Current Topics in Computer Vision and Machine Learning

Introductory Meeting

Prof. Dr. Bastian Leibe

Lehr- und Forschungsgebiet Informatik 8 (Visual Computing Institute)

RWTH Aachen University

http://www.vision.rwth-aachen.de

Organization

- **Reports**
  - English or German (depending on the supervisor)
  - 20 pages (+- 1 page) Bibliography counts – TOC does not
  - LaTeX is mandatory

- **Presentations**
  - In English
  - 45 minutes (+- 5 minutes)
  - Block at the semester end – 3 days of presentations
  - Slide Templates will be available on the webpage
  - Laptop can be provided for the presentation if necessary

Schedule

- Hand in Declaration of Compliance
  - as hardcopy – before outline
- Outline: Monday, May 23th
- Report: Monday, June 20th – graded version!
- Slides: Monday, July 18th
- Presentations: July – 27th-29th – 3 days
  - Turn in corrected report at presentation day

Hints for your report – DOs

- Read and understand your paper
- Write a report in your own words
- Search for additional literature
  - Take part in a library tour
  - Compare your paper to work of other authors
  - Explain the bigger picture
  - Describe something extra – content beyond the topic‘s original paper
- Discuss the advantages & disadvantages of the approach
- Make the reader understand the topic
  - Audience: Your fellow seminar participants
- Correctly cite all sources (also for all figures)

Hints for your report – DON‘Ts

- Do not simply copy or translate original text!
- Do not miss the deadlines
  - Penalty for every day you exceed any deadline
- We will check if you...
  - Have copied content / text from the paper or other sources
  - Have not correctly cited any material, etc.
  - If you do, you **immediately** fail the seminar

Reminder: How to cite

- General rule: For every piece of information it has to be clear if it is your own work or someone else‘s.
  - If your text contains “Our approach…”, “We propose...”, etc. you are doing it wrong...
- Direct Quote:
  - Smith et al. state that their “approach combines x and y in order to achieve z” [5].
  - You have to use direct quotes if you copy original text.
  - **Avoid** such direct quotes – and instead use your own words
- Indirect Quotes:
  - Smith et al. use an approach which combines x and y allowing to... [5].

Reminder: How to cite

- Mind credible sources
  - Papers published in journals or conference proceedings
    - Peer reviewed == reliable and good
  - arXiv.org
    - Depends!?
  - Wikipedia
    - Can be altered by anyone and it changes over time == not good
- Use the original sources
  - Instead of sources that only cites the original source
  - That requires to also look (and dig) for the original sources!
- Use BibTeX
  - Saves a lot of trouble
  - And good practice for your master thesis

Important Details – before we start...

- Declaration of Compliance
  - Read “Ethical Guidelines for the Authoring of Academic Work”
  - See seminar webpage for the document
  - Sign and hand in to me – as hardcopy – before outline deadline

- Send all submissions regarding the seminar to
  - seminar@vision.rwth-aachen.de

Topic 1 - Ishrat

- **Deep Sliding Shapes for Amodal 3D Object Detection in RGB-D Images**

  Song and Xiao, CVPR 2015

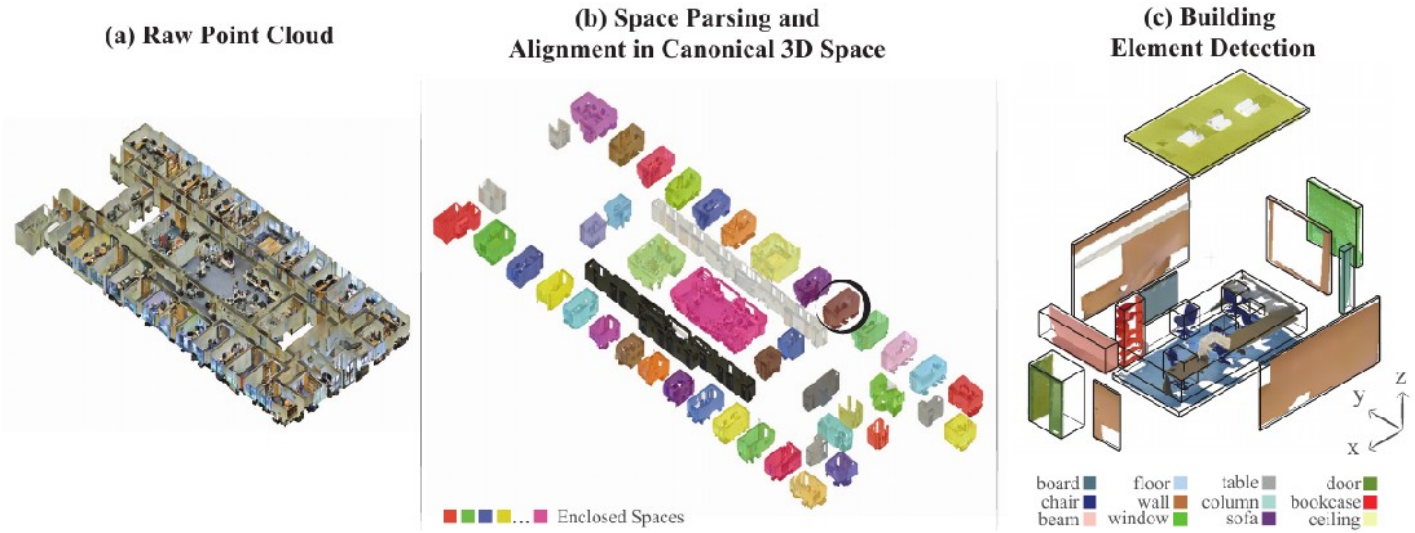
- **Goal:** 3D bounding boxes covering the full extend of the object even if occluded/truncated (amodel 3D object detection)
  1. Using CNNs for generating object proposals and classification in 3D
  2. Additionally exploiting existing CNN based 2D object detectors
  3. Combination of 1 and 2

