Segmentation of RGB-D Indoor Scenes
by stacking Random Forests and Conditional Random Fields

June 24, 2015

Mikkel Thøgersen

Department of Electronic Systems
Aalborg University
Denmark

AALBORG UNIVERSITY
DENMARK

Agenda

Motivation
Introduction

[1] http://asimo.honda.com/ASIMO_DCTM/News/images/highres/Meet_ASIMO.jpg
[2] http://www.toyota-global.com/innovation/partner_robot/images/family_img01.jpg
[3] https://www.willowgarage.com/sites/default/files/images/pr2Image.png

Motivation

Introduction

- The proportion of elderly people in developed countries

[1] http://asimo.honda.com/ASIMO_DCTM/News/images/highres/Meet_ASIMO.jpg
[2] http://www.toyota-global.com/innovation/partner_robot/images/family_img01.jpg
[3] https://www.willowgarage.com/sites/default/files/images/pr2Image.png

Motivation
Introduction

- The proportion of elderly people in developed countries
- Health care robots
- Service robots

1

2

3

[1] http://asimo.honda.com/ASIMO_DCTM/News/images/highres/Meet_ASIMO.jpg
[2] http://www.toyota-global.com/innovation/partner_robot/images/family_img01.jpg
[3] https://www.willowgarage.com/sites/default/files/images/pr2Image.png

Motivation
Introduction

- The proportion of elderly people in developed countries
- Health care robots
- Service robots

Challenges:

1

2

3

[1] http://asimo.honda.com/ASIMO_DCTM/News/images/highres/Meet_ASIMO.jpg
[2] http://www.toyota-global.com/innovation/partner_robot/images/family_img01.jpg
[3] https://www.willowgarage.com/sites/default/files/images/pr2Image.png

Motivation
Introduction

- The proportion of elderly people in developed countries
- Health care robots
- Service robots

Challenges:

- Hardware

1

2

3

[1] http://asimo.honda.com/ASIMO_DCTM/News/images/highres/Meet_ASIMO.jpg
[2] http://www.toyota-global.com/innovation/partner_robot/images/family_img01.jpg
[3] https://www.willowgarage.com/sites/default/files/images/pr2Image.png

Motivation
Introduction

- The proportion of elderly people in developed countries
- Health care robots
- Service robots

Challenges:

- Hardware
- Sensing/planning/executing

1

2

3

[1] http://asimo.honda.com/ASIMO_DCTM/News/images/highres/Meet_ASIMO.jpg
[2] http://www.toyota-global.com/innovation/partner_robot/images/family_img01.jpg
[3] https://www.willowgarage.com/sites/default/files/images/pr2Image.png

Motivation
Introduction

- The proportion of elderly people in developed countries
- Health care robots
- Service robots

Challenges:

- Hardware
- Sensing/planning/executing

Sensing the surroundings

1

2

3

[1] http://asimo.honda.com/ASIMO_DCTM/News/images/highres/Meet_ASIMO.jpg
[2] http://www.toyota-global.com/innovation/partner_robot/images/family_img01.jpg
[3] https://www.willowgarage.com/sites/default/files/images/pr2Image.png

Motivation
Introduction

The aim of the project is to create a vision based system, capable of semantically categorizing objects in indoor cluttered scenes using RGB-D cameras and computer vision techniques.

Semantic Segmentation
Introduction

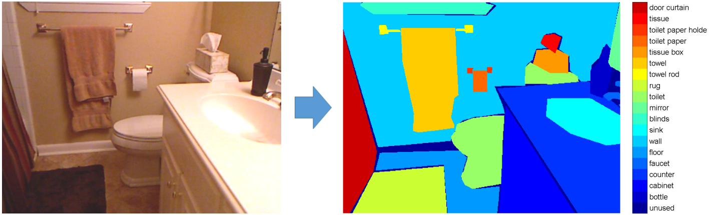
- Semantic - Meaningful

Semantic Segmentation
Introduction

- Semantic - Meaningful
- Segmentation - Division/Separation

Semantic Segmentation
Introduction

- Semantic - Meaningful
- Segmentation - Division/Separation



Dataset
Introduction

The literature suggest to use:

The NYU-v2 dataset[13]

Dataset
Introduction

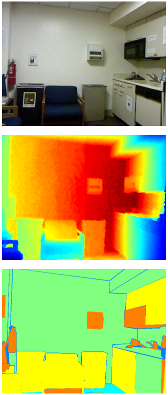
The literature suggest to use:

The NYU-v2 dataset[13]

About the dataset:

- 1449 densely labelled Kinect 360 RGB-D images
- 894 annotated classes
- Data split used by SOTA (795/654)
- Mapping into four semantic classes

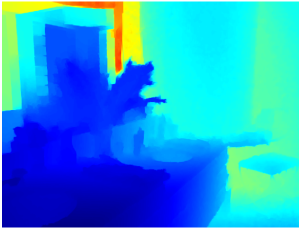
Floor	Structure	Furniture	Props	Unused



Dataset
Introduction

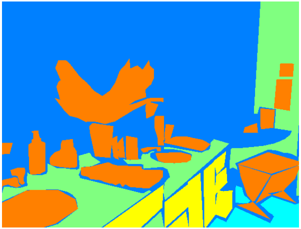
RGB image

Depth map



Using all available classes

Four semantic classes

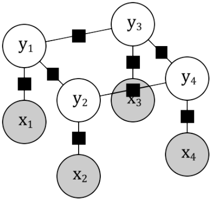


Choice of Models

Introduction

The literature points toward Conditional Random Fields.

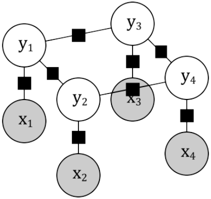
- Works that use the CRF shows SOTA results[2, 9, 11, 8].



Choice of Models
Introduction

The literature points toward Conditional Random Fields.

- Works that use the CRF shows SOTA results[2, 9, 11, 8].
- Structured prediction.

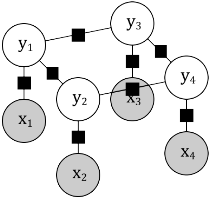


Choice of Models

Introduction

The literature points toward Conditional Random Fields.

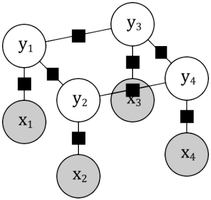
- Works that use the CRF shows SOTA results[2, 9, 11, 8].
- Structured prediction.
- Contextual.



Choice of Models
Introduction

The literature points toward Conditional Random Fields.

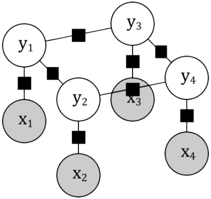
- Works that use the CRF shows SOTA results[2, 9, 11, 8].
- Structured prediction.
- Contextual.
- Independent of interest points.



Choice of Models
Introduction

The literature points toward Conditional Random Fields.

- Works that use the CRF shows SOTA results[2, 9, 11, 8].
- Structured prediction.
- Contextual.
- Independent of interest points.



Couprie *et. al* use Convolutional Neural Networks, however their results are worse than the previously mentioned.

