Image analysis based on probabilistic models

Christian Heipke

IPI - Institute for Photogrammetry and GeoInformation,
Leibniz Universität Hannover

Special thanks to

Prof. Franz Rottensteiner

Head, IPI Research Group
Photogrammetric image analysis

Lena Albert, Thorsten Hoberg, Joachim Niemeyer



Institut für Photogrammetrie und GeoInformation

Leibniz Universität Hannover

Table of contents

- Introduction
- Strategies in image analysis
- Examples
- (Some) current challenges
- Summary

Our tasks

- **Geometric parameter estimation**
  - Image orientation, 3D point determination (bundle adjustment)
  - Ego motion and obstacle avoidance (navigation, robotics)
  - Object surface reconstruction (DSM, DTM)
- **Image classification (machine learning)**
  - Foreground / background separation
  - Land cover / land use classification
  - Classification of surface patches (planar, quadratic, free form)
- **Interpretation, object extraction and tracking**
  - Scene understanding (“story telling”)
  - Façade interpretation
  - Tracking of moving objects (humans, animals, other targets)

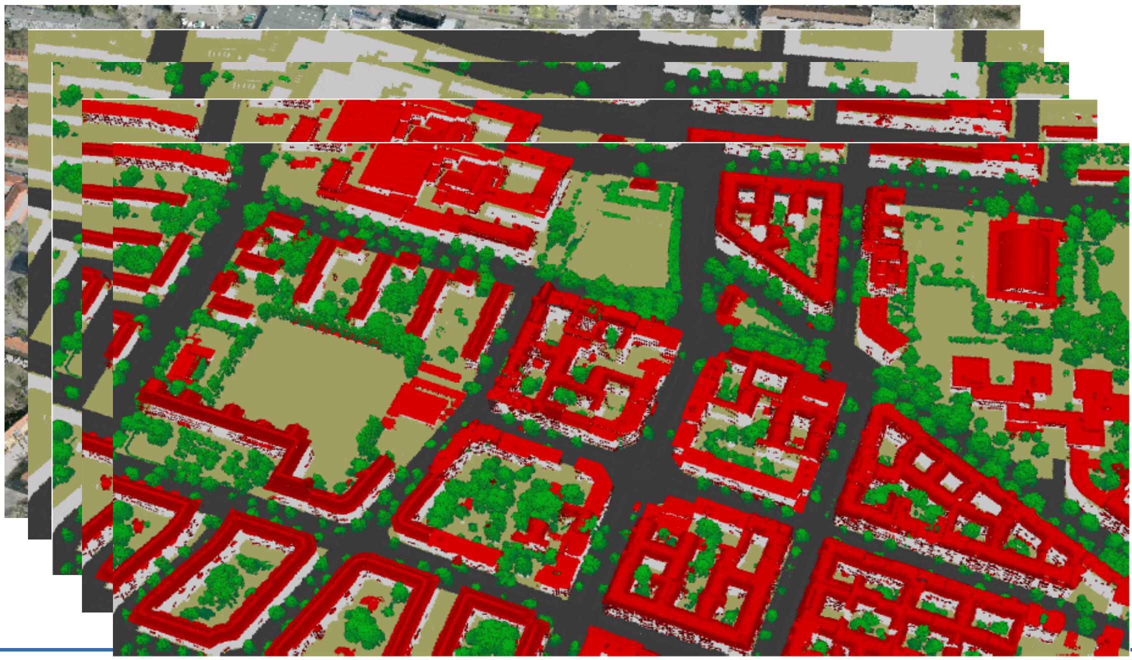
The goal


Source: DGPF

Source: OpenStreetMap

implicit information (aerial image) → explicit information (vector data)

The goal



Strategies in image analysis

Strategies in image analysis

- Automatic recognition of objects in sensor data requires

  knowledge about objects

- Differentiation according to use of knowledge:

Model based approaches

- Knowledge incorporated in the form of explicit object models
- Object classes are treated separately (no/little relations)

Statistical approaches

- Knowledge is learnt from examples
- Image is treated as a whole (incl. context)

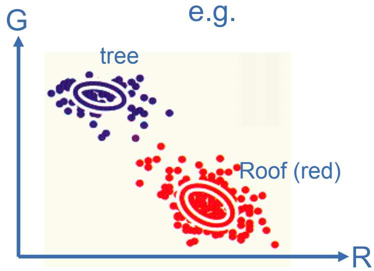
Model based approaches

- Sensor data are interpreted based on pre-defined object models
  - Example: building extraction from DSM
    - ➢ Buildings lie above the terrain (digital terrain model, DTM)
    - ➢ Buildings are characterised by planes (roofs, walls, …)
    - ➢ Buildings have a certain minimum height
  - For more complex tasks: e.g. Knowledge based systems
- Problem:
  - Complexity of pre-defined models
  - Transferability to new scenes

Statistical approaches

- Object knowledge coded in examples → training data
- Analysis based on similarity of features in feature space

e.g.

