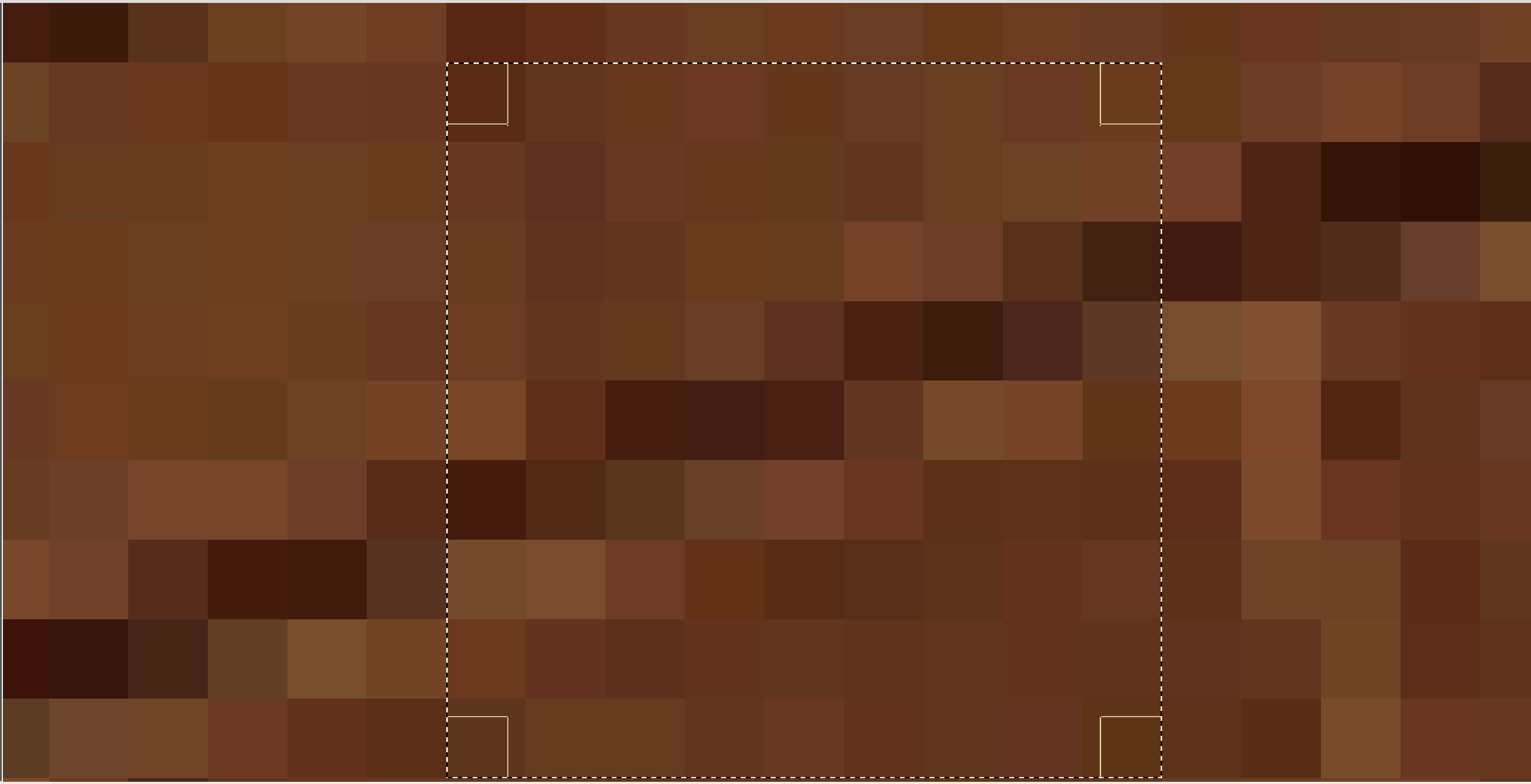
Global Probability of Boundary

Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues – Martin, Fowlkes, Malik

Using Contours to Detect and Localize Junctions in Natural Images – Maire, Arbalaez, Fowlkes, Malik.

presented by

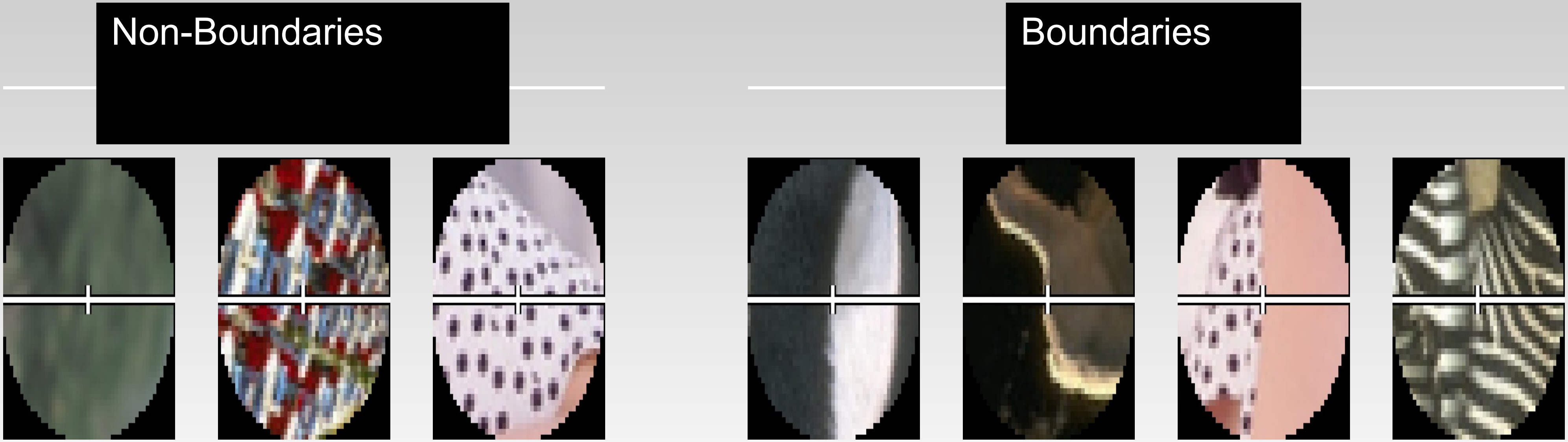
Varun Ramakrishna



Edge Detection Vs Boundary Detection

Edges- Abrupt change in some low-level image feature such as brightness or color

Boundary- Contour in image plane that represents a change in pixel ownership from one object to another



- Estimate Posterior probability of boundary passing though centre point based on local patch-based features
- Using a Supervised Learning based framework
- Boundary information integral to higher level tasks such as perceptual organization

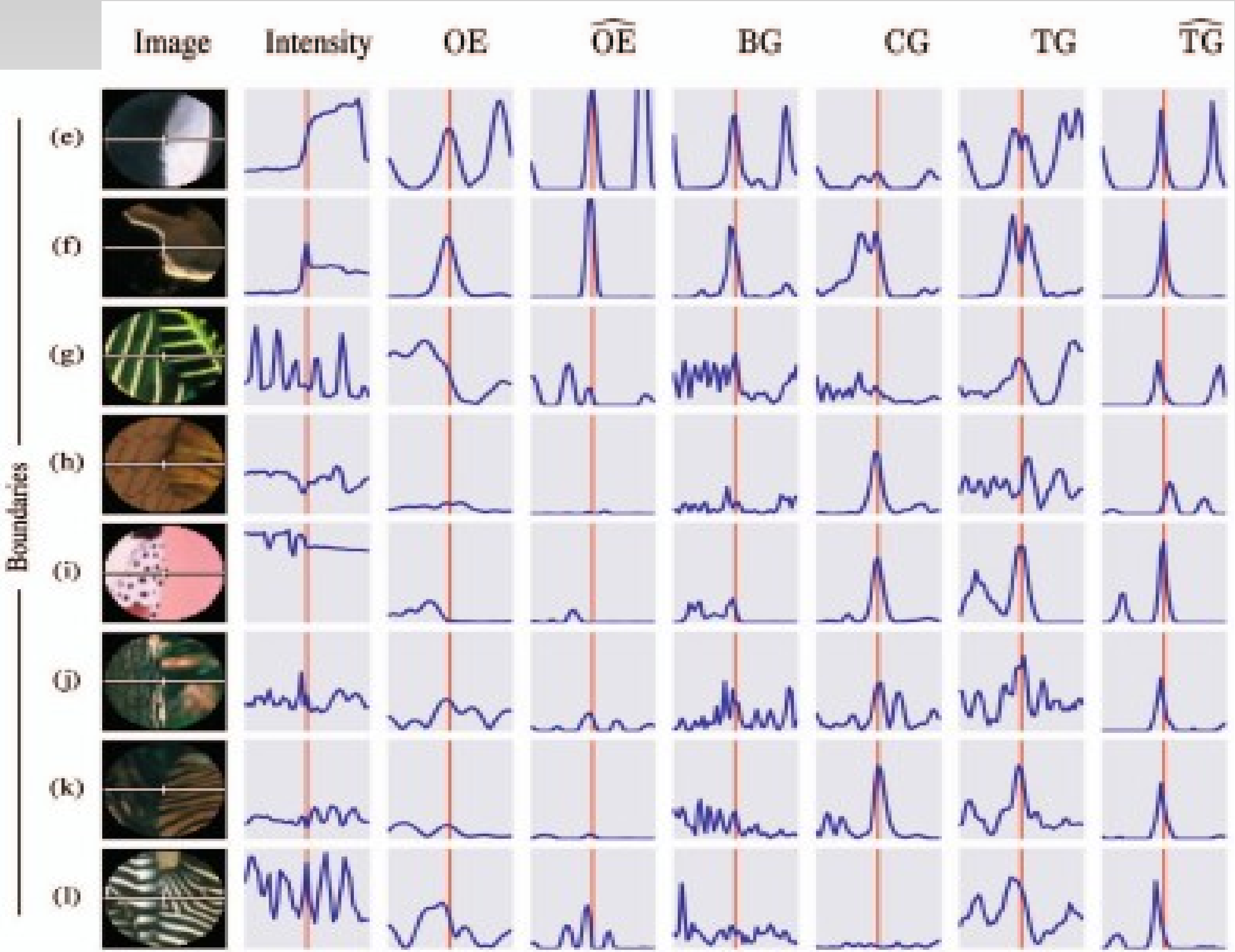


Image
Intensity
OE
$\widehat{OE}$
BG
CG
TG
$\widehat{TG}$
Boundaries
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)

Image Features

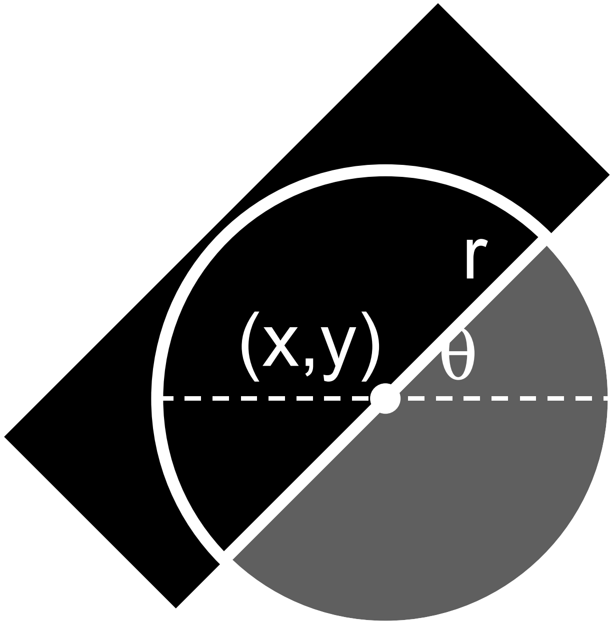
- Oriented Energy

$$\mathrm{OE}_{\theta,\sigma} = (I * f^{e}_{\theta,\sigma})^2 + (I * f^{o}_{\theta,\sigma})^2$$

- Gradient-Based Features

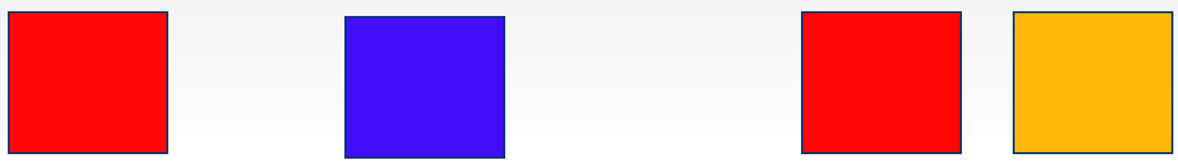
  Compare contents of the

  two disc halves



L*a*b* Colorspace

Which is more similar?



L*a*b* was designed to be uniform in that perceptual “closeness” corresponds to Euclidean distance in the space.

