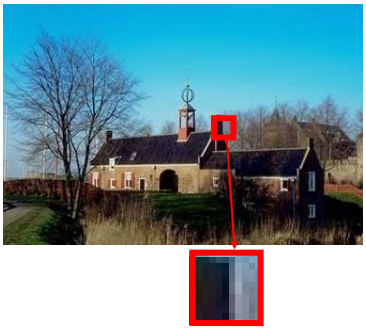
Martial Hebert
Sanjiv Kumar

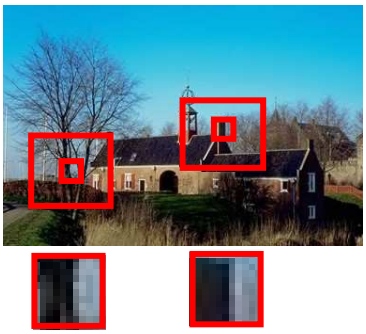
The Robotics Institute
School of Computer Science
Carnegie Mellon University

Motivation

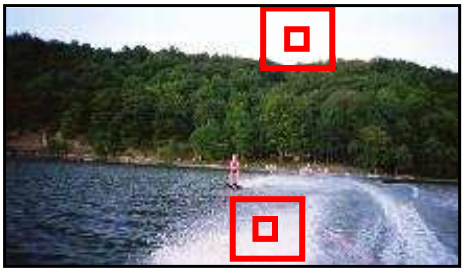


Map to appropriate feature space and classify

Motivation



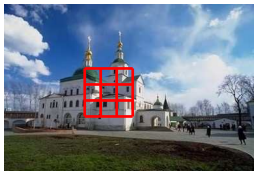
Need context from larger neighborhoods !

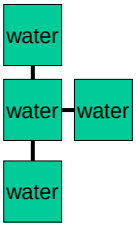


Context from whole image !

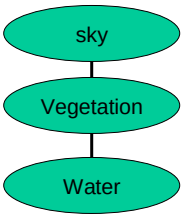
Context from Other Objects !

Interactions

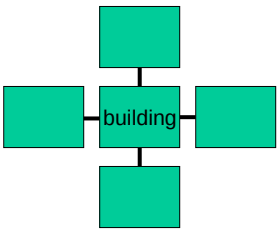




Pixel-Pixel



Region-Region



Patch-Patch

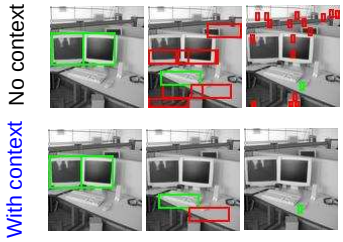
Interactions

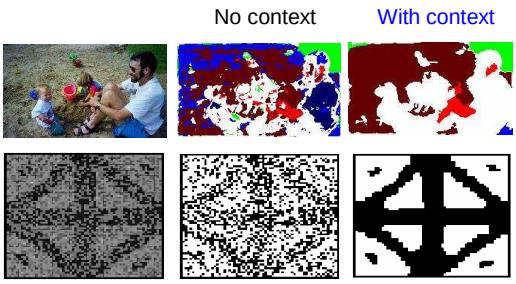


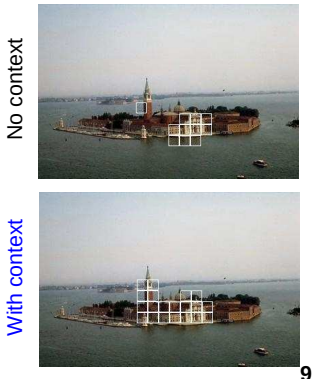
Part-Part **Object-Object** **Region-Object**

Modeling Interactions

- **Framework to learn all relevant contextual interactions in a single model automatically from training data**
- **Probabilistic models**
  - **Principled way to deal with ambiguities**
- **Graphical models**
  - **Powerful framework for ensuring global consistency using relatively local constraints**



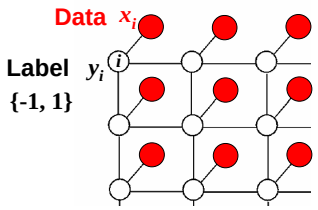


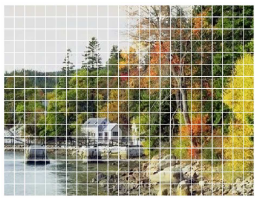


Outline

- **Background**
  - **Markov Random Fields (MRFs)**
- **Conditional Random Fields (CRFs)**
- **Multiclass**
- **Hierarchical Interactions**