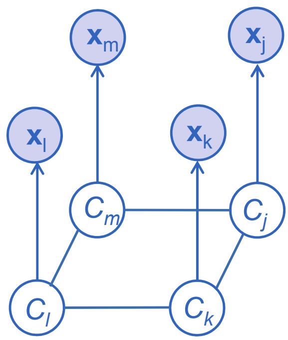


Institut für Photogrammetrie und GeoInformation
Leibniz Universität Hannover

Statistical approaches

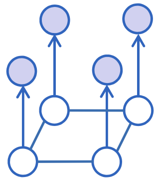
- **Probabilistic approaches:**
  - classification based on probabilities
    - ➢ Maximum Likelihood classification
    - ➢ Logistic regression
    - ➢ Random Fields
- **Non-probabilistic approaches:**
  - classification does not consider probabilities
    - ➢ Support Vector Machines (SVM)
    - ➢ Random Forests (RF)

Markov Random Fields (MRF)



- Motivation: consideration of **context**
- Class labels of neighbouring nodes are not independent -> smoothing
- Dependencies of the classes
  - → Simultaneous classification of all pixels
  - → Determination of a label image **C**, for which $p(\mathbf{C}|\mathbf{x}) \rightarrow$ max
- In general approximate training and classification only
  - Optimisation for graphs of certain structures

Markov Random Fields (MRF)

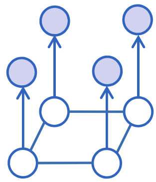


- $p(\mathbf{C}|\mathbf{x})$ can be described by a Gibbs-distribution (Li, 2009):

$$p(\mathbf{C}\,|\,\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \cdot \prod_{i \in n} \varphi_i(C_i, \mathbf{x}_i) \cdot \prod_{i,j \in e} \psi_{ij}(C_i, C_j)$$

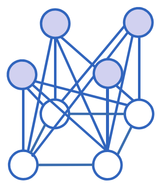
- Z… Partition function, normalisation, depends only on **x**
- data $\mathbf{x}_i$ of individual pixels, considered to be conditionally independent
- $\varphi_i(C_i, \mathbf{x}_i)$: Association potentials
  - relate data and classes
  - e.g. likelihood of a local classifier: $\varphi_i(C_i, \mathbf{x}_i) = p(\mathbf{x}_i \mid C_i)$
- $\psi_{ij}(C_i, C_j)$: Interaction potentials: context model, depends on classes
  - e.g. Potts model: it is most probable that two neighbouring pixels belong to the same class

Conditional Random Fields (CRF)



- Markov Random Fields:

$$p(\mathbf{C}\,|\,\mathbf{x}) = \frac{1}{Z} \cdot \prod_{i \in n} \varphi_i(C_i, \mathbf{x}_i) \cdot \prod_{i,j \in e} \psi_{ij}(C_i, C_j)$$



- Conditional Random Fields (Kumar & Hebert, 2006):

$$p(\mathbf{C}\,|\,\mathbf{x}) = \frac{1}{Z} \cdot \prod_{i \in n} \varphi_i(C_i, \mathbf{x}) \cdot \prod_{i,j \in e} \psi_{ij}(C_i, C_j, \mathbf{x})$$

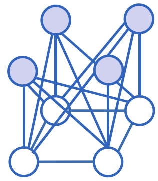
- $\varphi_i(C_i, \mathbf{x})$: Association potentials: data $\mathbf{x}_i$ of individual pixels are **not** considered to be conditionally independent
- $\psi_{ij}(C_i, C_j, \mathbf{x})$: Interaction potentials: context model also depends on data **x**

CRF: Potentials

- **Association potential:**
  - arbitrary local classifier with probabilistic output

    $$\varphi(C_i, \mathbf{x}) = p(C_i \mid \mathbf{f}_i(\mathbf{x}))$$

    with $\mathbf{f}_i(\mathbf{x})$ … feature vector for each node



- **Interaction potentials:**
  - Contrast sensitive Potts-Model: neighbouring pixels probably belong to the same class, if features are similar
    → smoothing depends on difference of features
  - Arbitrary local classifiers:

    $$\psi_{ij}(C_i, C_j, \mathbf{x}) = p(C_i, C_j \mid \mu_{ij}(\mathbf{x}))$$

    with $\mu_{ij}(\mathbf{x})$ … feature difference vector for each node ($\mu_i \neq \mathbf{f}_i$)

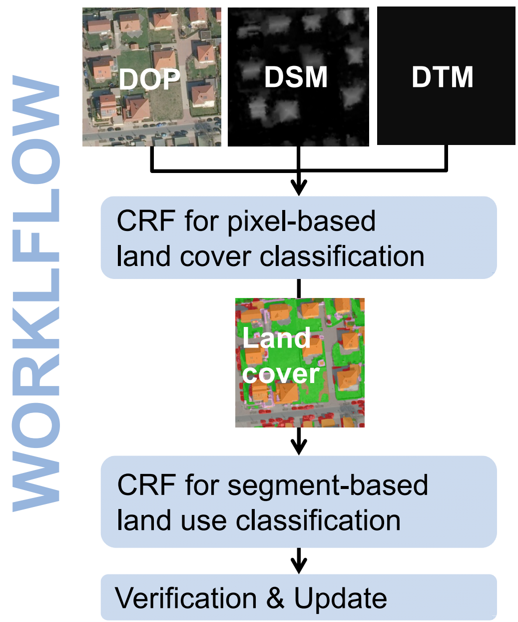
Examples

- High res. LU / LC update
- Road extraction considering occlusions
- Classification of urban LiDAR data
- Multi-temporal/multi-scale classification

