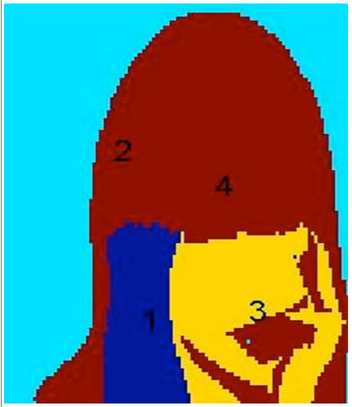
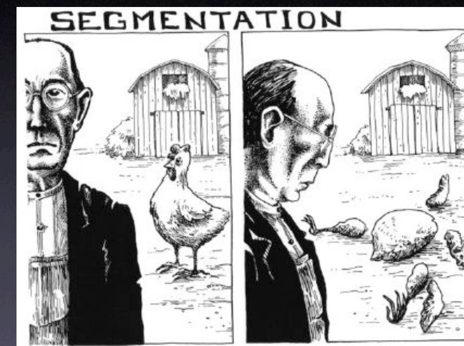


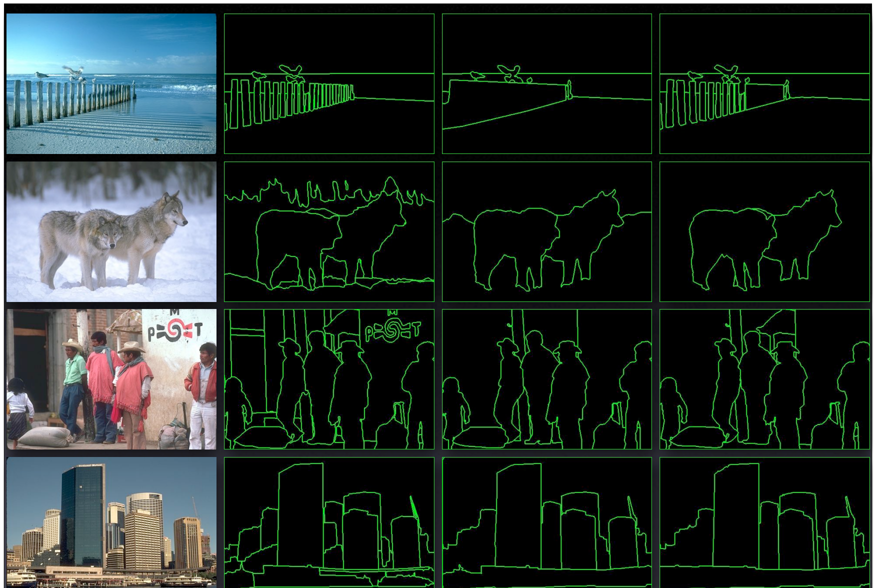
Visual Thinking via Graph Network

Jianbo Shi

Computer and Information Science
University of Pennsylvania

the whole is more than sum of ...





Berkeley Segmentation Database

A Hard Problem

Image Segmentation

Image = { pixels }
Pixel similarity

Segmentation = pixel partition

Cues: pixel similarity
object familiarity

- nocountryforoldmenjoelandethancoenschill ingconfrontationofadesperatemanwitharele ntlesskillerwontheacademyawardforbestpict ureonsundaynightprovidingamorethansatisf yingendingforthemakersofafilmthatmanybe lievedlackedoneevenasitenrichesarabrulerst herecentoilpriceboomishelpingtofuelanextr aordinaryriseinthecostoffoodandotherbasic goodsthatissqueezingthisregionsmiddlecla ssandsettingoffstrikesdemonstrationsandoc casionalriotsfrommoroccotothepersianfulf

- nocountryforoldmenjoelandethancoenschillingconfrontati onofadesperatemanwitharelentlesskillerwontheacademya wardforbestpictureonsundaynightprovidingamorethansati sfyingendingforthemakersofafilmthatmanybelievedlacked oneevenasitenrichesarabrulersthereccentoilpriceboomishel pingtofuelanextraordinaryriseinthecostoffoodandotherba sicgoodsthatissqueezingthisregionsmiddleclassandsettin goffstrikesdemonstrationsandoccasionalriotsfrommorocc otothepersianfulf
- “No Country for Old Men,” Joel and Ethan Coen’s chilling confrontation of a desperate man with a relentless killer, won the Academy Award for best picture on Sunday night, providing a more-than-satisfying ending for the makers of a film that many believed lacked one.
- Even as it enriches Arab rulers, the recent oil-price boom is helping to fuel an extraordinary rise in the cost of food and other basic goods that is squeezing this region’s middle class and setting off strikes, demonstrations and occasional riots from Morocco to the Persian Gulf.

