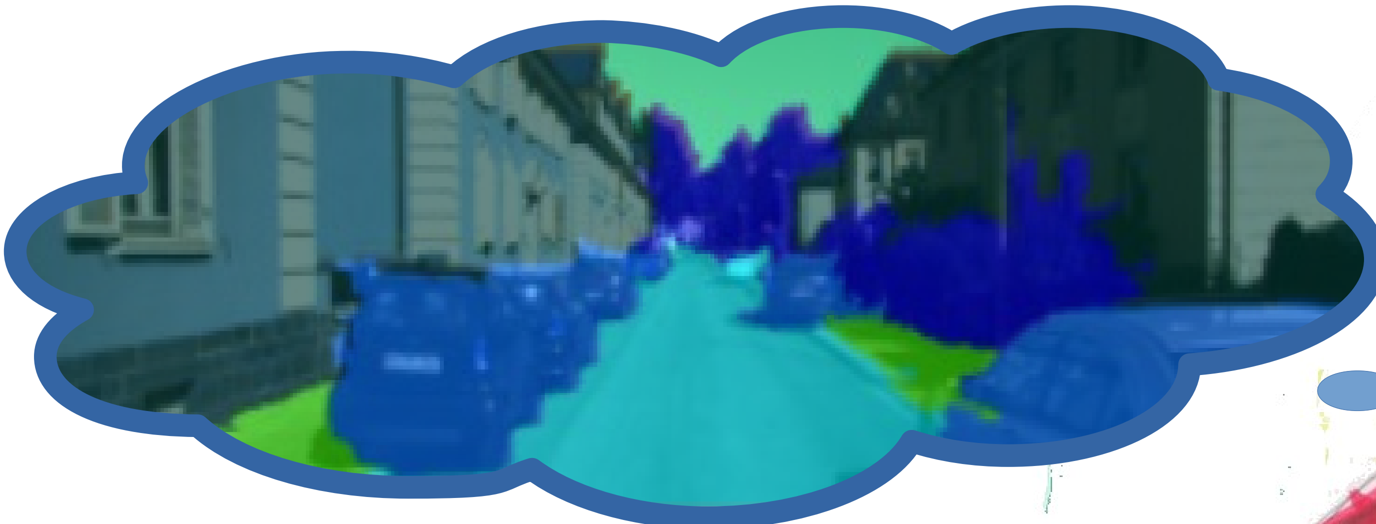
Multi class object detection for smart cars and safe roads



hrzz
Hrvatska zaklada za znanost

FER

Siniša Šegvić
UniZg - FER

About the project

Goal

- scalable multi-class object detection from the driver's perspective

Methodology

- state of the art classification and reconstruction approaches
  - Fisher vectors, DNN's, dense stereo and SfM

Frame:

- fully supported by Croatian Science Foundationunder the grant I-2433-2014
- 1 postdoc, 1 PhD student
- 01.10.2014 – 01.10.2017

Applications:

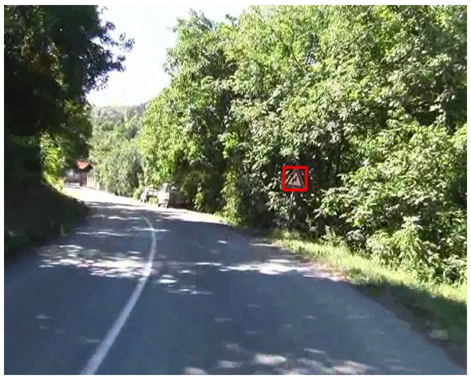
- smart vehicles (autonomous driving, driver's assistance)
- intelligent transportation systems (road safety inspection)