chosen to show the flexibility of CRF

High resolution land use / land cover update in urban areas

LU / LC update in urban areas

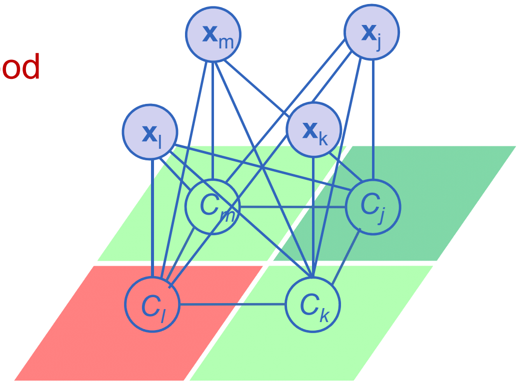


L. Albert et al., ISPRS TC IV 2014

Institut für Photogrammetrie und GeoInformation

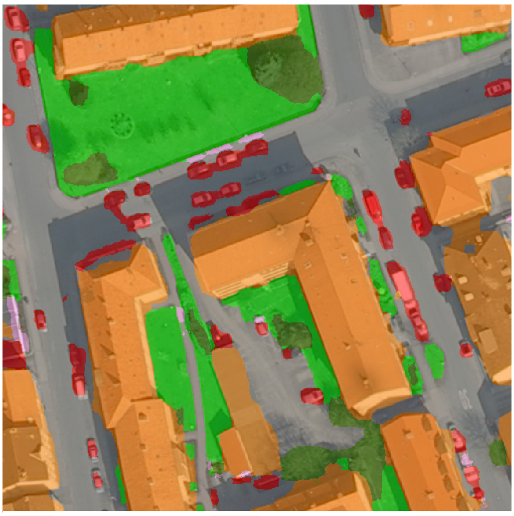
Step 1: Classification of land cover

- **Graphical model:**
  - Nodes: pixels
  - Edges: spatial 4-neighbourhood
- **Features:**
  - Spectral, 3D and texture
- **Association potentials:**
  - Random Forest
- **Interaction potential:**
  - Contrast sensitive Potts model

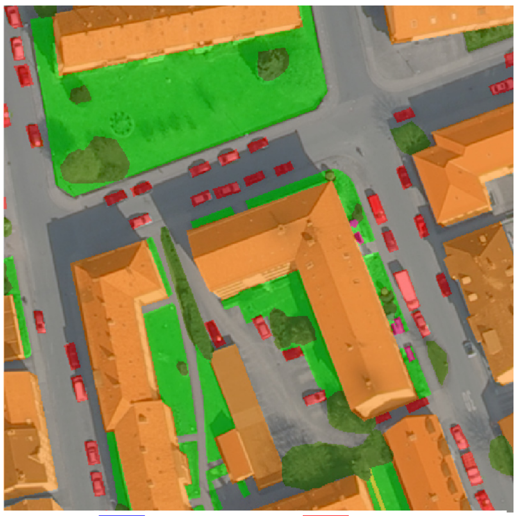


Land cover classification results

Classification result



Reference



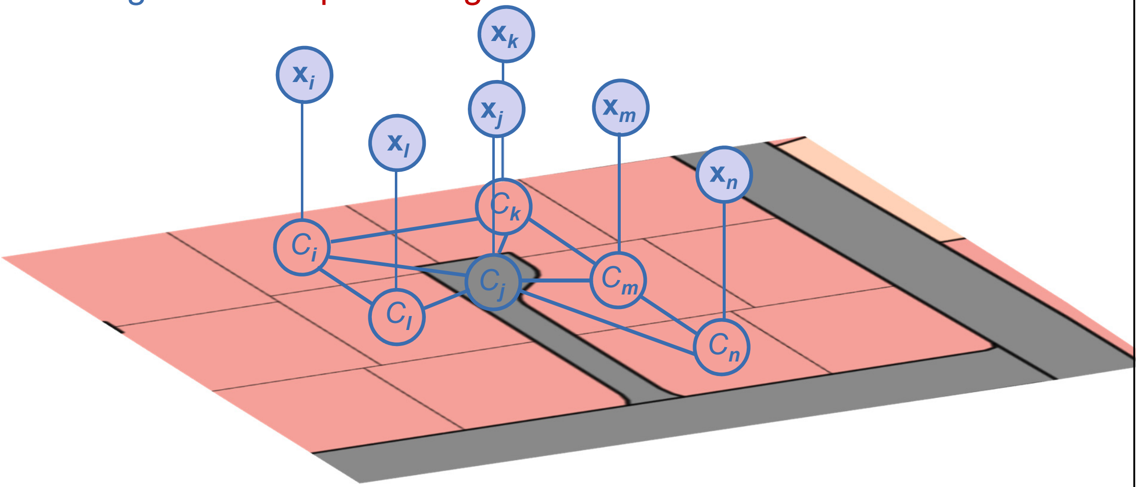
Building	Bare soil	Gras	Water	Car
Sealed s.	Agriculture	Tree	Rails	Other

