On the use of Regions for Semantic Image Segmentation

Rui Hu[1], Diane Larlus[2], Gabriela Csurka[2]

[1] University of Surrey, UK [2] Xerox Research Center Europe

Tuesday December 18th 2012
ICVGIP 2012

xerox

Table of contents

Semantic Segmentation

- Unsupervised Image segmentation

Semantic Segmentation

- Unsupervised Image segmentation


- Semantic Image segmentation


xerox

Semantic Segmentation

- State-of-the-art semantic segmentation methods usually leverage
  - Local appearance of objects (class likelihood maps)
  - Local consistency (constraining neighboring labels)
  - Global consistency (image level priors)
- That are combined in
  - unified CRF framework
    [Verbeek & Triggs 2007, Kohli et al 2009, Ladicky et al 2009]
  - sequential framework
    [Yang et al 2007, Csurka & Perronnin 2008]

Semantic Segmentation

Recent methods use unsupervised image partition in Regions, or Super-Pixels to enhance semantic segmentation:

- Local appearance is predicted based on region descriptors

  [Gu et al 2009, Lim et al 2009, Vijayanarasimhan & Grauman 2011, Lucchi et al 2011]
- Local consistency is enforced within regions:
  - in a post processing step

    [Csurka and Perronnin 2008]
  - or using higher order potentials in the CRF

    [Kohli et al 2009, Ladicky et al 2009, Gonfaus et al 2010]

Semantic Segmentation

What is the best way to use regions ?

We propose a benchmark studying the role and benefit of regions at different stages of the segmentation process.

Table of contents

Proposed Benchmark

We propose a benchmark based on 3 components

- A standard dataset:

  MSRC-21 dataset

- A standard super-pixel method:

  Berkeley segmentation approach

- A standard pipeline:

  Fisher-Vector based patch classification
  Condition Random Field

MRSC-21 dataset

- Standard benchmark, 591 images:
  - 276 images for training
  - 59 images for validation
  - 275 images for testing
- 21 classes:
  building, grass, tree, cow, sheep, sky, aeroplane, water, face, car, bicycle, flower, sign, bird, book, chair, road, cat, dog, body, boat
- Evaluate pixel-level classification
  - Average class-based accuracy

[Shotton et al, IJCV 2009]

xerox

Berkeley segmentation method

- Unsupervised segmentation of the image at multiple levels
  [Arbalaez et al, CVPR 2009]

xerox

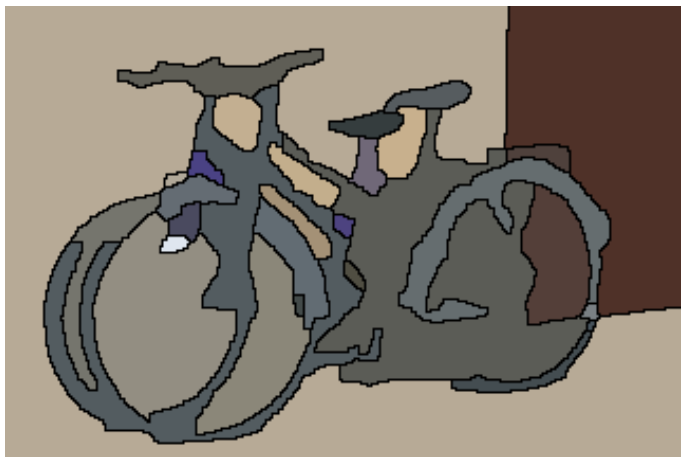
Berkeley segmentation method

- Unsupervised segmentation of the image at multiple levels
  [Arbalaez et al, CVPR 2009]

xerox

Berkeley segmentation method

- Unsupervised segmentation of the image at multiple levels
  [Arbalaez et al, CVPR 2009]



xerox

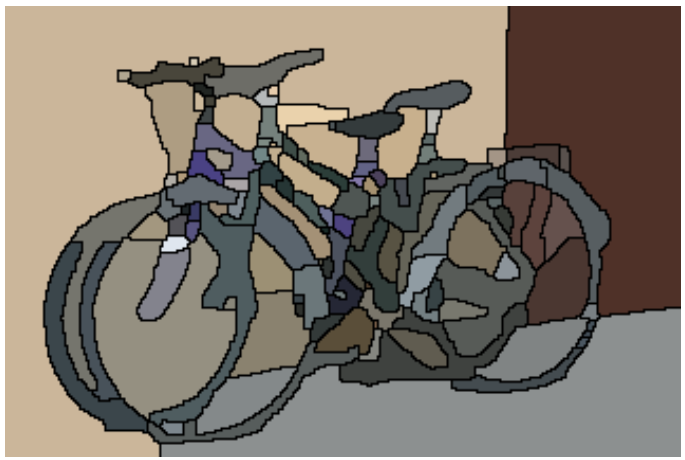
Berkeley segmentation method

- Unsupervised segmentation of the image at multiple levels
  [Arbalaez et al, CVPR 2009]

