

# A Tour of Image Segmentation

Dr. Scott Cohen | Adobe Research

# Image Segmentation

- Segment image pixels into different classes



- 2 Classes : Boy (Foreground), Not Boy (Background)

- Scott Cohen, Gregg Wilensky, Jeff Chien (Adobe)

# Segmentation Variations

- How many classes?
- How are the classes defined?
- What features are used to compute the segmentation?
- Hard Segmentation or Soft Segmentation?
  - Hard: a pixel is assigned to exactly one class
  - Soft: a pixel may be assigned to more than one class
- Automatic or Interactive computation?
  - What user input is provided?
- How many images are segmented?

# Interactive Binary Segmentation

- User Strokes (Scribbles) in **Foreground** and **Background**



User Intent

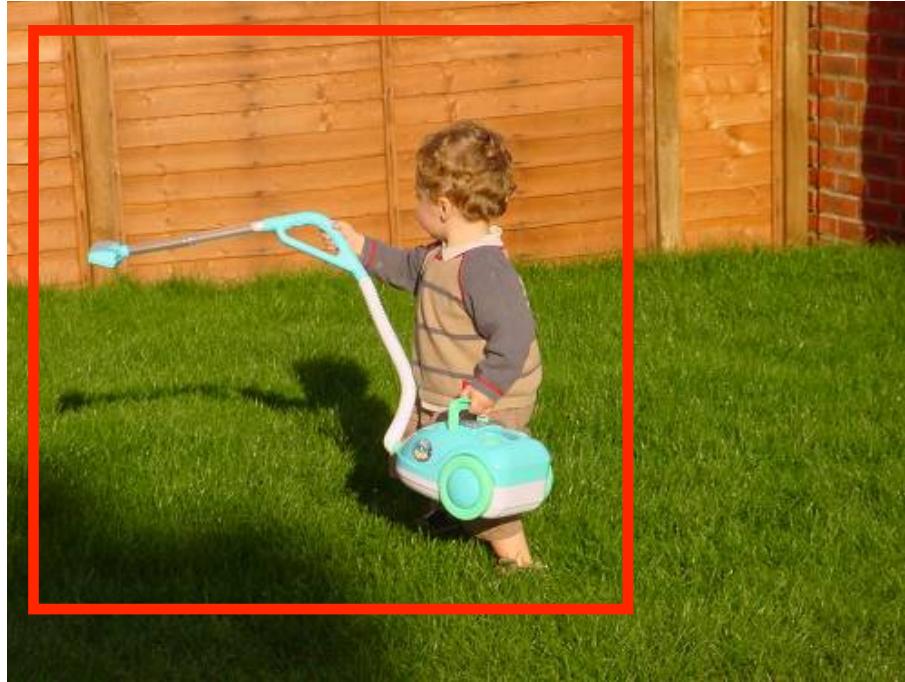


Desired Segmentation

Interactive Graph Cuts for Optimal Boundary & Region Segmentation  
Boykov, Jolly (ICCV 2001)

# Interactive Binary Segmentation

- GrabCut: Draw a rectangle around the object to select



GrabCut: Interactive Foreground Extraction using Iterated Graph Cuts  
Rother, Kolmogorov, Blake (Siggraph 2004)

# Interactive Binary Segmentation

- Magnetic Lasso: Trace around the object to select



Intelligent Scissors for Image Composition  
Mortensen, Barrett (Siggraph 1995)

# Automatic Binary Segmentation

- Segment the “Salient” Region



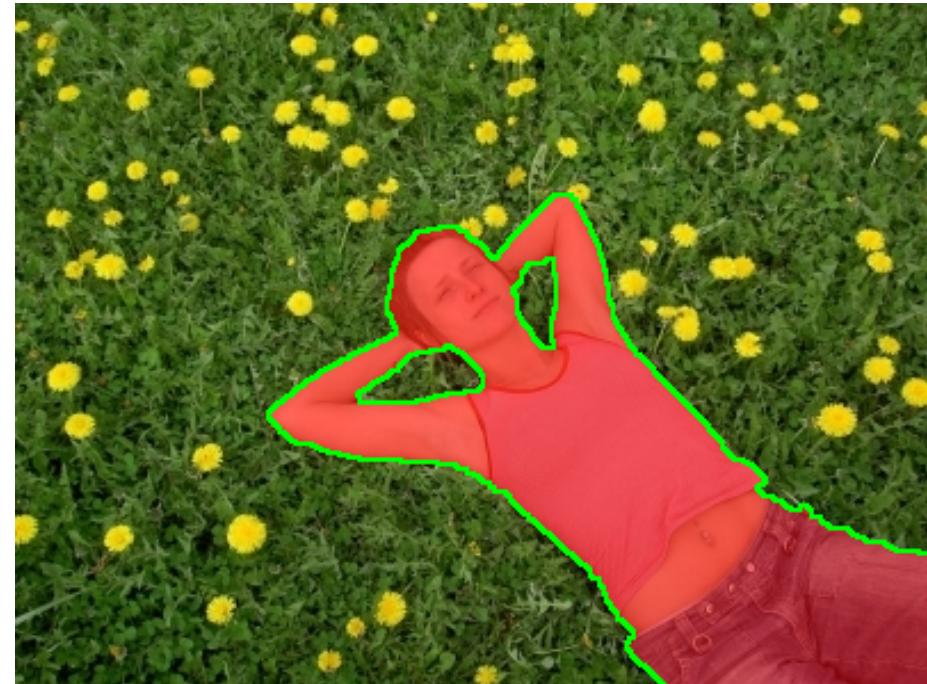
# Automatic Binary Segmentation

- Segment the “Salient” Region



# Automatic Binary Segmentation

- Segment the “Salient” Region



# Automatic Binary Segmentation

- Segment In-focus Regions



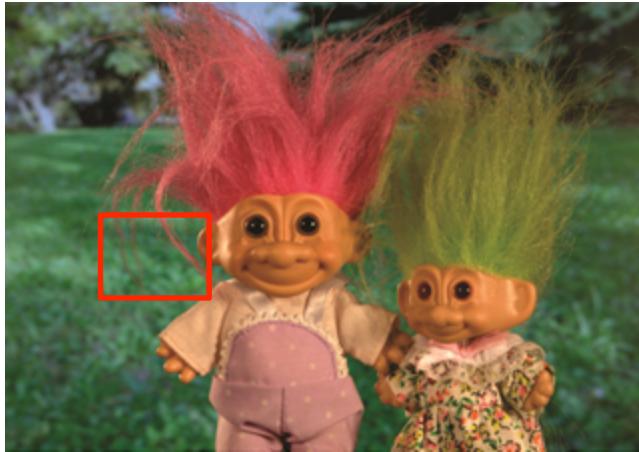
Input Image



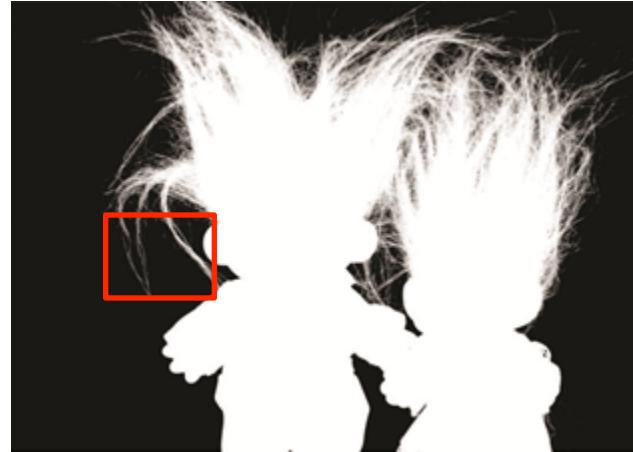
Output Segmentation

# Image Matting

- Soft Binary Segmentation  $I_p = \alpha_p F_p + (1 - \alpha_p) B_p$



Input Image  $I$



Output Segmentation  $\alpha \in [0, 1]$



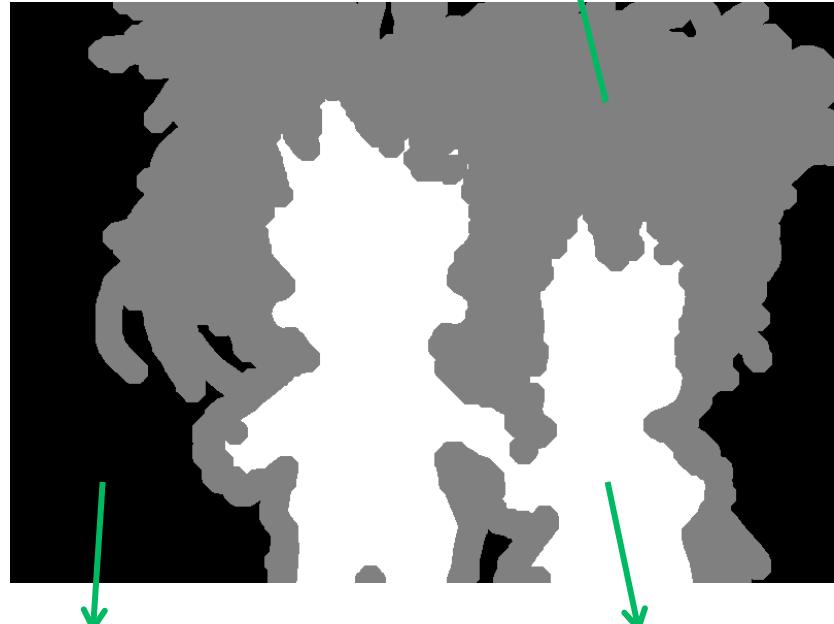
# Image Matting : Compositing Application

$$I_p = \alpha_p F_p + (1 - \alpha_p) B_p \quad \hat{I}_p = \alpha_p^{01} H_p + (1 - \alpha_p^{01}) \hat{B}_{pp}$$