Overall accuracy: 81,3 %

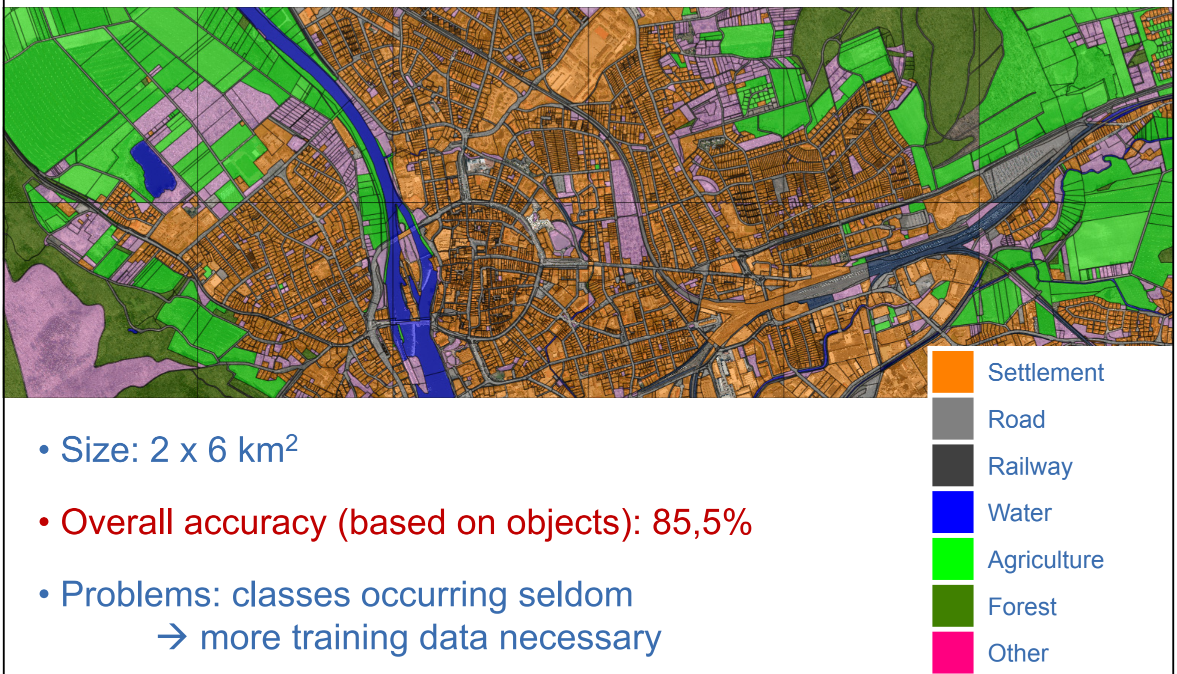
L. Albert et al., 2014

Step 2: Classification of land use

- **Graphical model:**
  - Nodes: segments (GIS-objects) from data base
  - Edges: direct spatial neighbourhood



Classification of land use: Hameln



- Size: 2 x 6 km^2
- Overall accuracy (based on objects): 85,5%
- Problems: classes occurring seldom
  → more training data necessary

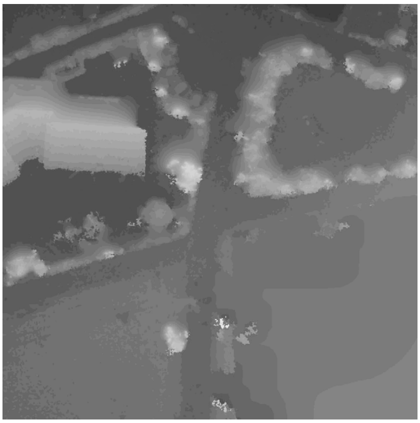
L. Albert et al., 2014

Road extraction considering occlusions

Motivation

- Road extraction: occlusion of road surface

Orthophoto (8 cm)

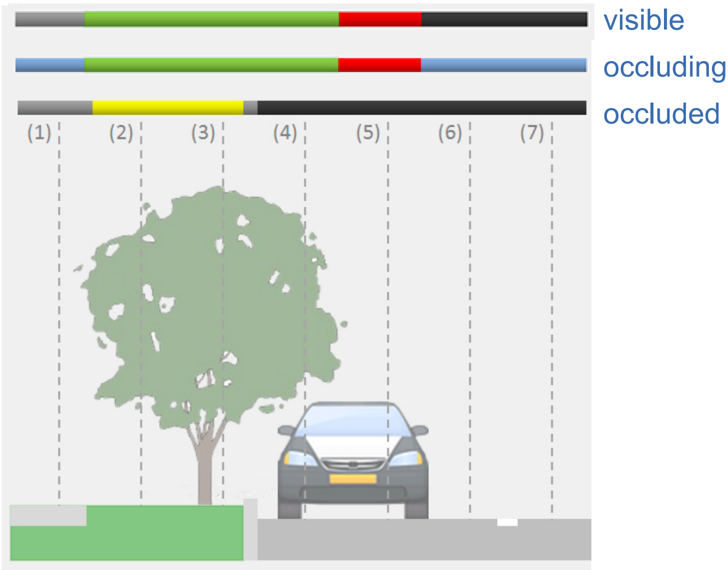


DSM

Institut für Photogrammetrie und GeoInformation

Leibniz Universität Hannover

Vertical scene structure

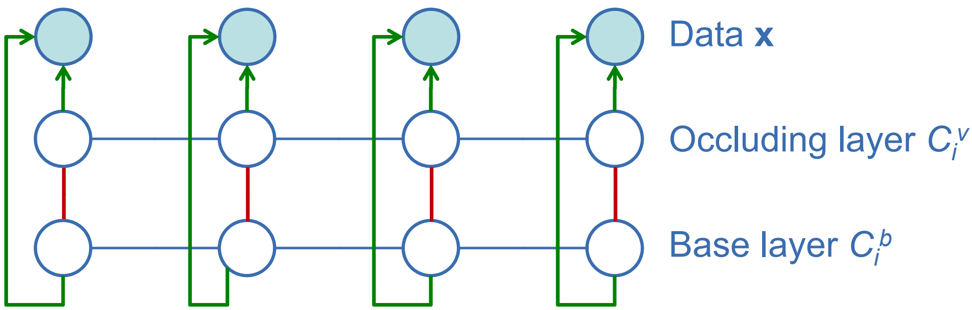


Graphical model

- Two layers with different class structure

1) Base layer C_i^b: can be occluded, does not occlude
   - Example: $C_i^b \in$ { *Sealed surface, Building, Gras, Farmland* }

