in the local patterns of image features to improve the accuracy that can be achieved in the classification of settlement areas. In order to do so, we want to use Conditional Random Fields (CRF) (Kumar & Hebert, 2006) because of their ability to consider contextual relations between both the class labels and the observed image features of the image sites (i.e., pixels or segments). For this purpose, we will use radiometric and texture features from multispectral Ikonos data, i.e. from imagery having a resolution of 4 m. The parameters of the CRF will be learned from training data, and we will assess the effects of using the context information on the classification results.

1.2 Related Work

The methods that can be applied to detect settlement areas in satellite images depend on the resolution of these images. In images having a resolution better than about 2.5 m, a settlement is decomposed into buildings, roads, vegetation, and other objects. Various classification techniques have been proposed to extract these object classes, e.g. (Gamba et al., 2007). In images of 2.5 – 10 m resolution, which are our main interest here, the individual objects can no longer be discerned except for large structures. Buildings, roads, and urban vegetation are merged into a class 'settlement' which is characterized by a very heterogeneous distribution of the spectral components of the respective pixels. Hyperspectral data may help to overcome this problem (Herold et al., 2003), but the more common approach is to introduce textural features into classification, because they are better suited to characterize settlements, e.g. (Cheriyadat et al, 2007; Zhong & Wang, 2007). Various textural features have been used for urban classification, e.g. features based on the Grey-Level Co-occurrence Matrix (GLCM) (Smits & Annoni, 1999; Cheriyadat et al., 2007; Zhong & Wang, 2007), normalised grey-level histograms (Shackelford & Davis, 2003), or features related to the distribution of gradient orientation (Zhong & Wang, 2007).

These features can be used in any classification scheme. In a Bayesian statistical setting, the features of individual image sites are considered to be conditionally independent, which leads to a separate classification of each of the individual sites (Bishop, 2006). This approach has been found to lead to a salt-and-pepper-like appearance of the classification results. In order to improve the situation, context can be taken into account in the classification process. The simplest way of doing so is by post-processing the original classification results, taking into account the distribution of class labels in a local neighbourhood, e.g. (Gamba & Dell'Acqua, 2003). A more sophisticated approach uses statistical models of context. Among these, Markov Random Fields (MRF) (Besag, 1986) have found many applications in pattern recognition and remote sensing, e.g. (Tupin & Roux, 2005; Gamba et al., 2007). MRF

* Corresponding author.

can be used for representing texture, e.g. (Paget & Longstaff, 1998). In a Bayesian context, the main contribution of MRF is to act as a smoothness term on the class labels via a model for their local statistical dependencies (Besag, 1986; Kumar & Hebert, 2006). The features extracted from different sites are still assumed to be conditionally independent, and the interaction between neighbouring image sites is restricted to the class labels. Conditional Random Fields (Kumar & Hebert, 2006) were developed to overcome these restrictions. CRF provide a discriminative framework that can also model dependencies between the data and interactions between the labels and the data. In their experiments with man-made structure detection in natural terrestrial images, Kumar and Hebert (2006) could show that CRF outperform MRF.

Up to now, hardly any work has been done on classifying remotely sensed data using CRF. Zhong and Wang (2007) analyse images from Quickbird and SPOT with a multiple CRF ensemble model for the detection of settlement areas. They apply CRF to five groups of texture features and then fuse these results. The fusion process itself is based on a MRF taking into account the conditional probabilities provided by each of the CRF. Lu et al. (2009) use CRF on LiDAR data for simultaneously classifying the LiDAR data into terrain- and off-terrain-points and estimating a Digital Terrain Model from the off-terrain points. He et al. (2008) use CRF for building extraction from SAR data. Of these works, our new method is most closely related to (Zhong & Wang, 2007). However, our model is simpler because it only employs a single CRF that is applied to a feature vector taking into account radiometric and textural characteristics of the image. As the local dependencies of image data and class labels are modelled by a CRF in a very general way (Kumar & Hebert, 2006), we do not think it is necessary to use a MRF in order to fuse the output of a set of CRF. In our experiments, the effects of including a statistical model of context based on CRF on the classification results will be assessed by comparing the results of our new method to a standard maximum likelihood classification based on the same set of features. The main focus of this paper is on the benefits of using CRF for modelling context in classification and not on finding an optimum set of features for describing settlements.

