Image Analysis Based on Probabilistic Models

Christian Heipke, Franz Rottensteiner, Hannover

ABSTRACT

This paper discusses random field based image classification methods, and in particular conditional random fields (CRF), for topographic mapping. A short review of the CRF principles reveals their main advantages, namely the possibility to incorporate local context into the classification to quantify the quality of the results in terms of probabilities. Three examples, the classification of point cloud data, multi-temporal and multi-scale classification of satellite images of different epochs and geometric resolution as well as the verification of existing land use data demonstrate the power and flexibility of CRF, but also its limitation in terms of capturing long range context. The paper closes with a short discussion on how to overcome this deficiency in the future.

1. INTRODUCTION

This paper deals with image analysis. Already in 1982, Arziel Rosenfeld, one of the pioneers of the field, defined image analysis as the "automatic generation of an explicit meaningful description of physical objects in the real world from images" (Rosenfeld 1982), where the input can be single images, stereo pairs or image sequences, and the term "image" refers to brightness images of a given number of spectral bands as well as to depth images, thus including multi- and hyperspectral images as well as 3D point clouds. This definition sets image analysis apart from an attempt to completely describe a scene depicted in images and also from understanding a whole process, of which the images depict only snapshots (Radig, 1993).

Applications of image analysis can be found wherever images are present – topographic mapping, environmental monitoring, precision agriculture, autonomous driving, industrial metrology, cultural heritage, medical imaging, surveillance are just a few examples. The reason is that image acquisition is typically not a task in its own right, but has the purpose of describing the depicted scene in terms of the objects present in that scene. While in the past, the images had to be interpreted by a human operator, due to significant scientific and technical developments in the last few decades the automatic extraction of an object description from the image data is in reach today. The type of description needed has to be defined beforehand and is dictated by both, the application and the image characteristics.

It is important to realise that in order to automatically recognise an object in images, geometric and radiometric knowledge about this object, e. g. the size and form and its appearance in the image is required. In the literature, mainly two different kinds of knowledge representations are reported. Model-based approaches incorporate the knowledge in the form of explicit object models. Knowledge based systems of sometimes very high complexity have been developed for this task (Niemann et al. 1990; Stilla et al. 1995; Pakzad et al. 2001). Also most of the work on building and road extraction in the photogrammetric community (Fischer et al. 1998; Mayer 1999; Heipke et al. 2000; Rottensteiner 2008) belongs to this group of approaches. While a convincing success could be achieved for individual tasks and examples, despite the often generic object models used it is rather difficult to transfer them to other domains or geographic areas. Another point of criticism has always been the issue of self-diagnosis: the algorithms do not provide any information about the quality of the obtained results. The first problem can be solved by combining different object models in a meta-approach (Ziems 2014), but the combination again needs the quality of the individual results as input in order to properly deal with potential conflicts.

Recent years have seen an increased interest in statistical methods being applied to image analysis. In essence rather than relying on pre-defined models these models are learnt from examples. In classical remote sensing such approaches have long been in use, the most well-known probably being maximum likelihood (ML) classification, with many others have been developed over the years, based on principles of pattern recognition and machine learning (Bishop 2006). Classification is based on a set of features derived from the image data. These features must be selected in a way to cluster the object classes in feature space and to separate different clusters in some optimal way. Instead of pre-defined object models, training data are needed, which must adequately represent all desired object classes in terms of the features used for classification. The object model is thus implicitly encoded in the training data. As a consequence statistical approaches can be more easily adapted to different domains by using a new set of training data.

Statistical classification can be subdivided into parametric and non-parametric approaches. Parametric approaches require the functional model of the probability density function (pdf) associated with the object classes to be known a priori, the corresponding parameters (e. g. the position and width of a Gaussian) are learnt during training. A particular advantage is that the classification result is expressed as a probability that the entities (individual pixels/points or segments) belong to certain classes given the observed features (and possibly additional evidence such as context information). Using the MAP (maximum a posteriori) criterion the most probable overall solution can thus be found, and the quality of the result in terms of the related probability can be stated. The Bayes classifier, logistic regression, Markov Random Fields (MRF) and Conditional Random Fields (CRF) are examples of probabilistic classifiers (see Bishop 2006 for a detailed description of these classifiers).

