



SUPERVISED AND UNSUPERVISED MRF BASED 3D
SCENE CLASSIFICATION IN MULTIPLE VIEW
AIRBORNE OBLIQUE IMAGES

Markus Gerke and Jing Xiao
(University of Twente, NL and Wuhan University, CN)



CONTENTS

- Introduction
- Classification in 3D object space
 - Basic idea
 - Supervised approach
 - Unsupervised approach
- Experiments: data and results
- Discussion

INTRODUCTION

MULTIPLE VIEW AIRBORNE OBLIQUE IMAGERY

- State-of-the-Art digital camera hardware and processing stimulate development of multi-head camera systems:
 - Midas Track'Air
 - IGI PentaCam
 - Hexacon/Leica RCD30 oblique
 - Microsoft Osprey
- Also known through
 - Pictometry/Bing Maps
 - BlomOblique
 -



Colored mesh of a 3D point cloud, derived from dense image matching

INTRODUCTION

MULTIPLE VIEW AIRBORNE OBLIQUE IMAGERY

- Main advantage: observability of lateral structures
- Main disadvantage: occlusions → multi view and high overlap
- Primary usage
 - automatic model texturing
 - manual interpretation
 - surface modeling (dense matching)
- Automatic scene interpretation:
Interesting, but not so much work yet done



Colored mesh of a 3D point cloud, derived from dense image matching

INTRODUCTION

MULTIPLE VIEW AIRBORNE OBLIQUE IMAGERY

- Main advantage: observability of lateral structures
- Main disadvantage: occlusions → multi view and high overlap
- Primary usage
 - automatic model texturing
 - manual interpretation
 - surface modeling (dense matching)
- Automatic scene interpretation:
Interesting, but not so much work yet done