# Image Matting : Trimap Input

$$I_p = \alpha_p F_p + (1 - \alpha_p) B_p$$



Known Background

Known Foreground

Unknown Region

## Demo: Interactive Matting

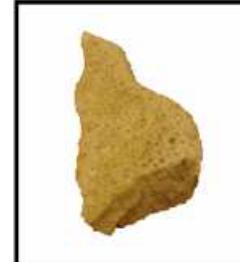
- Brian Price, Scott Cohen (Adobe)

# Co-Segmentation

- Segment the object in common in multiple images



Input Image Pair



Cosegmentation



Input Image pair



Cosegmentation

Cosegmentation of Image Pairs by Histogram Matching  
Rother, Kolmogorov, Minka, Blake (CVPR 2006)

# Co-Segmentation Methods

Method	Foregrounds	Backgrounds	Automatic or Interactive?	# of Images
Histogram Matching (CVPR06)	Same	Different	Automatic	2

# Co-Segmentation : Similar BGs, Interactive, Many Images

iCoseg: Interactive Co-segmentation with Intelligent Scribble Guidance  
Batra, Kowdle, Parikh, Luo, Chen (CVPR 2010)

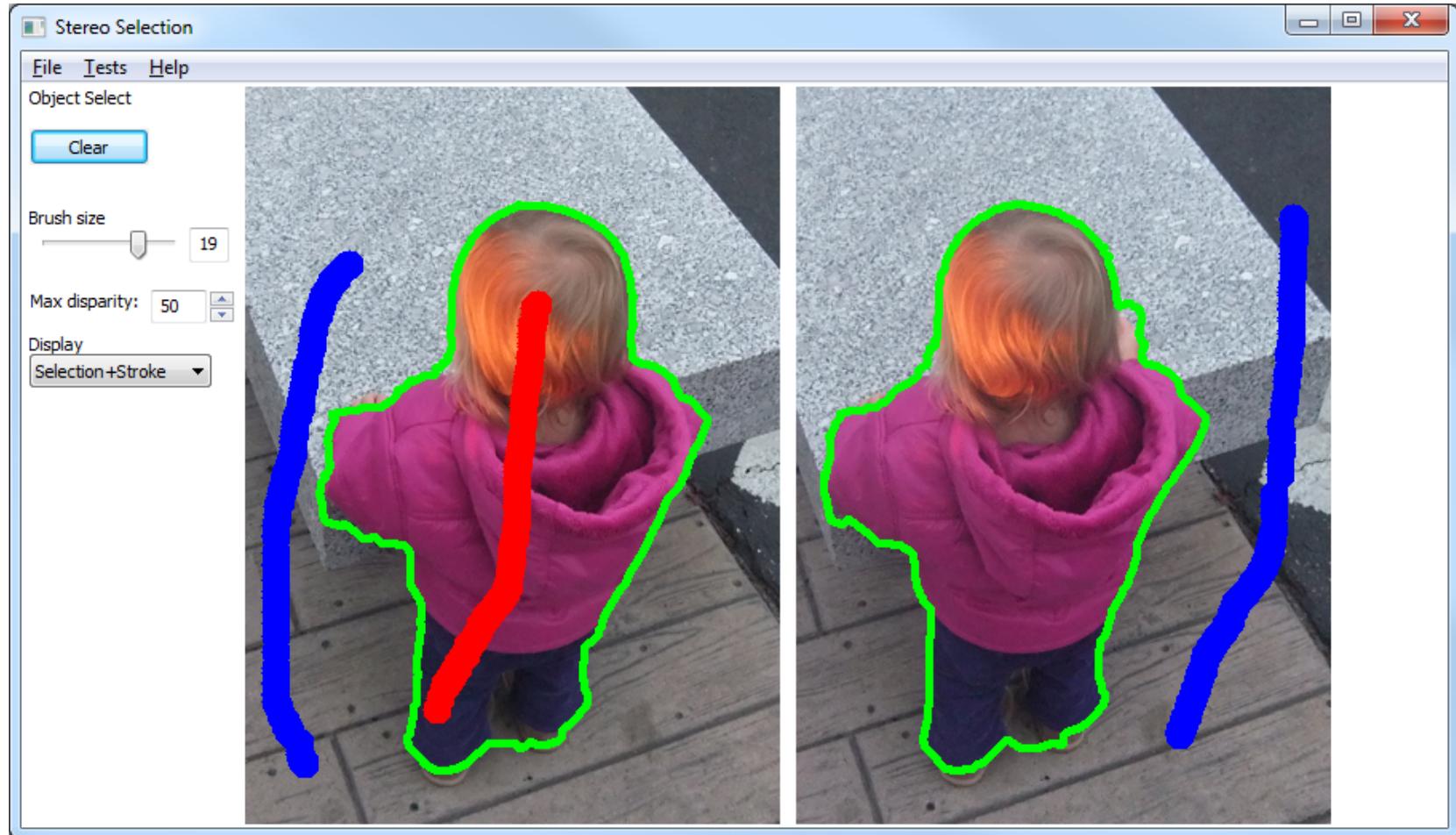


# Co-Segmentation Methods

Method	Foregrounds	Backgrounds	Automatic or Interactive?	# of Images
Histogram Matching (CVPR06)	Same	Different	Automatic	2
iCoseg (CVPR10)	Similar	Similar	Interactive	Many

# Stereo Co-Segmentation : Same BGs, Interactive, 2 Images

StereoCut: Consistent Interactive Object Selection in Stereo Image Pairs  
Price, Cohen (ICCV 2011)

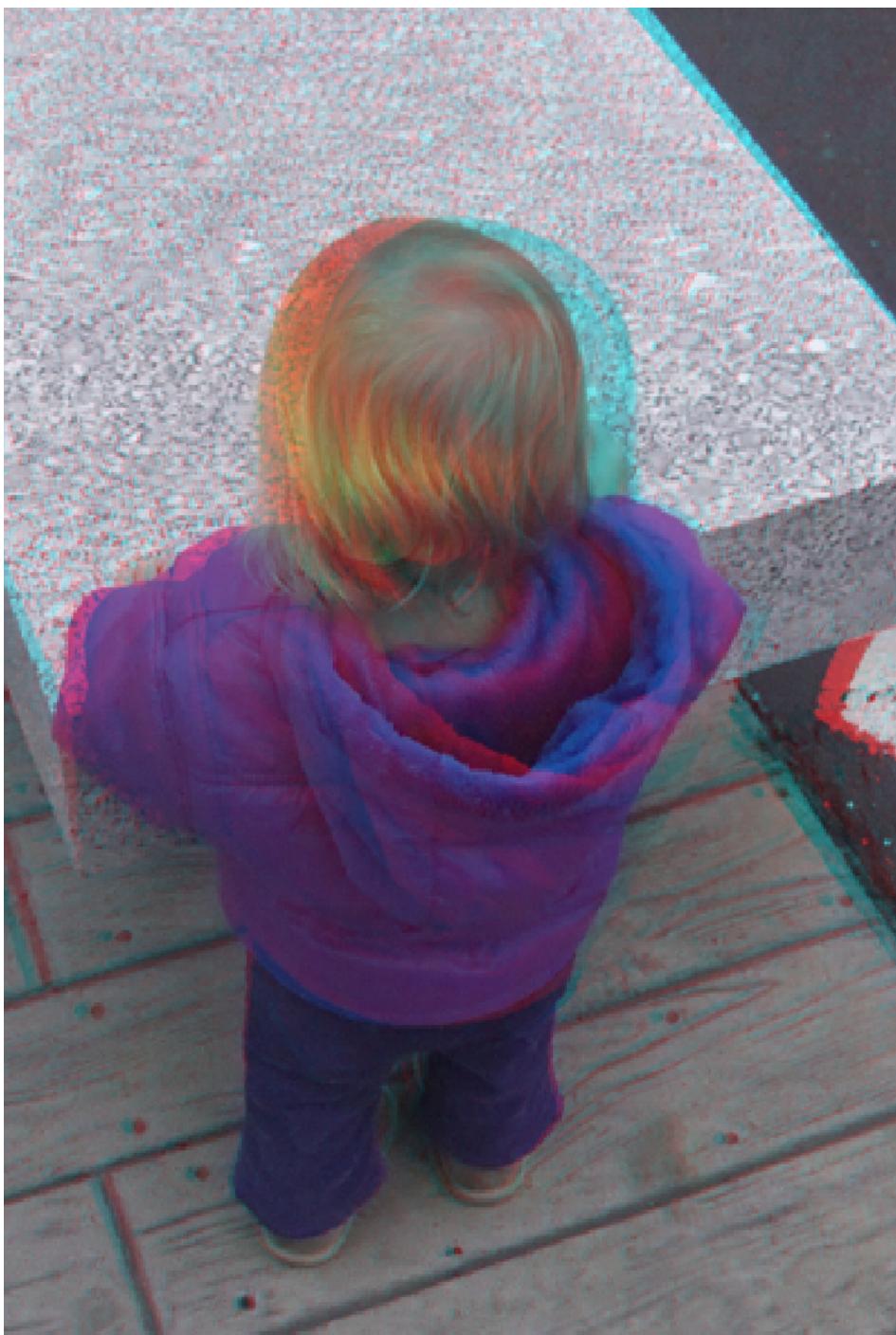
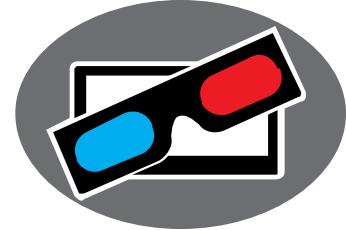


