PFG 2010 / 6, 439–449
Stuttgart, Dezember 2010

Article

A General Framework for Fast and Interactive Classification of Optical VHR Satellite Imagery Using Hierarchical and Planar Markov Random Fields

JENS KERSTEN, MONIKA GÄHLER & STEFAN VOIGT, Oberpfaffenhofen

Keywords: Markov random fields, optical remote sensing, region labeling, crisis mapping

Summary: In this paper a general framework for fast and interactive classification of very high resolution satellite imagery for emergency and crisis mapping applications as well as several other applications is proposed. Multiscale image information as well as hierarchical and spatial context information is incorporated into the classification process using a hybrid Markov model, which combines a hierarchical directed as well as a planar undirected Markov random field (MRF). Classification is carried out using the standard non-iterative maximum a posteriori (MAP) and marginal posterior mode (MPM) inference. Additionally, a modified MAP computation, which is able to outperform the original methods under certain conditions, is proposed. Here uncertain image information, for example from class transition areas, is not incorporated during the inference procedure. The effectiveness of both the framework and the modified MAP inference is demonstrated by two examples.

Zusammenfassung: *Ein allgemeines Rahmenwerk für eine schnelle und interaktive Klassifikation hochauflösender optischer Satellitenbilder mittels hierarchischen und planeren Markov-Zufallsfeldern*. In diesem Artikel wird ein Rahmenwerk von Methoden für eine rasche und interaktive Klassifikation hochauflösender optischer Satellitenbilder im Rahmen von Notfall- und Krisenkartierungen sowie einer Vielzahl anderer Anwendungen vorgestellt. Mittels eines hybriden Markov-Modells, welches ein hierarchisches, gerichtetes und ein ungerichtetes, planares Markov-Zufallsfeld (MRF) kombiniert, werden Bildinformationen auf mehreren Skalen sowie hierarchische und räumliche Kontextinformationen in den Klassifikationsprozess einbezogen. Die Klassifikation erfolgt mittels der bekannten Maximum a Posteriori (MAP) sowie der Marginal Posterior Mode (MPM) Inferenz. Des Weiteren wird ein modifizierter MAP Ansatz vorgestellt, welcher unter bestimmten Voraussetzungen bessere Ergebnisse als die ursprünglichen Methoden liefern kann. Dabei wird „unsichere" Bildinformation, wie zum Beispiel aus Bereichen, welche Mischklassen aufweisen, nicht in den Inferenzprozess einbezogen. Die Effektivität des Rahmenwerks sowie der modifizierten MAP Berechnung wird an zwei Beispielen demonstriert.

1 Introduction

This paper focuses on the challenge of interactive classification of very high resolution optical satellite imagery for emergency and crisis mapping applications. Due to near real time processing requirements, classification is here restated as a task of semantic annotation of square image regions. Furthermore, the transferability of the approach to various crisis scenarios and therefore to several other applications is desired. Representatives for recurrent tasks during crisis scenarios are given with the classification of water surfaces (flood events), urban (earthquakes) and burned areas (fires). The fact, that nearly every crisis situation is unique, often hinders an application of automatic image analysis methods and demands an application of manual processing steps in terms of visual interpretation. This

 www.schweizerbart.de
DOI: 10.1127/1432-8364/2010/0066 1432-8364/10/0066 $ 2.75

work aims on the combination of fast image analysis methods and the inherent image understanding of an image analyst, in order to minimize visual interpretation steps and to derive robust and reproducible results.

Contextual information can improve classification accuracy significantly, if such information can be well modeled (Khedam & Belhadj-Aissa 2003). Bayesian models form a natural framework for integrating both statistical models of image behaviour and prior knowledge about the contextual structure of semantic classes. The contextual structure is often modeled as a Markov random field (MRF). In order to capture the intrinsic hierarchical nature of remote sensing data, several efficient Markov image modeling approaches defined on tree structures were proposed during the last two decades (Bouman & Shapiro 1994, Fieguth et al. 1998, Lafferté et al. 2000, Wilson & Li 2003, Choi et al. 2008). In order to cope with the surrounding conditions accompanied by crisis mapping scenarios, a hierarchical Markov model is proposed in this article. The motivation for using such a model is to provide a general interactive and computational effective framework for classification or pre-classification of multispectral satellite imagery.

Compared to earlier approaches the proposed framework uses hierarchical semantic networks in order to capture the hierarchical nature of remote sensing data. Therefore, the definition and modeling of arbitrary thematic classes in arbitrary scales as well as the relationship between the classes in adjacent scales is provided. The general formulation of the framework aims on the transferability of this approach to several different image contents, thematic problems as well as products of different sensors. Due to the initial lack of training data the model parameters are learned in a sequential manner.

The article is organized as follows: In Section 2, the proposed general framework is introduced. In Section 3, the utilized hierarchical and planar Markov models as well as the parameter estimation and inference are presented. The relevance and efficiency of the proposed framework is demonstrated in Section 4 and discussed in Section 5. Conclusions are drawn in Section 6.