Topic 2 - Ishrat

- **3D Semantic Parsing of Large-Scale Indoor Spaces**

  Armeni et al., CVPR 2016

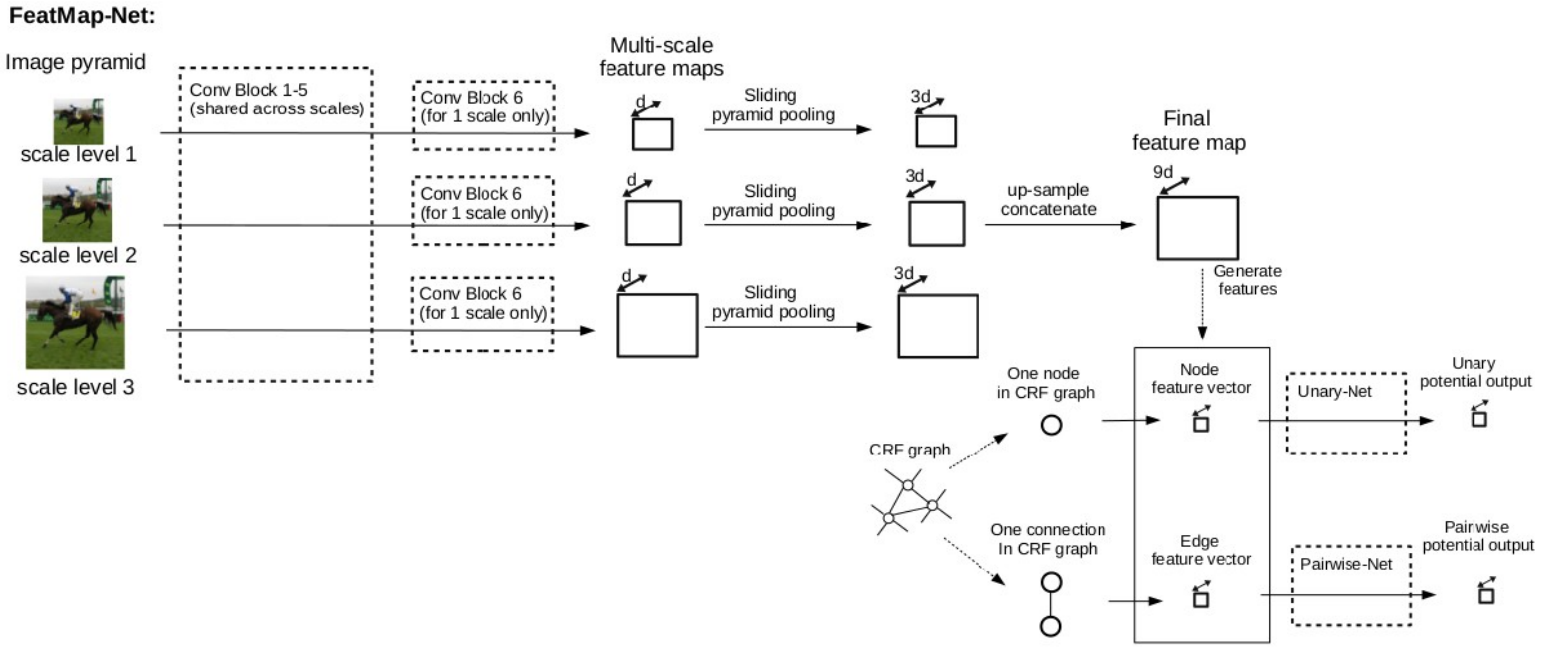
- **Input:** Point cloud of entire building
- **Output:** Parse of the building in disjoint spaces (mostly rooms) and parsing the room into semantic 3D elements like table, chair, door.
- Handling the complexity of such a huge point cloud hierarchically
- Incorporation of new semantics analyzing the space between rooms



Topic 3 - Alex

- **Exploring Context with Deep Structured models for Semantic Segmentation**
  Lin et al., arXiv 2016 (& CVPR16)

- Pixel level semantic segmentation often is post processed by CRFs
- CNNs and CRFs have been applied separately
- This method shows how to learn CRF potentials in CNNs