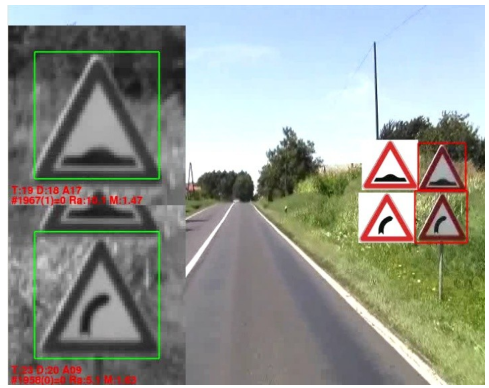
Members:

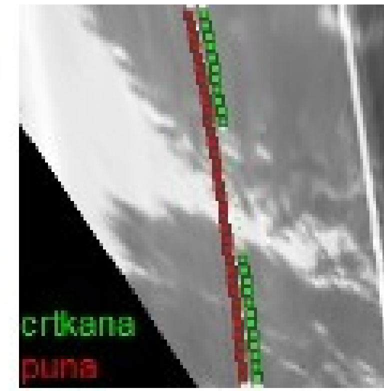
- Zoran Kalafatić, Josip Krapac, Ivan Krešo, Ivan Sikirć, Valentina Zadrija, Marko Ševrović, ...
- Axel Pinz, Albert Diosi, ...

Previous projects

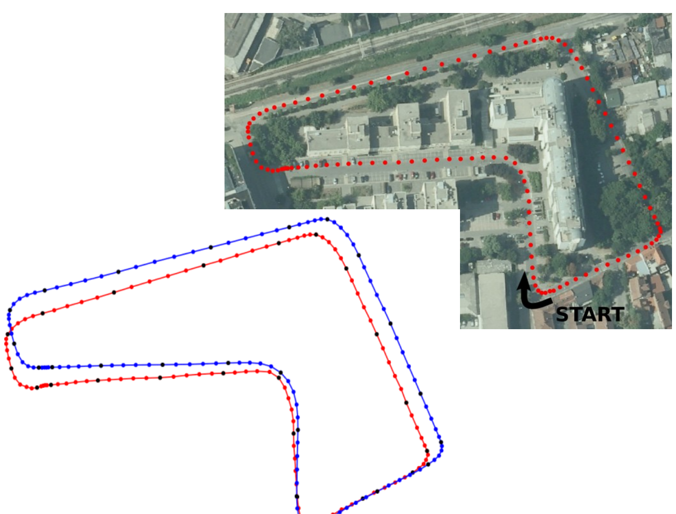
MASTIF - mapping and assessing traffic infrastructure (HrZZ - IPV, 2008-2011, 1 PhD student)

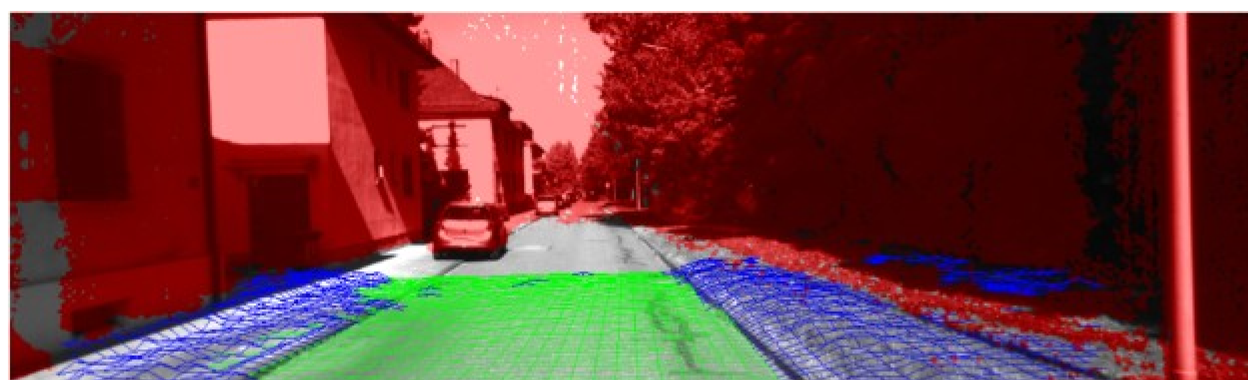






VISTA - computer vision innovations for safe traffic (EU IPA IIIc, 2013-2015, 2 PhD students)