L – lightness (white to black)

a – red-greeness

b – yellowness-blueness

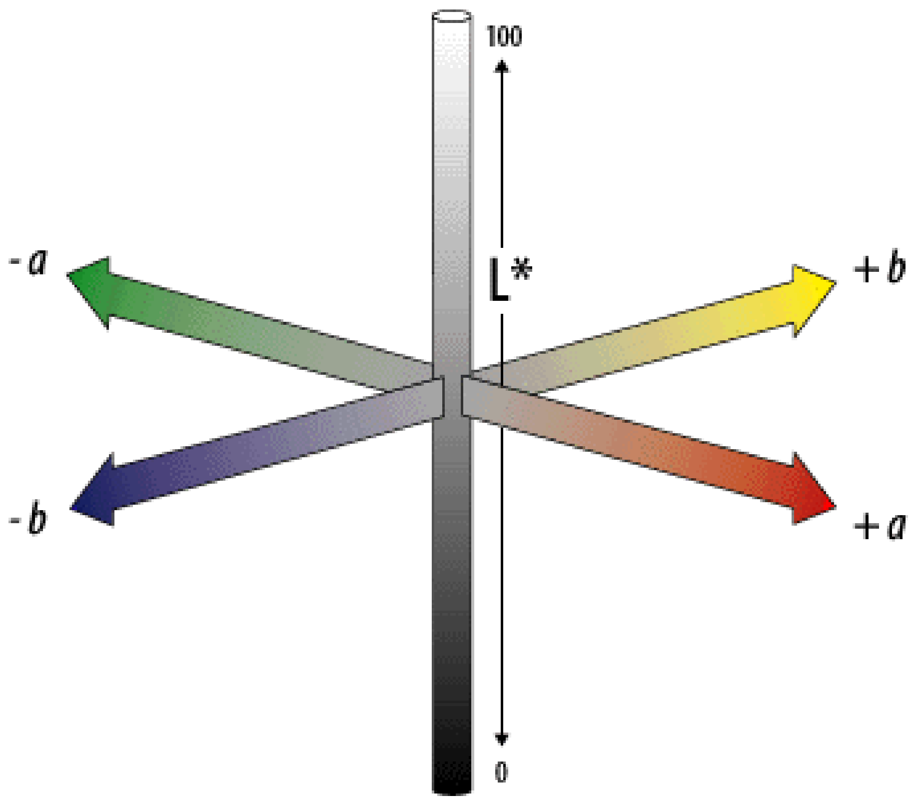


Image Features

- Work in L*a*b* Colorspace – distance between points is perceptually meaningful

Kernel density estimate followed by binning

- Brightness Gradient: Histogram of L* values
- Color Gradient : Histogram of a* b* values

Image Features

Comparison of Histograms

- L1 Norm

- Earth Mover's Distance

- **Chi-Squared Distance**

$$\chi^2(g,h) = \frac{1}{2}\sum\frac{(g_i - h_i)^2}{g_i + h_i}$$

Image Features

Texture Gradient

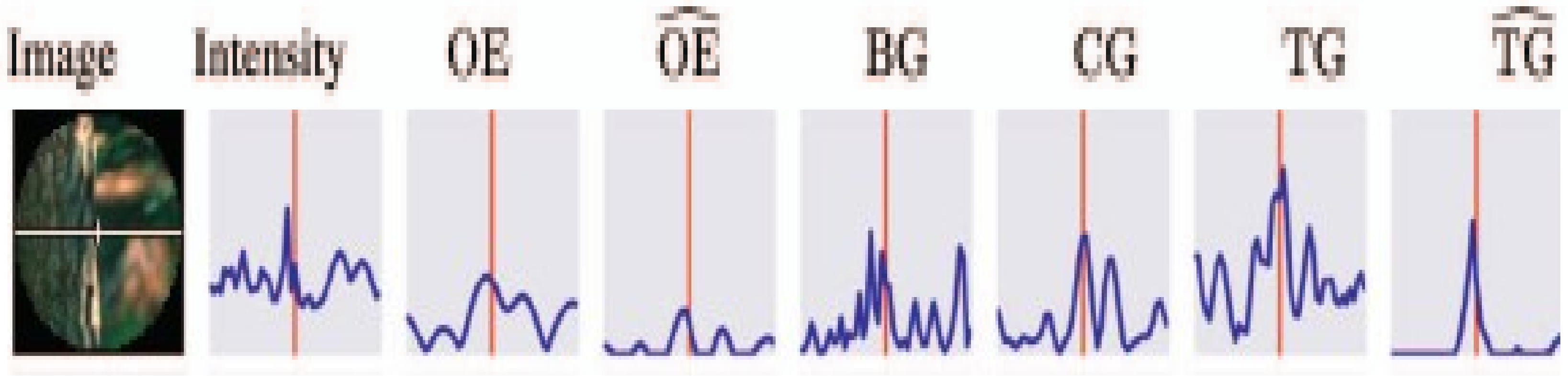
- 13 filter responses at each pixel
- Vector quantization using K-means
- Cluster centres define textons

- Chi-squared difference between texton distributions

Localization

- Underlying function should peak at human marked boundaries

- Spatially extended features
    - On and Off boundary pixels will have a high value



Localization

- Improve Localization by using derived feature
- Divide by distance to nearest maximum

$$\hat{f}(x) = \tilde{f}(x) \cdot \left(\frac{-f''(x)}{|f'(x)| + \epsilon} \right)$$

x → maxima, d(x) → 0, fnew → large

Image Features

- Brightness Gradient $BG(x,y,r,\theta)$
- Color Gradient $CG(x,y,r,\theta)$
- Texture Gradient $TG(x,y,r,\theta)$

Final set of features

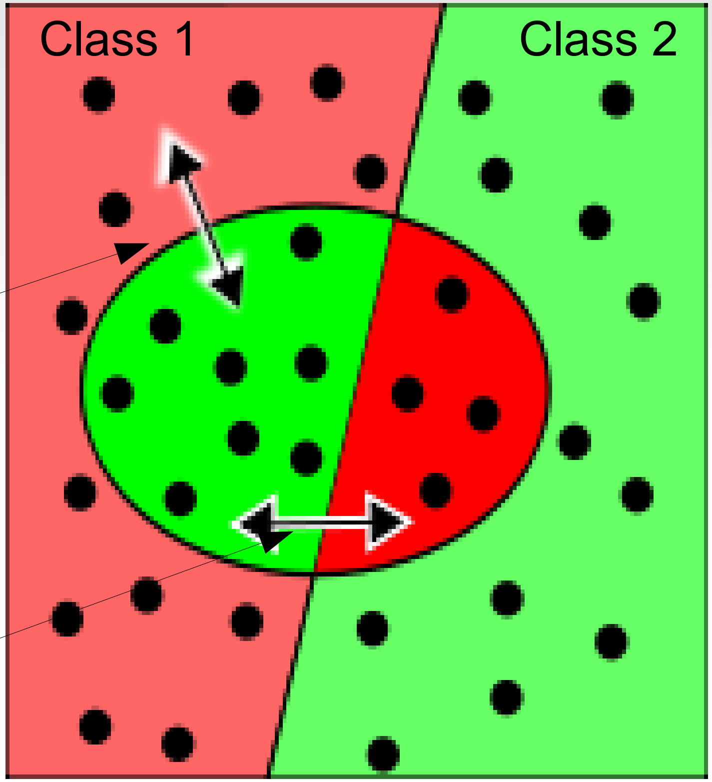
$\{OE', BG, CG, TG'\}$

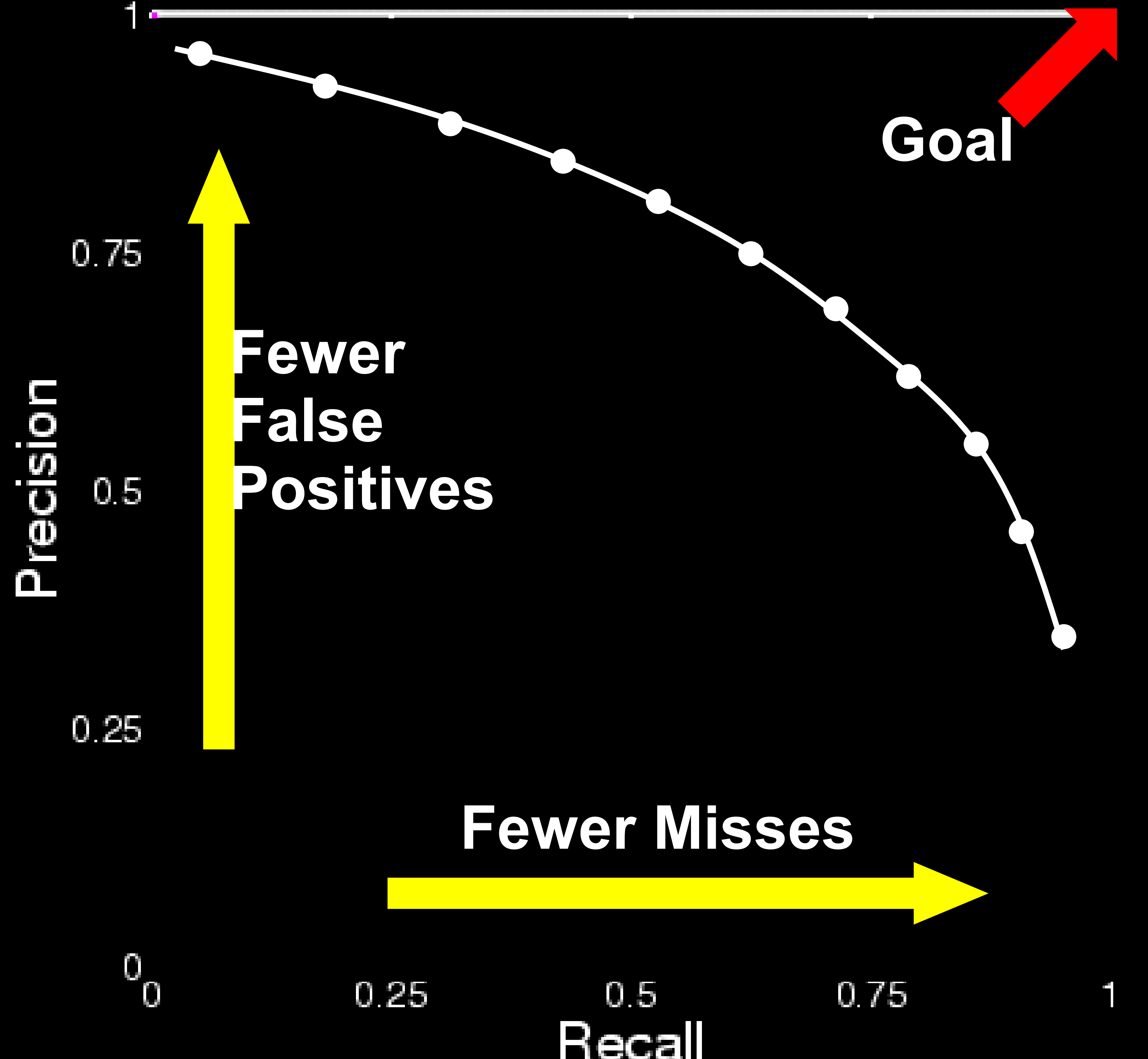
Precision-Recall vs ROC

- Framework to estimate quality of the boundary classifier
- Precision: True Positives / Hypothesized Class Total
- Recall: True Positives / True Class Total

Recall

Precision



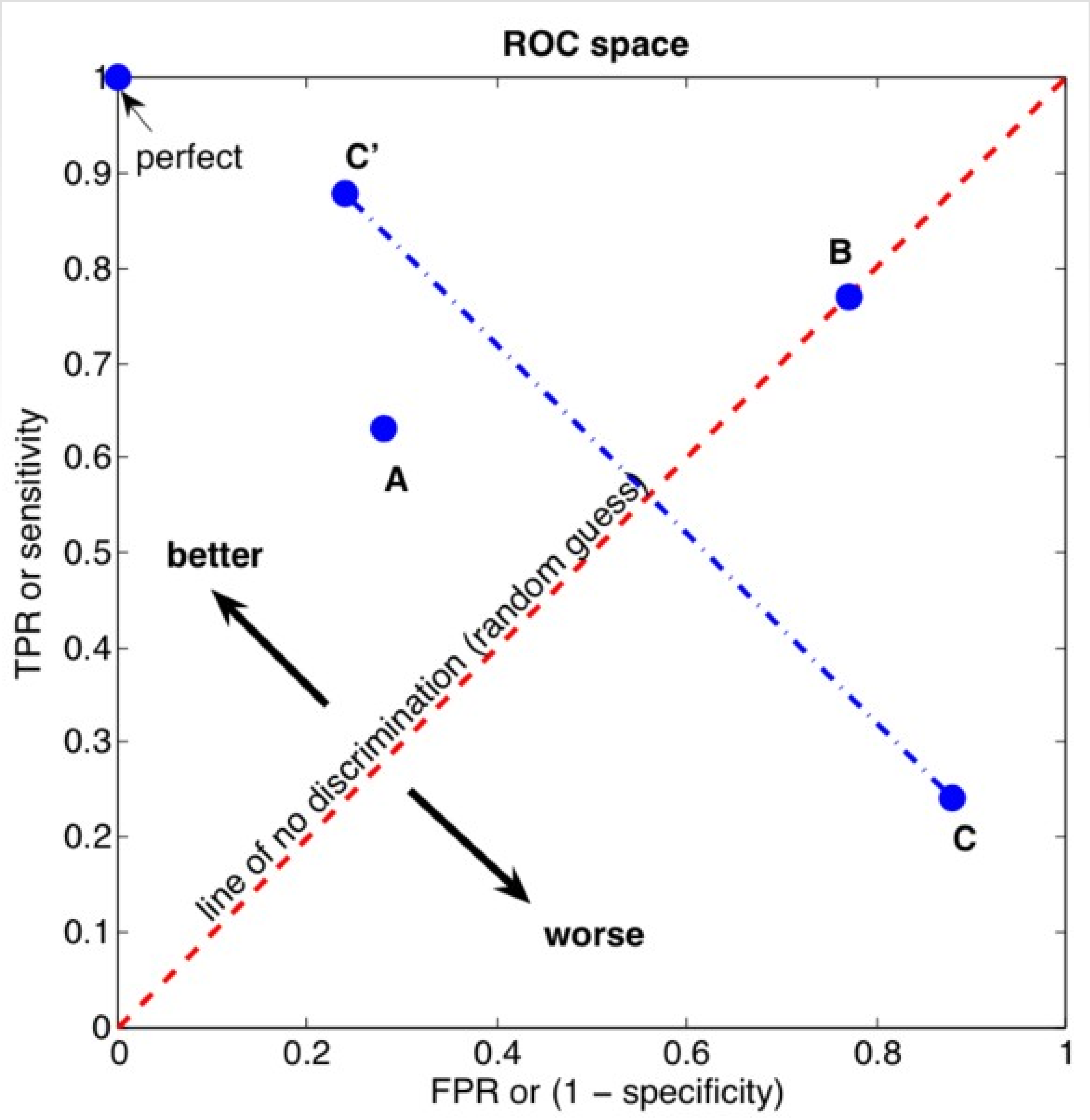


1
0.75
0.5
0.25
0
Precision
0
0.25
0.5
0.75
1
Recall
Goal
Fewer False Positives
Fewer Misses