Markov Random Field (MRF)





Generative Framework

$$P(y|x) \propto P(y,x) = \underbrace{P(x|y)}_{\text{Observation Model}}\underbrace{P(y)}_{\text{Prior Model (MRF)}}$$

$$P(y|x) = \frac{1}{Z}\exp\left(\sum_{i\in S}\log P(x_i|y_i) + \sum_{i\in S}\sum_{j\in N_i}\beta\, y_i y_j\right)$$

[Geman & Geman, '84]

Generative vs. Discriminative

We want : $P(y|x)$

- Generative Framework
  - Models $P(y,x)$ to get $P(y|x)$
  - Implicit modeling of observations
- Discriminative Framework
  - Models Conditional distribution $P(y|x)$ directly

Conditional Random Field (CRF)

- $P(y|x)$ is directly modeled as Gibbs Field
- Graphs on images with loops
- Clique potentials using local discriminative models

$$P(y|x) = \frac{1}{Z} \exp\left(\sum_{i \in S} A_i(y_i, x) + \sum_{i \in S} \sum_{j \in N_i} I_{ij}(y_i, y_j, x) \right)$$

Association Potential

Interaction Potential

[Lafferty et al., ICML'01] [Kumar, ICCV '03]

Comparison with MRF framework

MRF $$P(y|x) = \frac{1}{Z} \exp\left(\sum_{i \in S} \log P(x_i|y_i) + \sum_{i \in S} \sum_{j \in N_i} \beta\, y_i y_j \right)$$

CRF $$P(y|x) = \frac{1}{Z} \exp\left(\sum_{i \in S} A(y_i, x) + \sum_{i \in S} \sum_{j \in N_i} I(y_i, y_j, x) \right)$$

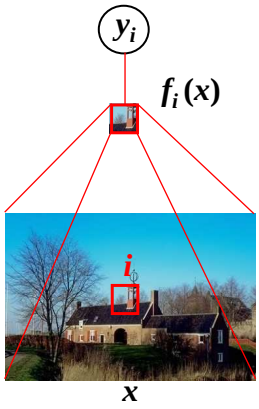
Data from multiple sites

Data-dependent label interactions

Example Association Potential $A(y_i, x)$

discriminative classifier

$$A(y_i, x) = \log P(y_i \mid f_i(x))$$
$$= \log \sigma(y_i w^{\mathrm{T}} f_i(x))$$



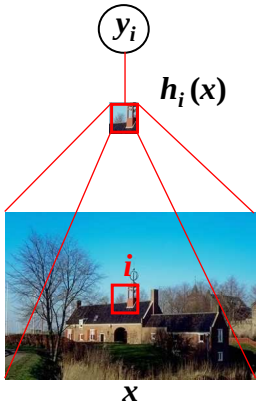
Probit, NN, boosting

[Szummer et al., 2004][He et al., 2004][Torralba et al., 2005]

Example Association Potential $A(x_i, y)$

discriminative classifier

$$A(y_i, x) = \log P(y_i \mid h_i(x))$$
$$= \log \sigma(y_i w^{\mathrm{T}} h_i(x))$$



Mapping to induce a nonlinear decision boundary:

$$h_i(x) = [1, \phi_1(f_i(x)), \ldots, \phi_R(f_i(x))]^{\mathrm{T}}$$

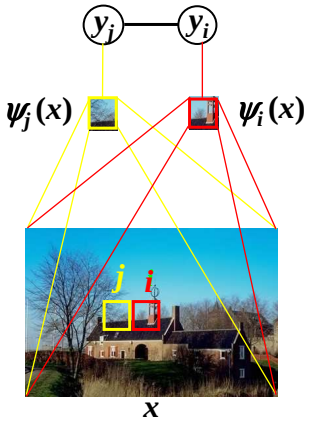
[Lafferty et al., ICML'04]

Example Interaction Potential $I(y_i, y_j, x)$

pairwise discriminative classifier

$$I(y_i, y_j, x) = \log P(y_i, y_j | \psi_i, \psi_j)$$

$$= y_i y_j \underbrace{v^T \mu_{ij}(x)}_{\beta}$$



[Kumar & Hebert, NIPS '04]

