Current Trends in Machine Learning

Preparation Meeting

Rudolph Triebel,
Vladislav Usenko,
Caner Hazirbas

What you will learn in the seminar

- Get an overview on current trends in machine learning
- Read and understand scientific publications
- Write a scientific report
- Prepare and give a talk

Important Dates

- First Meeting: 8.10.2014 (today)
  - Fix assignment of papers and date
- Choose your topic until 15.10.2014 (next week, first come first serve!)
- Deadline for the report: 27.02.2015
- Dates for the talks:
  - 7.01.2015
  - 14.01.2015
  - 21.01.2015
  - 28.01.2015

Preparation

- Please do not work on your topic completely alone

→ Meet at least twice with your supervisor

- Recommended schedule
  - 1 month before your talk: Meet your supervisor and discuss paper
  - 1 week before your talk: Meet your supervisor to discuss your slides
  - [optional] after the talk: Feedback of your supervisor regarding the talk
  - 1 week before 28.02.14: Submit a draft of your report

Report and Talk

- Send PDF (not PPTX, not DOC) via email to your supervisor, Latex template available on the web-page
- Recommended length: 6-8 pages
- Required: Minimum 6, Maximum 10 pages
- Language: English or German

Hints for Your Talk

- 20 min. + 5–10 min. for discussion
- Don’t put too much information on one slide
    - 1-2 min. per slide → 10-20 slides
- Recommended structure
    - Introduction, Problem Motivation, Outline
    - Approach
    - Experimental results
    - Discussion
    - Summary of (scientific) contributions

Evaluation Criteria

- Gained expertise in the topic
- Quality of your talk
- Quality of the report
- Active participation in the seminar is required (ask questions, comment talks)

Regular Attendance Is Required

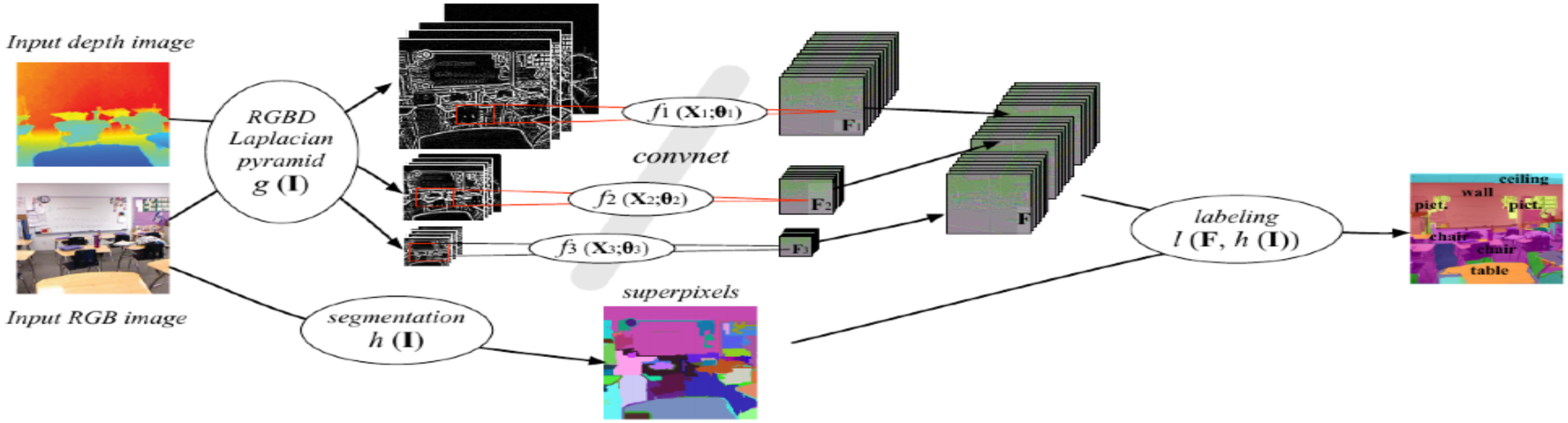
- Attendance at each appointment is necessary
- In case of absence: Medical attest

Overview of available Topics

Indoor Semantic Segmentation using Depth Information

[Couprie et al., ICLR 2013]