2 Description of the Framework

Modeling image characteristics in a hierarchical manner has shown to be valuable for many applications, e. g., image labeling and object detection (Awasthi et al. 2007), multiband segmentation of astronomical images (Collet & Murtagh 2004) as well as the unsupervised detection of flood-induced changes in SAR data (Martinis et al. 2010). As pointed out in (Pérez et al. 2000), MRFs defined on causal tree structures always enable computationally efficient and exact inference of the unknown class labels, which is quite appealing for an application in the field of rapid mapping. Motivated by this, the proposed model is defined on a causal quadtree.

Image classification should be possible even when no additional information like vector data or a digital elevation model (DEM) is available. Hence, only the image itself as well as the image analyst is required for the application of the framework. The complete workflow is illustrated in Fig. 1 and can be described as follows: 1) A quadtree image representation is instantiated at first. The size of the smallest region can be defined individually depending on the spatial resolution of the im-

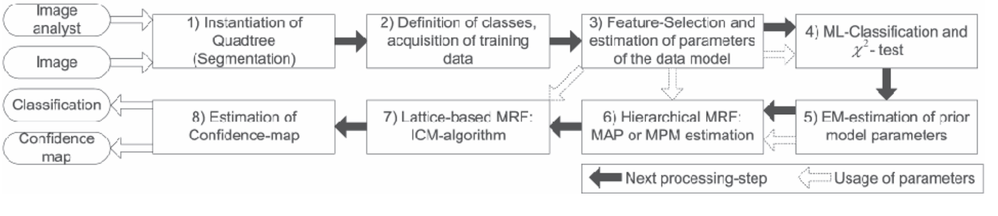

Fig. 1: Workflow for the proposed framework.

age as well as the structure of the thematic classes. 2) The framework allows an interactive definition of arbitrary semantic classes in different scales, i. e., the modeling of a hierarchical semantic network. For each class, the image analyst has to provide training data. 3) Based on the training data the parameters of Gaussian mixture models (data model) are estimated. Relevant features are identified through feature selection. 4) A constrained maximum likelihood classification is carried out in order to obtain training data (labeled regions) for the estimation of the prior model parameters via expectation maximization (step 5)). 6) Non-iterative hierarchical MAP or MPM inference is carried out using the parameters estimated in step 3 and 5. 7) A subsequent optimization step by incorporating spatial context concerning the finest quadtree level using a planar undirected MRF is carried out. 8) In order to obtain information concerning the confidence of the labeling process, a confidence map is computed.

Since some processing steps are optional, the classification can be carried out using several different combinations of methods. The steps 1, 2 and 3 are always required. Based on these computations the following combinations of processing steps are possible: (4), (4, 7), (4, 5, 6, (8)) and (4, 5, 6, 7, (8)), where step 8 is always optional. Interaction with the image analyst is required in the following steps: 1 (definition of the size of the smallest image regions), 2 (definition of classes and acquisition of training data), 3 (define the number of features and check estimation result), 4 (definition of the probability of error for the chi²-test) and 7 (definition of the weight of the context term).

3 Markov Image Modeling

Compared to other approaches, this framework allows to model arbitrary semantic classes in different scales as well as inter- and intra-scale context dependencies between the classes. In this section a complete overview of the methodology, namely hierarchical and lattice-based MRFs is given.

3.1 *Hierarchical Markov Image Models Defined on the Quadtree*

Problem Definition and Statistical Modeling

Given a set of variables (x,y), where x represents the unknown class labels and y the observed image data, the estimation of the "best" realization of x given y is desired. In a statistical process (x,y) represent occurrences of the vectors of random variables (X,Y). The so called Markovian independencies of (X,Y) are represented by an independence graph, which here is defined by a quadtree. As illustrated in Fig. 2, the components of X are indexed by the nodes of the quadtree, and additionally each (or a subset) of these nodes is associated with a component of the observation vector Y.

The set of all nodes of the quadtree is denoted S, and the set of nodes at a single level l is denoted S^l, with $l = 0,\dots,N$. The root node of the tree is denoted $S^0=r$. Each node $s \in S^{l\,=\,1,\dots,N}$ has a unique parent node s^- and each node $s \in S^{l\,=\,0,\dots,N-1}$ is equipped with a set of four child nodes t. Assuming both a first order top-down Markov chain structure, where each node in S^l depends only on its ancestor in S^{l-1}, as well as a standard site-wise factorization for the observation model $P(y|x)$, the joint distribution $P(x,y)$ is given as the factorization of local functions

$$P(x,y) = P(x_r)\prod_{s \neq r} P(x_s \mid x_{s^-})\prod_{s \in S} P(y_s \mid x_s) \quad (1)$$

and is entirely defined by the root prior $P(x_r)$, the parent-child transition probabilities $\{P(x_s|x_{s^-})\}_{s\neq r}$ and the data conditional likelihoods $\{P(y_s|x_s)\}_{s\in S}$.