Non-parametric approaches such as support vector machines (SVM; Vapnik 1998) and random forests (RF; Breiman 2001) do not make use of pdf. Classification is rather based on the estimation of a decision boundary between different classes (in case of SVM) or randomly chosen decision rules (RF). While it is harder to interpret the results of non-parametric approaches, they usually still yield acceptable results when only few training data are available. The reason is that only the decision boundary or the decision rules need to be learnt which typically needs less information than describing the relatively complex pdf.

This paper is structured as follows: in the next chapter we give an overview of Conditional Random Fields (CRF), a probabilistic approach to image analysis which incorporates local context into the classification process and has gained some popularity in recent years. We then present three examples, in which CRFs have been employed, to illustrate the flexibility and the limitations of this approach. We conclude with a few remarks on further research necessary to combine probabilistic approaches with those using stronger object models.

2. CONDITIONAL RANDOM FIELDS

This chapter contains a short introduction to Conditional Random Fields (CRF); it is based on Albert et al. (2014). CRF are a flexible framework for supervised contextual classification. They were first suggested for image classification by Kumar and Hebert (2006). CRF are undirected graphical models (Förstner 2013). The graph nodes represent the image sites, e.g. pixels or segments. The edges link adjacent nodes and model statistical dependencies between class labels and data at neighbouring image sites. The class labels of all image sites are combined into a label vector $\mathbf{y} = [y_1, \dots, y_i, \dots, y_n]$, where $i \in S$ is the index of an image site and S is the set of all image sites. The goal is to assign the most probable class labels $\mathbf{y}$ from a set of classes to all image sites simultaneously considering the data $\mathbf{x}$. CRF directly model the posterior probability $P(\mathbf{y}|\mathbf{x})$ of the label vector $\mathbf{y}$ given the observed data $\mathbf{x}$:

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z} \prod_{i \in S} \varphi(y_i, \mathbf{x}) \prod_{\{i,j\} \in E} \psi(y_i, y_j, \mathbf{x}). \tag{1}$$

In equation 1, $\varphi(y_i, \mathbf{x})$ are unary potentials, called *association potentials* and $\psi(y_i, y_j, \mathbf{x})$ are pairwise potentials, called *interaction potentials*. The partition function Z acts as a normalization constant which transforms the potentials into probabilities, E is the set of edges in the graph underlying the CRF, typically based on a neighbourhood system defined on the node set, and {i,j} is an edge between neighbouring sites i and j.

The association potential indicates how likely a node i belongs to a class y_i given the observations $\mathbf{x}$. The interaction potential models the relations between the labels y_i and y_j of adjacent nodes and the observations $\mathbf{x}$, this is where context information is introduced.

CRF represent a general framework, which allows introducing different functional models for both potentials (Kumar, Hebert 2006). It is possible to choose any arbitrary discriminative classifiers with a probabilistic output $P(y_i|\mathbf{x})$ for the association potential. This also applies to the interaction potential. Kumar and Hebert (2006) use a generalized linear model for the association potential, but several other classifiers have proven to work well, for instance a RF classifier (Schindler 2012). The models applied for the interaction potential are often simpler, favouring identical labels for neighbouring sites, penalising label changes, and thus effectively smoothing the result. Some approaches apply more complex models for the interaction potential in order to avoid over-smoothing. Depending on how exactly the interaction potential is defined, it may be necessary to provide fully labelled training data in order to learn all possible class transitions. Such a requirement may be hard to fulfil and should be taken into account when designing the interaction potential.