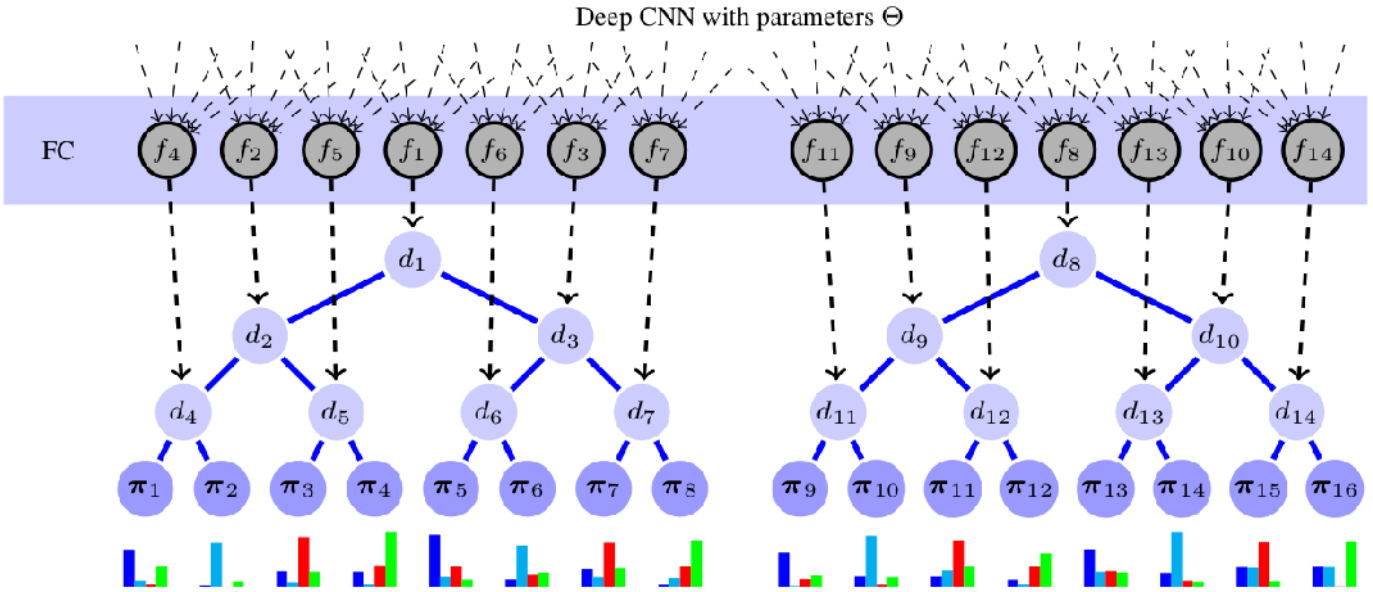


Topic 4 - Alex

- **Deep Neural Decision Forests**

  Kontschieder et al., ICCV 2015

- A combination of decision trees and deep neural networks
- Replaces a typical softmax output with a decision tree structure
- Joint optimization of the forest and the network



Topic 5 - Umer

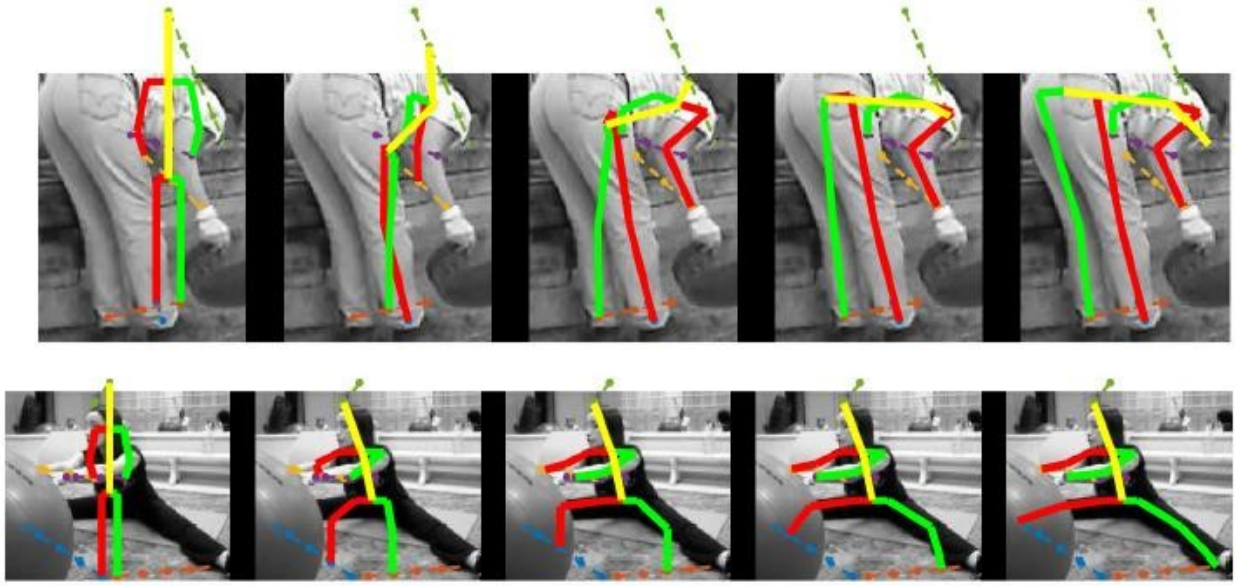
- **Pose Embeddings: A Deep Architecture for Learning to Match Human Poses**

  Mori et al., arxiv 2015
- Learn an embedding that aims to place images of people in similar poses near each other.
- Obviate the need for explicit Human Pose Estimation.
- Can be used to search for frames in videos where people are in similar pose, group activity analysis, ...

Topic 6 - Umer

- **Human Pose Estimation with Iterative Error Feedback**

  Carreira et al., CVPR 2016
- Use a self correcting model that when provided with an initial pose iteratively corrects itself.
- Encompasses both input and output spaces by using top-down feedback.

