Scene Segmentation with CRFs Learned from Partially Labeled Images

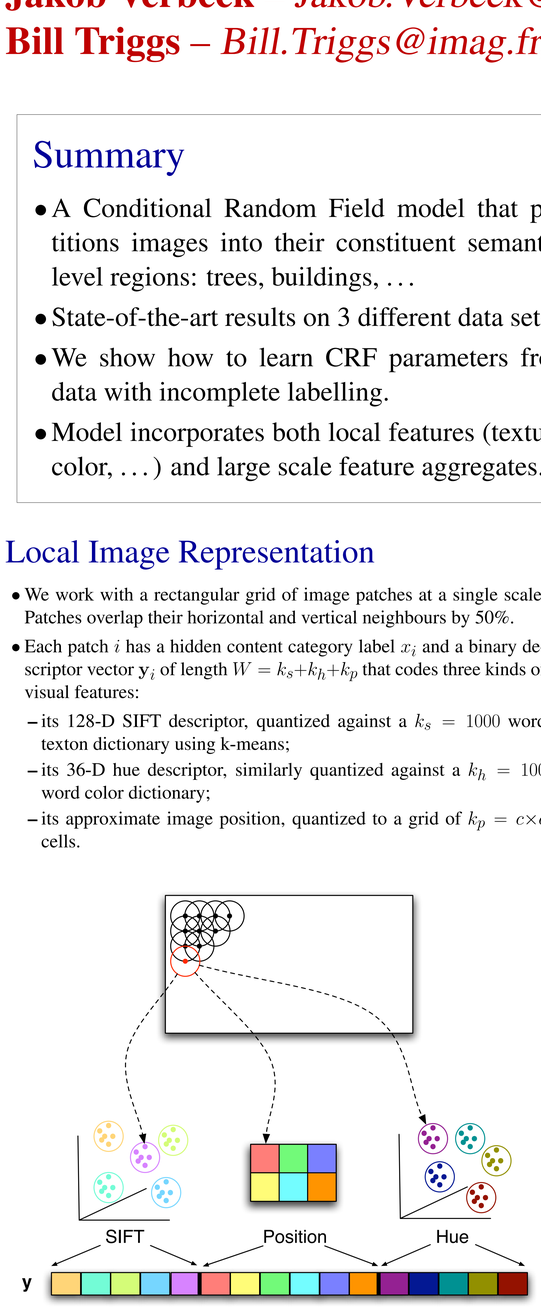
Jakob Verbeek – *Jakob.Verbeek@inria.fr, LEAR team, INRIA Rhône-Alpes, France*
Bill Triggs – *Bill.Triggs@imag.fr, Laboratoire Jean Kuntzmann, CNRS, France*

Summary

- A Conditional Random Field model that partitions images into their constituent semantic-level regions: trees, buildings, ...
- State-of-the-art results on 3 different data sets.
- We show how to learn CRF parameters from data with incomplete labelling.
- Model incorporates both local features (texture, color, ...) and large scale feature aggregates.

Local Image Representation

- We work with a rectangular grid of image patches at a single scale. Patches overlap their horizontal and vertical neighbours by 50%.
- Each patch i has a hidden content category label x_i and a binary descriptor vector $\mathbf{y}_i$ of length $W = k_s+k_h+k_p$ that codes three kinds of visual features:
  - its 128-D SIFT descriptor, quantized against a $k_s = 1000$ word texton dictionary using k-means;
  - its 36-D hue descriptor, similarly quantized against a $k_h = 100$ word color dictionary;
  - its approximate image position, quantized to a grid of $k_p = c\times c$ cells.



A schematic representation of local image feature extraction.

Aggregated Features

- To reduce the ambiguity of isolated patches, we also include aggregates of features (AF's) over larger image areas.
- On scale c, we divide the image into a regular $c\times c$ grid and average the local patch descriptors $\mathbf{y}_i$ within each grid cell.
- The aggregate feature $\mathbf{h}_{ic}$ for patch i is the average for the cell to which i belongs.