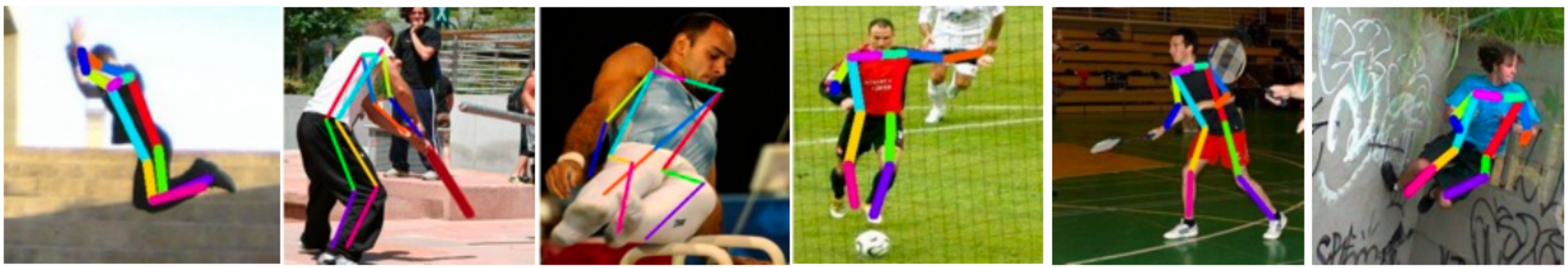
- Multi-class segmentation of indoor scenes with RGB-D inputs
- Multi-scale feature extraction (Convolutional Networks, Deep Learning)

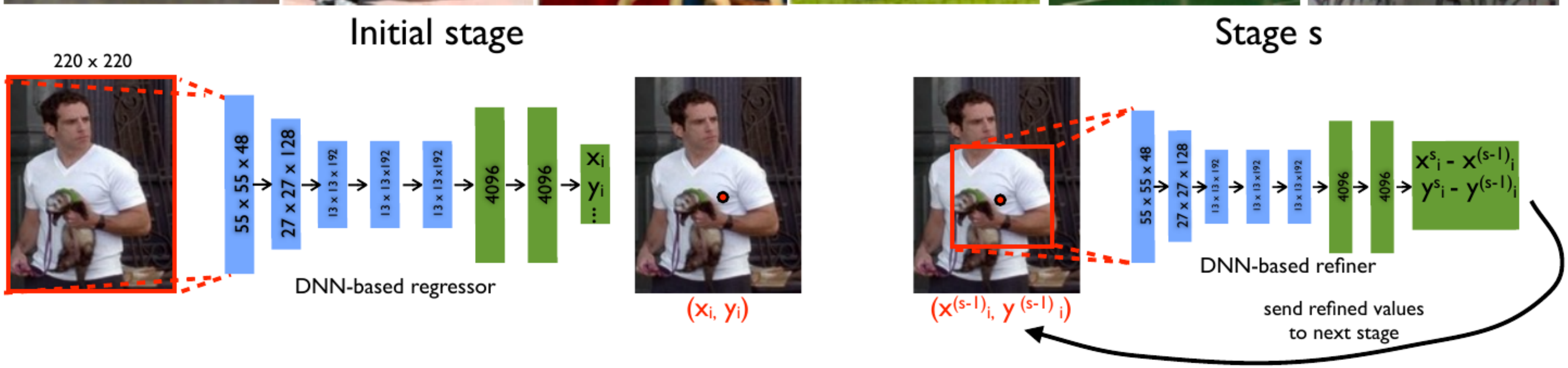


Deep Pose: Human Pose Estimation via Deep Neural Networks

[Toshev and Szegedy, CVPR 2014][Google]

