A Framework for Satellite Image Classification in the Context of Crisis Mapping using Markov Random Fields

J.Kersten[1] and M.Gähler[2]

German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Oberpfaffenhofen, Germany
[1]Email: jens.kersten@dlr.de
[2]Email: monika.gaehler@dlr.de

Abstract
In this contribution a general framework for classification of very high resolution optical satellite images is proposed and evaluated. This approach is designed in order to cope with the specific conditions accompanied by crisis mapping applications and is moreover well suited for several other applications. Multiscale image information (data model) as well as hierarchical and spatial context information (prior model) is incorporated into the classification process using a hybrid Markov model which combines a hierarchical directed as well as a planar lattice-based Markov Random Field (MRF). The modelling of arbitrary semantic classes in different scales enables the definition of a hierarchical semantic network representing the dependencies and relations between the classes in adjacent scales. Classification is carried out using non-iterative hierarchical maximum a-posteriori (MAP) or mode of posterior marginal (MPM) inference as well as a subsequent optimization step using a planar MRF. Additionally, a modified MAP inference which is able to outperform the original inference methods is proposed. The impact of incorporation of image data from multiple scales is evaluated in this contribution. Furthermore, the dependency between the quantity of the training data and the classification accuracy is analyzed.

1 Introduction
The (pre-) classification of very high resolution optical satellite images is a difficult challenge. Near real time processing as well as the transferability of the methods to various crisis scenarios is highly desired in emergency and crisis mapping applications. Representatives for recurrent tasks during crisis scenarios are given with the classification of water surfaces (flood events) as well as 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. The proposed general framework 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 classification results. Due to near real time processing requirements, classification is here restated as a task of semantic annotation of square image regions.

Contextual information can improve classification accuracy significantly, if such information can be well modelled (Khedam and 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 modelled as a Markov Random Field (MRF). In order to capture the intrinsic hierarchical nature of remote sensing data, several efficient Markov image modelling approaches defined on tree structures were proposed during the last two decades (Bouman and Shapiro, 1994; Fieguth *et al.,* 1998; Lafferty *et al.*, 2001; Wilson and 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 and explored in this article. The motivation for using such a model is to provide a general and computational effective framework for classification or pre-classification of multispectral satellite images.

In this paper, the framework is described briefly. Especially, the influence of the quantity of the training data as well as the design of the hierarchical semantic network on the classification accuracy is explored.

The article is organized as follows: In section 2, the proposed general framework is introduced briefly. In section 3, the experiments are described. The results of the experiment as well as the relevance and efficiency of the proposed framework are demonstrated and discussed in section 4. Conclusions are drawn in section 5.

2 The Framework

Modelling image characteristics in a hierarchical manner has shown to be valuable for many applications, e.g. image labelling and object detection (Awasthi *et al.*, 2007), multiband segmentation of astronomical images (Collet and 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), MRF´s 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.

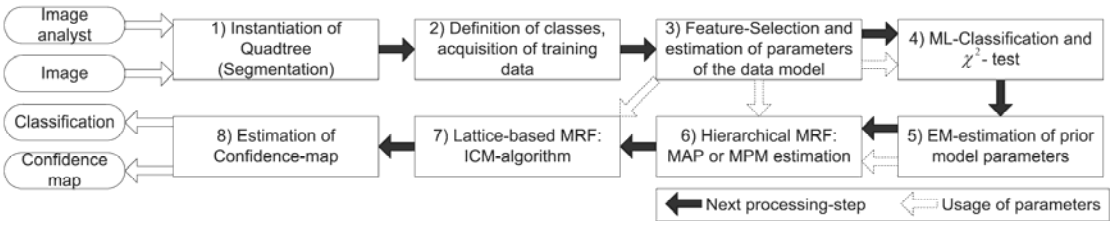

Figure 1: Workflow for the proposed framework.

Image classification should be possible even when no additional information like vector data or digital elevation models (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 figure 1 and can be described as follows: 1) First, a complete quadtree image representation is instantiated. The size of the smallest region can be defined individually depending on the spatial resolution of the image as well as the structure of the thematic classes. 2) The framework allows an interactive definition of arbitrary semantic classes in different scales (quadtree levels), i.e. the modelling of a hierarchical semantic network. For each class, the image analyst has to provide training data (image regions). 3) Based on the training data the parameters of Gaussian mixture models (data model) are estimated. Relevant features are identified through feature selection independently for each scale. 4) A constrained maximum likelihood classification (chi-square test) is carried out in order to obtain training data (labelled 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. Additionally, a modified MAP inference (Kersten *et al.*, 2010) is utilized. The image information of regions which exhibit low conditional likelihoods for all classes is not incorporated during the inference procedure (application of a chi-square test). 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 labelling process, a confidence map is computed.