Choice of Models
Introduction

Testing the CRF alone:

Choice of Models
Introduction

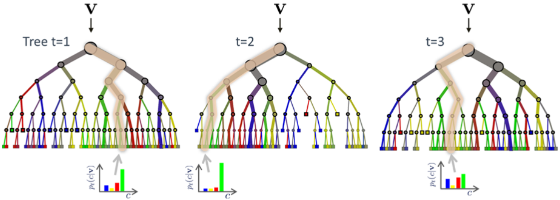
Testing the CRF alone: not good enough

Choice of Models

Introduction

Testing the CRF alone: not good enough

Stückler *et. al*[14] use a Random Forest with random offset features, which gives pixel wise predictions.



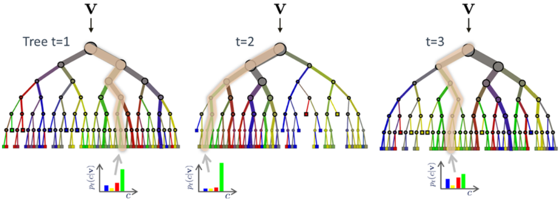
4

[4](Decision Forests for Computer Vision and Medical Image Analysis, Criminisi and Shotton, 2013, [4])

Choice of Models
Introduction

Testing the CRF alone: not good enough

Stückler *et. al*[14] use a Random Forest with random offset features, which gives pixel wise predictions.



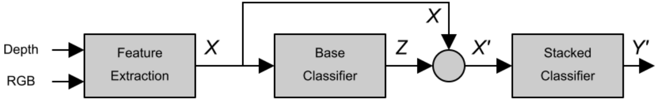
4

Use it as an input to the CRF.
A similar approach is adopted in [9] and shows good results.

[4](Decision Forests for Computer Vision and Medical Image Analysis, Criminisi and Shotton, 2013, [4])

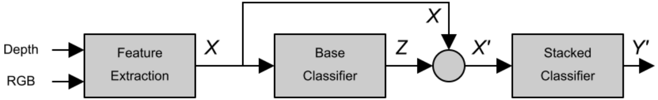
Choice of Models
Introduction

Lastly, the work of Gatta *et. al*[5] shows how stacking classifiers and using intermediate multi-scale decompositions can enhance performance of models. This is further shown by Sampedro *et. al* [12].



Choice of Models
Introduction

Lastly, the work of Gatta *et. al*[5] shows how stacking classifiers and using intermediate multi-scale decompositions can enhance performance of models. This is further shown by Sampedro *et. al* [12].

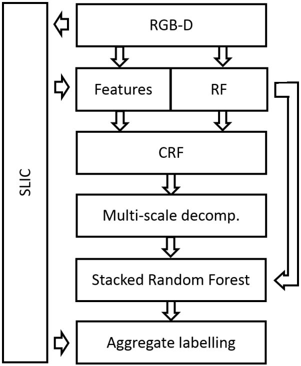


The included models are consequently:

- Conditional Random Field
- Random Forest with Random Offset Features
- A stacked classifier using the MSSL framework

Overview

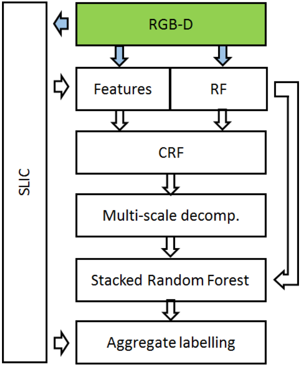
The model consists of three main elements:



Overview

The model consists of three main elements:

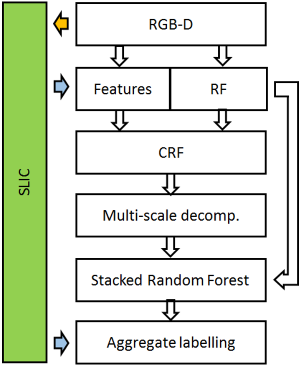
- SLIC, CRF and features



Overview

The model consists of three main elements:

- SLIC, CRF and features

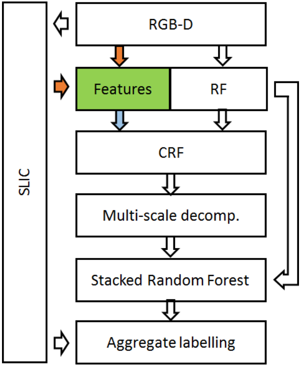


Overview

Segmentation of RGB-D Indoor Scenes
Mikkel Thøgersen

The model consists of three main elements:

- SLIC, CRF and features

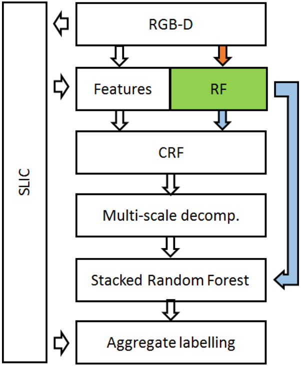


Dept. of Electronic Systems
Aalborg University
Denmark

Overview

The model consists of three main elements:

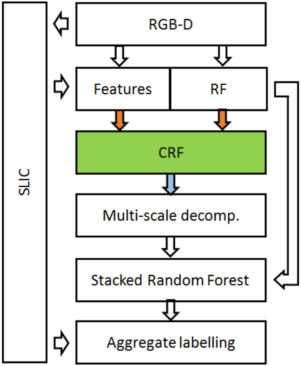
- SLIC, CRF and features
- Random Forest



Overview

The model consists of three main elements:

- SLIC, CRF and features
- Random Forest

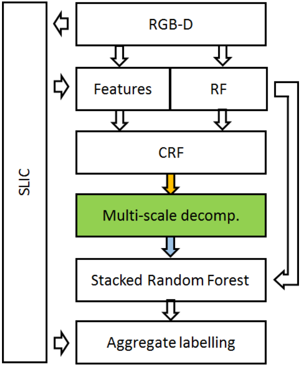


Overview

10

The model consists of three main elements:

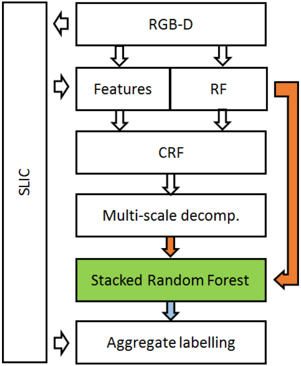
- SLIC, CRF and features
- Random Forest
- MMSSL using a Stacked Random Forest



Overview

The model consists of three main elements:

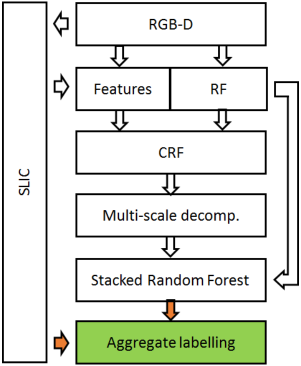
- SLIC, CRF and features
- Random Forest
- MMSSL using a Stacked Random Forest



Overview

The model consists of three main elements:

- SLIC, CRF and features
- Random Forest
- MMSSL using a Stacked Random Forest



AALBORG UNIVERSITY
DENMARK

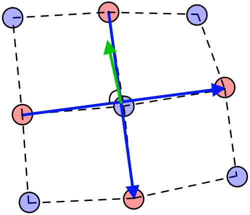
Extracting 3D Cartesian coordinates:

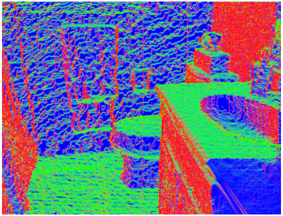
Dept. of Electronic Systems
Aalborg University
Denmark