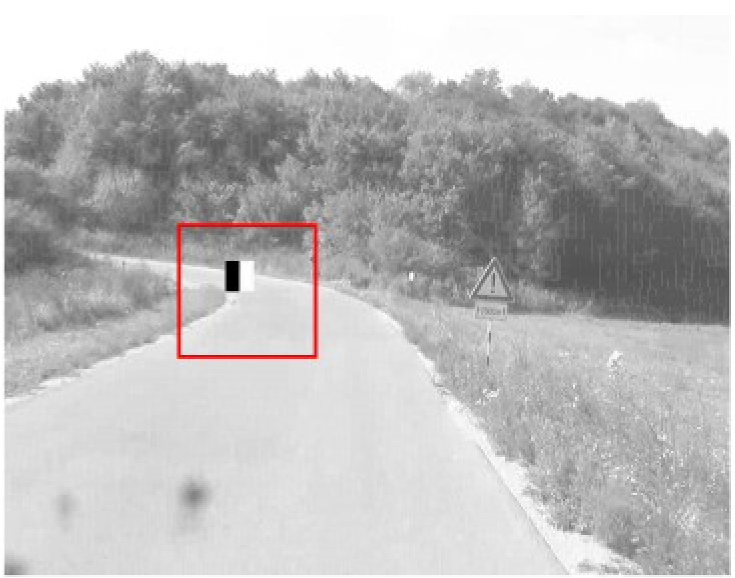


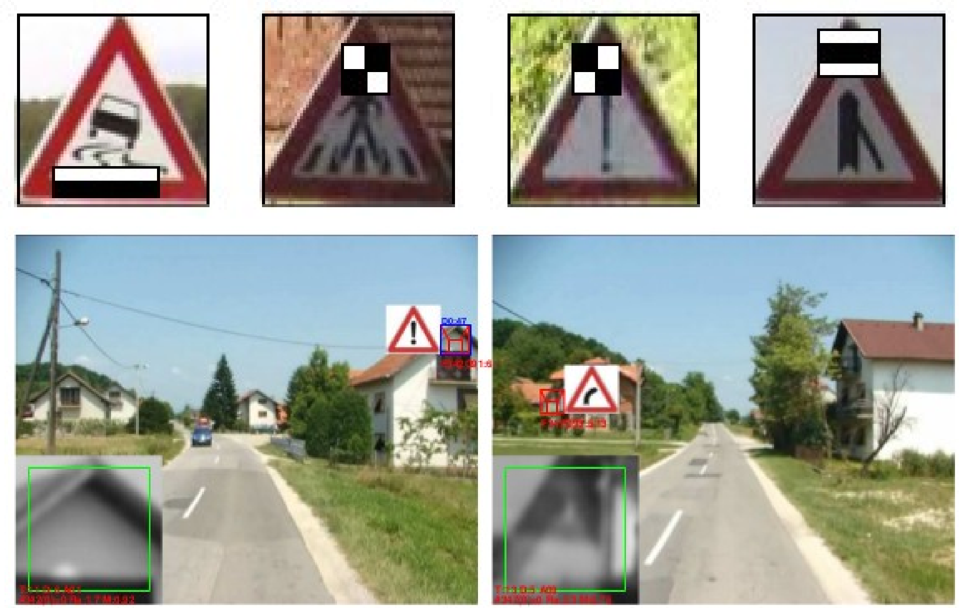
Agenda

- brief presentation of the project
- object localization
- weakly supervised localization with large Fisher vectors
- fast approximate GMM soft-assign
- semantic segmentation
- classification of traffic scenes
- discussion