Training and inference, respectively, are the tasks of determining the optimal parameters of the potentials and the optimal label configuration of the data. In the inference step, the most probable label configuration of the graphical model is determined for all nodes simultaneously. This is based on maximizing the posterior probability $P(\mathbf{y}|\mathbf{x})$ of the labels given the data. For larger graphs with cycles exact inference is computationally intractable and approximate methods have to be applied. In the investigations described in this paper we use Loopy Belief Propagation, a standard message passing algorithm (Frey, MacKay, 1998).

An example of the results which can be obtained using a CRF is depicted in Fig. 1. Fig. 1a shows the image to be classified, Fig. 1b the result of an ML classification, in Fig. 1c the CRF result is depicted. The ML result is much noisier, the overall accuracy based on independently acquired ground truth amounts to 73.2%. As was to be expected the CRF has a much smoother appearance, the obtained overall accuracy lies at 78.9%. This example illustrates the advantages of incorporating context into the classification procedure in the way described.
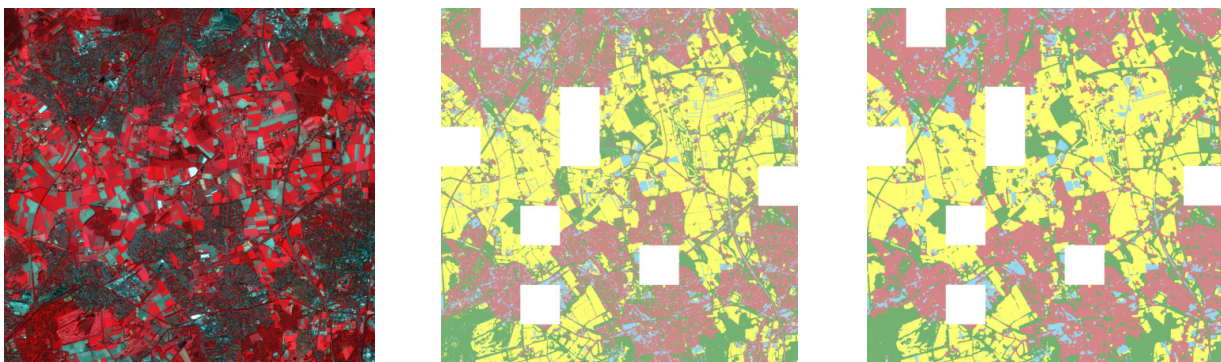

Figure 1: Classification of a satellite image (a) using the ML (b) and the CRF (c) approach. It can clearly be seen that the ML result looks noisier, it also has a smaller overall accuracy of 73.2% vs. 78.9% for CRF (the white areas represent training data).

3. APPLICATIONS

In this chapter three different applications are presented. They have been drawn from work carried out at the institute of the authors and have been selected to illustrate the possibilities of CRF and a number of possible extensions. These applications deal with:

- Classification of airborne laser scanning data (Niemeyer et al. 2014). In this example CRFs are used to classify irregularly distributed point clouds (as opposed to regularly distributed image data).
- Multi-temporal and multi-scale classification (Hoberg et al. 2015). Here, the basic CRF equation is extended to incorporate a temporal interaction potential and a hierarchical structure.
- Verification of large scale land use data (Albert et al. 2014). This example shows how multiple layer CRFs can be employed to classify both, land cover and land use, in a sequential manner.

3.1. Classification of airborne laser scanning data

Airborne laser scanning data constitute an important base for the generation of 3D city models. For this task the 3D point cloud first needs to be classified into different classes such as ground surface, roof surface, vegetation etc. In our case this task is carried out using a CRF (Niemeyer et al. 2014). The nodes of the graph on which the CRF is based are chosen to correspond to the 3D points. The neighbourhood is defined as the k nearest neighbours in the planimetric projection, which takes care of varying point density and at the same time conserves the relationship between points at different height levels, e. g. between a point on the roof and a neighbouring one on the ground.

The association potential is based on the results of a random forest classification with a large set of features incl. height above ground, intensity, and a geometric characterisation of the immediate neighbourhood. Also the interaction potential is based on a random forest classification, this time for different object class configurations rather than the class of individual points. The parameters of both potentials are trained separately.