xerox

Berkeley segmentation method

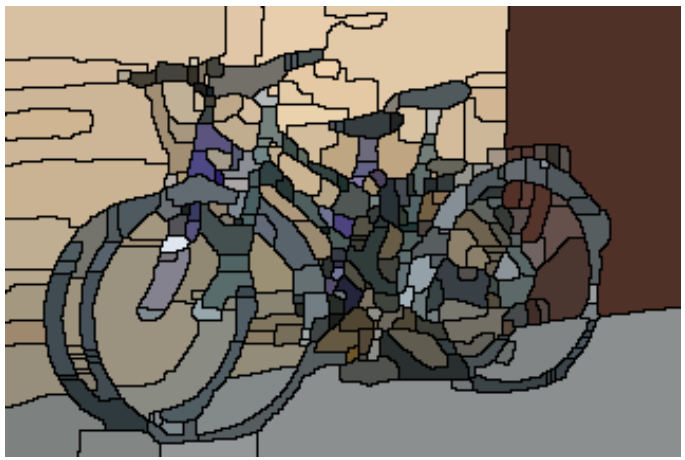
- Unsupervised segmentation of the image at multiple levels
  [Arbalaez et al, CVPR 2009]



xerox

Berkeley segmentation method

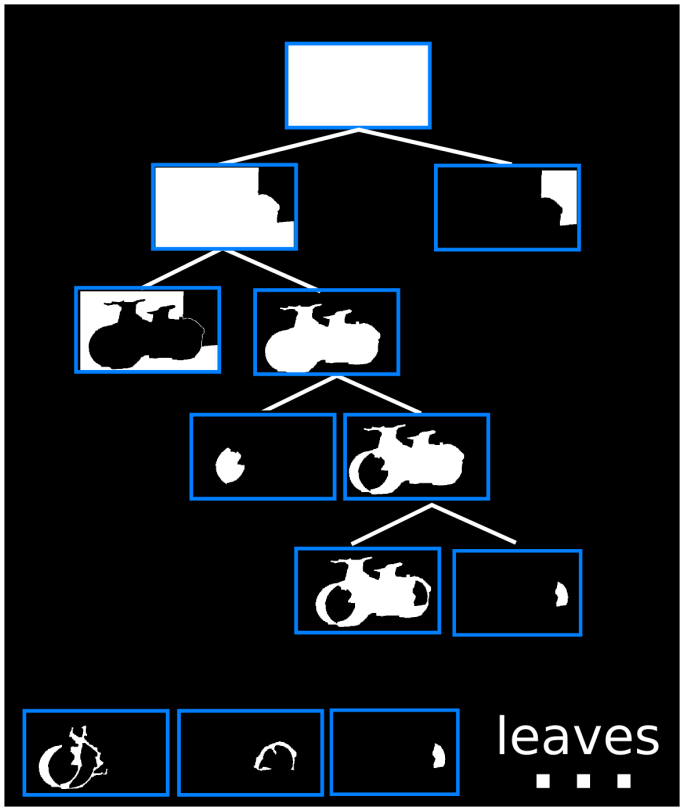
- Unsupervised segmentation of the image at multiple levels
  [Arbalaez et al, CVPR 2009]



xerox

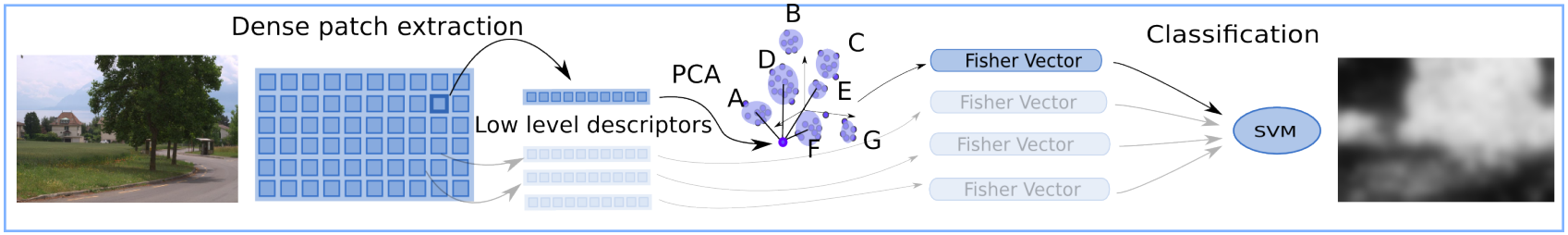
Berkeley segmentation method

- Unsupervised segmentation of the image at multiple levels
  [Arbalaez et al, CVPR 2009]