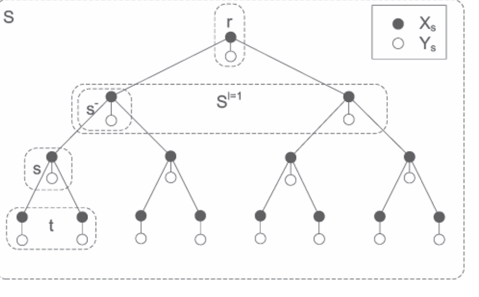

Fig. 2: Independence graph (the quadtree is illustrated as a dyadic tree).

Data Model

The proposed framework allows the definition of arbitrary semantic classes in arbitrary scales (i. e., quadtree levels). The classes are represented by Gaussian mixture models, where each involved level is treated independently. Various features like first and second order statistics, ratios like the well known NDVI, several colour space representations as well as texture and texture related features are available. An image element from a satellite image with four spectral channels can therefore be described by 70 features. In order to limit the computational efforts and to identify redundant and irrelevant features, the suboptimal feature selection method Sequential Forward Floating Search (SFFS) (Pudil et al. 1994) is utilized. The method starts with an empty set of features. Each iteration the most significant feature is added followed by a backtracking phase, where the optional exclusion of features is carried out for as long better variable subsets of the corresponding sizes are identified. The criterion for the evaluation of a distinct feature subset during the feature selection procedure is here defined as the mean Mahalanobis distance between the mixture components. Compared to the well known Bhattacharyya and Kullback-Leibler distance, we observed more reasonable and consistent results using the Mahanalobis distance.

Due to the fact that the true number of components of the desired mixture model is unknown, a split-based expectation maximization (EM) algorithm is applied. The decision, weather to split a component or not, depends on a multivariate normality criterion based on the Mahalanobis distance of a sample measurement vector from a certain Gaussian component center (Ververidis & Kotropoulos 2008). Using this criterion merging of components becomes obsolete.

Prior Model

For the prior model a Potts-like distribution is used in (Bouman & Shapiro 1994). This model favours identity between a node s and its parent node s^-. Since the proposed framework supports the definition of arbitrary semantic classes in each level, the definition of the involved transition probabilities is a more difficult task. Hence, an EM algorithm as described in (Feng et al. 2002) is utilized here to estimate the model parameters. A minimization of efforts concerning the acquisition of training data for the EM learning is achieved by a preliminary maximum likelihood (ML) classification based on the class feature densities of the data model. In order to avoid the incorporation of misclassifications, a chi² test of the form

$$\chi^2_{f,\alpha} \geq (y_s - \mu_x)^T \Sigma_x^{-1} (y_s - \mu_x) \qquad (2)$$

is applied, where f is the degree of freedom, α is the probability of error, y_s is the feature vector of node s and μ_x,Σ_x are the parameters of the Gaussian distribution of class x. Thus, the class label information of an image element s is involved in EM parameter estimation, if the right side of inequation (2) is smaller or equal than the corresponding quantile.

Inference

Since a distinct top-down pass through the quadtree hierarchy is a Markov chain in scale, a Viterbi-like algorithm (Forney 1973) can be applied for the exact and non-iterative inference of the class labels. The MAP estimate minimizes the probability that any image region will be misclassified. Wrong classifications are here penalized without any consideration about how many misclassifications occurred in total. A cost function which incorporates this aspect and is generally better behaved (Laferté et al. 2000) yields the MPM estimator. For further details concerning the derivation of the algorithms, the reader is referred to (Pérez et al. 2000) and (Laferté et al. 2000).

The data conditional likelihoods are modeled using multivariate Gaussian mixture models. Hence, missclassifications of image regions which e. g., represent class transition areas may be likely even when context information is included. Therefore, similar to a maximum likelihood classification, a chi² test (2) is proposed in order to detect image regions which exhibit low conditional likelihoods for all classes during the inference procedure. If the null hypothesis of this test is refused, only the prior model is incorporated for the inference of the class label of the appropriate image region. The chi[2] test is carried out in a preliminary computation step, which leads to the following MAP estimation:

1. Preliminary step. $\forall\ s{\in}S$, *iff* the null hypothesis of test (2) holds exclusively for one thematic class: classify s based on the data model. Else: $x_s = -1$.