Extracting Primitives
Preprocessing

Extracting 3D Cartesian coordinates:

Extracting normals:





Superpixel Segmentation

Preprocessing

A trend in literature is to use an over segmentation method to obtain superpixels and then label them. An over segmentation has some advantages:

Dept. of Electronic Systems
Aalborg University
Denmark

Superpixel Segmentation

Preprocessing

A trend in literature is to use an over segmentation method to obtain superpixels and then label them. An over segmentation has some advantages:

- Diminishing data

Superpixel Segmentation
Preprocessing

Segmentation of RGB-D Indoor Scenes
Mikkel Thøgersen

A trend in literature is to use an over segmentation method to obtain superpixels and then label them. An over segmentation has some advantages:

- Diminishing data
- Keeps object boundaries

Dept. of Electronic Systems
Aalborg University
Denmark

Superpixel Segmentation
Preprocessing

A trend in literature is to use an over segmentation method to obtain superpixels and then label them. An over segmentation has some advantages:

- Diminishing data
- Keeps object boundaries
- Enables local features in coherent regions

Superpixel Segmentation
Preprocessing

A trend in literature is to use an over segmentation method to obtain superpixels and then label them. An over segmentation has some advantages:

- Diminishing data
- Keeps object boundaries
- Enables local features in coherent regions
- Match very well with graph based methods

Superpixel Segmentation
Preprocessing

A trend in literature is to use an over segmentation method to obtain superpixels and then label them. An over segmentation has some advantages:

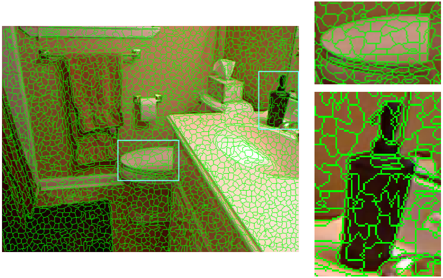
- Diminishing data
- Keeps object boundaries
- Enables local features in coherent regions
- Match very well with graph based methods

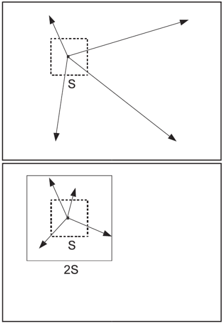
Achanta *et. al*[1] presents the SLIC segmentation:

- Superior in speed
- Superior in performance (mostly)
- Used in the works of Reza and Kosécka[11] and Müeller and Behnke[9].

SLIC Segmentation
Preprocessing

Is based on K-means clustering and limiting the search area for the clustering.



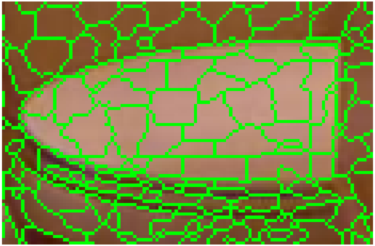


[1]

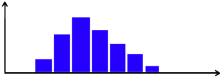
Introduction
Features

The features should describe the superpixels! Furthermore, Conditional Random Fields (CRFs) rely on contextual features.

- Node Features → Node Potentials
- Edge Features → Edge Potentials



Genreic Features
Features



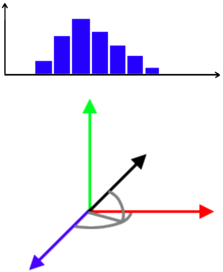
Primitive features:

- Color histograms

Genreic Features
Features

Primitive features:

- Color histograms
- Normals

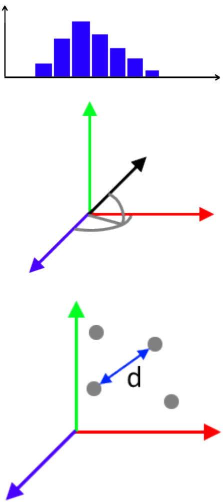


Genreic Features

Features

Primitive features:

- Color histograms
- Normals
- Spatial differences



Dominant Normals
Features

Finding the dominant normal directions:

- Up is generally up

Dept. of Electronic Systems
Aalborg University
Denmark

Dominant Normals

Features

Finding the dominant normal directions:

- Up is generally up
- Dominant directions
  $\rightarrow$ Manhattan assumption
  $\rightarrow$ scene coordinates

Dominant Normals

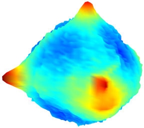
Features

Finding the dominant normal directions:

- Up is generally up
- Dominant directions
  $\rightarrow$ Manhattan assumption
  $\rightarrow$ scene coordinates

Method:

- Mean shift clustering
- Evaluate modes based on direction and support



Dept. of Electronic Systems
Aalborg University
Denmark

Dominant Normals

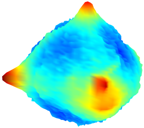
Features

Finding the dominant normal directions:

- Up is generally up
- Dominant directions
  $\rightarrow$ Manhattan assumption
  $\rightarrow$ scene coordinates

Method:

- Mean shift clustering
- Evaluate modes based on direction and support

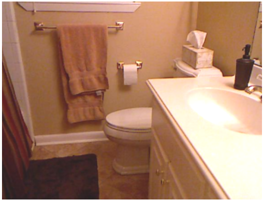


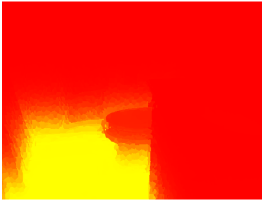
$$n_{\text{floor}} = \operatorname*{argmax}_{p \in \boldsymbol{P}} \exp\left[-\left(\frac{|p_\theta - \theta_{\text{std}}|}{180}\right)^{\lambda} \left(1 - \frac{p_\mu}{\sum_{p \in \boldsymbol{P}} p_\mu}\right)\right]$$

Dominant Normals
Features

Knowing the floor normal gives a range of valuable information:

- Find the floor





Dominant Normals
Features

Knowing the floor normal gives a range of valuable information:

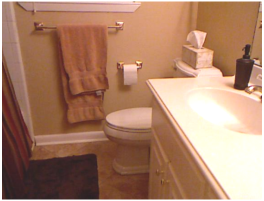
- Find the floor
- Comparing the vertical alignment

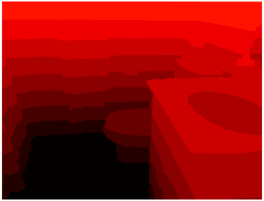
Dominant Normals

Features

Knowing the floor normal gives a range of valuable information:

- Find the floor
- Comparing the vertical alignment
- Getting height



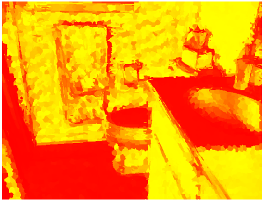


Dominant Normals

Features

Knowing the floor normal gives a range of valuable information:

- Find the floor
- Comparing the vertical alignment
- Getting height
- Helps to find the walls

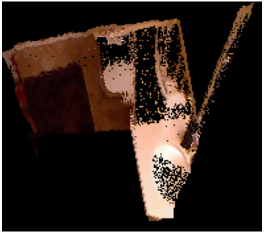


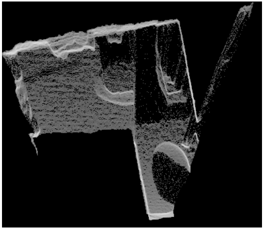
Dominant Normals

Features

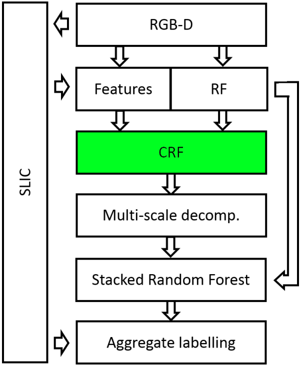
Knowing the floor normal gives a range of valuable information:

- Find the floor
- Comparing the vertical alignment
- Getting height
- Helps to find the walls
- Room Layout



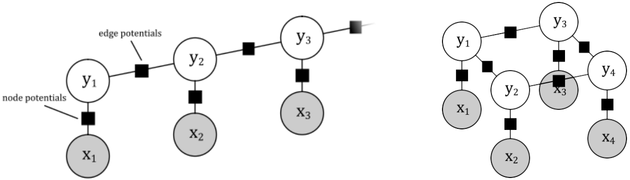


Overview
CRF Model

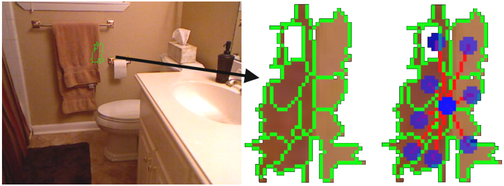


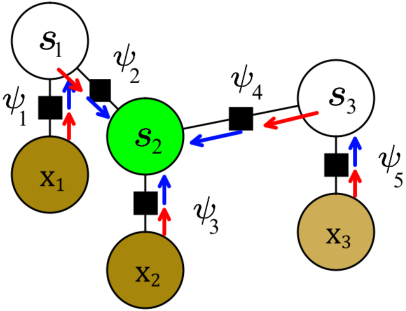
Graph Structure

CRF Model



Graph Structure
CRF Model





Parameterization

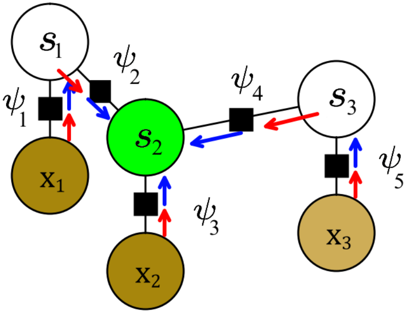
CRF Model

But what is ψ_α?

Parameterization
CRF Model

But what is ψ_α?

$$\psi_\alpha(s_{i-1}, k, s_i, c, \mathbf{w}) \rightarrow \exp\left(f_n(s_{i-1}, k, \mathbf{w}) + f_e(s_{i-1}, k, s_i, c, \mathbf{w})\right)$$

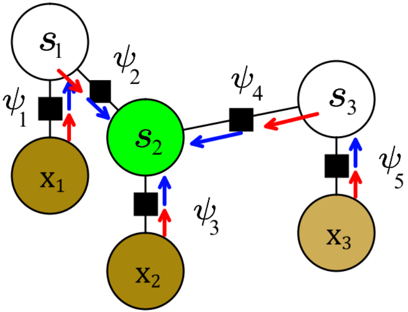


Parameterization

CRF Model

But what is ψ_α?

$$\psi_\alpha(s_{i-1}, k, s_i, c, \mathbf{w}) \quad \rightarrow \quad \exp\left(f_n(s_{i-1}, k, \mathbf{w}) + f_e(s_{i-1}, k, s_i, c, \mathbf{w})\right)$$



So what is f_n and f_e then?

All the node and edge feature functions combined!

$$f_n(s_{i-1}, k, \mathbf{w}) = w_k\ f_{\text{Nstd}}(s_{i-1}) + w_k\ f_{\text{D}\nabla}(s_{i-1}) + \ldots$$
$$f_e(s_{i-1}, k, s_i, c, \mathbf{w}) = w_{\mathbb{1}\{c=k\}}\ f_{\text{ColDiff}}(s_{i-1}, s_i) + \ldots$$

Learning
CRF Model

To fit the model to the data, we have to learn the $\mathbf{w}$-parameters. Several methods available, but they all optimize some form of:

$$\mathbf{w}^* = \underset{w}{\operatorname{argmax}} \prod_{n=1}^{N} p(y^n | x^n, \mathbf{w})$$

where y^n and x^n are the training samples.
Through the use of the Kullback-Leibler divergence, and by adding regularization parameters, an expression for the optimization of $\mathbf{w}^*$ is derived:

$$\mathcal{L}(\mathbf{w}) = \lambda ||\mathbf{w}||^2 + \sum_{n=1}^{N} \sum_{\psi \in \Psi} \psi(\mathbf{s}^n, \mathbf{s}^n) - \sum_{n=1}^{N} \log Z(x^n, w)$$

Which is the Regularized Maximum Conditional Likelihood Training.

Optimization
CRF Model

There are different ways of optimizing the parameters, $\mathbf{w}$.

Optimization
CRF Model

There are different ways of optimizing the parameters, $\mathbf{w}$.

From descriptions of the methods[10], the Pseudo-Likelihood method is chosen as it is fast to train.
Main benefit: Optimize over the individual nodes $\rightarrow$ fast.

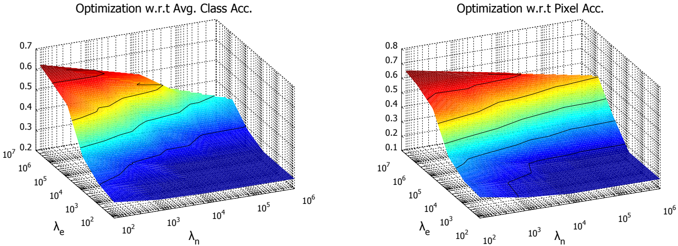
Optimization
CRF Model

The model is optimized w.r.t the regularization parameters, using a validation set and a 2D grid search:

Optimization

CRF Model

The model is optimized w.r.t the regularization parameters, using a validation set and a 2D grid search:



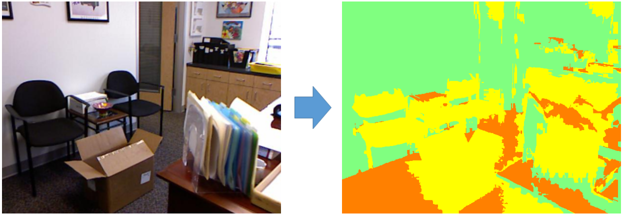
This shows an accuracy in the optimal points of:

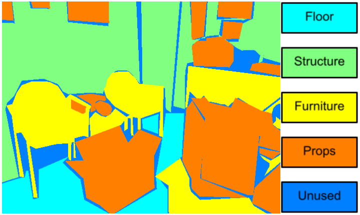
Opt. Param.	
λ_n	10^1
λ_e	10^6

Resulting accuracies	
Pix. acc.	70.9
avg. class acc.	67.7

AALBORG UNIVERSITY
DENMARK

Optimization
CRF Model





Dept. of Electronic Systems
Aalborg University
Denmark

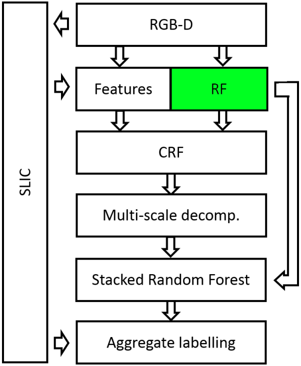
SOTA comparison

CRF Model

Work	Per class acc.	Pix. acc.
Müller and Behnke[9]	**71.9**	**72.3**
Couprie *et al*[3]	63.5	64.5
Khan *et al*[8]	65.6	69.2
Gupta *et al*[6]	65	64.9
Nico Höft *et al*[7]	62.0	61.1
This work	67.7	70.9

	floor	structure	furniture	props
floor	0.93	0.00	0.06	0.00
structure	0.01	0.79	0.17	0.04
furniture	0.02	0.11	0.80	0.07
props	0.08	0.28	0.46	0.18

Overview


SLIC
RGB-D
Features
RF
CRF
Multi-scale decomp.
Stacked Random Forest
Aggregate labelling

Random Offset Features

RF and ROF features

Random Offset Features:

- Similar to the well known Haar-features used by Viola-Jones[15]
- Can capture non-obvious features
- Fast
- Proven, used by Stückler *et al*[14].