- Human pose estimation based on DNN
- DNN based pose regression

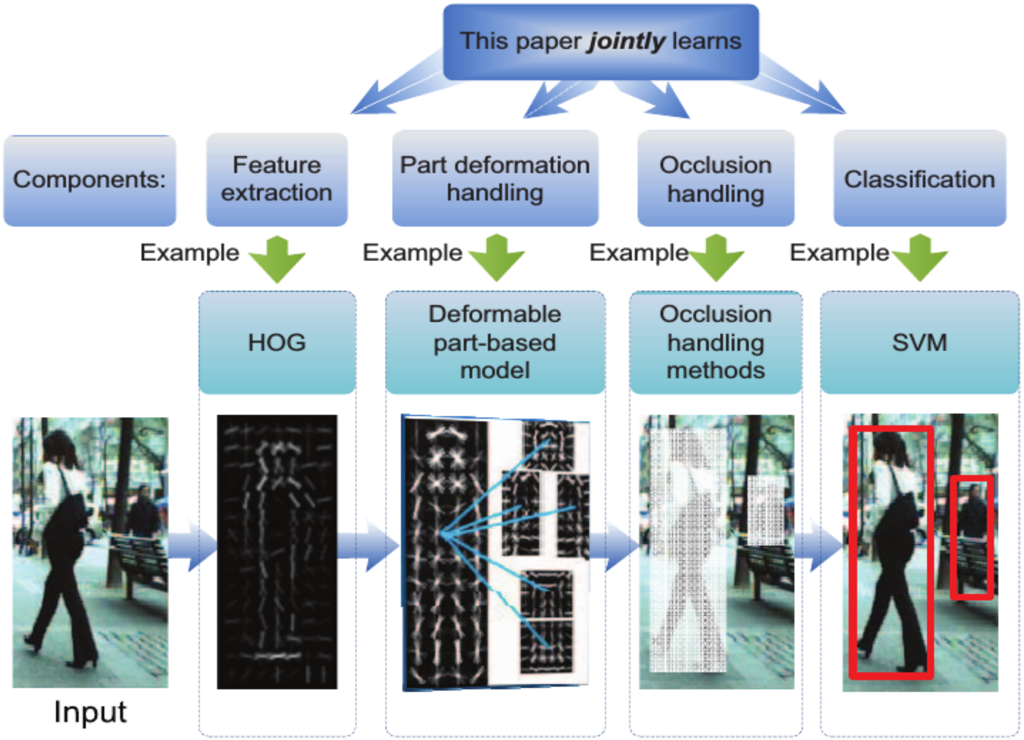




Joint Deep Learning for Pedestrian Detection

[Ouyang and Wang, ICCV 2013]

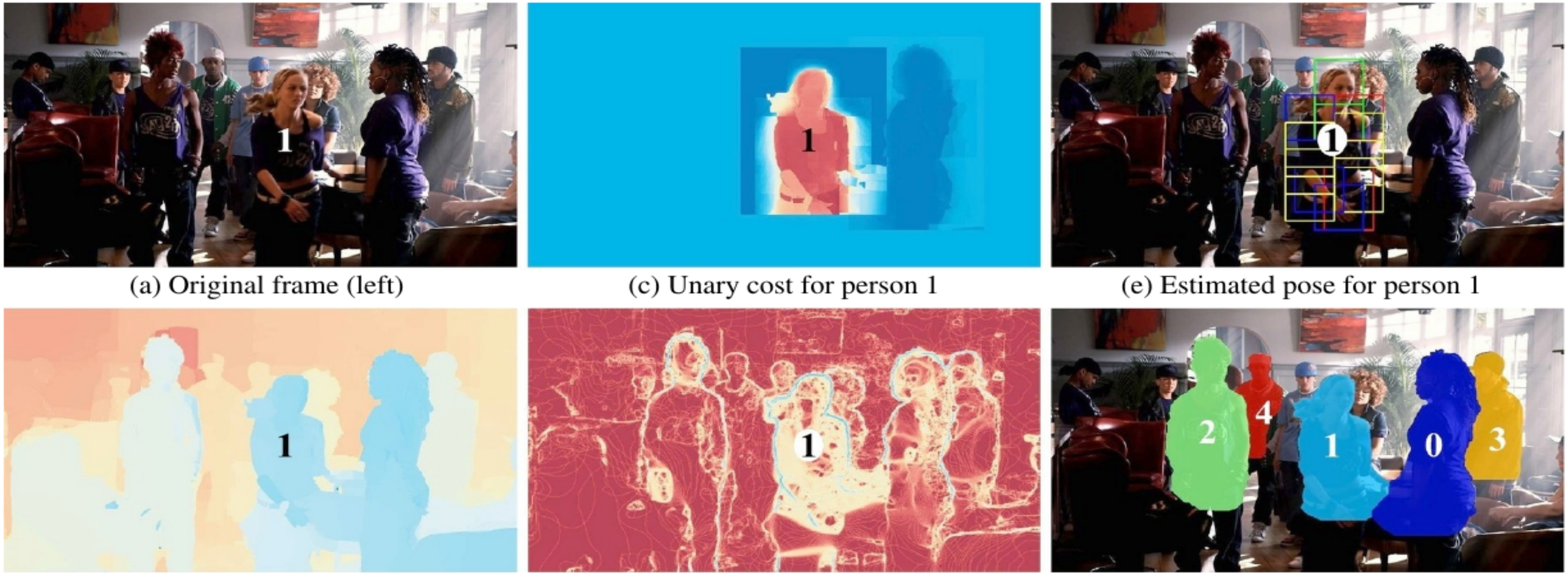
- Pedestrian Detection
- Deep Learning framework