2. Bottom-up pass.
Initialization (leaves $s{\in}S^N$) with $P(y_s|x_s) = 1$, *if* $xs = -1$:

$$P_s(x_{s-}) = \max_{x_s} P(y_s \mid x_s) P(x_s \mid x_{s-}) \tag{3}$$

$$x_s^*(x_{s-}) = \arg\max_{x_s} P(y_s \mid x_s) P(x_s \mid x_{s-}) \tag{4}$$

Recursion ($s{\in}S^{N-1}\dots S^1$) with $P(y_s|x_s) = 1$, *if* $xs = -1$:

$$P_s(x_{s-}) = \max_{x_s} P(y_s \mid x_s) P(x_s \mid x_{s-}) \prod_{t\in s^+} P_t(x_s) \tag{5}$$

$$x_s^*(x_{s-}) = \arg\max_{x_s} P(y_s \mid x_s) P(x_s \mid x_{s-}) \prod_{t\in s^+} P_t(x_s) \tag{6}$$

3. Top-down pass.
Initialization (root r) with $P(y_s|x_s) = 1$, *if* $xs = -1$:

$$\hat{x}_r = \arg\max_{x_r} P(y_r \mid x_r) P(x_r) \prod_{t\in r^+} P_s(x_r) \tag{7}$$

Recursion ($s \in S^1 \dots S^N$):

$$\hat{x}_s = x_s^*(\hat{x}_{s-}) \tag{8}$$

Until now, this modified inference is realized for the MAP criterion.

Marginal Posterior Entropy

Since a classification result may serve as a basis of decisions for civil and humanitarian relief organisations, a quantification of the uncertainty of the labeling process is highly recommended. As stated in (Pérez et al. 2000), the knowledge of the marginal posteriors allows to assess for each image region s the degree of confidence that can be associated to the estimated value. Therefore, the entropy H_s of the marginal posteriors can be computed as follows:

$$H_S(x_s \mid y) = -\sum_{k=1}^{K} P(x_s = k \mid y) \log_2 P(x_s = k \mid y), \tag{9}$$

where $k = 1,\dots K$ is the class index. Based on equation (9) a confidence map can be computed for the whole image. Image regions, which exhibit significant marginal posterior entropy are good indicators of misclassified image regions (Feng et al. 2002).

3.2 *Lattice-based Markov Image Modeling*

The above described hierarchical approach allows the modeling of relationships between semantic classes in adjacent scales but does not incorporate spatial context (i. e., intra-scale context). Furthermore, it is often stated that the prior quadtree structure induces non-stationarity in space (Laferté et al. 2000) due to the fact that spatially adjacent image regions may not have common neighbors at coarser scales. This leads to "blocky" artefacts in the estimation result. In order to remove these artefacts and to further improve the classification result, a subsequent incorporation of spatial context by using an undirected lattice-based MRF is proposed here.

Each image region $s{\in}S^N$ (i. e., all leave nodes of the quadtree) is connected with its four immediately adjacent regions leading to a first order neighborhood system. This limitation is reasonable under the consideration that the spatial dependencies between image regions rapidly decrease when the distance between the regions increases.

Gibbs random fields (GRF) provide global image models by specifying a probability mass function of the following form:

$$P(x) = Z^{-1} \exp(-U(x)), \text{ with } U(x) = \sum_{c\in C} V_c(x) \tag{10}$$

where C is the set of all cliques, Z is a normalization constant (i. e., partition function), *V* denotes the neighborhood system and *U(x)* is the Gibbs energy function. This function is defined as the sum of clique potentials $V_c(x)$ over all possible pairwise cliques. The application of the Hammersley-Clifford theorem (Besag 1974) allows the expression of the joint distribution over the labels *X* and the observations *Y* using Gibbs energy functions. The maximization of the posterior probability (MAP) is hereinafter equivalent to the minimization of the following energy function (Dubes & Jain 1989):

$$U(X,Y) = \sum_{1\le i\le I}\sum_{1\le j\le J}\begin{bmatrix} U_{data}\left(y_{(i,j)}\middle|x_{(i,j)}\right) + \\ U_{context}\left(x_{(i,j)}\middle|\left\{x_{(g,h)},(g,h)\in N_{(i,j)}\right\}\right)\end{bmatrix}, \quad (11)$$

where *(ij)* describes the position of the image element and $N_{(i,j)}$ denotes the neighborhood of element $s_{(i,j)}$. The energy term $U_{data}(\cdot)$ represents the class statistics and corresponds to the negative data conditional likelihoods of the data model (section 3.1). The energy term $U_{context}(\cdot)$ describes the interregional class dependence and is defined as follows:

$$U_{context}\left(x_{(i,j)}\middle|\left\{x_{(g,h)},(g,h)\in N_{(i,j)}\right\}\right) = \sum_{(g,h)\in N_{(i,j)}} \beta\cdot\delta_k\left(x_{(i,j)},x_{(g,h)}\right), \quad (12)$$

where β is a weighting factor and δ_k is the Kronecker delta function:

$$\delta_k\left(x_{(i,j)},x_{(g,h)}\right) = \begin{cases} -1, & if\ x_{(i,j)} = x_{(g,h)} \\ 0, & if\ x_{(i,j)} \ne x_{(g,h)} \end{cases} \quad (13)$$

The well known ICM algorithm (Besag 1986) is used to carry out the minimization of the energy function (11).

4 Examples and Results

In order to demonstrate the practicability and the relevance of the framework, two examples are examined. In the first experiment the classification of a scene representing a rural region near the city of Dresden, Germany is carried out. Because of the non-complex image content, several characteristics of the framework can be pointed out. A more realistic experiment is given with the second example (northeast Namibia). Here the image content is characterized by inhomogeneous areas and tiny structures like streets and houses.