2. MODELLING CONTEXT IN CLASSIFICATION USING CONDITIONAL RANDOM FIELDS

In many classification algorithms the decision for a class at a certain image site is just based on information derived at the regarded site, where a site might be a pixel, a square block of pixels in a regular grid or a segment of arbitrary shape. In fact, the class labels and also the data of neighbouring sites are often very similar or show characteristic patterns. Incorporating contextual information of neighbouring sites should improve the classification accuracy. The method described in this paper uses CRF for that purpose. In this section we want to give a brief overview on the CRF framework that is based on (Kumar & Hebert, 2006) and (Vishwanathan et al., 2006).

2.1 Conditional Random Fields (CRF)

The classification problem to be solved can be described as follows. We have observed image data $\mathbf{y}$. The image consists of image sites $i \in S$, where S is the set of all image sites. For each image site we want to determine its class x_i from a set of pre-defined classes. The class labels of all image sites can be combined in a vector $\mathbf{x}$ whose i^{th} component is the class of an individual image site i. Probabilistic classification methods determine the class labels so that they maximise the conditional probability $P(\mathbf{x} \mid \mathbf{y})$ of the class labels $\mathbf{x}$ given the observed data $\mathbf{y}$. CRF provide a discriminative framework for directly modelling $P(\mathbf{x} \mid \mathbf{y})$, which reduces the complexity of the involved models (Kumar & Hebert, 2006):

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

In Equation 1, $i \in S$ is the index of an individual image site, N_i is a certain neighbourhood of image site i, and thus j is an image site that is a neighbour to i. Z is a normalisation constant required to make $P(\mathbf{x} \mid \mathbf{y})$ a probability. The exact determination of Z is computationally intractable, which is the reason why approximate methods have to be used to determine the parameters of the model in Equation 1 and to maximise $P(\mathbf{x} \mid \mathbf{y})$ in the classification stage. In the exponent of Equation 1, the *association potential* A_i links the class label x_i of image site i to the data $\mathbf{y}$. Unlike with MRF, the association potential for an image site i may depend on the entire image $\mathbf{y}$. Thus, the data from neighbouring image sites are no longer considered to be conditionally independent. The second term in the exponent of Equation 1 is the *interaction potential* I_{ij}. It is responsible for modelling the dependencies between the labels x_i and x_j of neighbouring sites i and j and the data $\mathbf{y}$. This dependency of the interaction potential on the data is the second advantage of CRF over MRF. In MRF the interaction terms just depend on the labels, so that in many applications they only act as a kind of smoothness prior on the labels (Kumar & Hebert, 2006).

Any application of the CRF framework has to define what constitutes an image site and which classes are to be discerned. Furthermore, a model for the association and interaction potentials has to be found. We choose the image sites to be square blocks of pixels in a regular grid. The side length s of these squares is a parameter to be set by the user. We are only interested in a binary classification, so $x_i \in \{-1; 1\}$, where $x_i = 1$ means that image site i belongs to class *settlement* and $x_i = -1$ means that it belongs to the background. We model the CRF to be isotropic and homogeneous, hence the functions used for A_i and I_{ij} are independent of the location of image site i.

2.2 Association Potential

The association potential indicates how likely a site i is to belong to a label x_i given the observed data $\mathbf{y}$ and ignoring the other image sites. Kumar and Hebert (2006) suggest local discriminative classifiers for modelling the association potential by linking the association potential to the conditional probability $P'(x_i \mid \mathbf{y})$ of class x_i at image site i given the data $\mathbf{y}$:

$$A_i(x_i, \mathbf{y}) = log\, P'(x_i|\mathbf{y}) \quad (2)$$

The image data $\mathbf{y}$ are usually represented by image features that are determined from the original grey levels of the image. In order to put into practice the dependency of the association potential from the whole image, Kumar and Hebert (2006) define a site-wise feature vector $\mathbf{f}_i(\mathbf{y})$ which, though being computed specifically for site i, may depend on the entire image $\mathbf{y}$; usually the feature vector will be influenced by the data in a local neighbourhood that is not identical to the neighbourhood used for the interaction potential. Kumar and Hebert (2006) suggest using general linear models for $P'(x_i \mid \mathbf{y})$. For that purpose a feature space mapping $\mathbf{\Phi}(\mathbf{f})$ is required. It transforms the site-wise feature vectors $\mathbf{f}_i(\mathbf{y})$ into another feature space of

higher dimensions so that the decision surface becomes a hyperplane. Let $\mathbf{h}_i(\mathbf{y}) = \mathbf{\Phi}(\mathbf{f}_i(\mathbf{y}))$ be the site-wise transformed feature vector, with $\mathbf{\Phi}(\mathbf{f}_i(\mathbf{y})) = [1, \Phi_1(\mathbf{f}_i(\mathbf{y})), \ldots \Phi_N(\mathbf{f}_i(\mathbf{y}))]^T$ and Φ_k being arbitrary functions. The dimension of the transformed feature space is $N+1$. In a generalised linear model, the conditional probability $P'(x_i \mid \mathbf{y})$ is described by Equation 3:

$$P'(x_i|\mathbf{y}) = \frac{1}{1+e^{-x_i \cdot \mathbf{w}^T \cdot \mathbf{h}_i(\mathbf{y})}} \tag{3}$$

where $\mathbf{w}$ is a vector of dimension $N+1$. Its components describe the weights of the transformed features. These weights are the parameters of the association potential that have to be determined in a training phase. Fixing the first component of $\mathbf{h}_i(\mathbf{y})$ to 1 accommodates the bias parameter in the linear model in the exponent of Equation 3 (Bishop, 2006).

2.3 Interaction Potential

The interaction potential is a measure for the influence of the data $\mathbf{y}$ and the neighbouring labels x_j on the class x_i of site i. It can be linked to the conditional probability $P''(x_i = x_j \mid \mathbf{y})$ for the occurrence of identical labels at sites i and j given the data $\mathbf{y}$:

$$I_{ij}(x_i, x_j, \mathbf{y}) = log\, P''(x_i = x_j|\mathbf{y}) \tag{4}$$

In the interaction potential, the data are represented by site-wise feature vectors $\boldsymbol{\psi}_i(\mathbf{y})$, which may have a different functional form than the vectors $\mathbf{f}_i(\mathbf{y})$ used for the association potential in order to accommodate features that are typical for neighbourhood dependencies. From the feature vectors $\boldsymbol{\psi}_i(\mathbf{y})$ and $\boldsymbol{\psi}_j(\mathbf{y})$ of two neighbouring sites a new vector of relational features $\boldsymbol{\mu}_{ij}(\mathbf{y}) = \boldsymbol{\mu}_{ij}(\boldsymbol{\psi}_i(\mathbf{y}), \boldsymbol{\psi}_j(\mathbf{y}))$ can be derived. Kumar and Hebert (2006) suggest concatenating the two vectors $\boldsymbol{\psi}_i(\mathbf{y})$ and $\boldsymbol{\psi}_j(\mathbf{y})$ or using some distance function. The interaction potential can be modelled as

$$I_{ij}(x_i, x_j, \mathbf{y}) = x_i x_j \mathbf{v}^T \boldsymbol{\mu}_{ij}(\mathbf{y}) \tag{5}$$

In Equation 5, the vector $\mathbf{v}$ contains the feature weights. They are the parameters of the model of the interaction potential and have to be determined by training. Kumar and Hebert (2006) give a geometric interpretation of the interaction potential: It partitions the space of the relational features $\boldsymbol{\mu}_{ij}(\mathbf{y})$ betweens the pairs that have the same class labels and pairs that have different labels. Thus, unlike with the well-known Ising model for MRF (Besag, 1986), it will moderate smoothing of neighbouring labels if there is a discontinuity of the features between the two sites.