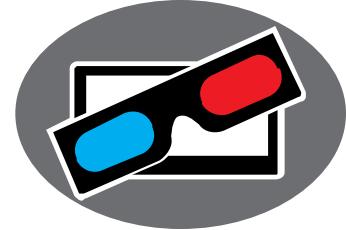
# Stereo Co-Segmentation Applications

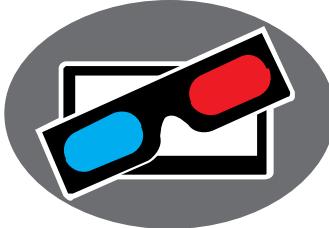
- Localized Stereo Editing
- Stereo Inpainting: remove Co-Segmented object from the stereo picture

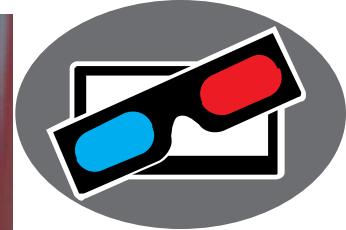


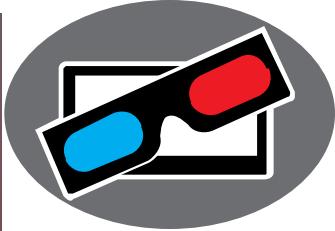
PatchMatch-based Content Completion of Stereo Image Pairs  
Morse, Howard, Cohen, Price (3DimPVT 2012)

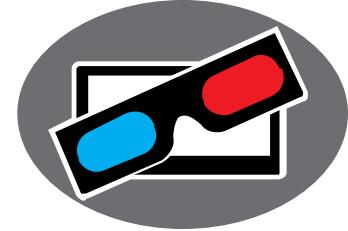














"Oo my first day, I  
got dihu baddly by a  
giggle here brother." - Lelala

## WHAT'S YOUR IDENTITY?

How do you define yourself by the clothes you wear and the products you buy. By the people you spend your free time with. By the things you do. By the way you express yourself and make yourself feel good.

### SOUTH CAMPUS COMMUNITY SCHOOL

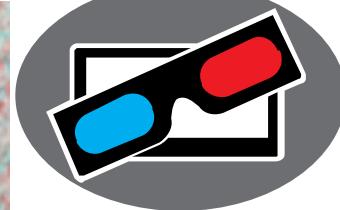
South Campus Community School is the oldest, most well-known private school in the region. It has a long history of being progressive and inclusive. It has a diverse student body, including the highest minority of students, and the National Honor Society. It is the largest school in the area. It offers a variety of academic and extracurricular activities. South Campus Community School has become well-known for its strong academic performance. In addition, it has a rich tradition of service, both locally and internationally. The school is located in the heart of the city, making it easily accessible to students and their families.

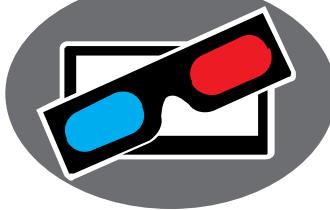
Lelala Spikes  
Instructor of Health & Wellness  
Communication Teacher  
page 200-200-200-200  
Email: [lspikes@scs.org](mailto:lspikes@scs.org)

### ST SEASIDE ELEMENTARY SCHOOL

St. Seaside Elementary School is a public elementary school located in the St. Seaside neighborhood of Seattle. It serves students from Kindergarten through 5th grade. The school has a diverse population of students from various backgrounds. The school is known for its focus on social-emotional learning and its commitment to equity and inclusion. The school has a strong tradition of academic achievement and community involvement.

Design for Social Justice  
Academy, Industrial Arts, and  
Performance and Production  
2007-2008-2009-2010-2011  
Tutoring





"Oh my first daya! I  
got hit badly by a  
girl in a little brother."

SOUTH CAMDEN COMMUNITY SCHOOLS

**Adobe Kindle Touch** [www.adobe.com/reader/kindle](#) Adobe's Kindle reader, with [Kindle Web Edition](#), provides a general reading experience that has established itself as one of the most popular e-reader platforms, and includes the digital version of the [National Bestseller List](#) and a host of other features.

**Lobster**  
Traditional South Carolina style  
Chesapeake Bay style  
New England style  
Vermont style

ST SEASIDE ELEMENTARY SCHOOL

...and have Wobbe and Meier made an *independence* committee there which regulates and sees to it that it matches up to the *independence* principle which we have. The *independence* committee has been established and it is now in operation. It is a *non-political* committee which consists of the *independence* members themselves. It is a *non-political* committee which consists of the *independence* members themselves. It is a *non-political* committee which consists of the *independence* members themselves.













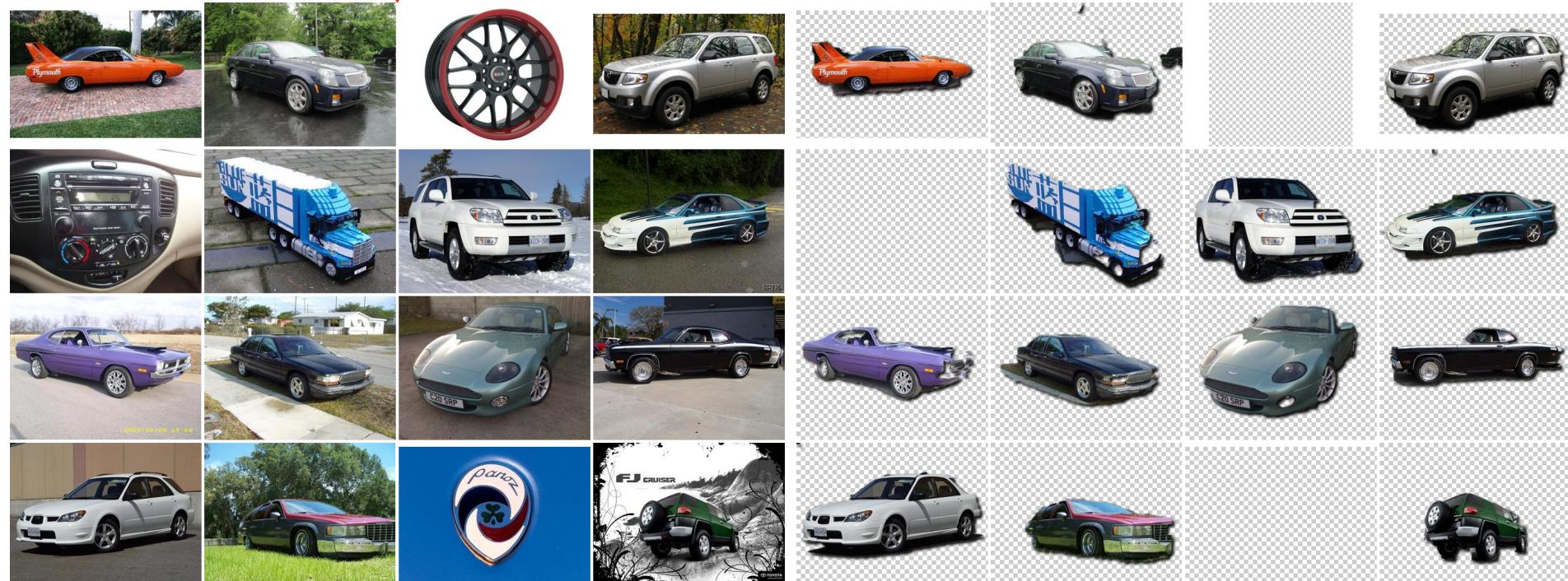
# Co-Segmentation Methods

Method	Foregrounds	Backgrounds	Automatic or Interactive?	# of Images
Histogram Matching (CVPR06)	Same	Different	Automatic	2
iCoseg (CVPR10)	Similar	Similar	Interactive	Many
Stereo (ICCV11)	Same	Same	Interactive	2

# Co-Segmentation : Similar BGs, Automatic, Many Images

Unsupervised Joint Object Discovery and Segmentation in Internet Images  
Rubinstein, Joulin, Kopf, Liu (CVPR 2013)

“Car” Internet Search



Input

Output

# Co-Segmentation Methods

Method	Foregrounds	Backgrounds	Automatic or Interactive?	# of Images
Histogram Matching (CVPR06)	Same	Different	Automatic	2
iCoseg (CVPR10)	Similar	Similar	Interactive	Many
Stereo (ICCV11)	Same	Same	Interactive	2
Object Discovery (CVPR13)	Similar, but more variation than iCoseg	Similar and Different	Automatic	Many