Pose Estimation and Segmentation of People in 3D Movies

[Alahari et al., ICCV 2013]

- Pixel-wise segmentation of multiple people
- Pose estimation
- Stereoscopic videos



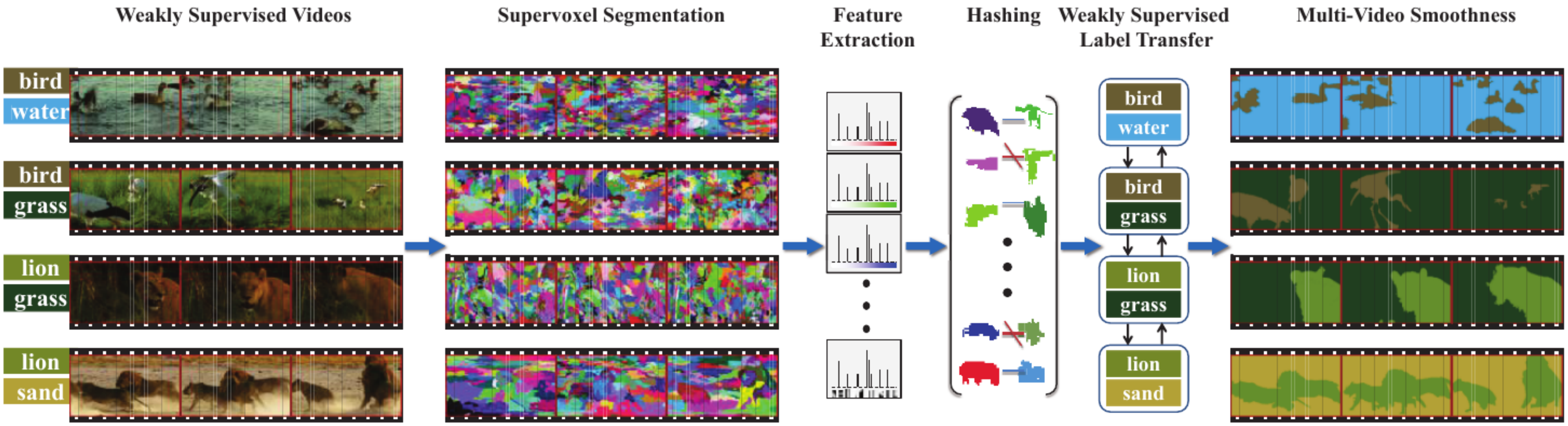
(a) Original frame (left) (c) Unary cost for person 1 (e) Estimated pose for person 1

(b) Disparity (d) Smoothness cost (f) Segmentation result

Weakly Supervised Multiclass Video Segmentation

[Liu et al., CVPR 2014]

- Nearest neighbor-based label transfer scheme for weakly supervised video segmentation
- Hashing for metric learning handles both metric and semantic similarity



Unsupervised Spectral Dual Assignment Clustering of Human Actions in Context

[Jones and Shao, CVPR 2014]

- Unsupervised human action clustering using contextual relations between actions and scenes
- Dual Assignment k-Means (DAKM)
- Spectral DAKM for realistic data

Diving | Basketball, Tennis, Volleyball | Basketball, Golf, Juggling | Biking, Walking, Riding, Juggling

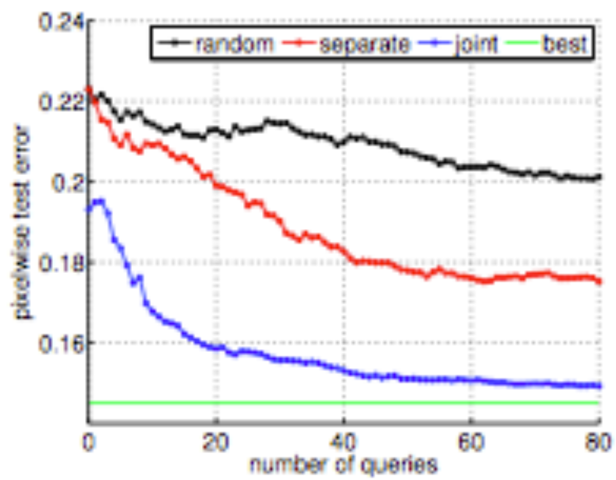
Latent Structured Active Learning

[Wenjie Luo, Schwing, Urtasun]