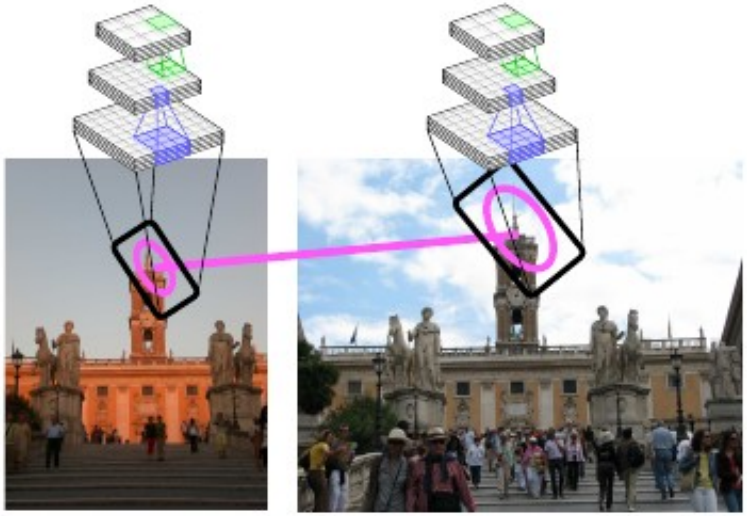


Topic 7 - Michael

- **Local Convolutional Features with Unsupervised Training for Image Retrieval**

  Paulin et al., ICCV 2015

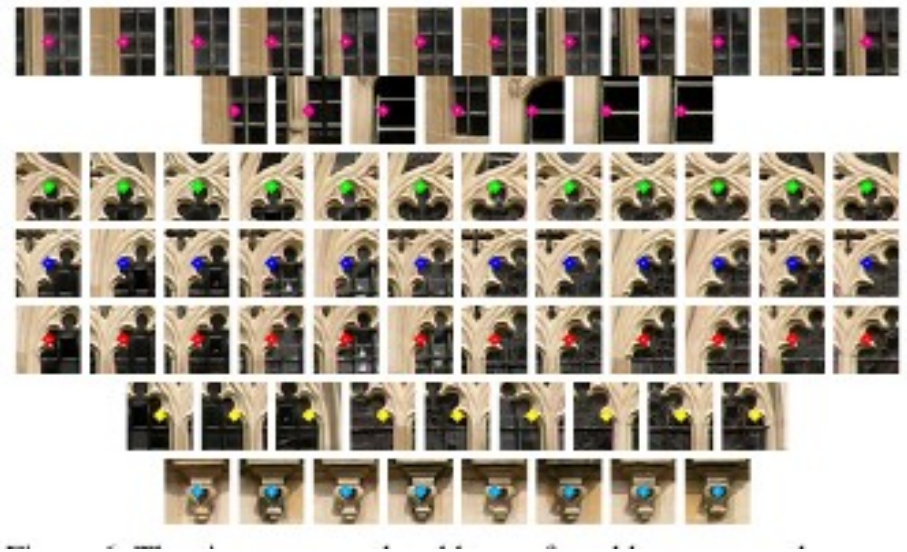
- Local features allow for robust retrieval under occlusions
- (Global) CNN features yield large improvement in retrieval
- Combination of local features and CNNs promises best of both worlds



Topic 8 - Michael

- **Early Burst Detection for Memory-Efficient Image Retrieval**
  Shi et al., CVPR 2015

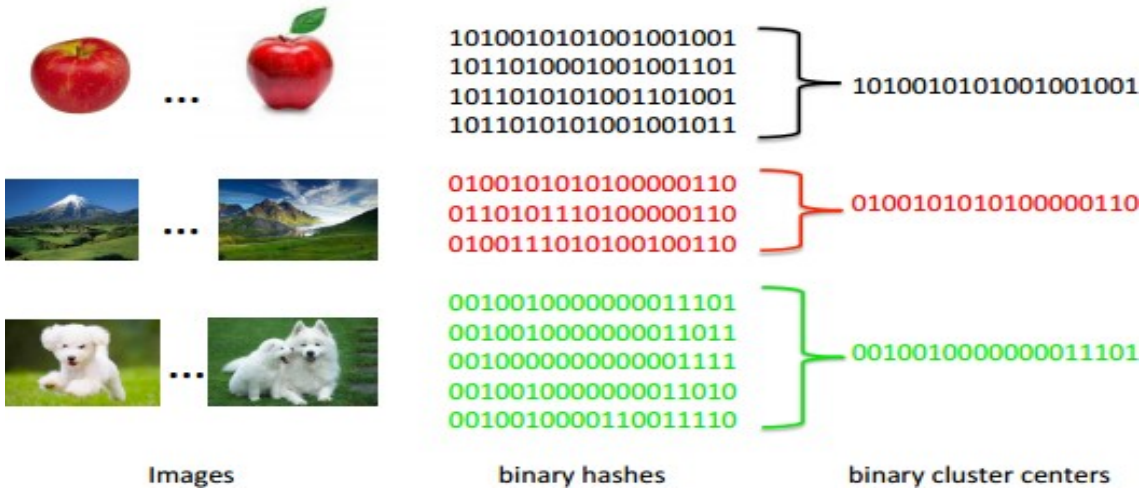
- Certain local features occur in high frequencies in images (bursts)
- idf normalization can only compensate a certain amount
- Early detection of bursts and aggregation into meta-features



Topic 9 - Dora

- **Web Scale Photo Hash Clustering on a Single Machine**
  Gong et al., CVPR 2015

- Clustering of hundreds of millions of photos per day
- Very interesting from application point of view
  (e.g. Facebook, Instagram etc.)
- Binary K-means in hash codes that retain similarity
- Online clustering on streaming data