Image Segmentation

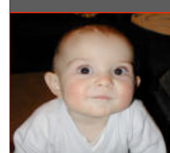
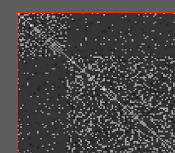
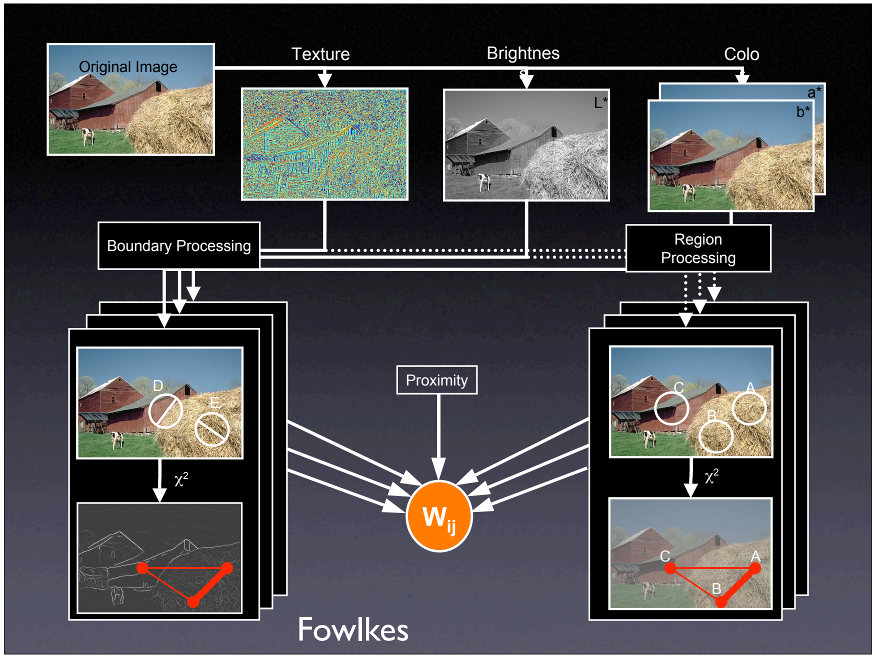


Image I →

Intensity
Color
Edges
Texture
...



Affinities W

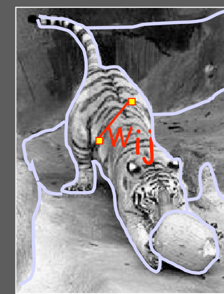


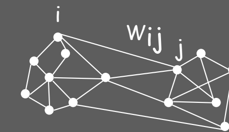
Intervening Contour

…turning a boundary map into Wij

1 - maximum P_b along the line connecting i and j

Graph Based Image Segmentation





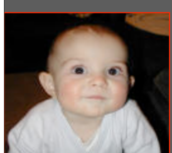
$$\mathbf{G} = \{\mathbf{V}, \mathbf{E}\}$$

Image = { pixels }
Pixel similarity

V: graph nodes
E: edges connection nodes

Segmentation = Graph partition

Graph Segmentation



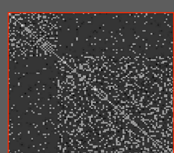


Image
I

Intensity
Color
Edges
Texture
...

Graph Affinities
W=W(I,Θ)

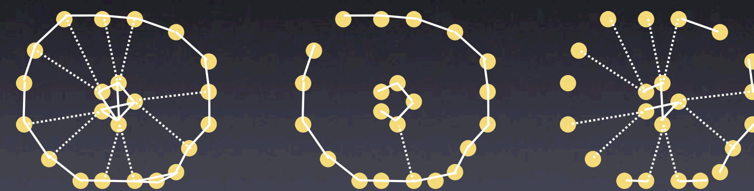
Right partition cost function?

Efficient optimization algorithm?