F-measure

- Harmonic mean of P and R
- Maximum value of F along the curve
- Quality Measure of the P_R curve

ROC Curves



ROC space
perfect
C'
B
A
C
better
worse
line of no discrimination (random guess)
TPR or sensitivity
FPR or (1 – specificity)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1

- **ROC** : TPR/FPR
- **PR** : Precision/Recall
- **TPR=Recall**= TP/(TP+FN)

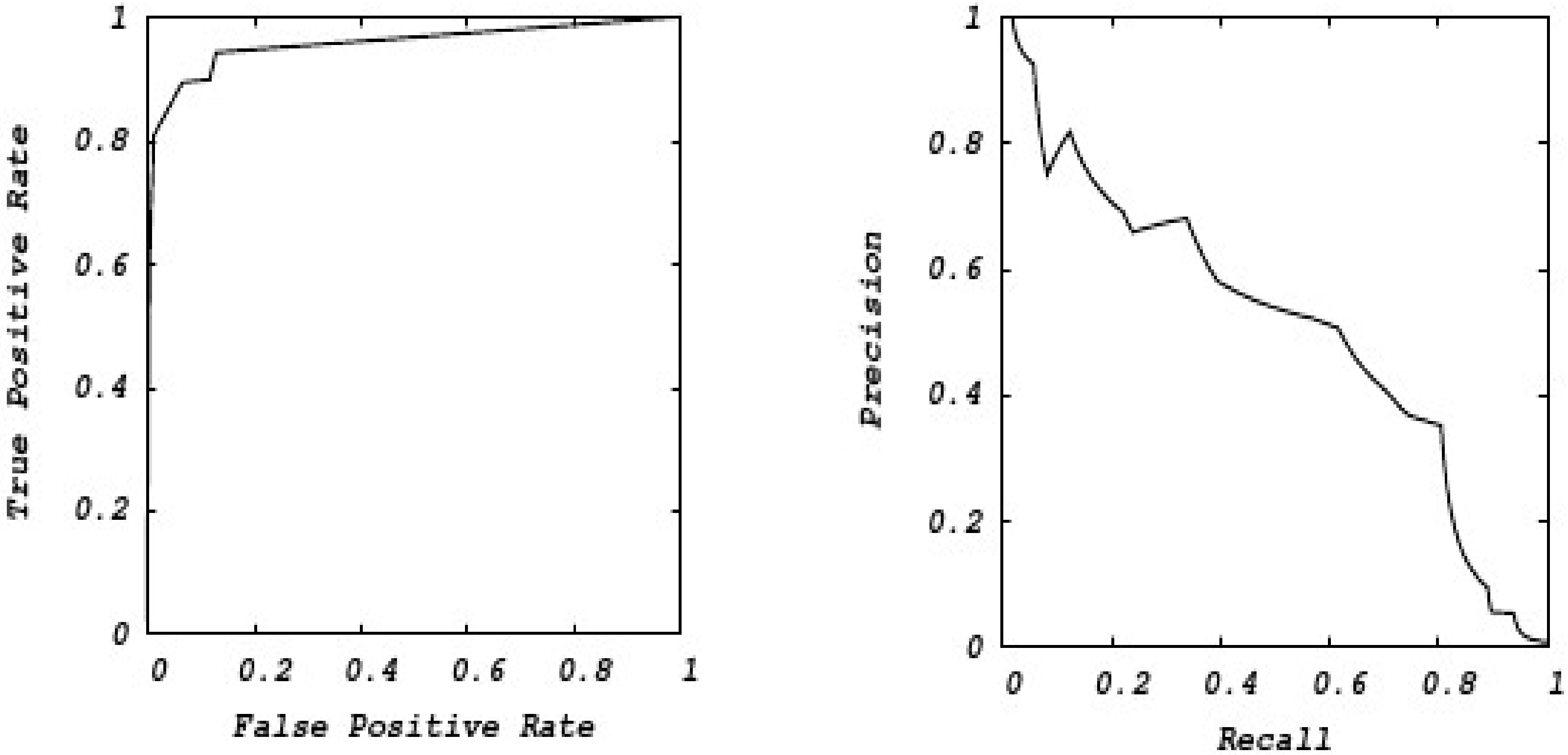
“ total positives”

- **FPR**= FP/(TN+FP)

“ total negatives”

- **Precision**= TP/(TP+FP)

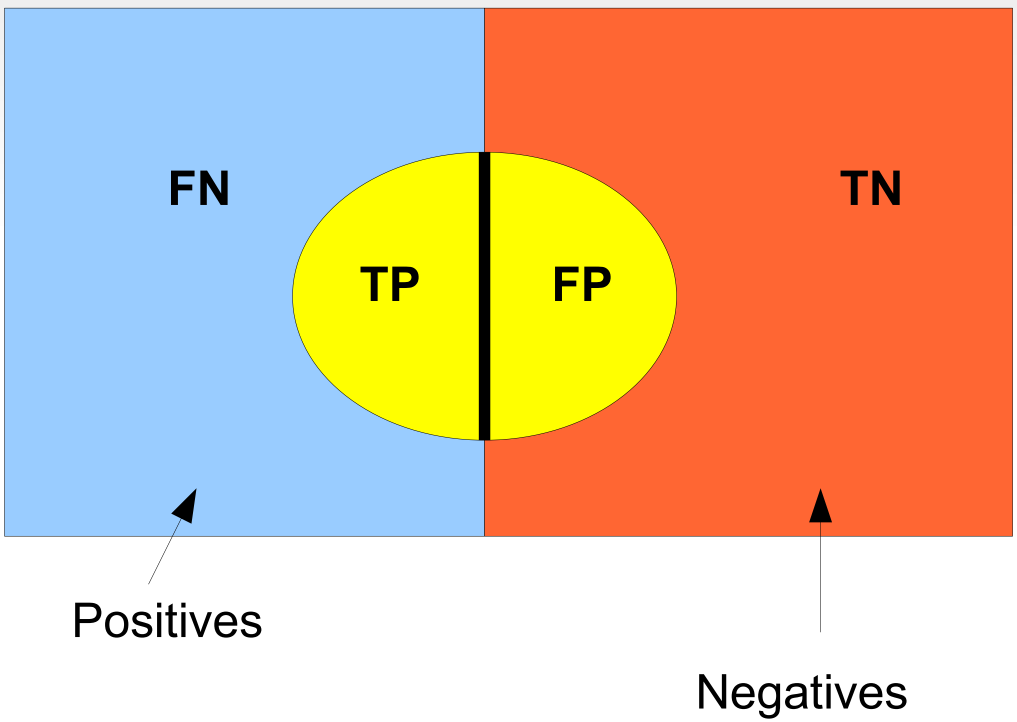
“predicted positives”



(a) Sample ROC curve

(b) Sample PR curve

Figure 1. The same curve shown in both ROC and PR space



Precision = TP/(TP+FP)

Recall = TP/(TP+FN)

TPR = TP/(TP+FN)= Recall

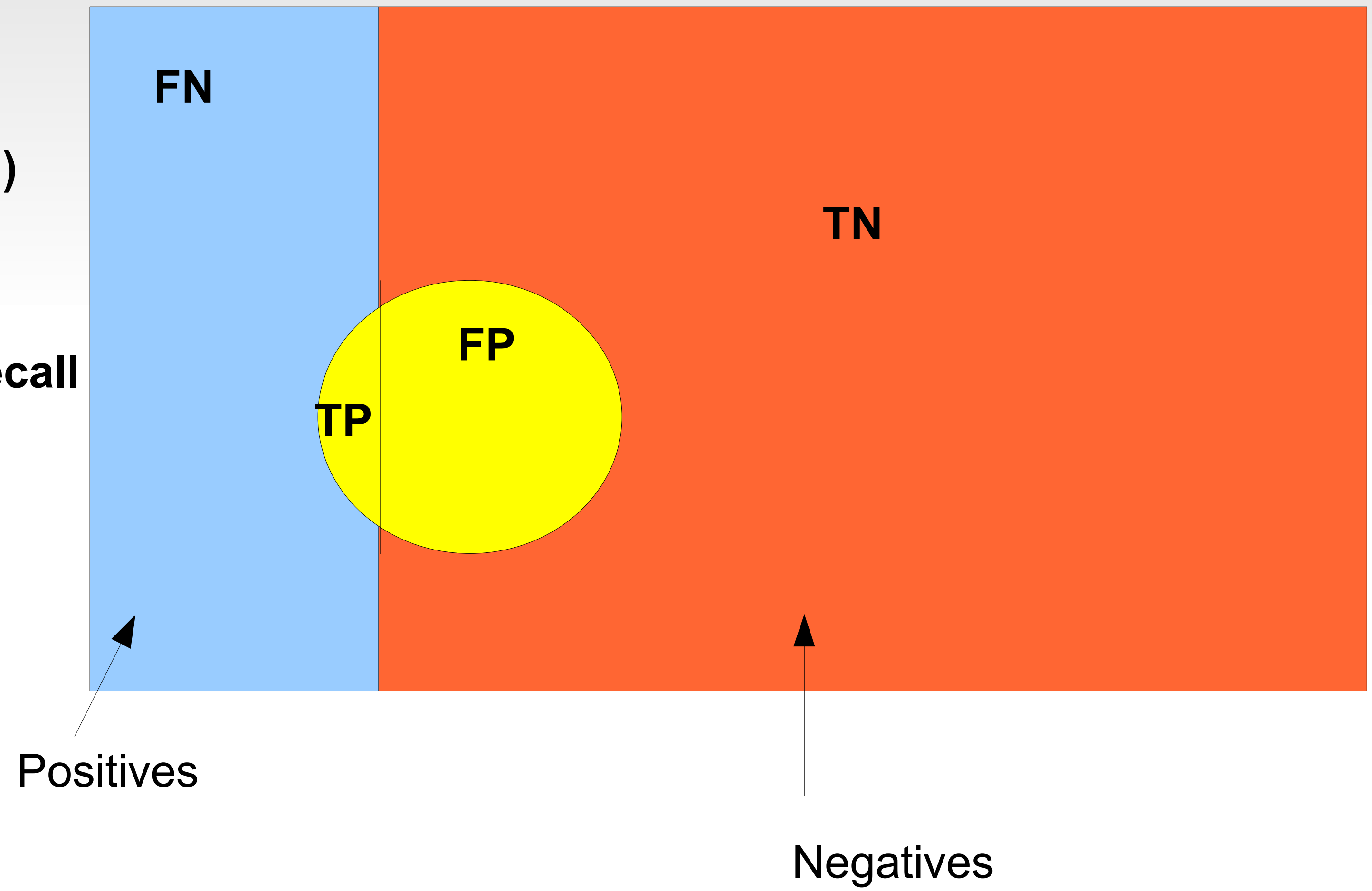
FPR = FP/(FP+TN)

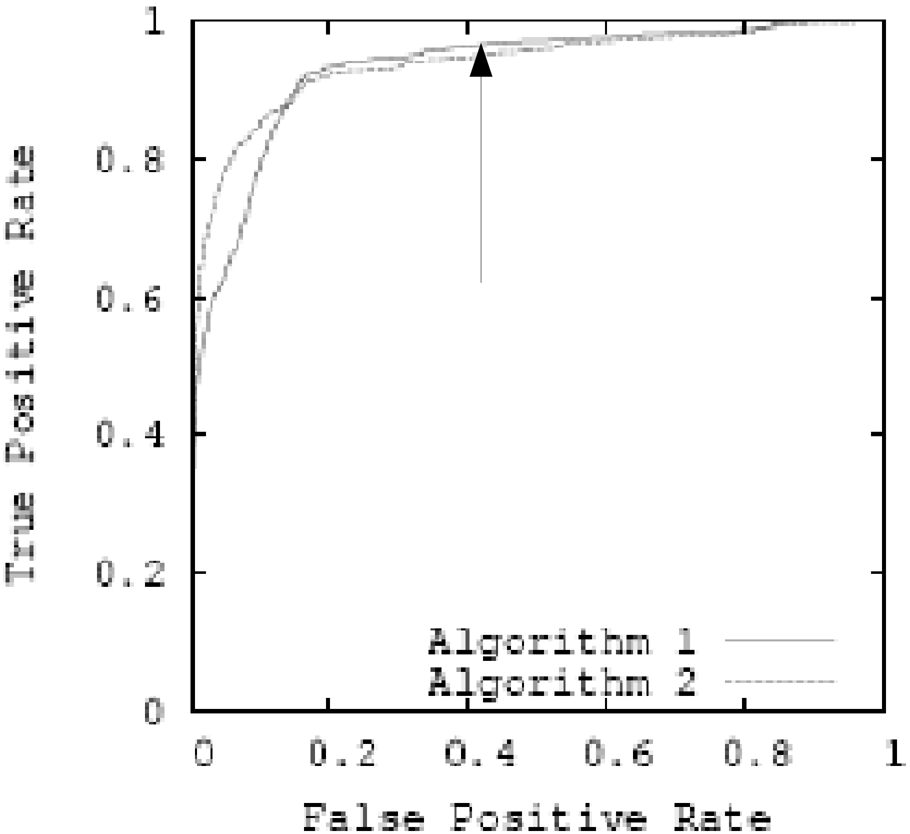
Precision = TP/(TP+FP)