Learning from Partially Labeled Images

- Hand-labeling every pixel in an image set is tedious and error-prone – we expect diminishing returns.
- Most existing methods ignore unlabeled pixels during learning, but CRF's can exploit them to improve the model.
- A partial labeling tells us that the true labeling X belongs to a constrained set A. We maximize the probability that the CRF assigns a labelling in A, approximating this using the **Bethe free energy contrast** between the unconstrained and constrained models:

$$L = \log p(X \in A|Y) = \log \sum_{X\in A} p(X|Y) \quad (1)$$

$$\approx F_{Bethe}(p(X|Y)) - F_{Bethe}(p(X|Y, X \in A)) \quad (2)$$

- Gradient descent learning requires single node and pairwise marginals. We approximate these using Loopy BP:

$$\frac{\partial L}{\partial \theta} = \sum_X \Big(p(X|Y) - p(X|Y, X \in A)\Big)\frac{\partial E(X|Y)}{\partial \theta} \quad (3)$$

CRF Energy Function

- The CRF energy function combines pairwise potentials on a regular 4-connected grid and terms quantifying how compatible the patch label is with the observed local image features and larger aggregates.
- E.g. if only global image-wide aggregates are used, the model is

$$p(X|Y) \propto \exp\big(-E(X|Y)\big), \quad (4)$$

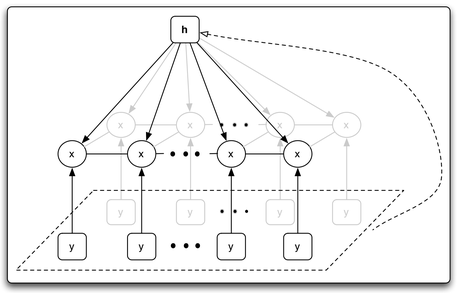
$$E(X|Y) = \sum_i \sum_{w=1}^{W} (\alpha_{wx_i} y_{iw} + \beta_{wx_i} h_w) + \sum_{i\sim j} \phi_{ij}(x_i, x_j) \quad (5)$$

where $i \sim j$ denotes neighboring patch pairs.

- We tested several pairwise potentials:

$$\phi_{ij}(x_i, x_j) = \gamma_{x_i,x_j}, \qquad \phi_{ij}(x_i, x_j) = (\sigma + \tau d_{ij})\delta(x_i - x_j), \quad (6)$$

where $d_{ij} = \exp(-\lambda\|z_i - z_j\|)$ and $z =$ average patch RGB value.



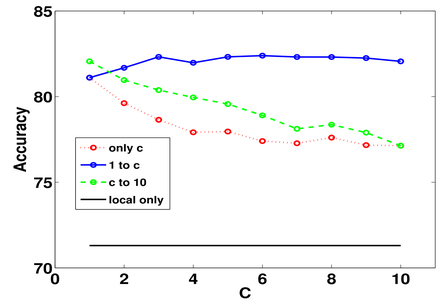
A sketch of the CRF model based on image-wide aggregates. The circular nodes represent the 4-connected grid of patch labels x_i. The square nodes represent feature functions. The dashed lines indicate the aggregation of the patch-level observations $\mathbf{y}_i$ into a global feature $\mathbf{h}$.

Experiments

Microsoft Research Cambridge data base: 240 images of 320×213 pixels; 9 classes: *building, grass, tree, cow, sky, plane, face, car, bike.*

Influence of Aggregate Features

- Performance of individual patch classifiers without pairwise potentials: without AFs (black); using AF's of a single c (red); using all AFs for $c' = 1, \ldots, c$ (blue); using all AFs for $c' = c, \ldots, 10$ (green).



- Incorporating AFs significantly improves performance. The largest scale AF's are the most informative.