2) Occluding layer C_i^v: can occlude other objects
   - Example: $C_i^v \in$ { Tree, Car, void }

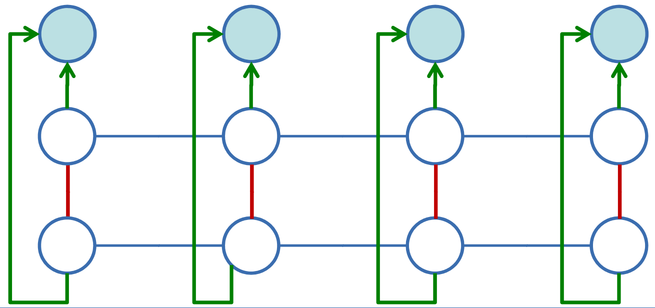


Graphical model

- Model for a posteriori probability:

$$p\left(\mathbf{C}^b,\mathbf{C}^v \mid \mathbf{x}\right)=\frac{1}{Z}\cdot\prod_{i\in n}\boxed{\xi_i\left(C_i^b,C_i^v,\mathbf{x}\right)}\cdot\left[\prod_{l\in\{b,v\}}\varphi_i^l\left(C_i^l,\mathbf{x}\right)\cdot\prod_{i,j\in e^l}\psi_{ij}^l\left(C_i^l,C_j^l,\mathbf{x}\right)\right]$$

- Standard CRF for each of the two layers
- new interaction potential $\xi_i(C_i^b, C_i^o, \mathbf{x})$ connecting layers



Results

- Simple interaction model (Kosov et al., 2013)

CRF: base, occlusion l.

Two layer model: base l., occlusion l.

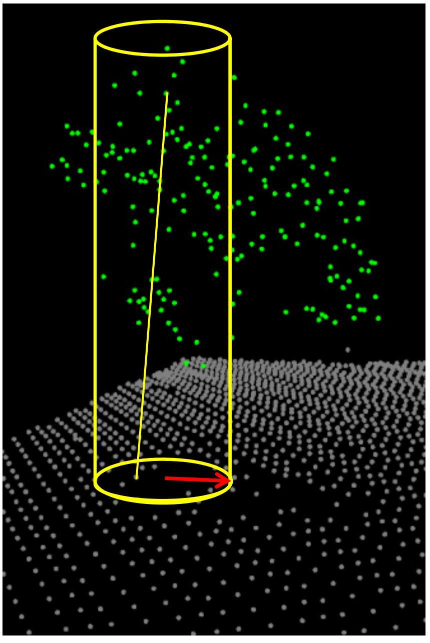
- Overall accuracy: Base layer: 85,6% (+14%)

  Occluding layer: 86,0% (±0%)

Classification of urban LiDAR data

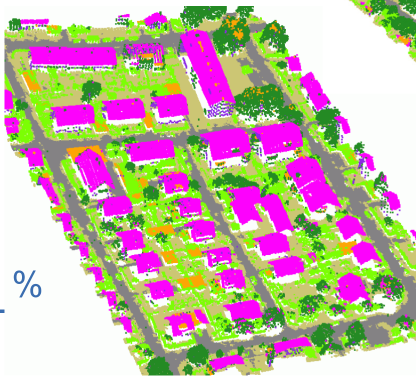
CRF for airborne LiDAR data

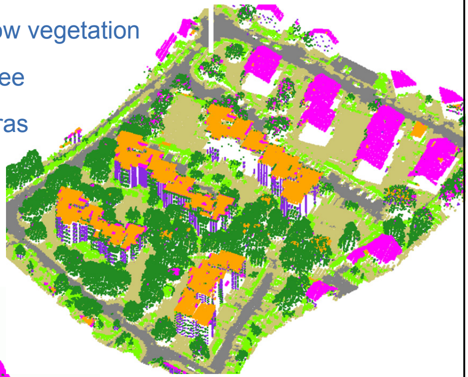
- Point wise classification (Niemeyer et al., 2014)
  - Nodes: 3D points
  - Features: extracted from local point neighbourhood
  - Edges: Nearest neighbourhood in 2D → cylinder
  - Potentials: Random Forests for association and interaction potentials



Results

• Test data „Vaihingen“ of ISPRS (Rottensteiner et al., 2014)





Overall accuracy: 83,4 %

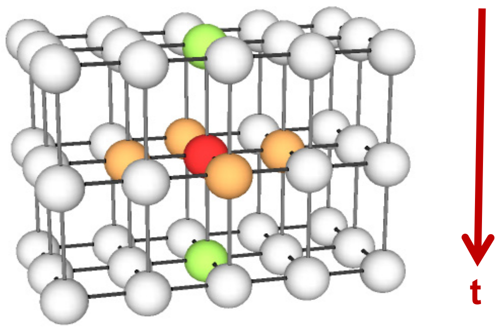
Multi-temporal / multi-scale classification of land cover