- Combine structured prediction problems with active learning
- Use weakly supervised learning
- Application to predict the 3D layout of rooms from single images

Algorithm 1 latent structured prediction

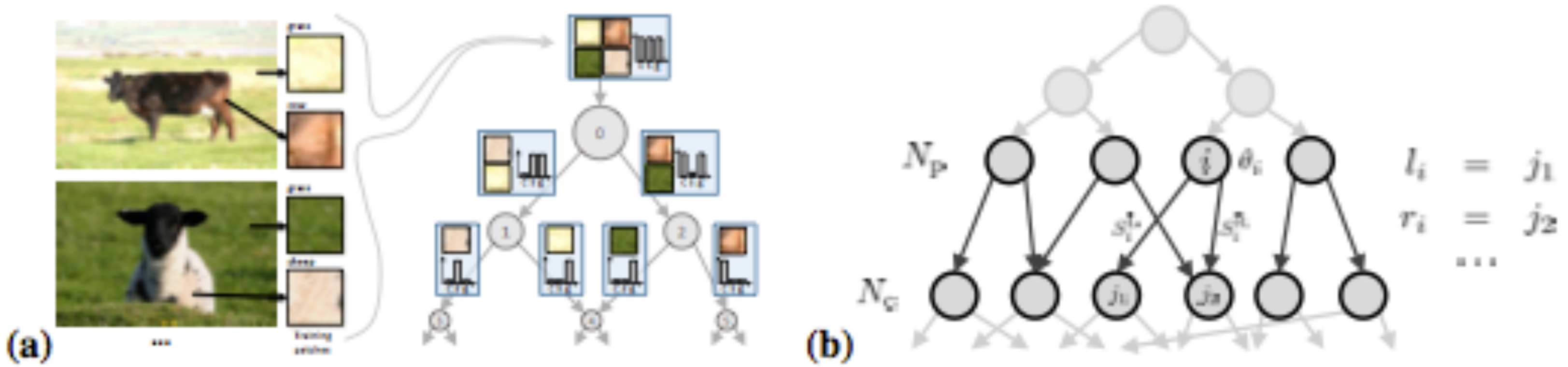
Input: data $\mathcal{D}$, initial weights w
repeat
  repeat
    //solve latent variable prediction problem
    $\min_d f_2 + f_3$ s.t. $\forall(x, y)\ d_{(x,y)} \in \mathcal{D}_{(x,y)}$
  until convergence
  //message passing update
  $\forall(x, y), i \in \mathbb{S} \quad \lambda_{(x,y),i} \leftarrow \nabla_{\lambda_{(x,y),i}}(f_1 + f_2) = 0$
  //gradient step with step size η
  $w \leftarrow w - \eta\nabla_w(f_1 + f_2)$
until convergence
Output: weights w, beliefs d



Decision Jungles: Compact and Rich Models for Classification

[Shotton et al.]

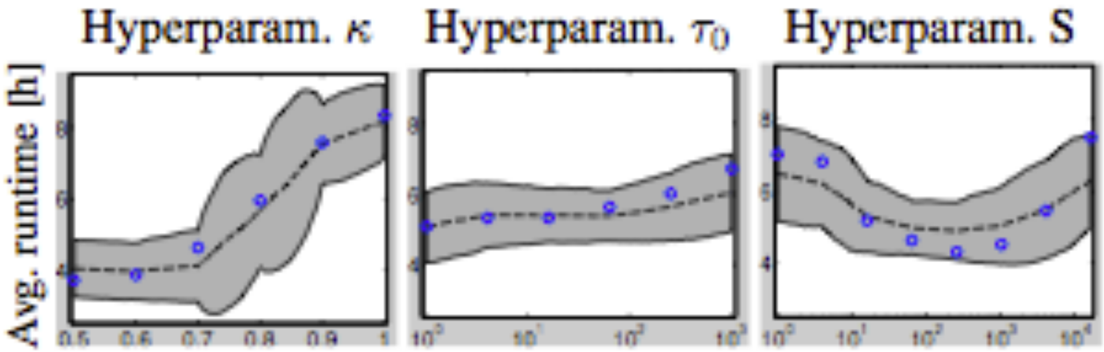
- Extension of Decision trees to DAGs
- Learning features and structure using node merging
- Results in less memory and better generalization



An Efficient Approach for Assessing Hyperparameter Importance

[Hutter el al.]