Object localization

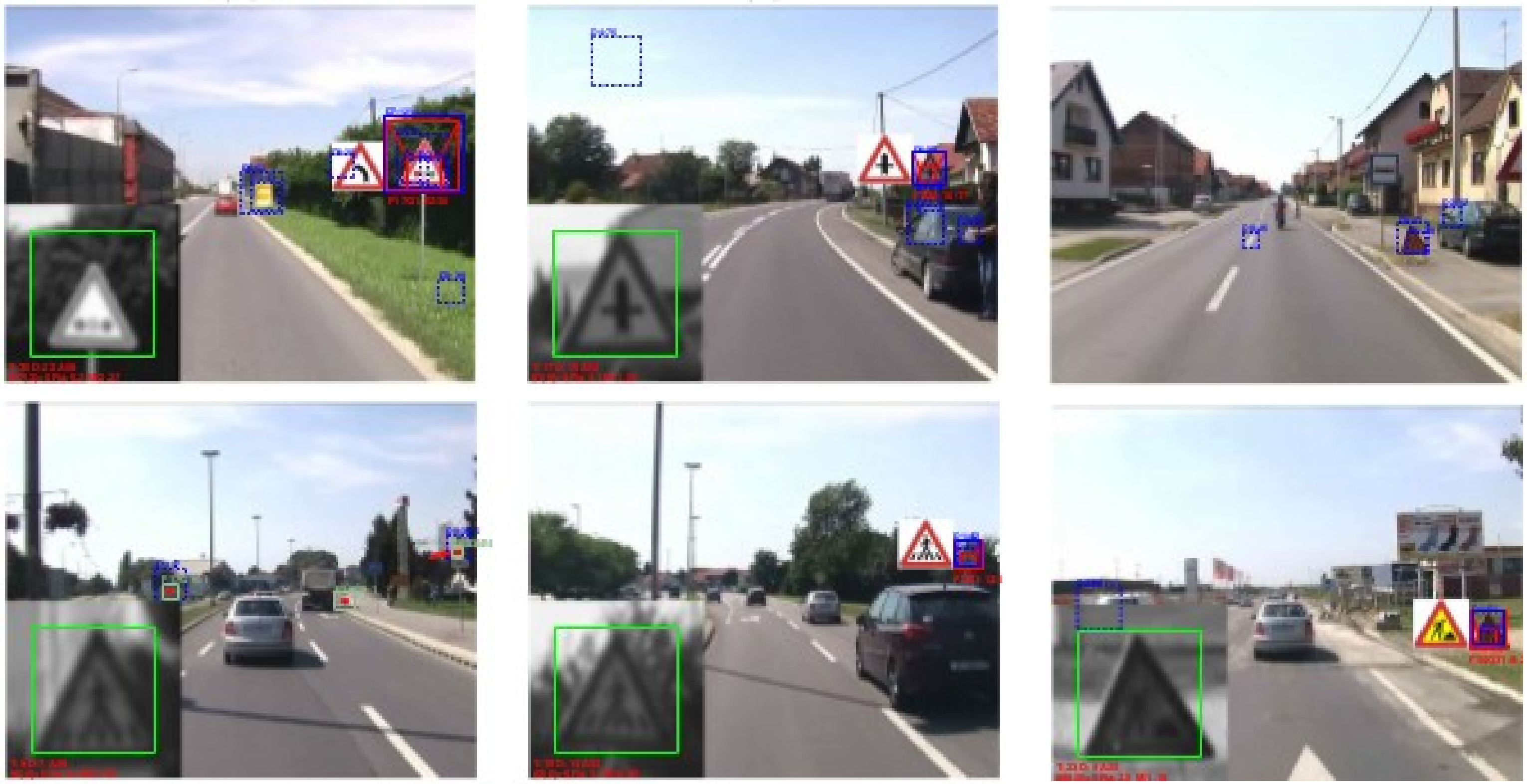
- Where are the objects of the desired class(es) (if any)?
- Classification in the sliding window
  - apply a fast classifier at 100000 image locations
  - simple features and cascaded classifiers
  - near ideal performance on traffic signs [syroco2012, mva2014]
  - further developments (DPM) allow deformable objects
  - poor performance in the multi-class scenario





