专题四 图像分割

张潇

航天航空学院

- David R. Martin, Charless C. Fowlkes, and Jitendra Malik. "Learning to detect Natural image boundaries using local brightness, color, and texture cues. “

- Arbelaez, P., Maire, M., Fowlkes, C., &Malik, J "Contour detection and hierarchical image segmentation. "

- Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. “Grabcut : Interactive foreground extraction using iterated graphcuts."

Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues

David R. Martin
Charless C. Fowlkes
Jitendra Malik

David R. Martin

before

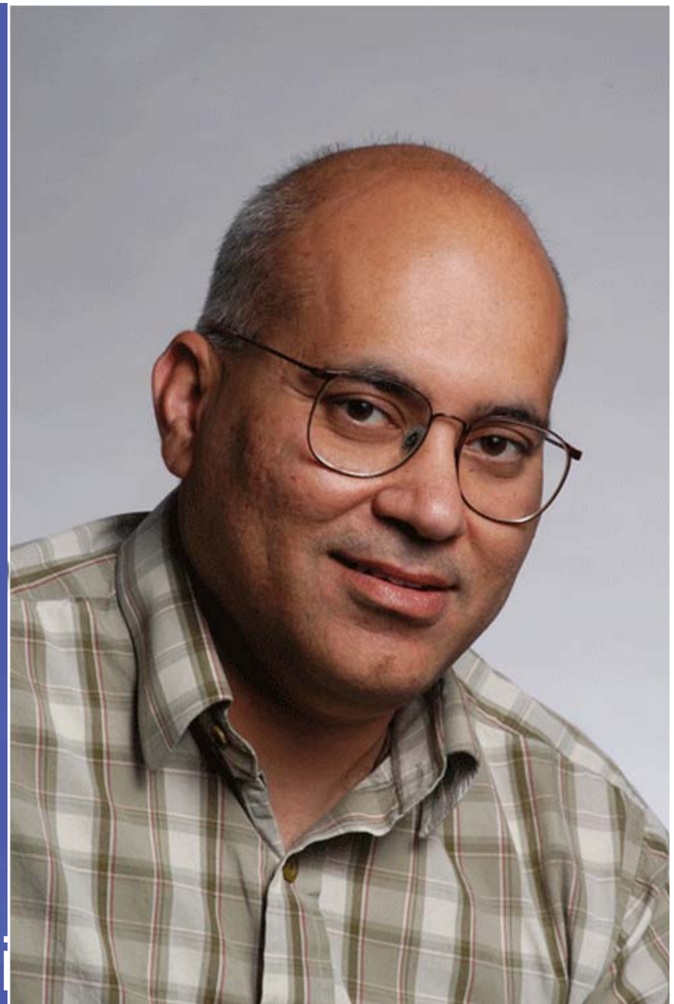
assistant professor
Computer Science Department
Boston College

now

3D Vision Group
Streetview project
Google

Jitendra Malik

Arthur J. Chick Professor
of EECS
University of California at Berkeley

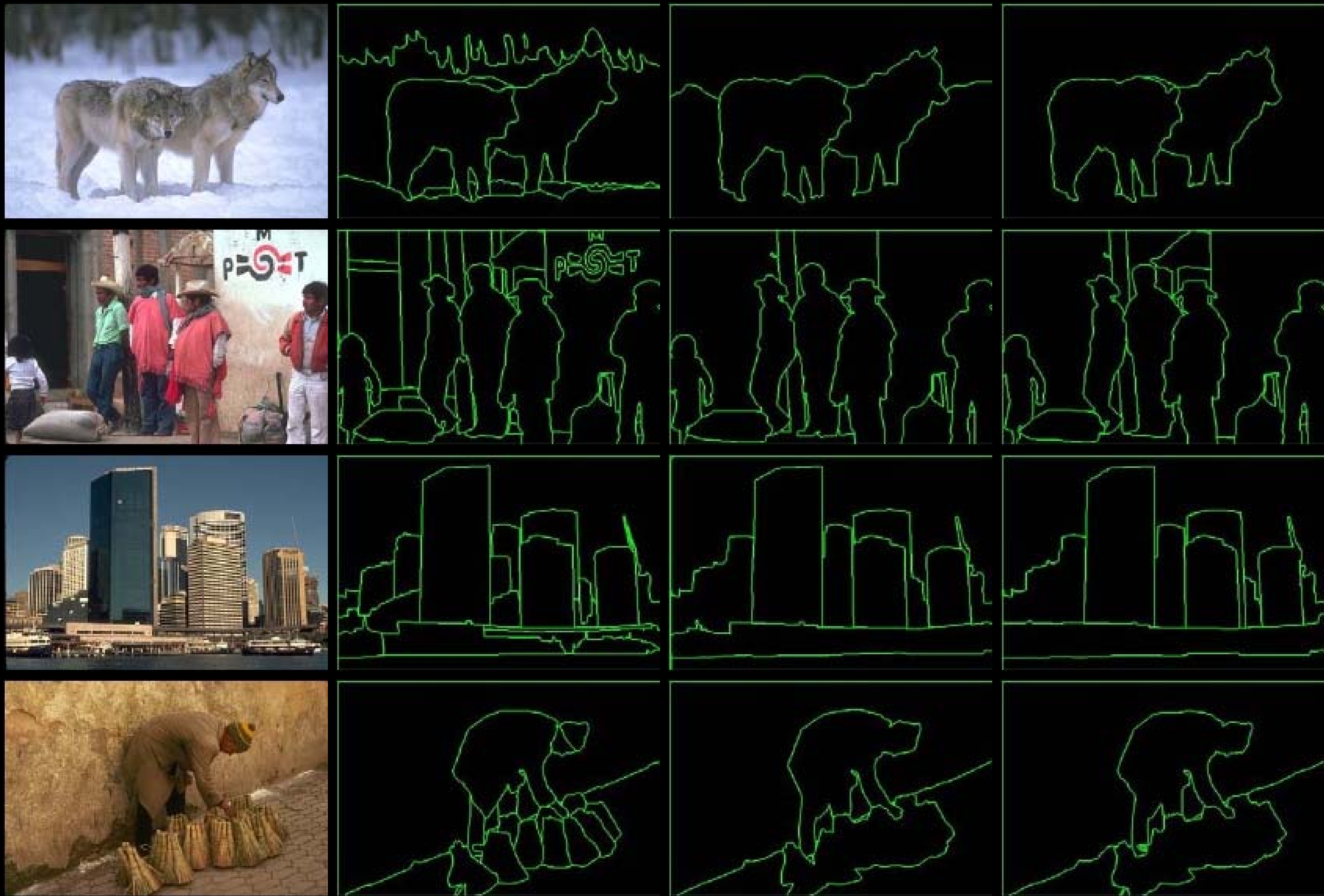


Computer vision

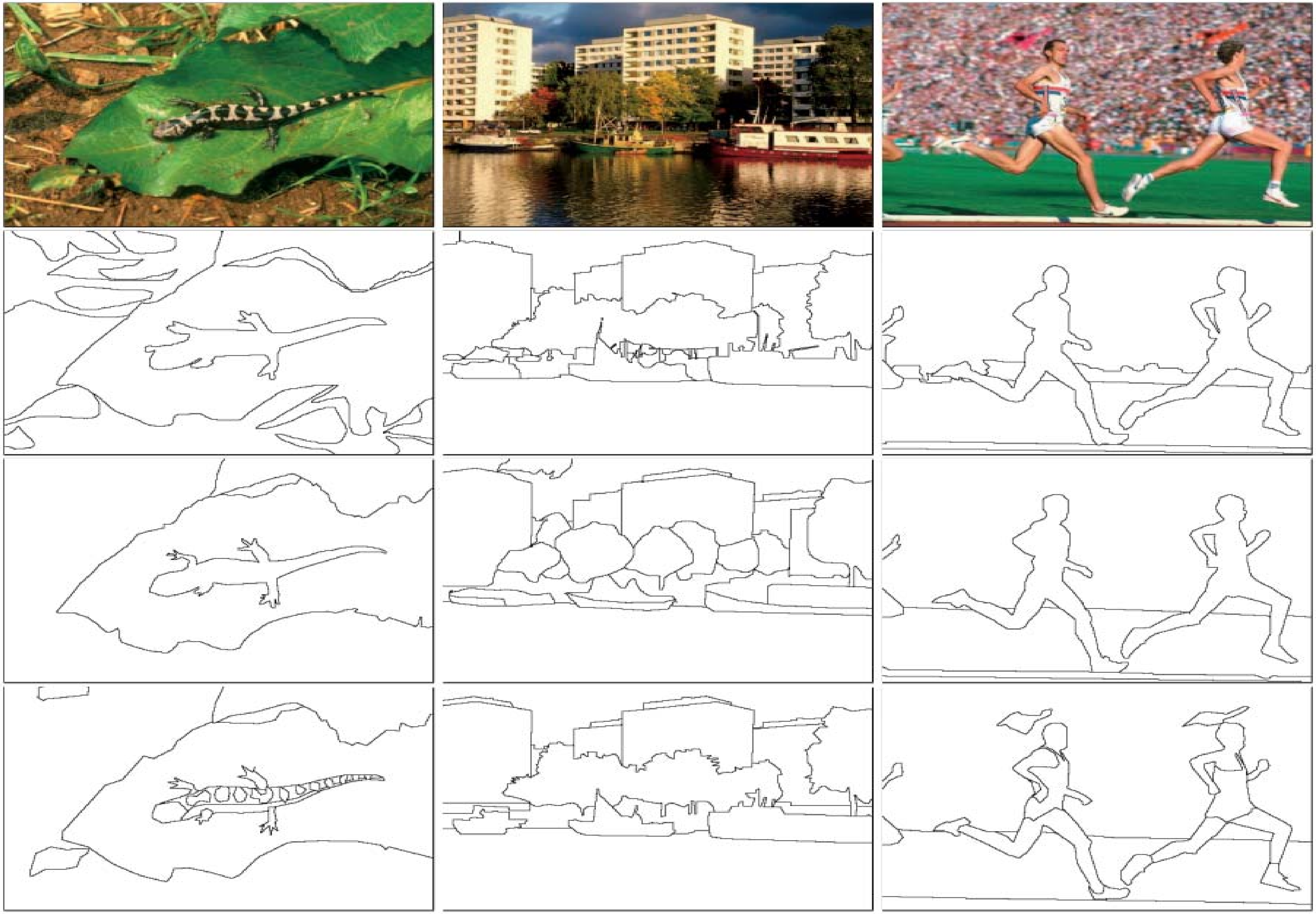
computational modeling of human vi
computer graphics
analysis of biological images

个人主页：http://www.cs.berkeley.edu/~malik/

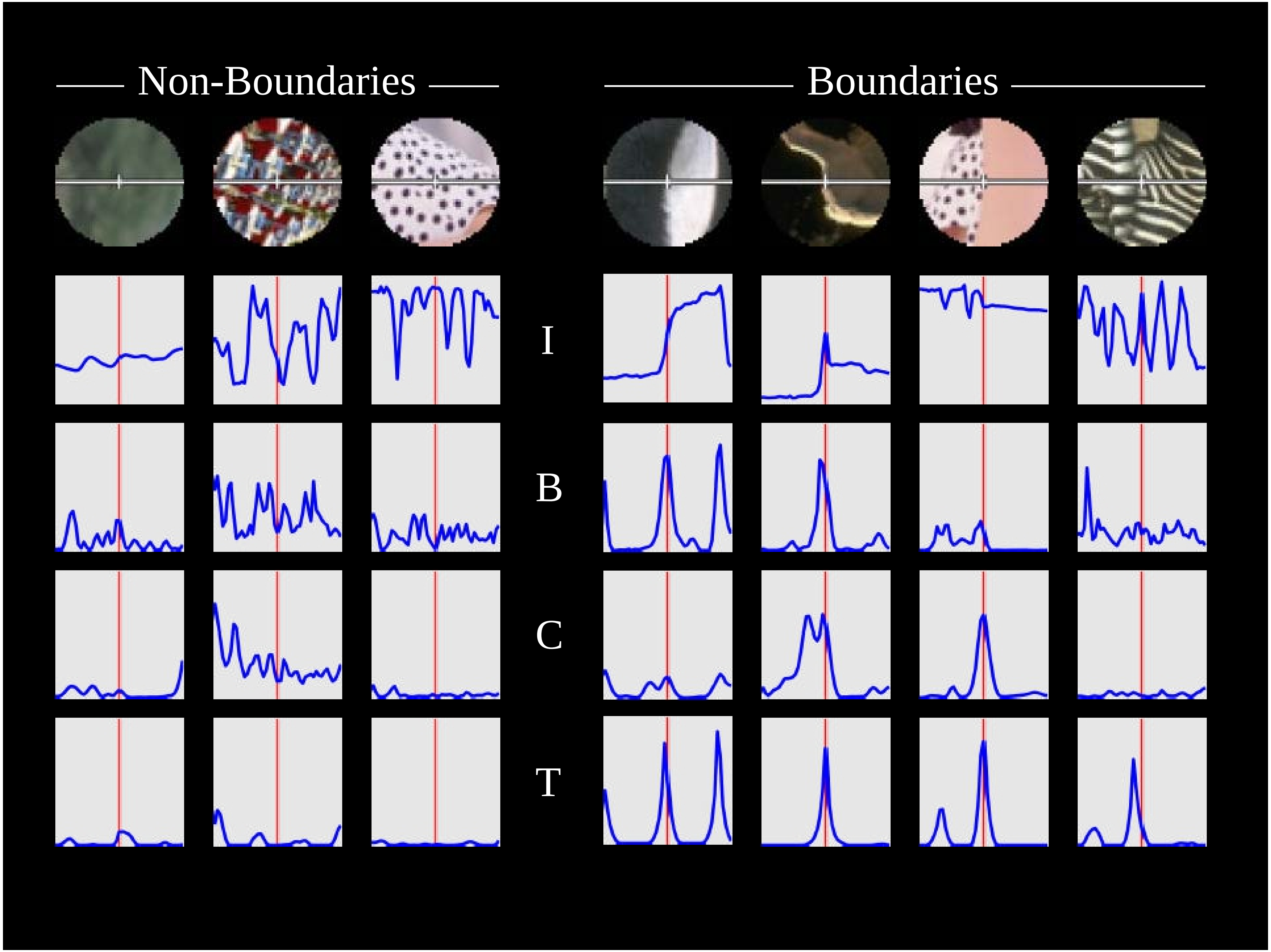
Human Segmentation Dataset

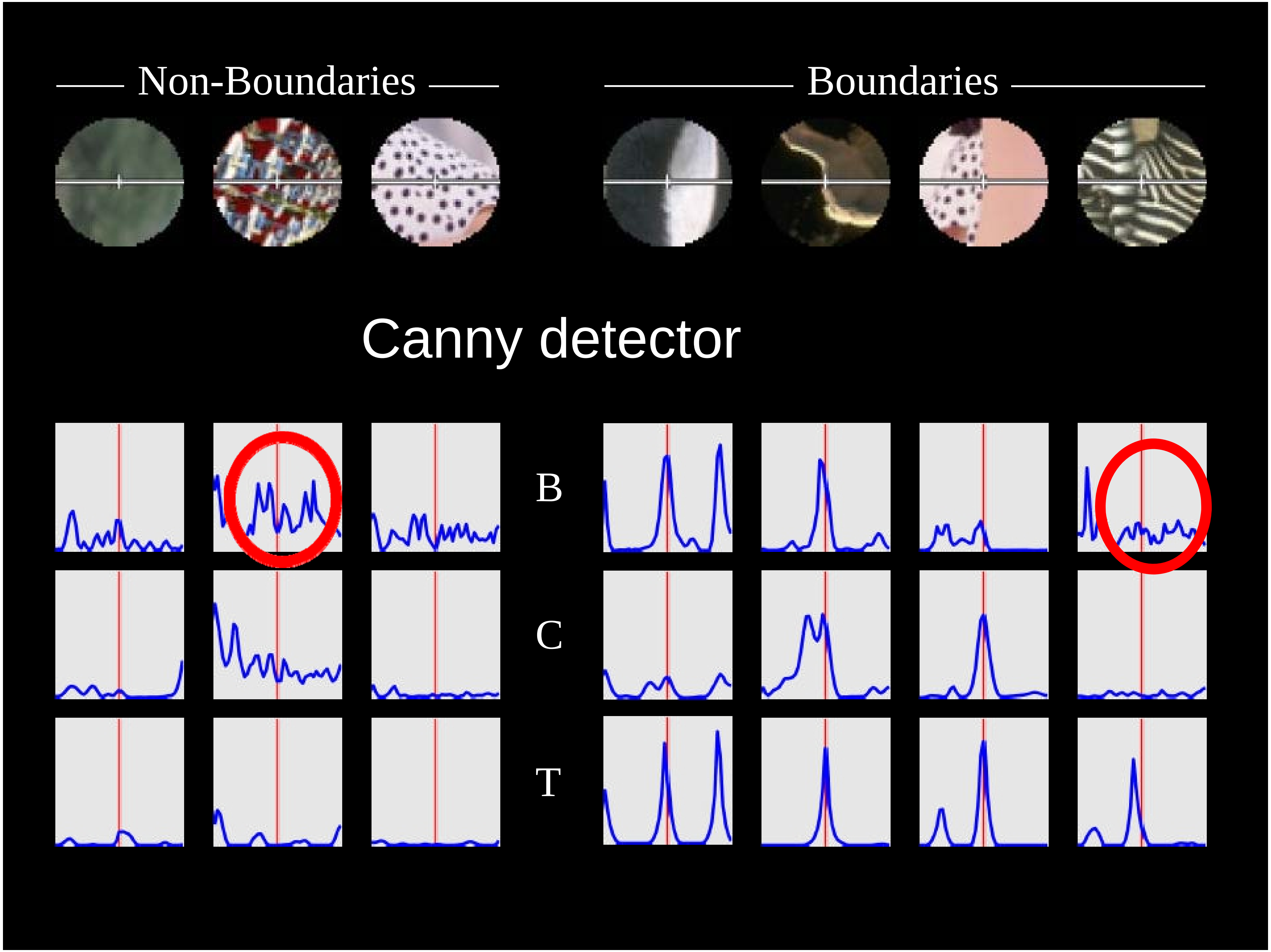


Berkeley Segmentation Dataset



Berkeley Segmentation Dataset


Non-Boundaries
Boundaries
I
B
C
T


Non-Boundaries
Boundaries
Canny detector
B
C
T

Brightness, Color and Texture Features

- Brightness Features
  - oriented energy
  - brightness gradient
- Color Features
  - color gradient
- Texture Features
  - texture gradient

Oriented Energy （OE）

在实际的自然图像中，亮度边缘不只是由简单的阶跃变化的像素点组成的集合。图像中会包含线、屋脊、阶跃以及其他特征，如阴影边界、镜面反射、相互照明、分布遮光。亮度梯度模型不能对这些特征进行完整的描述。

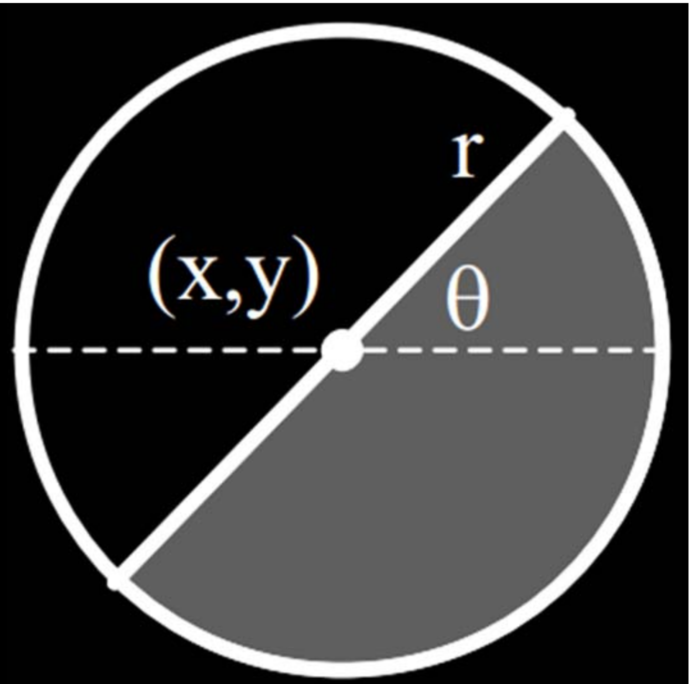
图像中傅里叶相位最大点，即局部能量的峰值为特征点，可以很好地描述出这些复杂的图像特征