4.1 *Example 1: Dresden, Germany*

The first example is given with a subset (512 × 512 pixels) from a multispectral IKONOS scene, which represents a rural region near Dresden, Germany. The image has the following product characteristics: pansharpened multispectral (four channels: blue, green, red, near-infrared), standard geometrically corrected, ground sampling distance (GSD) of PAN: 0.87 m cross scan and 0.92 m along scan, ground resolution: 1 m (cubic convolution), acquired on August 6th, 2007. The smallest image region size is defined by 2 × 2 pixels, which leads to a quadtree with nine levels.

The annotation of all image regions in the finest quadtree level to one of the following classes is desired: *dark field*, *vegetation*, *alley*, *street or green field*. Data models are trained in the three finest levels of the quadtree, where class *alley* is not modeled in the coarsest level, since the size of the square image regions is too large (8 × 8 pixels) to represent this class properly. Due to a reference provided by visual interpretation, the impact of context incorporation can be evaluated.

For each scale the feature selection algorithm identified a relevant feature subset, which basically consists of features from different colour space representations, e. g., the well known HLS or CIELUV colour space. Due to the simple image content, all derived classification results are very similar (Tab. 1). However, the overall accuracy increases with every processing step, which confirms the visual comparison of the results.

As expected, the non contextual ML classification yields the worst results. Confusions mainly occur between class *vegetation* and *green field* (Fig. 3.3). An application of the lattice-based MRF (β = 100.0) on the ML result (ML-MRF) causes the well known smoothing effect but cannot cope adequately with image areas, where many misclassifications occur in the ML result (Fig. 3.6). The MAP (Fig. 3.5) and the MPM (Fig. 3.4) estimator provide similar results, where the latter is slightly better. The lattice-based ML-MRF

Tab. 1: Overall Accuracys of different processing steps.

Domain	Regular 2d-grid (MRF)		Quadtree		Quadtree and 2d-grid	
Estimator	ML (3)	ML-MRF (6)	MAP (5)	MPM (4)	MPM-MRF (7)	Mod. MAP-MRF (8)
Overall accuracy	93.4 %	95.5 %	94.7 %	94.8 %	96.0 %	96.7 %