Random Offset Features

RF and ROF features

For a query point p:

- Generate random pixel offsets

Random Offset Features

RF and ROF features

For a query point p:

- Generate random pixel offsets
- Generate a randomly sized box around each offset

Random Offset Features

RF and ROF features

For a query point p:

- Generate random pixel offsets
- Generate a randomly sized box around each offset
- Choose a channel at random (Lab color or depth) for each offset and sum inside the box

Random Offset Features

RF and ROF features

For a query point p:

- Generate random pixel offsets
- Generate a randomly sized box around each offset
- Choose a channel at random (Lab color or depth) for each offset and sum inside the box
- Randomly take the absolute differences between the values

Random Offset Features

RF and ROF features

For a query point p:

- Generate random pixel offsets
- Generate a randomly sized box around each offset
- Choose a channel at random (Lab color or depth) for each offset and sum inside the box
- Randomly take the absolute differences between the values
- Done!

Dept. of Electronic Systems
Aalborg University
Denmark

Random Forest
RF and ROF features

To make these features work:

- Create a large set of Random Offset Features

Dept. of Electronic Systems
Aalborg University
Denmark

Random Forest

RF and ROF features

To make these features work:

- Create a large set of Random Offset Features
- Test them all

Random Forest

RF and ROF features

To make these features work:

- Create a large set of Random Offset Features
- Test them all
- Evaluate features

Random Forest
RF and ROF features

To make these features work:

- Create a large set of Random Offset Features
- Test them all
- Evaluate features
- Keep and use the best of them

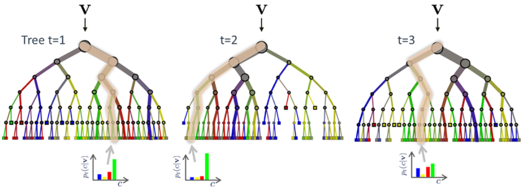
Random Forest

RF and ROF features

To make these features work:

- Create a large set of Random Offset Features
- Test them all
- Evaluate features
- Keep and use the best of them

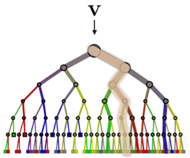
As a model for evaluating and using them, a Random Forest is chosen!



Random Forest

RF and ROF features

Decision Tree, consecutive split functions:

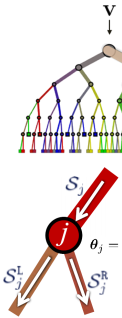


Dept. of Electronic Systems
Aalborg University
Denmark

Random Forest

RF and ROF features

Decision Tree, consecutive split functions:

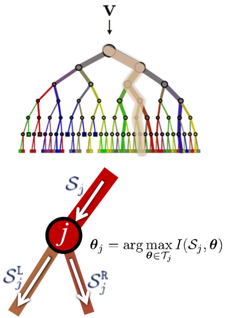


$$\boldsymbol{\theta}_j = \arg\max_{\boldsymbol{\theta}\in\mathcal{T}_j} I(\mathcal{S}_j, \boldsymbol{\theta})$$

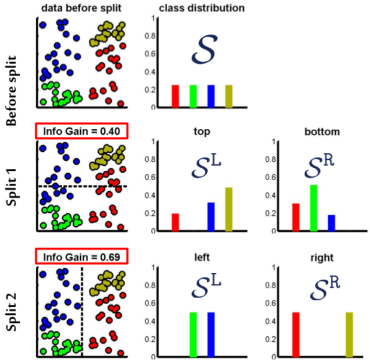
[4]

Random Forest
RF and ROF features

Decision Tree, consecutive split functions:



[4]

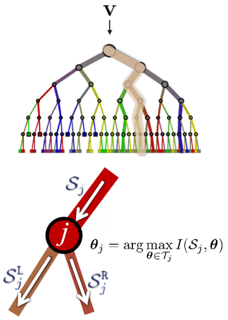


[4]

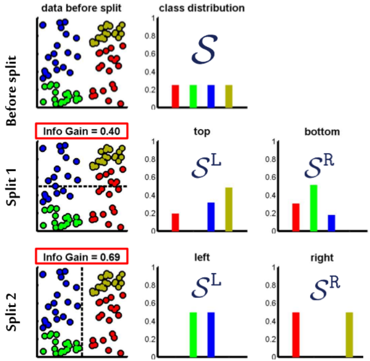
Random Forest

RF and ROF features

Decision Tree, consecutive split functions:



[4]



[4]

Information Gain (KL-divergence):

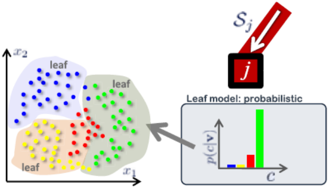
$$I(S,\theta) = H(S) - \sum_{i\in\{L,R\}} \frac{|S^i|}{|S|} H(S^i)$$

Shannon Entropy:

$$H(S) = -\sum_{c\in\mathcal{C}} p(c)\log(p(c))$$

AALBORG UNIVERSITY DENMARK

At the leaf nodes, a probability for each class is assigned.

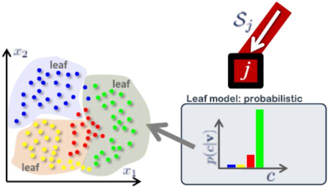


Dept. of Electronic Systems
Aalborg University
Denmark

Random Forest

RF and ROF features

At the leaf nodes, a probability for each class is assigned.



Depth of the tree is controlled usually by:

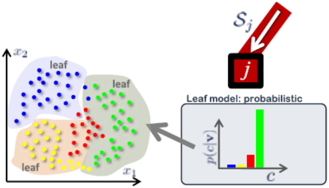
- Maximum depth
- Minimum number of samples at leaf

AALBORG UNIVERSITY
DENMARK

Random Forest

RF and ROF features

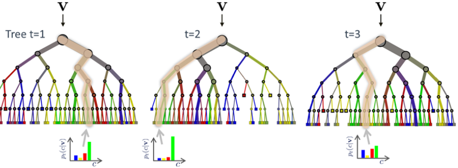
At the leaf nodes, a probability for each class is assigned.



Depth of the tree is controlled usually by:

- Maximum depth
- Minimum number of samples at leaf

Trees are gathered in ensembles and trained using Bagging. → Random Forests



Random Forest

RF and ROF features

A Random Forest is trained on three thousand random features.

Random Forest

RF and ROF features

A Random Forest is trained on three thousand random features.
Based on out-of-bag-samples, the features are evaluated using the trained model.
Following, the features are ranked.