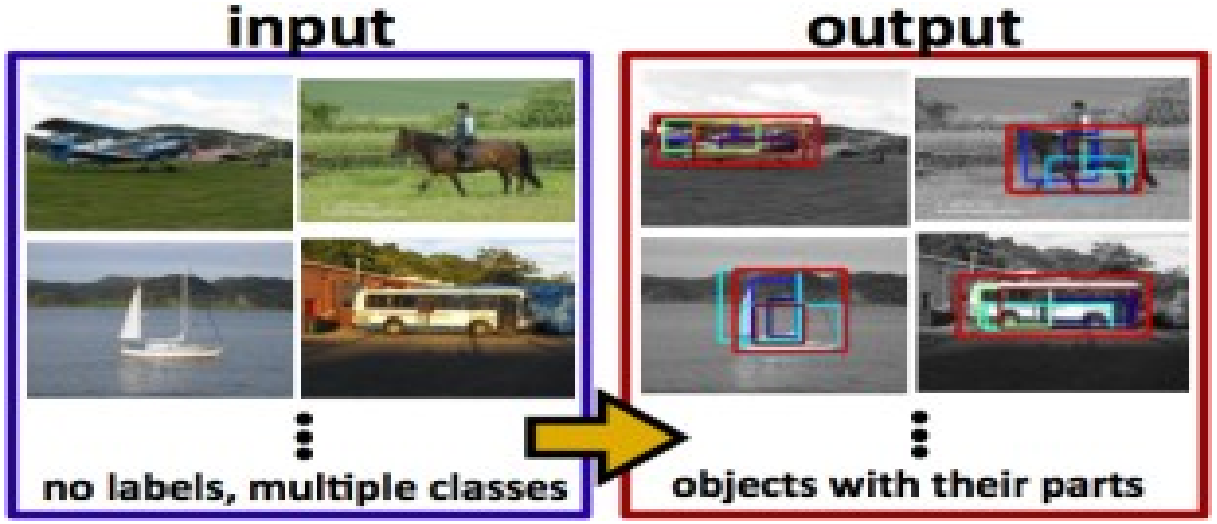


Topic 10 - Dora

- **Unsupervised Object Discovery and Localization in the Wild: Part-based Matching with Bottom-up Region Proposals**

  Cho et al., CVPR 2015

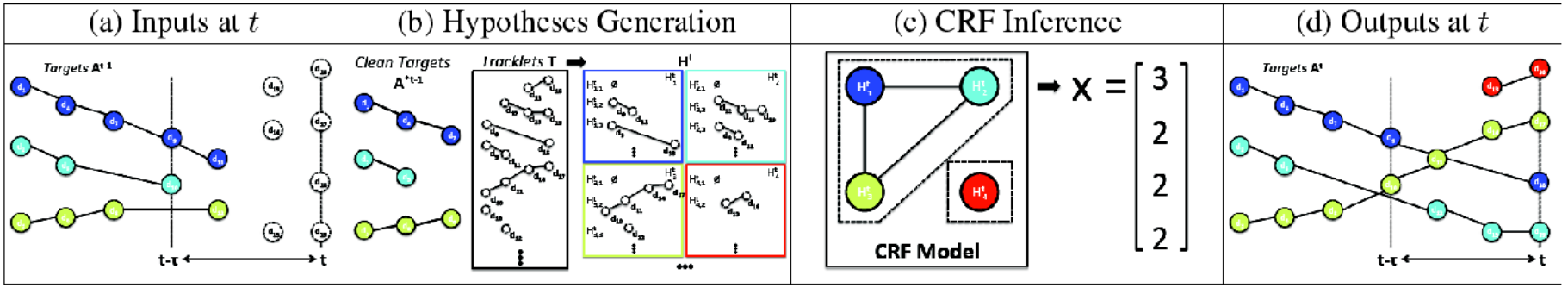
- Object discovery and localization
- Object proposal needed for speedups in many situations (e.g. CNN input)
- Fully unsupervised
- Part-based approach
- Foreground localization



Topic 11 - Aljosa

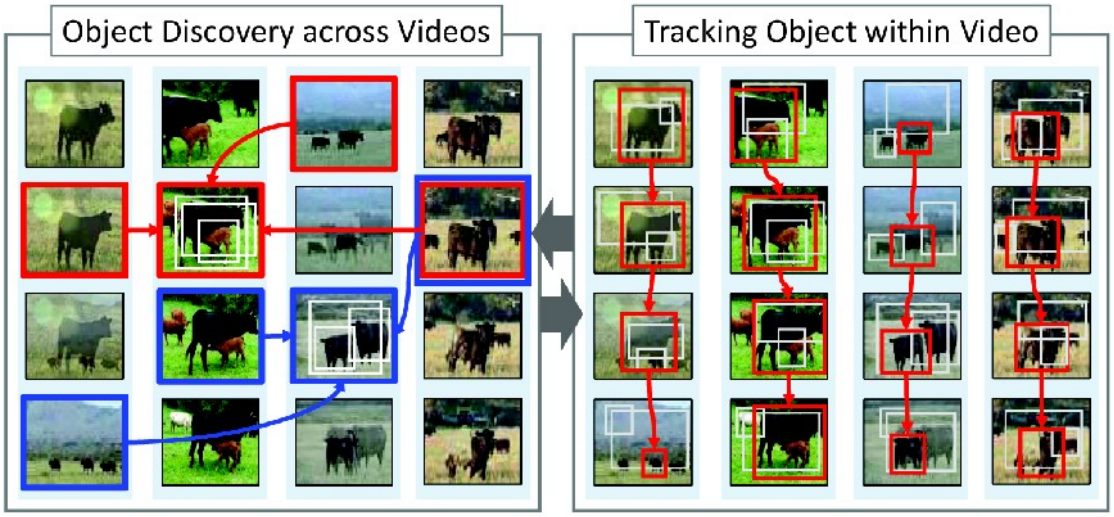
- **Near-Online Multi-target Tracking with Aggregated Local Flow Descriptor**