For simple cases, can try this:

Minimal/Maximal Spanning Tree

Tree is a graph G without cycle



Graph   Maximal   Minimal

Prim's algorithm

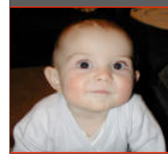
```
let T be a single vertex x
while (T has fewer than n vertices)
{
    find the smallest edge connecting T to G-T
    add it to T
}
```

Leakage problem in MST

local bad, global good

Example from Eitan Sharon

Graph Segmentation



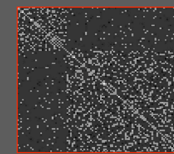
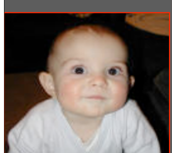


Image
I

Intensity
Color
Edges
Texture
...

Graph Affinities
W=W(I,Θ)

Spectral Graph Segmentation



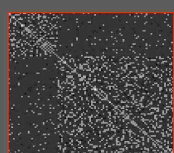
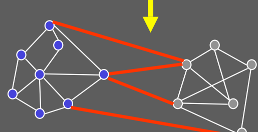


Image
I

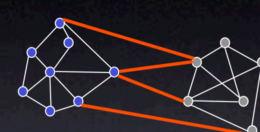
Intensity
Color
Edges
Texture
...

Graph Affinities
W=W(I,Θ)



Graph to encode
Gestalt:
Getting the big
picture of scene

Graph Cut



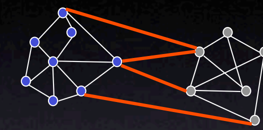
Global good, local bad

Problem with min cuts

Min. cuts favors isolated clusters

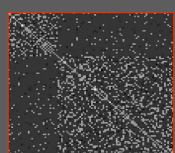
Normalize cuts in a graph

- (edge) Ncut = balanced cut

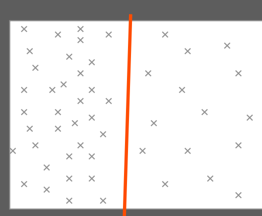


$$Ncut(A,B) = cut(A,B)\left(\frac{1}{vol(A)} + \frac{1}{vol(B)}\right)$$

Spectral Graph Segmentation



Graph Affinities
W=W(I,Θ)

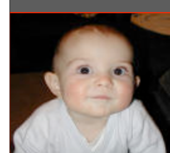


NP-Hard!

$$Ncut(A,B) = cut(A,B)\left(\frac{1}{vol(A)} + \frac{1}{vol(B)}\right)$$

measure both grouping and segmentation cost.

Spectral Graph Segmentation



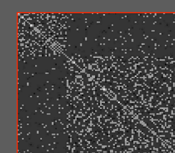


Image
I

Graph Affinities
W=W(I,Θ)

Eigenvector X
(W)

$$NCut(A,B) = \frac{cut(A,B)}{Vol\ A' \ Vol\ B}$$

$$WX = \lambda DX$$

$$X_A(i) = \begin{cases} 1 & \text{if } i \in A \\ 0 & \text{if } i \notin A \end{cases}$$

Representation

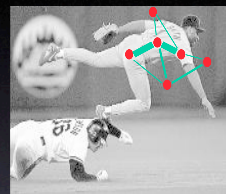
Partition matrix:

$X = [X_1, ..., X_K]$

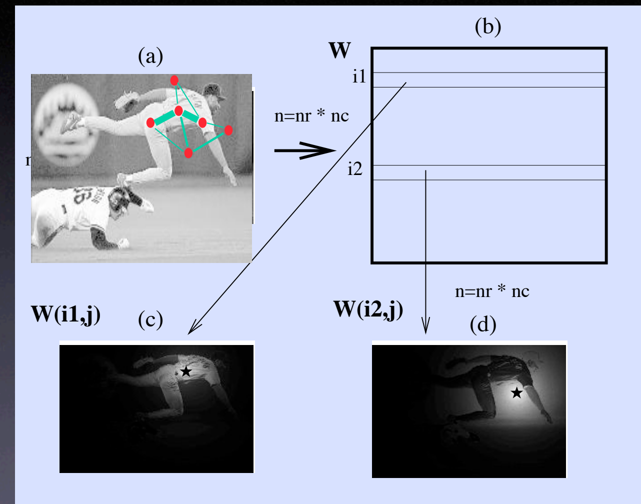
segments

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

pixels



Graph weight matrix W



Spectral Graph Segmentation

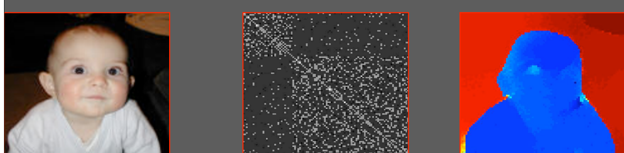


Image I → Graph Affinities W=W(I,Θ) → Eigenvector X (W)