- Hyperparameters are crucial elements in most learning approaches
- Propose an efficient marginalization technique for random forest predictions
- Use that to quantify the importance of hyper parameters



SOML: Sparse Online Metric Learning with Application to Image Retrieval

[Gao et al.]

- Learn sparse distance functions from high-dimensional data
- More efficient than non-sparse metric learning
- Application to Image Retrieval

Algorithm 1 SOML-TG—Sparse Online Metric Learning via Truncated Gradient

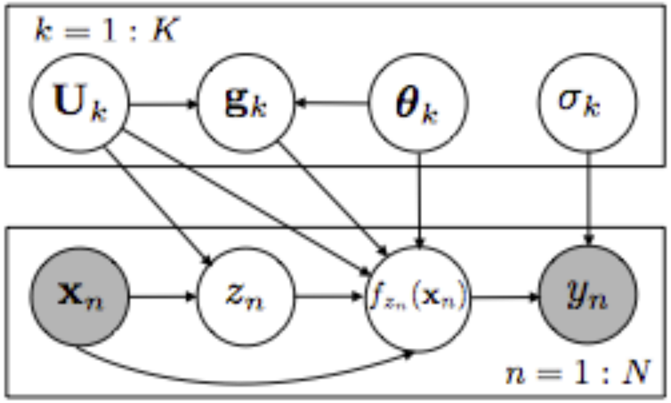
Input: Training Triplets: $(\mathbf{x}, \mathbf{x}_t^+, \mathbf{x}_t^-)$, $t = 1, \ldots, n$.
Output: The weight vector $\mathbf{w}$.

Initialize $\mathbf{w}_1 = 0$; $\alpha = \eta\lambda$
repeat
  Receive a triplet instance $(\mathbf{x}_t, \mathbf{x}_t^+, \mathbf{x}_t^-)$,
  Suffer loss $\mathcal{L}((\mathbf{x}_t, \mathbf{x}_t^+, \mathbf{x}_t^-); \mathbf{w}_t)$ measured by (7)
  if $\mathcal{L}((\mathbf{x}_t, \mathbf{x}_t^+, \mathbf{x}_t^-); \mathbf{w}_t) > 0$ **then**
    $\mathbf{v} = \mathbf{w}_t - \eta[\mathbf{x}_t \odot (\mathbf{x}_t^+ - \mathbf{x}_t^-)]$;
    for j=1 to m **do**
      if $\mathbf{v}_j \geq 0$ **then**
        $\mathbf{w}_{t+1,j} = \max(0, \mathbf{v}_j - \alpha)$;
      else
        $\mathbf{w}_{t+1,j} = \min(0, \mathbf{v}_j + \alpha)$;
      end if
    end for
  end if
until CONVERGENCE

Fast Allocation of Gaussian Process Experts

[Nguyen and Bonilla]

- Non-parametric Bayesian regression using a mixture of GPs
- Fast variational inference procedure for learning hyper parameters
- Results on large-scale data set with 10^5 training points



Automatic Construction and Natural-Language Description of Nonparametric Regression Models

[LLoyd et al.]

- Non-parametric (GP) formulation of functions that model high-level properties, e.g. smoothness
- Compositional structure of the “language of models”
- Results in an automated way to describe data

This component is approximately periodic with a period of 10.8 years. Across periods the shape of this function varies smoothly with a typical lengthscale of 36.9 years. The shape of this function within each period is very smooth and resembles a sinusoid. This component applies until 1643 and from 1716 onwards.

This component explains 71.5% of the residual variance; this increases the total variance explained from 72.8% to 92.3%. The addition of this component reduces the cross validated MAE by 16.82% from 0.18 to 0.15.

