Representation in Low-Level Visual Learning

Erik Sudderth

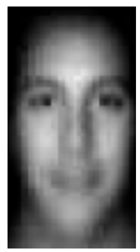
Brown University
Department of Computer Science

Generative Models: A Caricature

Turk & Pentland 1991, Moghaddam & Pentland 1995

Training Faces

$\approx$



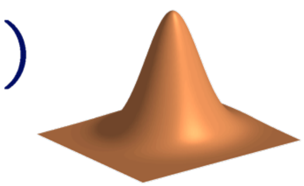
Mean Face

$+ \, w \times$

Eigenfaces

$$x_i = \mu + B w_i \qquad w_i \sim \mathcal{N}(0, I)$$

Gaussian Prior
“Knowledge”

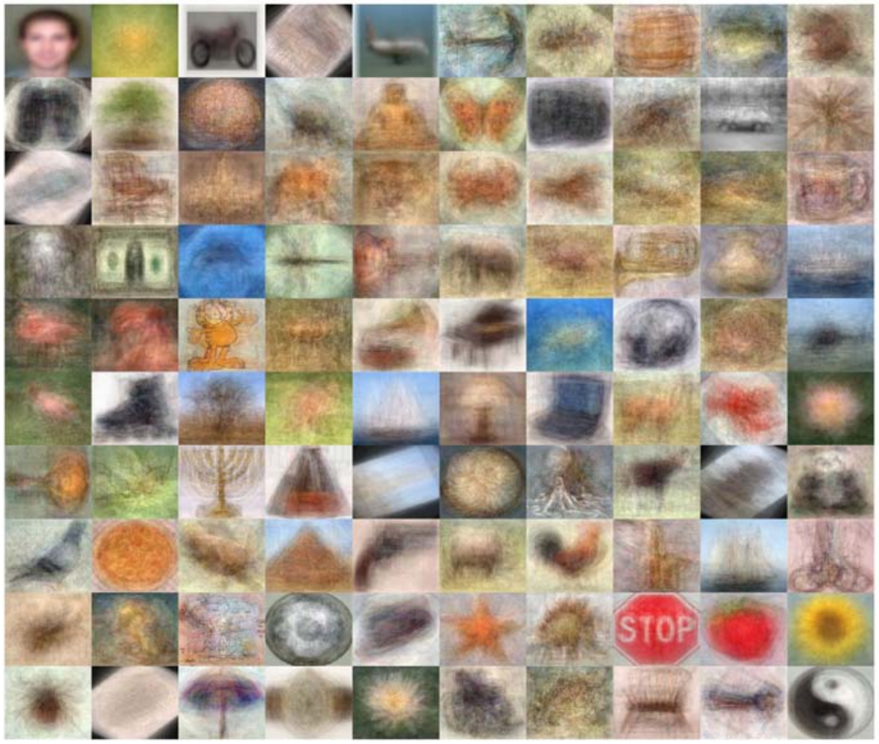


Most visual learning has used *overly simplified* models

What about Eigenbikes?

Representation Matters

The Traditional Solution: Dataset Selection

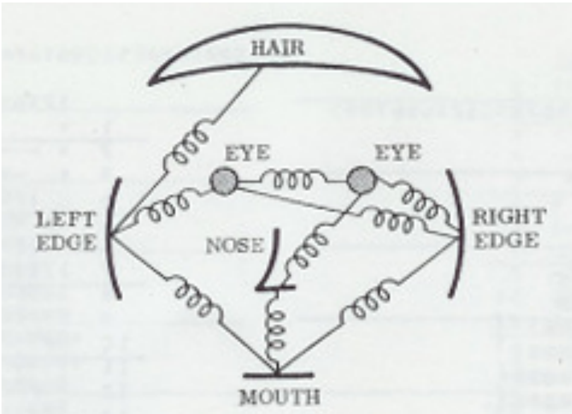


Caltech 101

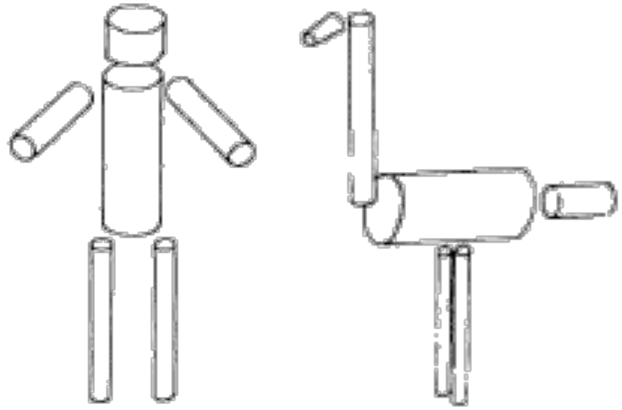
LabelMe *Excerpt, Sudderth et al., 2005*

Natural Scenes*, Olive & Torralba, 2001*

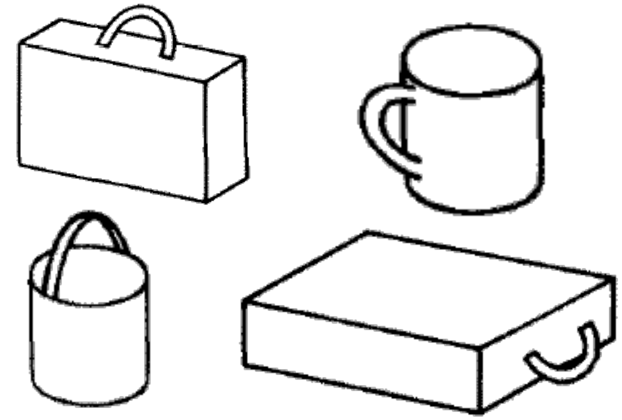
A Success: Part-Based Models



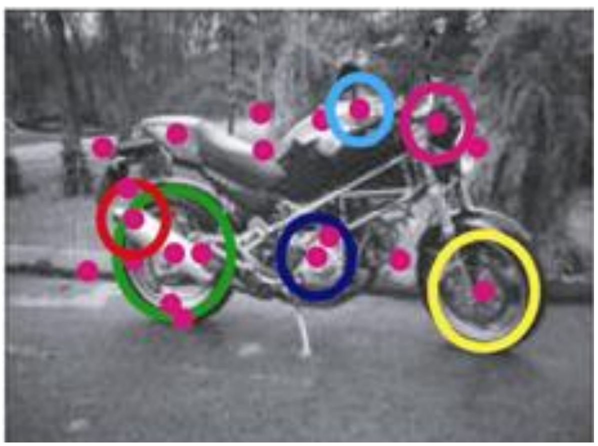
Pictorial Structures
Fischler & Elschlager, 1973



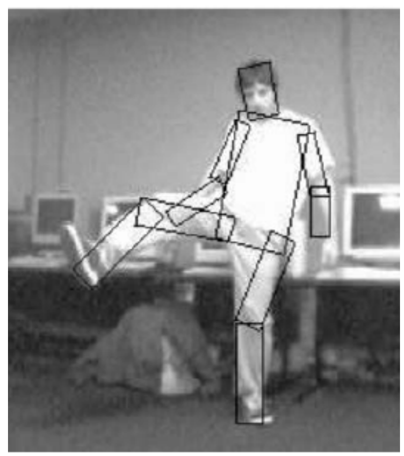
Generalized Cylinders
Marr & Nishihara, 1978



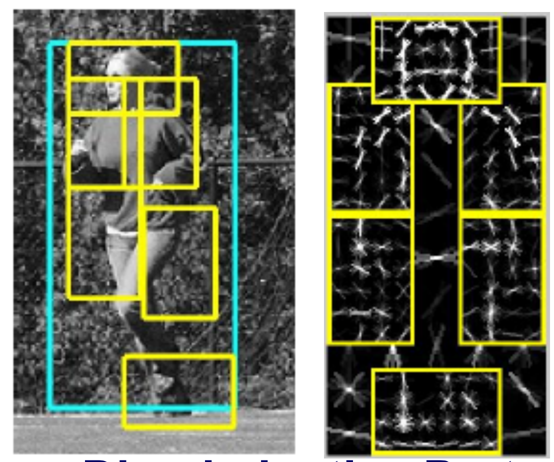
Recognition by Components
Biederman, 1987



Constellation Model
Perona, Weber, Welling, Fergus, Fei-Fei, 2000 to …



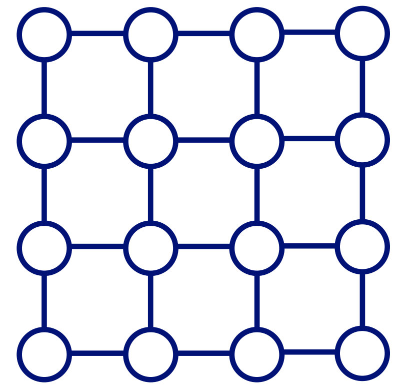
Efficient Matching
Felzenszwalb & Huttenlocher, 2005



Discriminative Parts
Felzenszwalb, McAllester, Ramanan, 2008 to …

Low-Level Vision: Discrete MRFs

Ising and Potts Markov Random Fields



$$p(z) = \frac{1}{Z(\beta)} \prod_{(s,t)\in E} \psi_{st}(z_s, z_t)$$

$$\log \psi_{st}(z_s, z_t) = \begin{cases} \beta_{st} > 0 & z_s = z_t \\ 0 & \text{otherwise} \end{cases}$$

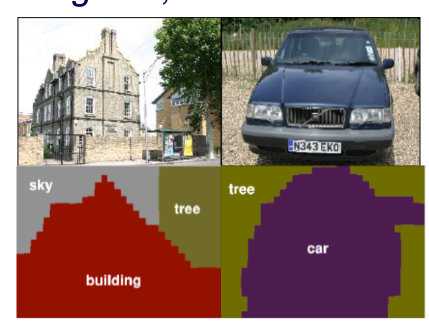
Maximum Entropy model with these (intuitive) features.