Using the Model

Parameter Learning

M **training images → parameters** $\theta = \{w, v\}$

Inference

Input image → Labels at every site y_i

Parameter Learning

- **Maximum Likelihood - Given *M* training images,**

$$l(\theta) = \sum_{m=1}^{M} \left\{ \sum_{i \in S} \log \sigma(y_i w^{\mathrm{T}} h_i(x)) + \sum_{i \in S} \sum_{j \in N_i} y_i y_j v^T \mu_{ij}(x) - \log Z \right\}$$

$$\frac{\partial l(\theta)}{\partial w} = \frac{1}{2} \sum_m \sum_i (y_i - \langle y_i \rangle_{P(y|x)}) h_i(x)$$

$$\frac{\partial l(\theta)}{\partial v} = \sum_m \sum_i \sum_j (y_i y_j - \langle y_i y_j \rangle_{P(y|x)}) \mu_{ij}(x)$$

- **Maximum pseudo-likelihood**
- **Contrastive Divergence**
- **Marginal Approximations**
- **Saddle Point Approximation**

[Hinton '86]

Parameter Learning

- **Maximum pseudo-likelihood (PL)**

$$\hat{\theta} = \arg\max_{\theta} \prod_{m=1}^{M} \prod_{i \in S} P(y_i^m \,|\, y_{N_i}^m, x^m, \theta)$$

- Overestimates interaction parameters [Kumar & Hebert '03]

- **Contrastive Divergence (CD) [Hinton '02]**

$$\langle y_i \rangle \approx (y_i)_{P_T^1} \quad \text{and} \quad \langle y_i y_j \rangle \approx (y_i y_j)_{P_T^1}$$

- Large bias [Williams et al. '02]

Marginals Based Approximation

- **Pseudo Marginal Approximation (PMA)**

Pseudo marginals (Belief Propagation)

$$\langle y_i \rangle = \sum_{y_i} y_i P(y_i|x) \approx \sum_{y_i} y_i b_i(y_i)$$

$$\langle y_i y_j \rangle = \sum_{y_i} \sum_{y_j} y_i y_j P(y_i, y_j|x) \approx \sum_{y_i} \sum_{y_j} y_i y_j b_{ij}(y_i, y_j)$$

- **Maximum Marginal Approximation (MMA)**

$$\tilde{y}_i = \arg\max_{y_i} b_i(y_i)$$

$$\langle y_i \rangle \approx \tilde{y}_i \text{ and } \langle y_i y_j \rangle \approx \tilde{y}_i \tilde{y}_j$$

Saddle Point Approximation (SPA)

- **Maximum Likelihood - Given *M* training images,**

$$\frac{\partial l(\theta)}{\partial w} = \frac{1}{2} \sum_m \sum_i (y_i - \langle y_i \rangle_{P(y|x)}) h_i(x)$$

$$\frac{\partial l(\theta)}{\partial v} = \sum_m \sum_i \sum_j (y_i y_j - \langle y_i y_j \rangle_{P(y|x)}) \mu_{ij}(x)$$

- **Saddle Point Approximation**

$$\hat{y} = \arg\max_{y} P(y|x, \theta)$$

$$\langle y_i \rangle \approx \hat{y}_i \ , \ \langle y_i y_j \rangle \approx \hat{y}_i \hat{y}_j$$

[Kumar & Jonas, EMMCVPR '05]

Saddle Point Approximation (SPA)

$$\hat{\boldsymbol{y}} = \arg\max_{\boldsymbol{y}} \boldsymbol{P}(\boldsymbol{y}|\boldsymbol{x},\theta)$$

$$\frac{\partial \boldsymbol{l}(\theta)}{\partial \boldsymbol{w}} \approx \frac{1}{2}\sum_{\boldsymbol{m}}\sum_{\boldsymbol{i}}(\boldsymbol{y}_{\boldsymbol{i}} - \hat{\boldsymbol{y}}_{\boldsymbol{i}})\boldsymbol{h}_{\boldsymbol{i}}(\boldsymbol{x})$$

$$\frac{\partial \boldsymbol{l}(\theta)}{\partial \boldsymbol{v}} \approx \sum_{\boldsymbol{m}}\sum_{\boldsymbol{i}}\sum_{\boldsymbol{j}}(\boldsymbol{y}_{\boldsymbol{i}}\boldsymbol{y}_{\boldsymbol{j}} - \hat{\boldsymbol{y}}_{\boldsymbol{i}}\hat{\boldsymbol{y}}_{\boldsymbol{j}})\mu_{\boldsymbol{ij}}(\boldsymbol{x})$$

Perceptron Learning Rules

[Freund et al. ML'99] [Collins, EMNLP'02]
[LeCun et al. IEEE Proc.'98]