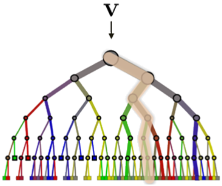
Random Forest

RF and ROF features

A Random Forest is trained on three thousand random features.
Based on out-of-bag-samples, the features are evaluated using the trained model.
Following, the features are ranked.

The final model can now be trained.

The model has a series of parameters:



[4]

Dept. of Electronic Systems
Aalborg University
Denmark

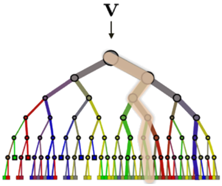
Random Forest

RF and ROF features

A Random Forest is trained on three thousand random features.
Based on out-of-bag-samples, the features are evaluated using the trained model.
Following, the features are ranked.

The final model can now be trained.

The model has a series of parameters:



[4]

Random Forest

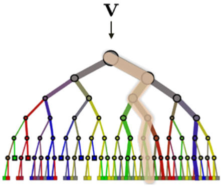
RF and ROF features

A Random Forest is trained on three thousand random features.
Based on out-of-bag-samples, the features are evaluated using the trained model.
Following, the features are ranked.

The final model can now be trained.

The model has a series of parameters:

- Number of splits at each branch



[4]

Random Forest

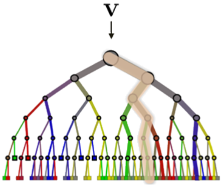
RF and ROF features

A Random Forest is trained on three thousand random features.
Based on out-of-bag-samples, the features are evaluated using the trained model.
Following, the features are ranked.

The final model can now be trained.

The model has a series of parameters:

- Number of splits at each branch
- Number of trees to train



[4]

Random Forest

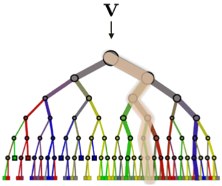
RF and ROF features

A Random Forest is trained on three thousand random features.
Based on out-of-bag-samples, the features are evaluated using the trained model.
Following, the features are ranked.

The final model can now be trained.

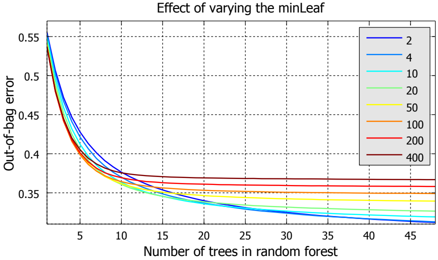
The model has a series of parameters:

- Number of splits at each branch
- Number of trees to train
- Minimum number of samples at Leaf



[4]

Optimization
RF and ROF features

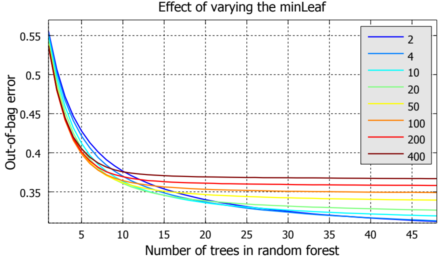


Parameter	value
nFeatures	40
nSamples	$1.56 \cdot 10^6$
nTrees	48
nSplitFeatures	7
minLeaf	-

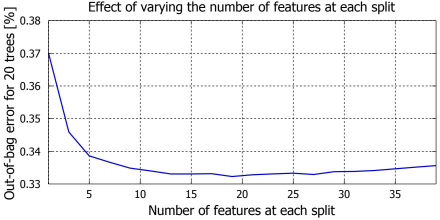
Dept. of Electronic Systems
Aalborg University
Denmark

Optimization

RF and ROF features



Parameter	value
nFeatures	40
nSamples	$1.56 \cdot 10^6$
nTrees	48
nSplitFeatures	7
minLeaf	-



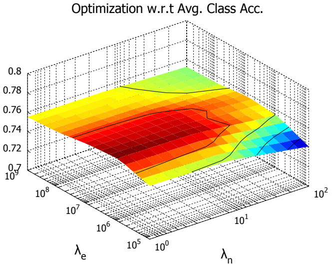
Parameter	value
nFeatures	40
nSamples	$1.56 \cdot 10^6$
nTrees	20
nSplitFeatures	—
minLeaf	10

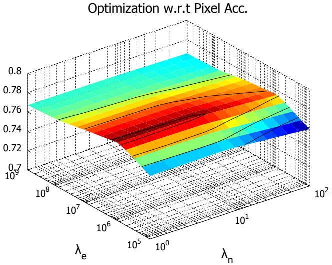
AALBORG UNIVERSITY
DENMARK

Optimization

RF and ROF features

To accommodate the Random Forests' predictions as input for the CRF, the regularization have to be re-optimized.





Optimal parameters are approximately:

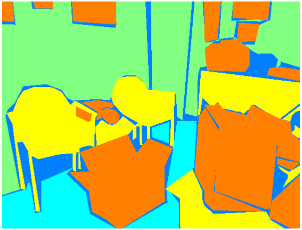
Optimal Parameters	
λ_n	10^0
λ_e	10^6

Dept. of Electronic Systems
Aalborg University
Denmark

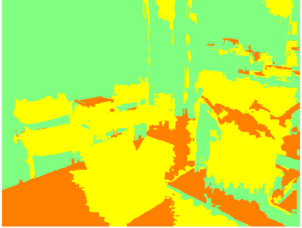
Optimization

RF and ROF features

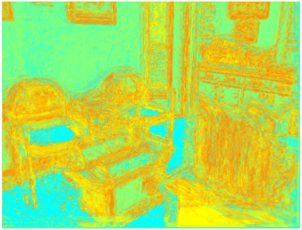
Original



Ground Truth



CRF labeling



RF labeling

Floor

Structure

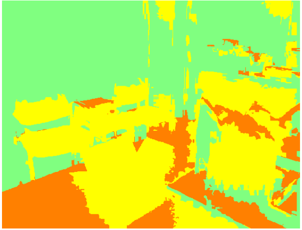
Furniture

Props

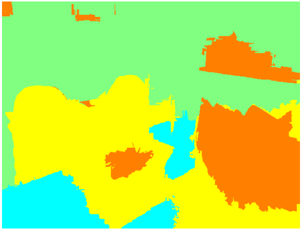
Unused

Optimization

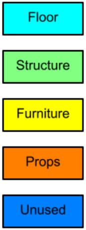
RF and ROF features

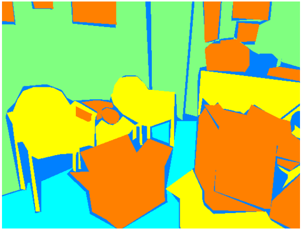


CRF



CRF + RF





Ground Truth

Dept. of Electronic Systems
Aalborg University
Denmark

SOTA comparison

RF and ROF features

Work	Per class acc.	Pix. acc.
Müller and Behnke[9]	**71.9**	**72.3**
Couprie *et al*[3]	63.5	64.5
Khan *et al*[8]	65.6	69.2
Gupta *et al*[6]	65	64.9
Nico Höft *et al*[7]	62.0	61.1
This work	70.0	71.5