方向能量（OE）为特定方向上的能量，也与相位一致性信息正相关，可以检测这些复杂的组合

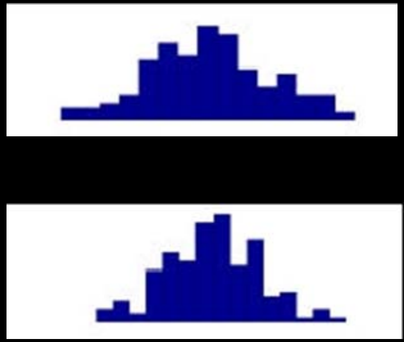
Oriented Energy （OE）

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

Gradient-Based Features



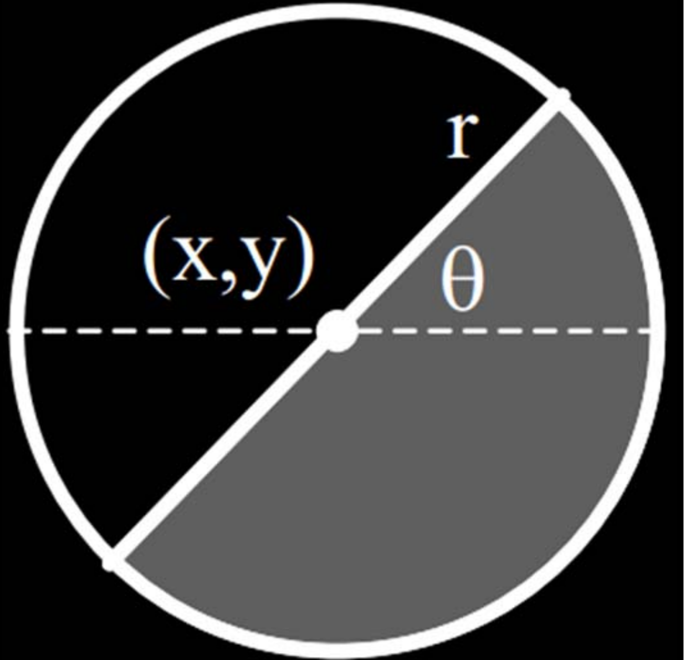
$$\boxed{G(x,y,\theta,r)}$$



- Compare histograms using chi-squared difference

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

Gradient-Based Features



- Brightness and Color Gradients
  - CIE L*a*b*   color-space
  - Estimate distributions of L*, a* and b* values inside analysis window
- Texture Gradient
  - Filter image with even/odd-symmetric filters which resemble V1 receptive fields.
  - Estimate distribution of vector-quantized filter responses inside analysis window

$$\boxed{G(x, y, \theta, r)}$$

- **Compare histograms using chi-squared difference**

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

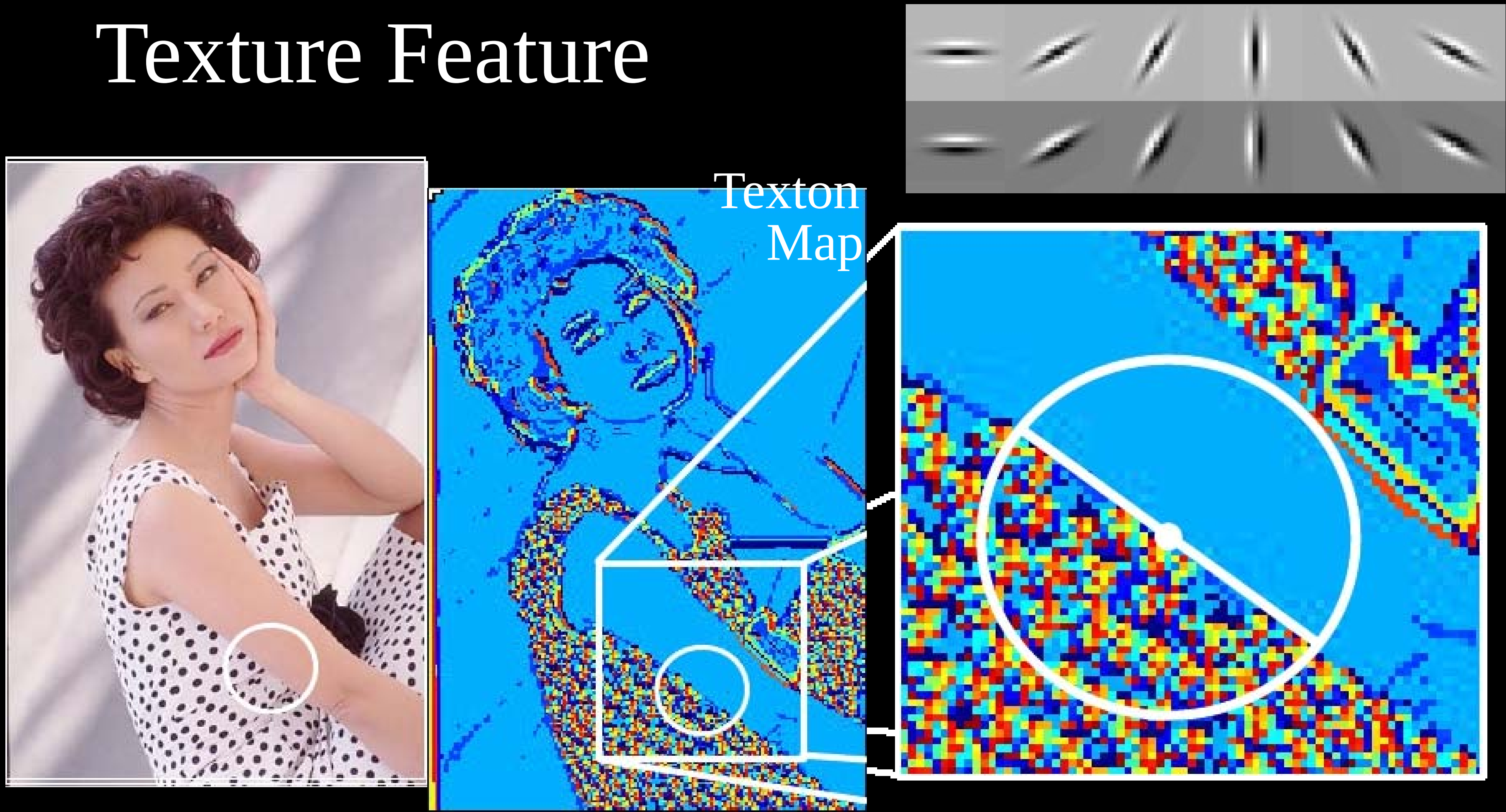
CIE L*a*b*

由国际照明委员会提出
Commission Internationale d'Eclairage

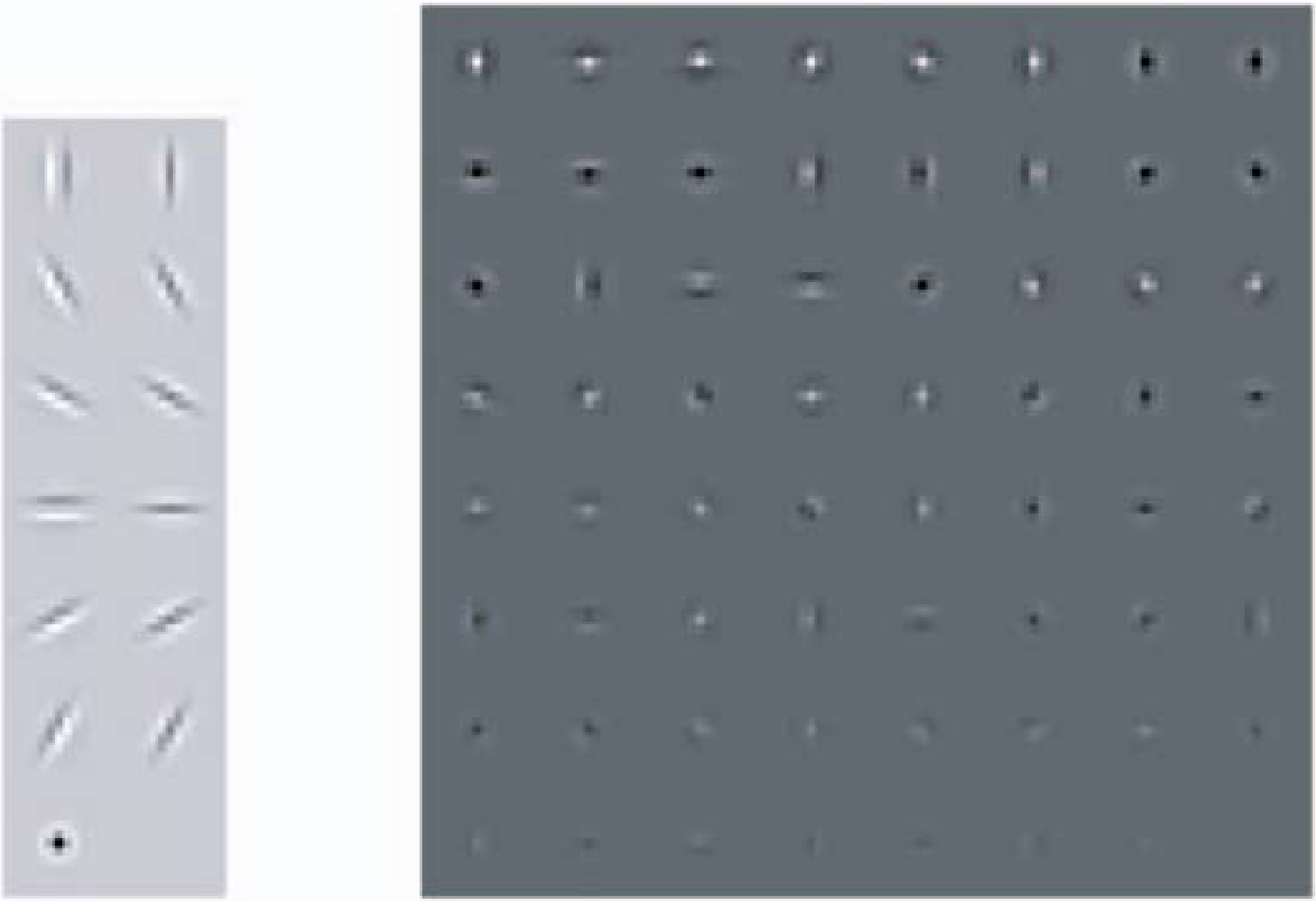
- L*表示颜色的亮度
  L* = 0指示黑色而L* = 100指示白色
- a*表示颜色在红色和绿色之间的位置
  a*负值指示绿色而正值指示红色
- b*表示颜色在黄色和蓝色之间的位置
  b*负值指示蓝色而正值指示黄色

RGB ⟶ XYZ ⟶ L*a*b*

Texture Feature

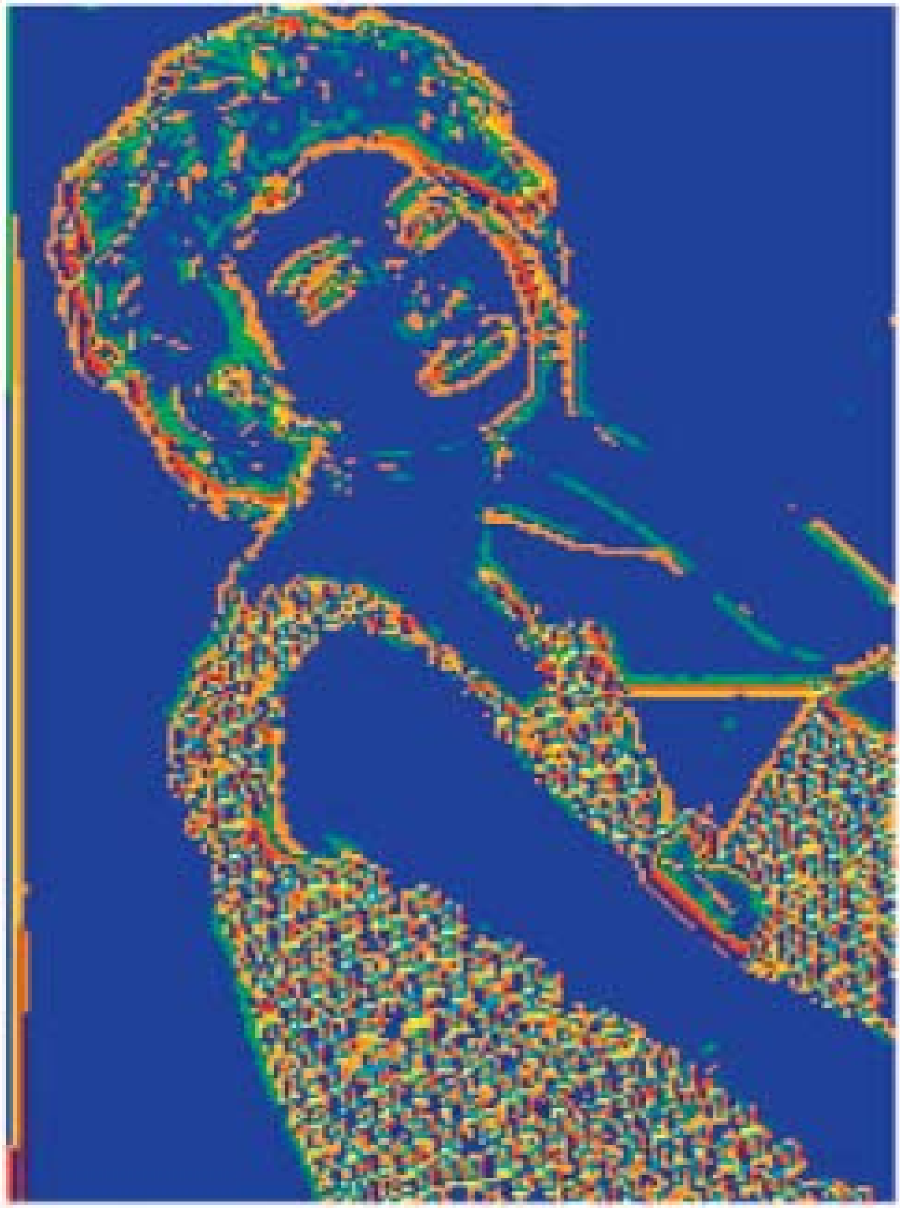


– Filter image with even/odd-symmetric filters which resemble V1 receptive fields.
– Estimate distribution of vector-quantized filter responses inside analysis window



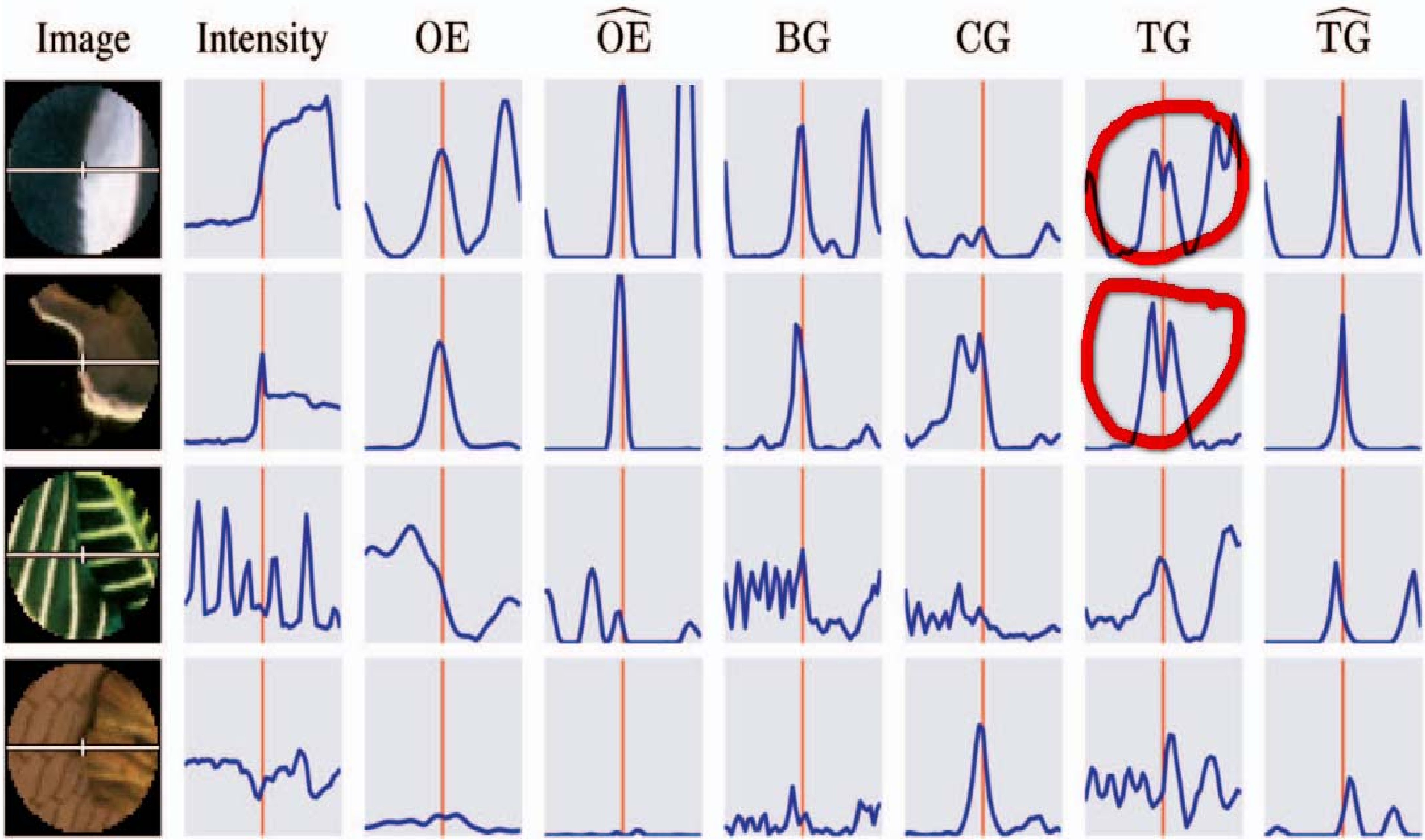
(a) (b)

(c)

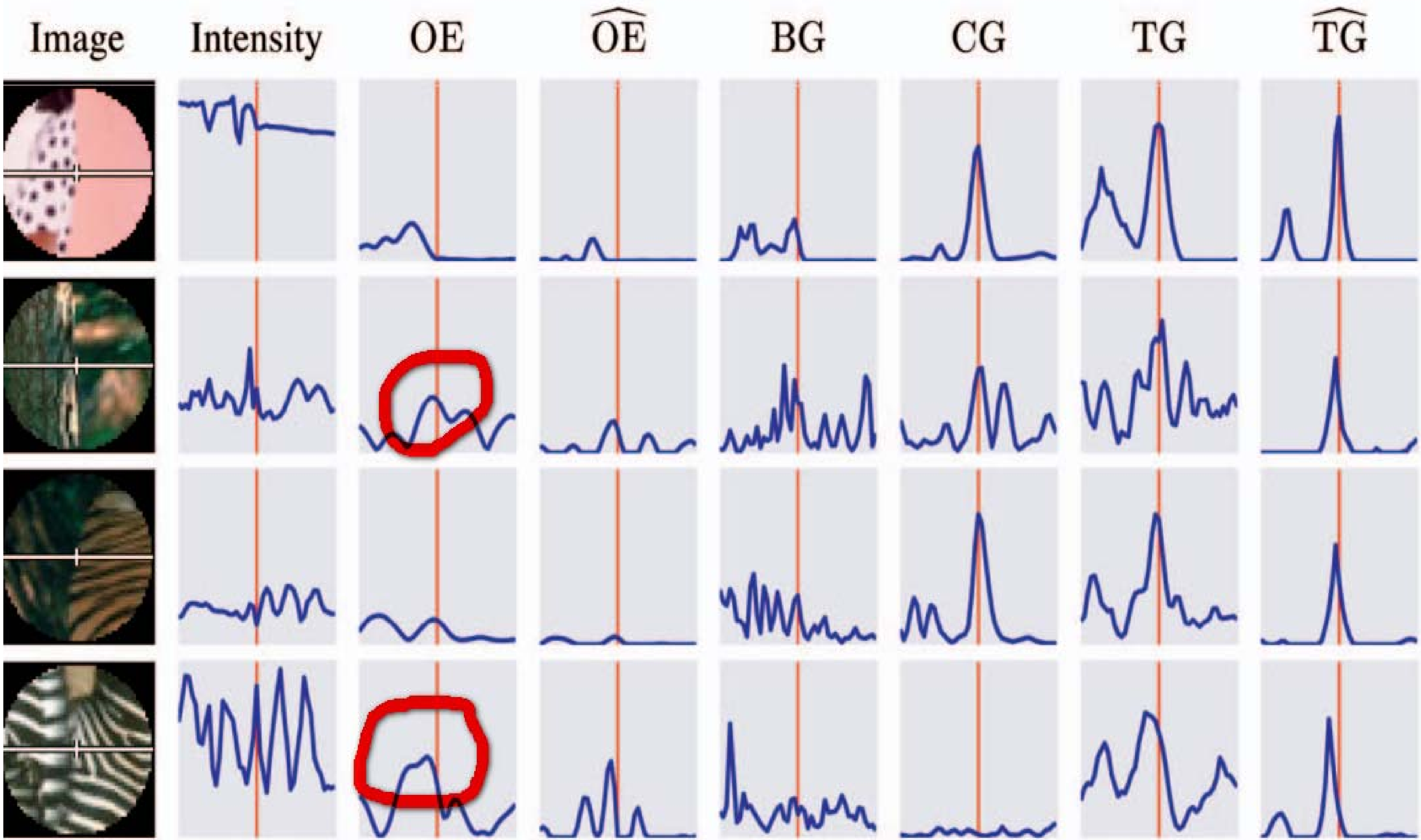


(d)

Localization



Localization



Localization

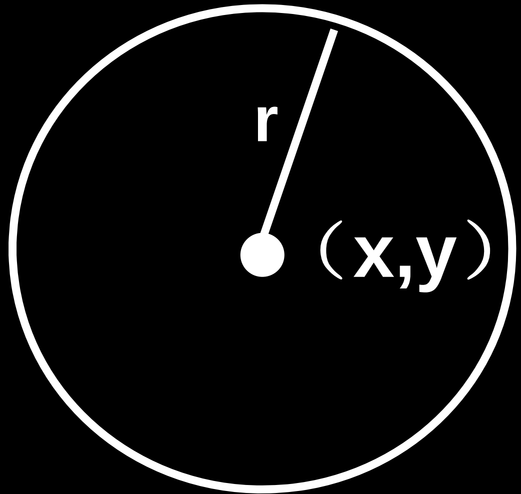
$$d(x) = -|f'(x)|/f''(x)$$

$$\hat{f}(x) = f(x)/d(x),$$

更加平滑稳定的版本

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

Localization



最小二乘法

抛物线拟合

ax^2+bx+c

$$f''(x) = 2a$$

$$f'(x) = b$$

$$\tilde{f}(x) = c$$

$$\hat{f} = -(2c^{+}a^{+})/(|b| + \epsilon)$$

CUE OPTIMIZATION

- optimal one octave range

In units of percentage of the image diagonal

OE, CG, and TG———1.4 to2.8 percent

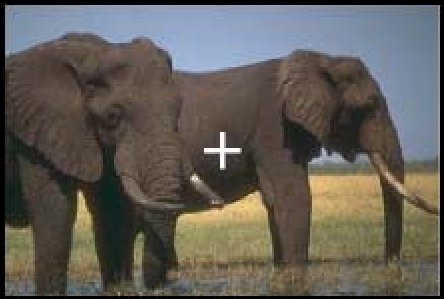
BG———0.75 to 1.5 percent

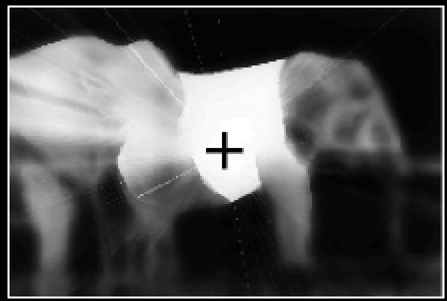
Classifiers for Cue Combination

- 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

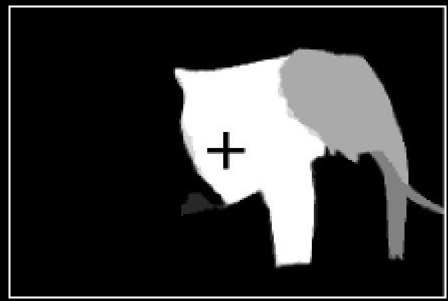
Range over bias, complexity, parametric/non-parametric