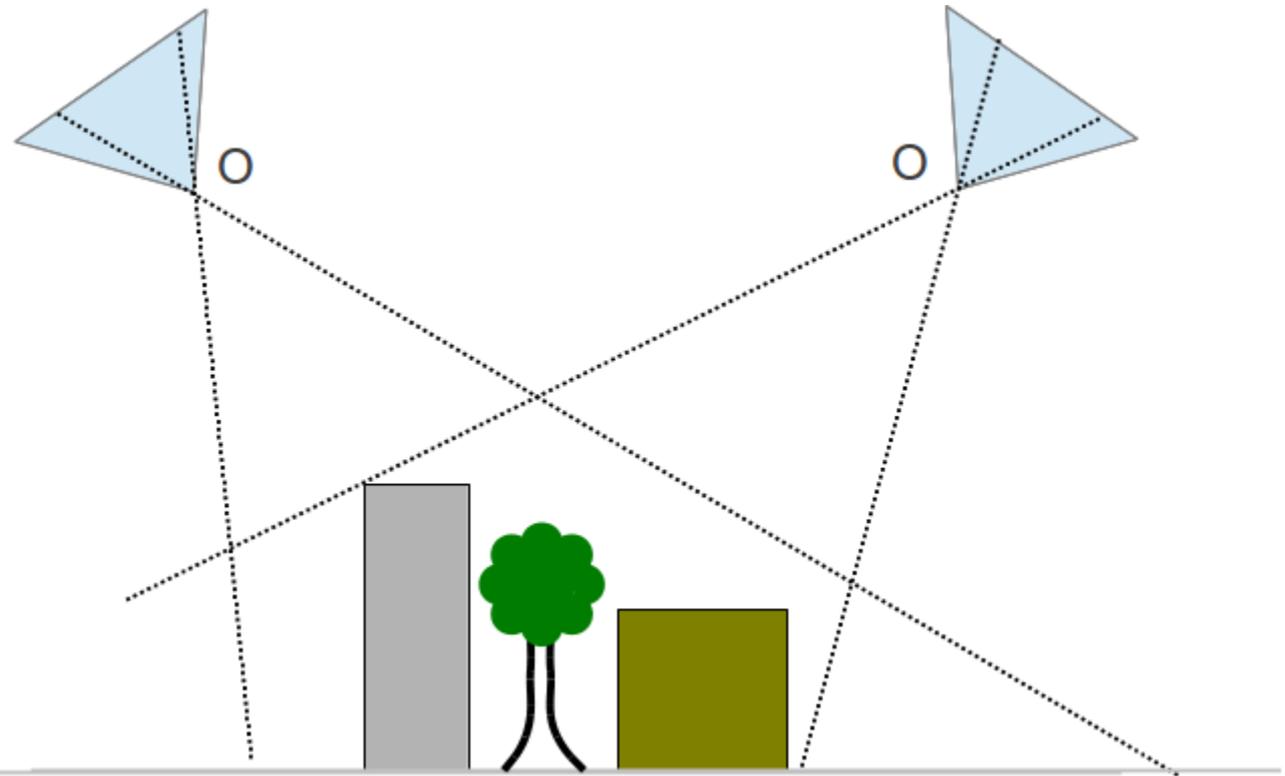


Colored mesh of a 3D point cloud, derived from dense image matching

Classical single image interpretation: 3D neighborhood and object properties not easily exploitable.

3D SCENE CLASSIFICATION

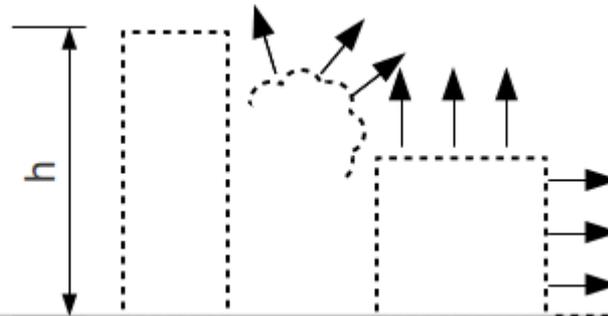
BASIC IDEA



Urban scene, observed from different angles

3D SCENE CLASSIFICATION

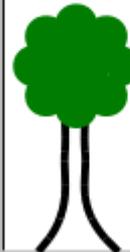
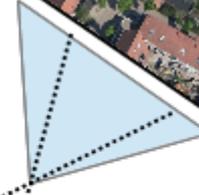
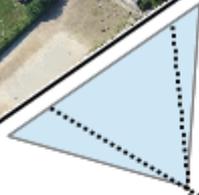
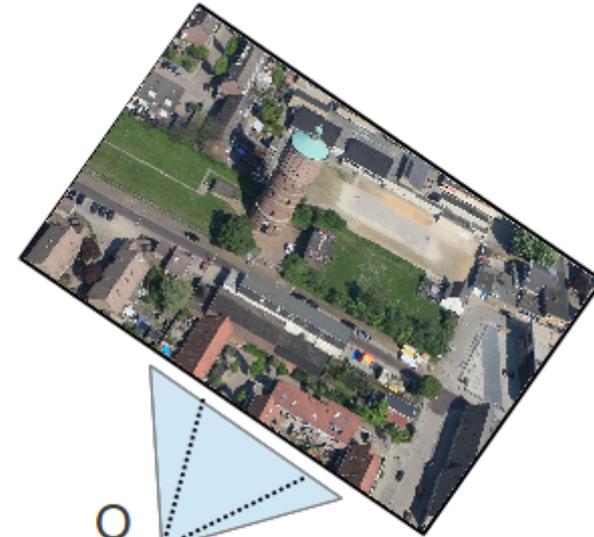
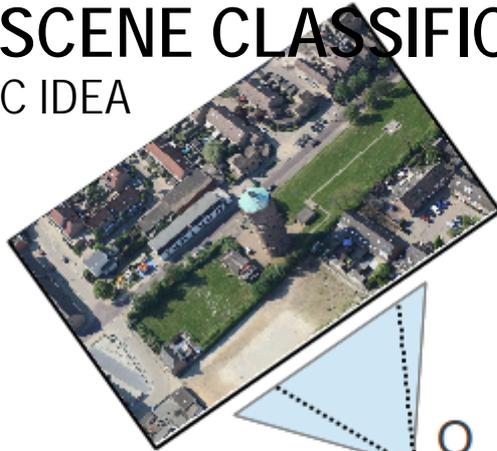
BASIC IDEA



Compute point cloud from images and derive 3D features:
normals (z component, residual), height above ground