# Scene Parsing | Semantic Segmentation

- Label each pixel in an image with its semantic category



Input Image

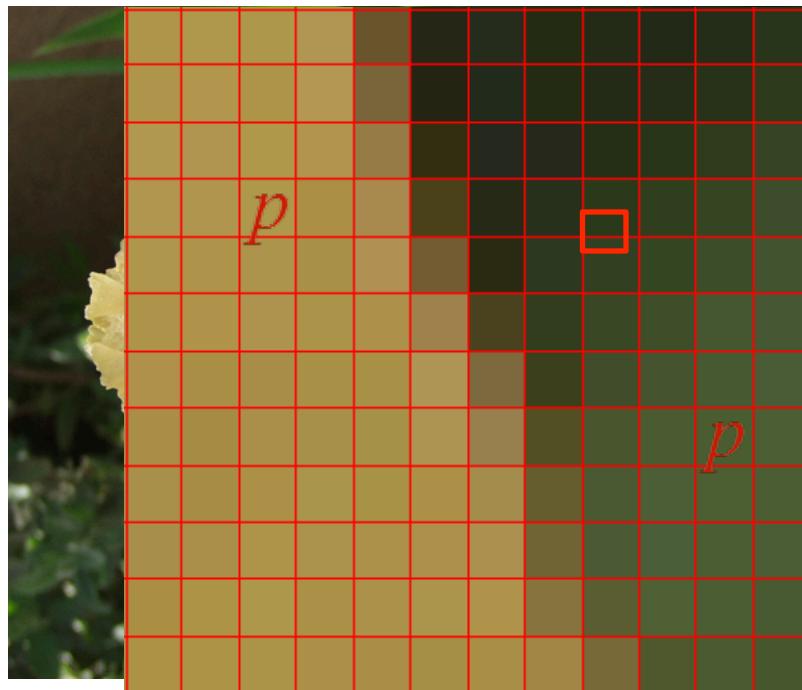


Desired Output

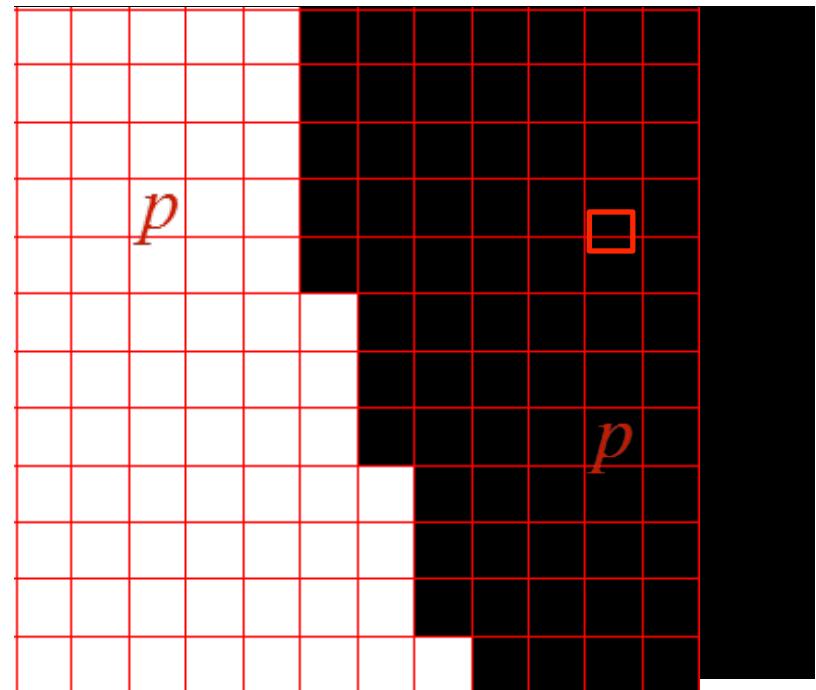
# Details About Some Segmentation Methods

- Common Framework
- Notation: Segmentation  $X = \{x_p\}$

$$I = \{I_p\}$$



$$x_p = 1 \quad X = \{x_p\} \quad x_p = 0$$

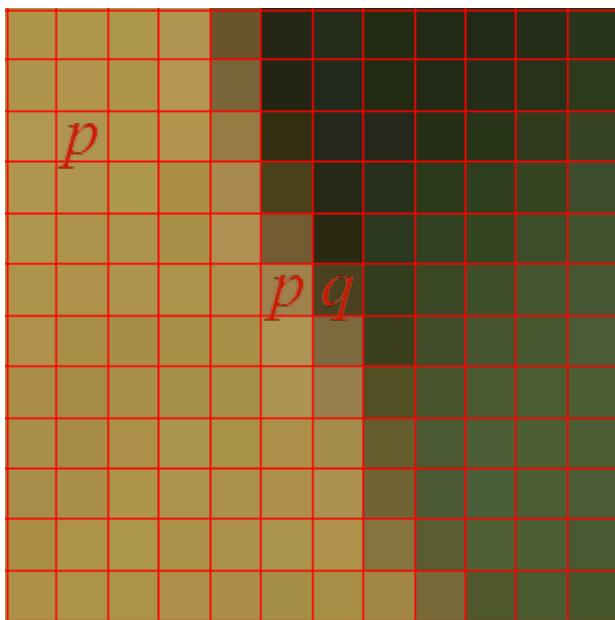


# Common Energy Minimization Framework

- Energy function to measure quality of segmentation  $X = \{x_p\}$

$$E(X) = \sum_p D_p(x_p) + \lambda \sum_{p,q \in N} V_{pq}(x_p, x_q)$$

- \* Global minimum found by min graphcut / maxflow algorithms



## Data Term

$D_p(0)$  = cost of labeling pixel p as BG

$D_p(1)$  = cost of labeling pixel p as FG

## Smoothness Term

$V_{pq}(1,0)$  = cost of p as FG, q as BG

$V_{pq}(0,1)$  = cost of p as BG, q as FG

$V_{pq}(0,0)$  = cost of p as BG, q as BG = 0

$V_{pq}(1,1)$  = cost of p as FG, q as FG = 0

# Energy E for Stroke-based Binary Segmentation using Color

- K-means on FG Strokes



- K-means on BG Strokes



- →Probability  $P_p(FG | I_p)$

$$D_p(FG) = 0 \text{ if } p \in F$$

$$= \infty \text{ if } p \in B$$

$$= -\log P_p(FG | I_p) \text{ otherwise}$$

---

$$D_p(BG) = 0 \text{ if } p \in B$$

$$= \infty \text{ if } p \in F$$

$$= -\log P_p(BG | I_p) \text{ otherwise}$$



# Energy E for Binary Segmentation: Smoothness

$$V_{pq}(BG, BG) = V_{pq}(FG, FG) = 0$$

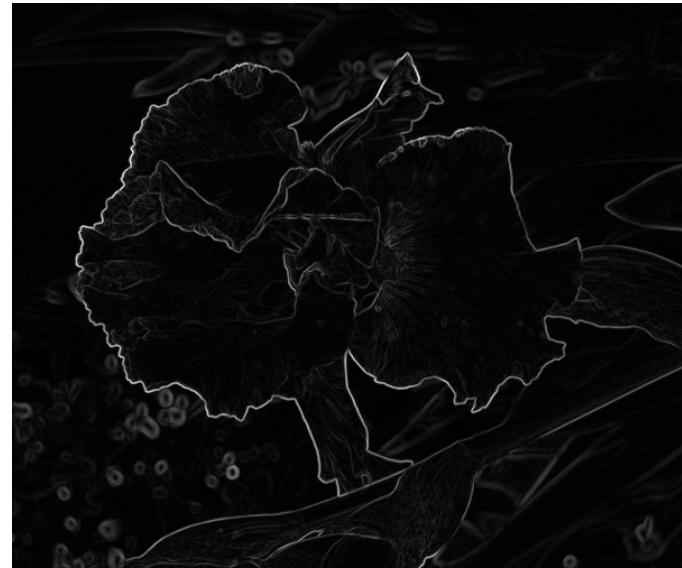
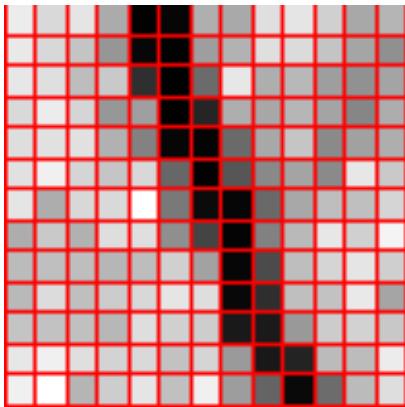
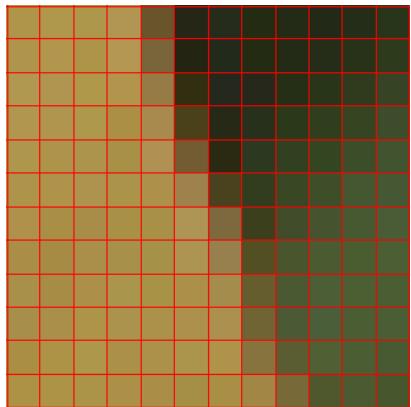
*I*