Two Evaluation Methods





Estimate W_{ij}



Groundtruth S_{ij}

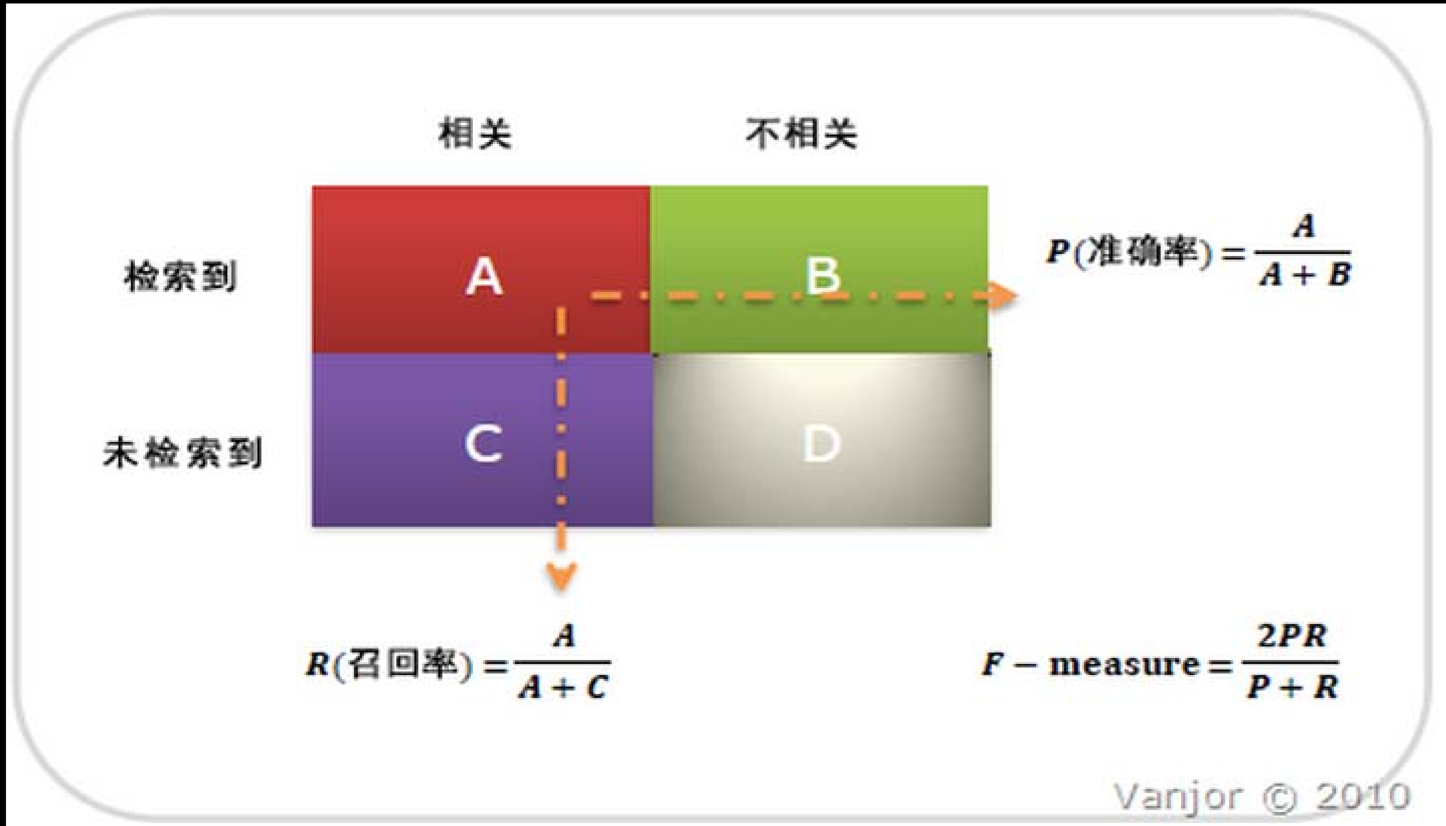
1. Precision-Recall of same-segment pairs
   - Precision is $\mathbf{P}(S_{ij}=1 \mid W_{ij} > t)$
   - Recall is $\mathbf{P}(W_{ij} > t \mid S_{ij} = 1)$
2. Mutual Information between W and S

$$\int p(s,w) \log p(s)p(w) / p(s,w)$$

Evaluation Methodology—— **precision-recall、F-measure**

- Precision is the fraction of detections that are true positives rather than false positives,
- Recall is the fraction of true positives that are detected rather than missed

Evaluation Methodology—— precision-recall、F-measure



召回率**(R) =** 系统检索到的相关文件 **/** 系统所有相关的文件总数

准确率**(P) =** 系统检索到的相关文件 **/** 系统所有检索到的文件总数

F-Measure是**Precision**和**Recall**加权调和平均

F=$PR/(P+R)$

Evaluation Methodology—— precision-recall、F-measure

- Precision is the probability that the detector’s signal is valid,

- Recall is the probability that the ground truth data was detected.

- F-Measure

$$\alpha = 0.5 F$$
$$= PR / (\alpha P + (1-\alpha) R)$$

Evaluation Methodology—— precision-recall、F-measure

$$P_b(x,y)$$ → a binary boundary map

↓

human boundary map

↓

P，R，F

Evaluation Methodology—— corresponding boundary map

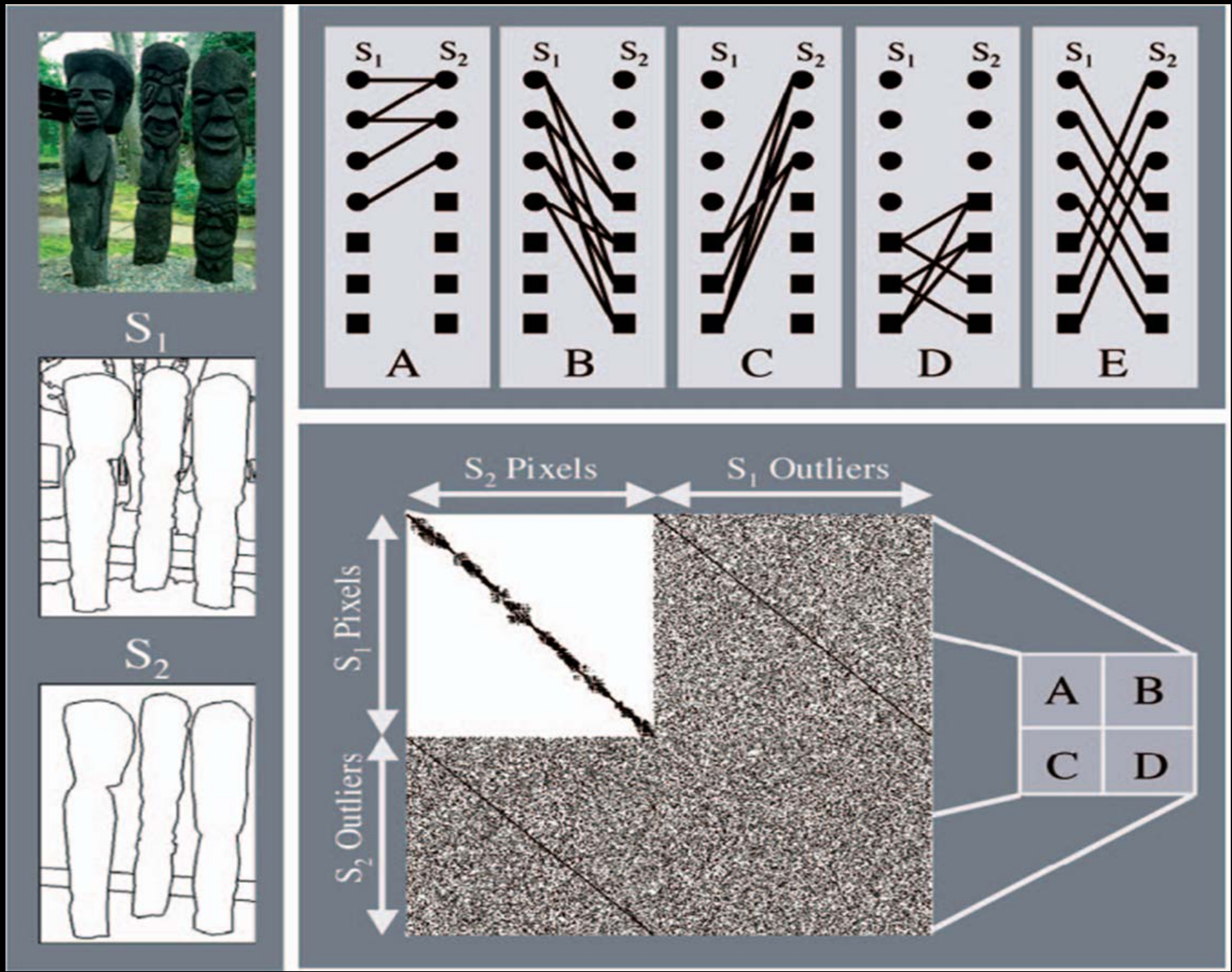
Goldberg’s CSA package

a sparse version

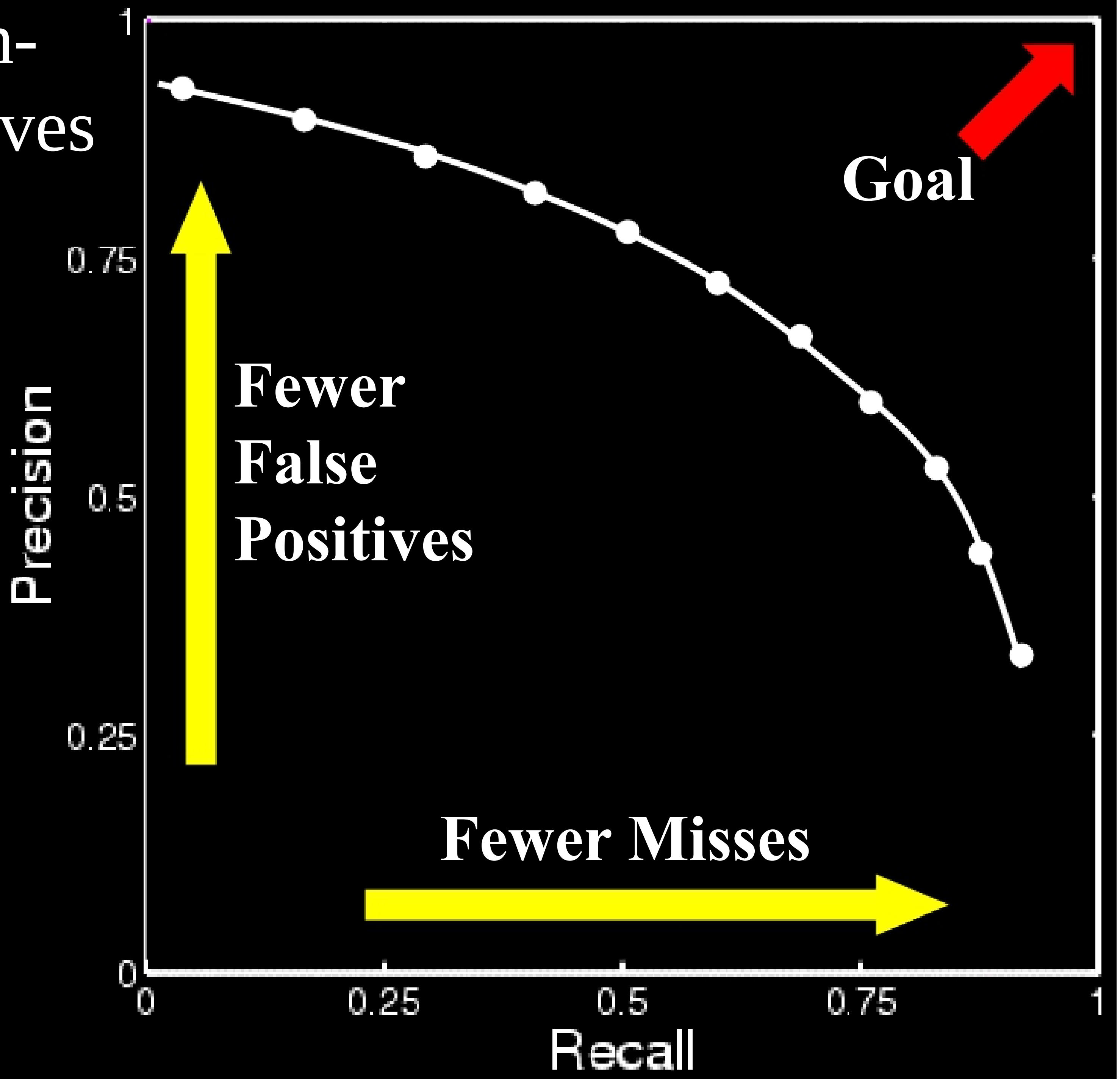
Weight &relative distance

$w > d_{\max}$ $w \leq d_{\max}$

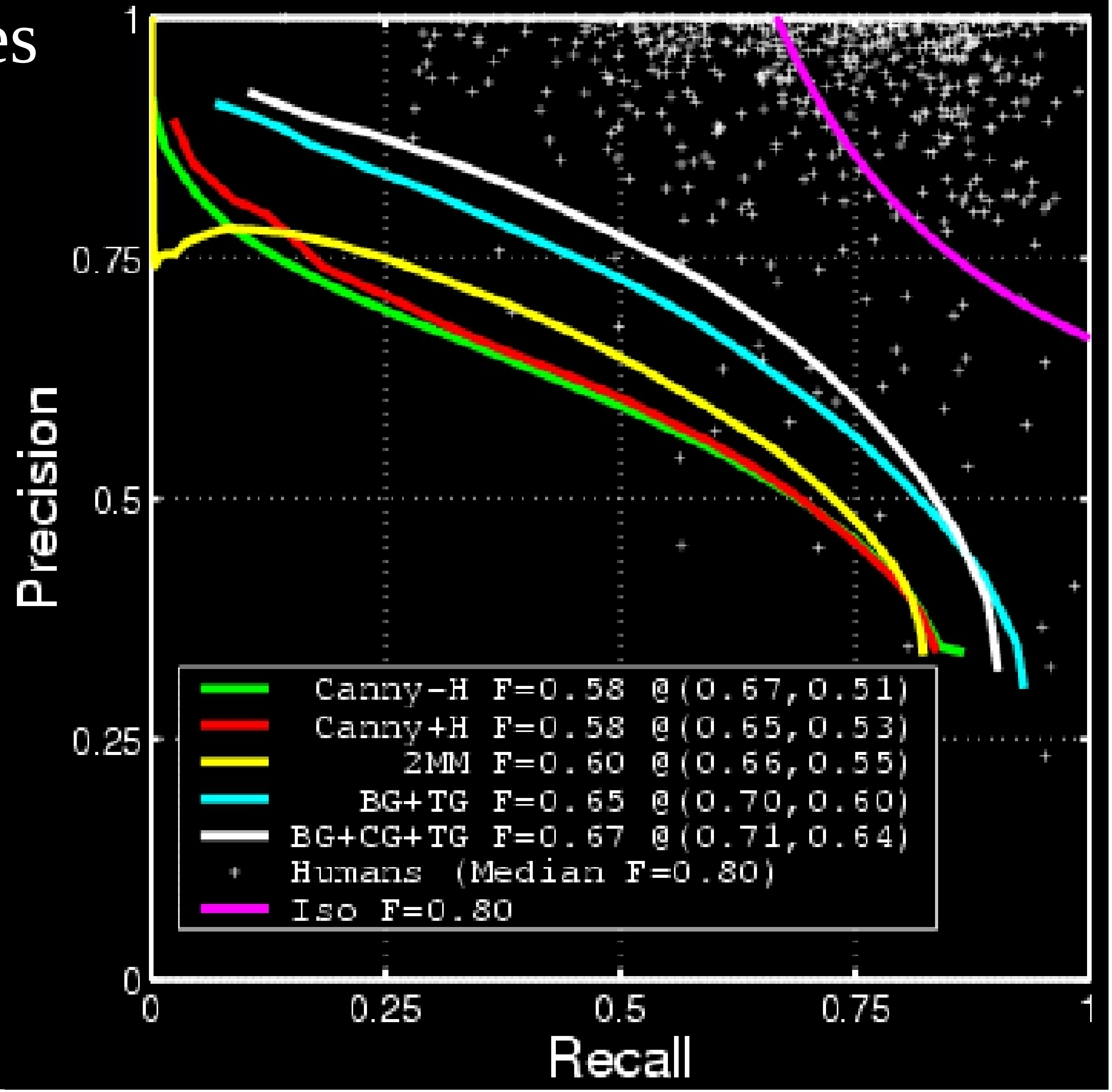
Evaluation Methodology—— corresponding boundary map