Multi-temp./multi-scale class. of land cover

- Goal: land cover classification, incorporating temporal knowledge
  - multi-scale model
  - can handle different sensors
- Data: Multi-spectral satellite images of different epochs
  - variable ground sampling distance: e.g. 4 m – 30 m
  - classes: *Settlement (S)*, *Industry* (I), *Forest* (*W*), *Agriculture* (*A*)
- Simultaneous classification of data of all epochs using CRF

CRF for multi-tem./multi-scale classification

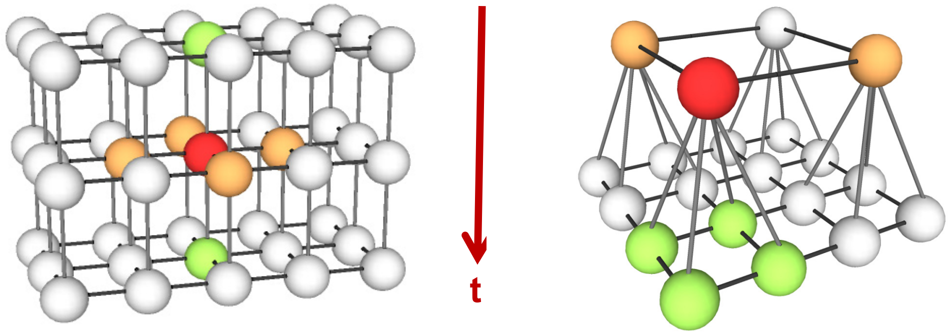
- Nodes: Pixel of images of all epochs
- Edges: direct spatial and temporal neighbourhood
- Graph structure:



for images with equal ground sampling distance

CRF for multi-tem./multi-scale classification

- Nodes: Pixel of images of all epochs
- Edges: direct spatial and temporal neighbourhood
- Graph structure:



for images with equal ground sampling distance

image with different ground sampling distance

Extended CRF model

- Standard CRF

$$p(\mathbf{C}\,|\,\mathbf{x}) = \frac{1}{Z} \cdot \prod_{i \in n} \varphi_i(C_i, \mathbf{x}) \cdot \prod_{i,j \in e} \psi_{ij}(C_i, C_j, \mathbf{x})$$

- Model for multi-temporal classification:

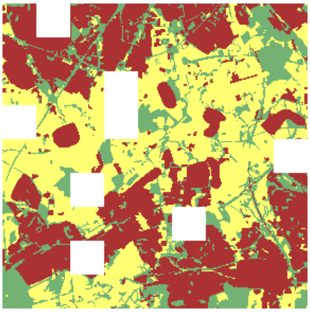
$$p(\mathbf{C}\,|\,\mathbf{x}) = \frac{1}{Z} \cdot \prod_t \left[\prod_{i \in n} \varphi_i^t(C_i^t, \mathbf{x}^t) \cdot \prod_{i,j \in e_s} \psi_{ij}^t(C_i^t, C_j^t, \mathbf{x}^t) \cdot \boxed{\prod_{i,l \in e_t} \tau^{tk}(C_i^t, C_l^k)} \right]$$

  - association potential: Normal distribution
  - ψ: spatial interaction potential: contrast sensitive Potts model
  - temporal interaction potential τ: transition probability $p(C_i^t \mid C_i^{t-1})$

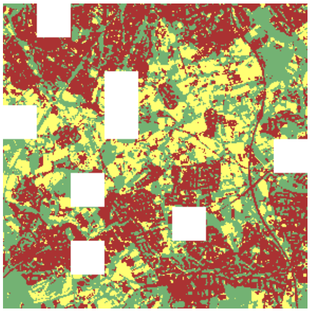
Institut für Photogrammetrie und GeoInformation
Leibniz Universität Hannover

Multi-temporal classification: results

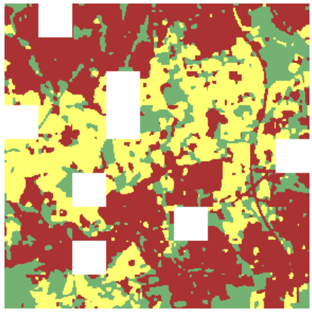
- Landsat 2010:



Reference



ML



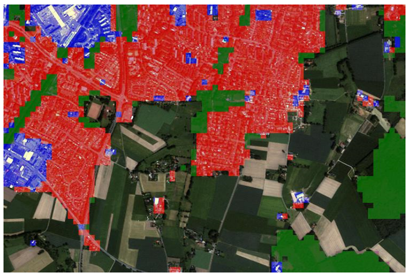
CRF multitemp

	Overall accuracy
Ikonos 2005 (4 m)	81.5 %
RapidEye 2009 (5 m)	80.6 %
Landsat 2010 (30 m)	80.4 %

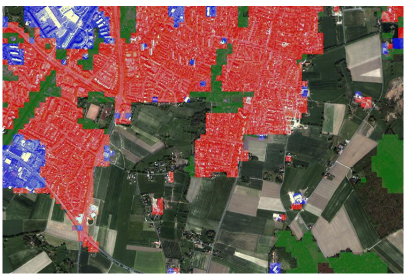
- allows also for detection of changes

Results – CRF multi-temporal

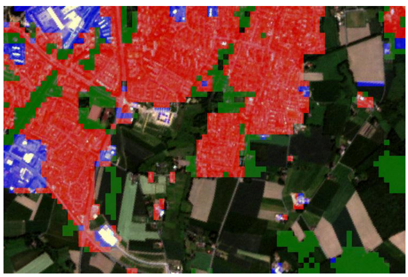
reference



Ikonos 2005 (81,5%)