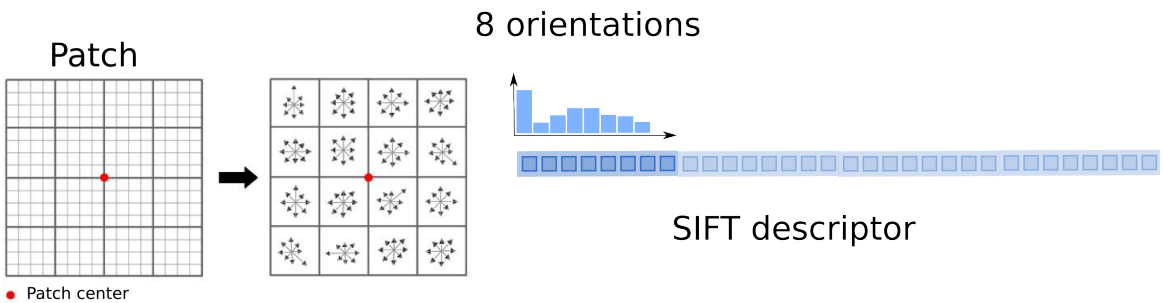
xerox

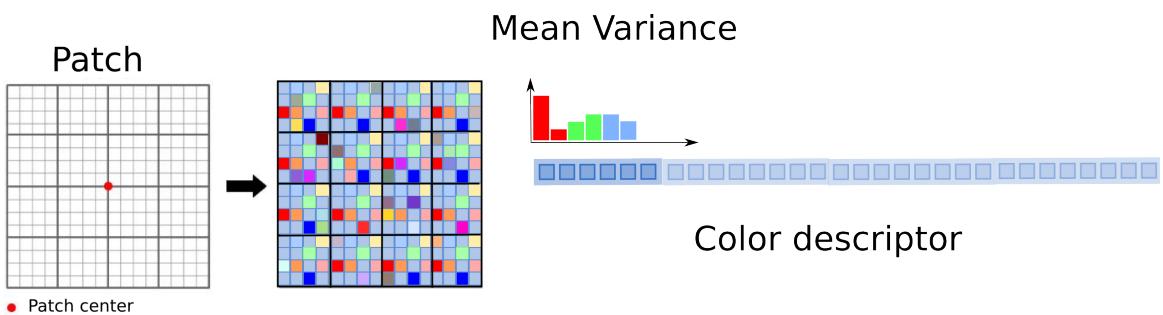
Patch-based Fisher Vector Representation



[Csurka and Perronnin, IJCV 2011]

Dense patch extraction at single scale or at 5 different scales, described using:





xerox

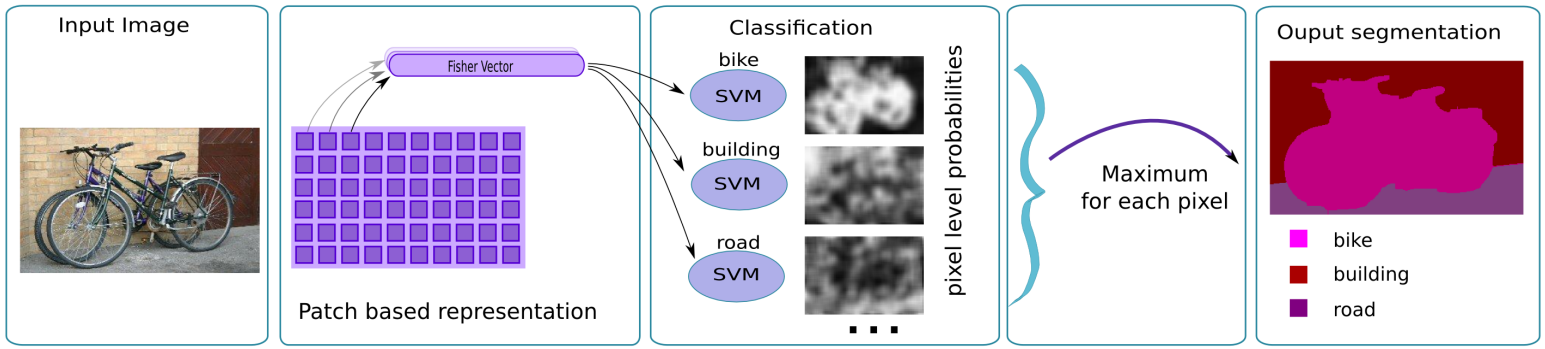
Conditional Random Field model

- Dense CRF model
  [Krähenbühl and Koltun, NIPS 2011]
  - Model with unary and pairwise potentials
    - Unary term: based on the patch-based FV classification
    - Pairwise term: all pairwise pixel connections are considered (not only 4 or 8 neighborhood systems)