$$NCut(A,B) = \frac{cut(A,B)}{Vol\ A'\ \ Vol\ B}$$

$$WX = \lambda DX$$

$$X_A(i) = \begin{cases} 1 & \text{if } i \in A \\ 0 & \text{if } i \notin A \end{cases}$$

Find Continuous Global Optima

$$\text{Ncut} = \frac{1}{K} \sum_{l=1}^{K} \frac{X_l^T (D - W) X_l}{X_l^T D X_l}$$

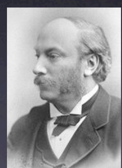
becomes

$$Ncut(Z) = \frac{1}{K} tr(Z^T W Z) \qquad Z^T D Z = I_K$$

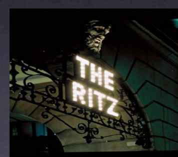
becomes

$$Ncut(Z) = \frac{1}{K} tr(Z^T W Z) \qquad Z^T D Z = I_K$$

We use the generalization of the Rayleigh-Ritz theorem to solve it.

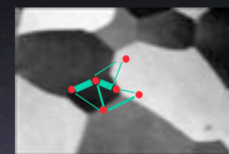


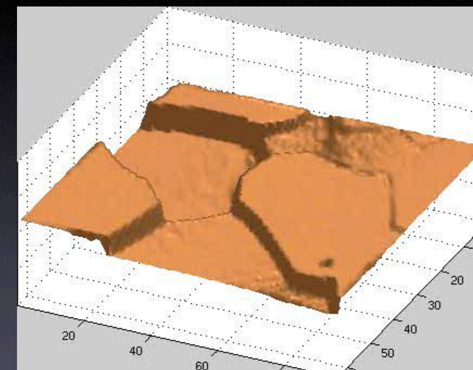
Rayleigh and...