Object localization (2)

- Some results [mva2014] (T: 2000, E: 1000):



- Other approaches
  - weakly supervised localization [visapp2015jk, gcpr2015vz]
  - bottom-up cues: bottom-up segmentation, trained objectness
  - classify each image pixel independently (=semantic segmentation!)
    - localize objects by detecting connected components with the corresponding labe

The quest: scalable multiclass performance

- cascades have poor representation capacity
- several alternative approaches:
  - powerful supervised features (e.g. overfeat [sermanet14iclr])
    - state of the art classifier (CNN) trained on ImageNet
    - huge complexity (can be alleviated with bottom-up proposals)
  - unsupervised middle-level features (FV) [cinbis13iccv]
    - considerable computational complexity
    - inherently multi-class due to generative modelling involved
    - classification models can be trained with **weak supervision**
  - pixel-level output (semantic segmentation) [farabet13pami]
    - considerable computational complexity
    - spatial layout is problematic
- in our research we explore the latter two options

Weakly supervised localization with unsupervised middle-level features

- train the model on images annotated with class labels
- use groundtruth locations only for performance evaluation
- benefit: able to collect more training samples with less effort
- difficulties: very small objects, rich background

Localization with weak supervision (2)

- Our approach:
  - disregard any bottom-up proposals
  - represent patch appearance with large Fisher vectors
  - speed-up the computations by:
    - sparse classification model [krapac15visapp]
    - **novel fast GMM soft-assign** [krapac15gcpr]
  - improve performance by:
    - taking into account non-linear normalizations [zadrija15gcpr]
    - represent patch layout with spatial Fisher vector [zadrija15gcpr]
    - enforcing the group sparsity of the localization model [krapac15gcpr]

Weakly supervised localization results

Table 2. Localization performance. T denotes the number of patches used to compute the object bounding box. K_w denotes the number of non-zero model components. p_{miss} denotes the miss frequency at the rightmost data point of the PR curve.

Nr.	Configuration	FV Normalization	Penalty	K_w	T	AP test	p_{miss}
1	S HOG [11]	-	l_2	-	-	**88**	**0.05**
2	M [22]	-	l_1	64	100	72	0.13
3	M	-	group	11	100	74	0.25
4	M	p, ℓ_2 intra	group	7	100	77	0.11
5	G	p, ℓ_2 intra	group	7	100	77	0.16
6	G + SH	p, ℓ_2 intra	group	7	all	75	0.14
7	G + SFV	p, ℓ_2 intra	group	7	all	**81**	**0.11**

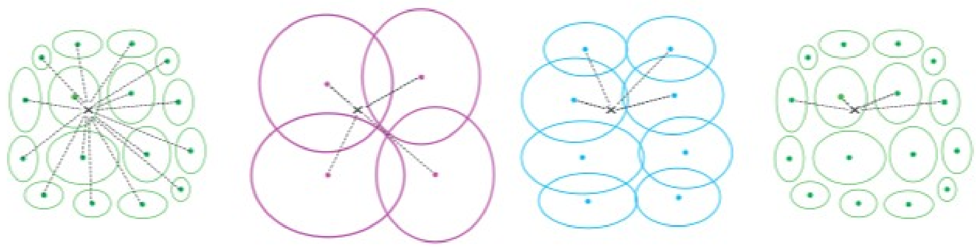
Current and future work

– make this work for articulated classes (cars, pedestrians, etc.)