For the evaluation of the method the ISPRS benchmark data set Vaihingen (Rottensteiner et al. 2014) was used containing three test areas with a point density of approximately 7 pts./m^2, see Fig. 2. Training and checking data were captured manually for the classes *road, inclined roof, flat roof, façade, grass, low vegetation* and *tree*.

Figure 2: Three test sites of Vaihingen: (a) Inner city, (b) High riser, (c) Residential.

The results in terms of a confusion matrix are shown in Tab. 1. Fig. 3 contains the visual result for the three test areas. Overall they can be characterised as good. A limiting factor was the lack of sufficient training data for some classes (e. g. façade) and class transitions. In particular, *road* and

street as well as *low vegetation* and *tree* were mixed in a number of cases. The improvement of using a CRF vs. taking the outputs of the random forest classifications as the final result was 2% in the overall accuracy, but more pronounced for the correctness and completeness of individual classes. A particular problem is roof superstructure, which leads to high roughness of some roof areas and as a consequence to a classification as *tree*. An improvement is possible by increasing the point density and by introducing context information over larger ranges.
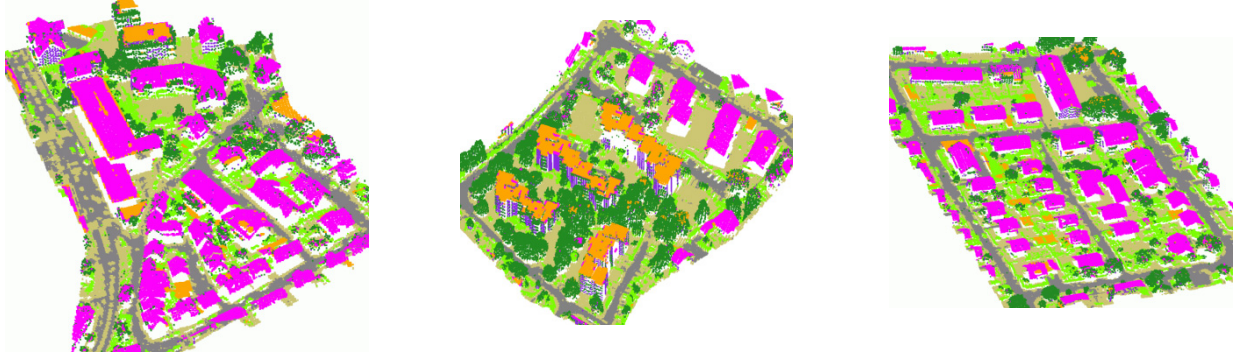

Figure 3: Results of the test areas: Inner city (a), High riser (b) and Residential (c): *road*: grey; *inclined roof*: purple; *flat roof*: orange; *façade*: violet; *grass*: brown; *low vegetation*: light green; *tree*: dark green (Niemeyer et al. 2014).

Class	Road	Inclined roof	Flat roof	Façade	Grass	Low veg.	Tree	Correctness [%]	
Road	187725	24	50	332	13182	274	182	93.0	(3.5)
Inclined roof	6	112116	742	2567	327	318	3568	93.7	(-0.2)
Flat roof	107	3909	44913	5043	2724	1372	958	76.1	(1.8)
Façade	116	1071	147	16932	662	1497	7971	59.6	(10.3)
Grass	26434	31	347	1338	149110	5166	1827	80.9	(1.8)
Low veg.	1198	549	1267	2863	10329	39107	21453	50.9	(0.7)
Tree	4	1993	1.231	3430	370	2607	101390	91.3	(-0.5)
Completeness [%]	87.1 (1.3)	93.7 (0.1)	92.2 (-1.0)	52.1 (-0.9)	84.4 (5.3)	77.7 (2.5)	73.8 (2.3)	Overall acc. 83.4	(2.0)

Table 1: Confusion matrix (columns: reference classes; lines: classification result) for CRF-based classification of the three test data sets (Niemeyer et al. 2014). Figures in brackets denote the difference to a classification without interaction potentials; positive values show an improvement.