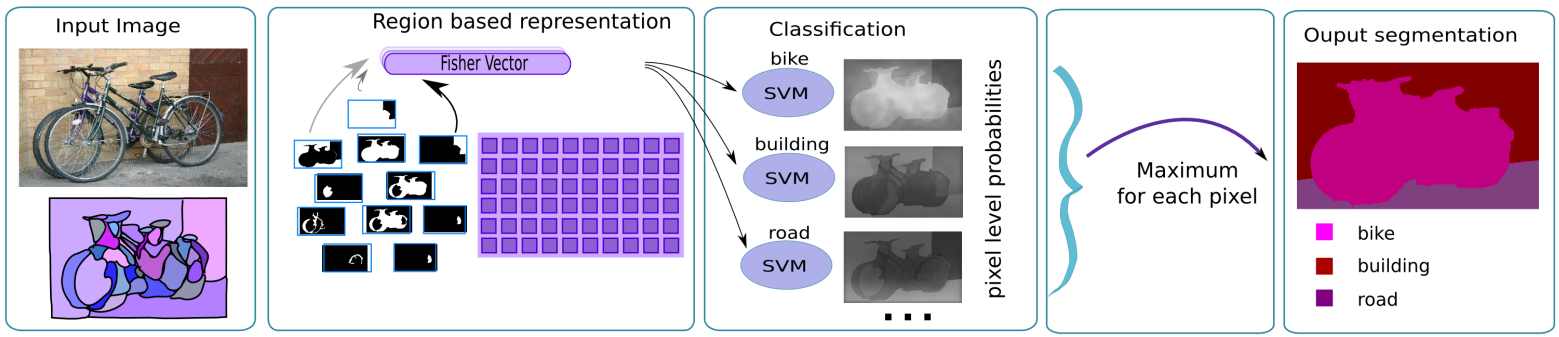
Table of contents

Appearance model

- Patch-based system: PB-SIS
  - Classify each patch individually
  - Accumulate patch probabilities at the pixel level



- Region-based system: RB-SIS
  - Aggregation of patches for each region of the hierarchy
  - Classify each region individually
  - Accumulate region information at the pixel level



Recognition model

Appearance only

- Patch-based semantic image segmentation: PB-SIS
- Region-based semantic image segmentation: RB-SIS

	One scale (1S)		Multi scale (MS)	
	PB-SIS	RB-SIS	PB-SIS	RB-SIS
COL	55.72	62.84	62.31	65.94
SIFT	46.10	61.98	54.29	65.44
APP (COL+SIFT)	63.63	**70.24**	69.98	**72.90**

Regions are great assets that improve local appearance based prediction.

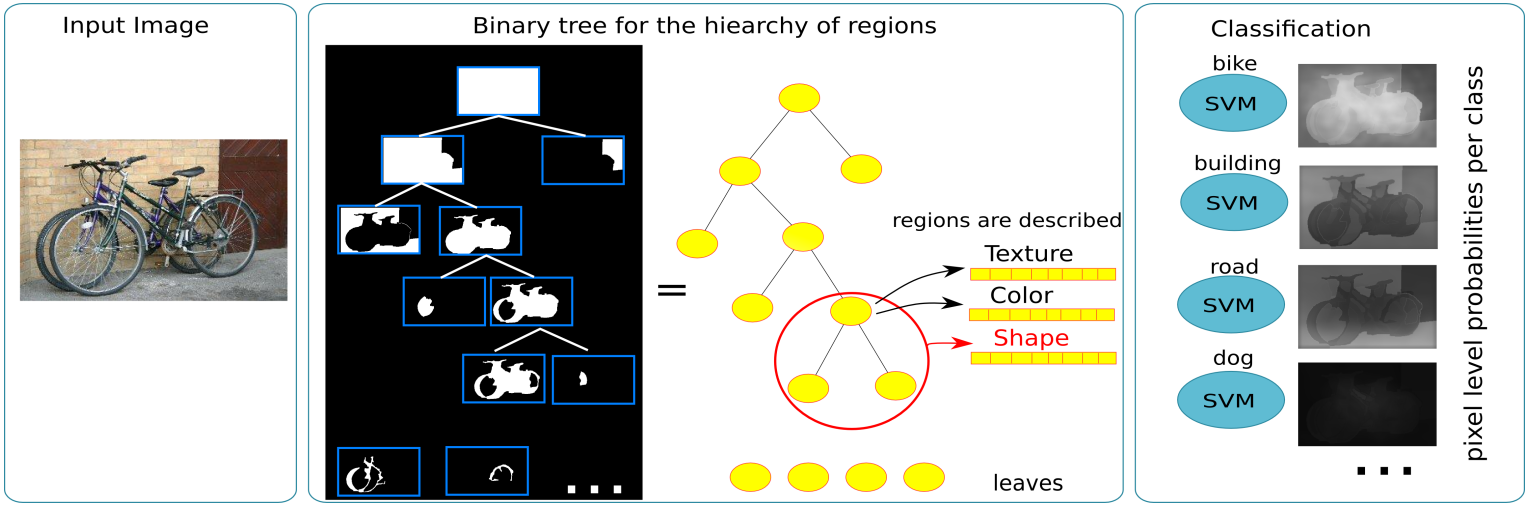
Exploiting the shape and the hierarchy of regions

For RB-SIS using regions, we can:

- use gPb as shape descriptor

  [Gu et al CVPR 2009, Lim et al ICCV 2009]

- exploit partially the hierarchy through Bags-of-Triplets



Exploiting the shape and the hierarchy of regions

RB-SIS: shape and bags-of-triplets

	shape only	+APP(1S)	+APP(MS)
BoR	34.77	70.35	71.85
BoR + BoT	42.70	71.18	**72.99**

- Shape alone performs poorly
- Hierarchy helps a lot for shape alone, but less when appearance is present

Table of contents

Image level prior

Appearance based predictions are combined with

- Global image classification (global Fisher Vector + SVM)
- Location prior (object location likelihood prior from training)