Recall = TP/(TP+FN)

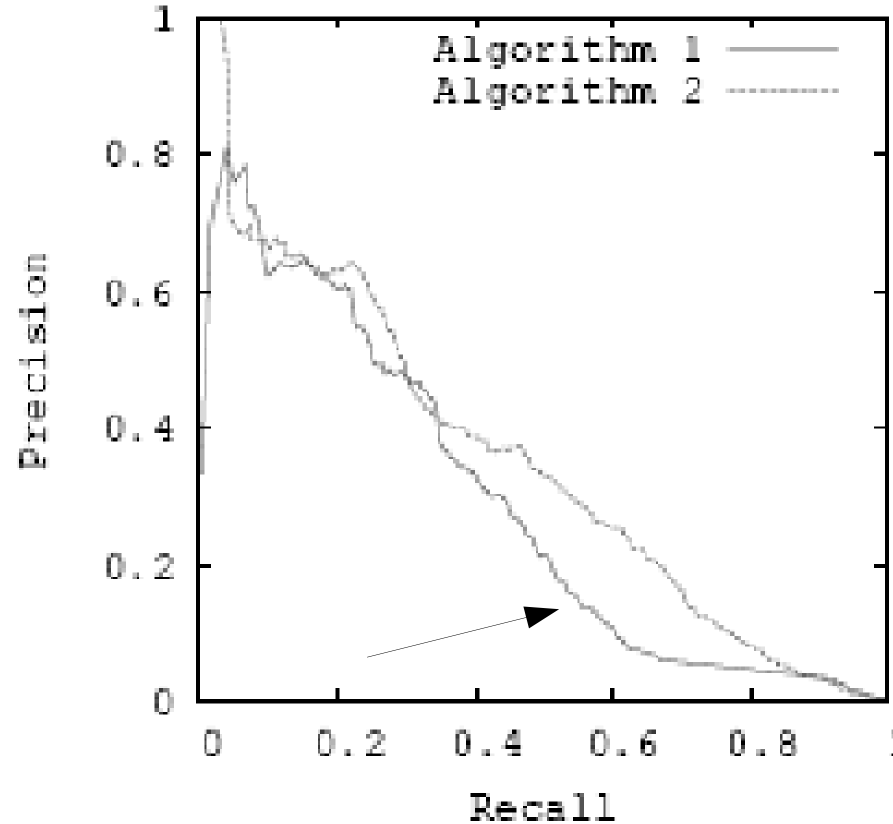
TPR = TP/(TP+FN)= Recall

FPR = FP/(FP+TN)





(a) Comparison in ROC space



(b) Comparison in PR space

ROC Curves

X-axis Fraction of false positives (fallout)

Y-axis Fraction of true positives (hit rate)

But true negatives grow as n^2, while true positives grow as n.

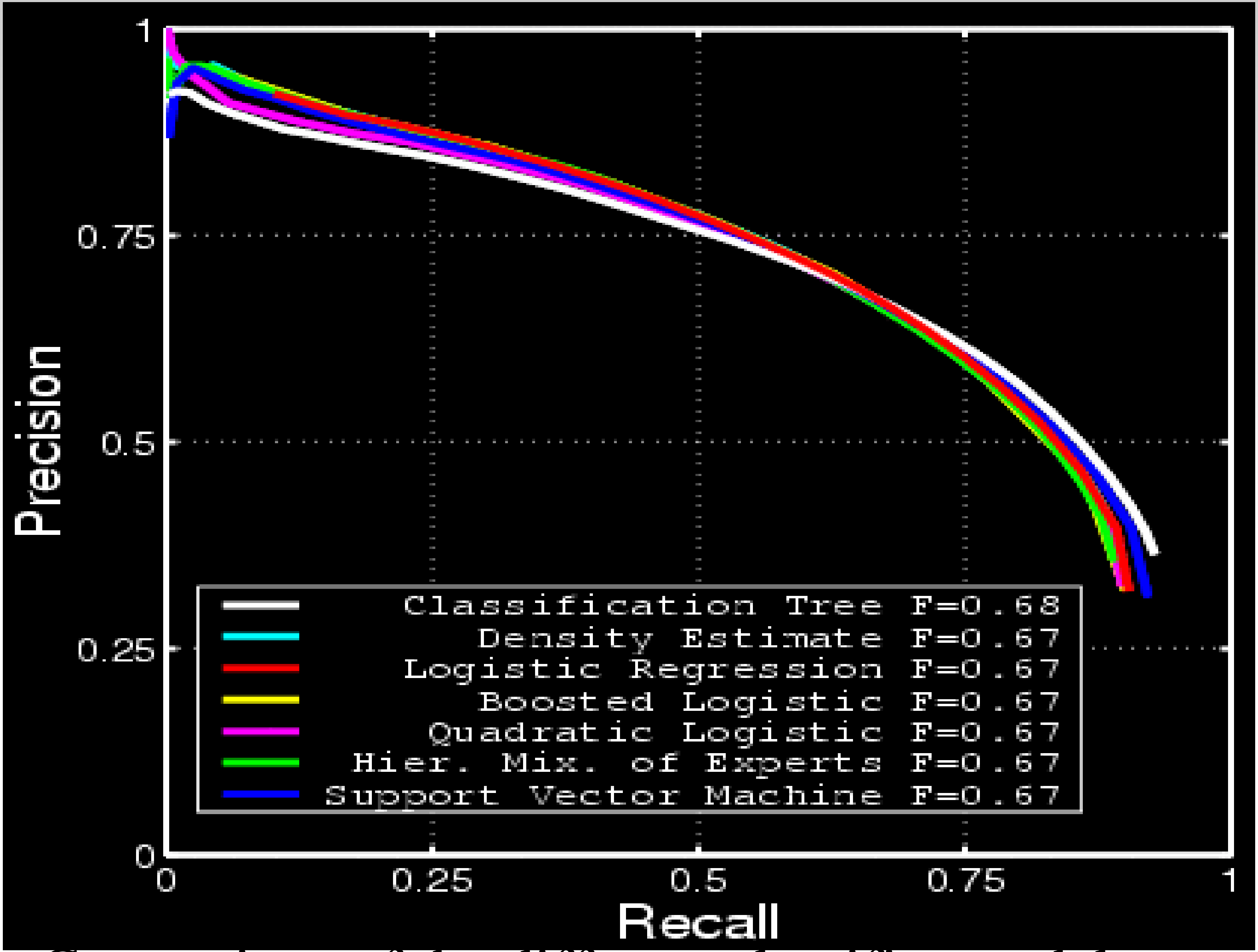
Fallout declines as 1/n, for a scaling of n of the image

Cue-Combination

- Classification Trees
  - Top-down splits to maximize entropy, error bounded
- Density Estimation
  - Adaptive bins using k-means
- Logistic Regression, 3 variants
  - Linear and quadratic terms
  - Confidence-rated generalization of AdaBoost (Schapire&Singer)
- Hierarchical Mixtures of Experts (Jordan&Jacobs)
  - Up to 8 experts, initialized top-down, fit with EM
- Support Vector Machines (`libsvm`, Chang&Lin)
  - Gaussian kernel, ν -parameterization

Range over bias, complexity, parametric/non-parametric

Training on 200 images from the BSDS

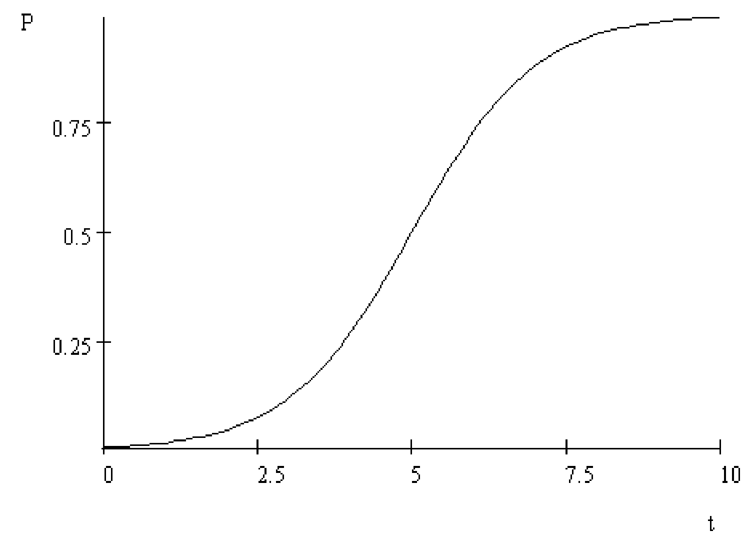


- Comparison of the different classifier models

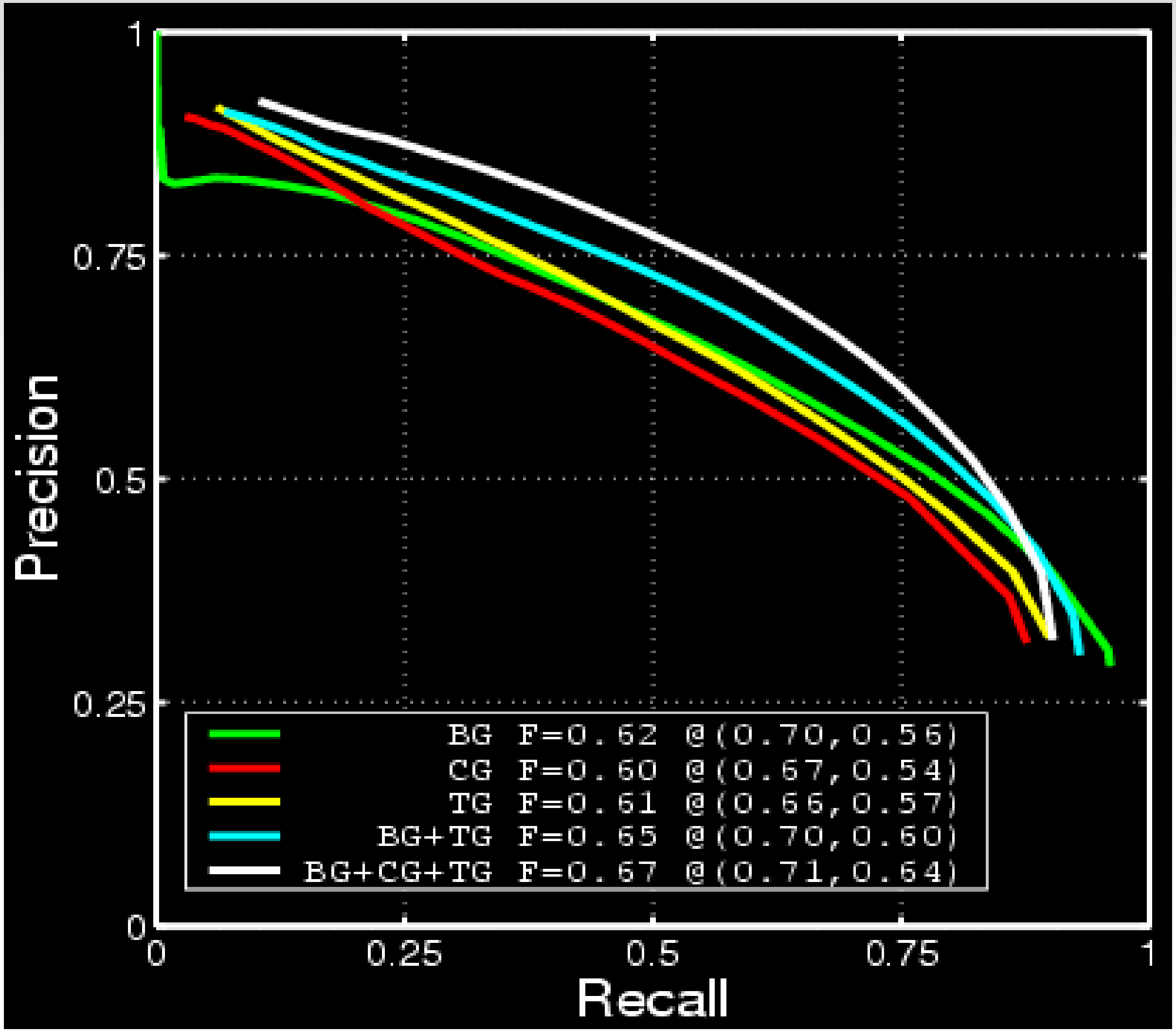
- Simple logistic regression model performs as well as more complex models
- Linear model supported by psychophysics (simple neuron model)