3.2. Multi-temporal multi-scale classification of optical satellite images

The goal of this project was the simultaneous classification of co-registered multi-temporal satellite orthophotos of different geometric ground sampling distance (GSD) in order to improve classification accuracy. For achieving these goals the CRF formulation needed to be extended to incorporate different epochs and a separate resolution for each epoch, see equation 2 and Fig. 4:

$$P(\mathbf{y}|\mathbf{x}) = \frac{1}{Z} \prod_{i \in S} \varphi(y_i^t, \mathbf{x}^t) \prod_{\{i,j\} \in E_s} \psi_s(y_i^t, y_j^t, \mathbf{x}^t) \prod_{\{k,l\} \in E_t} \tau^{mn}(y_k^m, y_l^n). \tag{2}$$
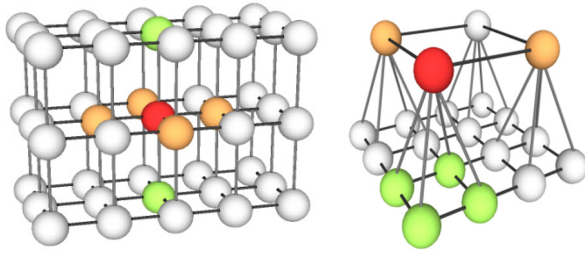
Figure 4: Extended graph structure of the CRF: (a) for multiple epochs, (b) for varying ground sampling distance. Red: pixel under consideration; orange: spatial neighbours; green: temporal neighbours (Hoberg et al. 2015).

In equation 2 subscript i denotes a particular pixel in a particular epoch and superscript t denotes the different epochs. E_S is the set of spatial edges, where each edge {i,j} corresponds to a pair of neighbouring pixels on the image lattice. The new temporal interaction potential $\tau^{mn}(y_k^m, y_l^n)$ links the class labels of pixels k and l between two different epochs m and n that are temporal neighbours and, thus, correspond to an edge in the set of temporal edges E_t (see Fig. 4a). Due to the potentially different resolutions of the images there may be more than one temporal neighbour for a given pixel (see Fig. 4b), also the class structure may have to be adapted.

The association potential is modelled as a multivariate Gaussian distribution, for the spatial interaction potential a contrast-sensitive Potts model is introduced. The key element of the temporal interaction potential is a transition matrix TM (Pakzad et al. 2001; Bruzzone et al. 2004). The elements of the TM depend on the probabilities of a change of class from one epoch to the next.

In order to investigate the characteristics of this multi-temporal and multi-scale CRF approach a number of experiments were conducted (see Hoberg et al. 2015). Here we report on a subset of these, restricting ourselves to the comparison of a standard classifier (we use ML for this purpose), a multi-temporal approach based on Hidden Markov Models (HMM) without spatial context (Leite et al. 2011), a mono-temporal CRF and a multi-temporal CRF. The data set consisted of an Ikonos image of 2005, a RapidEye image of 2009 and a Landsat image of 2010, depicting a scene of about 50 km^2 in Herne (Germany). Reference data for the classes *residential area*, *industrial area*, *forest* and *cropland* were digitised manually from the Ikonos image; for the Landsat image the two classes *residential area* and *industrial area* had to merged into *built-up area*. For the classification a large number of features were initially defined, subsequently a selection according to Hall (1999) was employed, which yielded a reduced set of features. This greedy approach selects features which are highly correlated with the class labels but uncorrelated with other features.

The results are depicted in Tab. 2. It can be seen that significant accuracy gains were achieved when simultaneously considering spatial and temporal context. Both, the mono-temporal and the multi-temporal CRF improve the results with respect to ML classification, while taking into account all scenes simultaneously yields an additional advantage.

Image	ML	HMM	CRF_mono	CRF_multi
Ikonos 2005	74.2	77.3	79.8	81.5
RapidEye 2009	64.4	75.9	68.6	80.6
Landsat 2010	71.7	79.2	75.9	80.4

Table 2: Overall accuracy of the different approaches in % (Hoberg et al. 2015).