We use $\boldsymbol{\psi}_i(\mathbf{y}) = \mathbf{f}_i(\mathbf{y})$, i.e. the features used for the interaction potential are identical to those used for the association potential. Furthermore, the component-wise absolute differences are used for the relational features $\boldsymbol{\mu}_{ij}$, i.e. $\boldsymbol{\mu}_{ij}(\mathbf{y}) = [1, |f_{i1}(\mathbf{y}) - f_{j1}(\mathbf{y})|, \ldots |f_{iR}(\mathbf{y}) - f_{jR}(\mathbf{y})|]^T$, where R is the dimension of the feature vectors $\mathbf{f}_i(\mathbf{y})$ and $f_{ik}(\mathbf{y})$ is the k^{th} component of $\mathbf{f}_i(\mathbf{y})$. The neighbourhood N_i of image site i consists of the four neighbouring image sites.

2.4 Parameter Learning and Classification

The parameters of the model for $P(\mathbf{x} \mid \mathbf{y})$ are the weights $\mathbf{w}$ and $\mathbf{v}$ of the association and interaction potentials, respectively. They can be combined to a parameter vector $\boldsymbol{\theta} = [\mathbf{w}^T, \mathbf{v}^T]^T$ that has to be estimated from training samples, i.e. a set $Y = \{\mathbf{y}_1, \ldots \mathbf{y}_M\}$ of M training images for which the class labels $X = \{\mathbf{x}_1, \ldots \mathbf{x}_M\}$ are known. If the parameters $\boldsymbol{\theta}$ are known, classification can be performed by maximising $P(\mathbf{x} \mid \mathbf{y})$ according to Equation 1. However, exact inference is computationally intractable for CRF (Kumar & Hebert, 2006). Vishwanathan et al. (2006) compare various methods for inference on CRF and come to the conclusion that Loopy-Belief-Propagation (LBP) (Frey & MacKay, 1998), which is a standard technique for performing probability propagation in graphs with cycles, provides the best results. It is thus used for classification in this work. In order to determine the parameters $\boldsymbol{\theta}$, $P(\mathbf{x} \mid \mathbf{y})$ is interpreted as $P(\mathbf{x} \mid \mathbf{y}, \boldsymbol{\theta})$, and $\boldsymbol{\theta}$ is estimated so that it maximises the conditional probability $P(\boldsymbol{\theta} \mid X, Y)$ or minimises the negative log-likelihood $L(\boldsymbol{\theta}) = -\log(P(\boldsymbol{\theta} \mid X, Y))$. An optimisation method that is frequently used is the BFGS Quasi-Newton method (Nocedal & Wright, 2006). If applied to minimise $L(\boldsymbol{\theta})$, it requires the computation of the gradients of $L(\boldsymbol{\theta})$, which in turn requires the selection of an approximate inference method (Vishwanathan et al., 2006). Following Vishwanathan et al. (2006), we use BFGS together with LBP for the simultaneous estimation of $\mathbf{w}$ and $\mathbf{v}$.

3. FEATURE EXTRACTION

In order to apply the CRF framework, the site-wise feature vectors $\mathbf{f}_i(\mathbf{y})$ that are used both for the association and the interaction potentials must be defined. It has to consist of appropriate features that can help to discriminate settlements from the background. In our application, we use two groups of features, namely gradient-based features $\mathbf{f}_{gi}(\mathbf{y})$ and colour-based features $\mathbf{f}_{ci}(\mathbf{y})$. Thus, the site-wise feature vector for site i contains both groups: $\mathbf{f}_i(\mathbf{y}) = [\boldsymbol{f}_{gi}(\mathbf{y})^T, \boldsymbol{f}_{ci}(\mathbf{y})^T]^T$. Both $\mathbf{f}_{gi}(\mathbf{y})$ and $\mathbf{f}_{ci}(\mathbf{y})$ contain features computed at two different scales λ_1 and λ_2. At scale λ_1, they are computed taking into account only the pixels inside the image site i (which is a square box of $s \times s$ pixels), whereas at scale λ_2 the pixels in a square of size $2 \cdot s$ centred at the centre of image site i are taken into account. Hence we do not only consider information derived at site i for the site-wise feature vectors $\mathbf{f}_i(\mathbf{y})$, but we also model dependencies between the image information of neighbouring sites. Of course, this principle could be expanded to a larger number of scales.