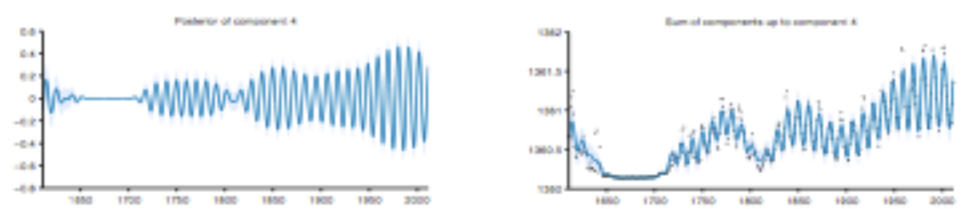
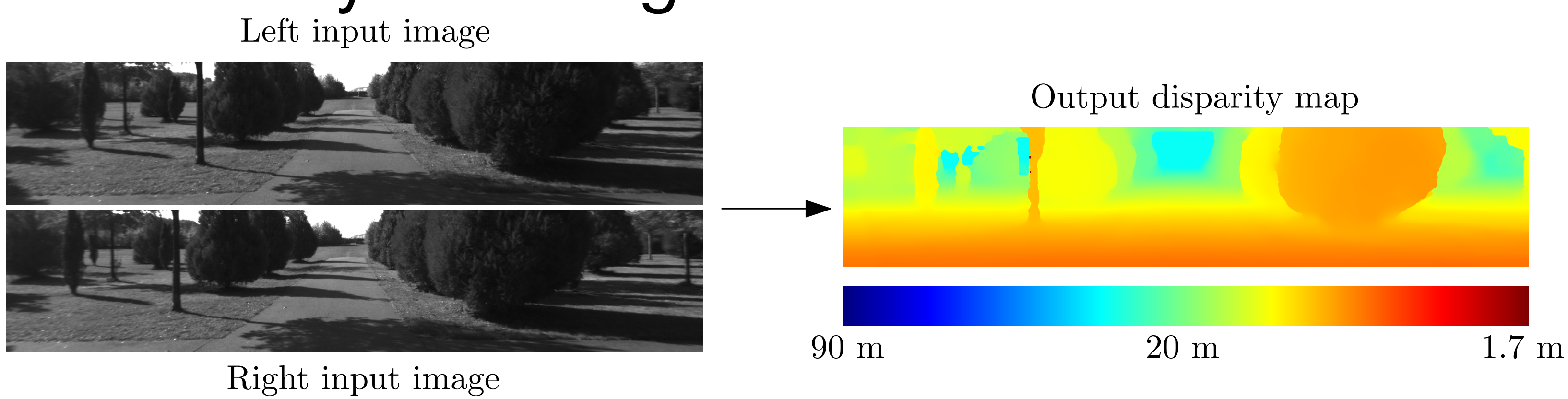


Figure 8: Pointwise posterior of component 4 (left) and the posterior of the cumulative sum of components with data (right)

Computing the Stereo Matching Cost with a Convolutional Neural Network

[Jure Zbotar, Yann LeCun]

- Estimate depth information for each pixel from stereo images
- Train Convolutional Neural Network to predict how well image patches match
- Refine disparity using semi-global matching and consistency checking



Autonomous Active Recognition and Unfolding of Clothes using Random Decision Forests and Probabilistic Planning

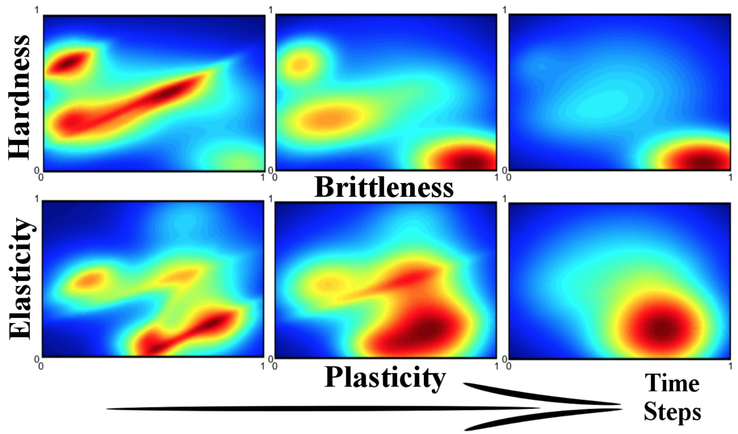
[Doumanoglou et al]

- Clothes recognition from depth images using Random Decision Forests
- Unfolding an article of clothing by estimating and grasping key points identified using Hough Forests

Learning Haptic Representation for Manipulating Deformable Food Objects

[Mevlana Gemici, Ashutosh Saxena]