$$f(z) = \frac{1}{1 + e^{-z}}$$

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k,$$



Cue-Combinations



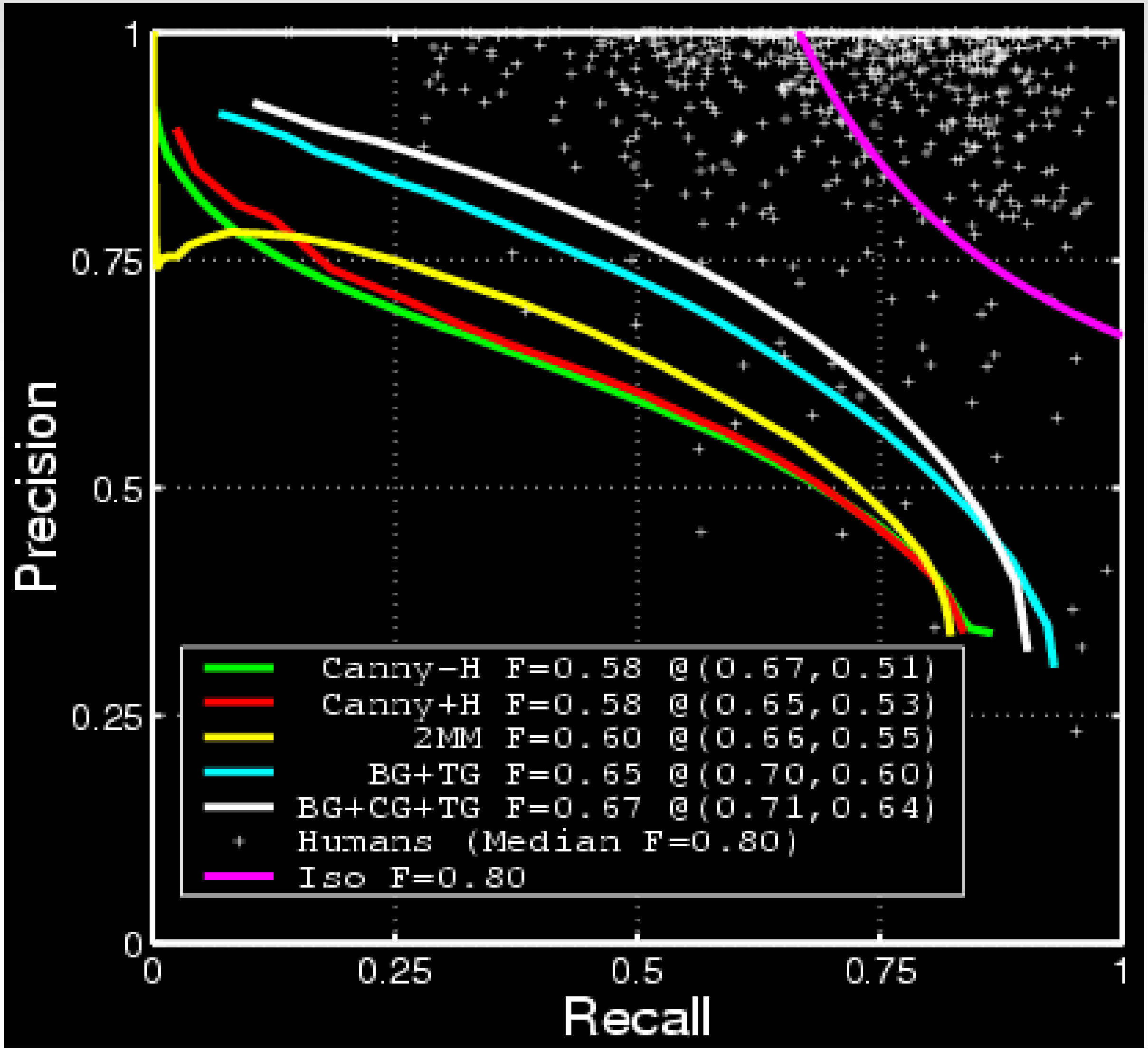
- Texture gradients are an important cue!

Berkeley Segmentation Dataset

- Human subjects presented with image
- Divide into a number of segments which represent ”things” or parts of ”things”
- 2-30 is a good number
- Segments should be approximately equally important
- 200 images for training, 100 images for testing

Image	Canny	2MM	Pb	Human

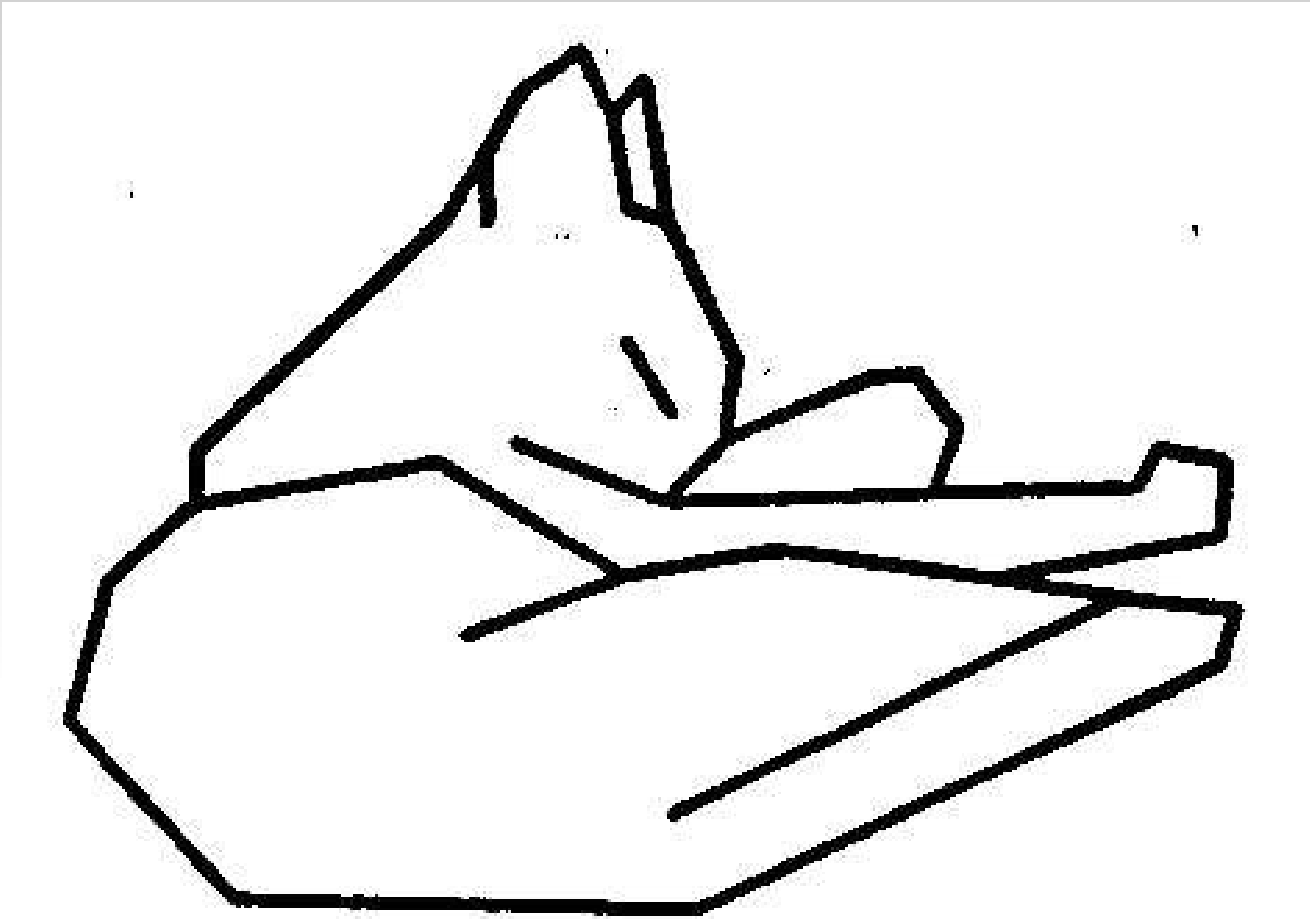
- **Testing on 100 images from the BSDS**



Key Results

- Simple model works well
- Texture gradient is important

.... Now combine the local cues with global cues....



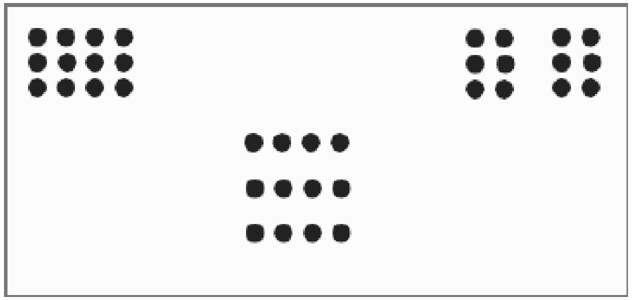
- Line drawings convey most information
- Goal of boundary detection → line drawings that would help in perceptual organization and hence object recognition

Perceptual Organization

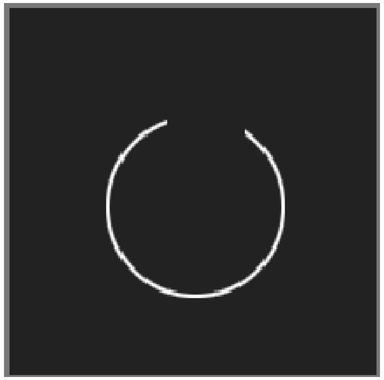
- Gestaltist view of Perceptual Organization
- The whole is different from the sum of the individual parts
- Integration of local cues as computed previously with global cues
- Global Framework : Mechanism for integration of local cues – Normalized Cuts

Perceptual Organization

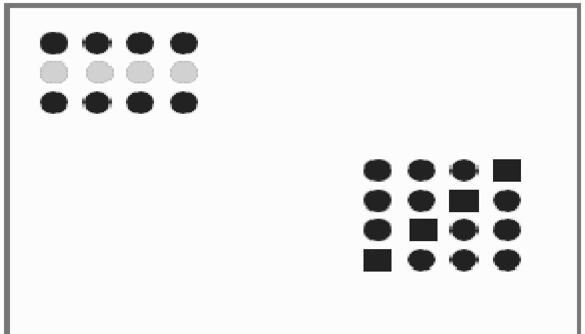
Proximity: Objects that are closer to one another tend to be grouped together.



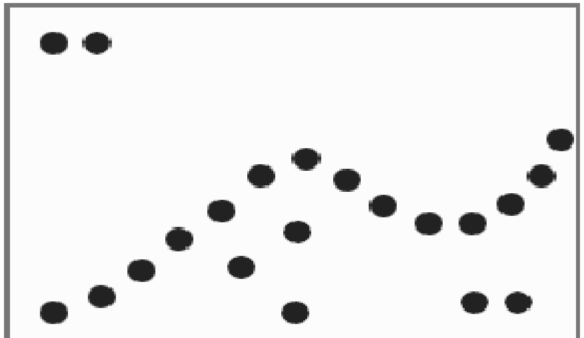
Closure: Humans tend to enclose a space by completing a contour and ignoring gaps.



Similarity: Elements that look similar will be perceived as part of the same form. (color, shape, texture, and motion).



Continuation: Humans tend to continue contours whenever the elements of the pattern establish an implied direction.