Ikonos 2007 (80,4%)

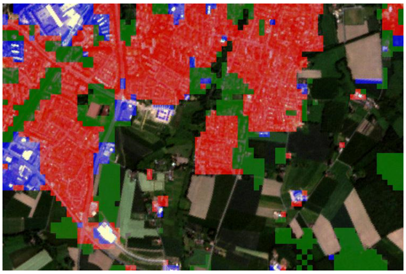


RapidEye 2009 (80,6%)

Results – RapidEye 2009

reference

ML (64,4%)



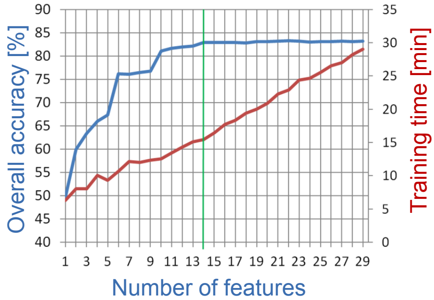
CRF mono (66,2%)

CRF multi (80,6%)

(Some) current challenges

Feature definition

- Success of classification depends on proper feature selection
- Few, but expressive features are often sufficient (e.g. Niemeyer et al., 2014):



… but which ones are best?

→ Approaches for automatic feature selection

→ Feature learning, e.g. „Convolutional Neural Networks“

Global context

- Interaction term of MRF and CRF are very local
  - local error clusters can't be avoided
  - more global interactions needed

- Approaches for modelling interactions

  across larger distances

  - Hierarchical models
  - Two-step procedure (pixels → segments)
  - Higher order potentials

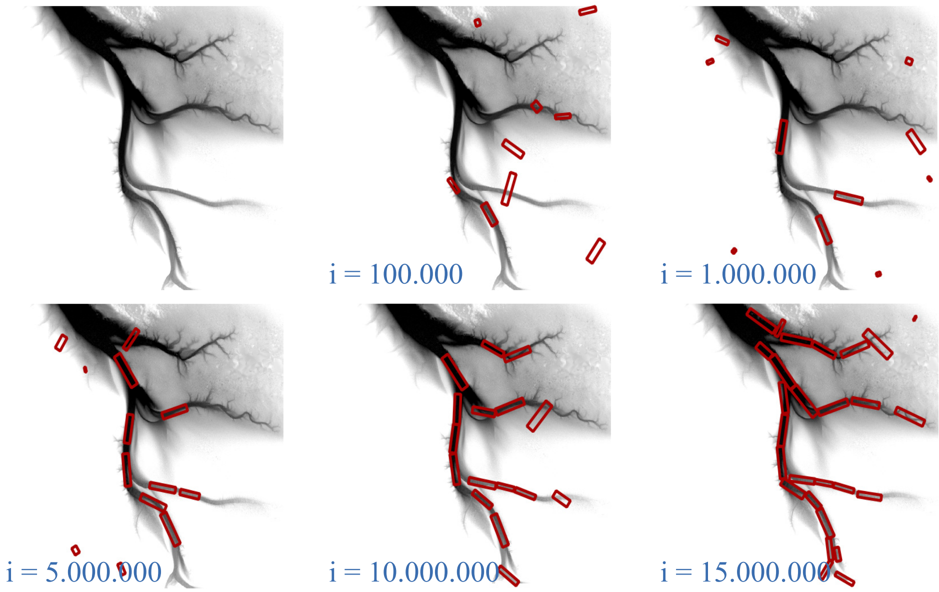
Integration of model knowledge

- Deterministic model knowledge cannot be integrated into CRF
  - e.g. knowledge about object form
- Stochastic approach with deterministic models:

  Marked Point Processes (Ortner et al., 2007)

Integration of model knowledge

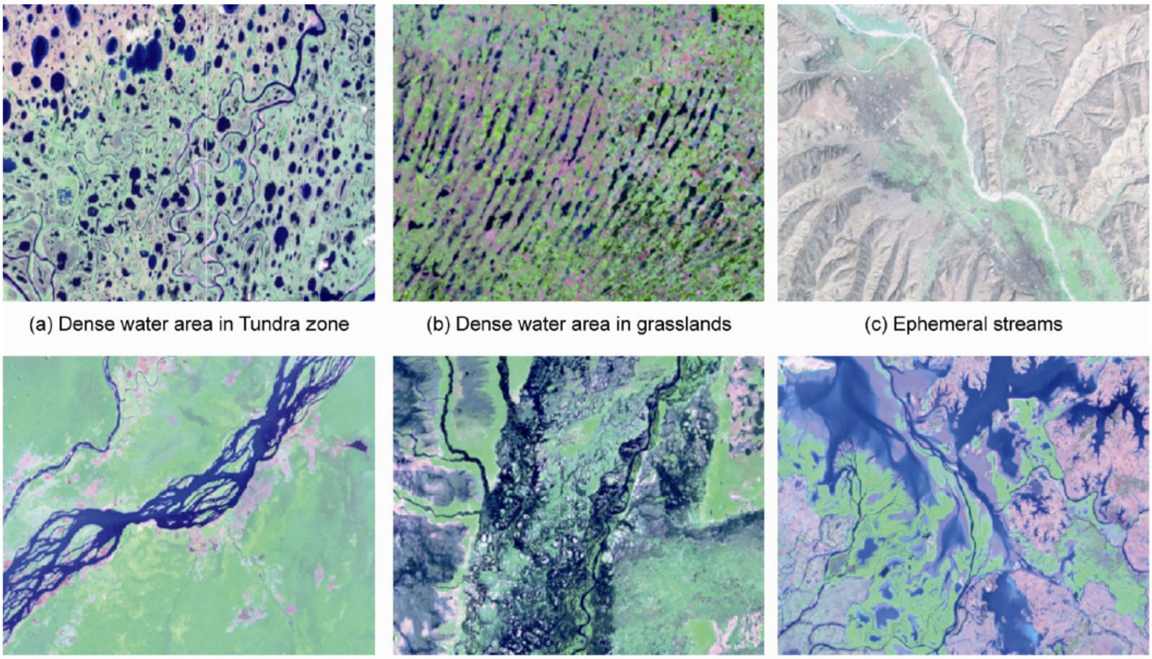
- Example: detection of water ways in laser scanner data (A. Schmidt)