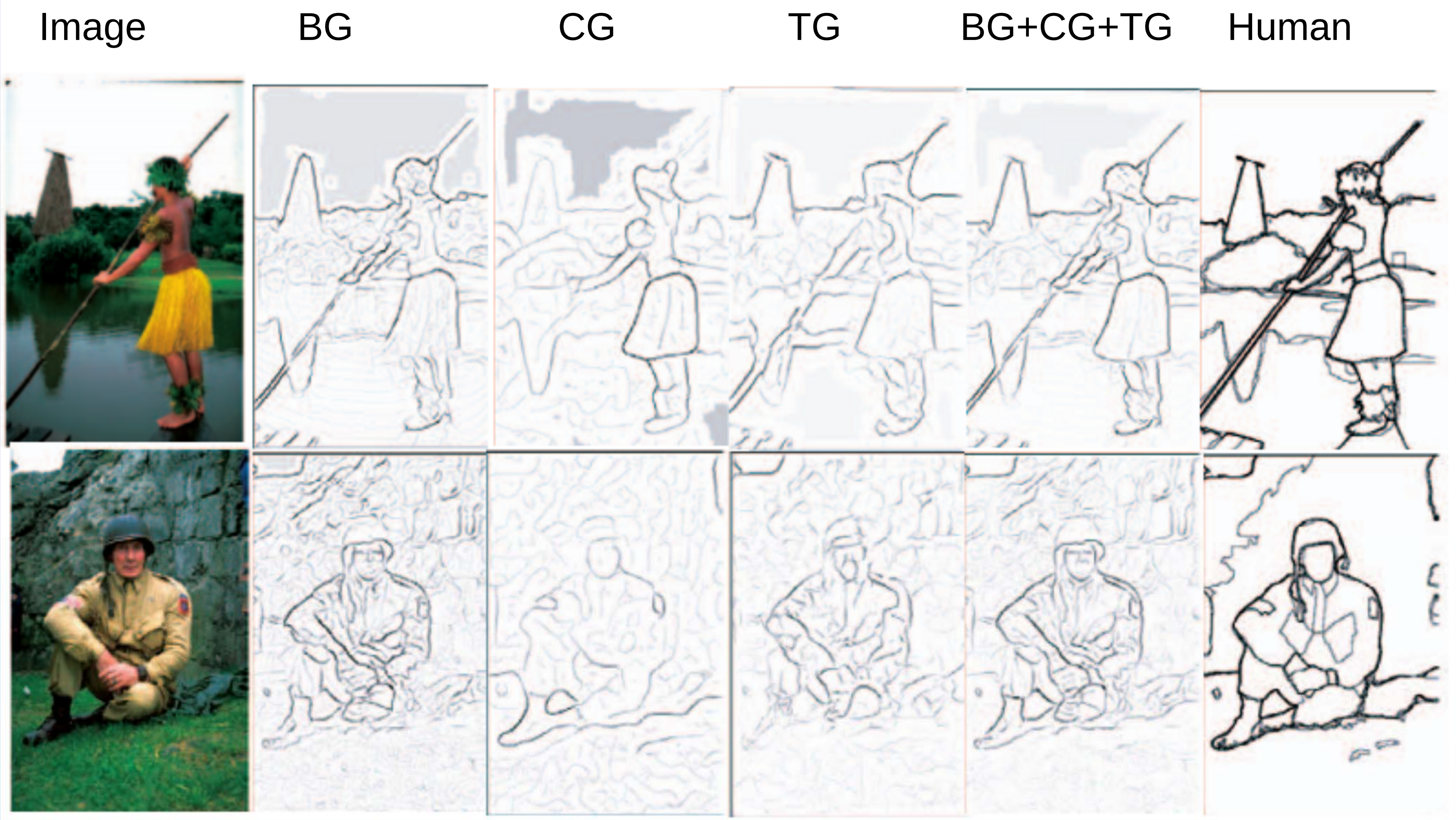
Precision-Recall Curves



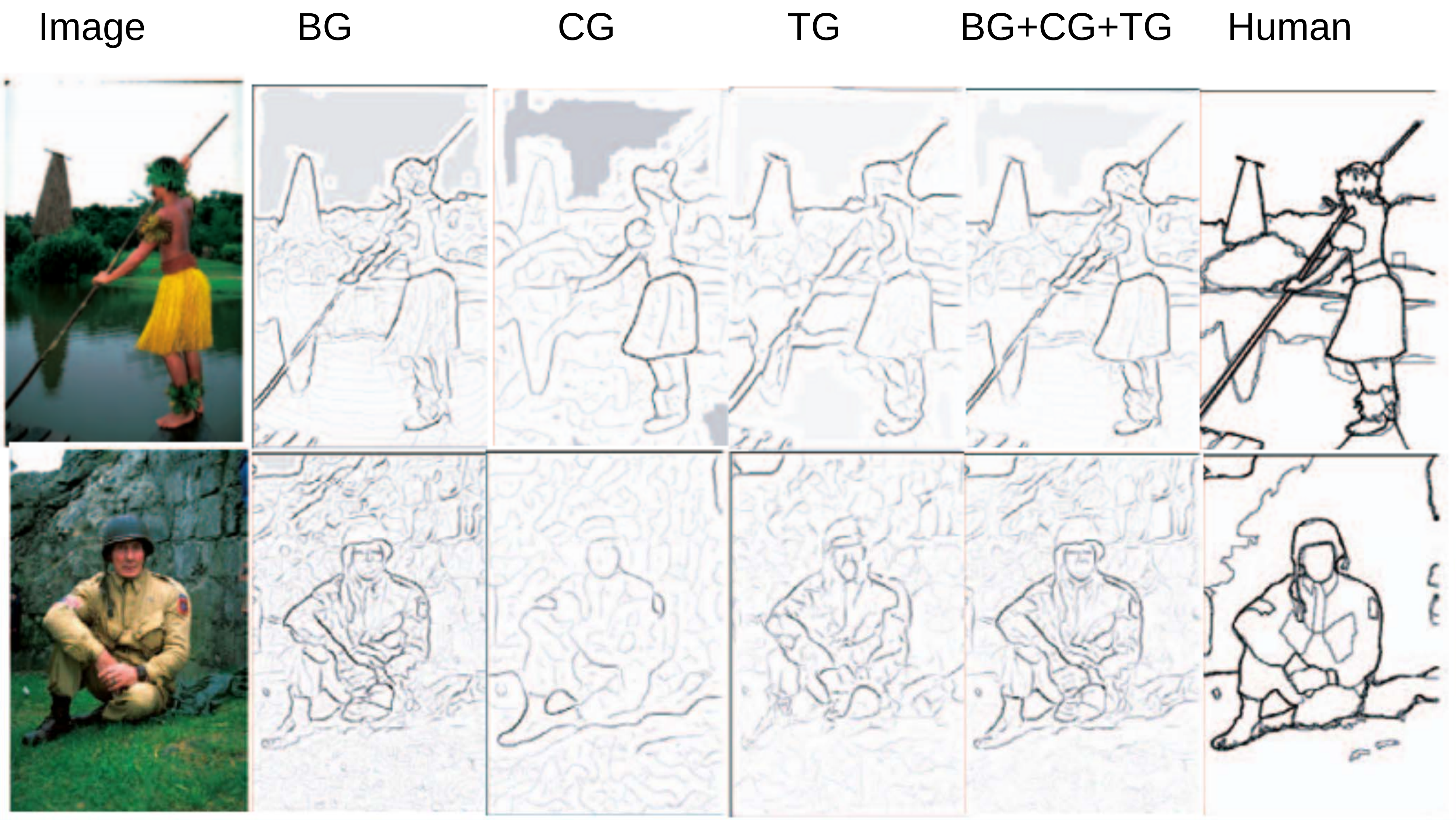
Two Decades of Local Boundary Detection



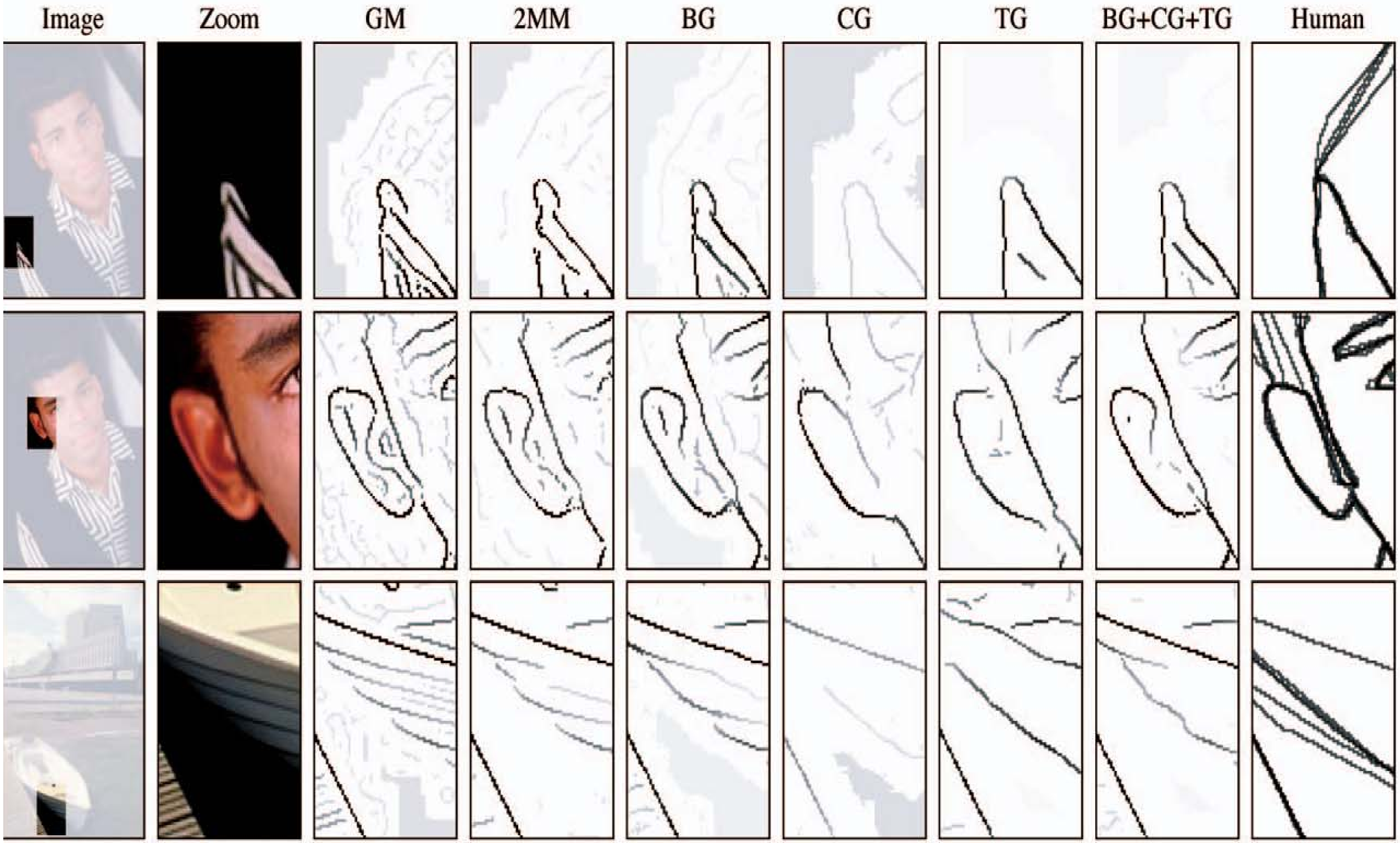
定性结果



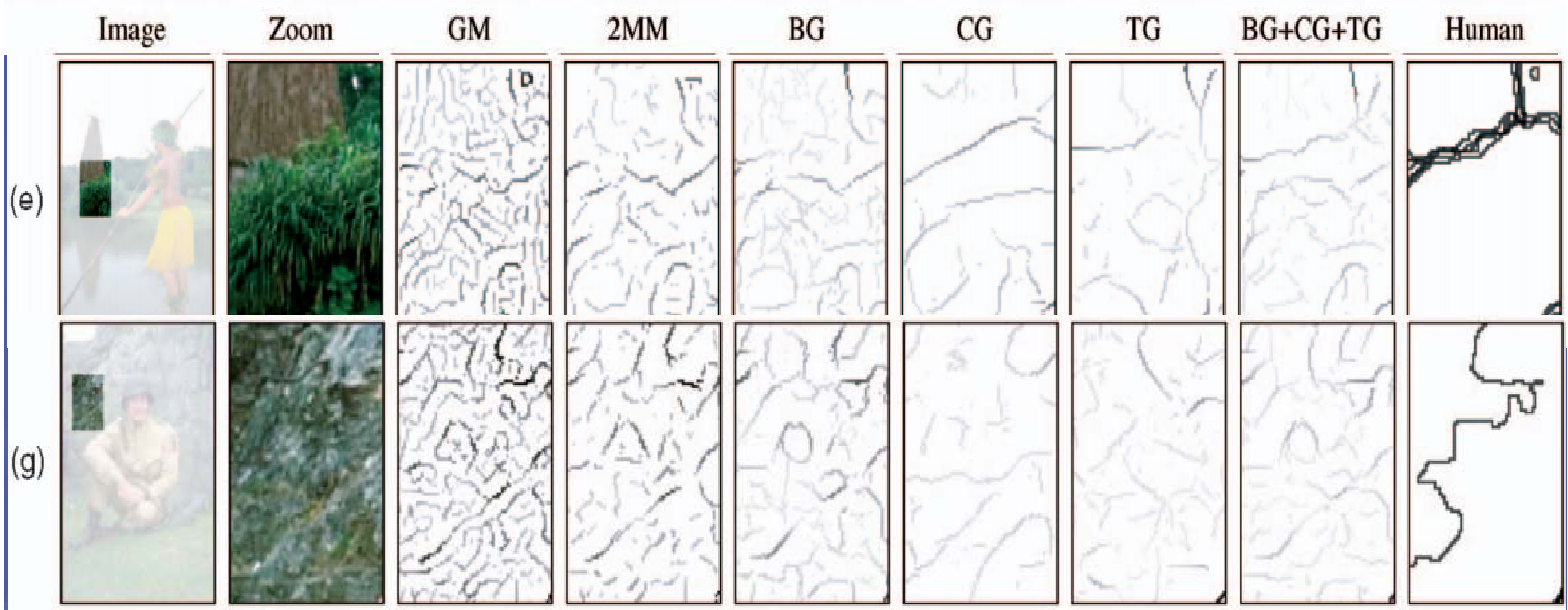
定性结果



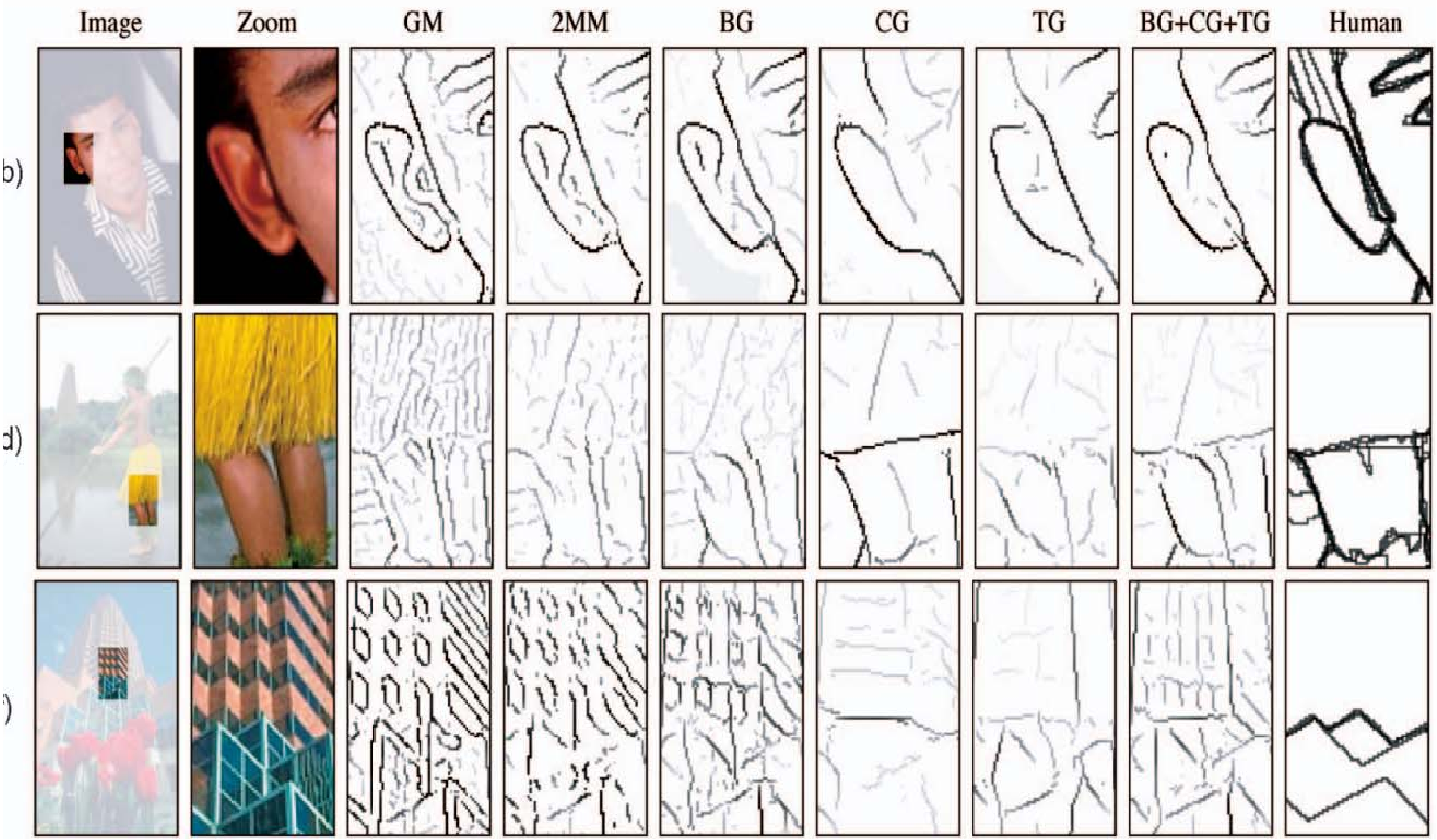
定性结果



定性结果



定性结果



Contour detection and hierarchical image segmentation

Arbelaez, P.
Maire, M
Fowlkes, C
Malik, J

Pablo Arbelaez

researcher
Computer Vision Group
UC Berkeley

Research field

modeling of visual perception
interaction between low and high-level information in vision
analysis of biological visual data

个人主页：http://www.cs.berkeley.edu/~arbelaez/

● gPb ——Contour Detection

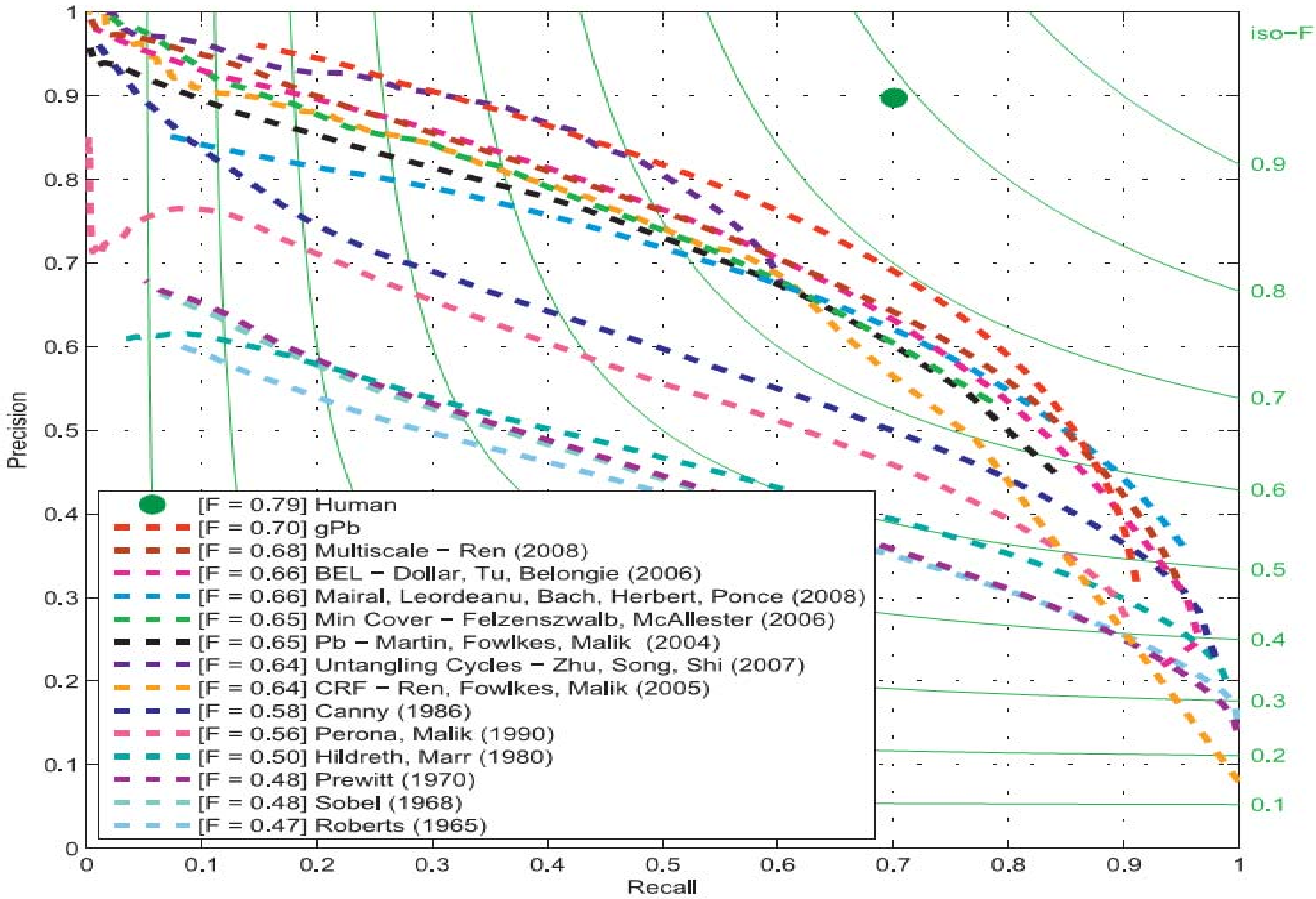
M. Maire, P. Arbela´ez, C. Fowlkes, and J. Malik, “Using Contours to Detect and Localize Junctions in Natural Images.”

Detection and segmentation happen in the same framework

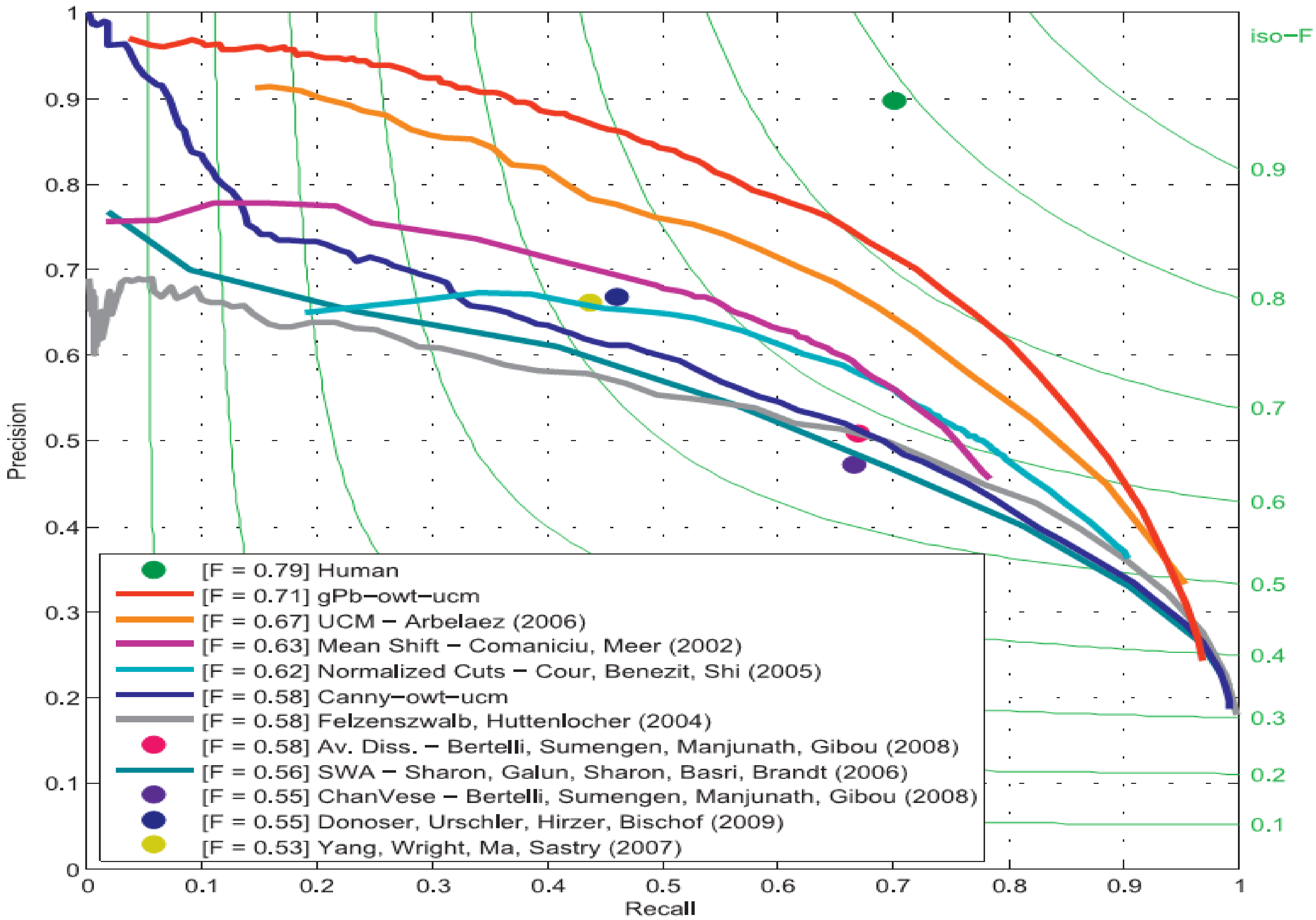
● gPb-owt-ucm ——segmentation algorithm

P. Arbela´ez, M. Maire, C. Fowlkes, and J. Malik, “From Contours to Regions: An Empirical Evaluation,”

Contour Detection (CVPR 2008)



Region Detection (CVPR 2009)



轮廓检测与图像分割

In general, contour detectors offer no guarantee that they will produce closed contours and hence do not necessarily provide a partition of the image into regions.