	REC	+ GL
PB-SIS	69.98	75.20
RB-SIS	72.99	75.88

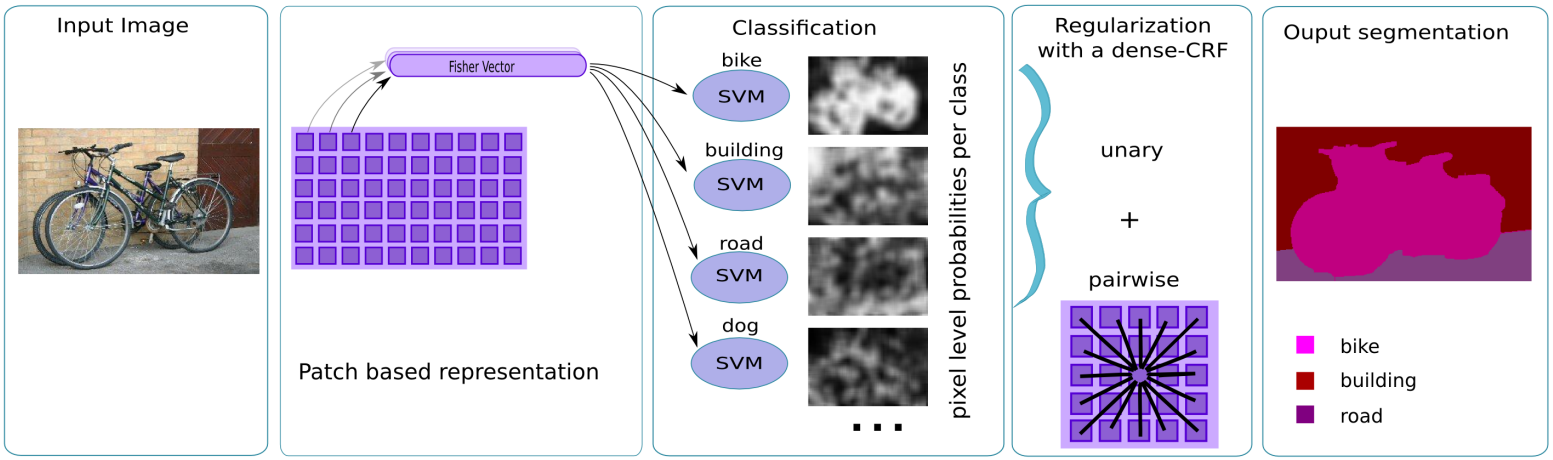
Recognition (REC) is enhanced with global and location (GL) priors

Table of contents

Conditional Random Field (CRF)

We use a dense CRF formulation

- unary potential: best recognition model enhanced with global and location priors
- pairwise potential: all pixel pairs are connected with pairwise
  - middle range regularization
  - longer range color-dependent regularization

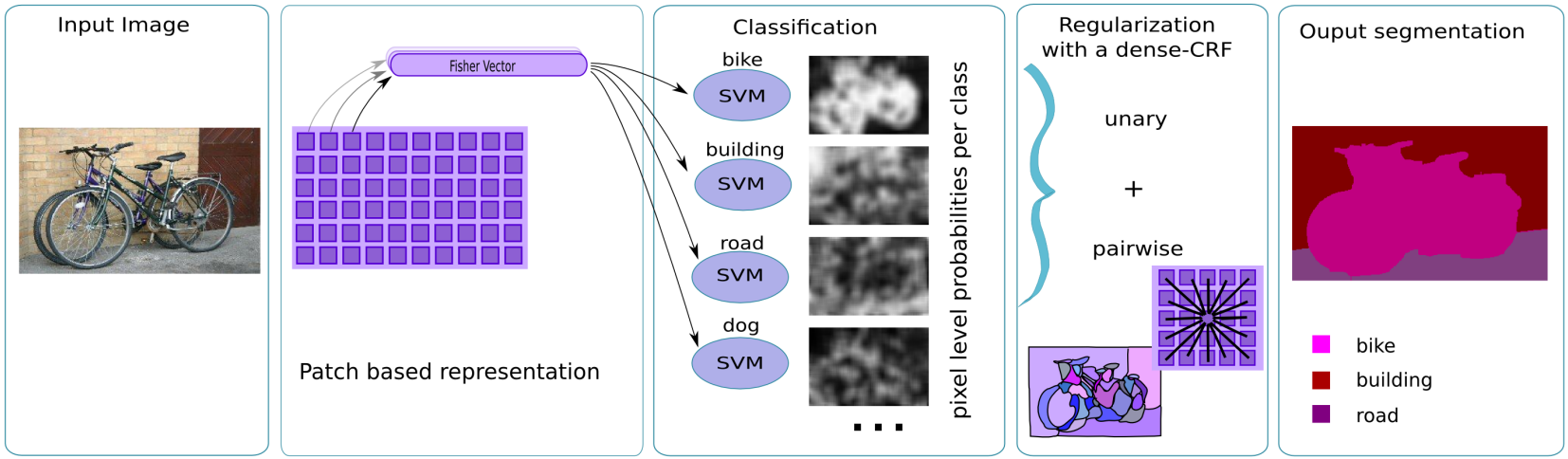


xerox

Conditional Random Field

We extend the dense CRF to use region information

- unary potential: best recognition model
- pairwise potential
  - middle range regularization
  - longer range color-dependent regularization
  - additional potential using leaf regions



xerox

Conditional Random Field

- Dense CRF results without (dCRF) and with (dCRFSP) region-based regularization

	REC	+ GL	dCRF	dCRFSP
PB-SIS	69.98	75.20	76.69	**77.25**
RB-SIS	72.99	75.88	75.80	76.02

- CRF regularization brings little improvement to RB-SIS
- PB-SIS benefits more from CRF, and outperforms RB-SIS