Choice of Pairwise Potential

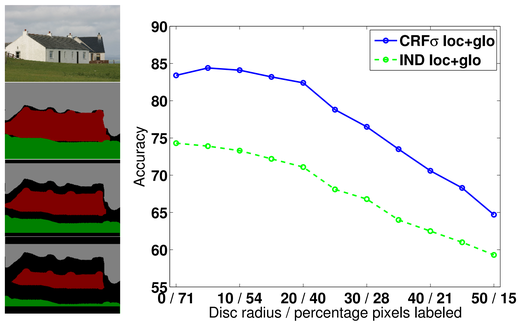
- Pairwise potential: *IND:* no pairwise coupling, independent patch-level classification; *CRFσ:* second form with $\tau = 0$; *CRFτ:* second form with both τ and σ nonzero; *CRFγ:* first form, class dependent.
- *local only:* no aggregate features; *local+global:* image-wide AFs.

AFs	Pairwise Potential			
	IND	CRFσ	CRFτ	CRFγ
local only	67.1	80.7	80.3	82.3
local+global	74.4	**84.9**	83.1	83.3

- Both CRF coupling and AFs produce significant performance gains. The exact form of the pairwise potential is less important.

Influence of Missing Training Labels

- We apply morphological erosion to the given training labels and plot performance as a function of the erosion radius r



Left: An image, its original labeling, and erosions of radius 10 & 20.

- Erosions with $r \leq 20$ reduce labelled pixels from 71% to 40% but have little impact on performance.

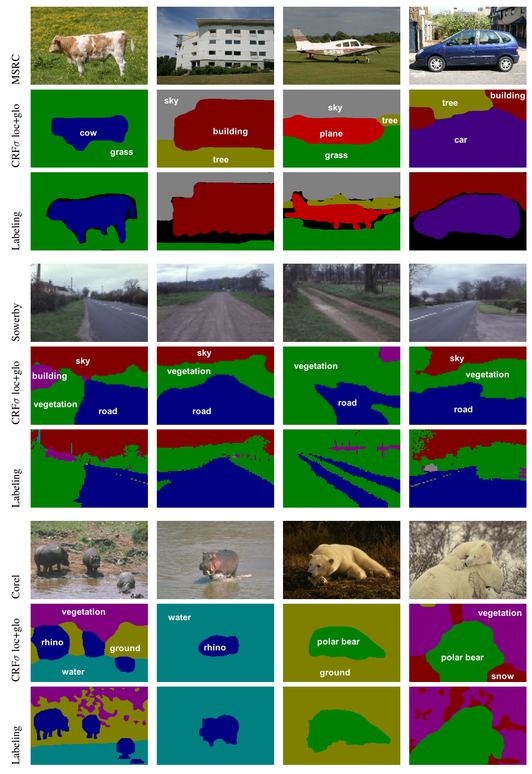
Comparison with Existing Methods

We compared recognition accuracies of our patch-level model with several state-of-the-art pixel-wise CRF models on two datasets

- Sowerby: 104 images, 96×64 pixels, 7 classes: *sky, vegetation, road marking , road surface, building, street objects, cars.*
- Corel: 100 images, 180×120 pixels, 7 classes: *rhino/hippo, polar bear, water, snow, vegetation, ground, sky.*

	Sowerby		Corel		MSRC
	IND	CRF	IND	CRF	
TextonBoost, Shotton [1]	85.6%	88.6%	68.4%	74.6%	-
M-CRF, He [2]	82.4%	89.5%	66.9%	80.0%	-
PLSA-MRF, Verbeek [3]	-	-	-	-	82.3
Schroff et al. [4]	-	-	-	-	75.2
CRFσ loc+glo	86.0%	87.4%	66.9%	74.6 %	84.9

Example Segmentations



[1] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost: joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *ECCV*, 2006.

[2] X. He, R. Zemel, and M. Carreira-Perpiñán. Multiscale conditional random fields for image labelling. In *CVPR*, 2004.

[3] J. Verbeek and B. Triggs. Region classification with Markov field aspect models. In *CVPR*, 2007.

[4] F. Schroff, A. Criminisi, and A. Zisserman. Single-histogram class models for image segmentation. In *ICCVGIP*, 2006.