3.1 Features Based on Gradients

For determining the gradient-based features, we start by computing the gradient magnitude (Figure 1) and orientation for each pixel of the input image. All the gradient-based features are derived from a weighted histogram of the gradient orientations computed for each image site at both scales. Each histogram has 30 bins, so that each bin corresponds to an orientation interval of 6° width. Each bin contains the sum of the magnitudes of all gradients having an orientation that is within the interval corresponding to the bin. Summing over the magnitudes and not just counting the numbers of gradients falling into each bin is necessary to maintain the impact of strong magnitudes.

Three examples for histograms of different land cover types are shown in Figure 2. It shows that due to the heterogeneity of settlement areas, there are several strong peaks in this class, whereas cropland is nearly homogeneous and has a histogram showing low magnitudes. Thus, that mean MG and the variance VG of the histogram magnitudes are chosen as features to distinguish between textured and homogeneous areas. The third

example in Figure 2 shows a road passing through cropland. In such a situation, the histogram shows only one strong peak as opposed to the settlement, where a larger diversity of orientations and thus a larger number of peaks can be observed. Thus, the number of bins *NG* with values above the mean was selected as the third gradient-based feature. All the features are normalised so that the values are in the interval [0, 1]. The gradient based feature vector $\mathbf{f}_{gi}(\mathbf{y})$ of image site *i* consists of six elements (three for each scale): $\mathbf{f}_{gi}(\mathbf{y}) = [MG_i^{(1)}, VG_i^{(1)}, NG_i^{(1)}, MG_i^{(2)}, VG_i^{(2)}, NG_i^{(2)}]^{\mathrm{T}}$, where the upper index indicates the scale. We also tried to use the main orientation of the image site and the angle between the two largest peaks of the histogram as additional features. Neither modification resulted in any significant improvement of the classification performance.

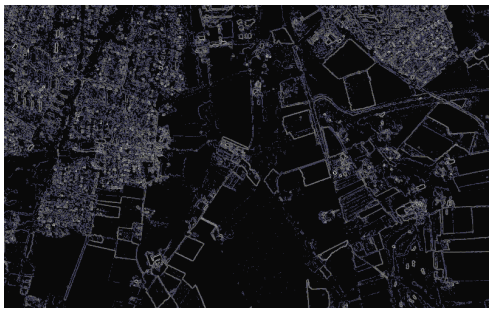

Figure 1. Gradient magnitude image of the test area.

3.2 Features Based on Colour

Figure 2 shows that in settlement areas we can expect a large variation of colours, whereas other land cover classes show a more homogeneous appearance. We carry out an IHS transformation and then proceed by analysing the hue image (Figure 3). For each image site *i* we compute the variance of the hue *VH* at both scales and normalise it so that its values are in the interval [0, 1]. The colour based feature vector of image site *i* has two components, namely *VH* for both scales: $\mathbf{f}_{ci}(\mathbf{y}) = [VH_i^{(1)}, VH_i^{(2)}]^{\mathrm{T}}$. We also tried to use the mean hue as an additional feature, but it did not improve our results. We also tried to use other bands or combinations of bands, but using the hue band showed better performance than any other single band, and the consideration of other bands did not improve the results significantly while increasing the computational costs.

3.3 Feature Space Mapping

The site-wise feature vectors $\mathbf{f}_i(\mathbf{y})$ have a dimension of 8. As in (Kumar & Hebert, 2006), the transformed feature vectors $\mathbf{h}_i(\mathbf{y})$ are obtained by a quadratic expansion of the feature vectors $\mathbf{f}_i(\mathbf{y})$ so that the functions $\Phi_k(\mathbf{f}_i(\mathbf{y}))$ include all the $l = 8$ components of $\mathbf{f}_i(\mathbf{y})$, their squares and all their pairwise products. The dimension of the transformed feature vectors $\mathbf{h}_i(\mathbf{y})$ is $l + 1 + l \cdot (l + 1) / 2 = 45$. In case of the interaction potential, no feature space mapping is used. The dimension of the relational feature vectors $\boldsymbol{\mu}_{ij}(\mathbf{y})$ is 9. Using a feature space mapping for these relational feature vectors degraded the results in our tests, maybe because the feature space becomes too high-dimensional.