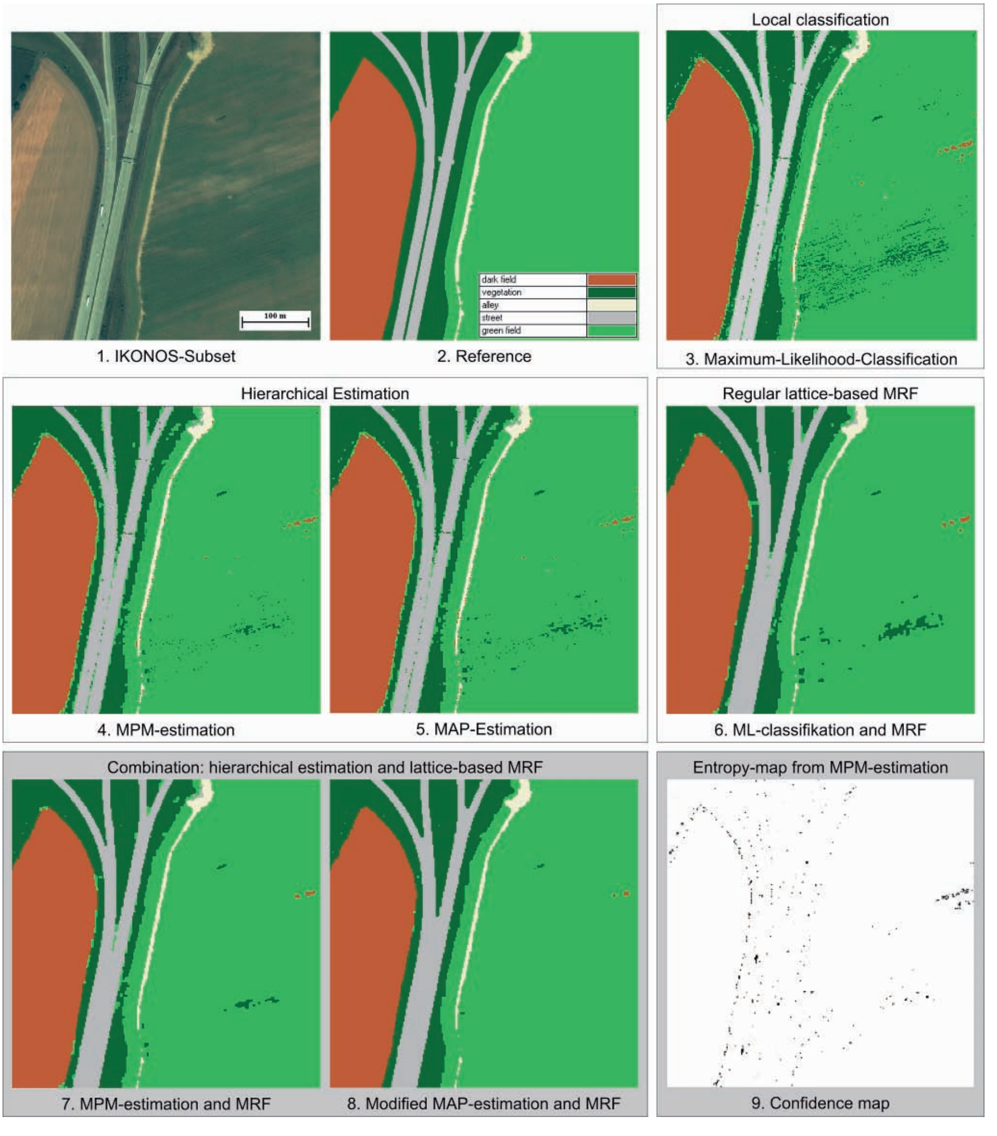

Fig. 3: Example 1: Intermediary results (white boxes) and final results (grey boxes). The local ML-classification (3) as well as (6) were derieved for comparison.

approach provides slightly better classification results than MAP or MPM estimation (see Tab. 1) but takes significantly more computational time than the non-iterative hierarchical approaches. A better classification result is carried out with the combination of the hierarchical MPM estimation and the subsequently applied grid-based MRF (MPM-MRF). The overall accuracy for the modified MAP estimation with $\alpha = 0.3$ is 95.7 %. However, this approach has shown to enforce the "blockiness" of the result. A subsequent MRF smoothing leads to a quite satisfying result with an overall accuracy of 96.7 % (Fig. 3.8). The confidence map shows the image regions, which may be misclassified using the MPM estimator (Fig. 3.9): The darker the regions the higher the uncertainty of the classification. This corresponds to the obviously visible misclassifications in the MPM result.

4.2 *Example 2: Caprivi Region, Northeast Namibia*

The second dataset is a multispectral QUICKBIRD scene from the Caprivi region, located in the northeast of Namibia, acquired on October 10th, 2009. The image has the following product characteristics: multispectral (four channels: blue, green, red, near-infrared), product level: standard imagery, ground resolution: 2.4 m. The size of the chosen subset is 1375 × 1135 pixels. In order to capture tiny structures like houses and streets, the smallest region-size is defined to be 2 × 2 pixels, which results in a quadtree with 11 levels. The classification of *water*, *streets*, *houses* and a class which represents the *rest* of the image is desired. Therefore, data models trained in two levels, namely the finest level S^{10} as well as level S^{8}, have shown to give good results. In the finest level all introduced classes are modeled. Since houses and streets are too small to represent them on the coarser level, only the class *water* and *rest* is modeled here. Due to similarities of the classes in the spectral domain, a local ML classification of the finest quadtree-level leads to confusions between the classes *water* and *street* as well as between *rest* and *house* (Fig. 4).

The application of the lattice-based MRF on the result of the ML classification as well as on all other intermediary results has shown to degrade the quality of the classification. A low weighting of the context term (e. g., $\beta = 2.0$) leads to a slightly smoothing, but mainly preserves misclassified areas. On the other hand, a relatively high weighting (e. g., $\beta = 15.0$) still preserves missclassified areas and partially leads to the omission of small structures like houses and streets. This effect is also caused by the application of the modified MAP estimation. The hierarchical MPM estimation yields the best result (Fig. 4.4) and is equivalent to the MAP estimation. The confidence map (Fig. 4.3) shows that especially image regions which represent streets as well as the land-water boundary are difficult to classify using the hierarchical MPM estimation.

5 Discussion

The experimental results show that the proposed framework allows an interactive classification of different types of high resolution optical satellite images by incorporating spatial and hierarchical context as well as image data from different levels of aggregation. Depending on the image content and the thematic focus, the desired result is obtained using different combinations of the methods utilized in this framework.

In the first experiment the image is characterized by large homogeneous areas. Here, the complete process chain (Fig. 1) is traversed and the best result is obtained using the modified MAP estimation and a subsequent lattice-based MRF smoothing. Using the original MAP or MPM estimation, the influence of the data model, which obviously cannot handle all image regions adequately, "votes down" the context information in some cases (e. g., the margin of the *dark field* is misclassified as *field*). The modified MAP computation has shown to be very useful to avoid misclassifications especially of mixed image regions in this example but enforces the "blockiness" of the classification result. The additional incorporation of spatial context counteracts this effect. In order to save computational time, a single iteration step of the ICM algorithm could be a