Normalized-Cuts review

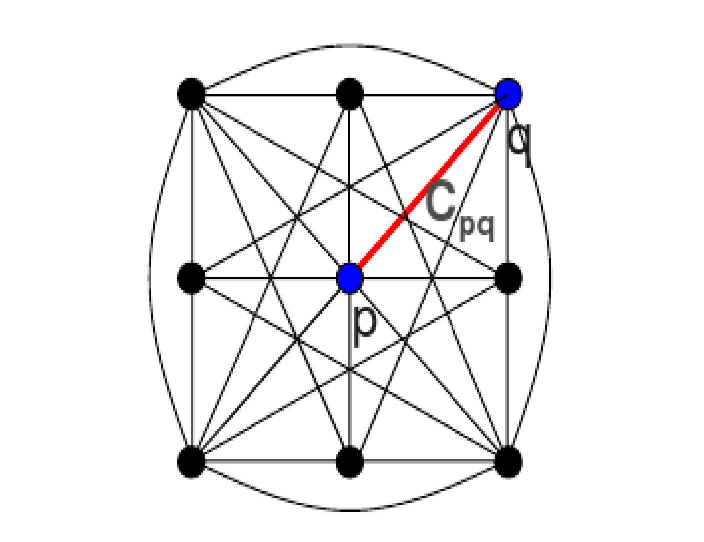


Image is modelled as a

fully connected graph

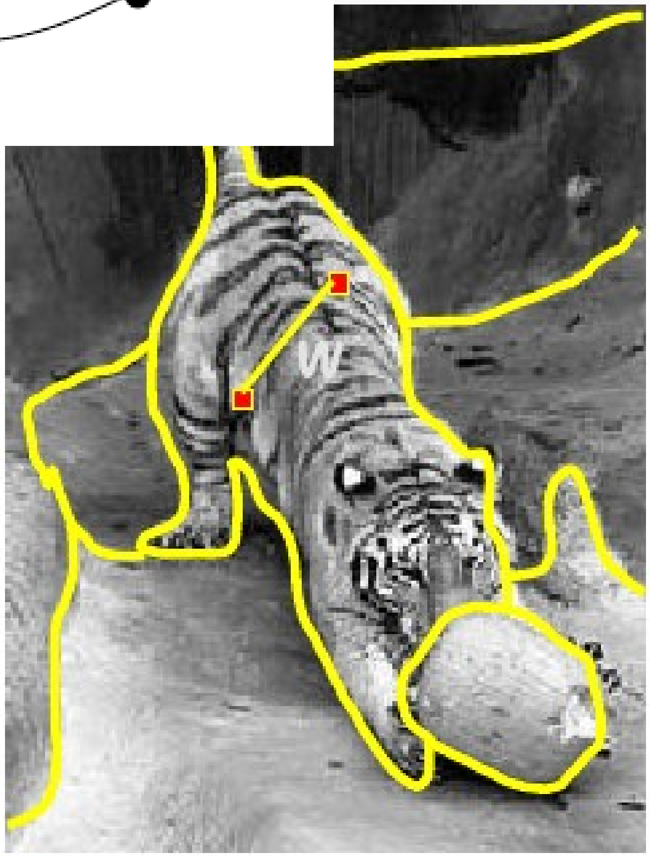
Each link between nodes

(pixels) associated with a cost

cpq - measures similarity-

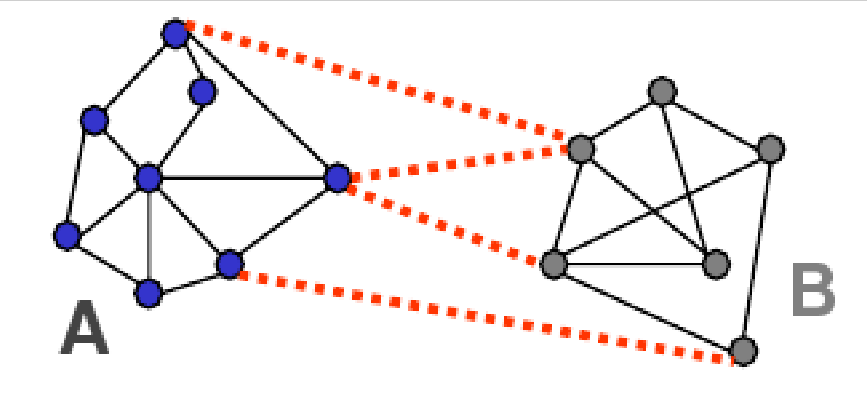
inversely proportional

to difference in feature



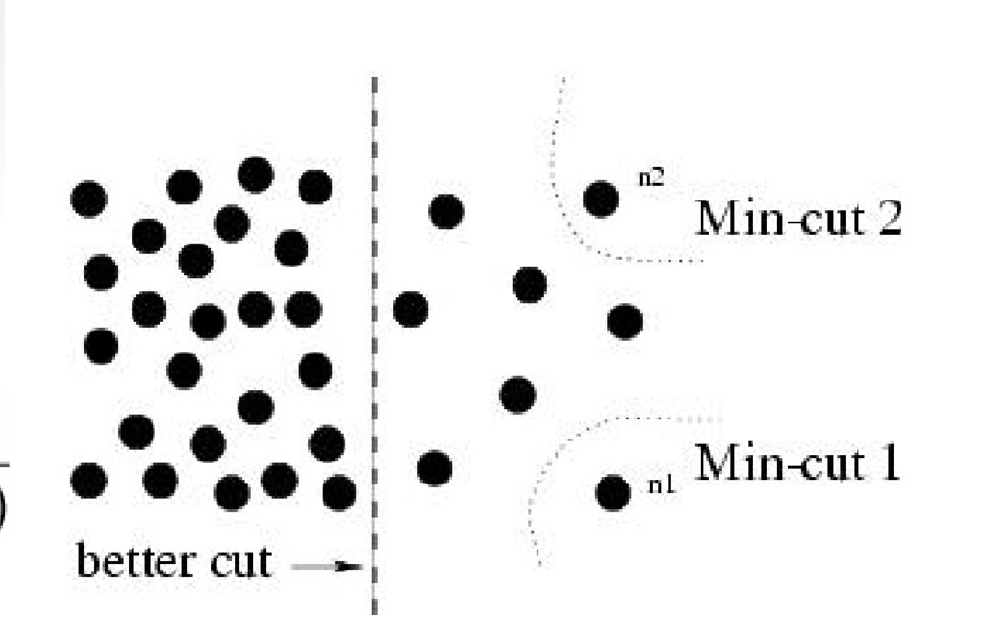
Find Cut that minimizes the

cost function



$$cut(A,B) = \sum_{p \in A, q \in B} c_{p,q}$$

However large segments are penalized, so fix by normalizing for size of segments



$$Ncut(A,B) = \frac{cut(A,B)}{volume(A)} + \frac{cut(A,B)}{volume(B)}$$

$$assoc(A,V) = \sum_{u \in A, t \in V} c(u,t)$$

- Solved by posing it as a generalized eigenvalue problem.

$$(\mathbf{D} - \mathbf{W})\mathbf{y} = \lambda \mathbf{D}\mathbf{y}$$

$\mathbf{W}$ is the cost matrix : $\mathbf{W}(i, j) = c_{i,j}$;

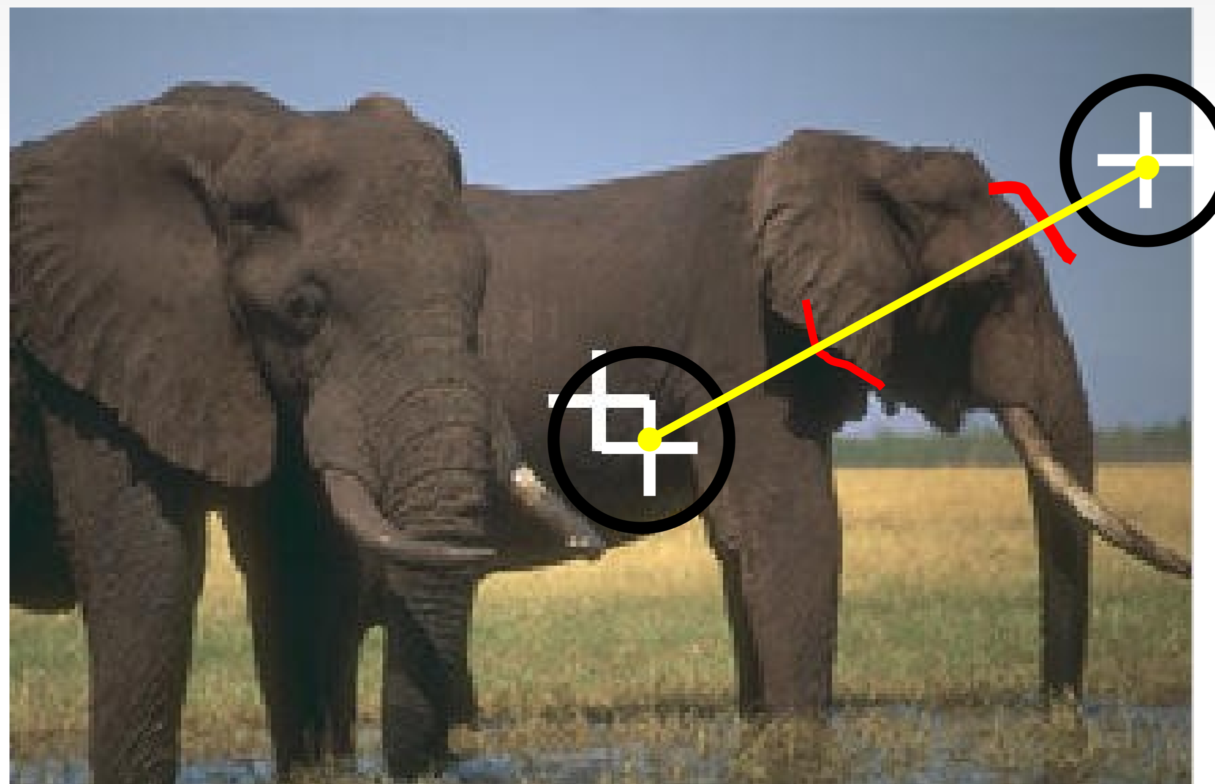
$\mathbf{D}$ is the sum of costs from node i : $\mathbf{D}(i, i) = \sum_j \mathbf{W}(i, j)$; $\mathbf{D}(i,j) = 0$

- **Maximum intevening contour cue**
- sij =Max (mPb(x,y,theta))

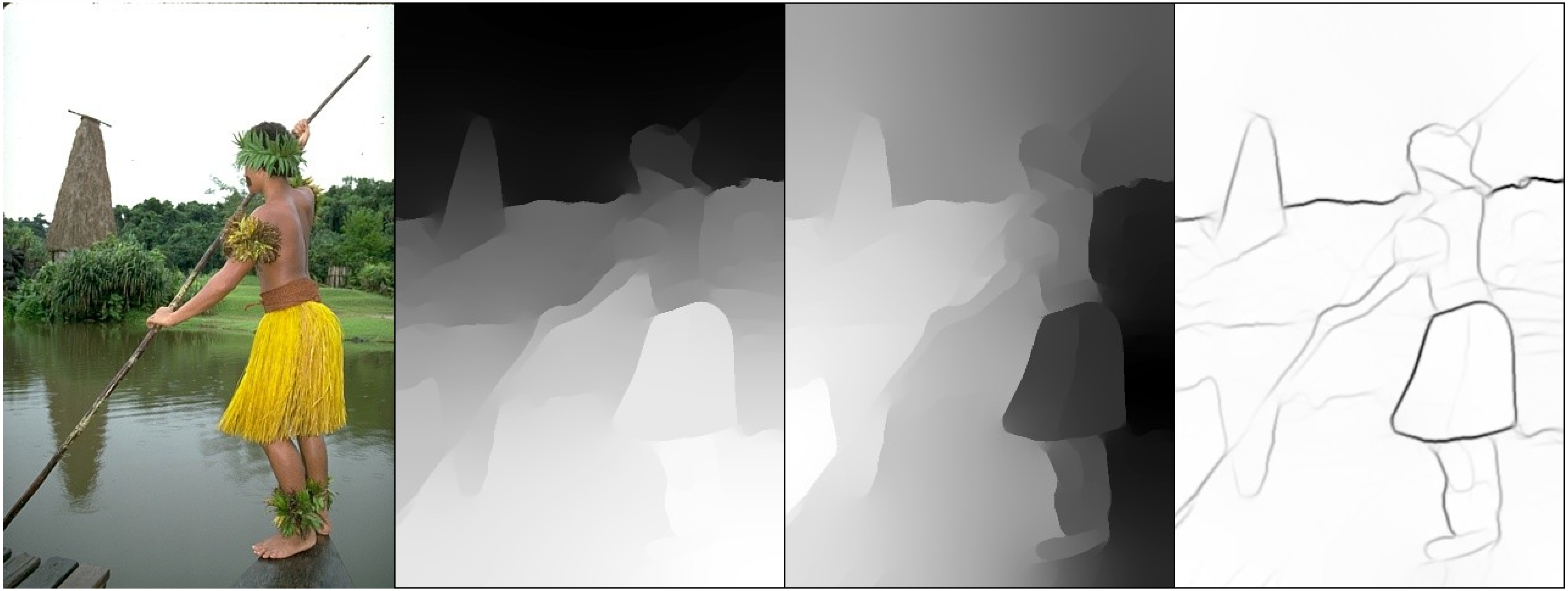
  on line segment between pixel i and j
- Wij= exp(-Cij/k)

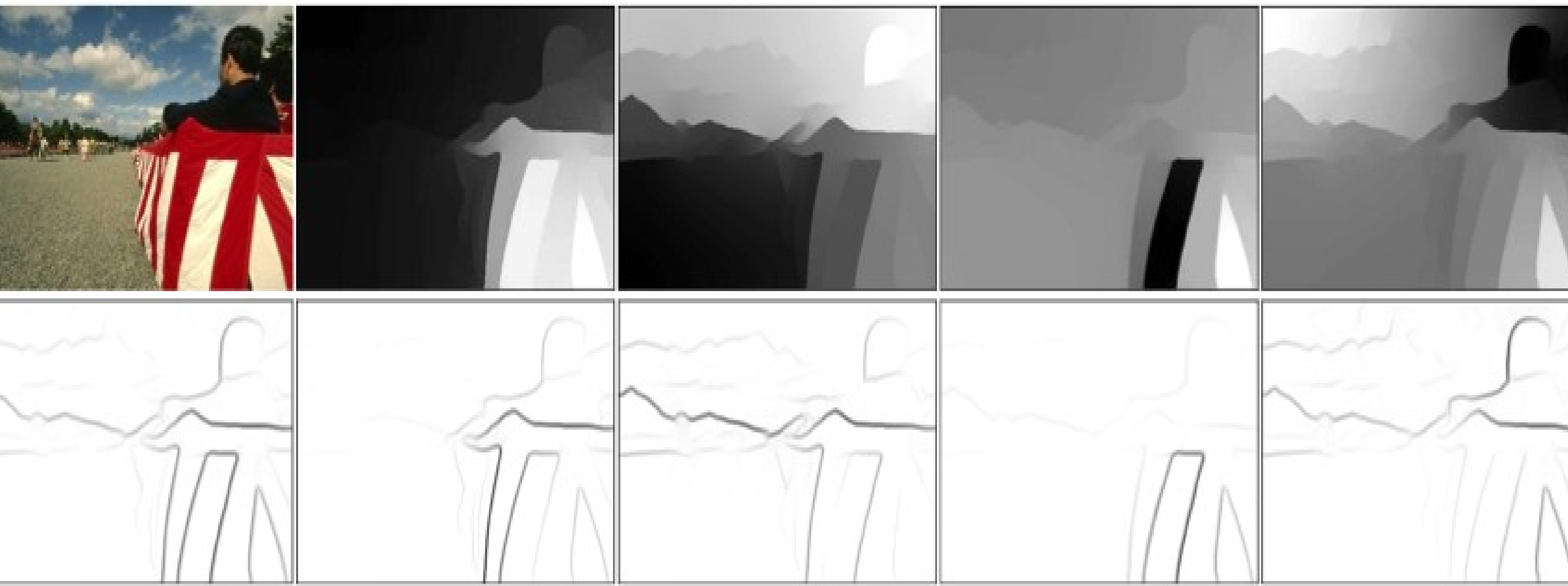
$$mPb(x,y,\theta) = \sum_{i=1}^{9} \alpha_i \cdot G_i(x,y,\theta)$$

Multiscale Pb



- Compute k+1 eigenvectors of the system arnd reshape in the size of original image – sPb
- Contours extracted by taking gaussian derivatives at multiple orientations