Comparative Experiments

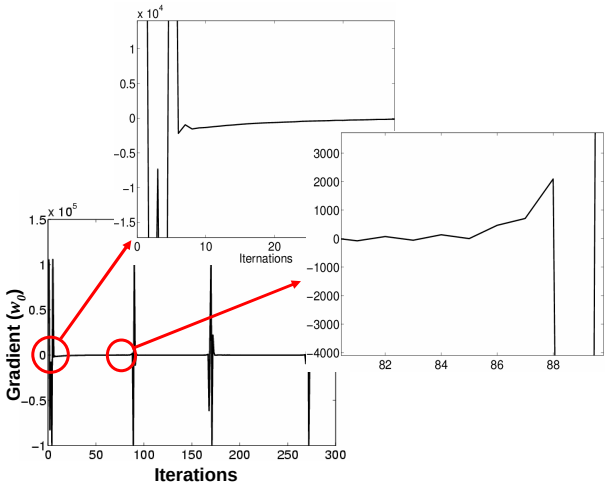
Original

Noisy

- **4 base images (64 x 64)**
- **Corrupted by two noise models**
  - **Gaussian (single Gaussian)**
  - **Bimodal (mixture of two Gaussians)**
- **Training – 10 images, Testing – 200 images**

Approximate Gradient Ascent

QuickTime™ and a
decompressor
are needed to see this picture.



[Kumar & Jonas, EMMCVPR '05]

Inference

Given input image x and $P(y|x)$, get the optimal labels y

- **Exact Maximum A Posteriori (MAP)**
  - **Mincut-maxflow** **[Greig et. al '89]**
- **Approximate Maximum Marginals**
  - **Loopy Belief Propagation (BP)** **[Frey & Mckay '98]**

Learning Experiments / Bimodal Noise

Inference

Parameter Learning	MAP (min-cut)	MPM (BP)	Local MAP (ICM)
SPA (min-cut)	5.82	19.19	14.88
PMA (BP)	6.45	5.48	17.39
MMA (BP)	26.53	5.70	16.00
CD	8.98	6.29	8.91
PL	17.69	7.30	22.21

Pixelwise error (%) on 200 test images

Learning $\overset{\text{Coupling}}{\longleftrightarrow}$ Inference

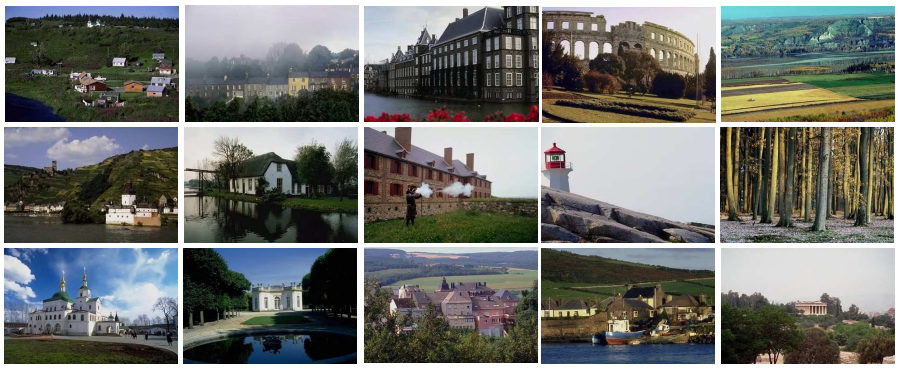
SPA (min-cut) $\longleftrightarrow$ MAP (min-cut)

PMA/MMA (BP) $\longleftrightarrow$ MPM (BP)

[Kumar & Jonas, EMMCVPR '05][Zhu at al., PAMI'02]

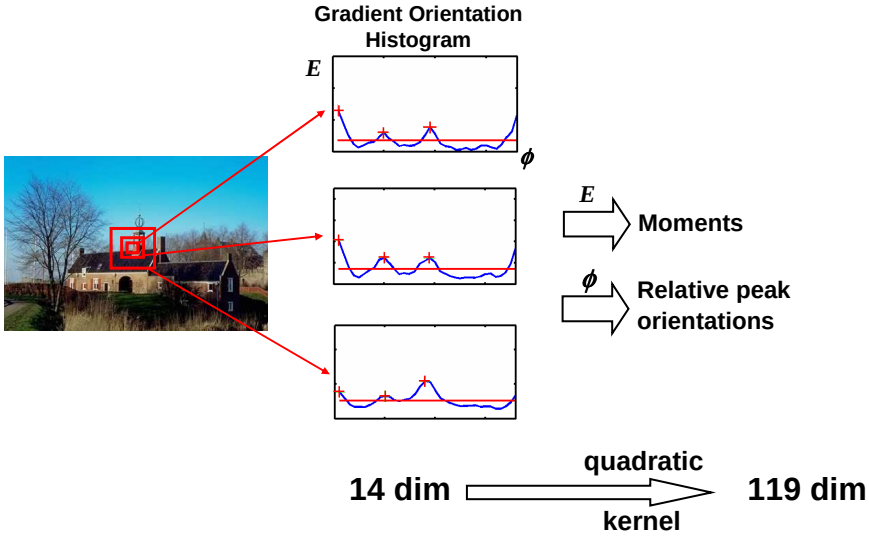
Example: Man-Made Structure Detection

Training Set – 130 images (256x384 pixels)