图像分割算法评价标准

Variation of Information

$$VI(S, S') = H(S) + H(S') - 2I(S, S'),$$

Rand Index

$$PRI(S, \{G_k\}) = \frac{1}{T} \sum_{i<j} [c_{ij} p_{ij} + (1 - c_{ij})(1 - p_{ij})],$$

Segmentation Covering

$$\mathcal{O}(R, R') = \frac{|R \cap R'|}{|R \cup R'|}$$

$$\mathcal{C}(S' \to S) = \frac{1}{N} \sum_{R \in S} |R| \cdot \max_{R' \in S'} \mathcal{O}(R, R'),$$

CONTOUR DETECTION

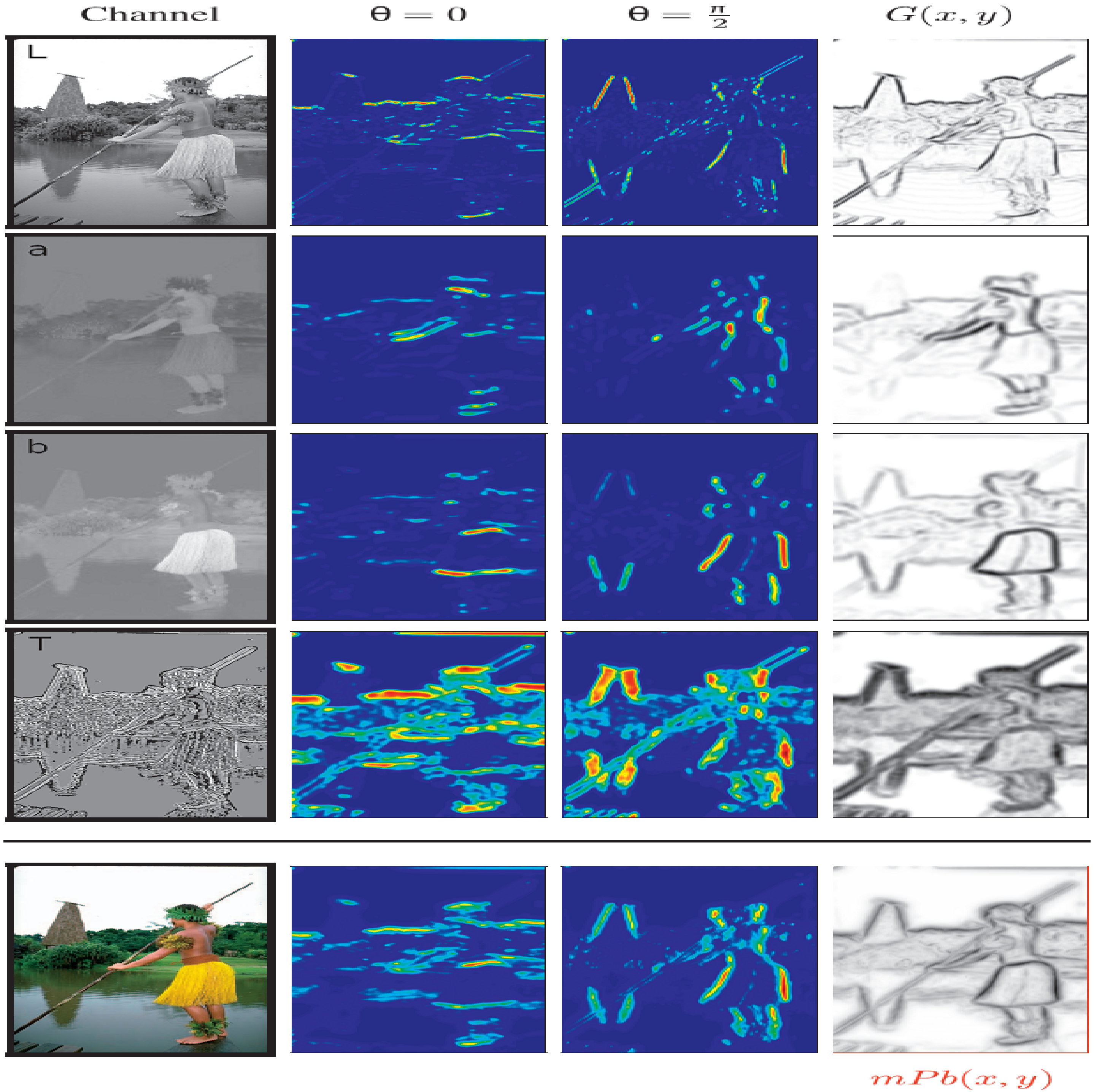
- Multiscale Cue Combination

$$G(x, y, \theta) \longrightarrow Pb(x, y, \theta)$$

In order to detect fine as well as coarse structures, we consider gradients at three scales: $\left[\frac{\sigma}{2}, \sigma, 2\sigma\right]$

$$mPb(x, y, \theta) = \sum_s \sum_i \alpha_{i,s} G_{i,\sigma(i,s)}(x, y, \theta),$$

$$mPb(x, y) = \max_\theta \{mPb(x, y, \theta)\}.$$


Channel
$\theta = 0$
$\theta = \frac{\pi}{2}$
$G(x, y)$
L
a
b
T
$mPb(x, y)$

CONTOUR DETECTION

- Globalization

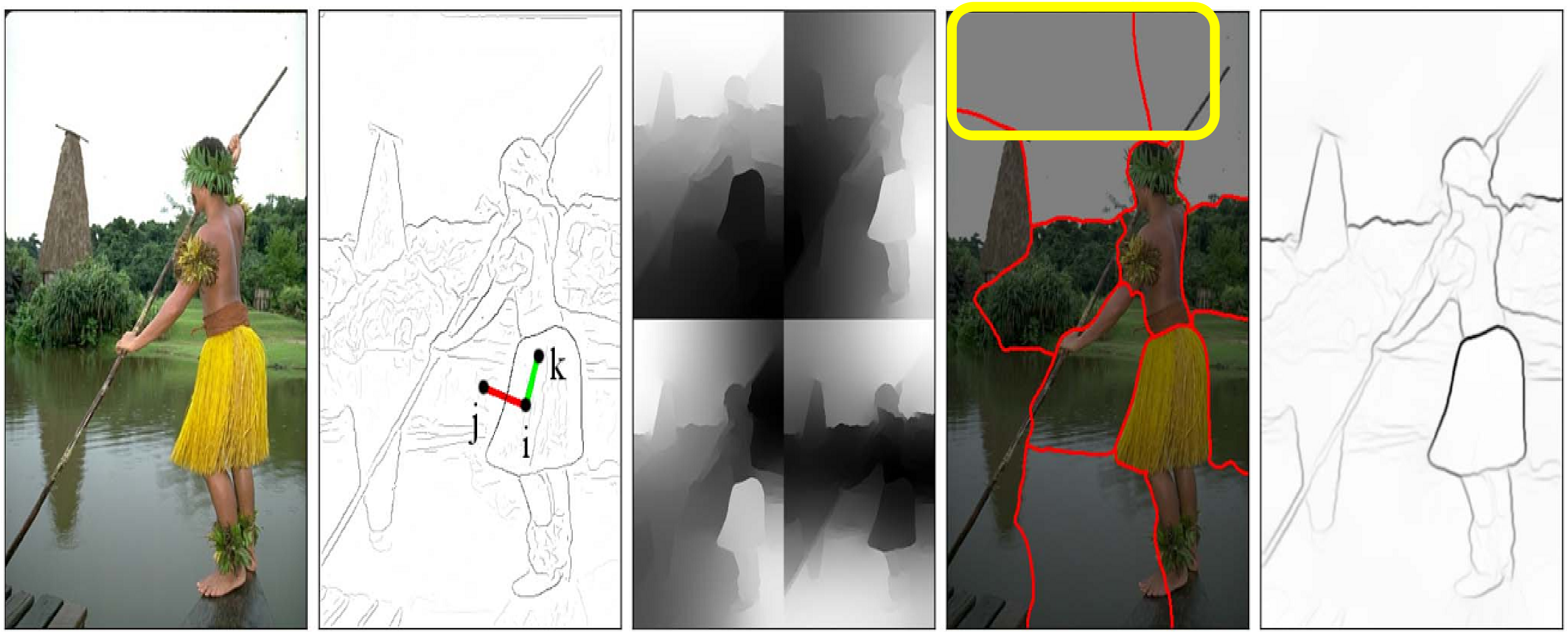
$$W_{ij} = \exp\left(-\max_{p \in \overline{ij}} \{mPb(p)\} / \rho \right)$$

$$D_{ii} = \sum_j W_{ij}$$

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

$$\{\mathbf{v_0}, \mathbf{v_1}, \ldots, \mathbf{v_n}\}$$

$$0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_n$$



(a) (b) (c) (d) (e)

CONTOUR DETECTION

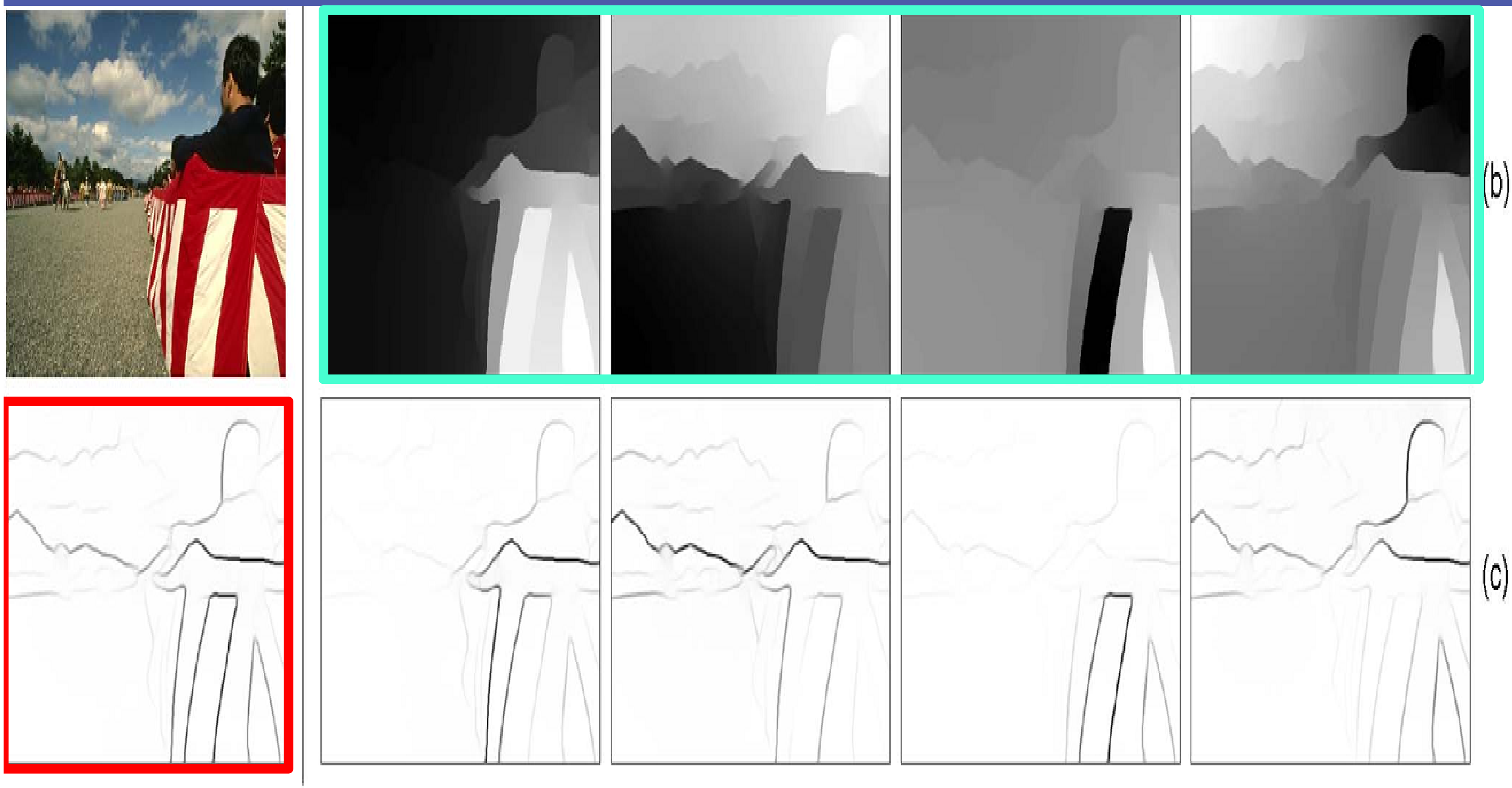
- Globalization

$$\{\mathbf{v_0}, \mathbf{v_1}, \ldots, \mathbf{v_n}\}$$

$$\{\nabla_\theta \mathbf{v_k}(x, y)\}$$

$$sPb(x, y, \theta) = \sum_{k=1}^{n} \frac{1}{\sqrt{\lambda_k}} \cdot \nabla_\theta \mathbf{v_k}(x, y)$$

$$\max_\theta\{\nabla_\theta \mathbf{v_k}(x,y)\}$$



$$sPb(x,y) = \max_\theta\{sPb(x,y,\theta)\}$$

CONTOUR DETECTION

- Multiscale Cue Combination

$$mPb(x,y,\theta) = \sum_s \sum_i \alpha_{i,s} G_{i,\sigma(i,s)}(x,y,\theta)$$

- Globalization

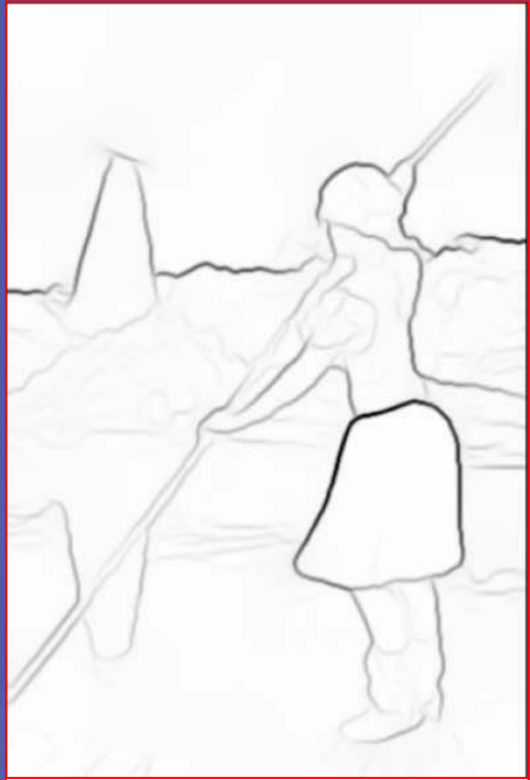
$$sPb(x,y,\theta) = \sum_{k=1}^{n} \frac{1}{\sqrt{\lambda_k}} \cdot \nabla_\theta \mathbf{v_k}(x,y)$$

globalized probability of boundary

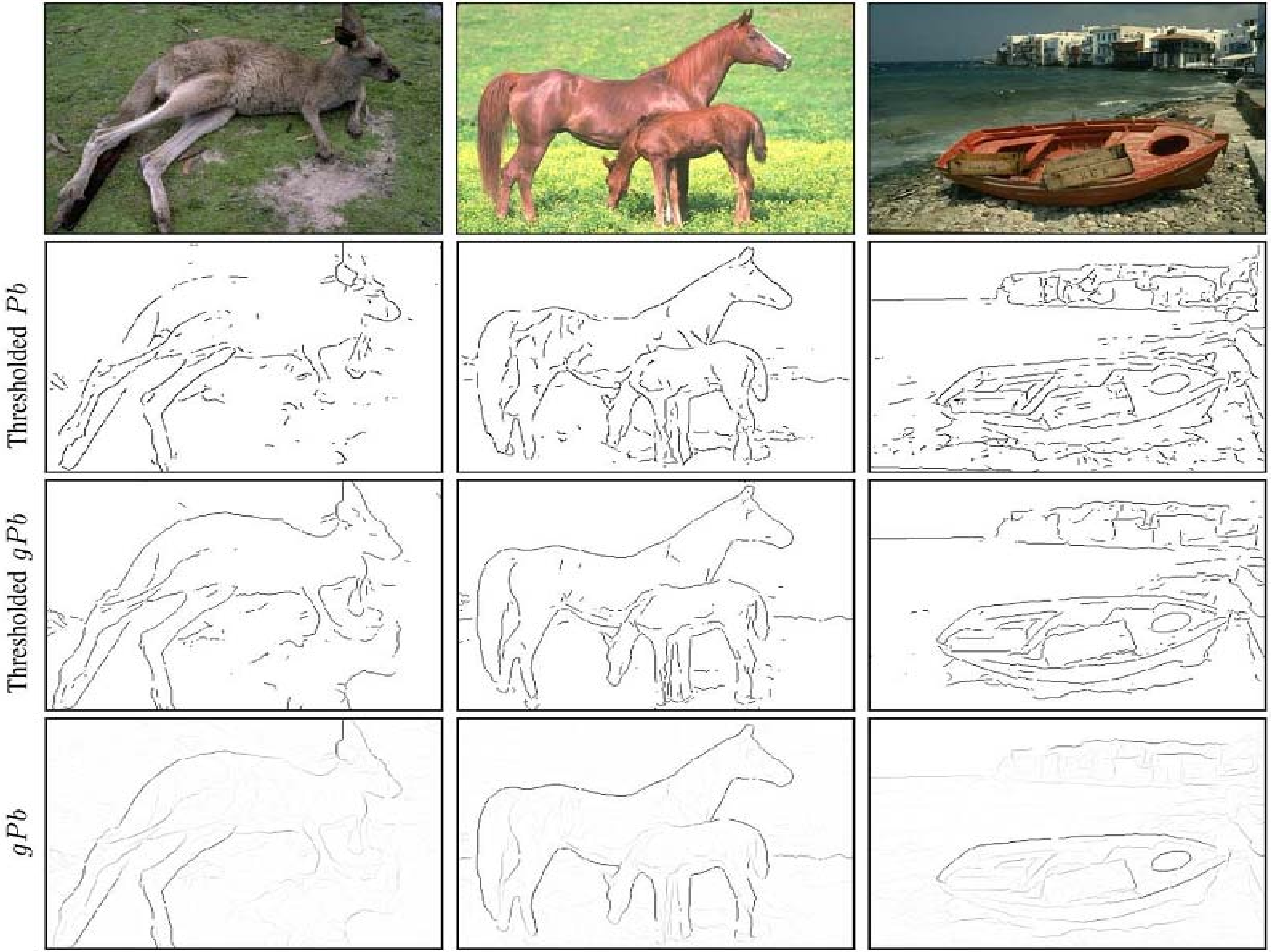
$$gPb(x,y,\theta) = \sum_s \sum_i \beta_{i,s} G_{i,\sigma(i,s)}(x,y,\theta) + \gamma \cdot sPb(x,y,\theta)$$

CONTOUR DETECTION

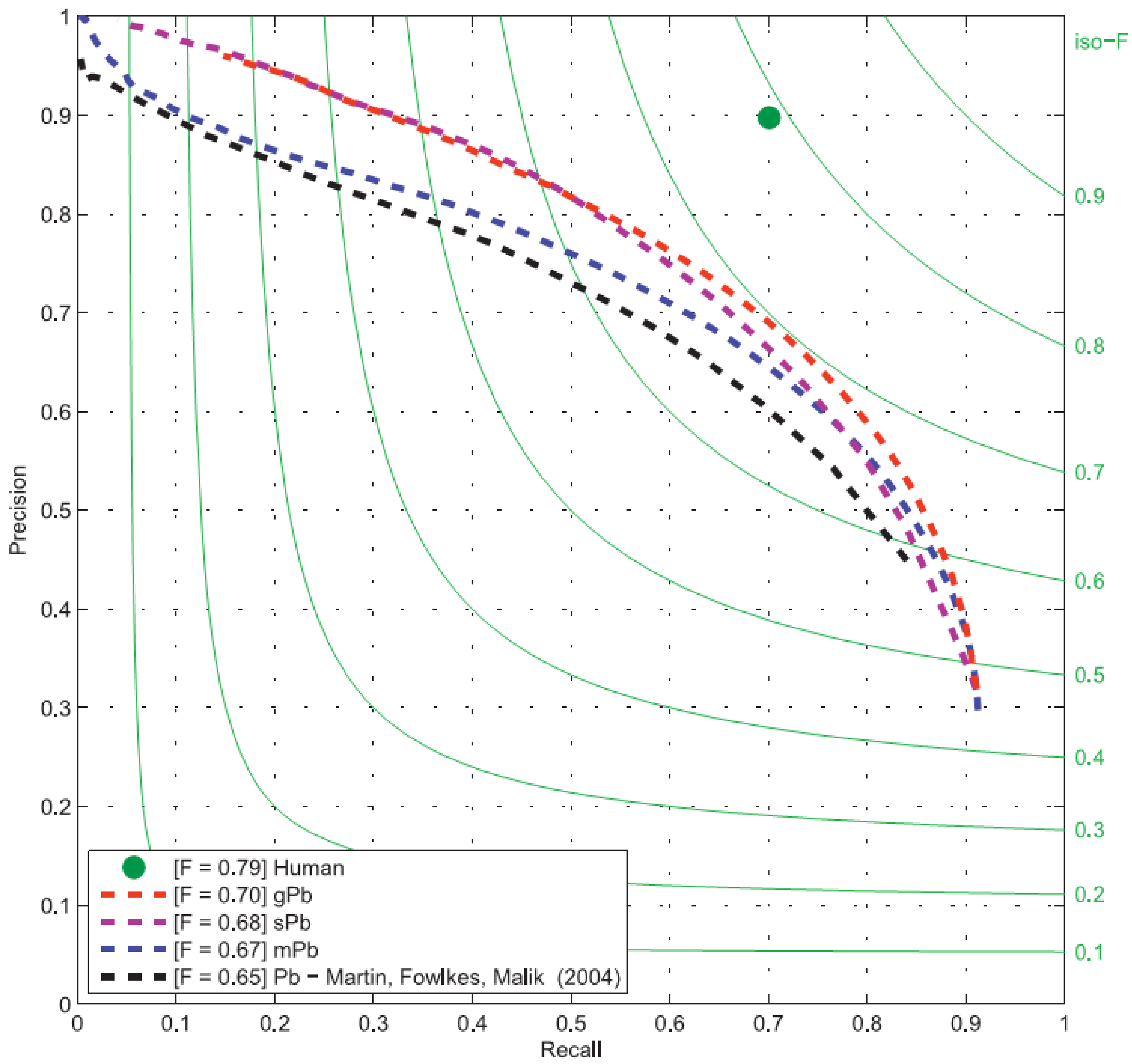
$mPb(x, y)$



$sPb(x, y)$



Thresholded Pb
Thresholded gPb
gPb



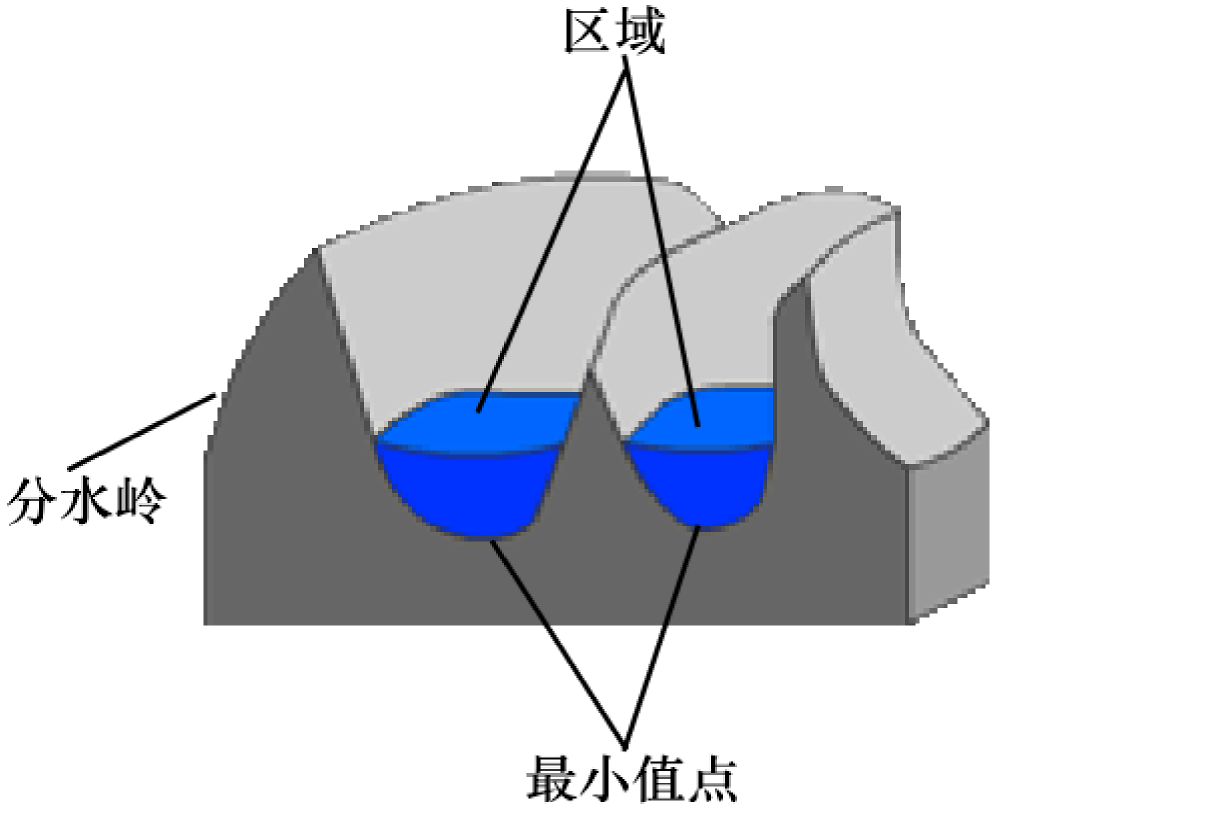
Precision
Recall
iso-F
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
[F = 0.79] Human
[F = 0.70] gPb
[F = 0.68] sPb
[F = 0.67] mPb
[F = 0.65] Pb – Martin, Fowlkes, Malik (2004)

