Current Topics in Computer Vision and Machine Learning

Introductory Meeting

Prof. Dr. Bastian Leibe

Lehr- und Forschungsgebiet Informatik 8 (Computer Vision)
RWTH Aachen University
http://www.vision.rwth-aachen.de
Organization

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

• **Presentations**
  – Preferred in English – German possible
  – 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 4th**

• Report: Monday, **June 8th** – graded version!

• Slides: Monday, **July 13th**

• Presentations: **July – 21-23** – 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 - 2

• Mind credible sources
  – Papers published in journals or conference proceedings
    ▪ Peer reviewed == reliable and good
  – Wikipedia
    ▪ Can be altered by anyone and it changes over time == not good

• Use the original sources
  – Instead of sources that only cite 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
Computing the Stereo Matching Cost with a Convolutional Neural Network
J. Zbontar, Y. LeCun, arXiv 2014

- **Input**: pair of images from the left and right camera
- **Output**: dense disparity map
- State-of-the-Art Results: 2nd on KITTI
Topic 2 - Francis

- Efficient Joint Segmentation, Occlusion Labeling, Stereo and Flow Estimation
  K. Yamaguchi, D. McAllester and R. Urtasun, ECCV 2014

- **Input**: Two pairs from stereo video sequence

- **Output**
  - Flow image
  - Boundary labels
  - Disparity image

- **State-of-the-Art**: 3rd on KITTI
• **Image Parsing via Stochastic Scene Grammar**  
  Zhao, Zhu, NIPS 11  
• Parsing indoor scenes:  
  – Compute 3D box layout of the room  
  – Detect 3D objects (e.g. furniture)  
  – Detect 2D faces (e.g. windows, door)  
• Stochastic scene grammar to represent hierarchical structure of visual entities  
• Hierarchical cluster sampling through Markov Chain Monte Carlo for inference
• **Understanding Indoor Scenes using 3D Geometric Phrases**  
  Wongun Choi, Yu-Wei Chao, Caroline Pantofaru, and Silvio Savarese  
  CVPR 2013

- Explain scene semantic, geometry, and object grouping in unified manner.
- Modelling of geometric and semantic relationships between co-occurring objects as a parse graph.
- Search of an hypothesis through energy minimization using Markov Chain Monte Carlo.
• **To Aggregate or Not to aggregate:**
  Selective Match Kernels for Image Search
  Giorgos Tolias, Yannis Avrithis, Hervé Jégou
  ICCV 2013

• Typical approaches for descriptor matching:
  – BoW + tf-idf score + spatial verification (or query expansion)
  – Hamming Embedding, binary codes for more precise description
  – VLAD for alternative encoding scheme

• **Paper’s approach:** Tries an aggregation of the state of the art above approaches to combine the best ingredients
Topic 6 - Theodora

- **DisLocation**: Scalable descriptor distinctiveness for location recognition
  R. Arandjelović, A. Zisserman
  Asian Conference on Computer Vision, 2014

- Feature matching problem
- Direct feature matching: expensive
- Alternatives: BoW, Hamming Embedding …
- **Paper's approach**: Defines descriptor “distinctiveness” based on density estimation and uses it to weight the descriptors
• **Stixmantics: A Medium-Level Model for Real-Time Semantic Scene Understanding**
  Timo Scharwächter, Markus Enzweiler, Uwe Franke, Stefan Roth
  ECCV‘14

• Visual semantic scene understanding of urban traffic scenes
  Important for car industry

• Joint scene, object class and motion inference

• Stixel-based scene representation
Topic 8 - Aljosa

- **Combining 3D Shape, Color, and Motion for Robust Anytime Tracking**
  RSS'14 (Robotics: Science and Systems)
- Joint 3D object tracking and shape integration (mainly) based on laser data
- Combine 3D shape, color and motion cues
- Real-time
- Does it generalize to 3D stereo data?

![Diagram](image)
Topic 9 - Umer

- **Pose Machines: Articulated Pose Estimation via Inference Machines**
  Ramakrishna, Munoz, Hebert, Bagnell, Sheikh, ECCV 14

- State-of-the-art approaches use part based models.
- Use tree structured models to make inference tractable.
- Inference often intractable or too slow in more complex models.
- Inference machines present an alternative to capture richer spatial interactions among multiple parts with efficient inference.
Topic 10 - Umer

- **Articulated Pose Estimation using Discriminative Armlet Classifiers**
  Gkioxari, Arbeleaz, Bourdev, Malik, CVPR 13

- Predicting the right pose of both arms still remains a challenging problem in Human Pose Estimation.
- Use discriminative armlet classifiers that are inspired by poselets.
- Integrate skin color, contour and contextual cues to get additional improvements in performance.
Occlusion Geodesics for Online Multi-Object Tracking
H. Possegger, T. Mauthner, P. M. Roth, H. Bischof
CVPR’14

occluded (undetected) pedestrians hard to track online
re-detection of lost pedestrians, solely by geometric cues
shortest paths through occlusion regions (occlusion geodesics)
• Context-Based Pedestrian Path Prediction
  J. F. P. Kooij, N. Schneider, F. Flohr, D. M. Gavrila
  ECCV 14

• intelligent vehicle should decide:
  is a pedestrian walking on the road? should I break?

• incorporate situational awareness, situation criticality, spatial layout

• inference on graphical model (Dynamic Bayesian Network, Switching Linear Dynamical System)
Hybrid Stochastic / Deterministic Optimization for Tracking Sports Players and Pedestrians
R. T. Collins, P. Carr, ECCV 14

- reliably track sport players using (static) RGB image, foreground + occupancy map
- split tracking problem in stochastic detection search and deterministic data association
- reversible jump move Markov Chain Monte Carlo method (RJMC MC)
Topic 14 - Wolfgang

- **Activity Forecasting**
  Kris M. Kitani, Brian D. Ziebart, James Andrew Bagnell, Martial Hebert
  ECCV 2012

- Learn about the structure of the scene (sidewalks, ...)
- Use knowledge to predict destinations/future paths of pedestrians

*Homotopy classes*  
*Our approach*
Topic 15 - Michael

- Re-ranking by Multi-feature Fusion with Diffusion for Image Retrieval
  Yang, Fan, Bogdan Matei, and Larry S. Davis

- Fusion of multiple retrieval methods
- Weighted graph-based approach for retrieval
- Data-driven query dependent weights
• **Neural Codes for Image Retrieval**
  Babenko, et al.
  ECCV 2014

• Learned features instead of designed
• Discriminative dimensionality reduction
• **ImageNet Classification with Deep Convolutional Neural Networks**
  Krizhevsky, Sutskever, Hinton, NIPS 2012

• Tackle a challenging large-scale benchmark task ILSVRC12
• Very influential work
• Investigation of a number of helpful techniques
• Foundation of a very successful line of work – GoogleLeNet
• Pedestrian Detection with Unsupervised Multi-Stage Feature Learning
  Sermanet, Kavukcuoglu, Chintala, LeCun, CVPR 13

• Initialization for Deep Learning setups is tricky
• Stagewise pre-training method – to initialize a neural net
• **Unsupervised** – no labeling costs – in contrast to supervised auxiliary task

Feature Map