Qualitative results

test image - groundtruth - PB-prior - RB-prior - PB-dCRFSP - RB-dCRFSP

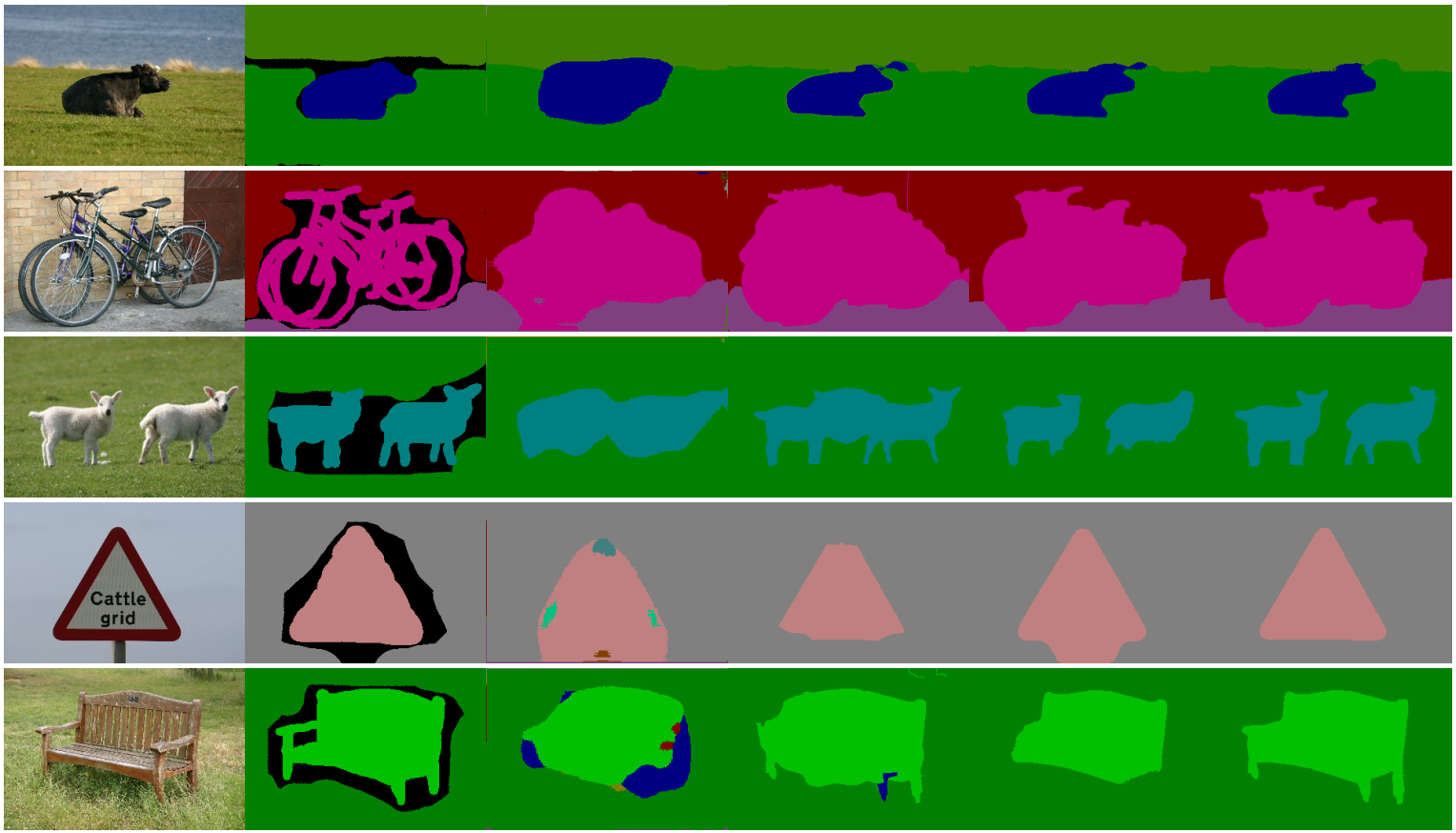


Table of contents

Conclusions

Proposed framework allows to evaluate the contribution of each component

Take Home Message:

- Simple recognition model using regions and global prior is already very competitive, no need for regularization
- When a CRF is considered, the patch-based model is enough, and regions could be used only at a later stage

Thanks for your attention !

Questions ?

- Backup-slides

Semantic Segmentation

Main limitation of an image partitioned into regions:

- No possible recovery if a region groups multiple classes.