- Encourage segmentation boundaries to occur at image edges

$$V_{pq}(FG, BG) = \exp\left(-\frac{\|I_p - I_q\|_2^2}{2\sigma^2}\right)$$

$$V_{pq}(BG, FG) = V_{pq}(FG, BG)$$

$$\|\nabla I\|$$



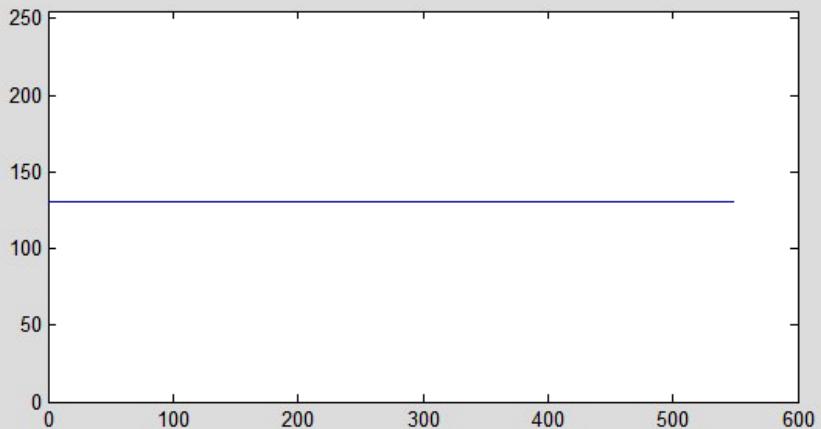
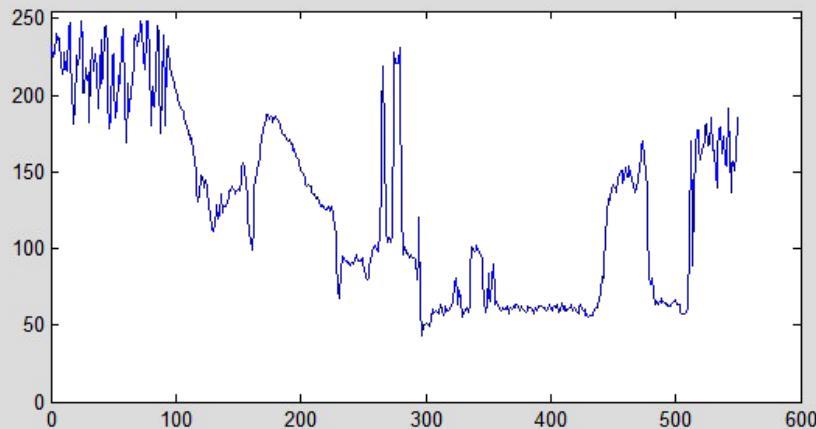
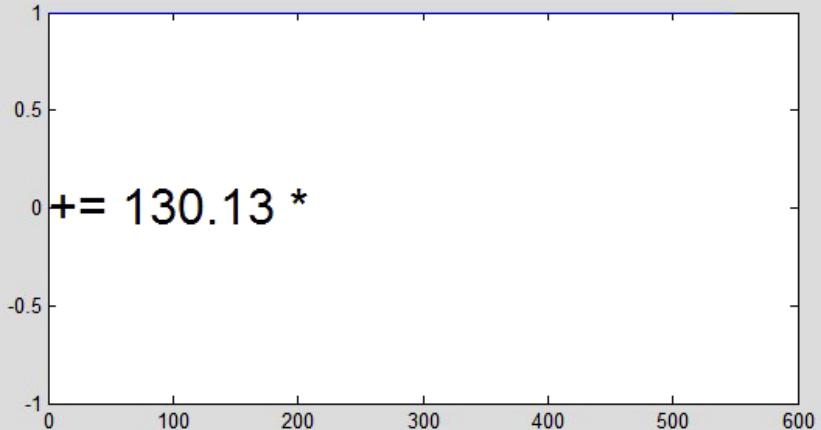
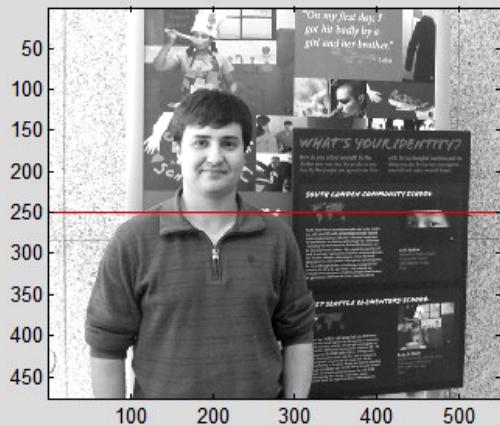
# Why is V called the smoothness term?

$$E(X) = \sum_p D_p(x_p) + \lambda \sum_{p,q \in N} V_{pq}(x_p, x_q)$$

$$P_p(\text{NEG}_{\text{SMO}}^{\text{VAL}})$$

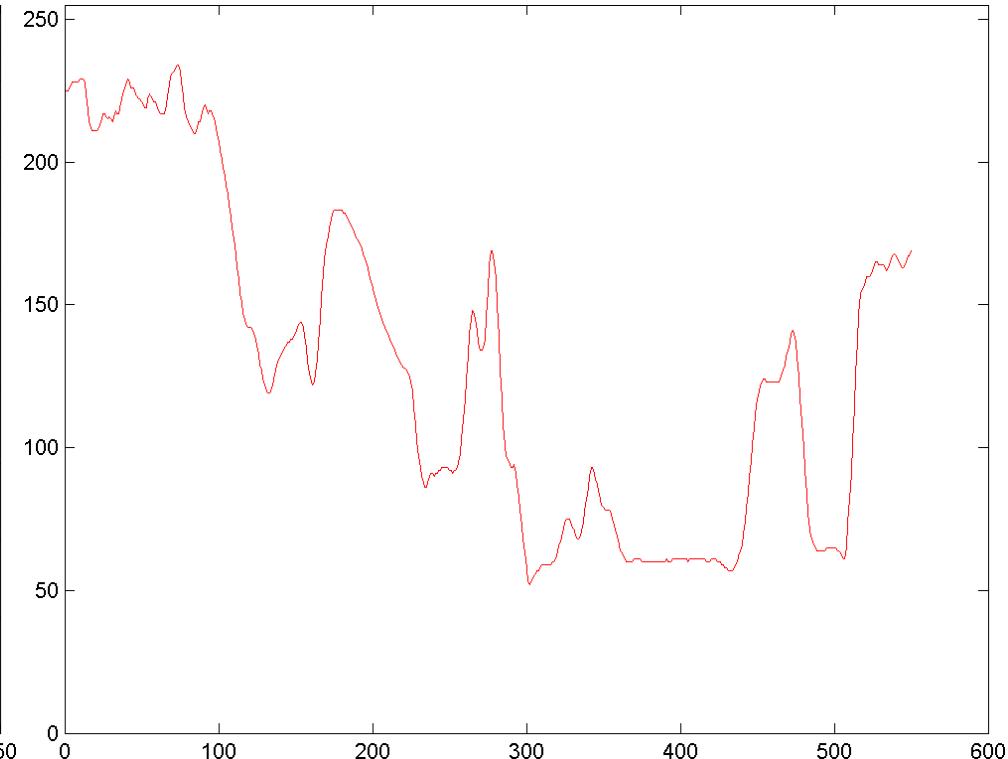
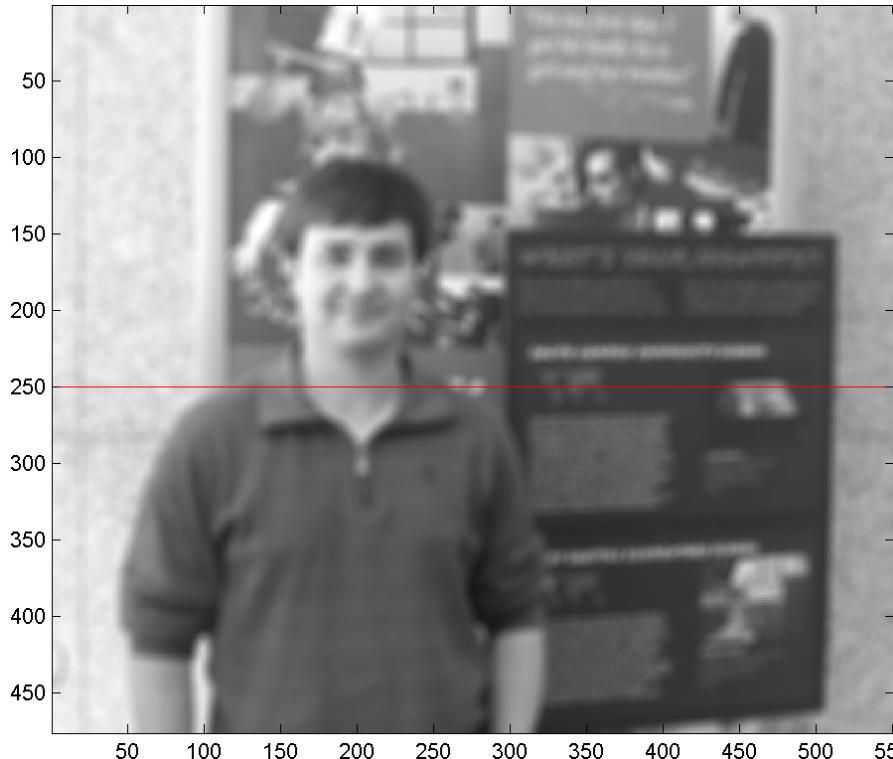


# In-Focus Segmentation: Frequency Decomposition

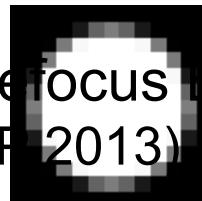


# In-Focus Segmentation: Modeling Defocus Blur

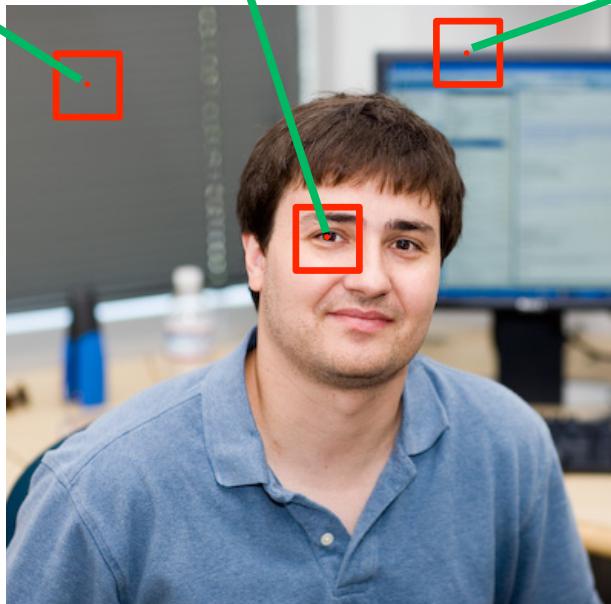
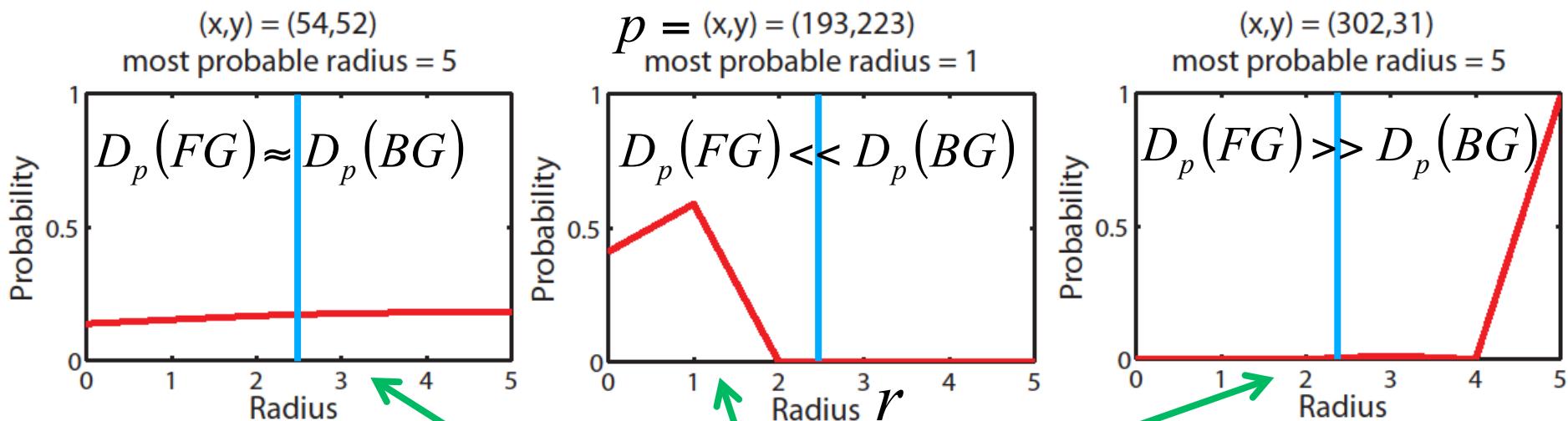
- Larger blurs remove higher frequencies