3.3. Verification of land use and land cover

In this section an approach for the verification of existing land use information of a cadastral data base is described (Albert et al. 2014). It is applied to check the land use data of ALKIS (Amtliches Liegenschaftskatasterinformationssystem; Authoritative Real Estate Cadastre Information System). Input data comprise recently acquired multi-spectral aerial images (0.2 m GSD) and a digital surface model (0.5 m GSD) generated from these images, and a digital terrain model with a post spacing of 5 m. In a pre-processing step orthophotos were derived from the original aerial images.

It is well known that images only allow determining land cover, not land use. To cope with this issue a two-stage approach was chosen: in the first step land cover was determined for each pixel, whereas in the second step land use was derived based on the existing cadastral area objects and the results of the previous step. For both classifications CRFs were used. In the context of this paper, this section illustrates the possibility of CRF to be employed for both, regularly spaced individual pixels as well as larger segments with an irregular shape and neighbourhood structure. The reason that the irregular neighbourhood structure can be mapped to CRF is of course the use of graph theory as the underlying concept.

For the association potential of the pixel-based land cover classification, containing a total of nine classes, random forests were employed, while for the interaction potential a contrast-sensitive Potts model was used. The following land use classification comprised seven classes and was based on random forests for both potentials.

The resulting confusion matrices are shown in Tab. 3 for land cover and in Tab. 4 for land use. They were produced in a test area of some 12 km^2 in the vicinity of the city of Hameln, Germany. For land cover the overall accuracy was 81.3%. The best completeness and correctness values were achieved for the class *building*, but also the classes *sealed area*, *grass*, *tree* and *water* yielded good completeness and correctness values. Lower values for the class *bare soil* are caused by an overall smaller number of training samples for this class, probably not representing the whole range of characteristics of this class to a sufficient degree. A problem turned out to be the discrimination of

		Classification							Com-pleteness [%]
		Buil-ding	Sealed	Soil	Grass	Tree	Water	Car	
Reference	Building	22.5	1.2	-	0.2	0.3	-	0.3	90.1
	Sealed	1.5	17.6	0.1	0.6	0.4	-	1.3	79.2
	Soil	-	0.2	1.9	0.4	0.2	-	-	68.7
	Grass	0.4	1.2	0.6	22.7	2.9	0.1	0.1	79.7
	Tree	0.2	0.6	0.2	2.2	13.2	0.1	-	79.3
	Water	-	0.1	-	0.1	0.1	1.8	-	85.9
	Car	0.1	0.2	-	-	-	-	1.2	76.0
Correctness [%]		90.7	82.3	66.9	86.4	76.8	89.5	38.7	

Table 3: Confusion matrix (all values in %) for land cover classification (Albert et al. 2014).

		Classification						Com-pleteness [%]
		Resi-dential	Street	Rail-way	Water	Agri-culture	Forest	
Reference	Residential	58.8	4.2	-	-	0.1	-	89.2
	Street	1.6	16.3	0.1	-	0.2	-	85.6
	Railway	0.1	0.5	0.2	-	-	-	21.1
	Water	0.1	0.1	-	0.4	-	-	51.2
	Agriculture	-	0.3	-	0.1	3.0	-	75.2
	Forest	-	0.2	-	-	0.2	0.6	38.4
Correctness [%]		96.1	73.3	65.0	74.6	81.4	68.5	

Table 4: Confusion matrix (all values in %) for land use classification (Albert et al. 2014).