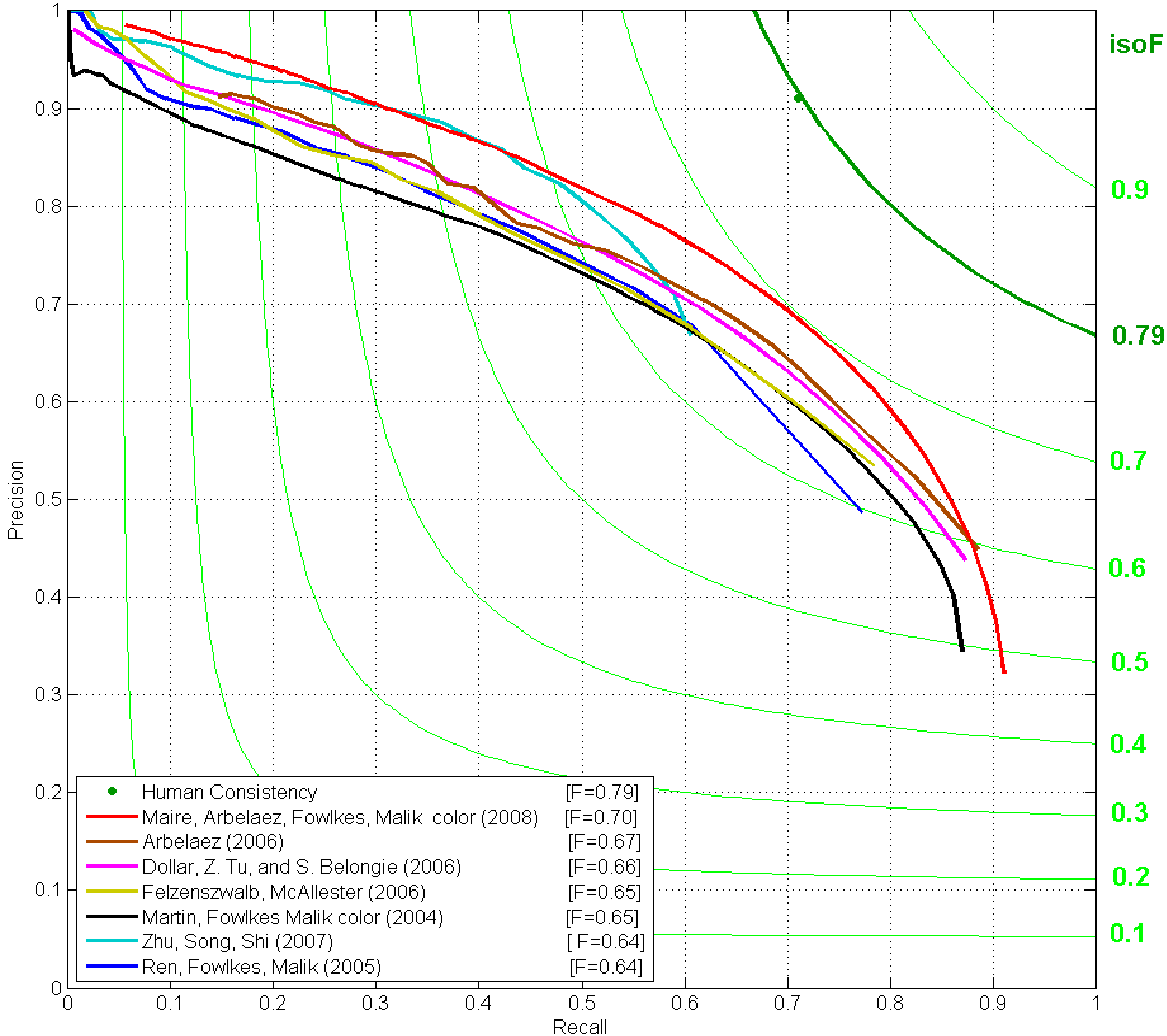
gure 1. **Top:** Original image and first four generalized eigenvectors. **Bottom:** Maximum response over orientations θ of $sPb(x, y, \theta)$,
d of $sPb_{\mathbf{v_j}}(x, y, \theta)$ for each eigenvector $\mathbf{v_j}$.

The signals mPb and sPb convey different information, so a linear combination is taken and the weights are learned from training data

- mPb fires at all edges
- sPb fires only at Salient curves

$$sPb(x, y, \theta) = \sum_{j=1}^{k} \frac{1}{\sqrt{\lambda_j}} \cdot sPb_{\mathbf{v_j}}(x, y, \theta)$$

$$gPb(x, y, \theta) = \sum_{i=1}^{9} \beta_i \cdot G_i(x, y, \theta) + \gamma \cdot sPb(x, y, \theta)$$



isoF
0.9
0.79
0.7
0.6
0.5
0.4
0.3
0.2
0.1
Precision
Recall
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Human Consistency [F=0.79]
Maire, Arbelaez, Fowlkes, Malik color (2008) [F=0.70]
Arbelaez (2006) [F=0.67]
Dollar, Z. Tu, and S. Belongie (2006) [F=0.66]
Felzenszwalb, McAllester (2006) [F=0.65]
Martin, Fowlkes Malik color (2004) [F=0.65]
Zhu, Song, Shi (2007) [F=0.64]
Ren, Fowlkes, Malik (2005) [F=0.64]

Experiments with LabelMe

- Goal of using the boundary detection for generating line drawings that would be useful for object recognition
- Interesting to see how well certain object boundaries as detected by gPb correspond with human segmentations

LabelMe Dataset

- Images with objects labelled
- gPb computed for images with certain objects
- Boundaries in region of object extracted from complete image using the object mask
- Problems with Dataset

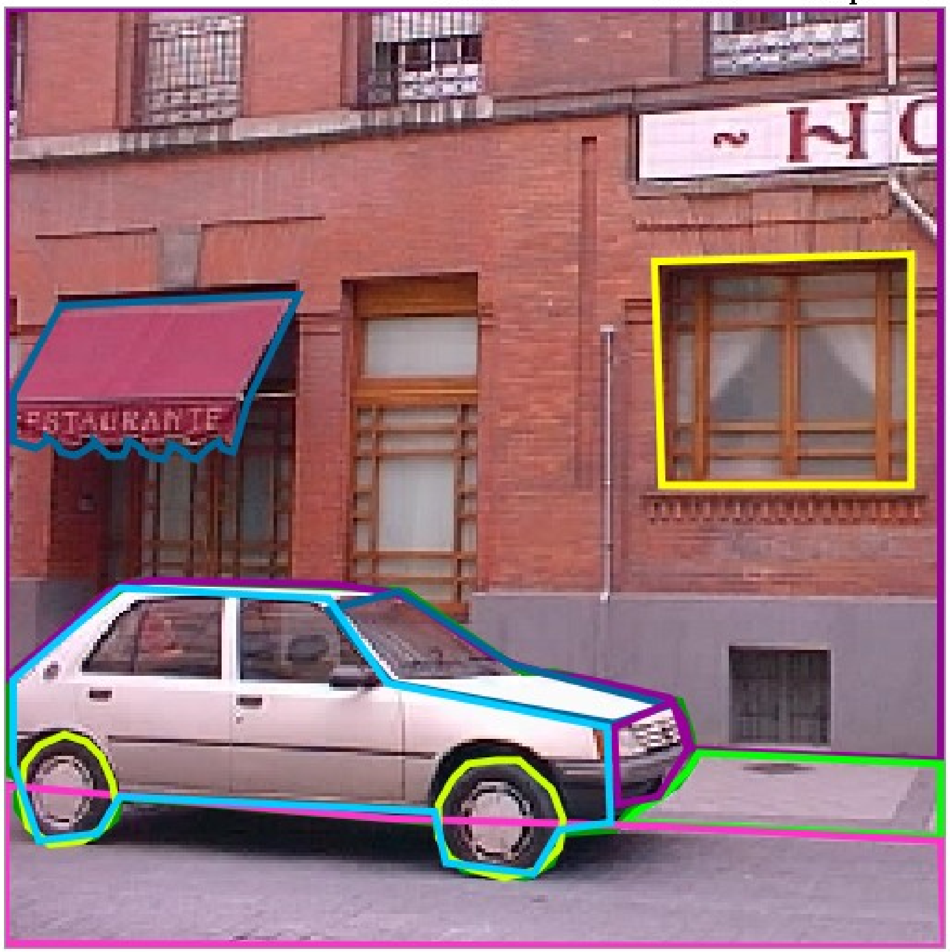
Zoom Erase Help Make 3D Upload image Show me another image

Sign in (why?)

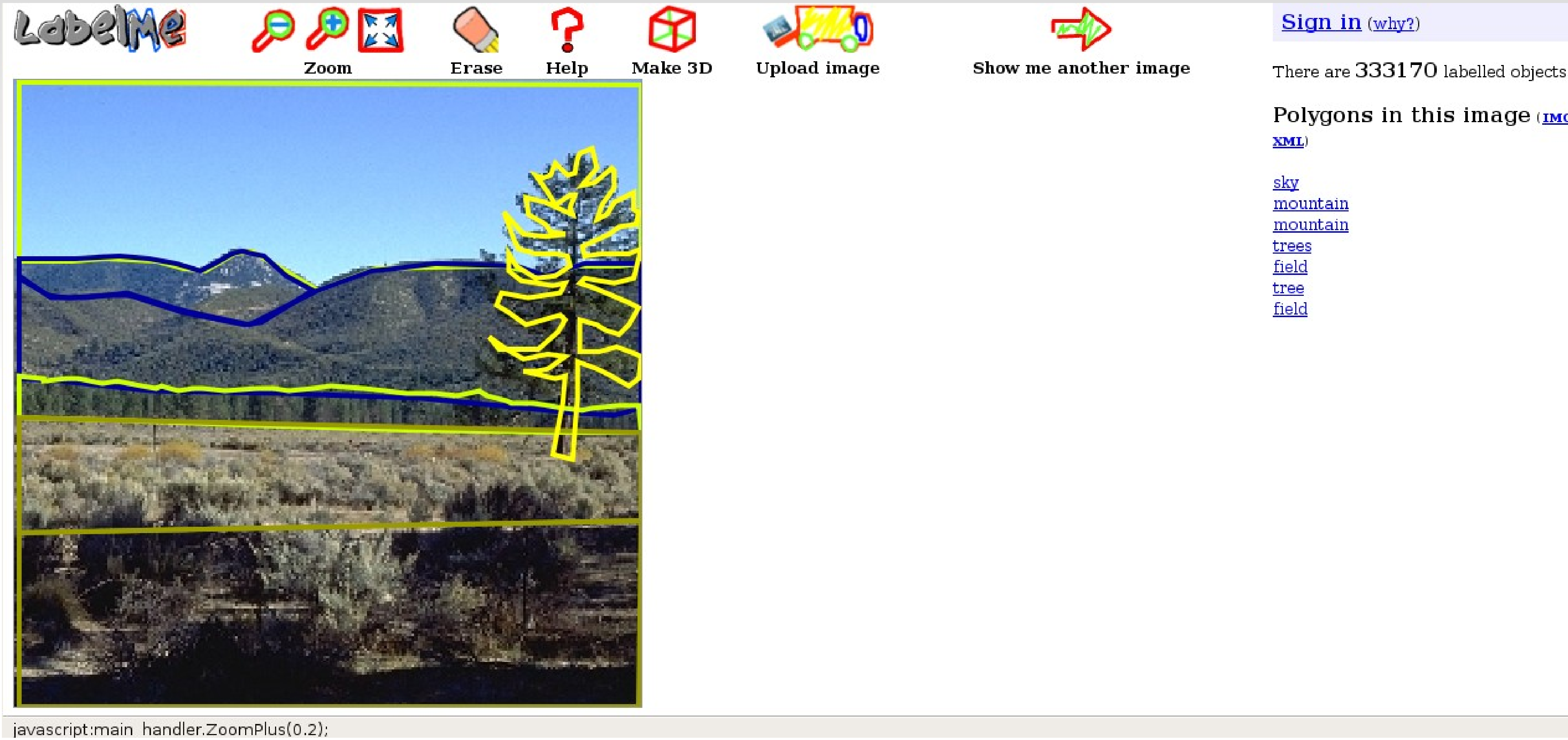
There are 33317

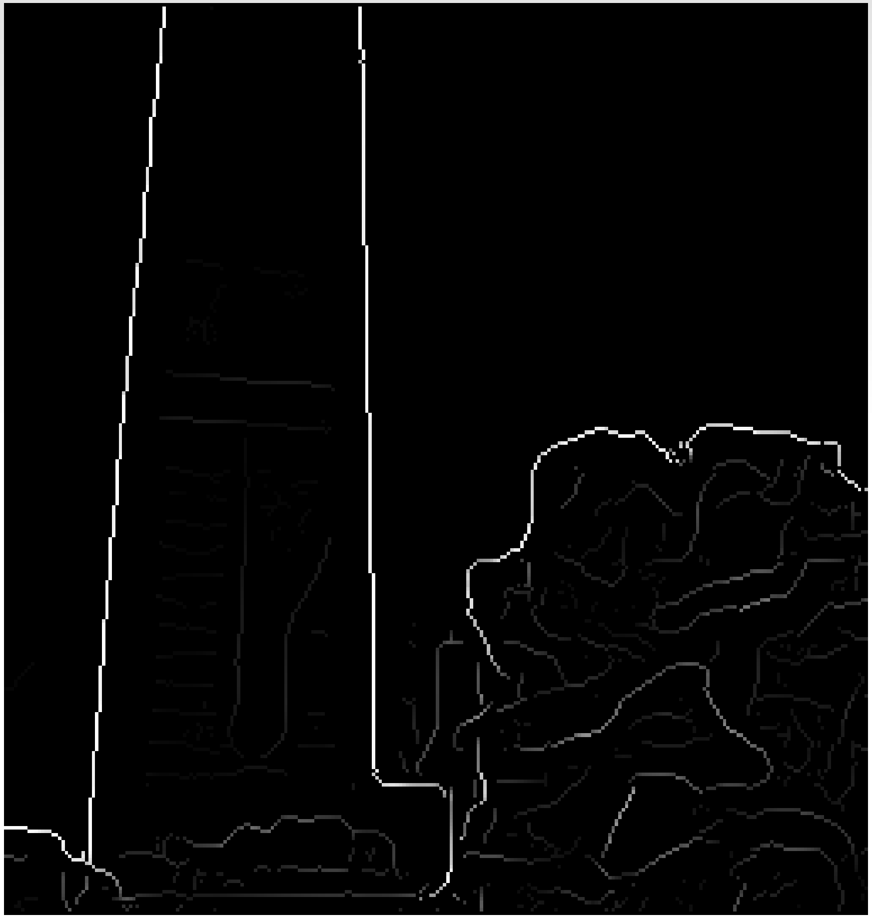
Polygons in th

XML)