3D SCENE CLASSIFICATION

BASIC IDEA



From images also compute features and assign to points
(need for visibility check)

3D SCENE CLASSIFICATION

BASIC IDEA



Color features, like Excess of Green Index

3D SCENE CLASSIFICATION

BASIC IDEA



Straight lines

3D SCENE CLASSIFICATION

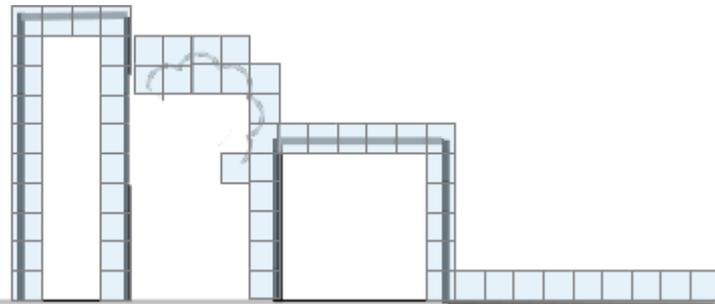
BASIC IDEA



Standard deviation in a sliding window

3D SCENE CLASSIFICATION

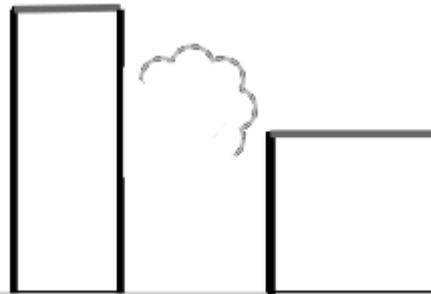
BASIC IDEA



Spatial enumeration (voxelisation)