Scale variations | Illumination variations | Pose variations | Non-linear structures | Negative samples

Design of Feature Vector (h_i)

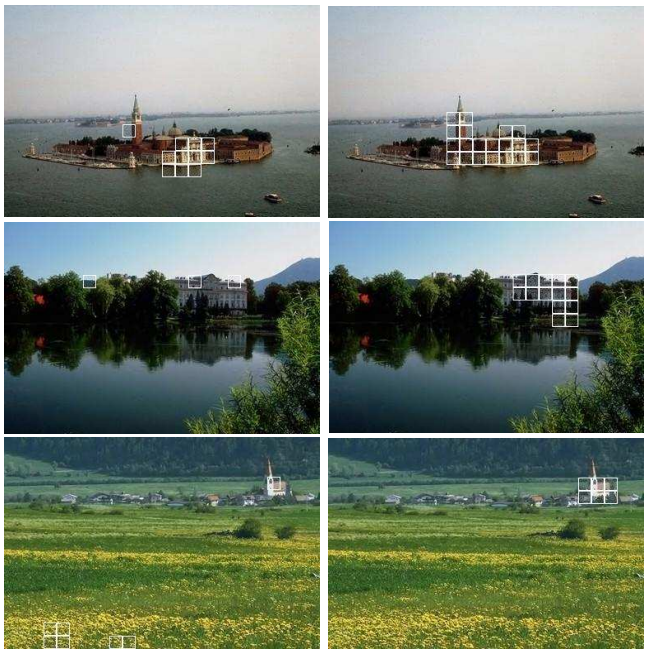


[Kumar & Hebert CVPR'03]

[Lafferty et al., ICML'04]

Performance Evaluation

MRF CRF



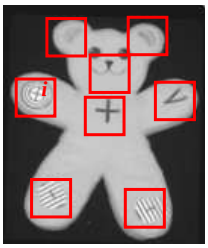
So Far…

- Background
  - Markov Random Fields (MRFs)
- Conditional Random Fields (CRFs)
- Multiclass
- Hierarchical Interactions

Multiclass Extension

- **Labels** $y_i \in \{1,...,C\}$
- **Association Potential**

$$A(y_i, x) = \log P(y_i | x)$$



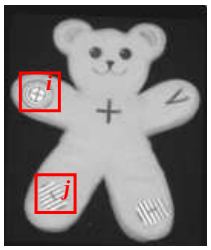
$$P(y_i = k | x) = \begin{cases} \dfrac{\exp(w_k^T h_i(x))}{1 + \sum_{l=1}^{C-1} \exp(w_l^T h_i(x))} & \text{if } k < C \\ \dfrac{1}{1 + \sum_{l=1}^{C-1} \exp(w_l^T h_i(x))} & \text{if } k = C \end{cases}$$

[Kumar & Hebert, ICCV'05][Quattoni et al., NIPS'04]

Multiclass Extension

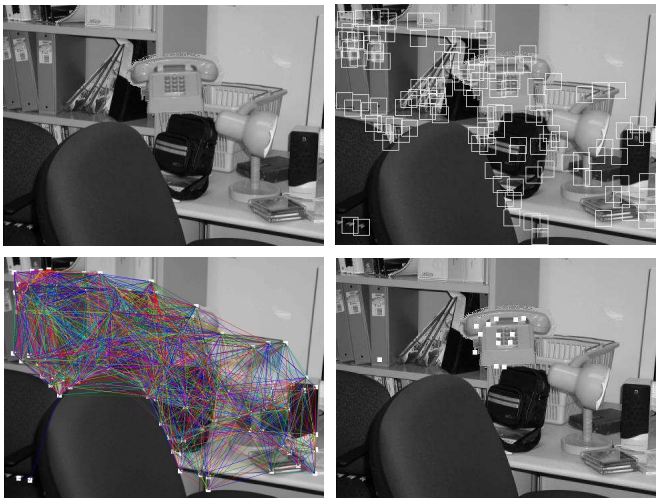
- **Interaction Potential**

$$I(y_i, y_j, x) = \sum_{k=1}^{C} \sum_{l=1}^{C} v_{kl}^T \mu_{ij}(x) \delta(y_i = k) \delta(y_j = l)$$