– evaluate multiclass performance

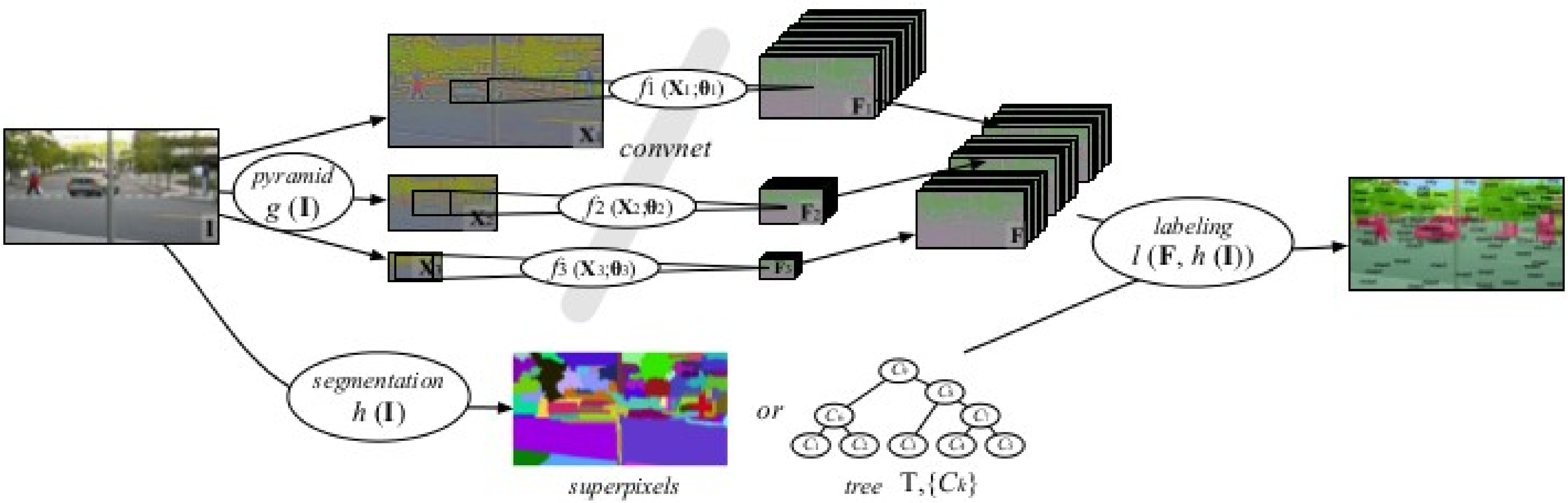
Fast approximate GMM soft assign

- GMM soft assign: the bottleneck of many classification and localization approaches based on visual words and Fisher vectors
- the idea: represent the GMM as a tree obtained by recursive agglomerative clustering
- improve the performance for an order of magnitude without a noticable performance hit