SEGMENTATION

- Oriented Watershed Transform (OWT)
- Ultrametric Contour Map (UCM)

OWT-UCM

分水岭法（Meyer，1990）是一种基于拓扑理论的数学形态学的分割方法，其基本思想是把图像看作是测地学上的拓扑地貌，图像中每一点像素的灰度值表示该点的海拔高度，每一个局部极小值及其影响区域称为集水盆，而集水盆的边界则形成分水岭。



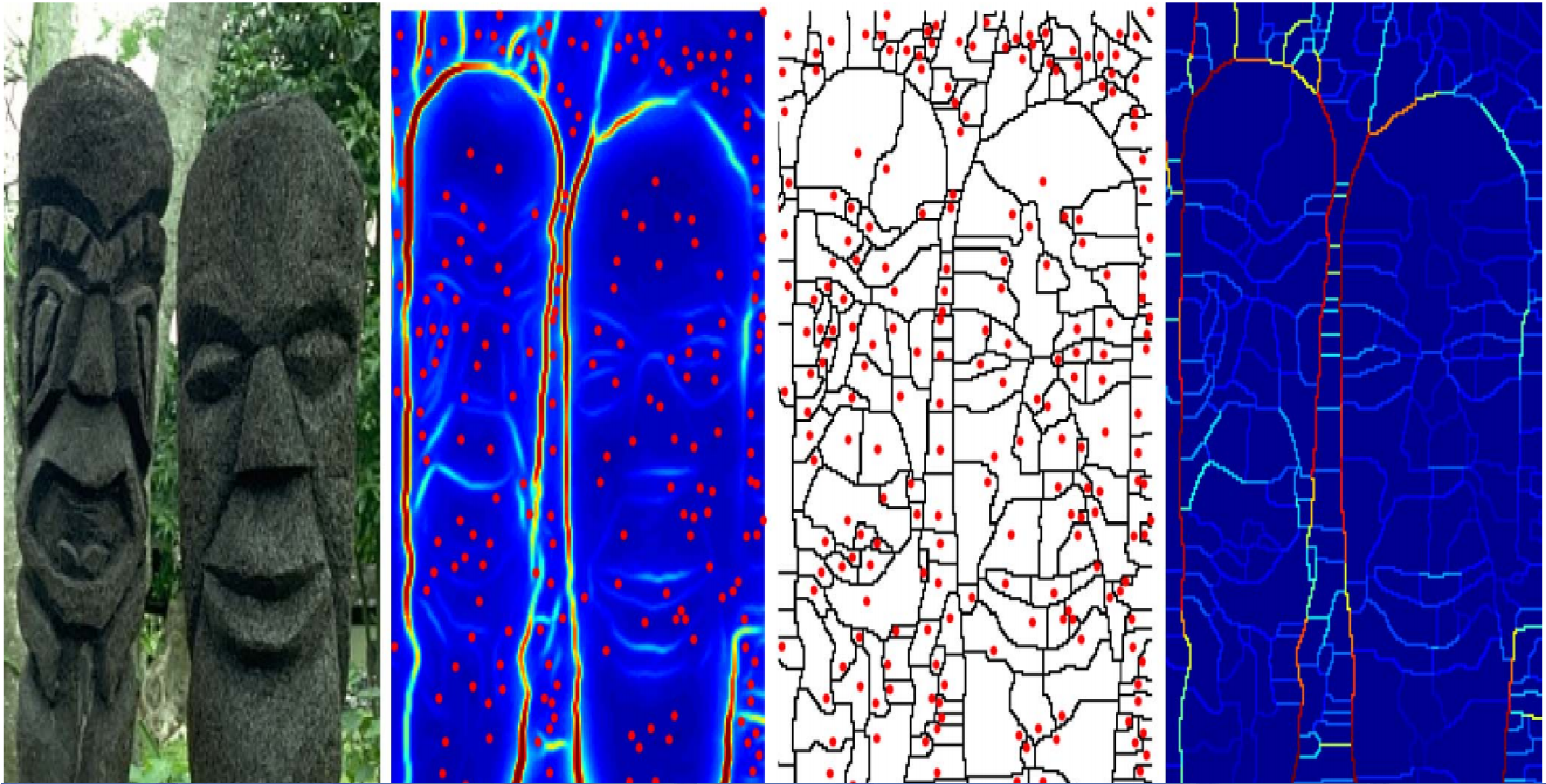
图像三维表示

Oriented Watershed Transform

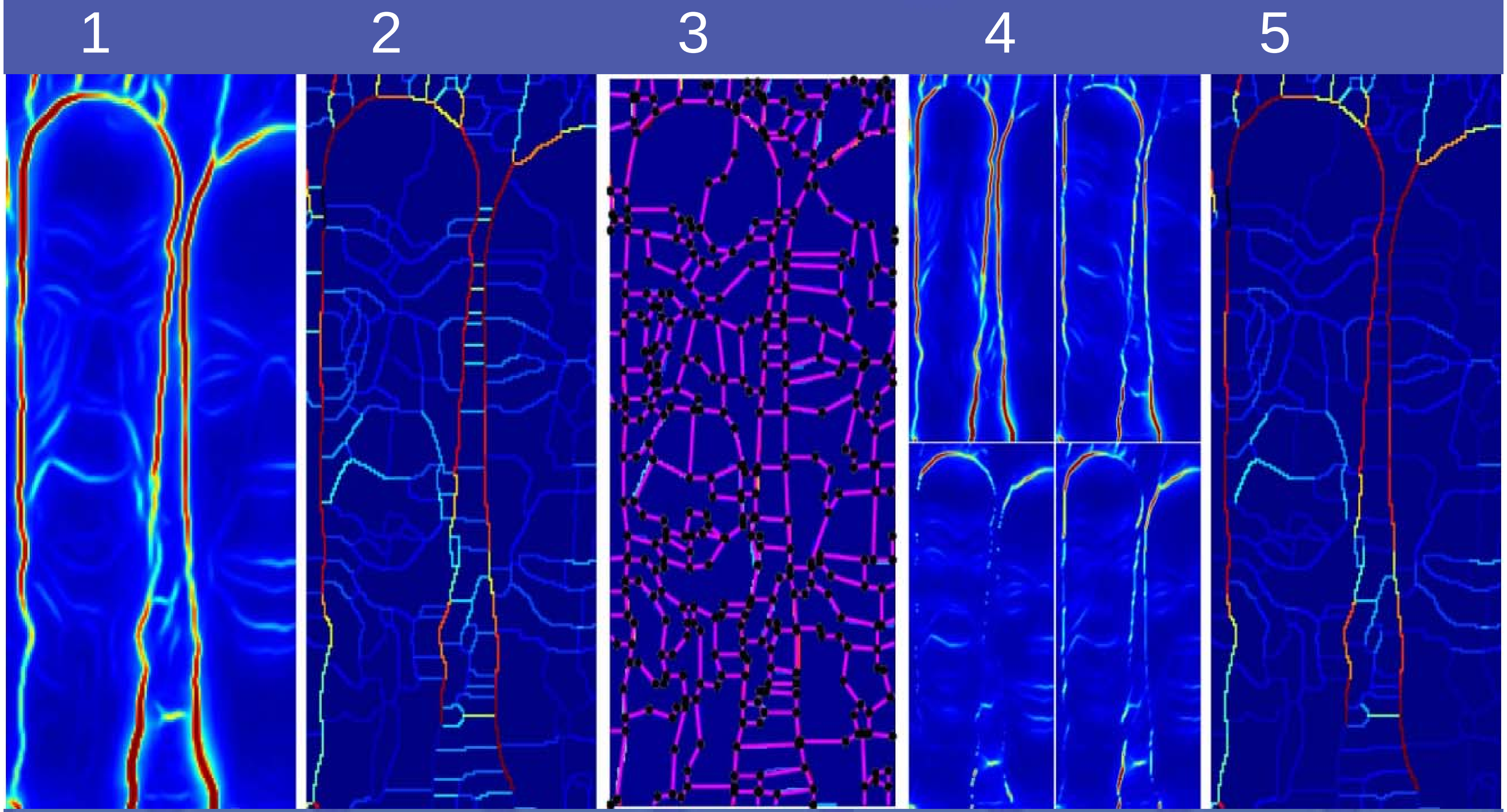
$E(x, y, \theta)$ predicts the probability of an image boundary at location (x, y) and orientation θ.

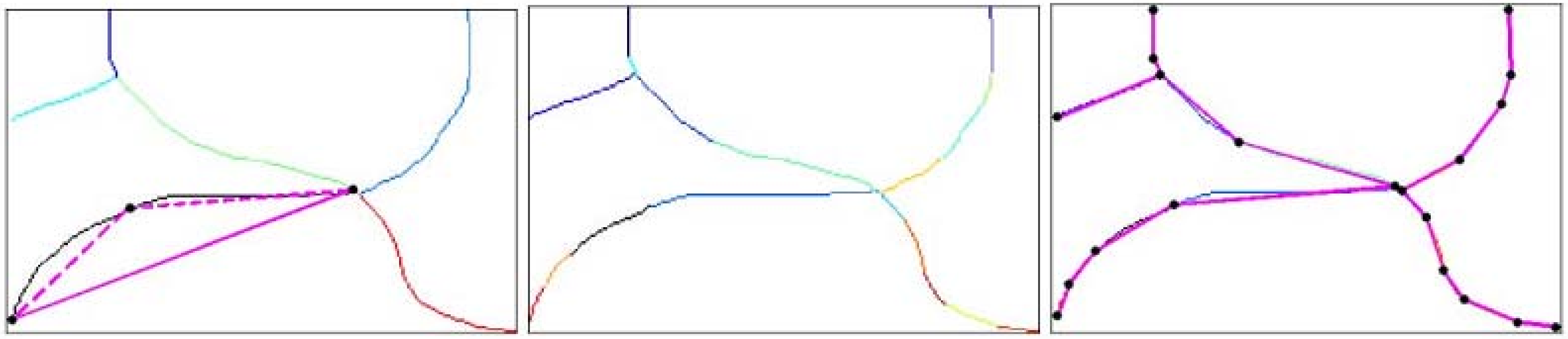
$$E(x, y) = \max_{\theta} E(x, y, \theta)$$

Watershed Transform



Oriented Watershed Transform (OWT)





(a) (b) (c)

Ultrametric Contour Map

$$G = (\mathcal{P}_0, \mathcal{K}_0, W(\mathcal{K}_0))$$

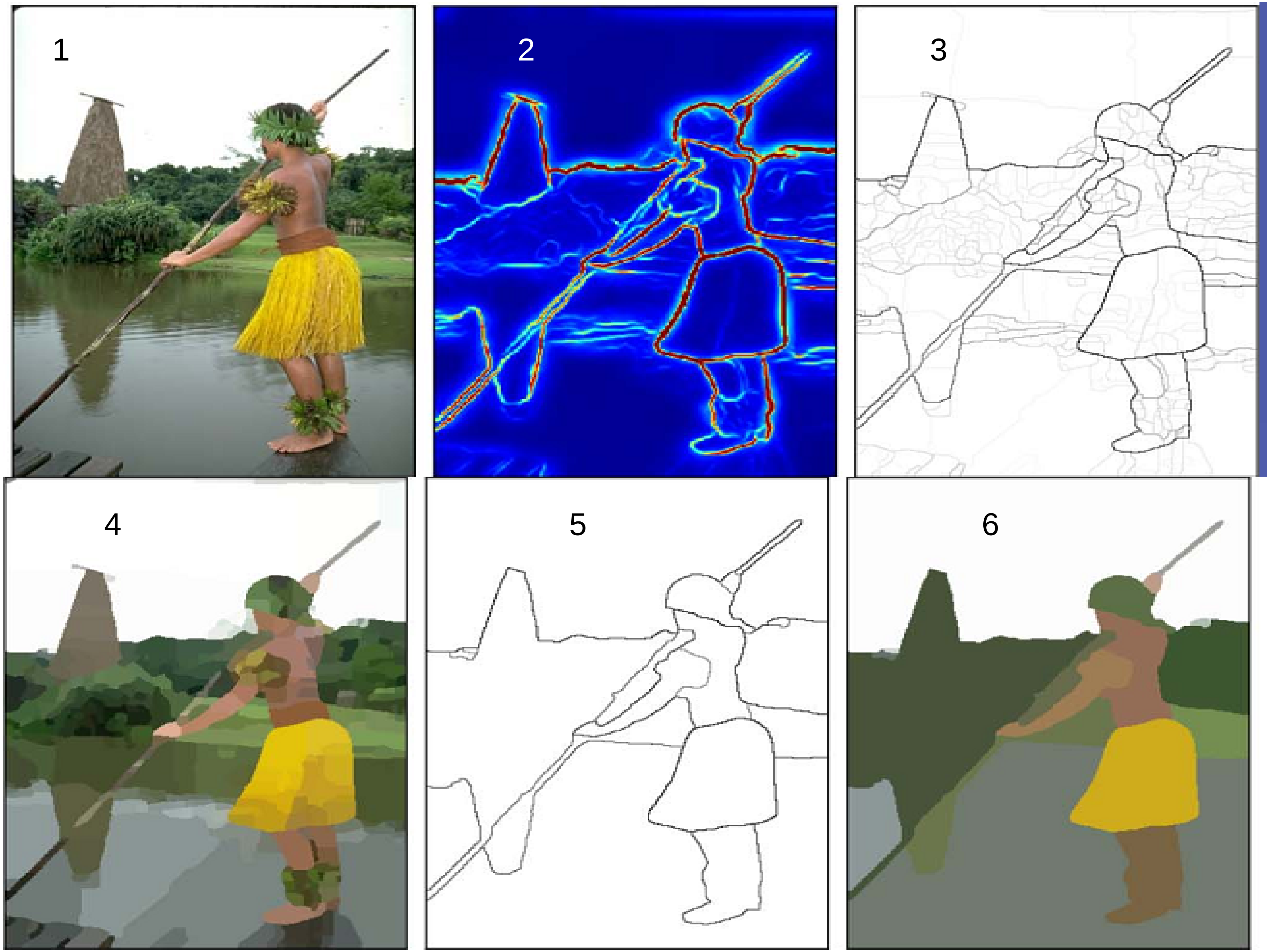
1. Select minimum weight contour:

$$C^* = \arg\min_{C \in \mathcal{K}_0} W(C).$$

2. Let $R_1, R_2 \in \mathcal{P}_0$ be the regions separated by C^*.
3. Set $R = R_1 \cup R_2$, and update:

$$\mathcal{P}_0 \leftarrow \mathcal{P}_0 \backslash \{R_1, R_2\} \cup \{R\} \quad \text{and} \quad \mathcal{K}_0 \leftarrow \mathcal{K}_0 \backslash \{C^*\}.$$

4. Stop if $\mathcal{K}_0$ is empty.
   Otherwise, update weights $W(\mathcal{K}_0)$ and repeat.


1
2
3
4
5
6

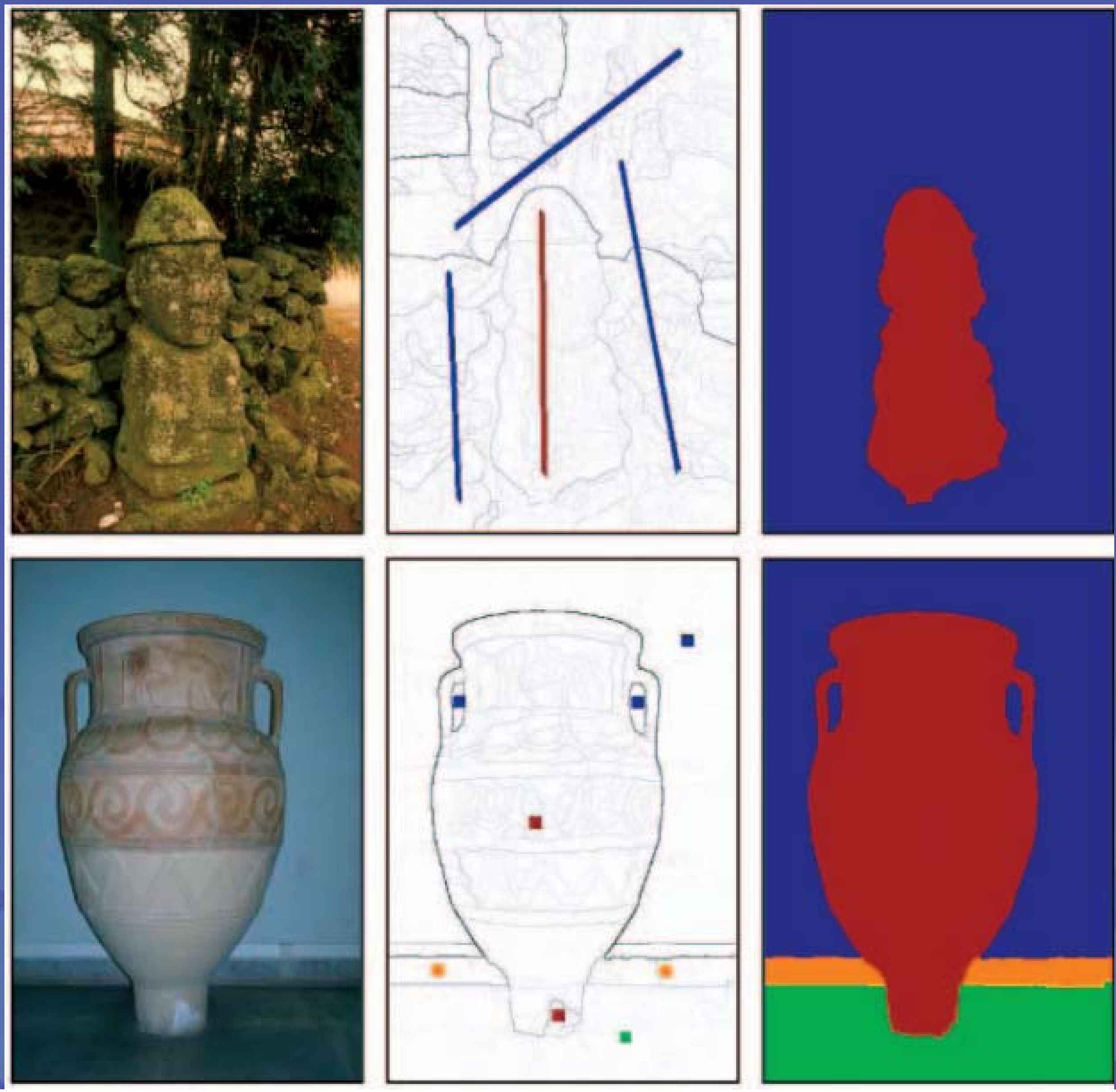
Boundary Benchmarks

	BSDS300			BSDS500		
	ODS	OIS	AP	ODS	OIS	AP
Human	0.79	0.79	–	0.80	0.80	–
gPb-owt-ucm	**0.71**	**0.74**	**0.73**	**0.73**	**0.76**	**0.73**
[34] Mean Shift	0.63	0.66	0.54	0.64	0.68	0.56
[33] NCuts	0.62	0.66	0.43	0.64	0.68	0.45
Canny-owt-ucm	0.58	0.63	0.58	0.60	0.64	0.58
[32] Felz-Hutt	0.58	0.62	0.53	0.61	0.64	0.56
[31] SWA	0.56	0.59	0.54	–	–	–
Quad-Tree	0.37	0.39	0.26	0.38	0.39	0.26
gPb	0.70	0.72	0.66	0.71	0.74	0.65
Canny	0.58	0.62	0.58	0.60	0.63	0.58