3D SCENE CLASSIFICATION

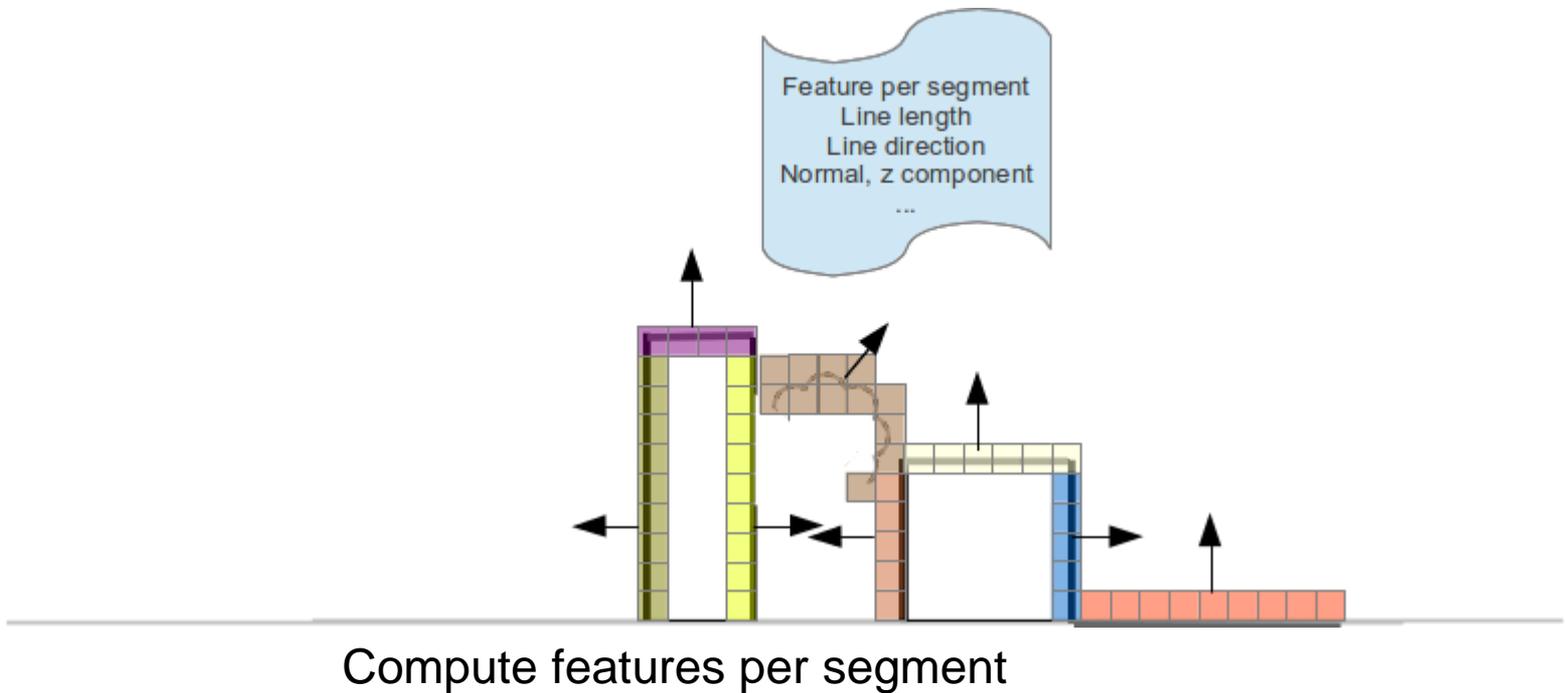
BASIC IDEA



Segment the point cloud: planarity and connected components → entity to be classified

3D SCENE CLASSIFICATION

BASIC IDEA

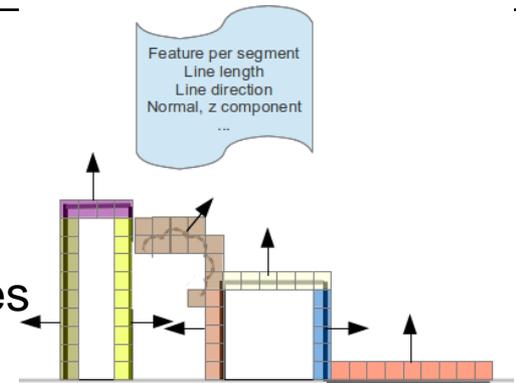


3D SCENE CLASSIFICATION

BASIC IDEA

So we

- combine 3D and image features
- exploit 3D point cloud for scene segmentation, entities defined in 3D object space
- and finally will use standard methods for classification
 - Random Trees (supervised)
 - MRF-approach, using graph-cuts for energy minimization (unsupervised)





3D SCENE CLASSIFICATION

SUPERVISED APPROACH

- Reference labeling in images, project labels into 3D space and assign to segments
- Sample from the reference segments some 20% for the training (Random Forest, opencv implementation)
- Evaluation of result using the remaining reference labels
- Repeat to check for overfitting
- → the final label information per segment is transferred to voxels.

Classes defined: *roof, tree, facade, vegetated ground, sealed ground and roof destroyed/rubble* (for damage mapping application case)



3D SCENE CLASSIFICATION

UNSUPERVISED APPROACH

- Using graph cuts in the 3D lattice (voxels)
- Total energy to be minimized $E = \sum_{p \in P} D_p(f_p) + \sum_{(p,q) \in N} V_{pq}(f_p, f_q)$

with $D_p(f_p)$: Data energy at voxel p for label f ,

and V_{pq} : pairwise interaction potential (Potts here: label smoothness),
and N : local neighborhood, here 3x3x3

- $D_p(f_p)$ is composed out of factors, derived from the features. Example:
factor from local normal, z component (for the others see paper)

$$S_Z = \begin{cases} m_Z, & \text{if } f_p = \text{façade} \\ 1 - m_Z, & \text{if } f_p = \text{sealed_grnd} \\ \min(1 - m_Z, |0.5 - m_Z|), & \text{if } f_p = \text{roof} \\ \min(m_Z, 1 - m_Z, |0.5 - m_Z|) + C, & \text{else} \end{cases}$$

EXPERIMENTS

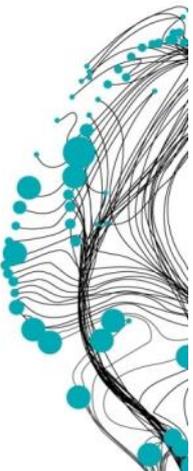
- Here only dataset Enschede. For Haiti damage mapping refer to paper
- Flight by Slagboom en Peeters BV
 - Altitude 400m, overlap > 60% in all directions
 - 5 Canon MarkII cameras (36mm x 24mm), mounted in a custom head
 - Oblique views: 45deg tilt, GSD 5-8 cm