Institut für Photogrammetrie und GeoInformation
Leibniz Universität Hannover

Training: Transfer learning and label noise

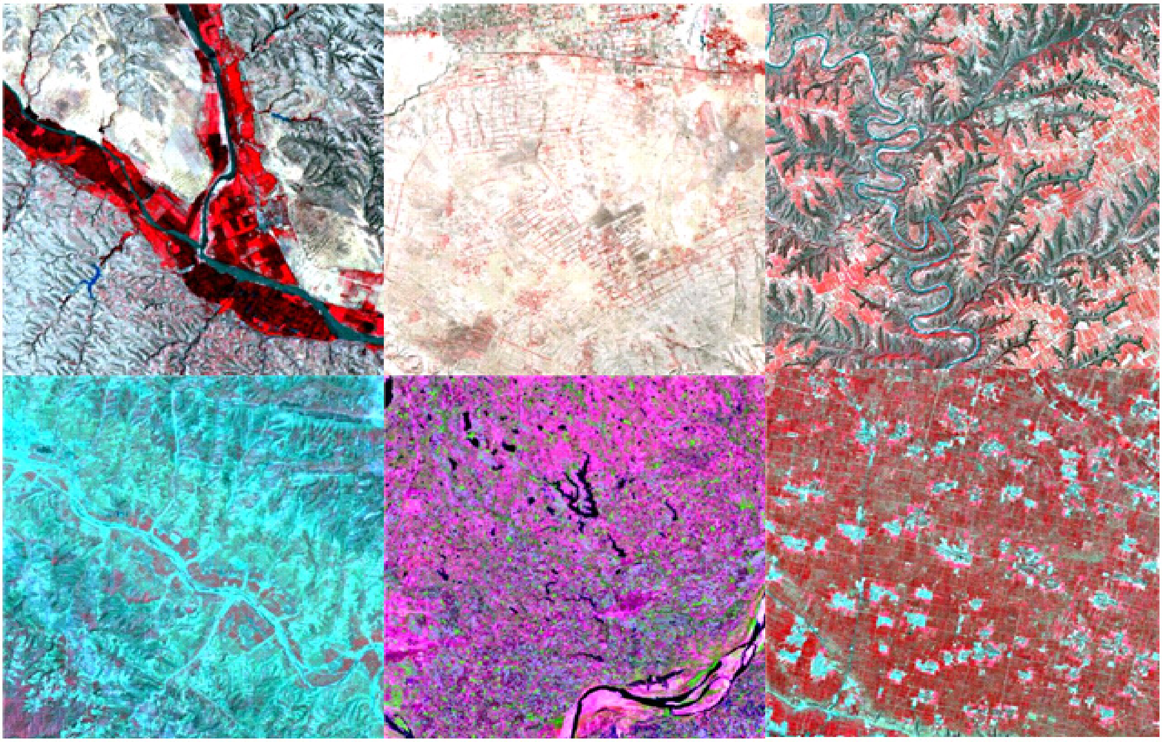
- **Problem:** Acquisition of sufficient amounts of representative training data
  - \+ Training allows simple adaptation to sensor data
  - – Acquisition of training data is time and cost intensive
  - – Training data are not necessarily without error
- Transfer learning: Adaptation of a learnt classifier to data of a different distribution (Pang & Yang, 2010)
  - – With (few) or without any new training data
- Classification under label noise: Model for errors in training data for which parameters are also being estimated (Frénay & Verleysen, 2014)



(a) Dense water area in Tundra zone
(b) Dense water area in grasslands
(c) Ephemeral streams
(d) Braided drainage
(e) Flooded river
(f) Ephemeral lakes

Transfer learning for water areas?

Liao et al., 2014



Transfer learning for crop land areas?

Chen Jun, 2012

Transfer learning for road extraction?

M. Ziems, 2014

Summary

Summary

- Probabilistic models (MRF / CRF – also Bayes networks)
  - Considerable improvement of results through context
  - Very flexible tool for image analysis (and beyond …)
    - ➢ image/lidar data – pixels/segments
    - ➢ occlusion modelling
    - ➢ multi-temporal/multi-scale data
    - ➢ (object tracking, moving objects, …)
- Focus of this talk on models for classification:
  - Graph structure and modelling of probability density functions
  - Actual 3D reconstruction not discussed