For further information concerning the methodology utilized in the framework the reader is referred to (Kersten *et al.*, 2010).

3 Description of Experiments

For each image all parameters for the inference of the class labels are learned individually based on the training data provided by the image analyst. In this paper the influence of the cardinality of training data (experiment 1) as well as the impact of the incorporation of image data from multiple scales (experiment 2) concerning the overall classification accuracy is evaluated. For each of the two experiments the classification of landcover of the following two images is carried out.

Figure 2: Images for the experiments. Left: image 1, right: image 2. The scale bar in image 1 is valid for both images.

The images are subsets (512x512 pixels) from a pan-sharpened multispectral IKONOS scene with a spatial resolution of 1 m acquired on August 6th, 2007. The classification of the following classes is desired: *dark field, field, street* and *vegetation* (image 1) as well as *vegetation, house* and *rest* (image 2). A quantitative evaluation can be done based on references provided by visual interpretation. The results of the experiments should serve as a guideline for the application of this framework in emergency- and crisis mapping activities as well as a basis for further modifications of the framework.

3.1 Experiment 1: data model

The first experiment focuses on the robustness of the estimation of the data model parameters (step 3 in figure 1) with respect to both the cardinality of the training data and the number of dimensions in the feature space domain. Therefore, a leave-x-out cross validation with x=10 is carried out using 50, 80, 110, 140, 170 and 200 training segments for each class in the finest quadtree level (image element size: 4x4 pixels). Furthermore, for each number of training segments 3, 5, 7 and 10 "optimal" features were identified by feature selection individually. The resulting overall accuracies are average values of 20 runs for each constellation. Hence, 480 training and classification processes were carried out for each image. The amount of all image segments is divided into the groups of test and training data, where only image elements which are not used for the training of the data model are incorporated into the evaluation. Since this experiment focuses on the parameters of the data model, the results of maximum likelihood classifications using the different numbers of training segments are carried out and compared (see section 4.1).

3.2 Experiment 2: hierarchical inference

In this experiment the impact of incorporation of image information from multiple scales (quadtree levels) as well as the appropriate modelling of contextual dependencies between thematic classes in adjacent scales (i.e., a hierarchical semantic network) is evaluated. Therefore, data models in different combinations of scales are trained based on the results of experiment 1. For both images a quadtree representation with nine levels is instantiated. Hence, the finest image element has a size of 2x2 pixels and is denoted as S^9. Data models are trained in the following combinations of scales: (S^9); (S^9, S^8); (S^9, S^8, S^7) as well as (S^9, S^7). The overall accuracy of the hierarchical mode of posterior marginals (MPM) inference is compared. The modified MAP inference is computed for the best combination of scales. Additionally, a subsequent optimization step using a lattice-based MRF is carried out for both the best MPM result and the result of the modified MAP inference.

4 Results and discussion

4.1 Experiment 1: data model

Due to the non-complex image content the mean overall classification accuracy for image 1 lies between 92.0 % and 95.2 % (figure 3, left). A feature space representation using 3 features has shown to be suboptimal for the classification compared to the other configurations (5, 7 and 10 features). Furthermore, for this image a number of training segments greater than 80 for each thematic class has shown to provide relatively robust parameter estimation. The error bars in figure 1 express the standard deviation of the overall accuracy (scaled by a factor of 0,1 due to graphical clarity) estimated based on the results of the cross validation. In this experiment, the standard deviation lies between 0.23 and 4.18. Using 200 training segments for each class yields the lowest standard deviation for all numbers of features.

For the second image the usage of 3 features has also shown to give no reasonable results. On the other hand, using 5 and 7 features yields results between 73.6 % and 77.9 % mean overall accuracy. A 10-dimesional feature-space leads to poorer classification results than lower dimensional representations. One reason for this is given with the well known "curse of dimensions". Compared to the first image, the standard deviation of the overall accuracy is relatively high (2.07 – 7.46) and generally decreases with an increasing number of training segments.