cars from sealed area. Although most of the car pixels in the ground truth were found, as confirmed by a good completeness, most of the classified car pixels actually do not correspond to a car, reflected in a low correctness value. This is amongst other factors caused by the fact that rows of individual cars were merged, incorrectly including sealed area in between. Fig. 5 depicts an example of the classification result, showing that most derived classes closely match the reference.
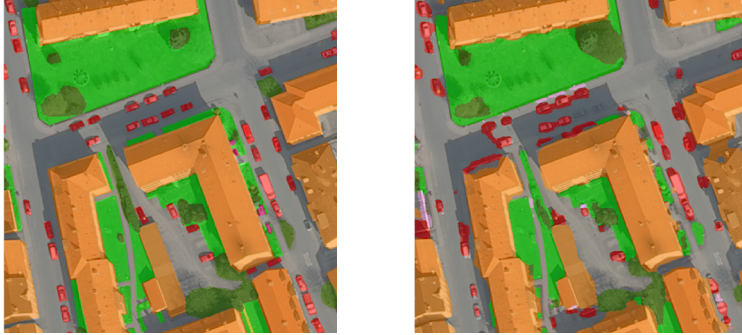

Figure 5: Land cover classification – left: reference, right: CRF result. (Albert et al. 2014).

For the land use classification the CRF approach achieves a mean overall accuracy of 85.5%. The results for the class *residential* are quite good, with completeness and correctness values better than 85%. The completeness value of the class *street* also reached this value, but only about 70% of these objects are correct. This is caused by 4.2% of objects classified as *street* while actually corresponding to the class *residential*. Lower correctness and especially completeness values can be explained by the fact that the amount of training data was not sufficient for an adequate discrimination of the corresponding classes. In total, only about 5% of the objects in the training data belong to the classes *railway*, *water*, *agriculture* and *forest*. Also, the amount of training data for the interaction potential was not sufficient for some classes to allow for a reliable classification. Therefore, in future all the existing ALKIS data rather than only some ALKIS objects should be used to determine the parameters of the various potentials.

4. CONCLUSIONS

This paper demonstrated the possibilities and some limitations of using context sensitive classification methods for various tasks of topographic mapping. Advantages with respect to local approaches could be identified, and it became clear that such methods based on random fields are very flexible and can be adapted to different situations such as point and area entities as input, regular and irregular distribution of entities, as well as multi-temporal and multi-scale classification. A further possibility, not discussed here, consists in using a multi-layer CRF which can e. g. be used to model occlusions (Kosov et al. 2013).

One issue with regard to context sensitive classification is the need for a sufficient amount of training data. This can be a problem, in particular when modelling spatial class transitions, where each combination of adjacent classes is modelled as an extra class. In this case, existing data (as in the case of the land use verification example) can be used. However, these existing data are of course not free of errors, which needs to be taken into account. Transfer learning (Pan, Yang 2010) can be employed to further reduce the amount of required training classes, however this is currently still a topic of intensive research and development.

A general limitation of random fields is the difficulty to incorporate long range context and more explicit object models. For long range context, hierarchical CRFs are a suitable solution, as are higher order potentials (Kohli et al. 2009). Another interesting avenue is the combination of random fields and marked point processes (Chai et al. 2013), which allow to introduce relatively strong knowledge about the shape of objects.

In any case, these probabilistic classification methods hold a lot of promise for further development in the field, as they are inherently able to capture the notion that all observations are uncertain and thus need to be considered in conjunction with appropriate accuracy measures.

5. ACKNOWLEDGEMENTS

6. REFERENCES

Albert L., Rottensteiner F., Heipke C., 2014: Land Use Classification Using Conditional Random Fields for the Verification of Geospatial Databases. ISPRS Annals II-4, 1-7.

Bishop C. M., 2006: Pattern Recognition and Machine Learning. Springer, New York (NY), USA.

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

Bruzzone L., Cossu R., Vernazza G., 2004 : Detection of land-cover transitions by combining multidate classifiers. Pat. Rec. Let. 25 (13), pp. 1491-1500.

Chai D., Förstner W., Lafarge F., 2013: Recovering line-networks in images by junction-point processes. In Proceedings of the IEEE CVPR, pp. 1894-1901.