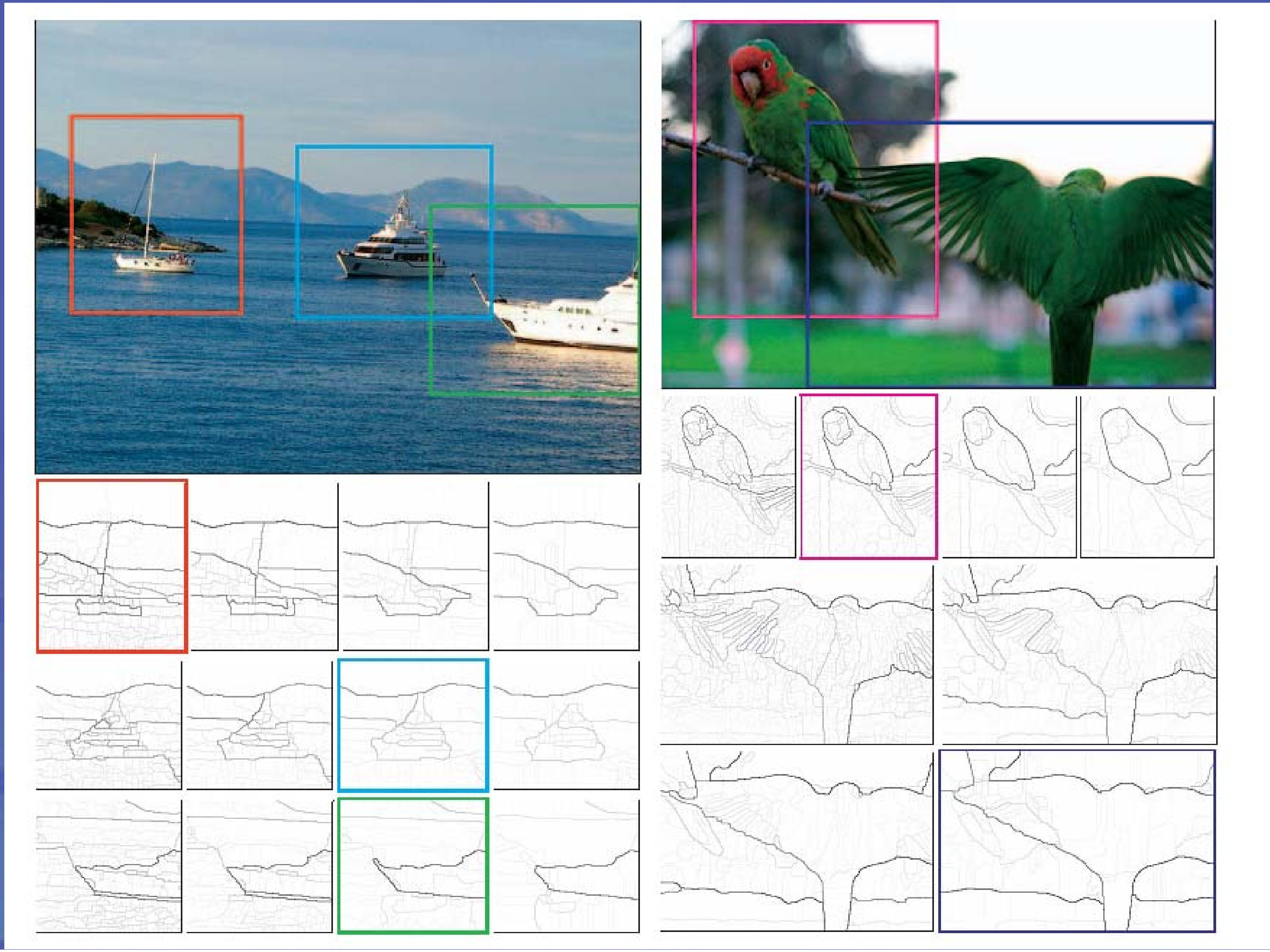
Region Benchmarks

	BSDS300						
	Covering			PRI		VI	
	ODS	OIS	Best	ODS	OIS	ODS	OIS
Human	0.73	0.73	–	0.87	0.87	1.16	1.16
gPb-owt-ucm	**0.59**	**0.65**	**0.75**	**0.81**	**0.85**	**1.65**	**1.47**
[34] Mean Shift	0.54	0.58	0.66	0.78	0.80	1.83	1.63
[32] Felz-Hutt	0.51	0.58	0.68	0.77	0.82	2.15	1.79
Canny-owt-ucm	0.48	0.56	0.66	0.77	0.82	2.11	1.81
[33] NCuts	0.44	0.53	0.66	0.75	0.79	2.18	1.84
[31] SWA	0.47	0.55	0.66	0.75	0.80	2.06	1.75
[29] Total Var.	0.57	–	–	0.78	–	1.81	–
[70] T+B Encode	0.54	–	–	0.78	–	1.86	–
[30] Av. Diss.	0.47	–	–	0.76	–	2.62	–
[30] ChanVese	0.49	–	–	0.75	–	2.54	–
Quad-Tree	0.33	0.39	0.47	0.71	0.75	2.34	2.22

INTERACTIVE SEGMENTATION



MULTISCALE FOR OBJECT ANALYSIS



GrabCut
Interactive Foreground Extraction using Iterated Graph Cuts

Carsten Rother
Vladimir Kolmogorov
Andrew Blake

Microsoft Research Cambridge-UK

Photomontage

SIGGRAPH2004

What GrabCut does

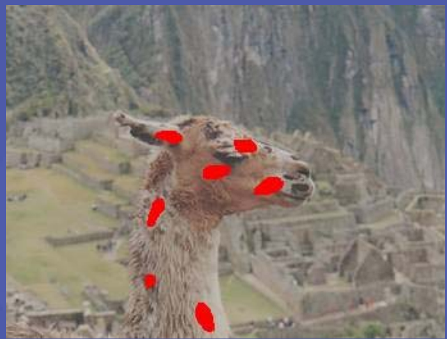
SIGGRAPH2004

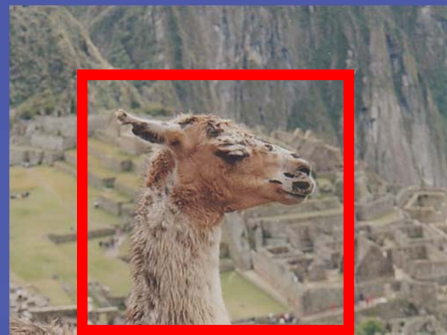
Magic Wand
(198?)

Intelligent Scissors
Mortensen and Barrett (1995)

GrabCut

User Input





Result

Regions

Boundary

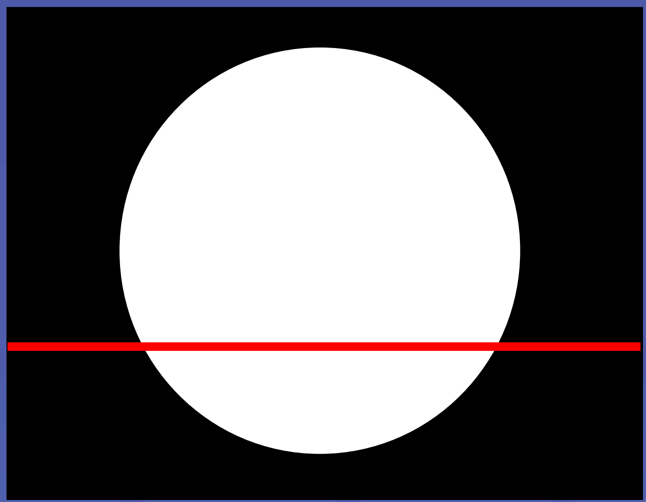
Regions & Boundary

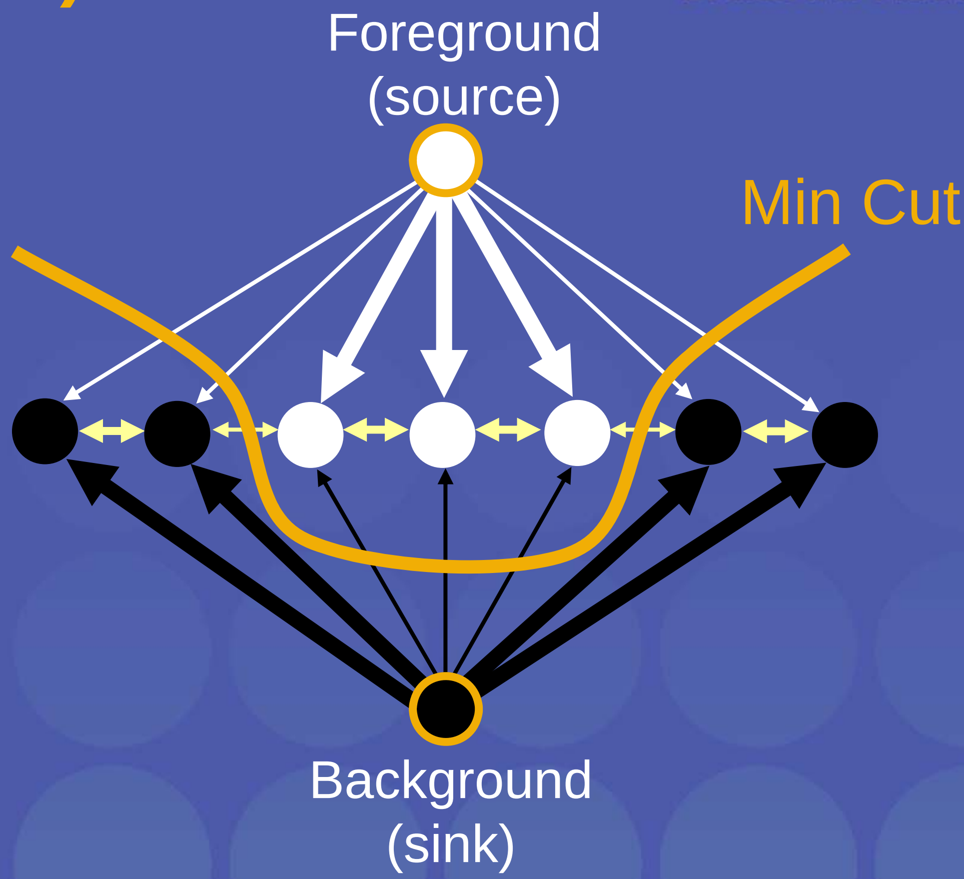
Graph Cuts

Boykov and Jolly (2001)

SIGGRAPH2004

Image





Cut: separating source and sink; Energy: collection of edges

Min Cut: Global minimal enegry in polynomial time

Graph Cuts

- Image segmentation

Image $\mathbf{z} = (z_1, \ldots, z_n, \ldots, z_N)$

Segmentation $\underline{\alpha} = (\alpha_1, \ldots, \alpha_N)$

$$\underline{\theta} = \{h(z;\alpha), \alpha = 0, 1\}$$

Graph Cuts

- Segmentation by energy minimisation

$$\mathbf{E}(\underline{\alpha},\underline{\theta},\mathbf{z}) = U(\underline{\alpha},\underline{\theta},\mathbf{z}) + V(\underline{\alpha},\mathbf{z})$$

$$U(\underline{\alpha},\underline{\theta},\mathbf{z}) = \sum_{n} -\log h(z_n;\alpha_n)$$

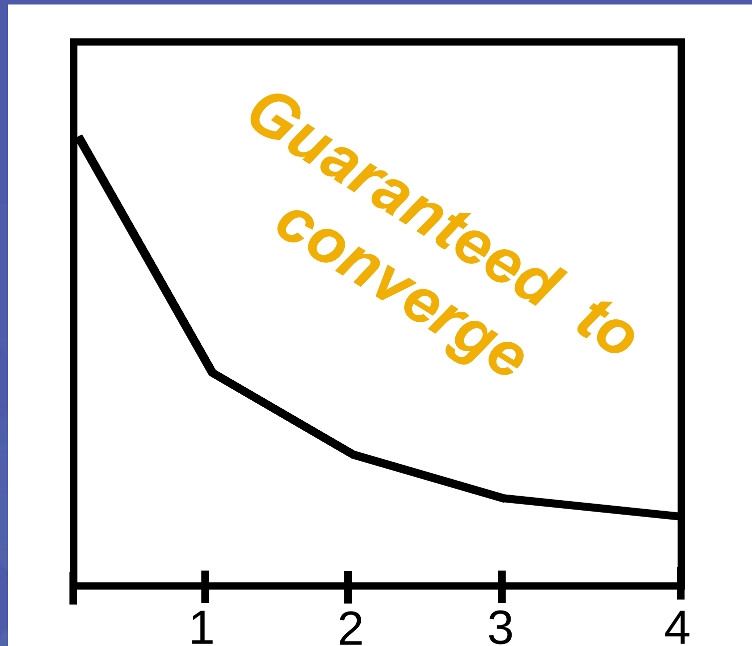
$$V(\underline{\alpha},\mathbf{z}) = \gamma \sum_{(m,n)\in\mathbf{C}} dis(m,n)^{-1}\,[\alpha_n \neq \alpha_m] \exp -\beta (z_m - z_n)$$

$$\beta = \left(2\left\langle (z_m - z_n)^2 \right\rangle\right)^{-1}$$

$$\underline{\hat{\alpha}} = \arg\min_{\underline{\alpha}} \mathbf{E}(\underline{\alpha},\underline{\theta})$$

Iterated Graph Cuts

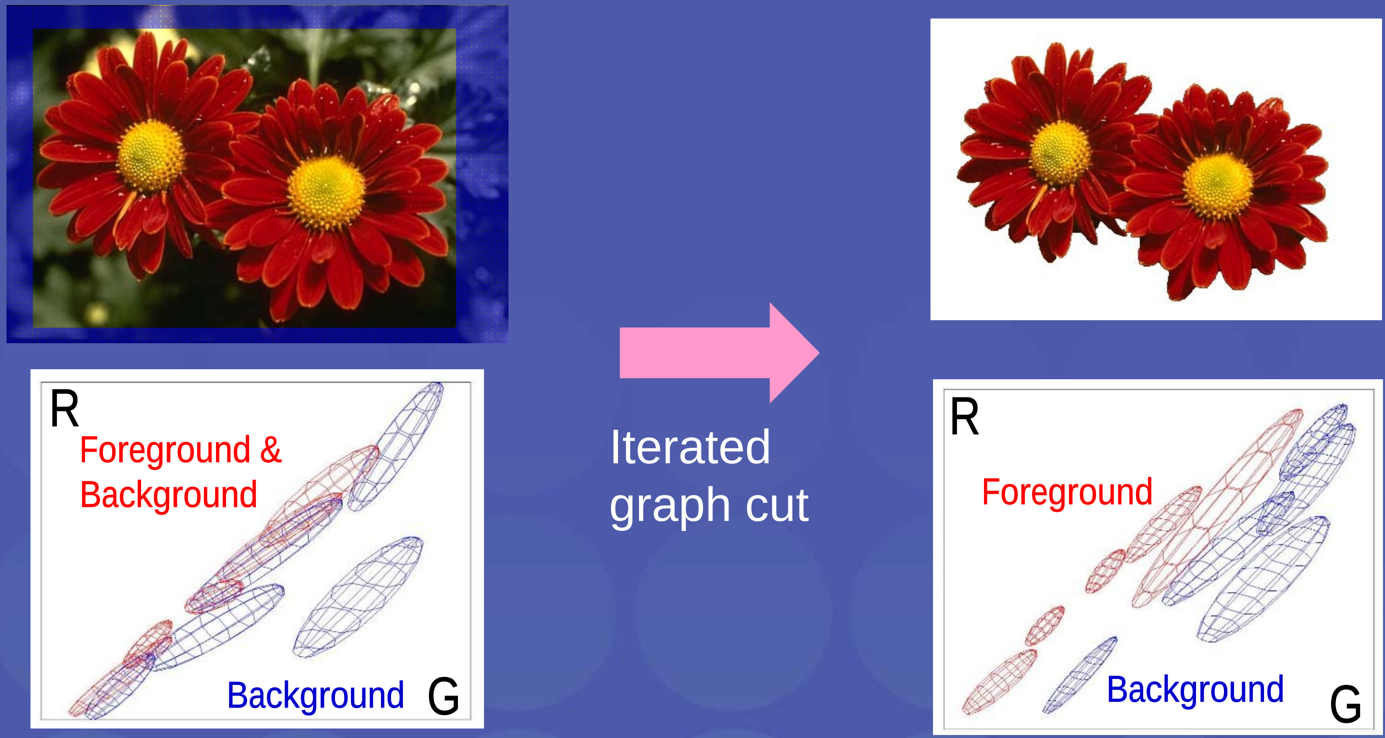
SIGGRAPH2004



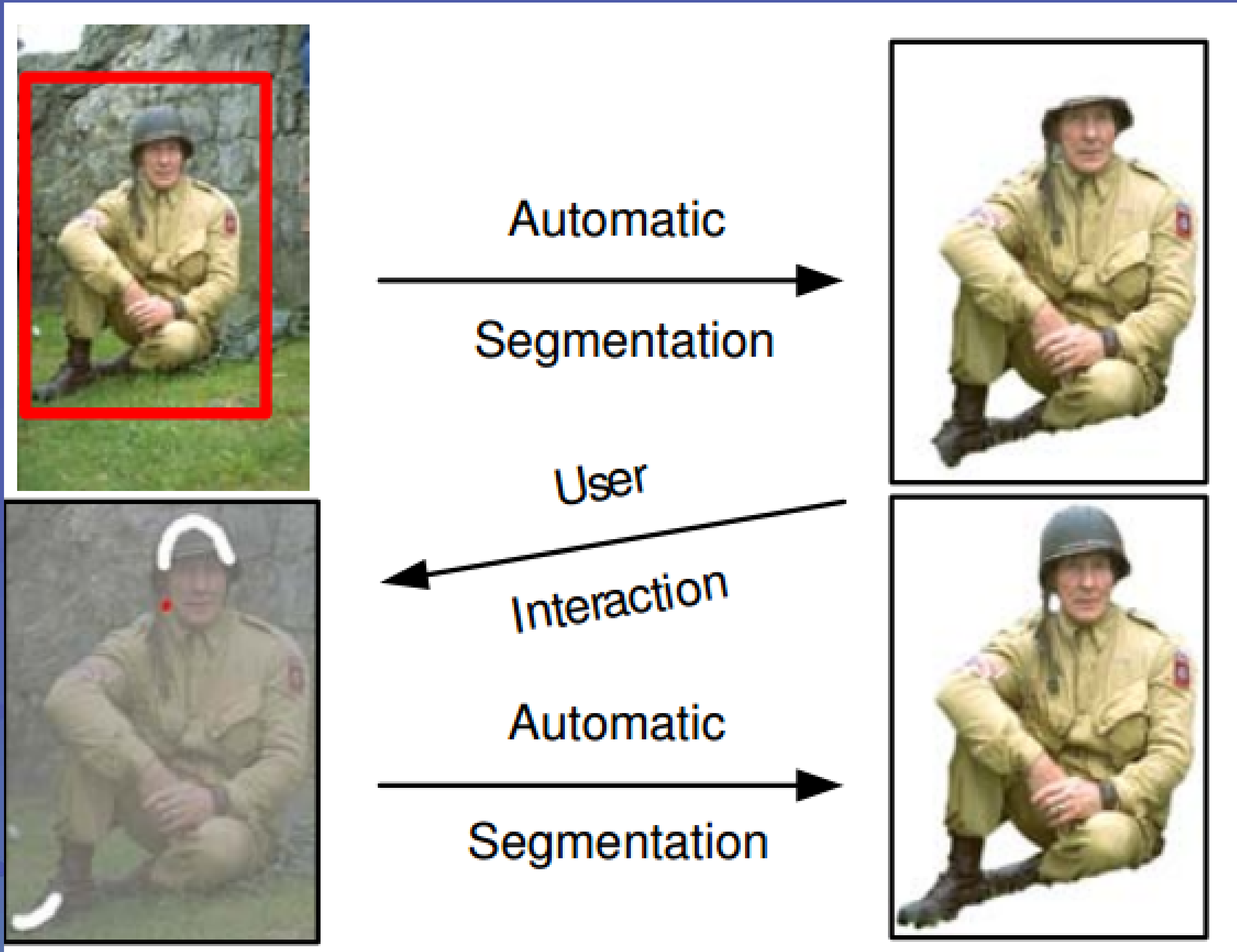
Result

Energy after each Iteration

Colour Model



Gaussian Mixture Model (typically 5-8 components)


Automatic
Segmentation
User
Interaction
Automatic
Segmentation

004

Iterative image segmentation in GrabCut

Initialisation

- User initialises trimap T by supplying only T_B. The foreground is set to $T_F = \emptyset$; $T_U = \overline{T}_B$, complement of the background.
- Initialise $\alpha_n = 0$ for $n \in T_B$ and $\alpha_n = 1$ for $n \in T_U$.
- Background and foreground GMMs initialised from sets $\alpha_n = 0$ and $\alpha_n = 1$ respectively.

Iterative minimisation

1. *Assign GMM components to pixels:* for each n in T_U,
$$k_n := \arg\min_{k_n} D_n(\alpha_n, k_n, \theta, z_n).$$
2. *Learn GMM parameters* from data $\mathbf{z}$:
$$\underline{\theta} := \arg\min_{\underline{\theta}} U(\underline{\alpha}, \mathbf{k}, \underline{\theta}, \mathbf{z})$$
3. *Estimate segmentation:* use min cut to solve:
$$\min_{\{\alpha_n :\, n \in T_U\}} \min_{\mathbf{k}} \mathbf{E}(\underline{\alpha}, \mathbf{k}, \underline{\theta}, \mathbf{z}).$$
4. Repeat from step 1, until convergence.
5. Apply border matting (section 4).

User editing

- *Edit:* fix some pixels either to $\alpha_n = 0$ (background brush) or $\alpha_n = 1$ (foreground brush); update trimap T accordingly. Perform step 3 above, just once.
- *Refine operation:* [optional] perform entire iterative minimisation algorithm.