- **Inference : Loopy Belief Propagation (BP)**
  - **Alternatives: CCCP [Yuille '02], EP [Minka '01]**

Parts-Based Object Detection

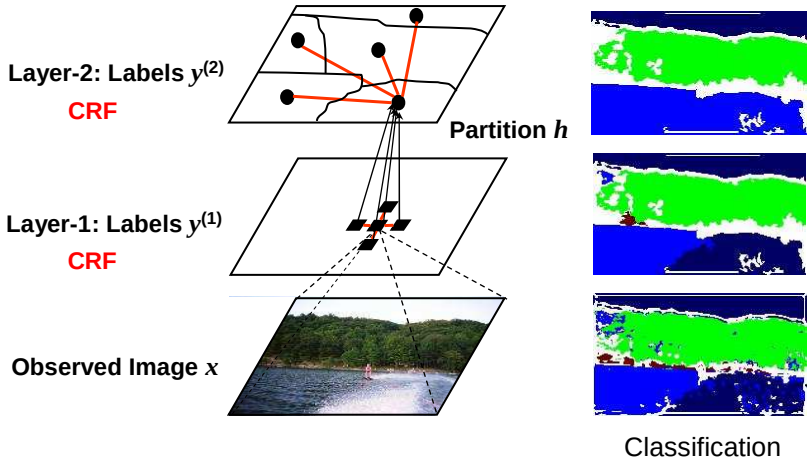


Inference (Loopy BP) time: 1.35 sec

Outline

- Background
  - Markov Random Fields (MRFs)
- Conditional Random Fields (CRFs)
- Multiclass
- Hierarchical Interactions

Hierarchical Interactions



Hierarchical Formulation

We want: Labels given input image

Layer-2 labels given mapping and labels 1

Mapping from layer-1 to 2

Layer-1 labels given input image

CRF

CRF

$$P(y \mid x) \cong \sum_{h,\, y^{(1)}} P(y^{(2)} \mid h,\, y^{(1)}) P(h \mid y^{(1)}) P(y^{(1)} \mid x)$$

Design Issues

$P(y^{(2)} \mid h, y^{(1)})$
$P(y^{(1)} \mid x)$

Designed conditional random fields at layer 1 and layer 2

$\sum_{y^{(1)}}$

Leads to combinatorial explosion and cannot be computed exactly

Designed approximation (using the fact that $P(y^{(1)} \mid x)$ is close to 0 almost everywhere)

$P(h \mid y^{(1)})$ Designed a simple distribution model

(An Attempt at a) Hierarchical Formulation

We want: Labels given input image

Layer-2 labels given mapping and labels 1

Mapping from layer-1 to 2

Layer-1 labels given input image

CRF

CRF

$$P(y \mid x) \cong \sum_{h,\, y^{(1)}} P(y^{(2)} \mid h, y^{(1)}) P(h \mid y^{(1)}) \underbrace{P(y^{(1)} \mid x)}_{\delta(y^{(1)} - \tilde{y}_b)}$$

b = Marginals at layer 1 $b_i = P(y_i^{(1)} \mid x)$

$\tilde{y}_b$ = Max. marginal at layer 1 (estimated through BP)

(An Attempt at a) Hierarchical Formulation

We want: Labels given input image

Layer-2 labels given mapping and labels 1

Mapping from layer-1 to 2

Layer-1 labels given input image

CRF

CRF

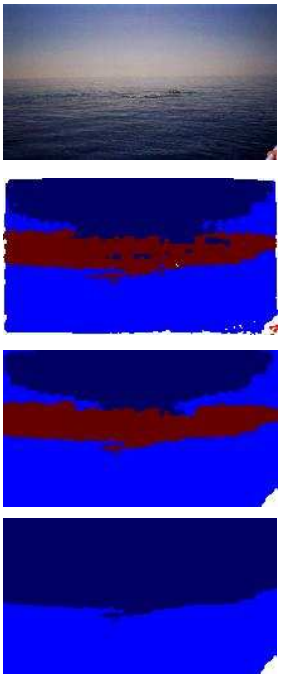
$$P(y \mid x) \cong \sum_{h,\, y^{(1)}} P(y^{(2)} \mid h,\, y^{(1)}) P(h \mid y^{(1)}) P(y^{(1)} \mid x)$$