GrabCut: *Rother, Kolmogorov, & Blake 2004*

Previous Applications

- Interactive foreground segmentation
- Supervised training for known categories

…but very little success at segmentation of unconstrained natural scenes.



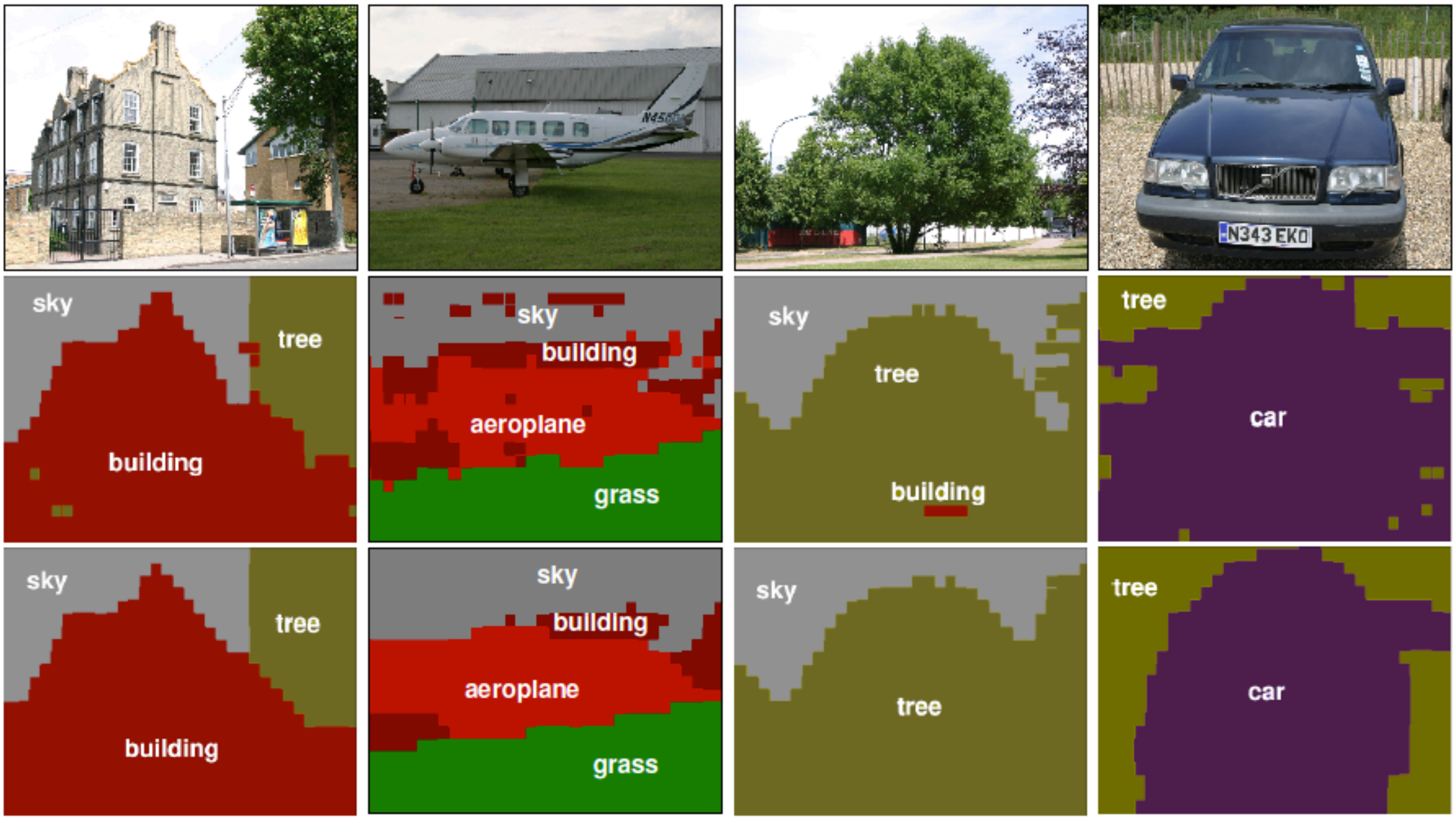
Verbeek & Triggs, 2007

Region Classification with Markov Field Aspect Models

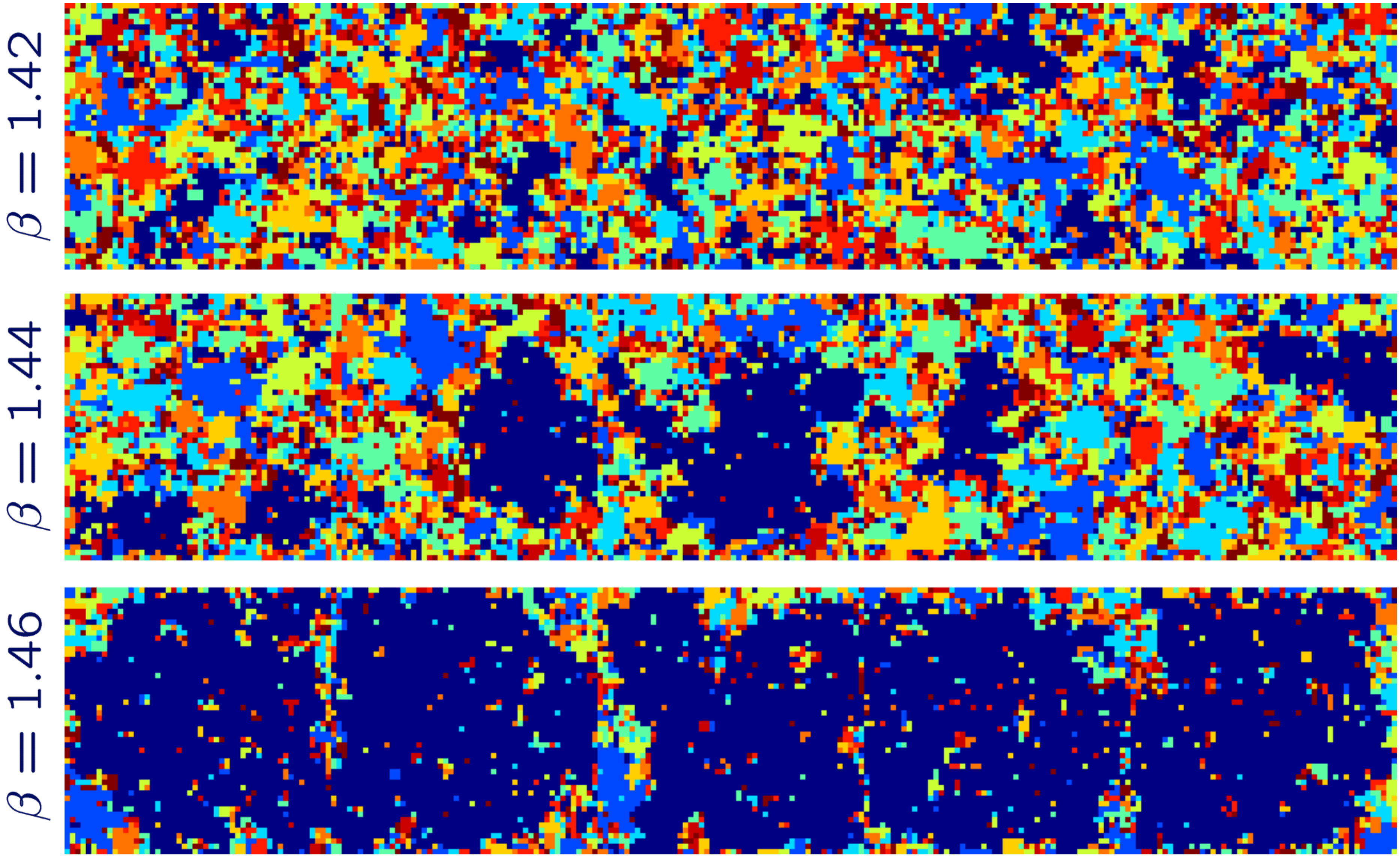
Verbeek & Triggs, CVPR 2007

Local: 74%

MRF: 78%



10-State Potts Samples



States sorted by size: largest in blue, smallest in red

1996 IEEE DSP Workshop

The Ising/Potts model is not well suited to segmentation tasks

R.D. Morris *X. Descombes*
J. Zerubia
INRIA, 2004, route des Lucioles, BP93, Sophia Antipolis Cedex, France.