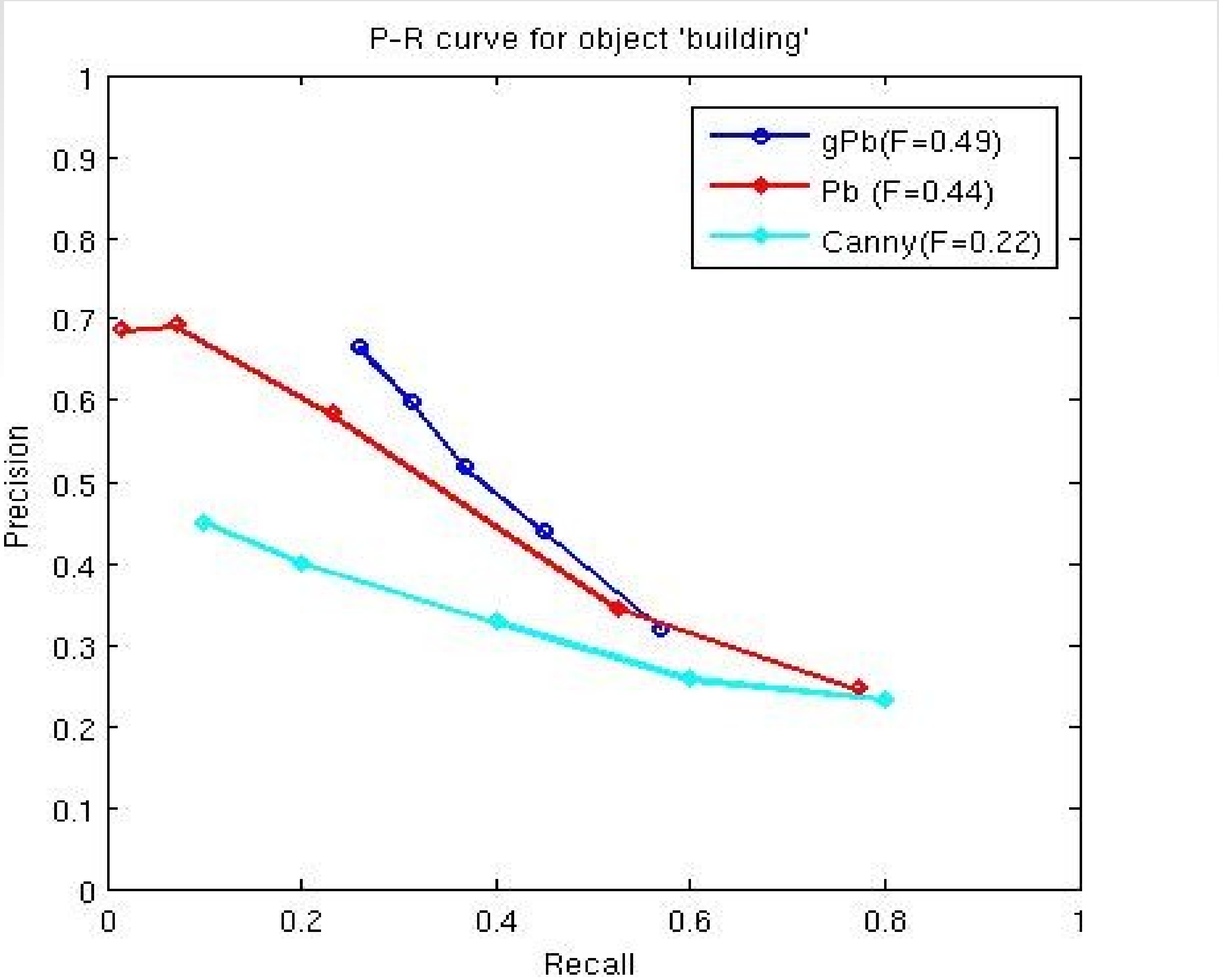


car side
building
sidewalk
road
window
wheel
wheel
awning
car_top_front
car_front
car_right


LabelMe
Zoom
Erase
Help
Make 3D
Upload image
Show me another image
Sign in (why?)
There are 333170 labelled objects
Polygons in this image (IMG, XML)
sky
mountain
mountain
trees
field
tree
field
javascript:main_handler.ZoomPlus(0.2);

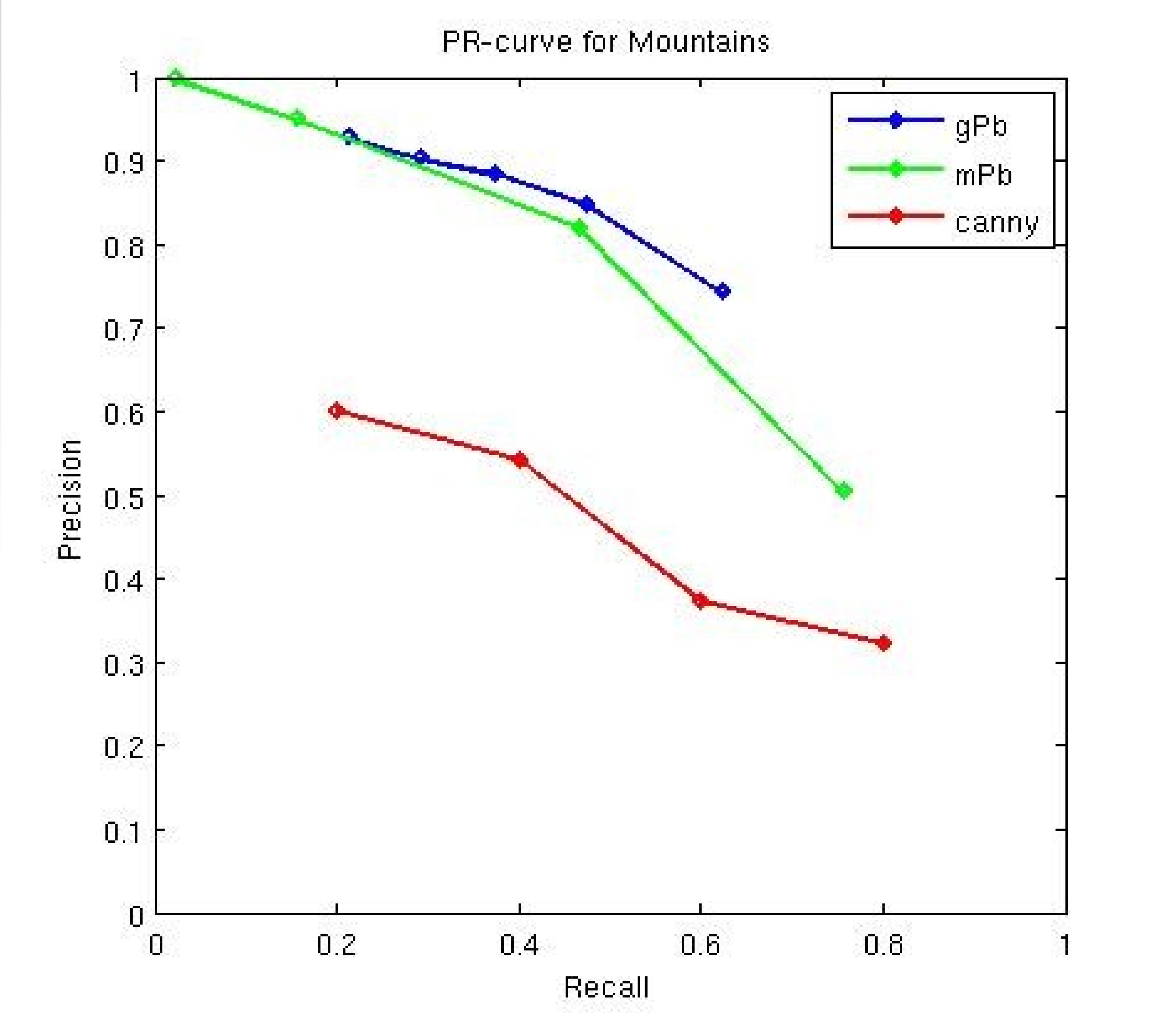


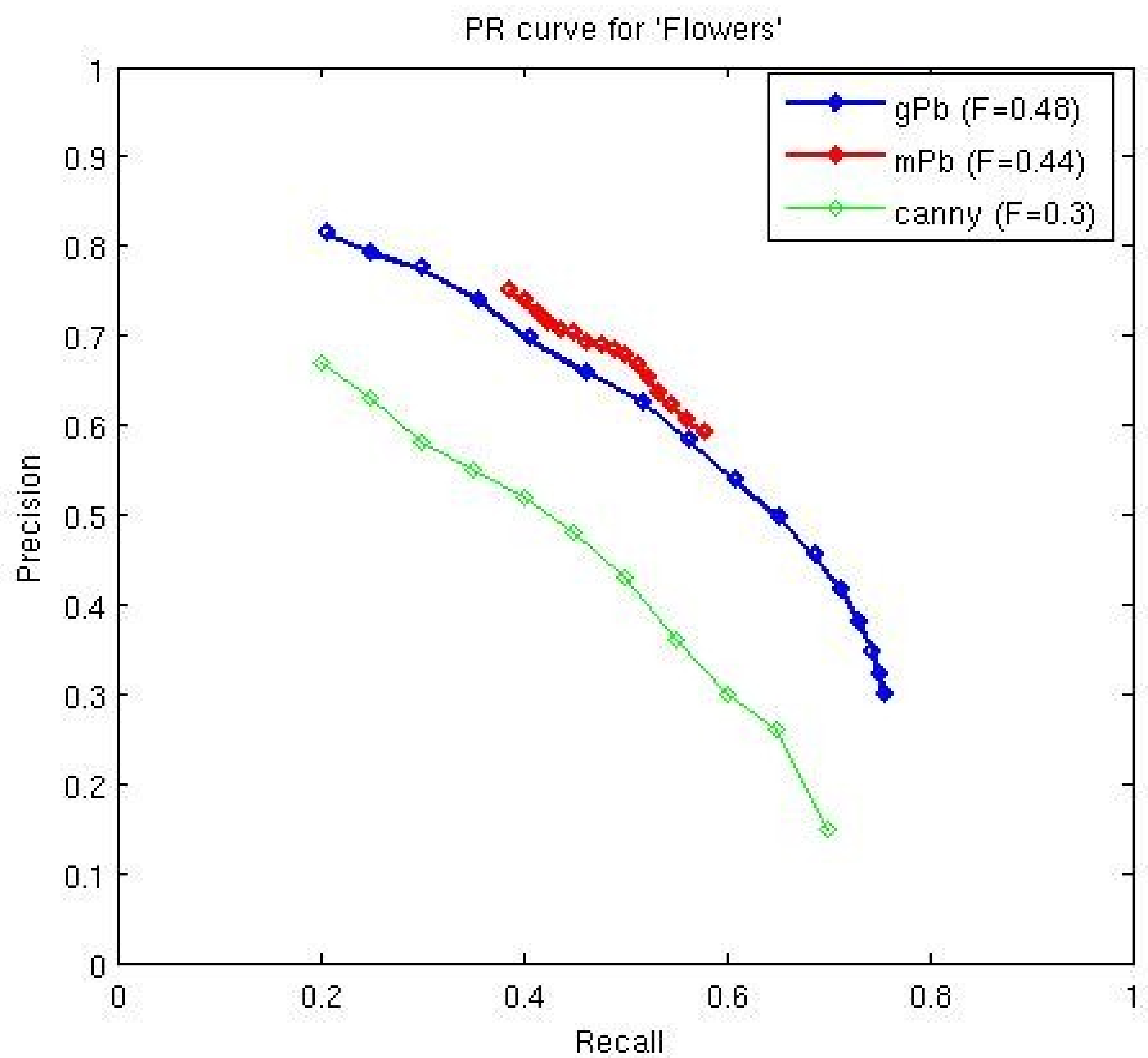
Results



P-R curve for object 'building'
gPb(F=0.49)
Pb (F=0.44)
Canny(F=0.22)
Precision
Recall
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1

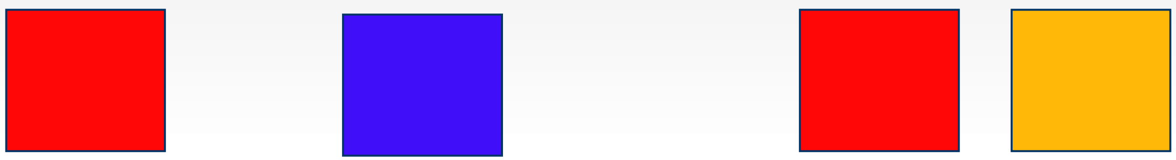



PR-curve for Mountains
Precision
Recall
gPb
mPb
canny
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1


PR curve for 'Flowers'
gPb (F=0.48)
mPb (F=0.44)
canny (F=0.3)
Precision
Recall
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1

Thank You

Which is more similar?



L*a*b* was designed to be uniform in that perceptual “closeness” corresponds to Euclidean distance in the space.

Color