Estimating Spatially Varying Defocus Blur from A Single Image  
Zhu, Cohen, Schiller, Milanfar (TIP 2013)



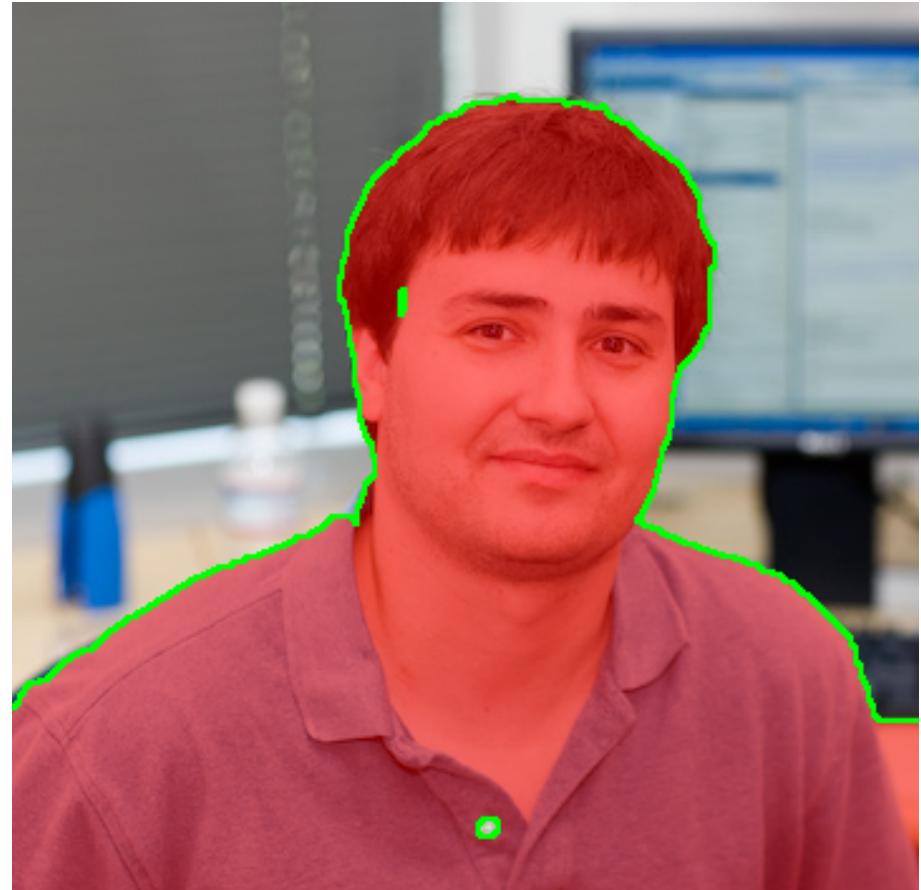
# Out-of-Focus Blur Estimation



- Measure power across frequencies

# Automatic In-Focus Segmentation Results

$$E(X) = \sum_p D_p(x_p) + \lambda \sum_{p,q \in N} V_{pq}(x_p, x_q)$$



# Automatic In-Focus Segmentation Results



# Automatic In-Focus Segmentation Results

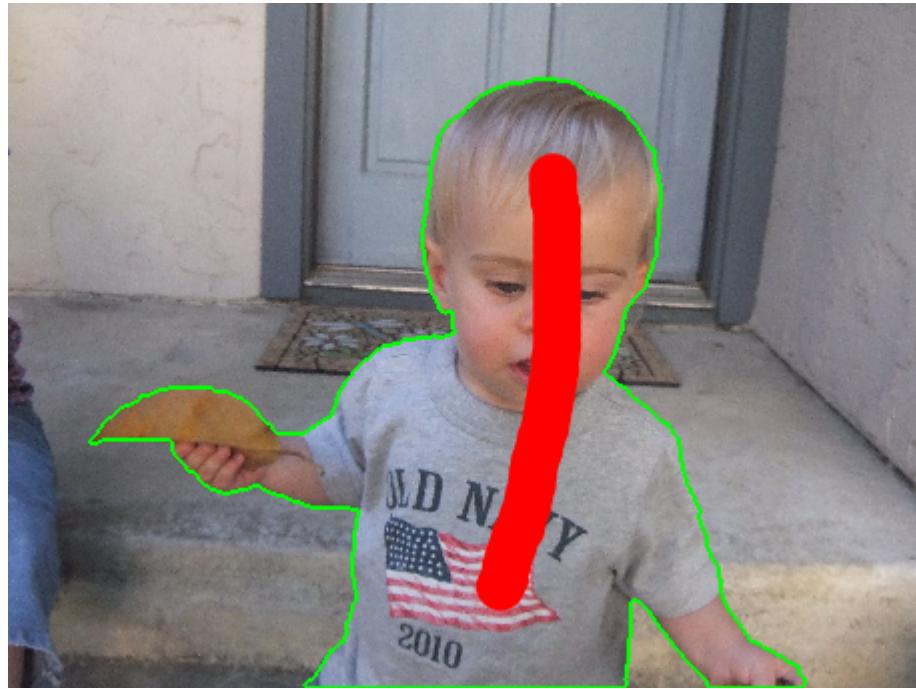
- Works when foreground and background colors are similar