4. EXPERIMENTS

For our experiments we used the RGB bands of a multi-spectral Ikonos scene of a rural region near Herne, Germany. The resolution is 4 m. Two test areas having a similar type of land cover were cut out of the scene, each covering an area of 3.2 x 2.0 km^2. Ground truth was obtained by manually labelling these test areas on a pixel-level. In order for an area to be labelled as a settlement, it had to contain at least four houses; smaller groups of houses were ignored. One of the test areas and the related ground truth were used for training, whereas the other one served as our test scene. For the test scene, the ground truth could be used to evaluate the results.

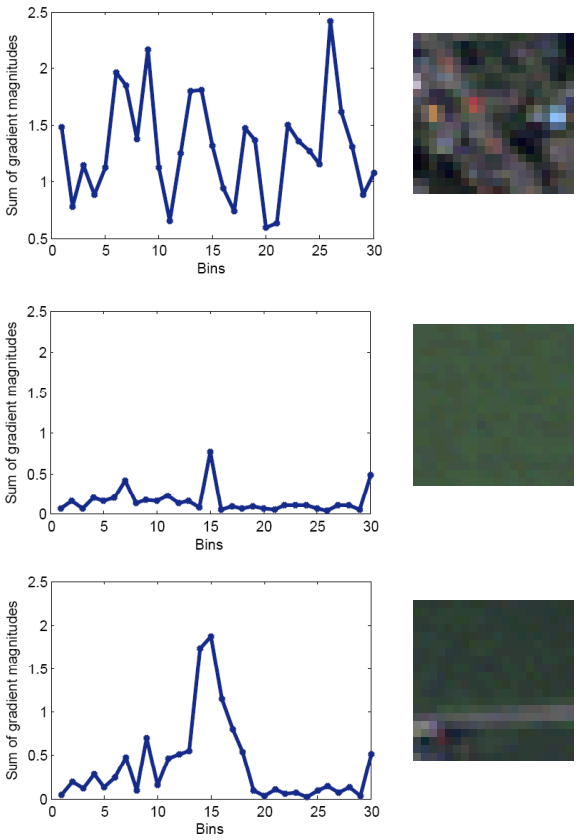

Figure 2. Gradient orientation histograms and the image patches they were computed from (s = 20 pixels). Upper row: settlement; centre: cropland; last row: cropland intersected by a road.

Figure 3. Hue image of the test area.

After having defined the size *s* of an image site, the features and the class labels were determined for all the image sites of the training area. An image site was labelled as belonging to class settlement if more than 50% of its pixels belonged to the settlement class. The features and the class labels for the image sites of the training area were used to determine the parameters of the CRF. After that, the test scene was also subdivided into image sites of size *s*, the features were extracted for all image sites, and the parameters learned from the training data were

used to determine the class of each image site by maximising $P(\mathbf{x} \mid \mathbf{y})$ using LBP. A reference classification was determined from the ground truth in the same way as the class labels for training were generated, i.e. by majority voting of the pixels in each image site. After that, completeness, correctness and quality (Heipke et al., 1997) were computed based on a comparison of the class labels of the image sites.
This procedure was applied using three different block sizes s, namely s = 4, 10, and 20 pixels, which resulted in 25000, 4000, and 1000 image sites, respectively. By using different block sizes, it should be possible to assess the influence of this parameter on the results. Furthermore, we carried out a standard Maximum-Likelihood (ML)-classification using $s = 4$ and $s = 10$ pixels and the same features as for the CRF, but only for the scale λ_1; using also λ_2 deteriorated the ML results. In the ML classification we used a normal distribution for the likelihood model $P(\mathbf{f}_i(\mathbf{y}_i) \mid x_i)$, determining the mean and covariance function from the training data. A comparison of the ML classification results and the results achieved by using CRF should highlight the influence of the statistical model of context.