$$\approx \sum_{h} P\left(y^{(2)} \mid h, b\right) P\left(h \mid b\right)$$

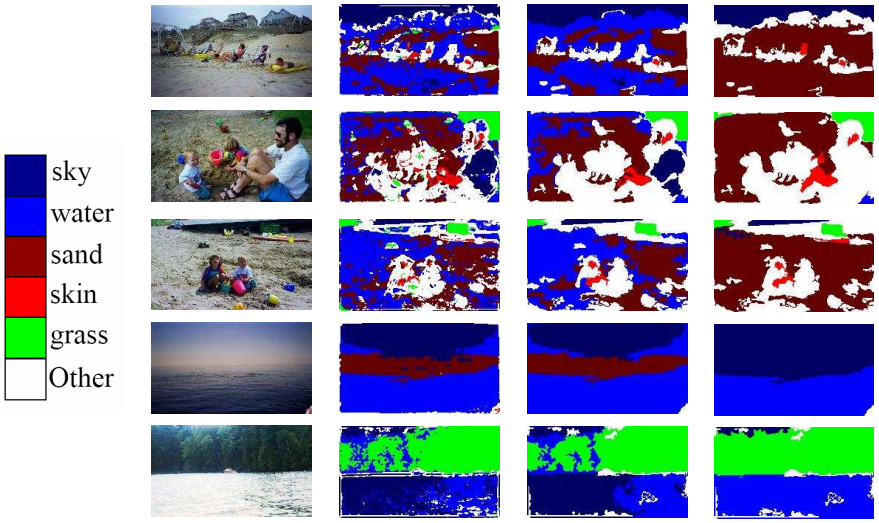
- Sequential Training – Pseudo marginals (BP)
- Inference – Max marginals (BP)

Region Classification

- Data sets - Images of outdoor scenes
  - Beach Dataset: 123 images (124x218 pixels), 6 classes
  - Sowerby Dataset: 104 images (64x96 pixels), 7 classes
- Layer 1 Features
  - HSV/Luv Color, texture moments, oriented DoG filters
- Layer 2 features
  - Relative configuration of regions (*above, beside, enclosed*)
- Training Time: ~ 10 min on 2.8 GHz machine



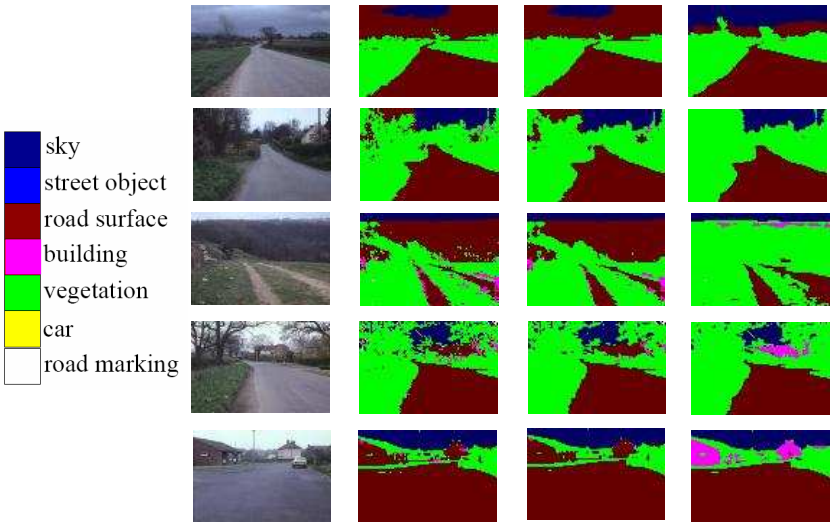
Beach Dataset



Input image | Softmax (no context) 62.3 % | Layer-1 (label smoothing only) 63.8 % | Layer-2 (full model) 74 % (~ 2 Sec)

[Singhal et al., CVPR'04] Previous best 64 %

Sowerby Dataset



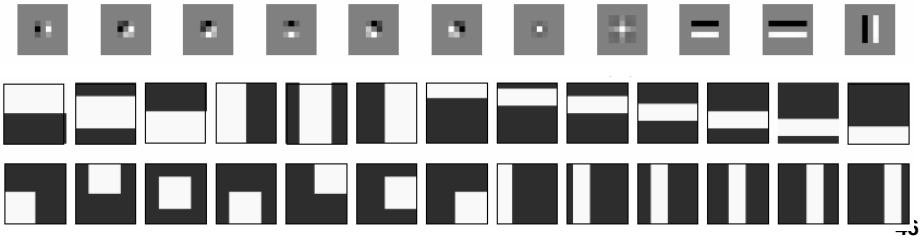
Input image | Softmax (no context) 85.4 % | Layer 1 (label smoothing only) 85.8 % | Layer 2 (full model) 89.3 %