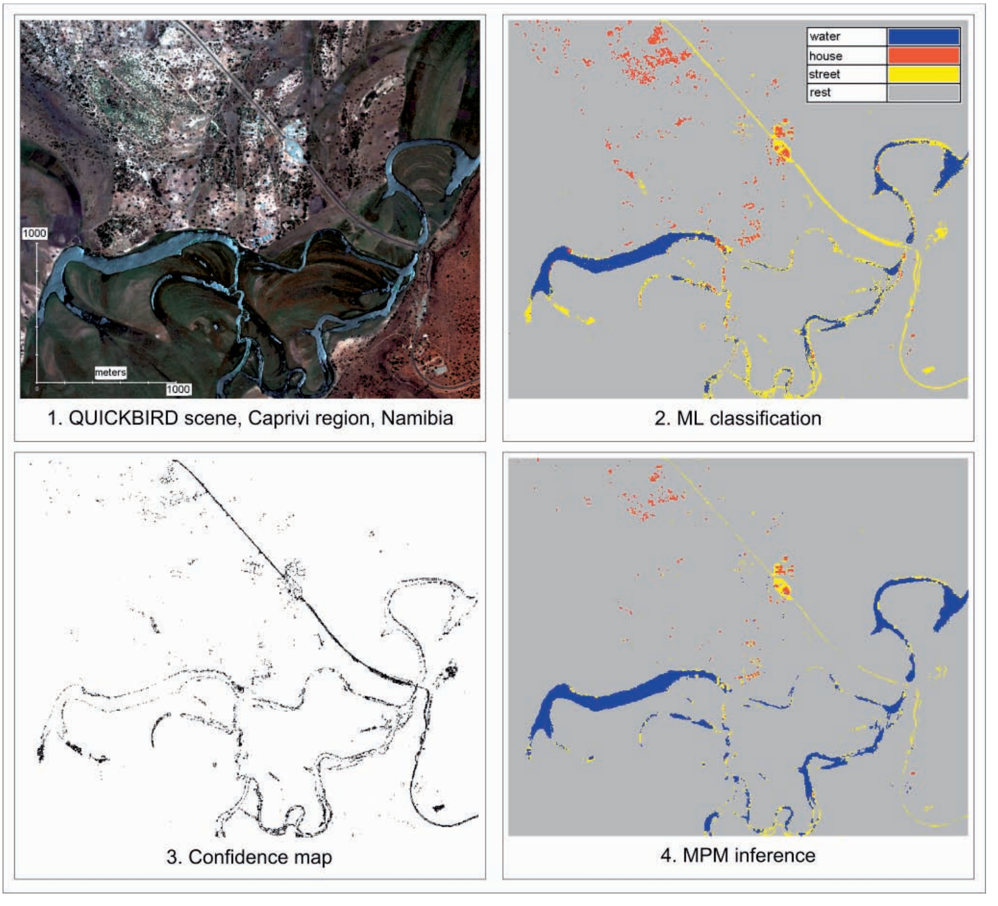

Fig. 4: Results for example 2.

reasonable and sufficient approach for this task. Other ways to tackle this problem are given with the application of the ICM algorithm only for those elements which exhibit other class memberships than their corresponding parent segment (MARTINIS et al. 2010) or for those elements where the posterior marginal entropy exceeds a defined threshold.

In the second experiment the incorporation of spatial context degrades the classification result since the tiny structures of houses and streets tend to disappear by increasing the weight of the context term. The application of the modified MAP estimation also leads to the omission of tiny structures. These observations lead to the conclusion, that the finest chosen region size (2 × 2 pixels) is too coarse to capture the tiny structures like houses and streets using square image regions, since these objects exhibit approximately the same respective a slightly bigger size. The original hierarchical MPM estimation yields the best but not the optimal result, whereas confusions between the class *water* and *street* as well as *rest* and *house* (many false positives for the class *street* and *house*) occur. The utilization of data models in more than two levels has shown to degrade the classification result further. The additional incorporation of a data model in level S^9 obviously provides no additional information and leads to more misclassifications. Instead of that, an additional data model in a coarser scale has shown to be useful to solve ambiguities.

The result of the modified MAP estimation in example 2 is an effect of both the limited separability provided by the data model as well as the chosen class hierarchy. The class *house* and *street* are only modeled in the finest level, whereas the class *rest* dominates the image. Hence, if the data model is not incorporated into the modified MAP computation (chi^2 test), the class *rest* is favoured for the corresponding image element in the finest level.

Compared to the grid-based estimation, the tree-based non-iterative estimators needed significantly less computational time (experiment 1: ML-MRF: 238 sec., MPM: 82 sec.). It has to be mentioned, that the parameter estimation for the prior model took 363 sec., which is quite acceptable compared to the grid-based MRF approach where the appropriate weight for the context term is usually identified in a trial-and-error manner.

The experiments were carried out using an Intel Xeon CPU with 2.33 GHz and 3.25 GB RAM. The whole framework is implemented in Interactive Data Language (IDL). The usage of other programming languages like C++ can accelerate the computational times further.

6 Conclusion

In this paper, a general approach for interactive classification of optical very high resolution satellite images is proposed. The framework is designed in order to cope with the surrounding conditions of rapid mapping activities and is moreover suitable for the task of satellite image classification in general. Hierarchical and spatial context information as well as image data from different levels of aggregation is incorporated into the classification process using hierarchical and lattice-based Markov random fields.