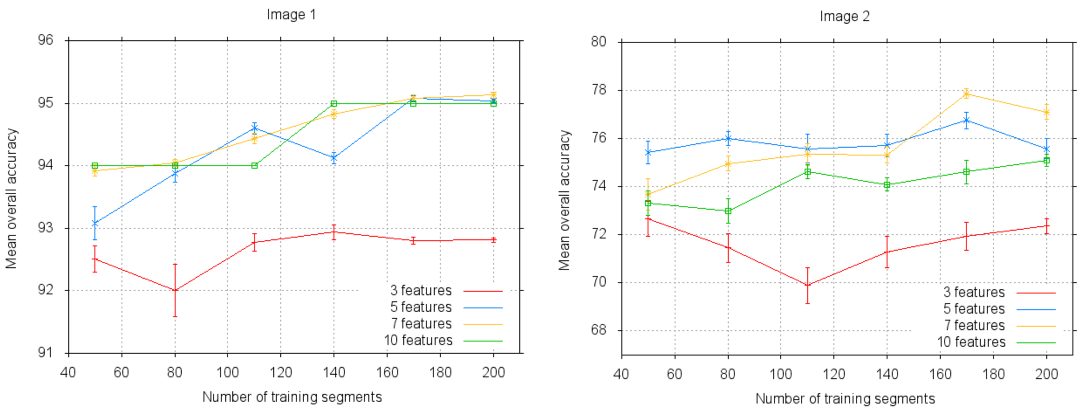

Figure 3: Results of experiment 1. Mean overall accuracy over the number of training segments per thematic class. Error bars: standard deviation of the overall accuracy (scaled by a factor of 0.1 due to graphical clarity).

This experiment demonstrates that the estimation of the parameters of the data model (Gaussian mixture model) is a difficult task which depends on several aspects. In this contribution the impact of the cardinality of the training data as well as of the features is evaluated. For the first image expected and reasonable results are obtained, since the mean classification accuracy increases respective remains relatively constant when the amount of training data per thematic class is increased. Furthermore, the variation of the overall accuracy decreases for large sets of training data. This observation also holds for the second image. Compared to the first image, the separation between the thematic classes in the feature space domain is obviously more complex here. The overall accuracy is lower than in the first image and a moderate feature number is required (between 5 and 7) in order to provide the best possible separability between the classes. An increasing number of training segments leads to more robust results (lower standard deviations) while the improvement of the overall accuracy is moderate.

As expected, the two examples in this experiment point out that the performance of the data model incorporated into the framework heavily depends on the amount of training data as well as the number of dimensions of the feature space. Basically it could be shown, that a large amount of training data per class (greater than 80 segments) leads to good (i.e. consistent) results, if the number of features is well chosen (here: 5-10 features).

In this experiment variations of selected features occur during the cross validation, which demonstrate the difficulty of model detection. One reason for this is given with the suboptimal feature selection method.

4.2 Experiment 2: hierarchical inference

As described in section 3.2 four different data models (i.e., different combinations of incorporated quadtree levels) are trained for each image. The three involved quadtree levels consist of image elements of the sizes 2x2, 4x4 and 8x8 pixels. In table 1 the results for image 1 are shown. Taking into account the results of experiment 1 as well as the computational efforts, five features and 100 training segments per class and level are used for the training of the data models.

A non-contextual maximum likelihood (ML) classification provides a good result (94.99 %) which can be further improved through incorporating multiscale image and context information. All thematic classes are trained in each incorporated level. Using the non-iterative hierarchical MPM inference, the best result is carried out by incorporating all three levels (S^9, S^8, S^7).

Table 1. Results of experiment 2: image 1.

Description	ML	S^9	S^9, S^8	S^9, S^8, S^7	S^9, S^7	MPM-MRF S^9, S^8, S^7	Mod. MAP-MRF S^9, S^8, S^7
Overall Accuracy	94.99 %	95.44 %	95.26 %	96.18 %	94.90 %	96.59 %	96.75 %

The subsequent application of the lattice-based MRF (weight of context term β=100.0 and first order neighbourhood) in the finest quadtree level is able to improve this result (MPM-MRF S^9, S^8, S^7). The modified MAP inference (Kersten *et al.*, 2010) combined with a subsequent lattice-based MRF (Mod. MAP-MRF S^9, S^8, S^7) is able to slightly outperform the standard inference methods and yields the best overall classification accuracy here (chi² test with a probability of error α = 0.3).

For the second image the three desired classes are modelled in the two finest quadtree levels. Due to the large size of the image elements in the coarse level S^7, only the classes *house* and *rest* are modelled here. In order to keep the computational efforts low, five features and 100 training segments per class are used here. Similar to the first image, the incorporation of all three levels yields the best MPM result and obviously increases the ability of resolving local ambiguities (see table 2).

Table 2. Results of experiment 2: image 2.

Description	ML	S^9	S^9, S^8	S^9, S^8, S^7	S^9, S^7	MPM-MRF S^9, S^8, S^7	Mod. MAP-MRF S^9, S^8, S^7
Overall Accuracy	70.50 %	71.67 %	74.38 %	75.78 %	71.70 %	77.45 %	79.43 %

The application of the lattice-based MRF (smoothing factor β=20.0) on the best MPM result (S^9, S^8, S^7) improves the overall accuracy significantly. The modified MAP inference yields the best result with an overall accuracy of 79.43 % (with α = 0.3).