The completeness, correctness, and quality achieved for the test scene in our experiments are shown in Table 1. The CRF-based method achieves completeness and correctness values of 90% and better in all cases except for s = 4, where completeness is slightly smaller. In comparison, the ML method also achieves 90% completeness, but correctness is very low (76%) for $s = 10$ pixels. For $s = 4$ pixels, the results are even worse. Using the CRF framework with its statistical model of context in the classification process significantly increases the quality of the results.

Method	s [pixel]	Completeness	Correctness	Quality
ML	4	77.6%	68.2%	57.0%
ML	10	90.7%	75.8%	70.3%
CRF	4	89.6%	90.3%	81.7%
CRF	10	92.9%	90.0%	84.2%
CRF	20	94.4%	91.6%	86.9%

Table 1. Evaluation of the classification results achieved for ML and for CRF using different block sizes s.

Figure 4 shows the ground truth and the results achieved both for CRF and ML classification for $s = 10$ pixels. The CRF results achieved for $s = 20$ pixels are shown in Figure 5. Examining these figures, it is obvious that the CRF approach tends to result in compact shapes. It works very well on the larger settlement areas. However, the smoothing effects of the context model cause small settlement areas to be missed. Small patches of non-settlement areas surrounded by settlement are also misclassified. These over-smoothing effects indicate that the impact of the interaction potential might be too strong. On the other hand, comparing the results of the CRF and ML classification results in Figure 4, the benefits of considering context become obvious. The ML results are much noisier. Large structures in settlements are not correctly detected, and there are many small false positives related to groups of trees. For the CRF method, there is a minor effect of the block size on the quality of the results: using $s = 20$ pixels, the completeness is 5% larger than for $s = 4$ pixels, because the features can be extracted more reliably if the block size is larger. However, a larger block size will reduce the level of detail of the results. Our experiments indicate that a value between s = 4 and s = 10 pixels might be optimal. Figure 6 shows a part of the test area for the CRF and ML classification using $s = 4$ pixels.

Figure 4. Test scene for $s = 10$ pixels. Class *settlement* is superimposed to the image in red. First row: ground truth; second row: CRF; third row: ML.

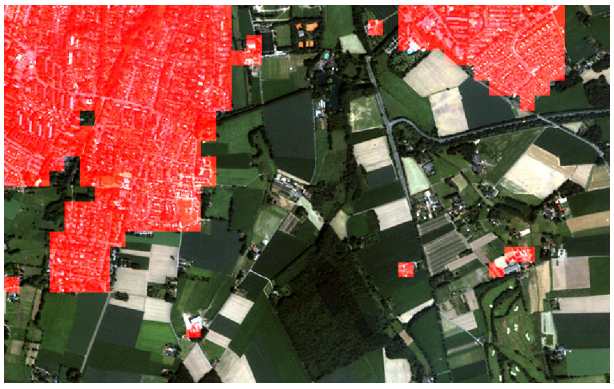

Figure 5. Results of CRF classification using $s = 20$ pixels. Class *settlement* is superimposed to the image in red.

Despite the in general somewhat poorer results of the CRF approach compared to larger block size, the shape of the settlement is well-preserved, whereas a reliable classification can not be achieved using the ML approach.

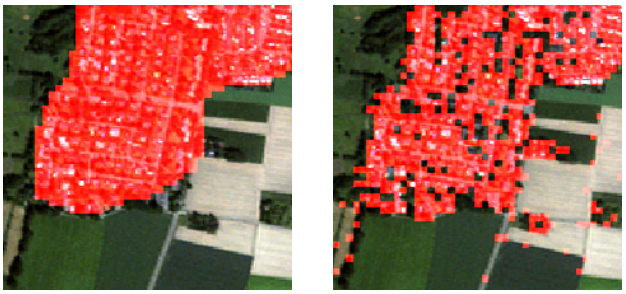

Figure 6. Section of the results of the Maximum-Likelihood-classification and the CRF-classification for $s = 4$.

Our results are quite promising, even more so because they were achieved using only a small set of features and a relatively simple model for the interaction potential. Using better features or a better context model could still improve the results.