# Automatic In-Focus Segmentation Results



# Interactive Stereo Co-Segmentation



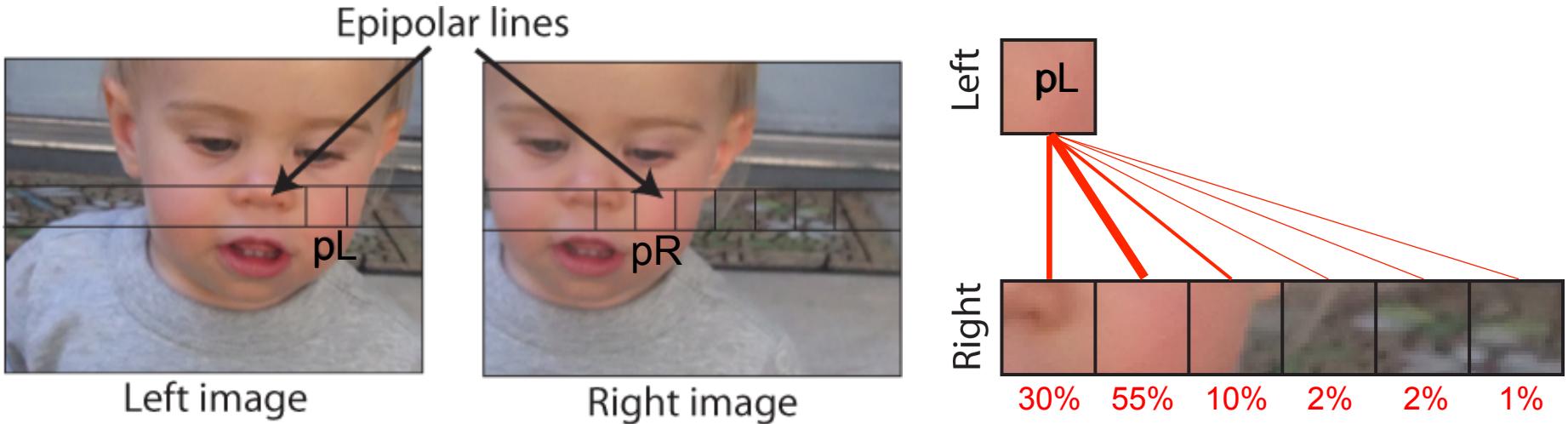
Left



Right

# Energy E for Stereo Segmentation

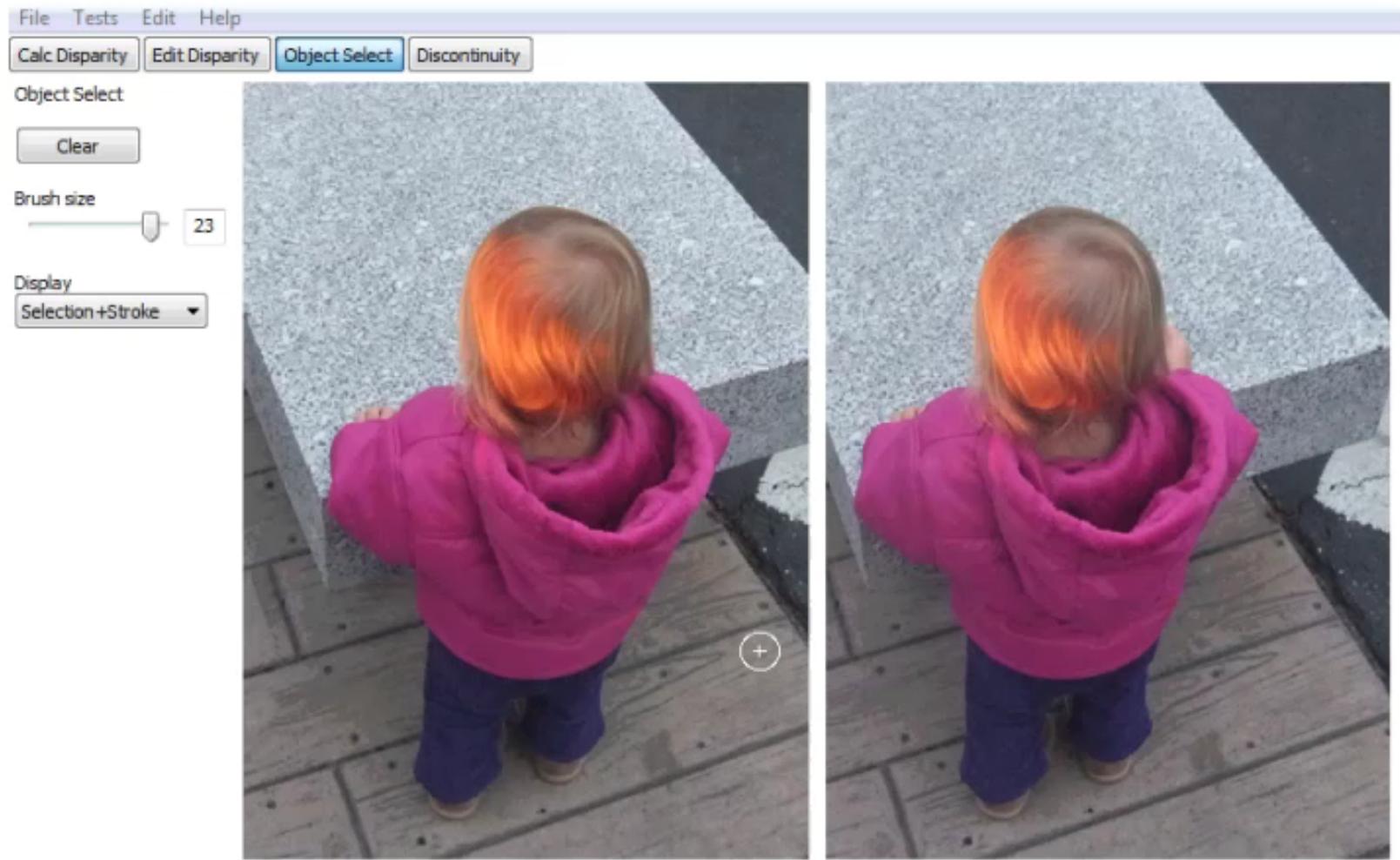
- Add an energy term to include correspondence information



- If  $(p_L, p_R)$  likely correspond, then their labels should be same

$$E(X) = \sum_p D_p(x_p) + \lambda \sum_{p,q \in N} V_{pq}(x_p, x_q) + \mu \sum_{pL, pR} P_{match}(pL, pR) |x_{pL} - x_{pR}|$$

# Stereo Segmentation Results

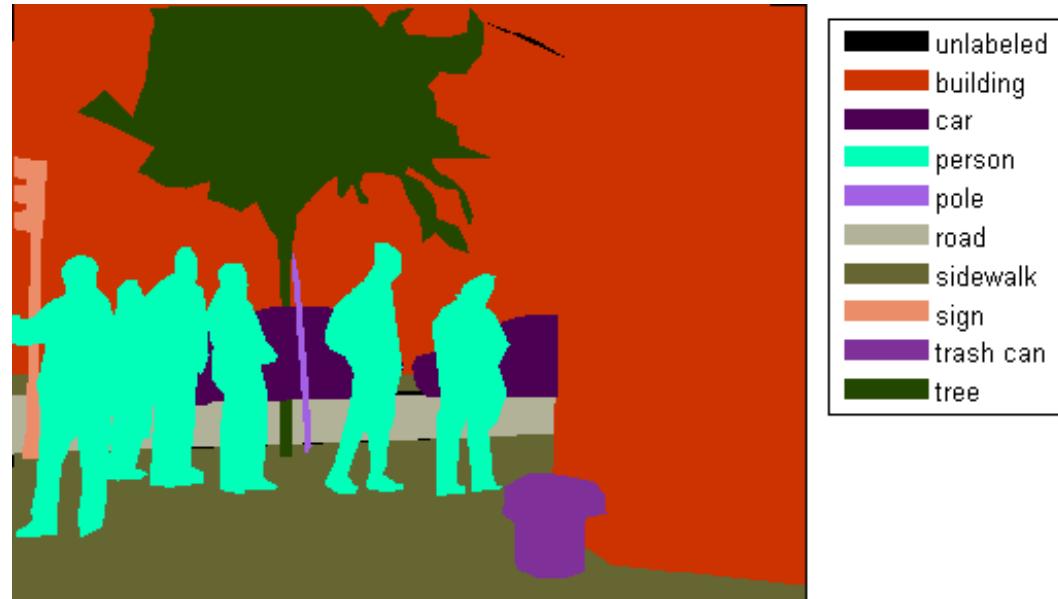


# Scene Parsing | Semantic Segmentation

- Label each pixel in an image with its semantic category



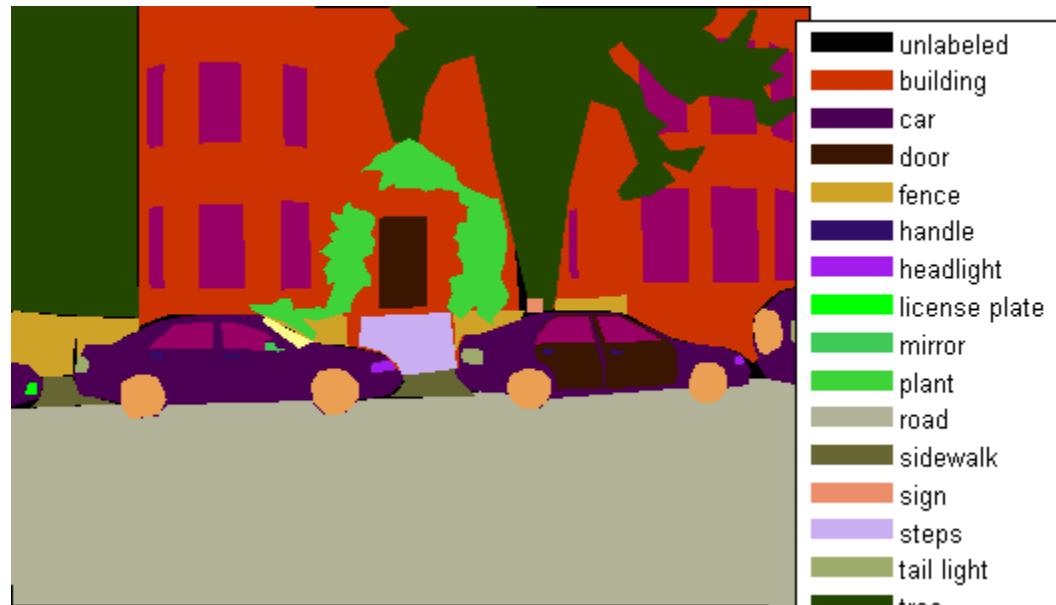
Input Image



Desired Output

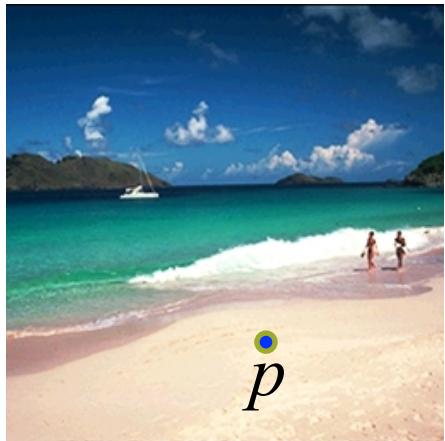
# Scene Parsing | Semantic Segmentation

- Training Data : Labeled Images (Input)
- SIFTflow : 2488 Labeled Training Images, 33 Classes
- LMSun : 45176 Labeled Training Images, 232 Classes