[He et al., CVPR'04][Feng et al., PAMI'02]

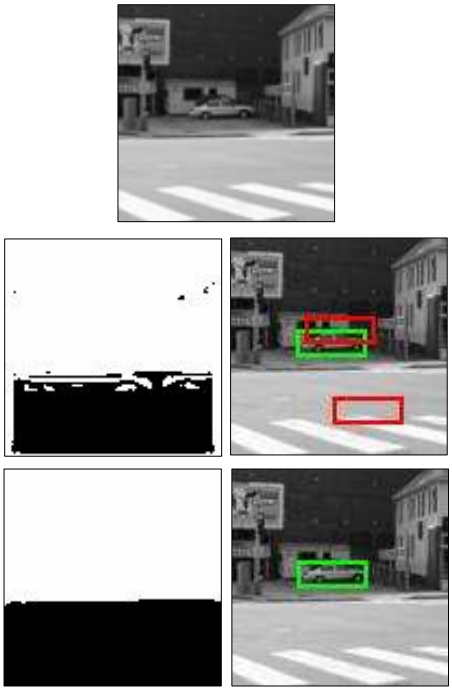
Object-Region Interactions

- **MIT database**
  - ~100 outdoor images cars, buildings and roads (100x100 pixels)
- **Region features**
  - Intensity, oriented DoG filters for texture
- **Object features**
  - Detector score trained with Gentle Boosting

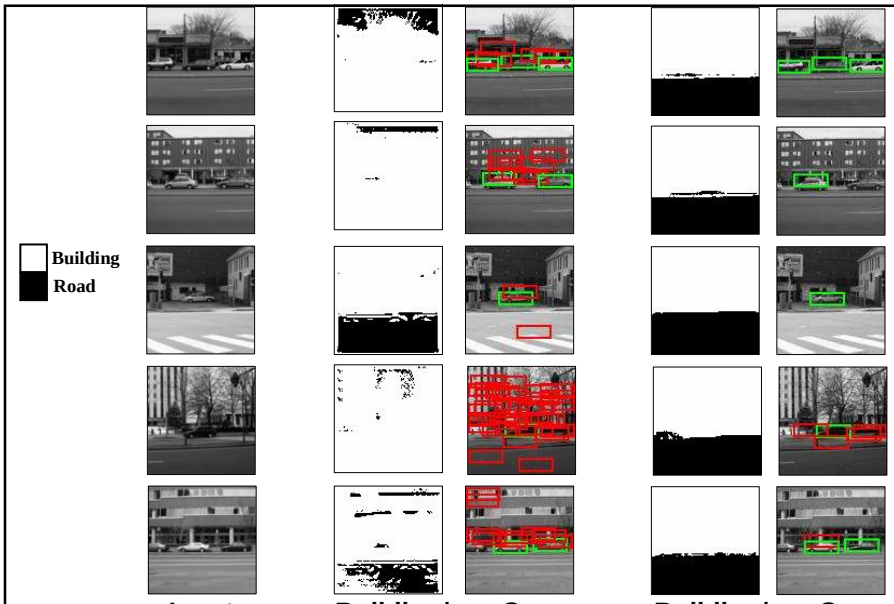
Dictionary (500 filters and templates)



[Torralba et al., NIPS'04]



47



[Torralba et al., NIPS'04]

48

Object-Object Interactions

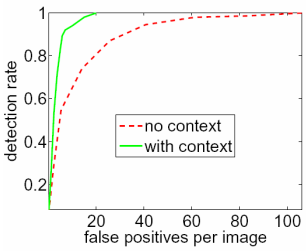
- MIT context database: 164 images (100x100 pixels)
- Very small objects (8x5 pixels) → High false positives
- Initial object detectors trained with Gentle Boosting

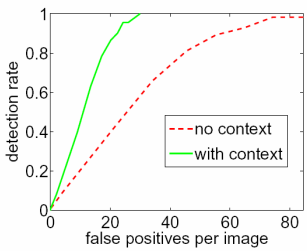





Input image | Without Interactions | With Interactions

Keyboard





Mouse

Summary

- Background
  - Markov Random Fields (MRFs)
- Conditional Random Fields (CRFs)
- Multiclass
- Hierarchical Interactions