5. CONCLUSION AND OUTLOOK

We have presented a new CRF-based approach for the classification of settlements in high resolution optical satellite imagery. CRFs allow incorporating contextual information into the classification process. The focus of this paper was on the impact of the context information on the classification results and not on a sophisticated selection of features. Tests on a multispectral Ikonos scene of 4 m resolution containing settlement areas of different size have shown that our CRF-based approach can achieve completeness and correctness values of over 90% for settlement areas and that it clearly outperforms ML classification based on the same set of features. Further research will focus on the extension of the framework to a classification of an arbitrary number of classes. The necessity of this already becomes obvious when trying to classify Ikonos panchromatic data of 1 m resolution with our approach. Settlements and forests are much harder to distinguish, which leads to unsatisfactory results. The situation could be improved by considering at least one more class, namely *forest*. Moreover the CRF framework should be applied to the results of a preliminary segmentation in order to obtain a more precise determination of the class boundaries. In this way, the problem of several classes existing in one site could also be reduced. Another goal for the future is an extension of the CRF framework to make it applicable to multi-temporal interpretation by considering spatial as well as temporal context, e.g. by introducing an additional temporal interaction potential.

ACKNOWLEDGEMENTS

The implementation of our method uses the CRF Toolbox for Matlab by K. Murphy & M. Schmidt (Vishwanathan et al., 2006): http://www.cs.ubc.ca/~murphyk/Software/CRF/crf.html

REFERENCES

Besag, J. 1986. On the statistical analysis of dirty pictures. J. Royal Statistical Soc. Series B (Methodological) 48(3):259-302.

Bishop, C. M., 2006. Pattern recognition and machine learning. 1st edition, Springer New York, 738 pages.

Cheriyadat, A., Bright, E., Potere, D., Bhaduri, B., 2007. Mapping of settlements in high resolution satellite imagery using high performance computing. GeoJournal 69(1/2):119-129.

Frey, B. J. and MacKay, D. J., 1998. A revolution: belief propagation in graphs with cycles. In: Advances in Neural Information Processing Systems, 10, MIT Press, pp. 479-485.

Gamba, P., Dell'Acqua, F., 2003. Increased accuracy multiband urban classification using a neuro-fuzzy classifier. Int. J. Remote Sensing 24(4):827-834.

Gamba, P., Dell'Acqua, F., Lisini, G., Trianni, G., 2007. Improved VHR urban area mapping exploiting object boundaries. IEEE-TGARS 45(8):2676-2682.

He, W., Jäger, M., Reigber, A., Hellwich, O., 2008. Building extraction from polarimetric SAR data using mean shift and conditional random fields. In: Proc. of European Conference on Synthetic Aperture Radar, 3, pp. 439-442.

Heipke, C., Mayer, H., Wiedemann, C., Jamet, O., 1997. Evaluation of automatic road extraction. In: *IntArchPhRS* XXXII (3-4W2), pp. 151–160.

Herold, M., Gardner, M. E., Robert, D. A., 2003. Spectral resolution requirements for mapping urban areas. IEEE TGARS 41(9):1907-1919.

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

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

Nocedal, J. and Wright, S. J., 2006. Numerical Optimization. 2nd edition, Springer New York, 664 pages.

Paget, R. and Longstaff, I. D., 1998. Texture synthesis via a noncausal nonparametric multiscale Markov Random Field. IEEE Transactions on Image Processing 7(6):925-931.

Shackelford, A. K. and Davis, C. H., 2003. A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas. IEEE TGARS 41(9):1920-1932.

Smits, P. C. and Annoni, A., 1999. Updating land-cover maps by using texture information from very high-resolution space-borne imagery. IEEE TGARS 37(3):1244-1254.

Tupin, F. and Roux, M., 2005. Markov Random Field on region adjacency graph for the fusion of SAR and optical data in radargrammetric applications. IEEE TGARS 43(8):1920-1928.

Vishwanathan, S., Schraudolph, N. N., Schmidt, M. W., Murphy, K. P., 2006. Accelerated training of conditional random fields with stochastic gradient methods. In: 23rd International Conference on Machine Learning, pp. 969-976.

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