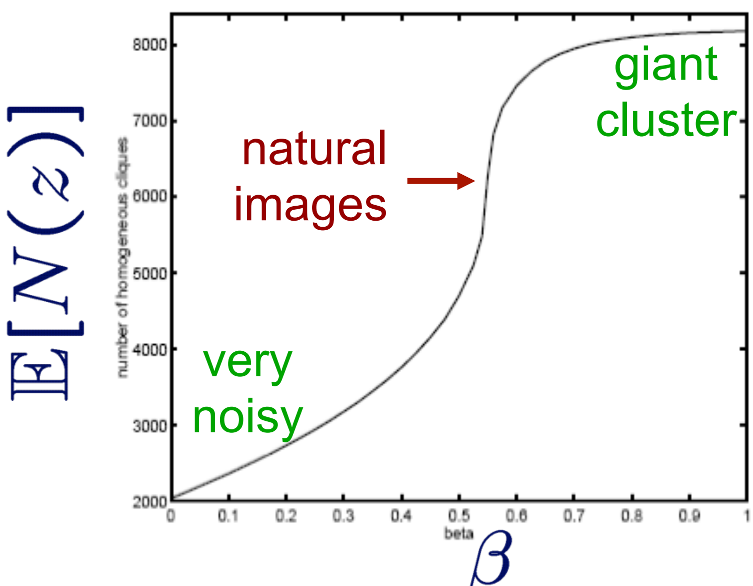


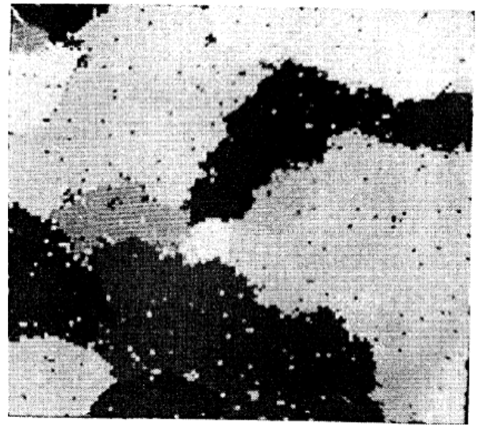
Figure 1. $< N(\mathbf{x}) >$ vs β for $64 \times 64 \times 4$-state Potts model

$N(z) \rightarrow$ number of edges on which states take same value

$\beta \rightarrow$ edge strength

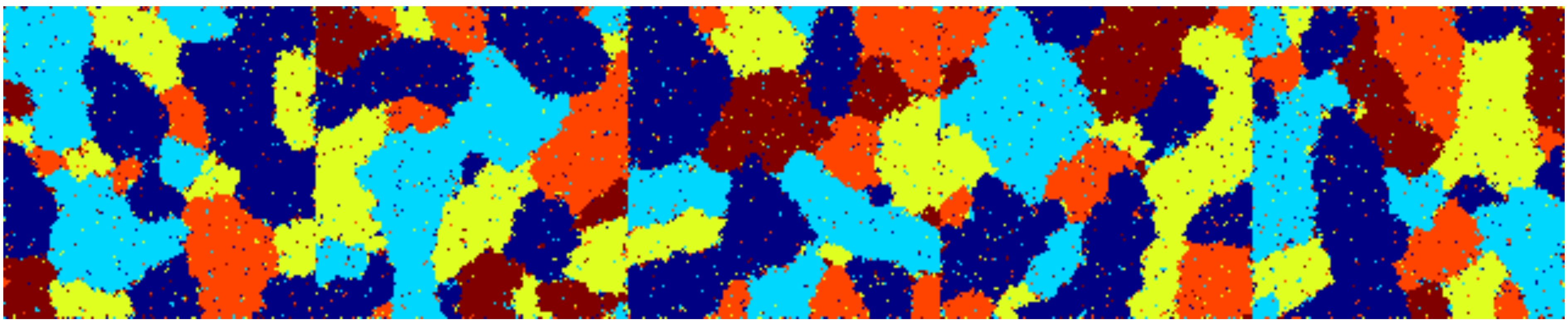
Even within the *phase transition* region, samples lack the *size distribution* and *spatial coherence* of real image segments

Geman & Geman, 1984

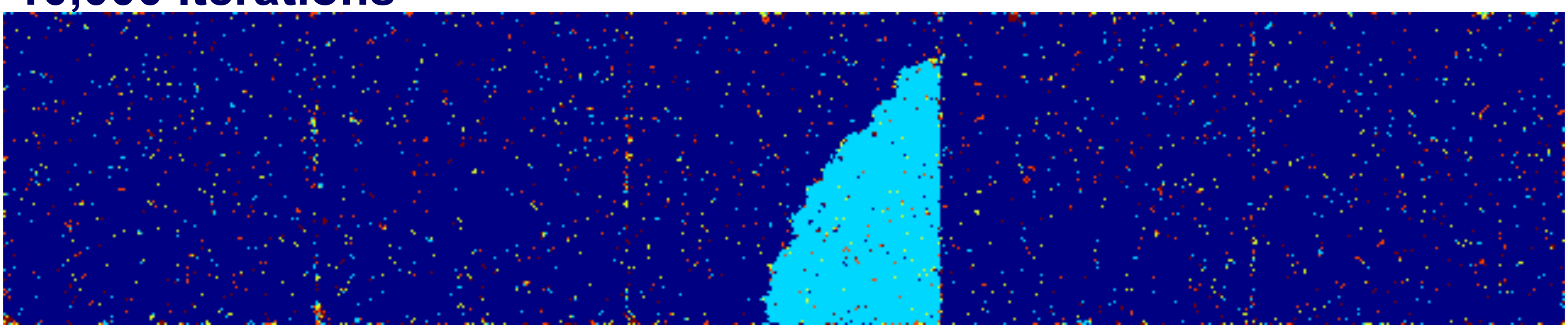


128 x128 grid
8 nearest neighbor edges
K = 5 states
Potts potentials: $\beta = 2/3$