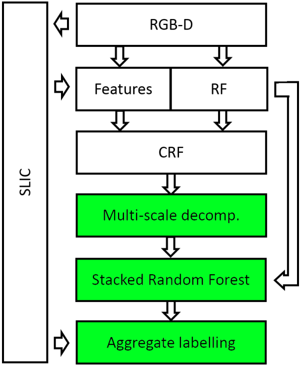
SOTA comparison
RF and ROF features

Work	Per class acc.	Pix. acc.
Müller and Behnke[9]	**71.9**	**72.3**
Couprie *et al*[3]	63.5	64.5
Khan *et al*[8]	65.6	69.2
Gupta *et al*[6]	65	64.9
Nico Höft *et al*[7]	62.0	61.1
This work	70.0	71.5

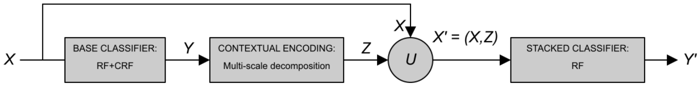
Not good enough

Overview

Multi-scale Multi-class Stacked Sequential Learning



Multi-scale Multi-class Stacked Sequential Learning



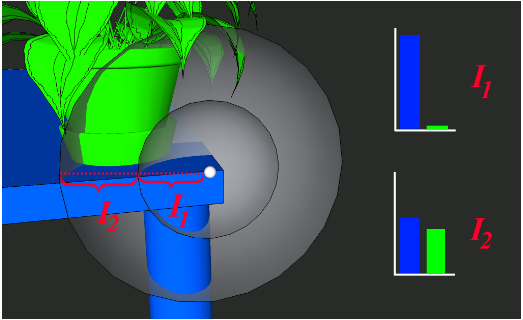
Consists of:

- A multi-scale decomposition
- A stacked classifier

Dept. of Electronic Systems
Aalborg University
Denmark

Multi-scale Decomposition

Multi-scale Multi-class Stacked Sequential Learning

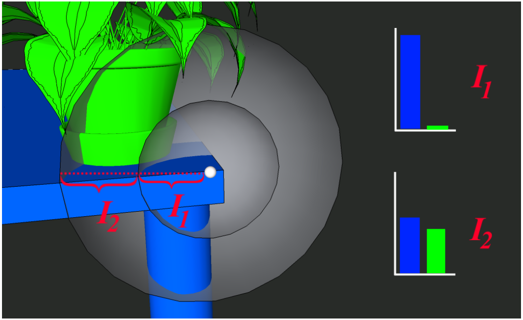


From a query point:

Dept. of Electronic Systems
Aalborg University
Denmark

Multi-scale Decomposition

Multi-scale Multi-class Stacked Sequential Learning

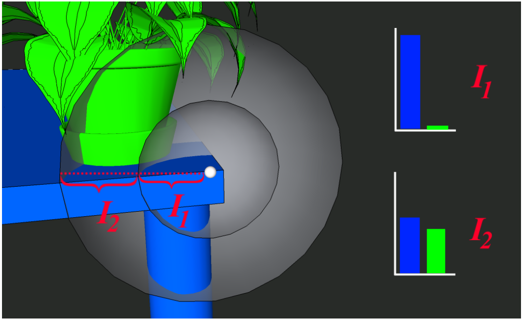


From a query point:

- Define a set of distance intervals, I. (4 chosen)

Multi-scale Decomposition

Multi-scale Multi-class Stacked Sequential Learning

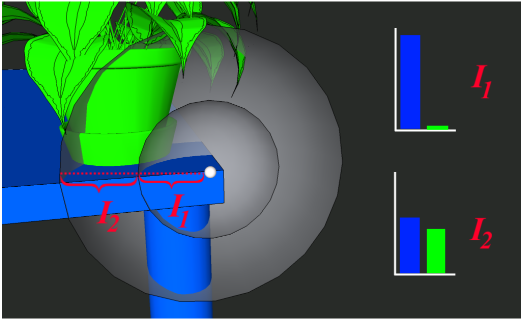


From a query point:

- Define a set of distance intervals, I. (4 chosen)
- For each interval sum over each of the confidence maps.

Multi-scale Decomposition

Multi-scale Multi-class Stacked Sequential Learning

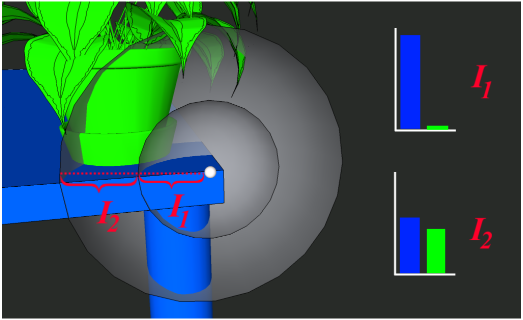


From a query point:

- Define a set of distance intervals, I. (4 chosen)
- For each interval sum over each of the confidence maps.
- Normalize over each interval.

Multi-scale Decomposition

Multi-scale Multi-class Stacked Sequential Learning

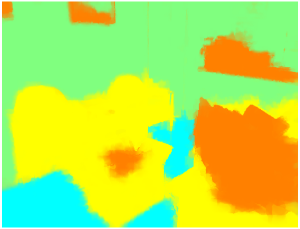


From a query point:

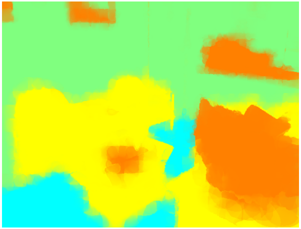
- Define a set of distance intervals, I. (4 chosen)
- For each interval sum over each of the confidence maps.
- Normalize over each interval.
- Result: a CI dimension feature vector (16-dim).

Multi-scale Decomposition
Multi-scale Multi-class Stacked Sequential Learning

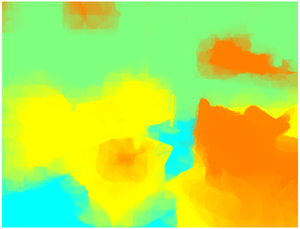
Decomposition images



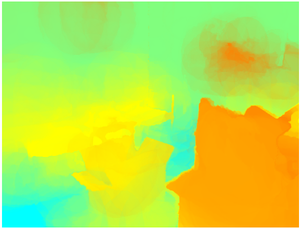
2 cm



5 cm



11 cm



30 cm

Stacked Random Forest

Multi-scale Multi-class Stacked Sequential Learning

Another Random Forest

Acts on the features of the CRF *and* the decomposition

Dept. of Electronic Systems
Aalborg University
Denmark

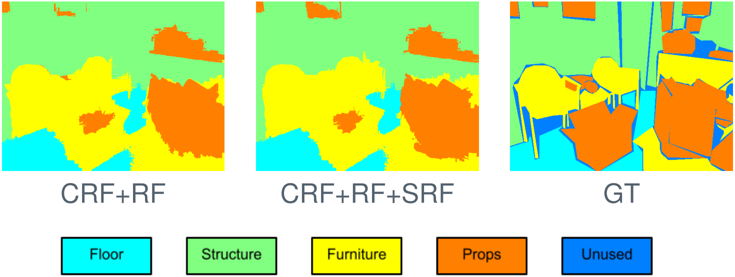
Stacked Random Forest

Multi-scale Multi-class Stacked Sequential Learning

Another Random Forest

Acts on the features of the CRF *and* the decomposition

Sample:



Final Results
Results

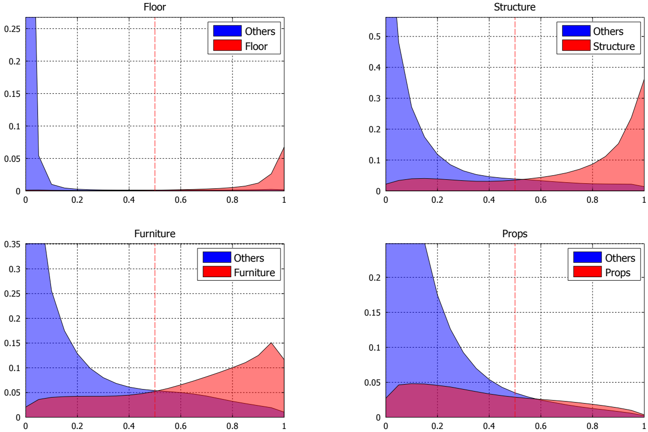
Work	Floor	Struct.	Furn.	Props	Cl. acc.	Pix. acc.
Müller and Behnke[9]	94.9	78.9	71.1	42.7	71.9	72.3
Couprie *et al*[3]	87.3	87.8	45.3	35.5	63.5	64.5
Khan *et al*[8]	87.1	**88.2**	54.7	32.6	65.6	69.2
Gupta *et al*[6]	82	73	64	37	65	64.9
Nico Höft *et al*[7]	77.9	65.4	55.9	**49.9**	62.0	61.1
This work	**95.5**	80.5	**77.1**	35.3	**72.1**	**73.8**

	floor	structure	furniture	props
floor	0.95	0.00	0.04	0.00
structure	0.01	0.80	0.12	0.07
furniture	0.02	0.11	0.77	0.10
props	0.09	0.21	0.35	0.35

Measures
Results

Measure	Floor	Struct.	Furn.	Props	Average
Precision	83.1	78.1	69.0	54.4	71.1
Recall	94.9	82.8	82.1	19.0	69.7
Specificity	97.3	86.7	80.1	96.9	90.2

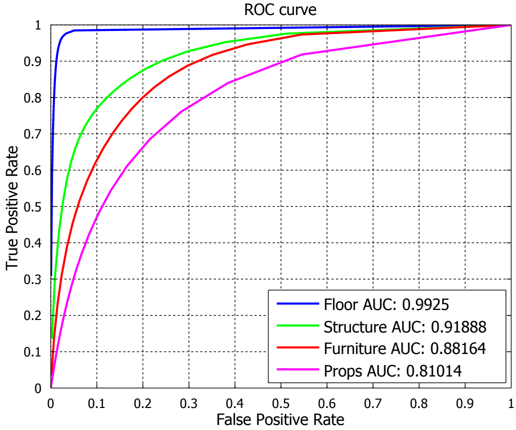
The posterior class distributions:



Measures

Results

A common method to evaluate methods is the Receiver Operating Curve:



Conclusion

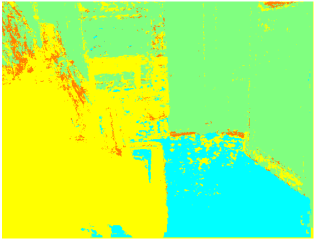
Results

The project showed:

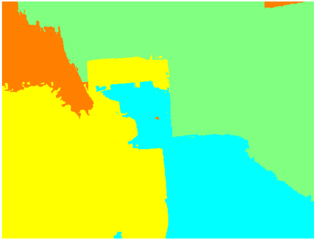
- State-of-the-art results
- MMSSL works
- CRF is powerful
- Dominant normal directions is powerful
- Paper pending! BMVC
- Journal paper

Sample results

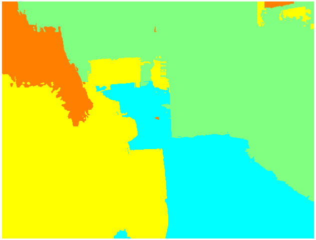
Segmentation of RGB-D Indoor Scenes
Mikkel Thøgersen



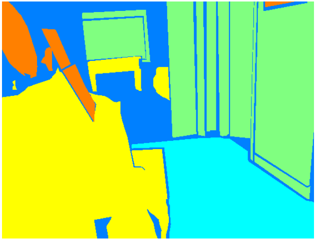
RF



CRF + RF



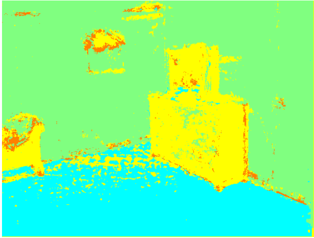
RF + CRF + RFS



GT

Dept. of Electronic Systems
Aalborg University
Denmark

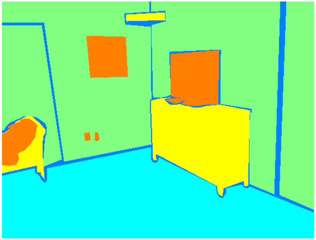
Sample results



RF

CRF + RF

RF + CRF + RFS



GT

References I

[1] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11):2274–2282, Nov 2012.

[2] Byung-soo, Pushmeet Kohli, and Silvio Savarese. 3d scene understanding by voxel-crf. In *IEEE International Conference on Computer Vision*, 2013.

[3] Camille Couprie, Clément Farabet, Laurent Najman, and Yann LeCun. Indoor semantic segmentation using depth information. *CoRR*, abs/1301.3572, 2013.

[4] Antonio Criminisi and Jamie Shotton. *Decision Forests for Computer Vision and Medical Image Analysis*. Springer, 2013.

References II

[5] Carlo Gatta, Eloi Puertas, and Oriol Pujol. Multi-scale stacked sequential learning. *Pattern Recognition*, 44, 2011.

[6] Saurabh Gupta, Pablo Arbelaez, and Jitendra Malik. Perceptual organization and recognition of indoor scenes from rgb-d images. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 564–571. IEEE, 2013.

[7] Nico Höft, Hannes Schulz, and Sven Behnke. Fast semantic segmentation of rgb-d scenes with gpu-accelerated deep neural networks. *Advances in Artificial Intelligence*, 8736:80–85, 2014.

[8] Salman Hameed Khan, Mohammed Bennamoun, Ferdous Sohel, and Roberto Togneri. Geometry driven semantic labeling of indoor scenes. In *Computer Vision ECCV*. 2014.

References III

[9] Andreas Christian Müller and Sven Behnke. Learning depth-sensitive conditional random fields for semantic segmentation of rgb-d images. *ICRA*, 2014.

[10] Sebastian Nowozin and Christoph Lampert. *Structured Prediction and Learning in Computer Vision*. NOW - the essence of knowledge, 2011.

[11] Md. Alimoor Reza and Jana Kosecka. Object recognition and segmentation in indoor scenes from rgb-d images. 2013.

[12] Frederic Sampedro, Sergio Escalera, and Anna Puig. Iterative multi-class multi-scale stacked sequential learning: Definition and application to medical volume imaging. *Pattern Recognition Letters*, 46:1–10, 2014.

[13] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgbd images. In *Computer Vision ECCV*, 2012.

References IV

[14] Jorg Stückler, Benedikt Waldvogel, Hannes Schulz, and Sven Behnke. Dense real-time mapping of object-class semantics from rgb-d video. *Journal of Real-Time Image Processing*, pages 1–11, 2013.

[15] Paul Viola and MichaelJ. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

Thank you for listening!

AALBORG UNIVERSITY
DENMARK

Thank you Jordi!