# Scene Parsing | Semantic Segmentation

Input Image



$p$

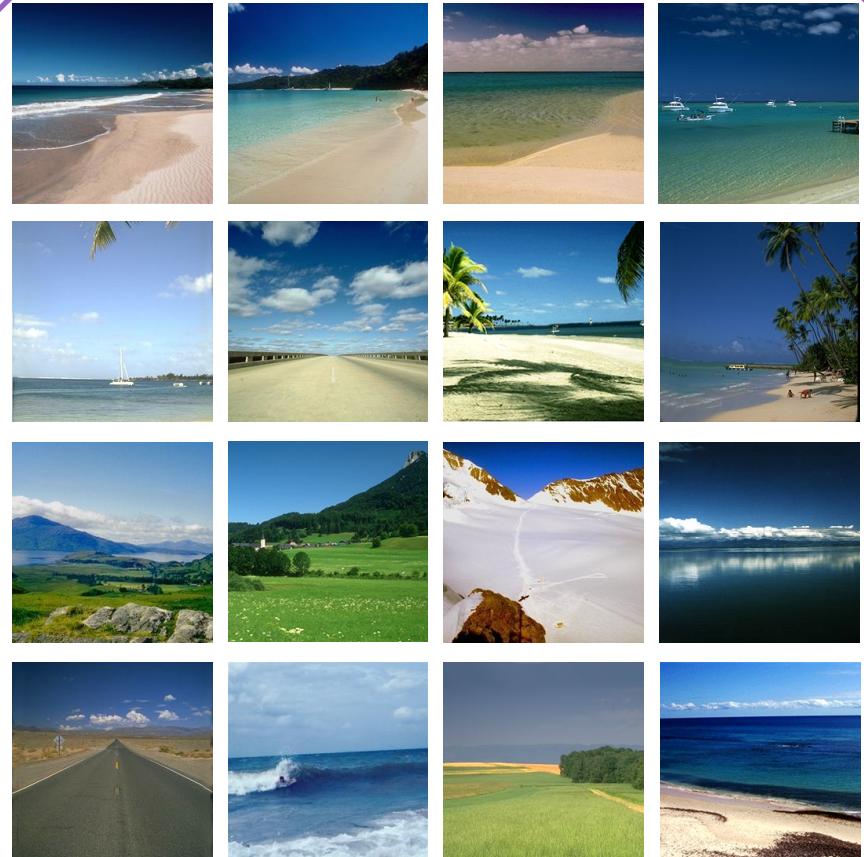
Image Retrieval



Sea, sand, sky, mountain,  
field, tree, rock, plant, road,  
grass, boat, river, person

$$P_p(x_p = \text{sea}), P_p(x_p = \text{sand}), \dots$$

Similar Training Images



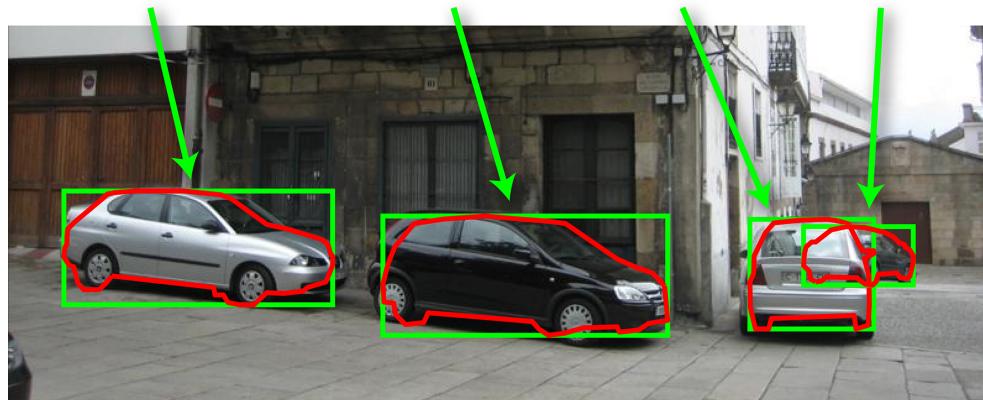
# Scene Parsing | Semantic Segmentation

- Tighe, Lazebnik. Finding things: Image parsing with regions and per-exemplar detectors. CVPR13

Cars from Similar Training Images

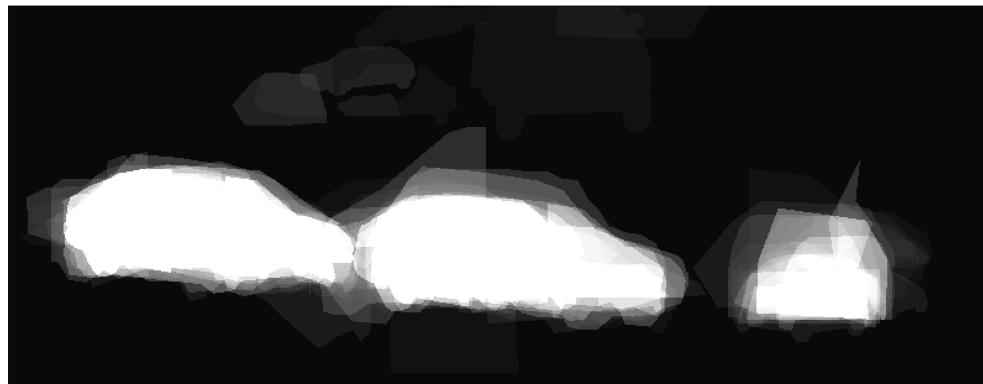


Test Image



Likelihood of Car

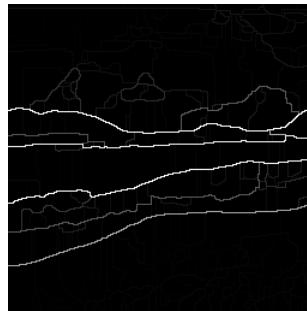
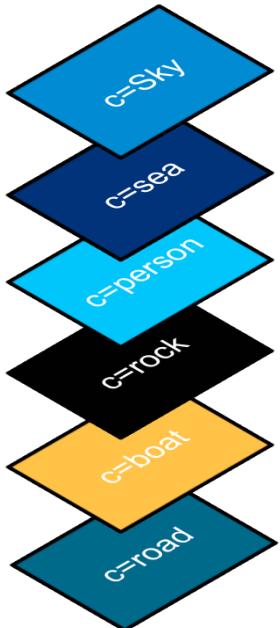
$$P_p(x_p = \text{car})$$



# Energy E for Scene Parsing | Semantic Segmentation

$$E(X) = \sum_p D_p(x_p) + \lambda \sum_{p,q \in N} V_{pq}(x_p, x_q)$$

$$P_p(x_p = c)$$



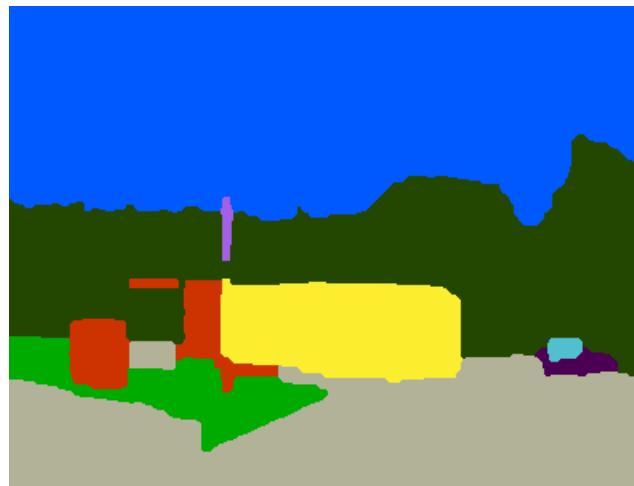
# Scene Parsing | Semantic Segmentation Results



Image



Human Annotation



Tighe CVPR13

# Scene Parsing | Semantic Segmentation Results



Image



Human Annotation

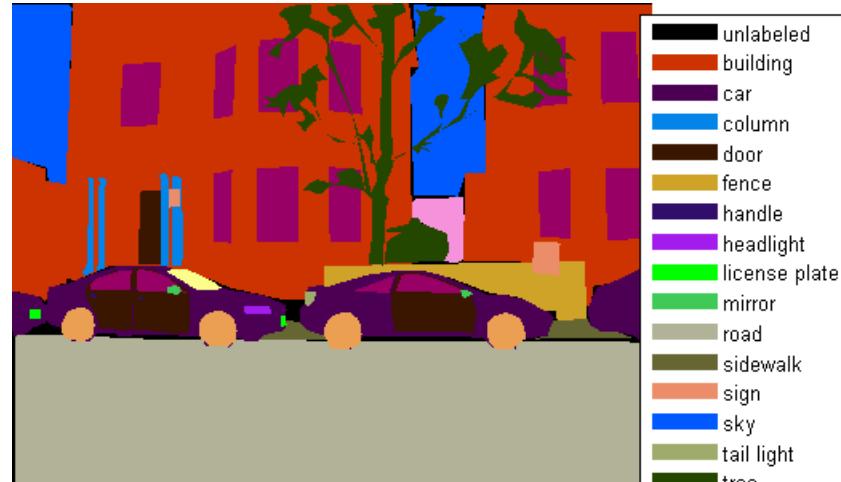


Tighe CVPR13

# Scene Parsing | Semantic Segmentation Results



Image



Human Annotation

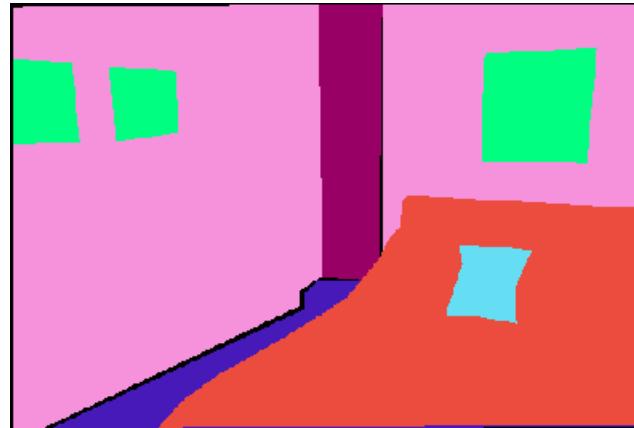


Tighe CVPR13

# Scene Parsing | Semantic Segmentation Results



Image



black	unlabeled
red	bed
purple	floor
green	painting
light blue	pillow
pink	wall
dark purple	window



red	bed
light blue	ceiling
blue	curtain
dark teal	mountain
dark blue	sky
pink	wall
dark purple	window

# Summary and Conclusion

- There are many variations of segmentation problems
  - How many classes? What are the classes?
  - What features are used?
  - Hard or Soft Segmentation?
  - Automatic or Interactive? What User Input?
  - How many images are segmented?
- Semantic Segmentation: still a lot of work to be done