200 Iterations

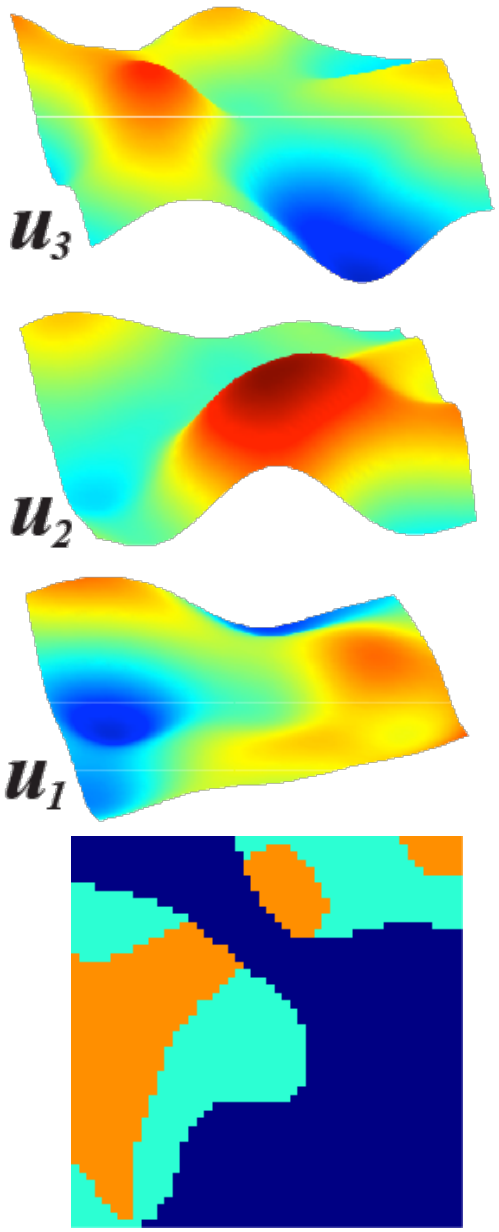


10,000 Iterations



Spatial Pitman-Yor Processes

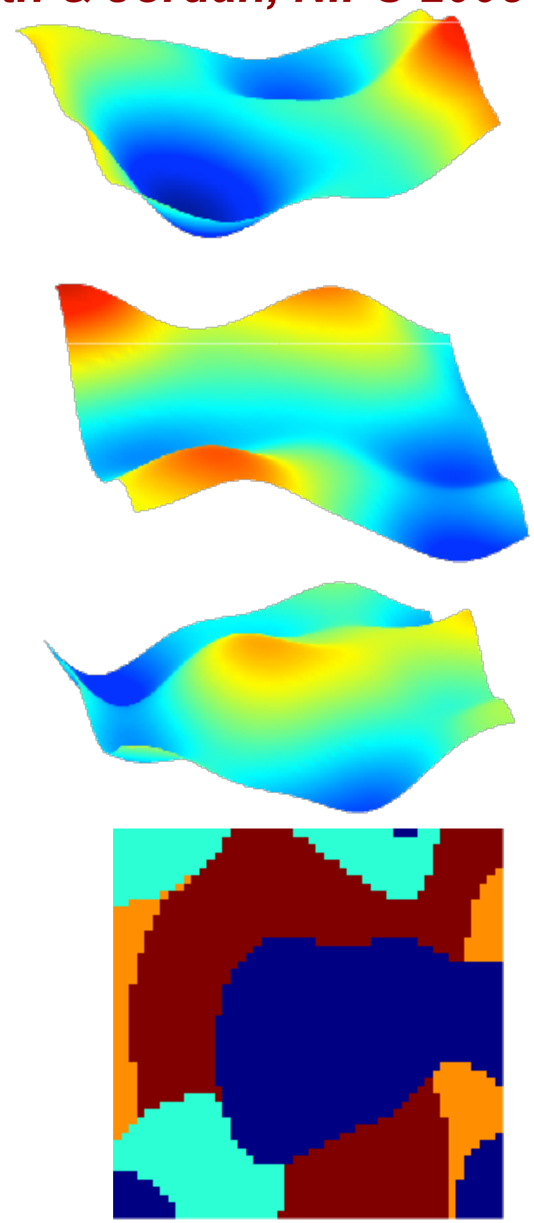
Sudderth & Jordan, NIPS 2008



- Cut random *surfaces* (Gaussian processes) with *thresholds*
- Surfaces define *layers* that occlude regions farther from the camera

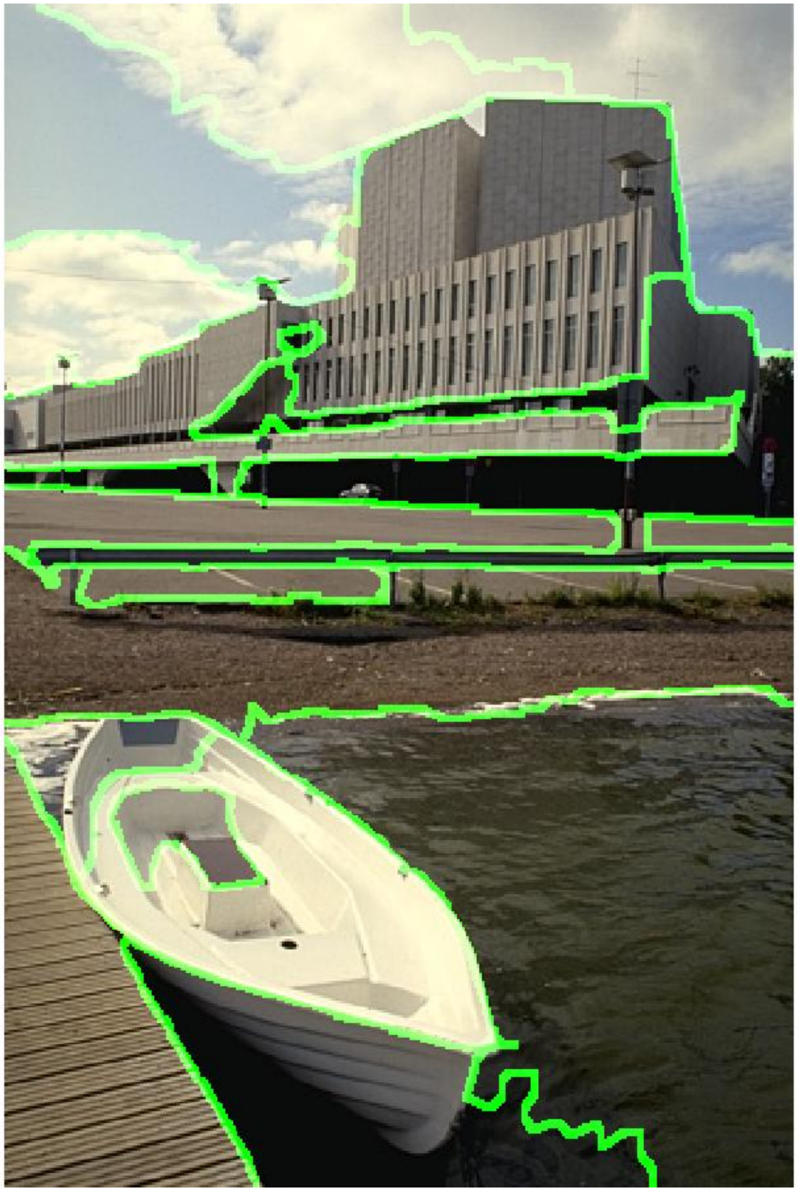
Technical Challenges

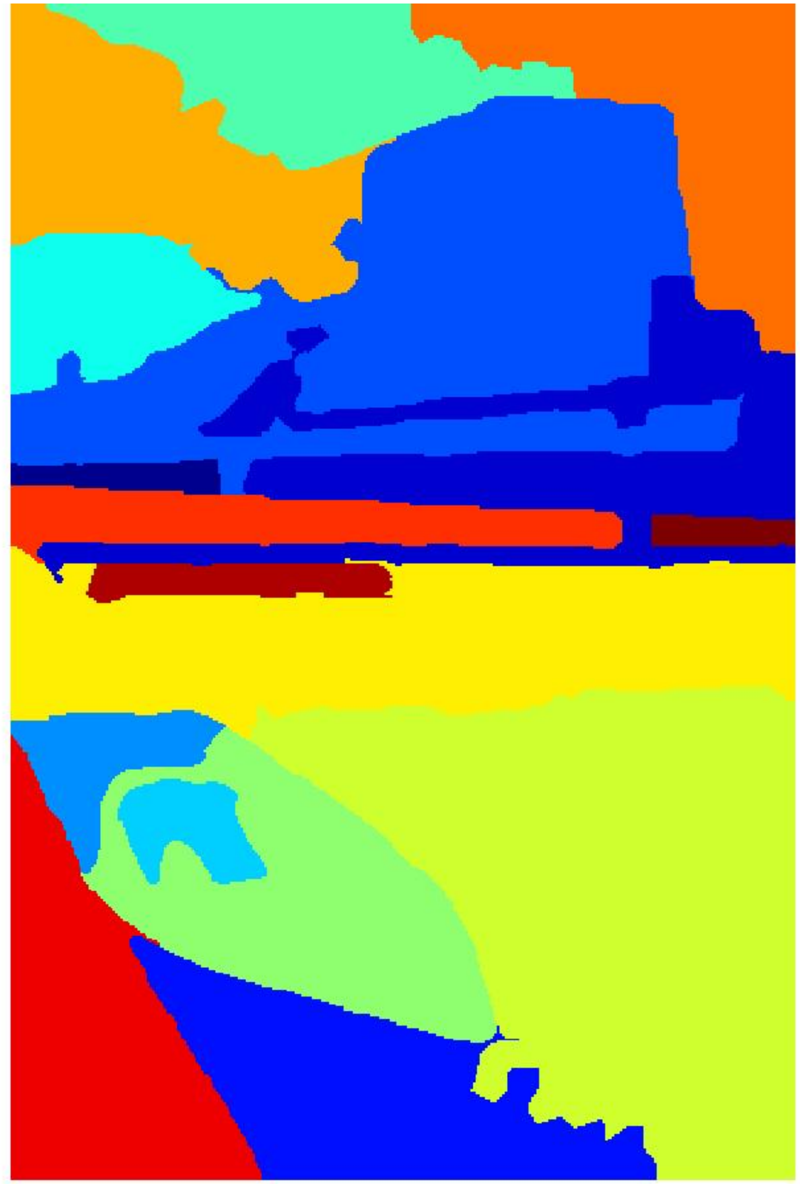
- Learn statistical *biases that are consistent* with human segments
- *Inference problem:* find the latent segments underlying an image