Fischer A., Kolbe T., Lang F., Cremers A. B., Förstner W., Plümer L., Steinhage V., 1998: Extracting buildings from aerial images using hierarchical aggregation in 2D and 3D, Computer Vision and Image Understanding, 72 (2), 1998, pp. 185-203.

Förstner W., 2013: Graphical models in geodesy and photogrammetry. PFG, 4/2013, pp. 255-268.

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

Hall M. A., 1999: Correlation-based Feature Subset Selection for Machine Learning. PhD thesis, Department of Computer Science, University of Waikato, Hamilton, New Zealand. www.cs.waikato.ac.nz//ml/publications/1999/99MH-Thesis.pdf (visited March 31, 2015).

Heipke C., Pakzad K., Straub B.-M., 2000: Image analysis for GIS data acquisition. Photogrammetric Record, 16 (96), pp. 963-985.

Hoberg T., Rottensteiner F., Feitosa R. Q., Heipke C., 2015: Conditional Random Fields for multitemporal and multiscale classification of optical satellite imagery. IEEE Trans. GRS, 53 (2), pp. 659-673.

Kohli P., Ladický L., Torr P., 2009: Robust higher order potentials for enforcing label consistency. IJCV, 82, pp. 302-324.

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

Kumar S., Hebert, M., 2006: Discriminative Random Fields. IJCV, 68 (2), pp. 179-201.

Leite P. B. C., Feitosa R. Q., Formaggio A. R., Costa G. A. O. P., Pakzad, K., 2011: Hidden Markov models for crop recognition in remote sensing image sequences. Pat. Rec. Let., 32 (1), pp.19-26.

Mayer H., 1999: Automatic object extraction from aerial imagery – a survey focussing on building. Computer Vision and Image Understanding, Vol. 74, No. 2, pp. 138-149.

Niemann H., Sagerer G., Schroeder S., Kummert F., 1990: ERNEST: A semantic network system for pattern understanding. IEEE Trans. PAMI, 12 (9), pp. 883-905.

Niemeyer, J., Rottensteiner, F., Sörgel, U., 2014: Contextual classification of lidar data and building object detection in urban areas. ISPRS JPhRS, 87, pp. 152-165.

Pakzad K., Growe, S., Heipke C., Liedtke C.-E., 2001: Multitemporale Luftbildinterpretation: Strategie und Auswertung. Künstliche Intelligenz, (15) 4, pp. 10-16.

Pan, S. J., Yang, Q., 2010: A survey on transfer learning. IEEE Tans. KDE, 22 (10), pp. 1345-1359.

Radig B. (Hrsg), 1993: Verarbeiten und Verstehen von Bildern. Oldenbourg, München.

Rosenfeld A., 1982: Computer image analysis: an emerging technology in the service of society, Computer Science Technical Reports TR-1177, MCS-79-23422, University of Maryland.

Rottensteiner F., 2008: Automatic extraction of buildings from airborne laser scanner data and aerial images. Habilitation, Institute for Photogrammetry, TU Vienna.

Rottensteiner F., Sohn G., Gerke M., Wegner J. D., Breitkopf U., Jung J., 2014: Results of the ISPRS benchmark on urban object detection and 3D building reconstruction. ISPRS JPhRS, 93, pp. 256-271.

Schindler K., 2012: An overview and comparison of smooth labeling methods for land-cover classification. IEEE Trans. GRS, 50, pp. 4534-4545.

Stilla U, Michaelsen E, Lütjen K., 1995: Structural 3D-analysis of aerial images with a blackboard-based production system. In: Grün A, Kübler O., Agouris P. (Eds.) Automatic extraction of man-made objects from aerial and space images. Birkhäuser, Basel, pp. 53-62.

Vapnik V. N., 1998: Statistical Learning Theory. Wiley, N.Y.

Yang M. Y., Förstner W., 2011: A hierarchical conditional random field model for labeling and classifying images of man-made scenes. ICCV Workshop on CV for RSE, IEEE, pp. 196-203.

Ziems M., 2014: Automatic verification of road databases using multiple road models. DGK-C, No. 724, Beck, München.