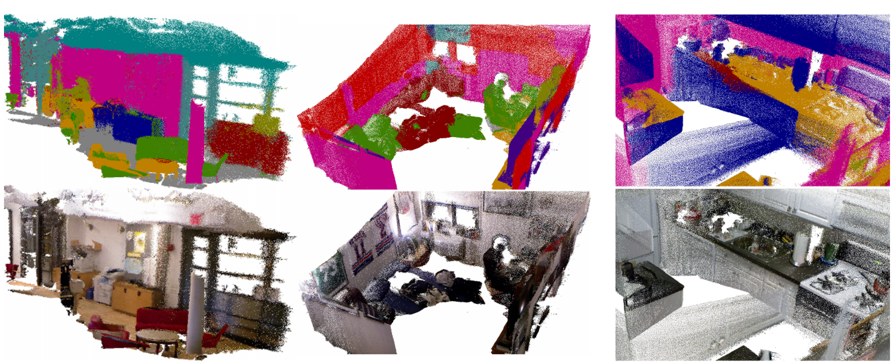
- Design actions that obtain haptic data with information about physical properties of the object
- Extract features based on manipulators configuration, effort and dynamics
- Discover Haptic Categories through Dirichlet Process

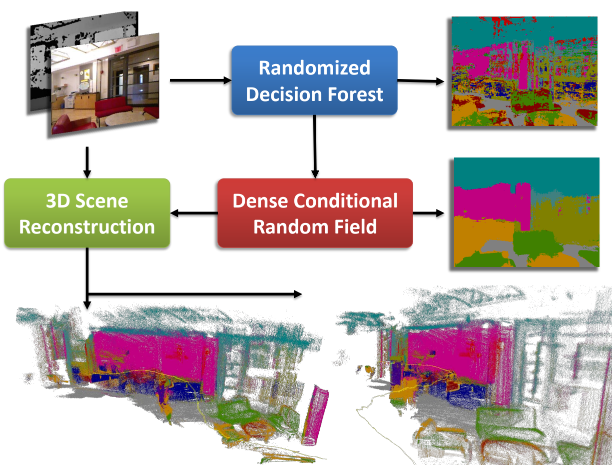


Dense 3D Semantic Mapping of Indoor Scenes from RGB-D Images

[Alexander Hermans, Georgios Floros, Bastian Leibe]

- Consistent 3D semantic reconstruction of indoor scenes
- Fast 2D semantic segmentation approach based on Randomized Decision Forests
- 2D-3D label transfer based on Bayesian updates and dense pairwise 3D Conditional Random Fields.

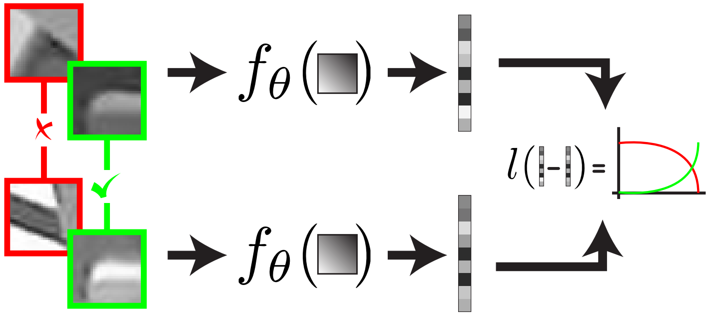


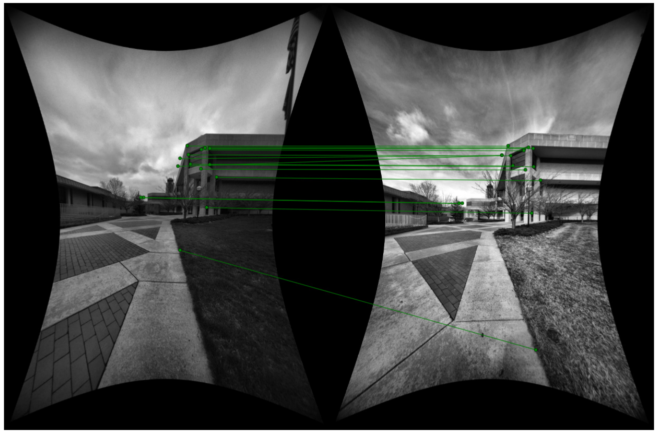


Learning Visual Feature Descriptors for Dynamic Lighting Conditions

[Nicholas Carlevaris-Bianco, Ryan Eustice]

- Use Machine Learning to find descriptors that can be used for long-term outdoor deployments
- Track feature points in time-lapse videos to acquire training data
- Convolutional multi-layer perceptron used as descriptor





Enjoy the seminar!