  Choi et al., ICCV 2015
- Multi-target tracking-by-detection in street scenes
- Tracking as model-selection
- Solve inference problem in graphical model
- State-of-the-art results on KITTI

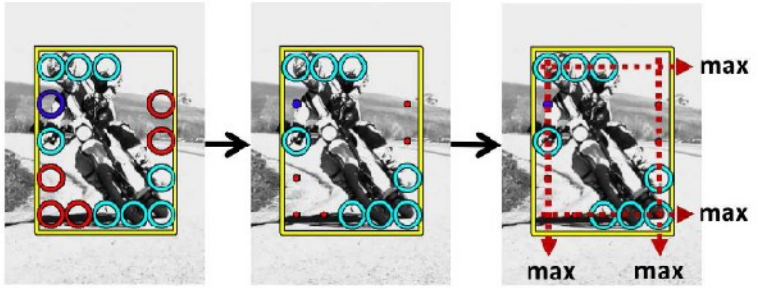


Topic 12 - Aljosa

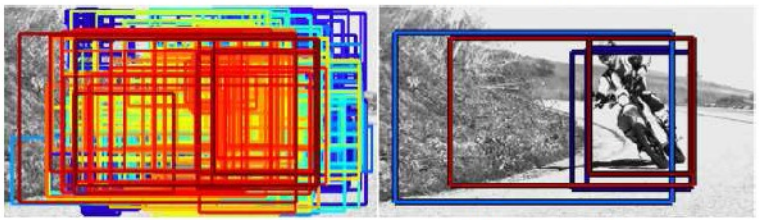
- **Unsupervised Object Discovery and Tracking in Video Collections**
  Kwak et al., ICCV 2015
- Un-supervised object discovery and tracking in video collections
- Automatically localize dominant objects
- What is an object? Challenging problem!
- Tracking helps object discovery
  (and vice versa), iterative process




(a) Video frame and its color-coded motion clusters.

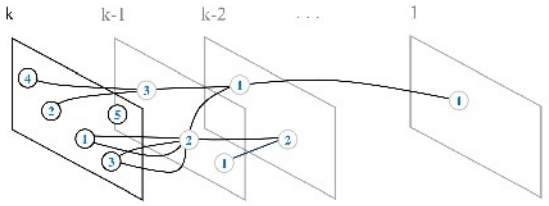


(b) Measuring the motion coherence score for a box.

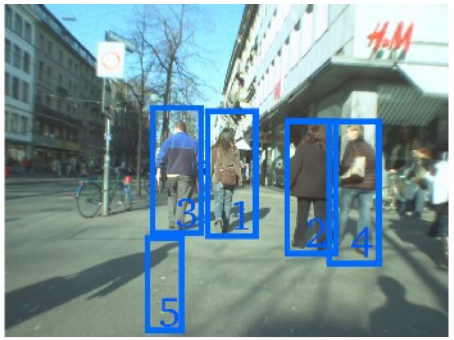

(c) Heat map of the scores and the top 5 boxes.

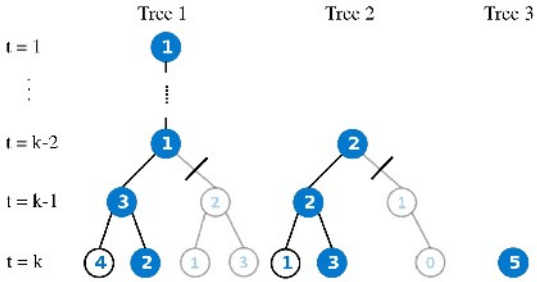
Topic 13 - Stefan

- **Multiple Hypothesis Tracking Revisited**
  Kim et al., ICCV 2016
- Track multiple objects by building up possible track trees
- Performance depends on maintaining small set of tracks
- Here: Exploit on-line appearance models for each hypothesis branch