Possible solutions:

- Multiple segmentation to obtain overlapping sets of regions

  [Pantofaru et al 2008, Gould et al 2009]

- Exploiting a hierarchy of regions

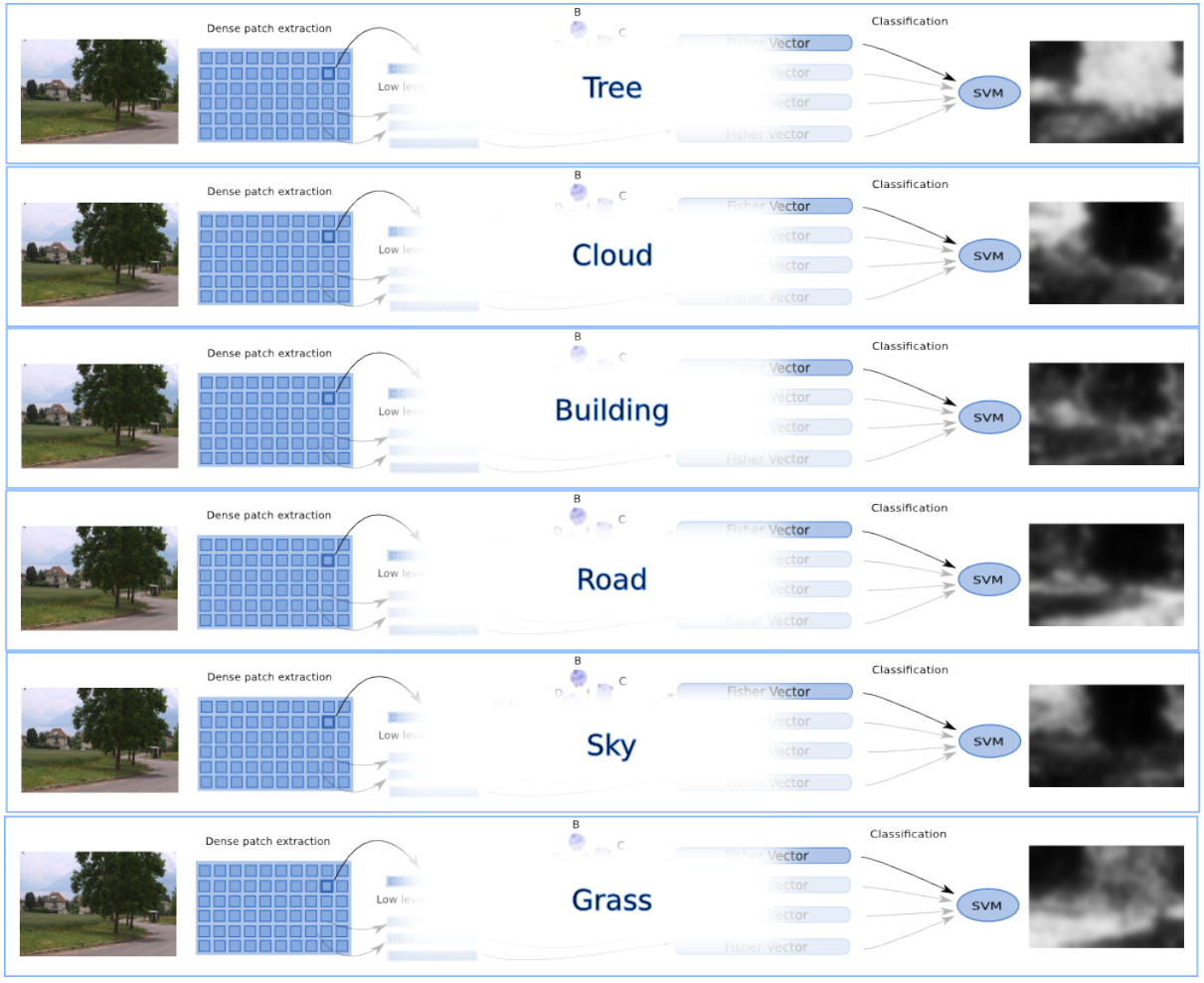
  [Ladicky et al 2009, Gu et al 2009, Lim et al 2009, Munoz et al 2010]

- Graph of regions

  [Chen et al 2011]

Patch-based Fisher Vector Representation

No regularization: simple patch voting



Final Segmentation

Ground truth for evaluation

xerox

Conditional Random Field (CRF)

We use a dense CRF formulation:

- CRF based regularization: dCRF

$$E(x) = \sum_i \psi_u(x_i) + \sum_{i<j} \delta_{x_i,x_j} \; \psi_p(x_i, x_j),$$

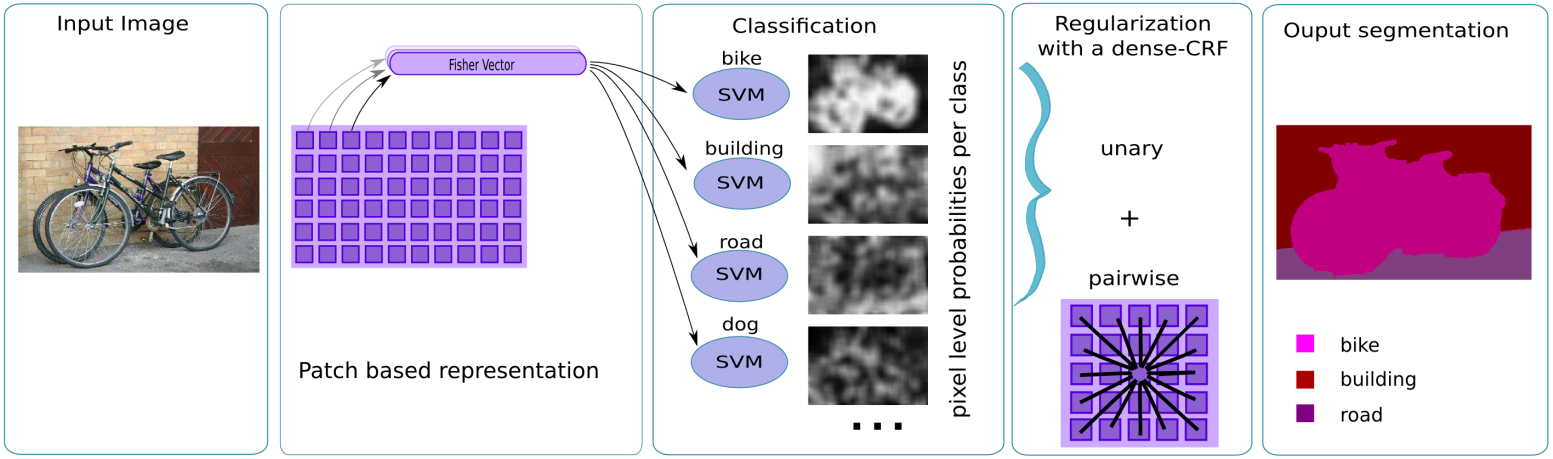
- Pairwise potential

$$\begin{aligned} \psi_p(x_i, x_j) &= \omega_1 \, exp\left(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2}\right) \\ &+ \omega_2 \, exp\left(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2}\right) \end{aligned}$$

with p_i and I_i being the position and RGB value of pixel x_i respectively.

Conditional Random Field

- CRF based regularization: dCRF



xerox

Conditional Random Field (CRF)

We extend the dense CRF to use region information:

- CRF based regularization: dCRFSP

$$E(x) = \sum_i \psi_u(x_i) + \sum_{i<j} \delta_{x_i,x_j}\ \hat{\psi}_p(x_i, x_j),$$

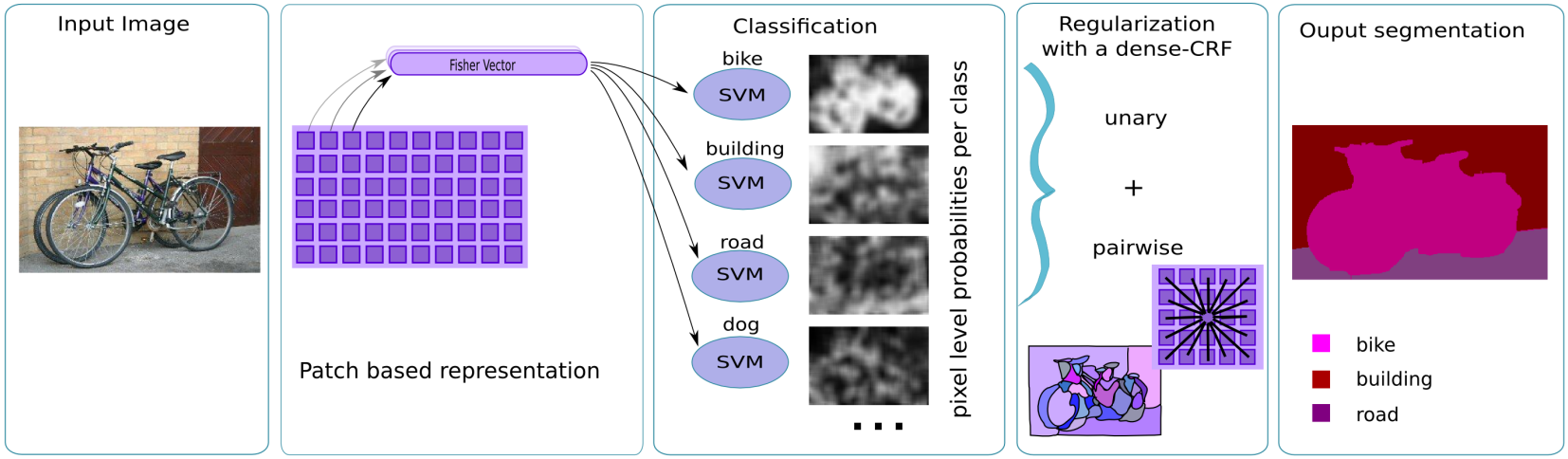
- Pairwise potential

$$\hat{\psi}_p(x_i, x_j) = \psi_p(x_i, x_j) + \omega_3\ exp\left(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|R_i - R_j|^2}{2\theta_\delta^2}\right)$$

with position p_i, RGB value of pixel I_i and the leaf region that contains x_i, R_i.

Conditional Random Field

- CRF based regularization: dCRFSP



xerox