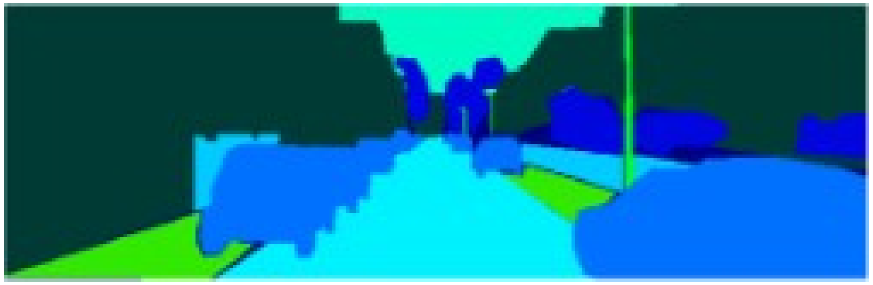
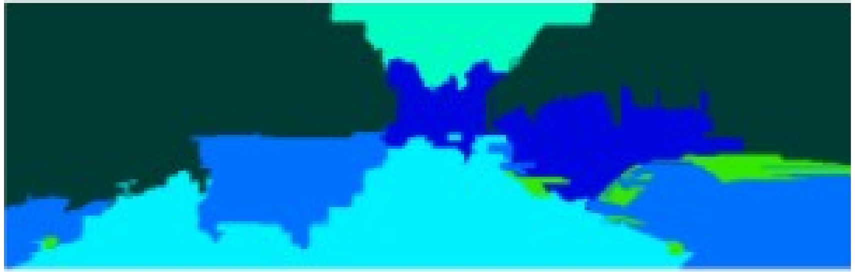
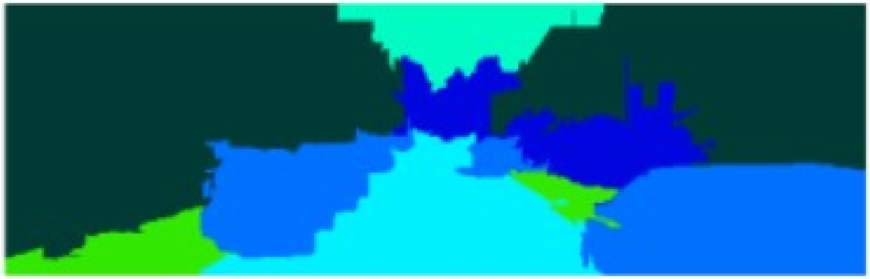
Semantic segmentation

- Assign a semantic label to each pixel
  - related to object classification
  - state of the art: texton boost, CNN
- State of the art CNN architecture [farabet13pami]
  - similar to the vanilla classification CNN
  - differences: preprocessing, multi-scale, no fully connected layer, postprocessing



Semantic segmentation (2)

- CNN results on RGBD KITTI (T: 100, E: 46):
  - 620 x 192 pixels, 8 classes: vehicle, road, building, sidewalk, vegetation, sky, traffic sign, pedestrian
  - RGB pixel accuracy: 73.6%, 75.1%
  - RGBD pixel accuracy: 78.5%, 79.1%
  - class accuracy: below 50%
  - execution speed (GTX980): 20Hz, 10 Hz

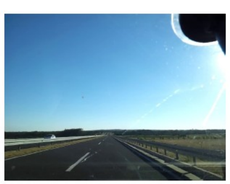
Impact of the hardware

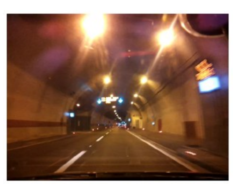
- Desktop GPU (GTX 980, 500W): 20 Hz
- Laptop CPU (Core i5, 100 W): 0.2 Hz
- Laptop GPU (100W): 4 Hz
- NVIDIA Tegra X1 (5W): 4 Hz (estimate)

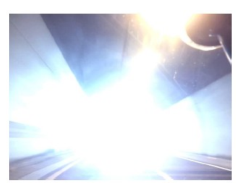
Image classification (2)

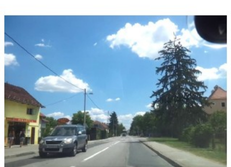
- State of the art approaches:
  - soft quantization (GMM), representation in the gradient space of the generative model (FV) [bmvc2011,iccv2011]
  - convolutional neural networks (CNN) jointly train the quantization and classification
- Example: classification on a representation budget [iv2014]
  - 8 scene classes important for fleet management
  - best performance (AP): SFV + RF, GIST + SVM
  - AP over 90% even for rb=64 (T: 1500, V: 1500, E: 3000)