Improved Learning & Inference

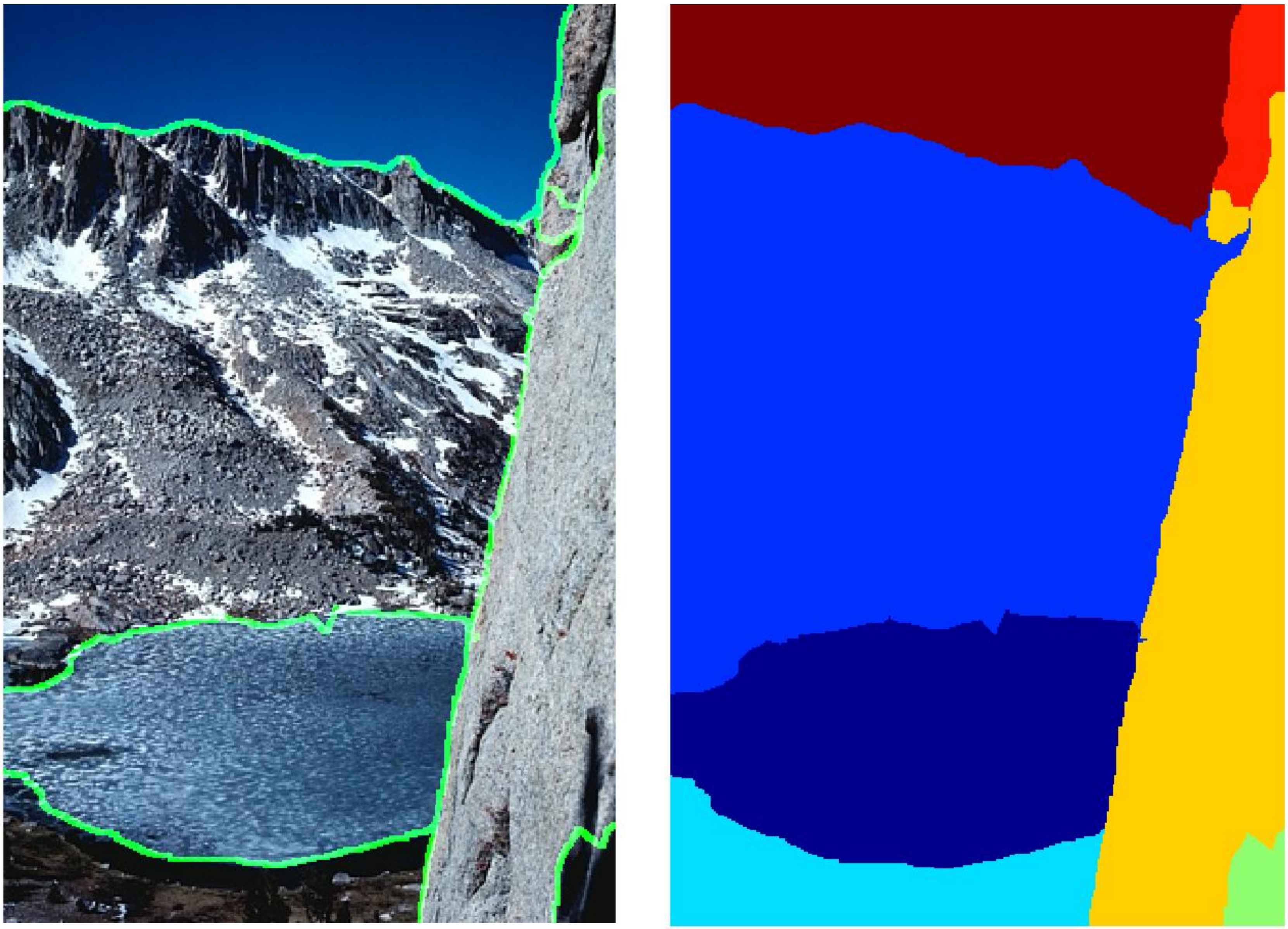
Ghosh & Sudderth, in preparation, 2011 (image from Berkeley Dataset)





Improved Learning & Inference

Ghosh & Sudderth, in preparation, 2011 (image from Berkeley Dataset)

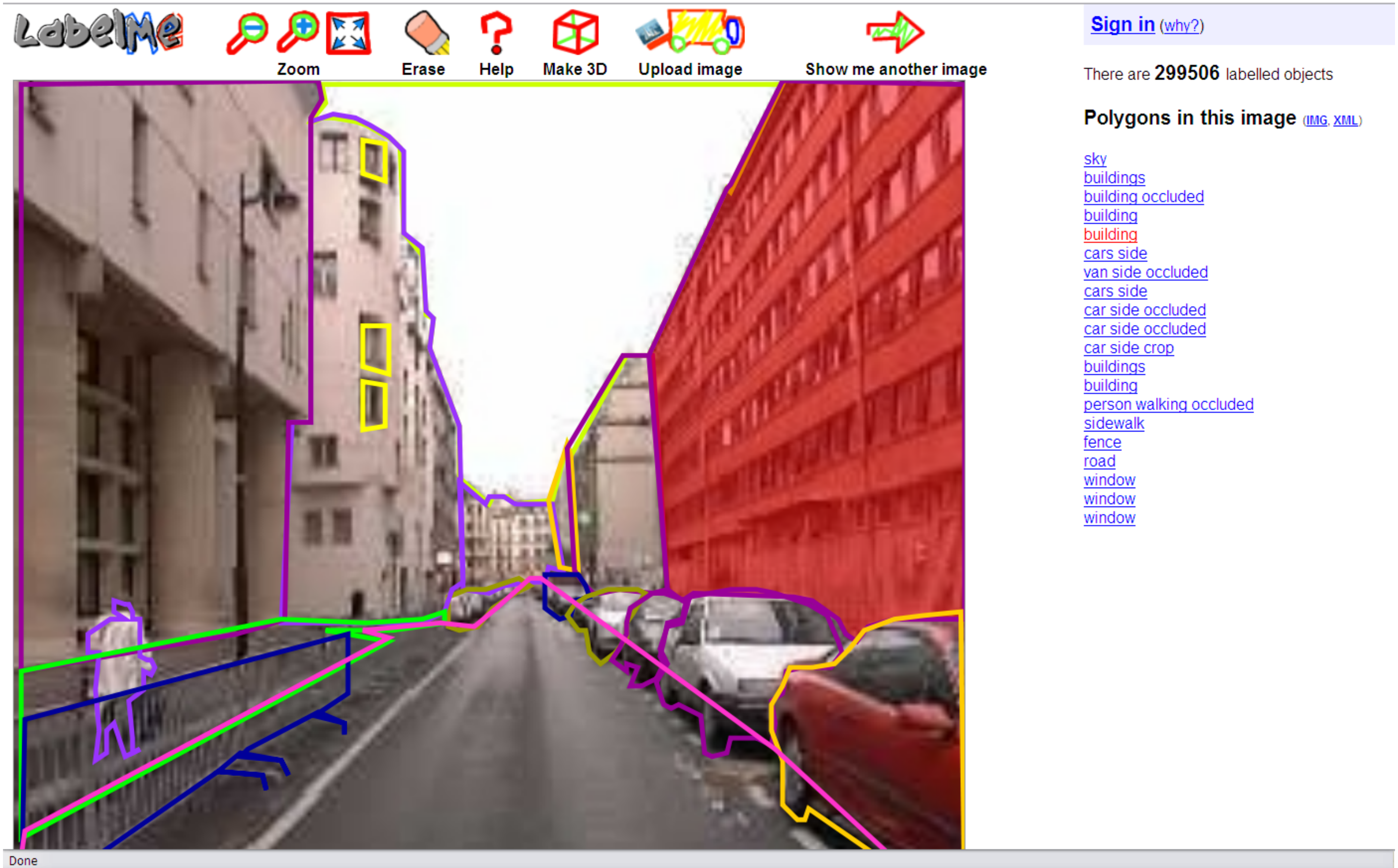


Improved Learning & Inference

Ghosh & Sudderth, in preparation, 2011 (image from Berkeley Dataset)

Showing only most likely mode, but model provides posterior distribution over (non-nested) segmentations of varying resolution and complexity.