Moderately straightforward examples

SIGGRAPH2004

… GrabCut completes automatically

Difficult Examples

SIGGRAPH2004

Camouflage & Low Contrast

Fine structure

No telepathy

Initial Rectangle

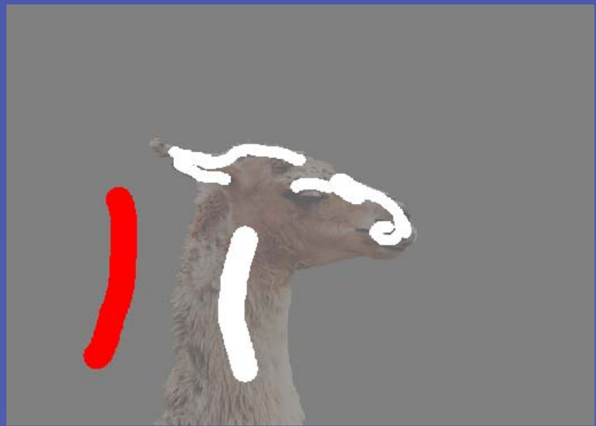
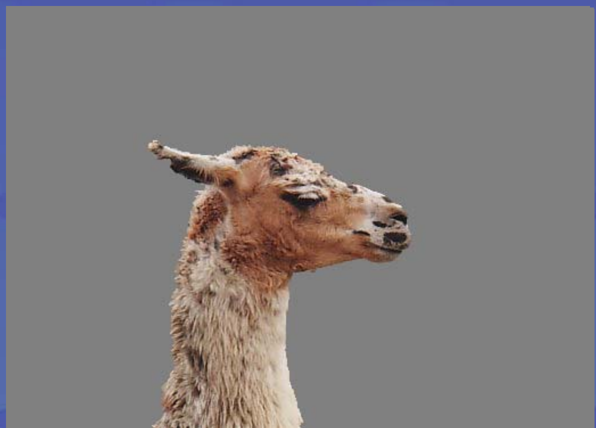
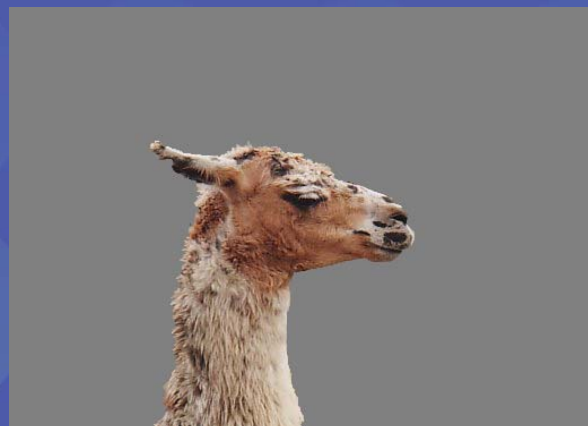
Initial Result

Evaluation – Labelled Database

Available online: http://research.microsoft.com/vision/cambridge/segmentation/

Comparison

SIGGRAPH2004

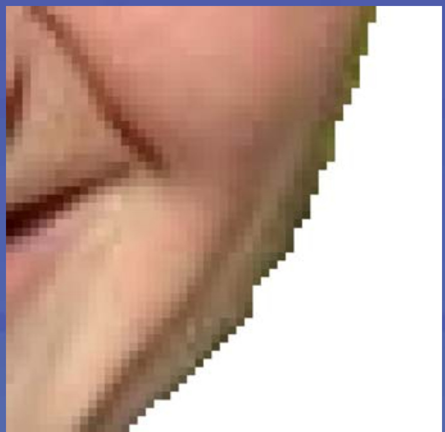
	Boykov and Jolly (2001)	GrabCut
User Input		
Result		
	Error Rate: 0.72%	Error Rate: 0.72%

Comparison

	LazySnapping	GrabCut
Smart Initialisation		Rectangle or Lasso
Editing	Boykov-Jolly Brushing (global) Boundary editing (local)	Boykov-Jolly Brushing (global) Boundary editing (local)
Speed	Interactive, due to segmentation into regions	Interactive, due to multiple image resolution
Pre-processing (Image loading)	Segmentation into regions	

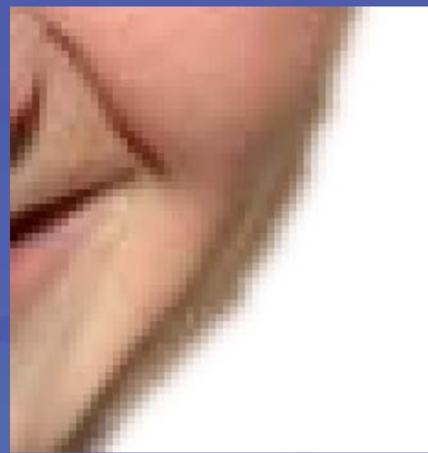
Li et al. (2004), **LazySnapping**

Border Matting



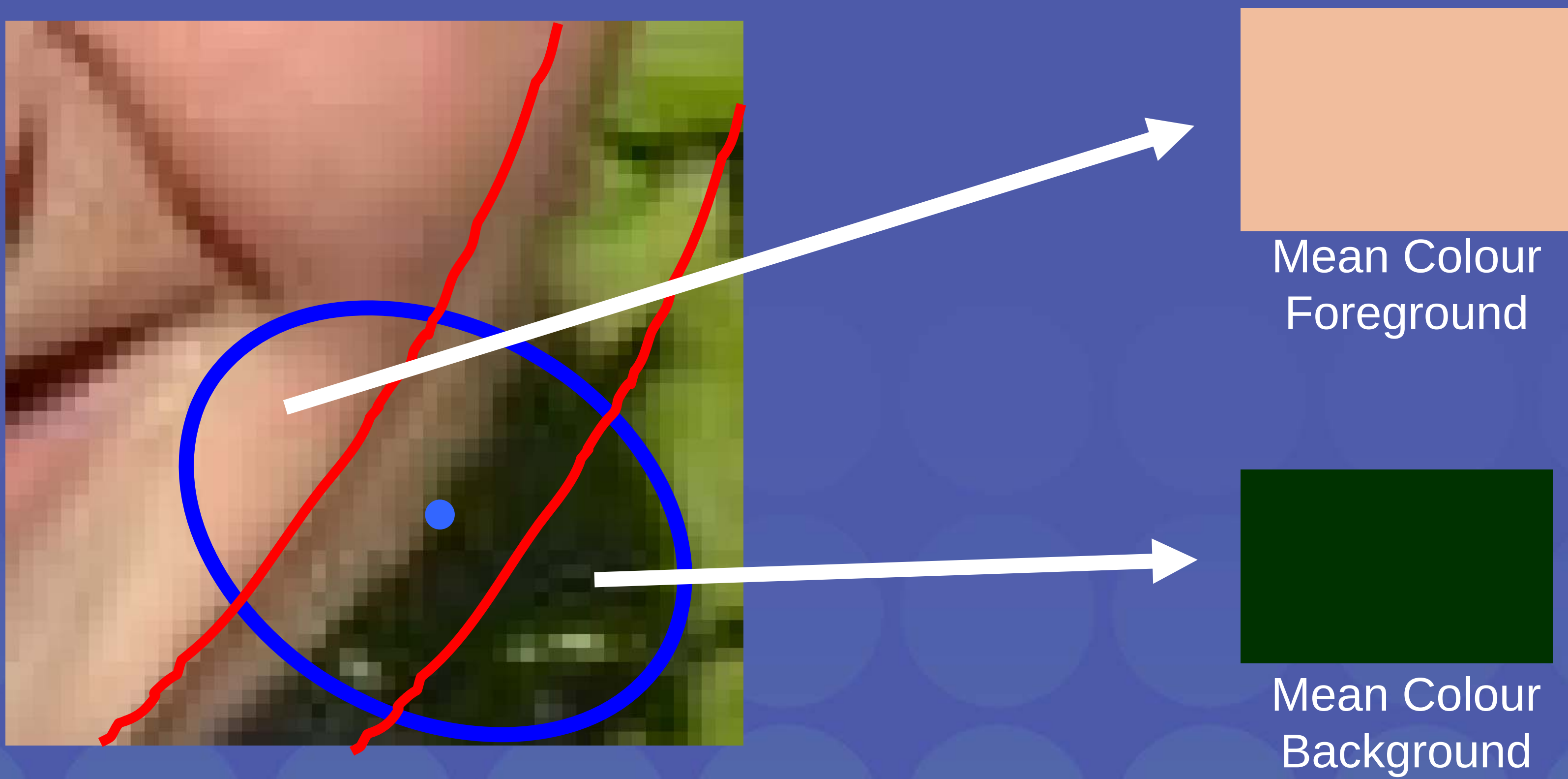
Hard Segmentation

Automatic Trimap



Soft Segmentation

Natural Image Matting

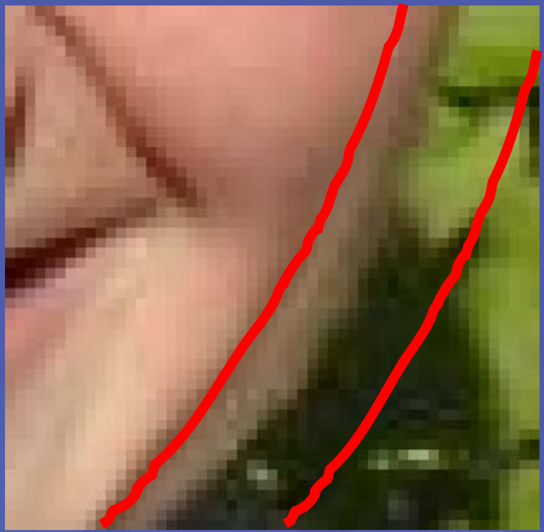


Solve

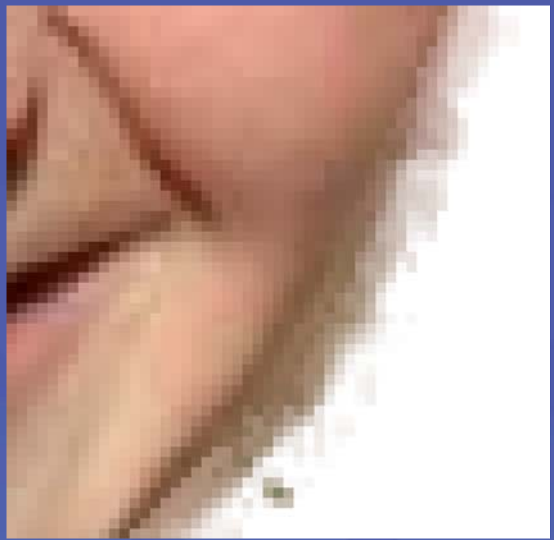
Ruzon and Tomasi (2000): Alpha estimation in natural images

Comparison

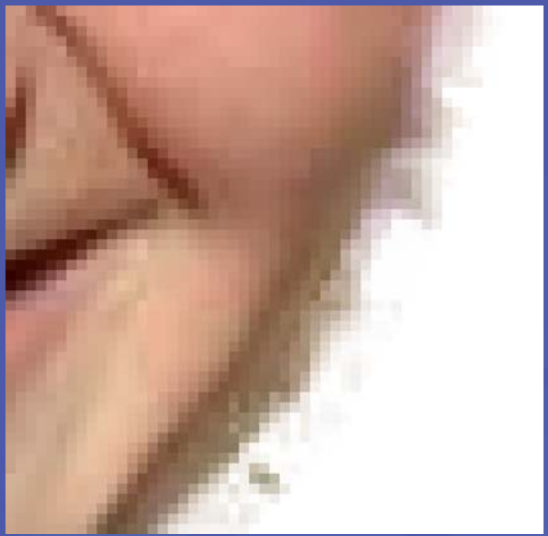
With no regularisation over alpha



Input



Bayes Matting
Chuang et. al. (2001)

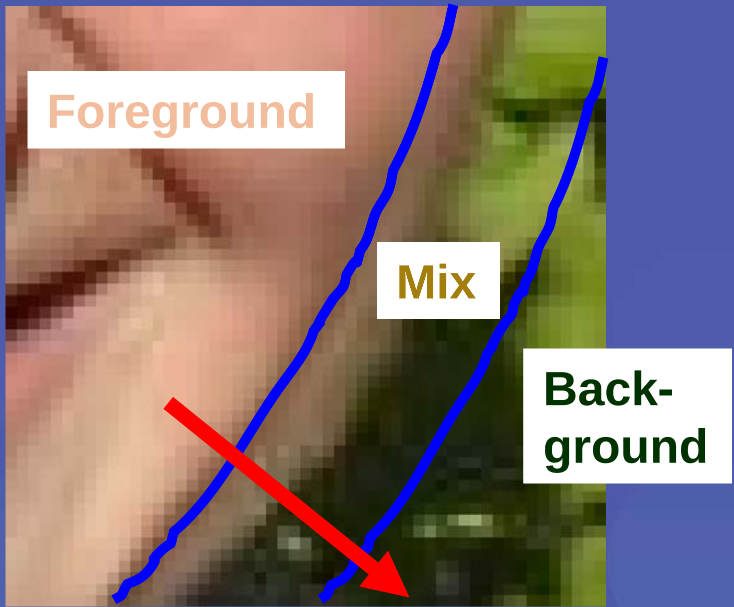


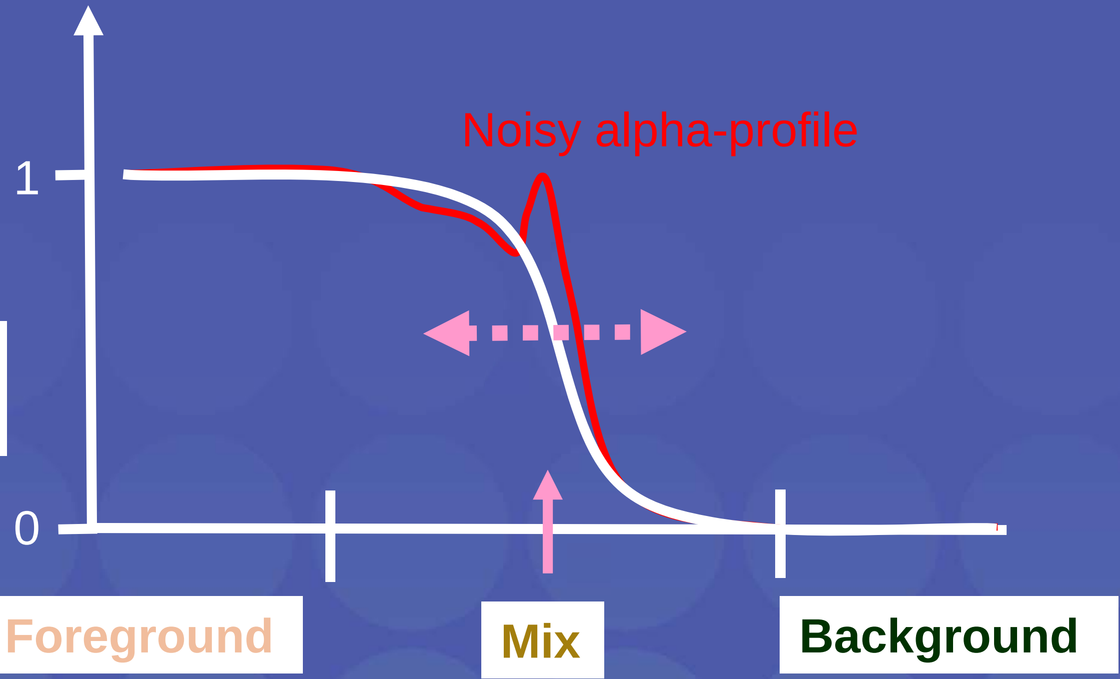
Knockout 2
Photoshop Plug-In

Shum et. al. (2004): Coherence matting in “Pop-up light fields”

Border Matting

SIGGRAPH2004

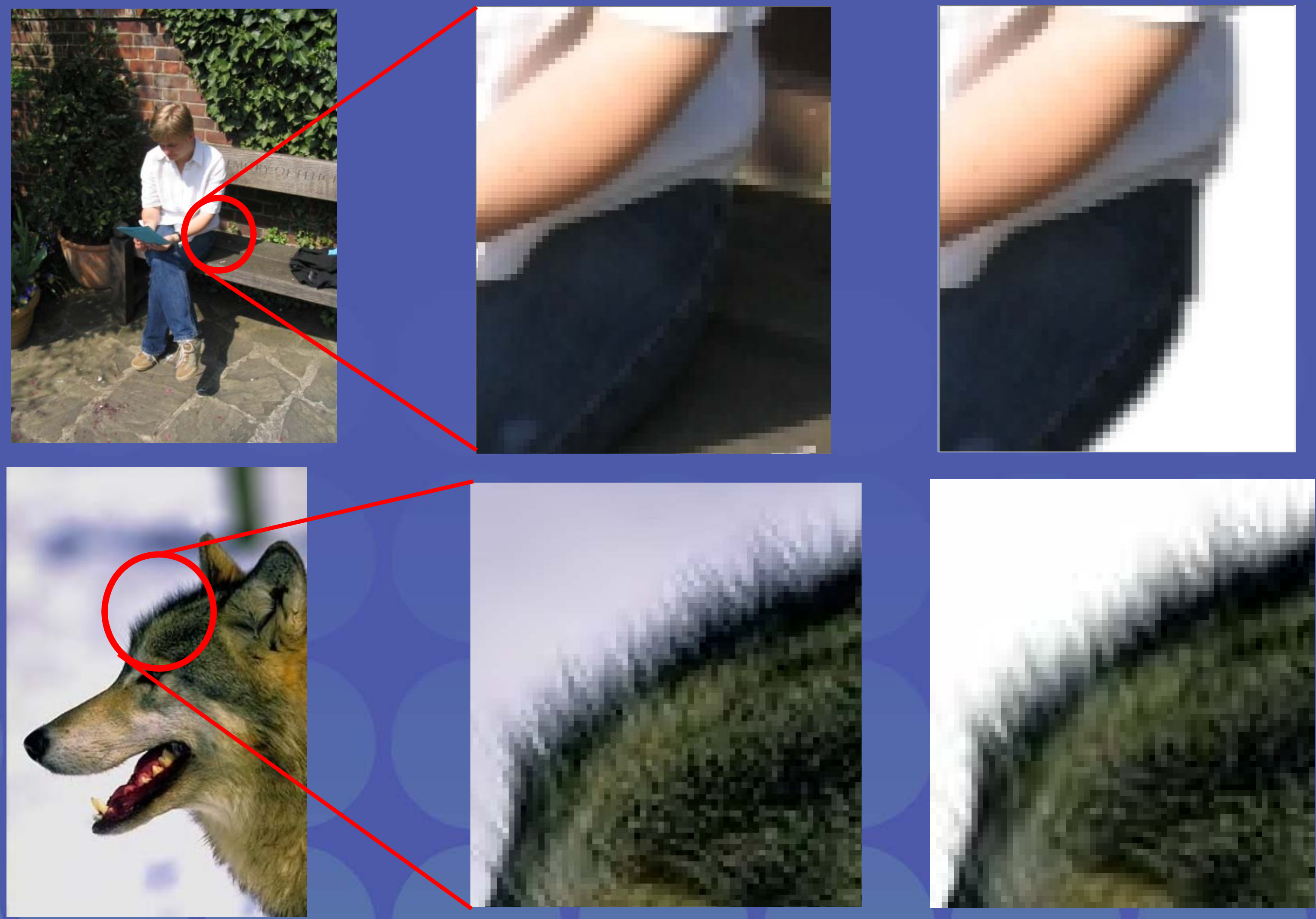




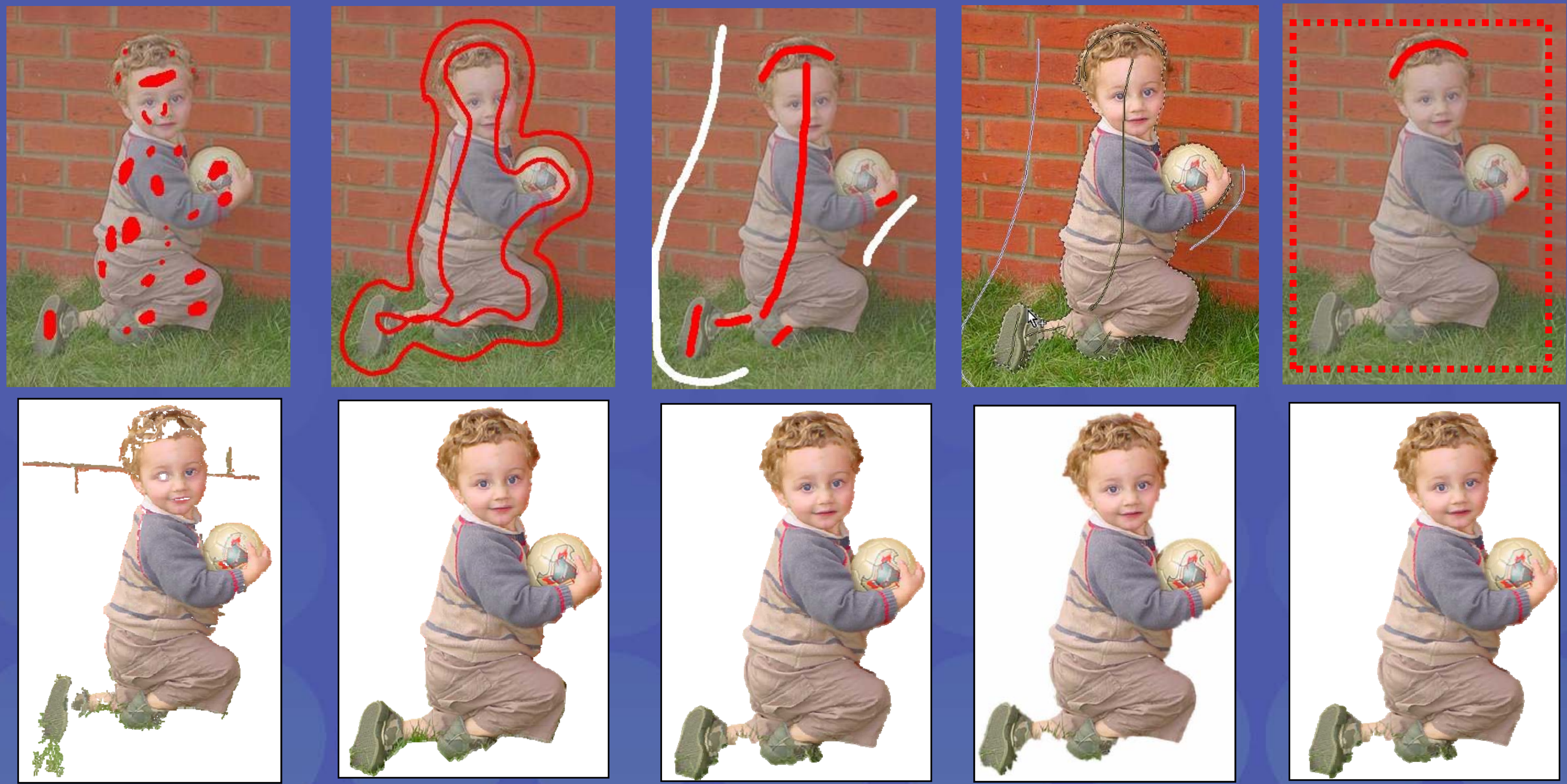
Fit a smooth alpha-profile with parameters

Results

SIGGRAPH2004



Summary



Magic Wand (198?)

Intelligent Scissors Mortensen and Barrett (1995)

Graph Cuts Boykov and Jolly (2001)

LazySnapping Li et al. (2004)

GrabCut Rother et al. (2004)

Conclusions

SIGGRAPH2004

- GrabCut – powerful interactive extraction tool
- Iterated Graph Cut based on colour and contrast
- Regularized alpha matting by Dynamic Programming

Demo

SIGGRAPH2004

The End !