The experimental results confirm the effectiveness as well as the transferability of the framework on different datasets. Since the thematic classes are parameterized using Gaussian mixture models, the classification of image regions which represent a mixture of classes has shown to be very difficult even when hierarchical context information is incorporated. In order to decrease misclassifications, a new hierarchical MAP estimation is proposed. This modified MAP estimation has shown to outperform the original MAP and MPM estimation under the condition that the square image regions of the finest quadtree level are sufficiently smaller than the image areas of the particular thematic class. The results in (Kersten & Gähler 2010) also confirm these results.

The estimation of model parameters is a crucial task and depends on several aspects. Hence, especially the influence of the cardinality of the training data as well as the features in the feature space domain is addressed in current research activities. Furthermore, the application of the framework on several different image contents and thematic problems is currently evaluated.

The impact of incorporating different hierarchy levels should be analyzed in detail. The robustness of the parameter estimation with respect to other determining factors like the evaluation function applied in the feature selection method will be a topic in further investigations. Furthermore, the potential of detecting true misclassifications using the confidence map should be treated in the future work.

References

Awasthi, P., Gagrani, A. & Ravindran, B., 2007: Image Modeling using Tree Structured Conditional Random Fields. – International Joint Conference on Artificial Intelligence: 2060–2065.

Besag, J., 1986: On the statistical analysis of dirty pictures. – Joint Royal Statistic Society B **48** (3): 259–302.

Besag, J., 1974: Spatial Interaction and the Statistical Analysis of Lattice Systems. – Journal of the Royal Statistical Society, Series B **36**: 192–236.

Bouman, C. & Shapiro, M., 1994: A multiscale image model for Bayesian image segmentation. – IEEE Transactions on Image Processing **3:** 162–177.

Choi, M.J., Chandrasekaran, V., Malioutov, D.M., Johnson, J.K. & Willsky, A.S., 2008: Multiscale stochastic modelling for tractable inference and data assimilation. – Computer Methods in Applied Mechanics and Engineering **197:** 3492–3515.

Collet, C. & Murtagh, F., 2004: Multiband segmentation based on a hierarchical Markov model. – Pattern Recognition **37:** 2337–2347.

Dubes, R. & Jain, A.K., 1989: Random Fields in Image Analysis. – Journal of Applied statistics **16:** 131–164.

FENG, X., WILLIAMS, C.K.I. & FELDERHOF, S.N., 2002: Combining belief networks and neural networks for scene segmentation. – IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (4): 467–483.

FIEGUTH, P., KARL, W. & WILLSKY, A., 1998: Efficient multiresolution counterparts to variational methods of surface reconstruction. – Computer Vision and Image Understanding **70** (2): 157–176.

FORNEY, G.D., 1973: The Viterbi algorithm. – IEEE Proceedings **61:** 268–278.

KERSTEN, J. & GÄHLER, M., 2010: A Framework for Satellite Image Classification in the Context of Crisis Mapping using Markov Random Fields. – Remote Sensing and Photogrammetry Society Annual Conference 2010 (accepted).

KHEDAM, R. & BELHADJ-AISSA, A., 2003: Contextual fusion by genetic approach applied to the classification of satellite images. – Remote sensing in transition. Rotterdam' Millpresse.

LAFERTÉ, J.M., PÉREZ, P. & HEITZ, P., 2000: Discrete Markov Image Modeling and Inference on the Quadtree. – IEEE Transactions on Image Processing **9** (3).

MARTINIS, S., TWELE, A. & VOIGT, S., 2010: Unsupervised extraction of flood-induced backscatter changes in SAR data using Markov image modeling on irregular graphs. – IEEE Transactions on Geoscience and Remote Sensing **48** (12) (accepted).

PÉREZ, P., CHARDIN, A. & LAFERTÉ, J.M., 2000: Noniterative manipulation of discrete energy-based models for image analysis. – Pattern Recognition **33** (4): 573–586.

PUDIL, P., NOVOVICOVA, J. & KITTLER, J., 1994: Floating Search Methods in Feature Selection. – Pattern Recognition Letters **15** (11): 1119–1125.

VERVERIDIS, D. & KOTROPOULOS, C., 2008: Gaussian Mixture Modeling by Exploiting the Mahanalobis Distance. – IEEE Transactions on Signal Processing **56** (74): 2797–2811.

WILSON, R. & LI, C., 2003: A class of discrete multiresolution random fields and its application to image segmentation. – IEEE Transactions on Pattern Analysis and Machine Intelligence **25:** 42–56.

Address of the Authors:

Dipl.-Ing. JENS KERSTEN, Tel.: +49-8153-28-3384, Dr. MONIKA GAEHLER, Tel.: -3309, Dr. STEFAN VOIGT, German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), D-82234 Oberpfaffenhofen, Fax: -1445, e-mail: Name.Surname@dlr.de

Manuskript eingereicht: Juni 2010
Angenommen: September 2010