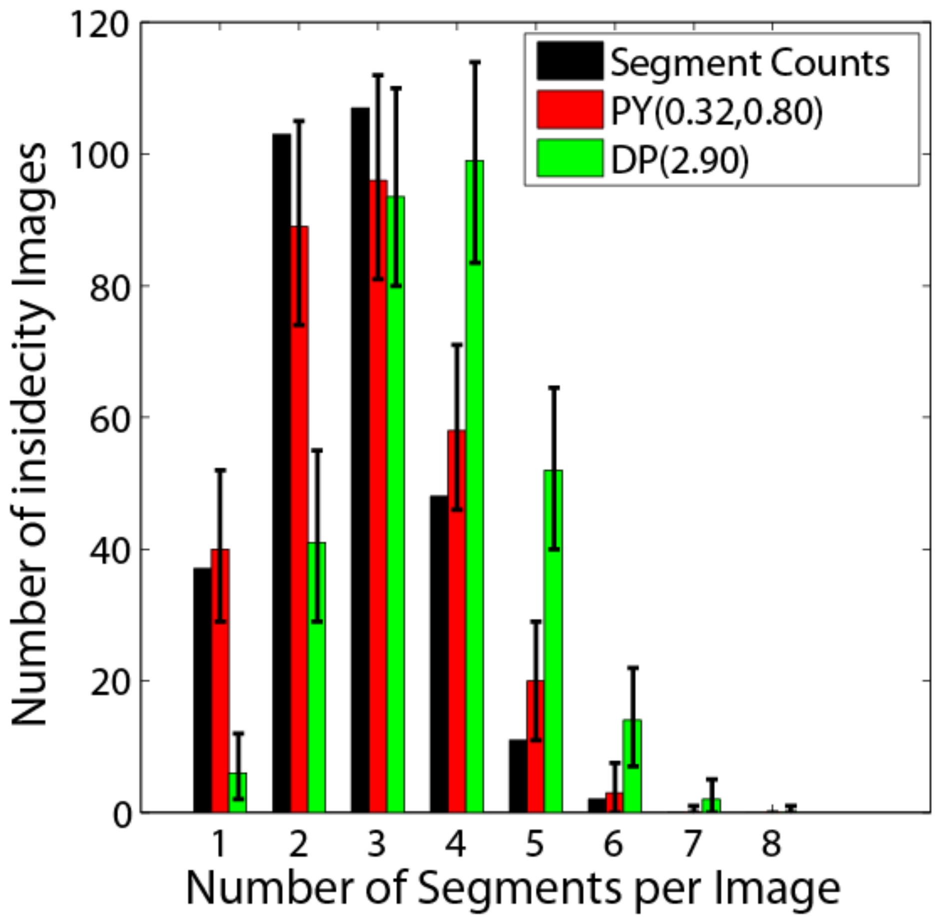
Human Image Segmentations



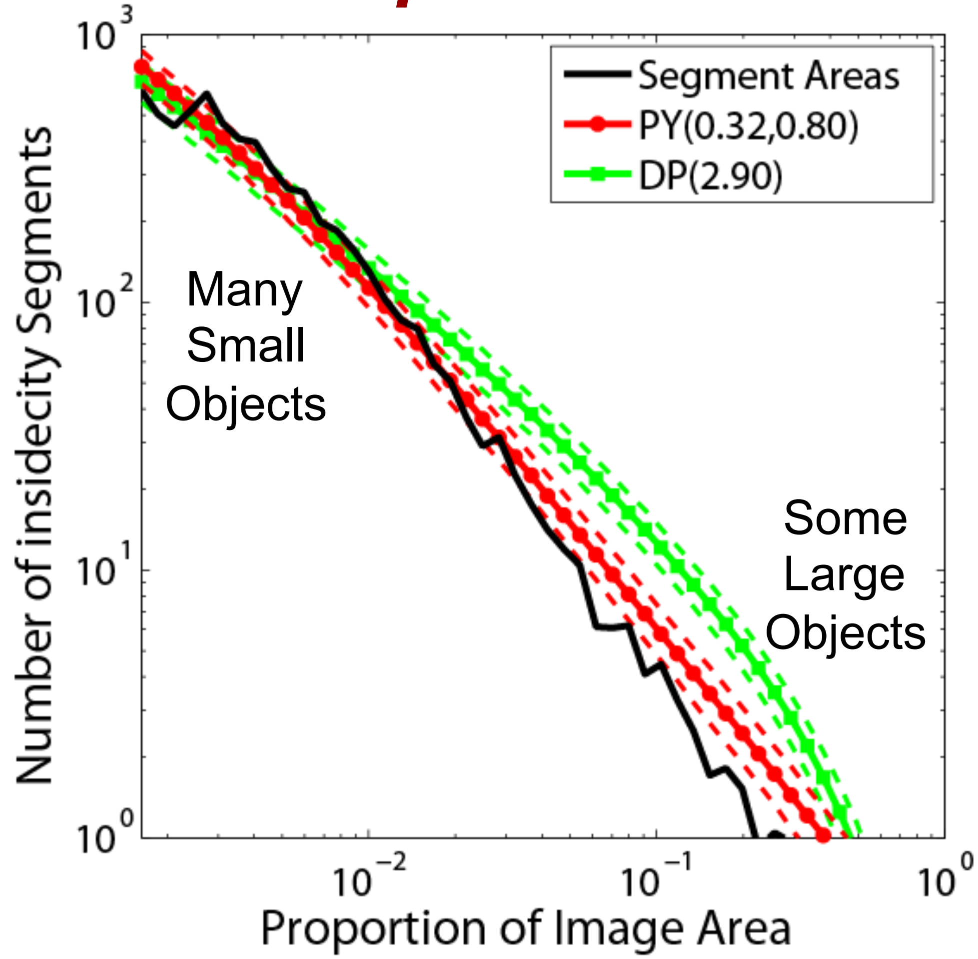
Labels for more than 29,000 segments in 2,688 images of natural scenes

Statistics of Human Segments

How many objects are in this image?