Pictometry

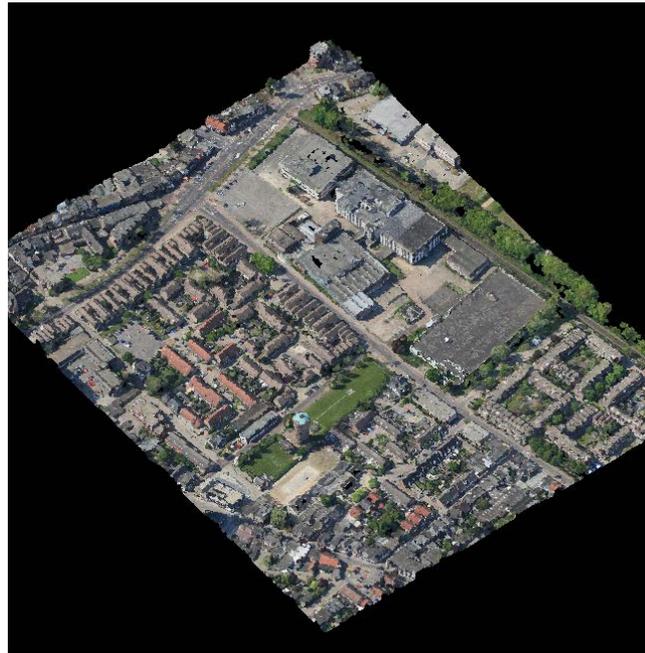
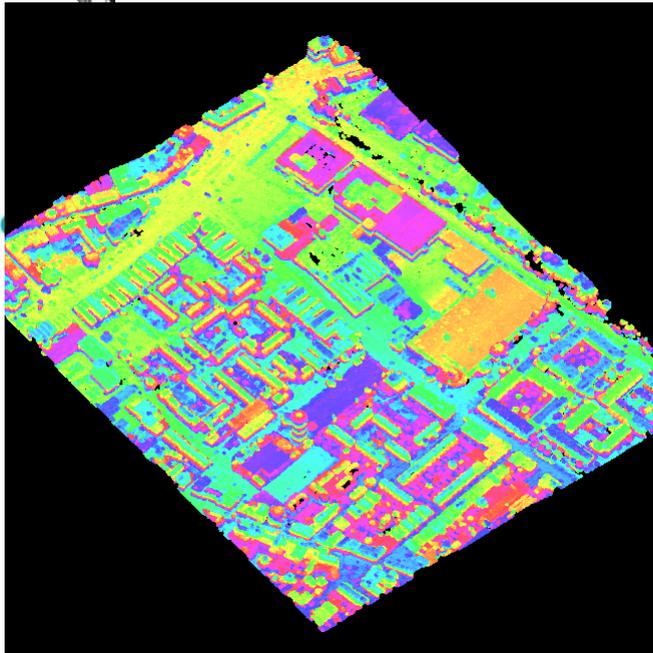