(a) autocesta

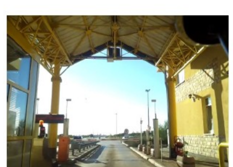

(b) cesta


(c) tunel

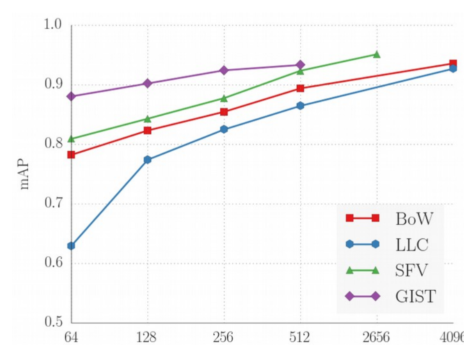

(d) izlaz iz tunela

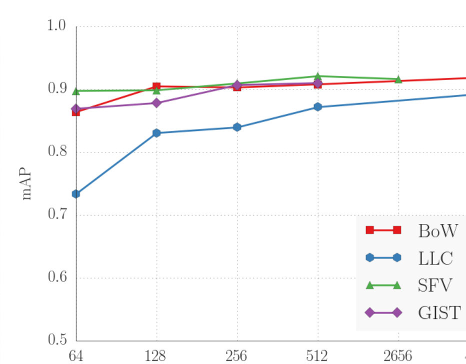

(e) naselje


(f) nadvožnjak


(g) naplatna kućica


(h) promet





Discussion

Bibliography

[eccv2006] Sinisa Segvic, Anthony Remazeilles, François Chaumette: Enhancing the Point Feature Tracker by Adaptive Modelling of the Feature Support. ECCV (2) 2006: 112-124

[cviu2009] Sinisa Segvic, Anthony Remazeilles, Albert Diosi, François Chaumette: A mapping and localization framework for scalable appearance-based navigation. Computer Vision and Image Understanding 113(2): 172-187 (2009)

[itsc2010] Sinisa Segvic, Karla Brkic, Zoran Kalafatic, Vladimir Stanisavljevic, Marko Sevrovic, Damir Budimir and Ivan Dadic. A computer vision assisted geoinformation inventory for traffic infrastructure. ITSC 2010: 66-73

[cvpr2010] Josip Krapac, Moray Allan, Jakob J. Verbeek, Frédéric Jurie: Improving web image search results using query-relative classifiers. CVPR 2010: 1094-1101

[iccv2011] Josip Krapac, Jakob J. Verbeek, Frédéric Jurie: Modeling spatial layout with fisher vectors for image categorization. ICCV 2011: 1487-1494

[bmvc2011] Josip Krapac, Jakob J. Verbeek, Frédéric Jurie:Learning Tree-structured Quantizers for Image Categorization. BMVC 2011: 1-11

[syroco2012] Sinisa Segvic, Zoran Kalafatic, Ivan Kovacek: Sliding Window Object Detection without Spatial Clustering of Raw Detection Responses. SyRoCo 2012: 114-119.

[iv2014] Ivan Sikiric, Karla Brkic, Josip Krapac, Sinisa Segvic: Image representations on a budget: Traffic scene classification in a restricted bandwidth scenario. Intelligent Vehicles Symposium 2014: 845-852

[mva2014] Sinisa Segvic, Karla Brkic, Zoran Kalafatic, Axel Pinz: Exploiting temporal and spatial constraints in traffic sign detection from a moving vehicle. Mach. Vis. Appl. 25(3): 649-665 (2014)

[icmr2014] Josip Krapac, Florent Perronnin, Teddy Furon, Hervé Jégou: Instance classification with prototype selection. ICMR 2014: 431

[visapp2015ik] Ivan Kreso, Sinisa Segvic: Improving the Egomotion Estimation by Correcting the Calibration Bias. VISAPP (3) 2015: 347-356

[visapp2015jk] Josip Krapac, Sinisa Segvic: Weakly Supervised Object Localization with Large Fisher Vectors. VISAPP (2) 2015: 44-53

[gcpr2015vz] Valentina Zadrija, Josip Krapac, Jakob Verbeek, Sinisa Segvic: Patch-level Spatial Layout for Classification and Weakly Supervised Localization. GCPR 2015

[gcpr2015jk] Josip Krapac, Sinisa Segvic: Fast Approximate GMM Soft-Assign for Fine-Grained Image Classification with Large Fisher Vectors. GCPR 2015