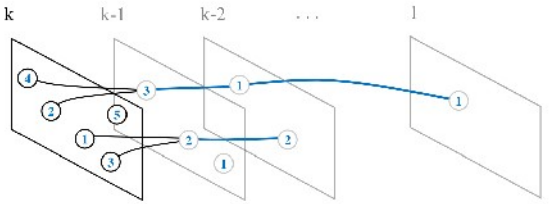


(a) Tracks in Video Frames





(b) N -scan Pruning

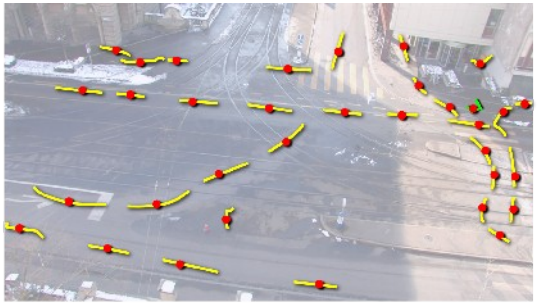


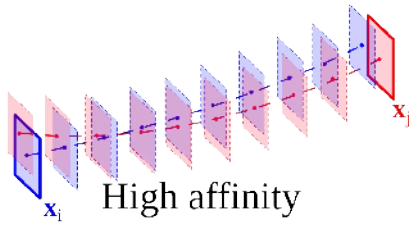
(c) Remaining Track Hypotheses

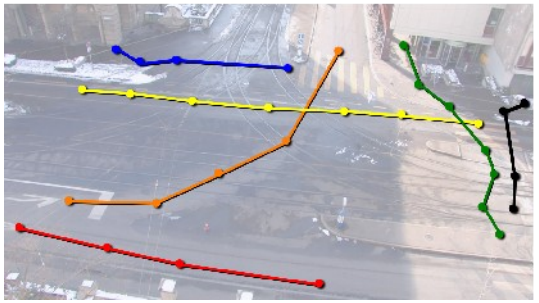
Topic 14 - Stefan

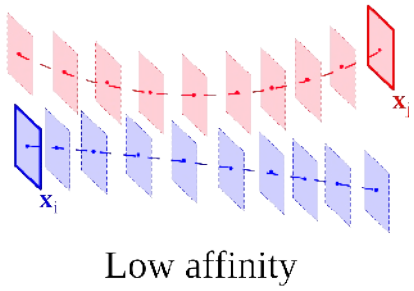
- **Leveraging Single- for Multi-Target Tracking Using a Novel Trajectory Overlap Affinity Measure**

  Manen et al., WACV 2016
- Use single-target tracking to find short robust tracklets
- Link tracklets to continuous tracks for multi-target tracking
- Here: Use trajectory overlap affinity measure for reasonable tracks









Topic 15 - Wolfgang

- **Co-operative Pedestrians Group Tracking in Crowded Scenes Using an MST Approach**

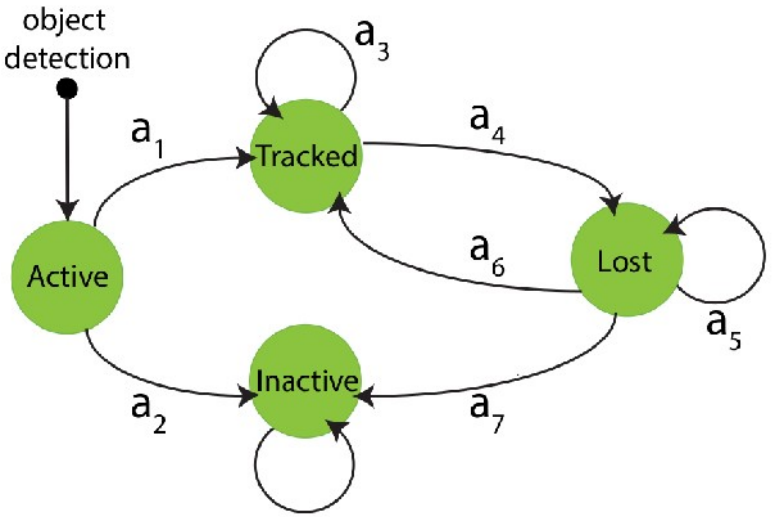
  Setia and Mittal, WACV 2015

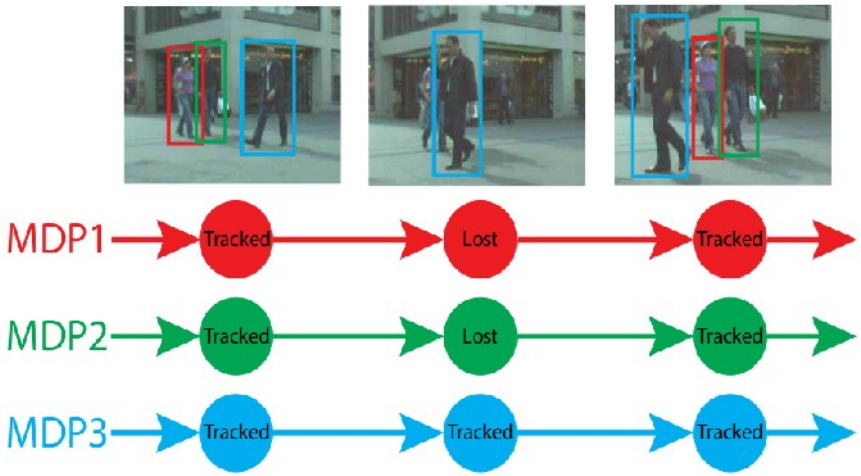
- In Tracking-by-Detection, groups can cause huge problems...
  - Because of how detectors work
- Try and address groups explicitly to overcome this

Topic 16 - Wolfgang

- **Learning to Track: Online Multi-Object Tracking by Decision Making**
  Xiang et al., ICCV 2015

- Model Tracking as a Markov Decision Process
- Explicitly handle Birth/Death/Occlusion/Reappearance
- Learn the Policies by Reinforcement Learning
  - Rather than designing them by hand





Topic 17 - Jörg

- **DynamicFusion: Reconstruction and Tracking of Non-rigid Scenes in Real-Time**

  Newcombe, Fox, and Seitz, CVPR 2015

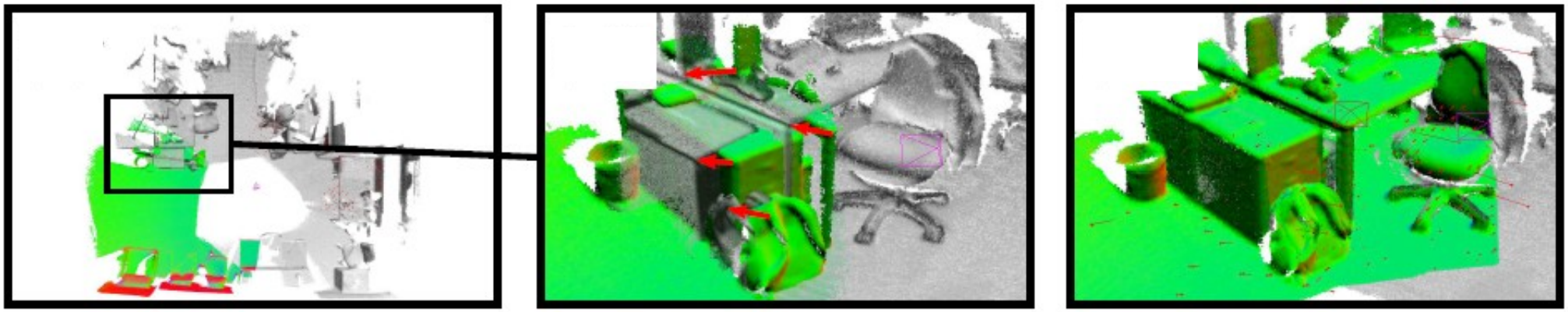
- **Aim:** Tracking and mapping of deformable objects with RGB-D sensors
- **Approach:**
  - Incremental fusion in a canonical TSDF representation
  - Estimation of deformation between canonical TSDF and current frame