Slagboom en
Peters



EXPERIMENTS

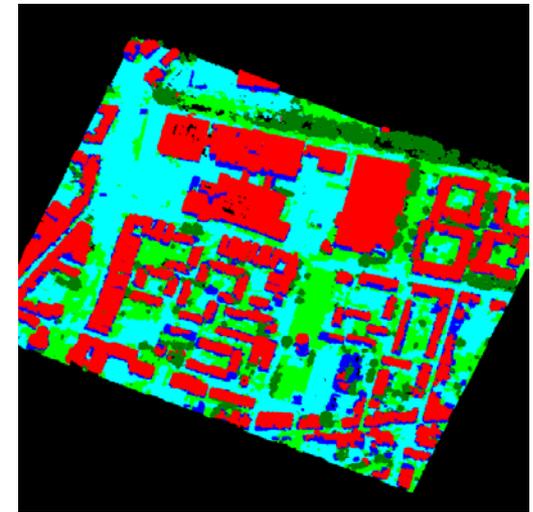
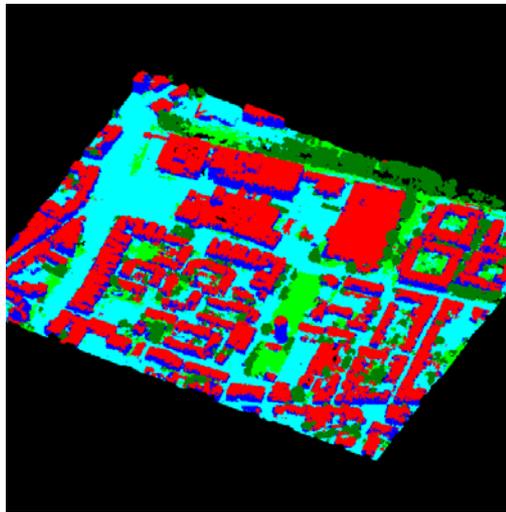
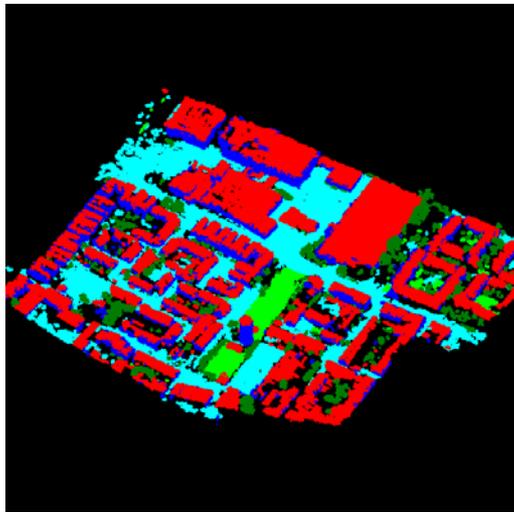
- Dense point cloud matching using PMVS2 (Furukawa and Ponce, 2011)



Test area: mixed settlement and industry, small green areas

EXPERIMENTS

Example extraction



Tree Roof Facade Veg Ground Sealed Ground

Example: reference labels, extraction Random Trees,
extraction Graph Cut



EXPERIMENTS

Confusion matrices: Random Trees, overall accuracy 84.7%

REF→	Facade	Roof	Seal_Grd	Veg_Grd	Tree
Facade	0.806	0.122	0.054	0.002	0.015
Roof	0.075	0.876	0.036	0.001	0.013
Seal_Grd	0.032	0.009	0.905	0.043	0.012
Veg_Grd	0.000	0.000	0.069	0.862	0.069
Tree	0.019	0.019	0.049	0.143	0.771

Confusion matrices: Graph Cut, overall accuracy 78.3%

REF→	Facade	Roof	Seal_Grd	Veg_Grd	Tree
Facade	0.628	0.231	0.073	0.051	0.018
Roof	0.066	0.838	0.071	0.006	0.020
Seal_Grd	0.002	0.001	0.800	0.196	0.000
Veg_Grd	0.000	0.000	0.029	0.968	0.003
Tree	0.012	0.045	0.037	0.303	0.602



EXPERIMENTS

Confusion matrices: Random Trees, overall accuracy 84.7%

REF→	Facade	Roof	Seal_Grd	Veg_Grd	Tree
Facade	0.806	0.122	0.054	0.002	0.015
Roof	0.075	0.876	0.036	0.001	0.013
Seal_Grd	0.032	0.009	0.905	0.043	0.012
Veg_Grd	0.000	0.000	0.069	0.862	0.069
Tree	0.019	0.019	0.049	0.143	0.771

Confusion matrices: Graph Cut, overall accuracy 78.3%

REF→	Facade	Roof	Seal_Grd	Veg_Grd	Tree
Facade	0.628	0.231	0.073	0.051	0.018
Roof	0.066	0.838	0.071	0.006	0.020
Seal_Grd	0.002	0.001	0.800	0.196	0.000
Veg_Grd	0.000	0.000	0.029	0.968	0.003
Tree	0.012	0.045	0.037	0.303	0.602

EXPERIMENTS

Example Façade/roof confusion



Main problem: structures on the façade (balconies, sun blinds, small roofs above entry) not modeled (façade in reference, but roof in extraction)