This experiment demonstrates the additional benefit of the incorporation of multiscale and context information into the classification process. The quality of the result depends on the modelled class hierarchy, where the definition of an appropriate class hierarchy is a difficult task. In this paper the incorporation of image data from three quadtree levels yields the best result using the hierarchical MPM inference. A subsequent application of a lattice-based MRF further improves these results.

The modified MAP inference outperforms the standard inference methods in both experiments. Since the data conditional likelihoods are modelled using multivariate Gaussian mixture models, misclassifications of image regions which e.g. represent class transition areas may be likely even when context information is included. The modified MAP inference avoids the incorporation of image information in the case of low conditional likelihoods for all classes.

An experimental version of the framework is implemented in Interactive Data Language (IDL). In table 3 the computational times for the processing steps and configurations respective image 1 are presented.

Table 3. Computational times in seconds for experiment 2, image 2.

Configuration	S^9	S^9, S^8	S^9, S^8, S^7	S^9, S^7
Feature selection and estimation of data model parameters	283.1	504.9	585.7	372.9
EM-estimation of prior model parameters	421.1	691.5	515.5	907.5
MPM inference	47.4	61.9	72.2	46.5
Maximum likelihood classification	48.4	-	-	-
Lattice-based MRF	-	-	149.2	-
Modified MAP inference	-	-	40.2	-

The best classification result is here obtained in 23 minutes. Additionally, the acquisition of training data takes approximately 10 minutes for this experiment. An optimization of the implementation as well as the usage of other programming languages like C++ can further accelerate the computational times. Concerning an application of the framework in operational rapid mapping activities a maximum overall processing time of one hour for arbitrary image sizes is desired.

5 Conclusion

In this paper a general framework for fast classification or pre-classification of optical satellite images is briefly described and evaluated. This approach is designed in order to cope with the specific conditions accompanied by crisis mapping activations and is moreover well suited for several other applications. Two experiments are carried out in order to point out the performance of the framework respective parameter estimation of the data model as well as the incorporation of data and context from multiple scales (i.e. quadtree levels). Basically the first experiment shows that a large amount of training data per class (greater than 100 segments) leads to robust results, if the number of features is well chosen (here: 5-10 features). In the second experiment the additional benefit of the incorporation of spatial and multiscale context and image information is demonstrated. The hierarchical quadtree model allows fast and non-iterative inference of the unknown class labels. In addition, a subsequent application of a lattice-based MRF improves the result of the hierarchical inference methods. The modified MAP estimation has shown to outperform the standard inference methods.

Concerning the data model there are several factors affecting the parameter estimation in this framework, e.g. the normalization of the features, the split-based clustering algorithm (criterion for cluster splitting and the initialization of new clusters), feature selection (e.g. evaluation criterion for a distinct feature combination) as well as the choice of training data. Small variations concerning the training data (cross validation) obviously have a significant influence on the resulting data model (see standard deviations of overall accuracy). Hence, further analysis of the framework will be a topic of investigations in the future.

6 Acknowledgements

This document is an extension of the publication (Kersten *et al.*, 2010).

References

AWASTHI, P., GAGRANI, A., and RAVINDRAN, B., 2007. Image Modeling using Tree Structured Conditional Random Fields. *IJCAI-2007:* 2060-2065.

BOUMAN, C., and 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., and 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., and MURTAGH, F., 2004. Multiband segmentation based on a hierarchical Markov model. *Pattern Recognition*, **37**: 2337-2347.

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

KERSTEN, J., GÄHLER, M., and VOIGT, S., 2010. A Framework for fast and interactive Classification of optical VHR Satellite Imagery using Markov Random Fields. *In: PFG 6/2010, "DeSecure". E. Schweizerbart'sche Verlagsbuchhandlung (Nägele u. Obermiller) Stuttgart.* (Submitted)

KHEDAM, R., and BELHADJ-AISSA, A., 2003. Contextual fusion by genetic approach applied to the classification of satellite images. *R. Goossens (Ed.), Remote sensing in transition. Rotterdam' Millpresse.*

LAFFERTY, J., MCCALLUM, A., and PEREIRA, F., 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. *In Proceedings of International Conference on Machine Learning.*

Martinis, S., Twele, A. and 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., and LAFERTÉ, 2000. Noniterative manipulation of discrete energy-based models for image analysis. *Pattern Recognition,* **33**(4): 573-586.

WILSON, R., and 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.