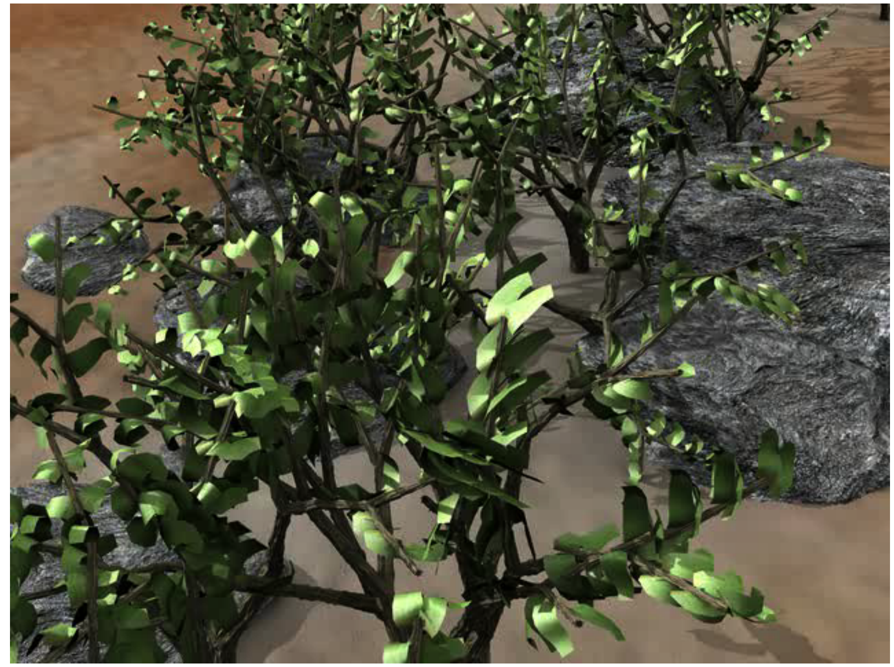


Object sizes follow a power law

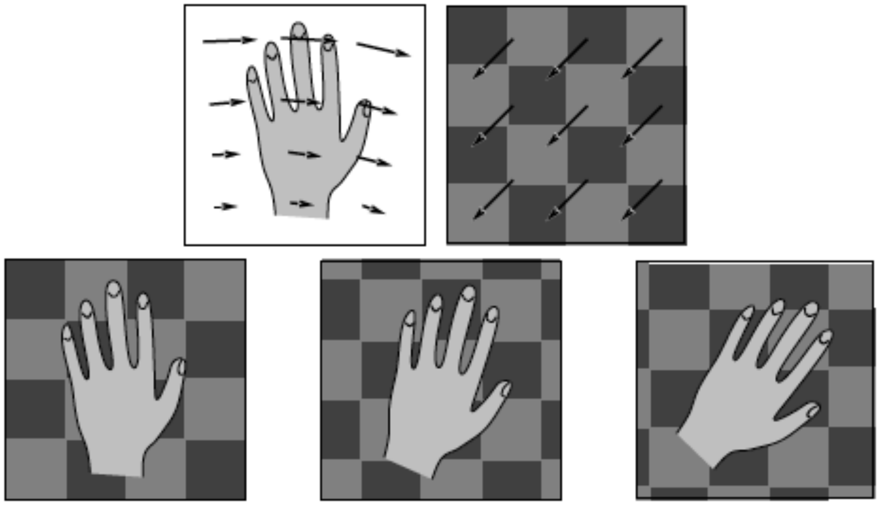


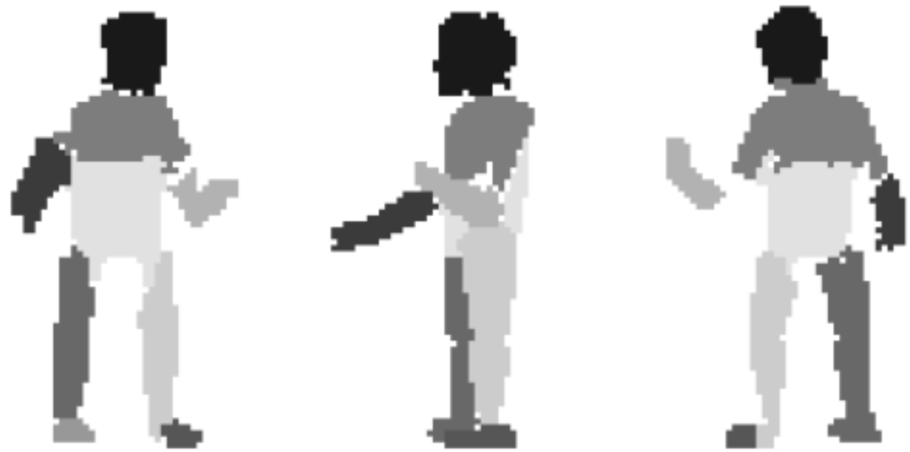
Labels for more than 29,000 segments in 2,688 images of natural scenes

Estimating Image Motion



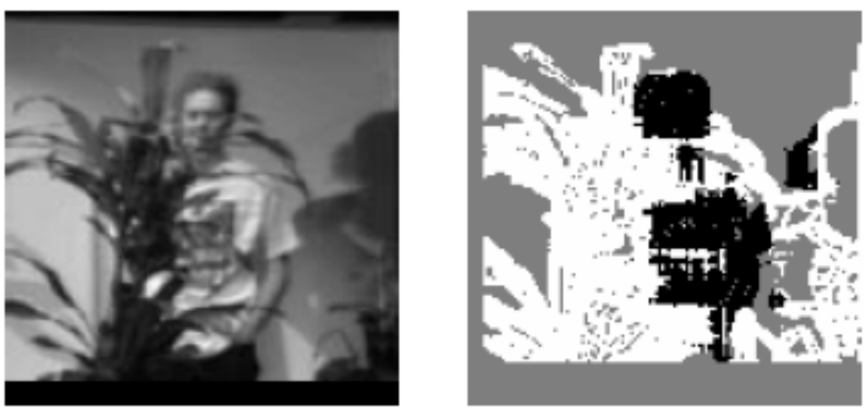
Motion in Layers





Wang & Adelson, 1994

Darrell & Pentland, 1991, 1995

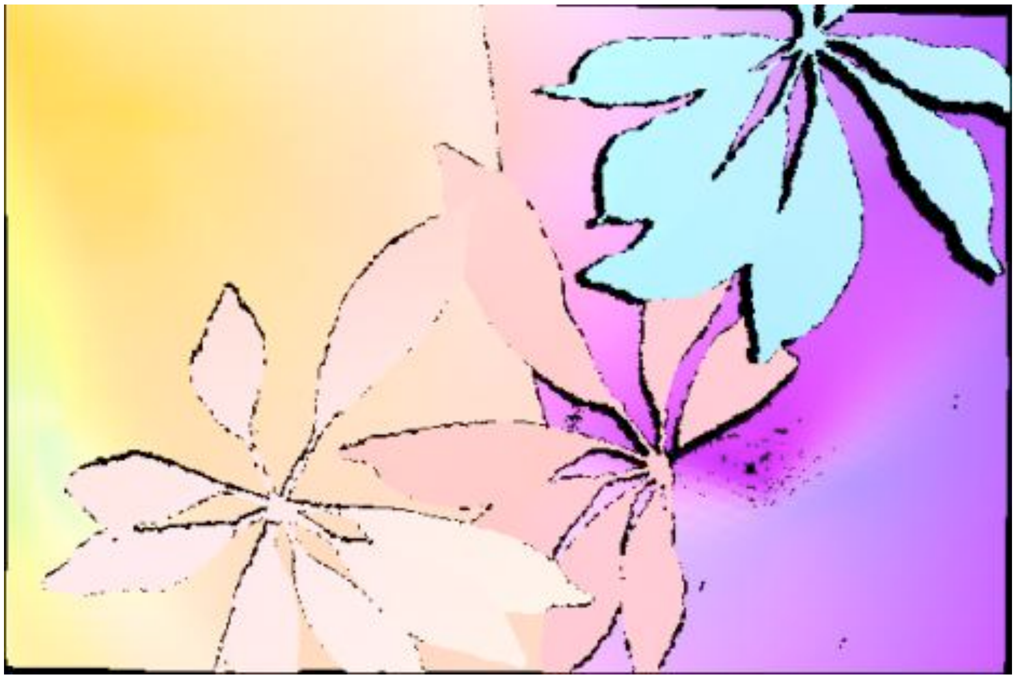


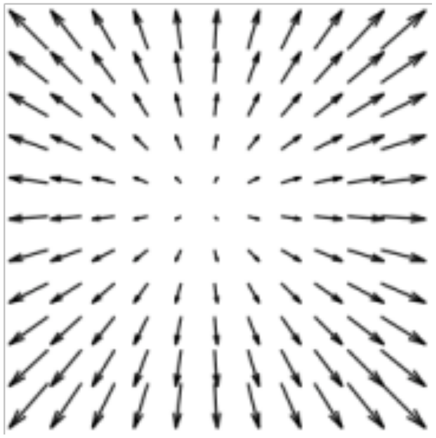
Weiss 1997

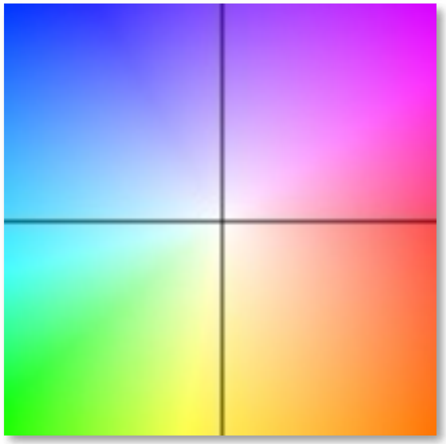
Jojic & Frey, 2001

Optical Flow Estimation

Middlebury Optical Flow Database (Baker et al., 2011)



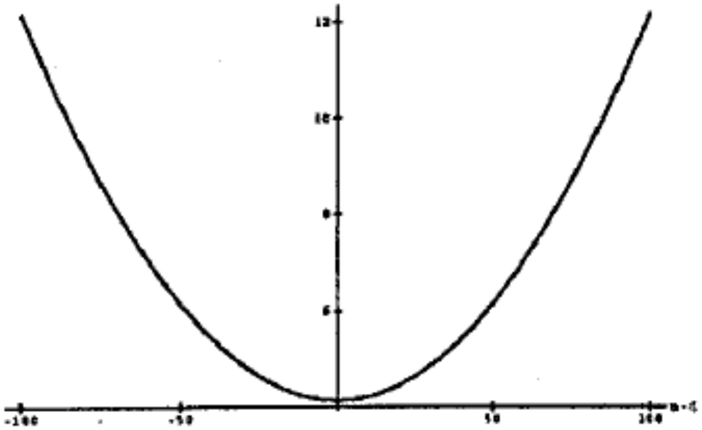


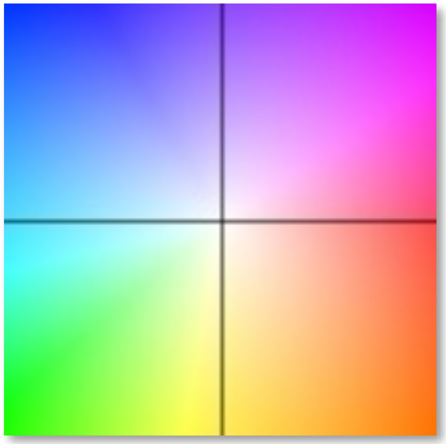


Ground truth optical flow (occluded regions in black, error not measured)