Topic 18 - Jörg

- **ElasticFusion: Dense SLAM Without A Pose Graph**

  Whelan, et al., RSS 2015

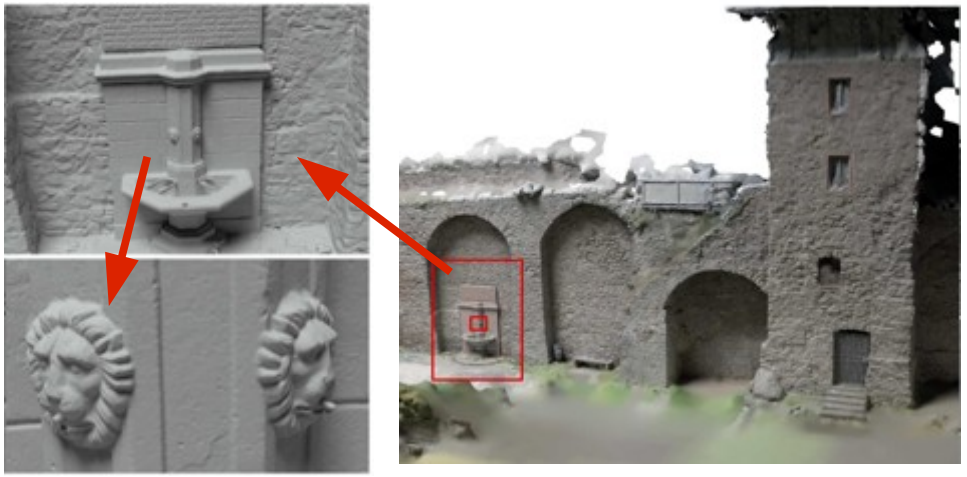
- **Aim:** RGB-D SLAM with consistent maps
- **Approach:**
  - Fusion in surfel maps
  - “Elastic” map adjustment after loop-closures

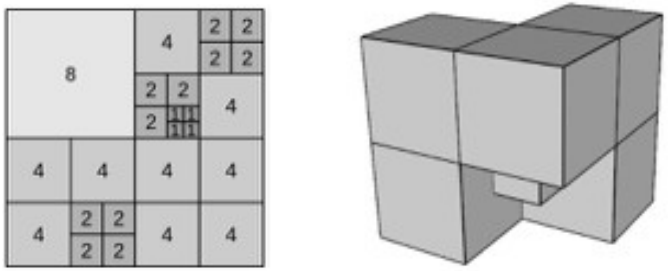


Topic 19 - Francis

- **Global, Dense Multiscale Reconstruction for a Billion Points**
  Ummenhofer and Brox, ICCV 2015

- **Goal:** Reconstruct large-scale scenes with objects of very different sizes
- **Motivation:** Very detailed surface reconstruction at different scales
  - How to represent the scene?
  - Efficient storage?
  - How to perform globally optimal inference?





signed distance function

$$E(u, \mathbf{v}) = \lambda_1 E_{\text{data}_u} + \lambda_2 E_{\text{data}_\mathbf{v}} + \alpha_1 E_{\text{coupling}} + \alpha_2 E_{\text{smooth}}$$

normal vector field

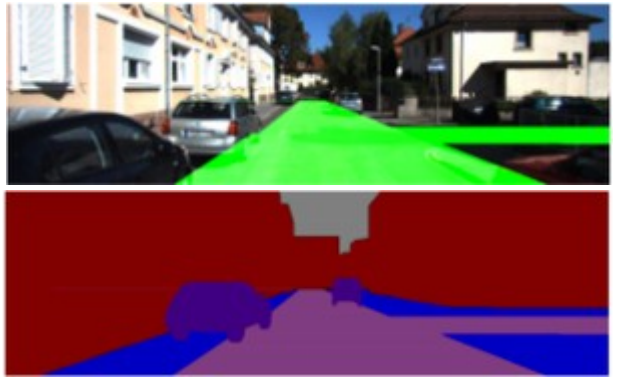
Topic 20 - Francis

- **Enhancing Road Maps by Parsing Aerial Images Around the World**
  Mattyus et al., ICCV 2015

- **Goal:** Exploit aerial images to enhance roads maps (e.g. OpenStreetMap)
- **Motivation:** Self-driving cars need detailed maps for navigation and localization
  - Allows for improved scene understanding
  - Formulated as energy minimization problem
  - Parameterized using centerline and width of road


(Toronto: Pearson Airport)


(Mexico City)



Assignment

- Pick three topics you might find interesting
  - No preference, just pick three
- Then we assign the topics
  - *Which algorithm can we use?*
- Let us quickly review the topics...
  - Remember the numbers!