EXPERIMENTS

Confusion matrices: Random Trees, overall accuracy 84.7%

REF→	Facade	Roof	Seal_Grd	Veg_Grd	Tree
Facade	0.806	0.122	0.054	0.002	0.015
Roof	0.075	0.876	0.036	0.001	0.013
Seal_Grd	0.032	0.009	0.905	0.043	0.012
Veg_Grd	0.000	0.000	0.069	0.862	0.069
Tree	0.019	0.019	0.049	0.143	0.771

Confusion matrices: Graph Cut, overall accuracy 78.3%

REF→	Facade	Roof	Seal_Grd	Veg_Grd	Tree
Facade	0.628	0.231	0.073	0.051	0.018
Roof	0.066	0.838	0.071	0.006	0.020
Seal_Grd	0.002	0.001	0.800	0.196	0.000
Veg_Grd	0.000	0.000	0.029	0.968	0.003
Tree	0.012	0.045	0.037	0.303	0.602

EXPERIMENTS

Example Tree/ground vegetation, especially in unsupervised method



Main problem: low vegetation (bush, shrub): is it ground vegetation or tree?



EXPERIMENTS

Confusion matrices: Random Trees, overall accuracy 84.7%

REF→	Facade	Roof	Seal_Grd	Veg_Grd	Tree
Facade	0.806	0.122	0.054	0.002	0.015
Roof	0.075	0.876	0.036	0.001	0.013
Seal_Grd	0.032	0.009	0.905	0.043	0.012
Veg_Grd	0.000	0.000	0.069	0.862	0.069
Tree	0.019	0.019	0.049	0.143	0.771

Confusion matrices: Graph Cut, overall accuracy 78.3%

REF→	Facade	Roof	Seal_Grd	Veg_Grd	Tree
Facade	0.628	0.231	0.073	0.051	0.018
Roof	0.066	0.838	0.071	0.006	0.020
Seal_Grd	0.002	0.001	0.800	0.196	0.000
Veg_Grd	0.000	0.000	0.029	0.968	0.003
Tree	0.012	0.045	0.037	0.303	0.602

EXPERIMENTS

Example sealed/vegetated ground confusion, especially in unsupervised



Main problem: color information in shadow areas misleading



DISCUSSION

- Both approaches perform similar, specific problems
 - Not modeled objects
 - Confusion vegetation (ground vs tree)
 - Shadow areas (sealed vs. vegetation)
- The two latter ones more serious in unsupervised approach
- Noise and gaps in point cloud another problem (especially at facades)

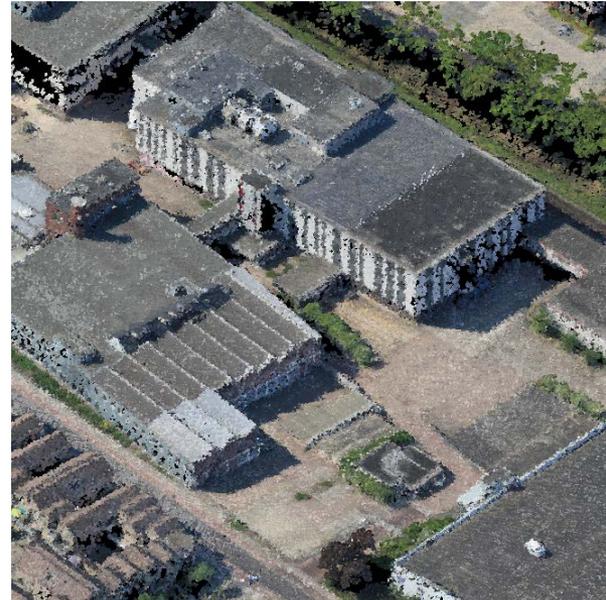


OUTLOOK

- New midformat systems come with NIR, → better for vegetation detection
- Still noise in stereo-views, thus need to filter
- Need to model structures at facades

OUTLOOK

- New midformat systems come with NIR, → better for vegetation detection
- Still noise in stereo-views, thus need to filter
- Need to model structures at facades
- First tests with SURE (SGM) done, typical result: denser, but more noise especially in stereo views → more sophisticated denoising necessary





Thank you for your attention