Optical Flow: A Brief History

Quadratic (Gaussian) MRF: Horn & Schunck, 1981

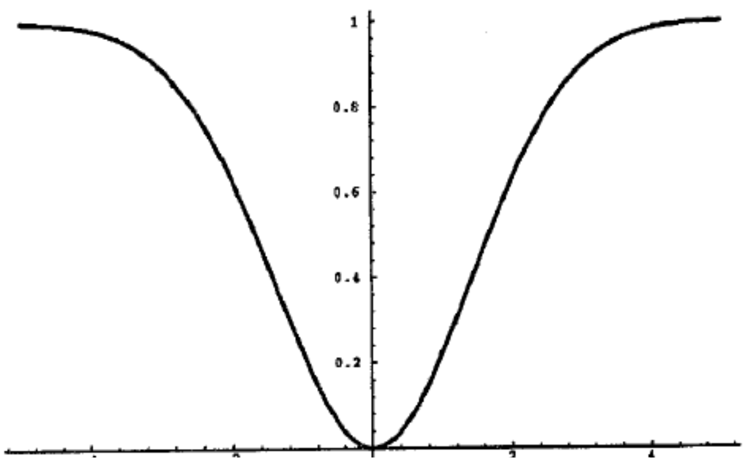


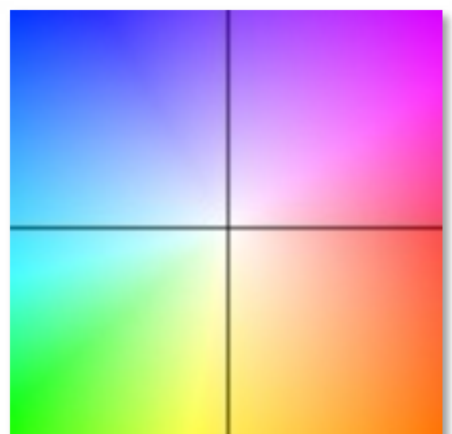


Their model with modern parameter tuning and inference algorithms

Optical Flow: A Brief History

Robust MRF: Black & Anandan, 1996; Black & Rangarajan, 1996

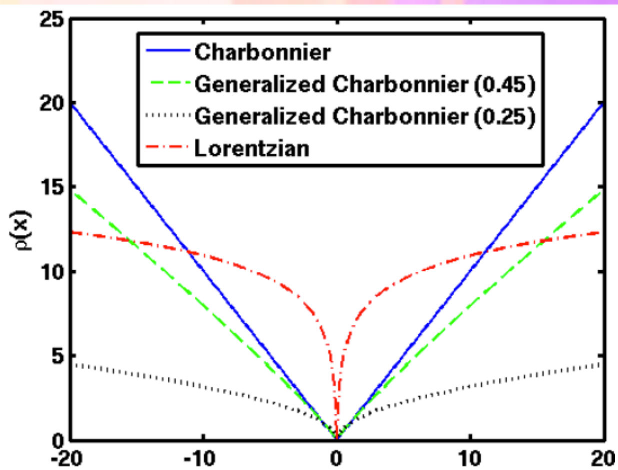


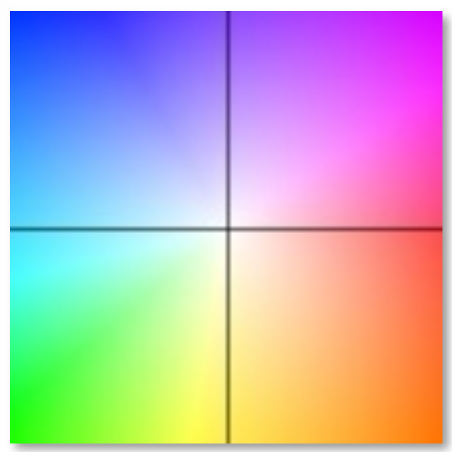


Their model with modern parameter tuning and inference algorithms

Optical Flow: A Brief History

Refined Robust MRF: Sun, Roth, & Black, 2010



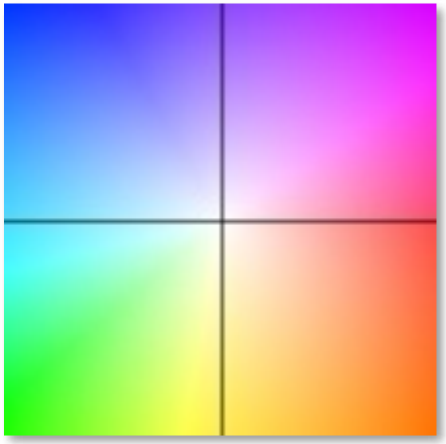


Middlebury benchmark leader in mid-2010

Optical Flow in Layers

Sun, Sudderth, & Black, NIPS 2010

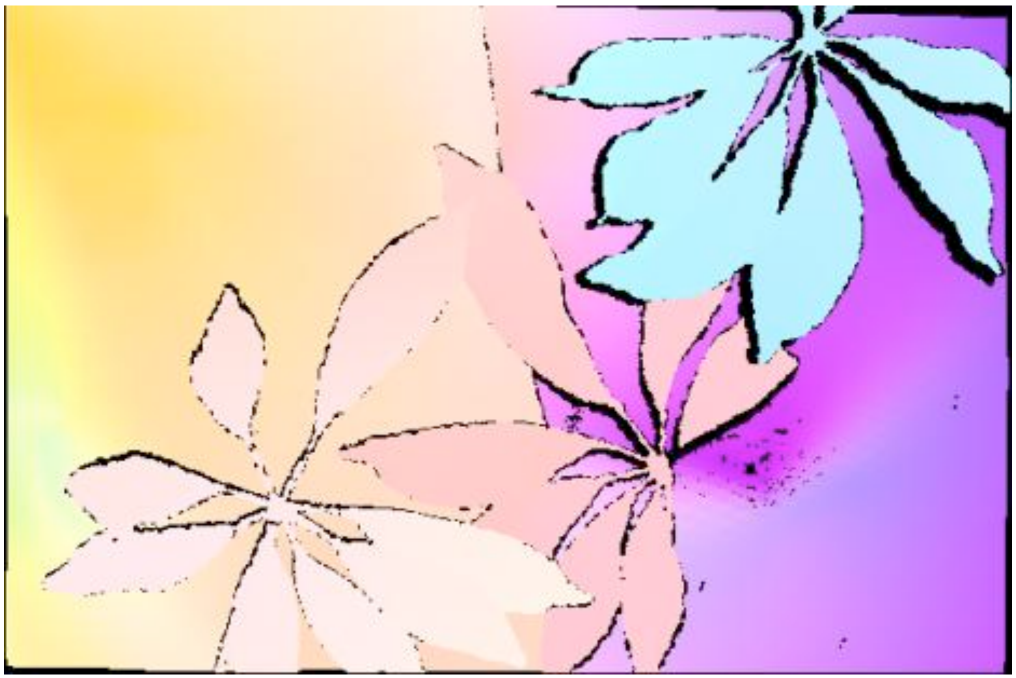
Explicitly models occlusion via support of ordered layers, rather than treating as unmodeled outlier.

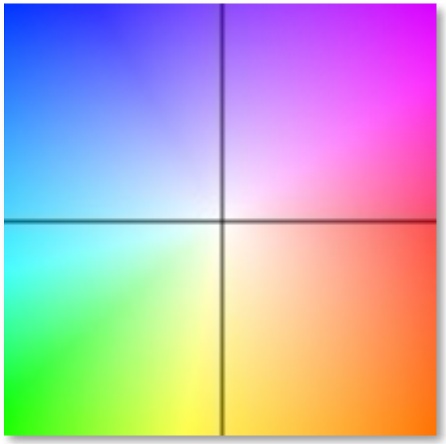


Current lowest average error on Middlebury benchmark

Optical Flow Estimation

Ground Truth: Middlebury Optical Flow Database

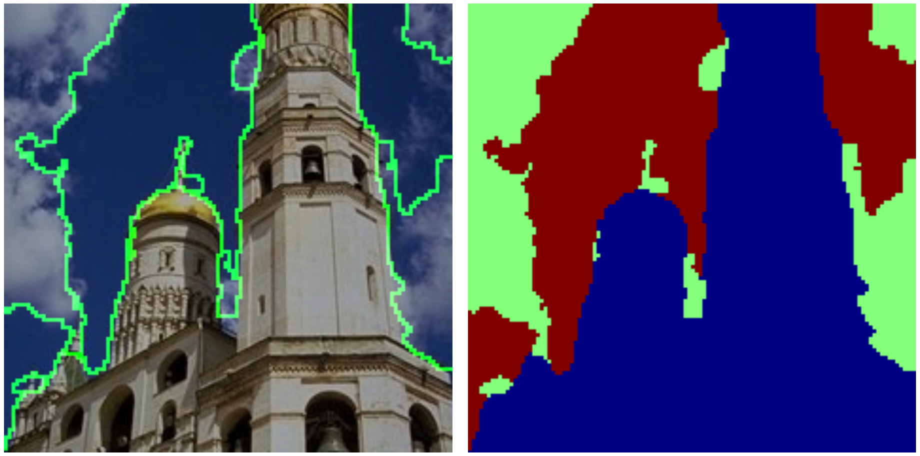




Ground truth optical flow (occluded regions in black, error not measured)

Layers, Depth, & Occlusion

Older layered models had unrealistically simple models of layer flow & shape, or did not explicitly capture depth order when modeling occlusions.



Questions?