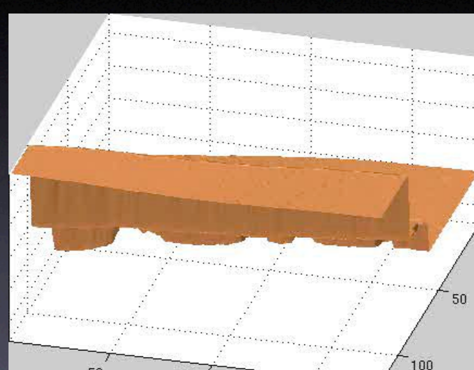
Ritz

Interpretation as a Dynamical System





Interpretation as a Dynamical System



Spectral Graph Segmentation

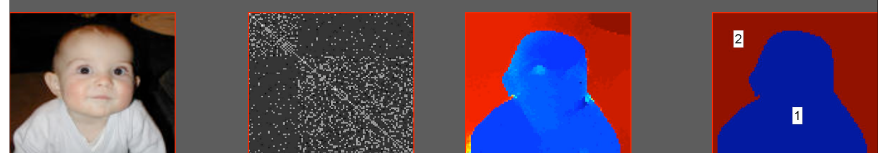


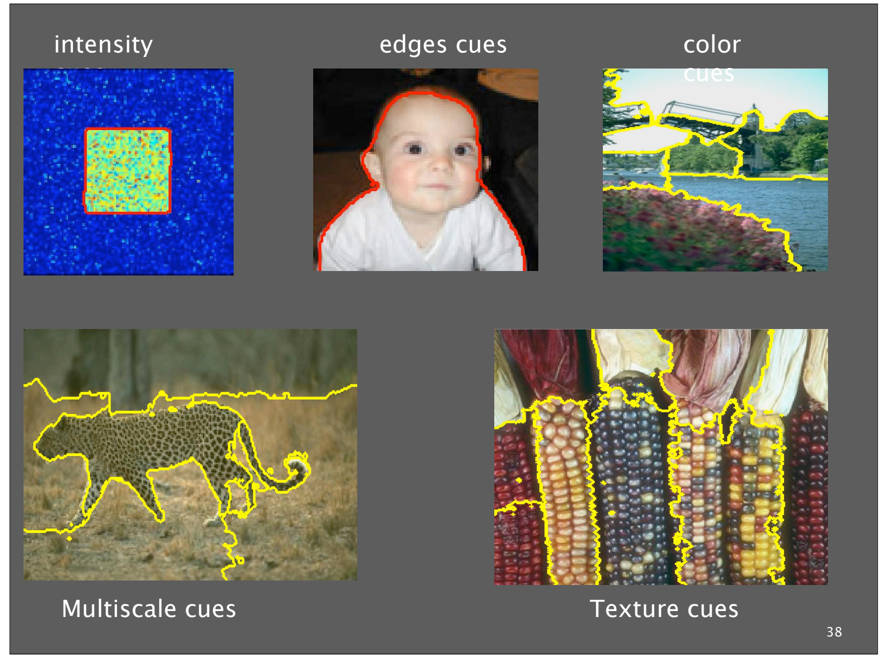
Image I → Graph Affinities W=W(I,Θ) → Eigenvector X (W) → Discretisation

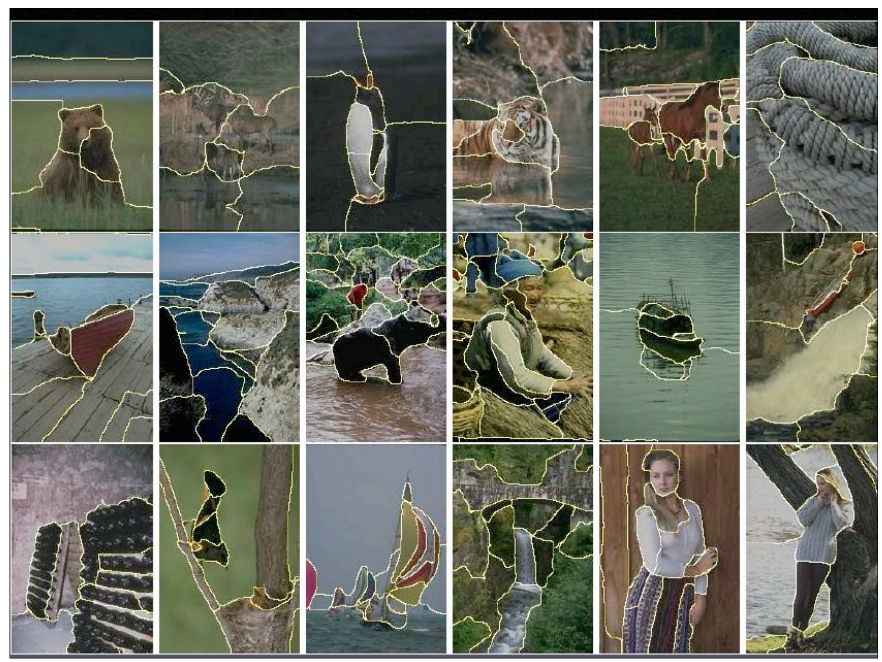
$$NCut(A,B) = \frac{cut(A,B)}{Vol\ A'\ \ Vol\ B}$$

$$WX = l\ DX$$

$$X_A(i) = \begin{cases} 1 & \text{if } i \in A \\ 0 & \text{if } i \notin A \end{cases}$$

Brightness Image Segmentation





Shi&Malik,'97

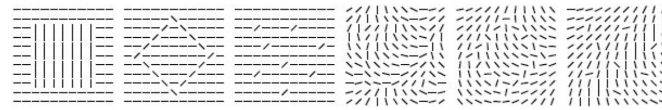
1) Encoding of basic visual cues
- a) Texture, belongie&malik '98
- b) Intervening Contour, leung&malik'98
- c) Motion, shi&malik '98
- d) Texture+contour, Malik et.al. '99, '03

2) Graph encoding of grouping constraints
- a) Complex value graph, figure-ground, Yu&shi'00
- b) Grouping with repulsion, popout, Yu&shi'01
- c) Grouping with bias, attention, Yu&shi'02

3) Learning shape based segmentation
- a) Learning Random walk, feature combination, Meila&shi'
- b) Object specific segmentation, object specific,Yu&shi'
- C) Learning Spectral graph partitioning,

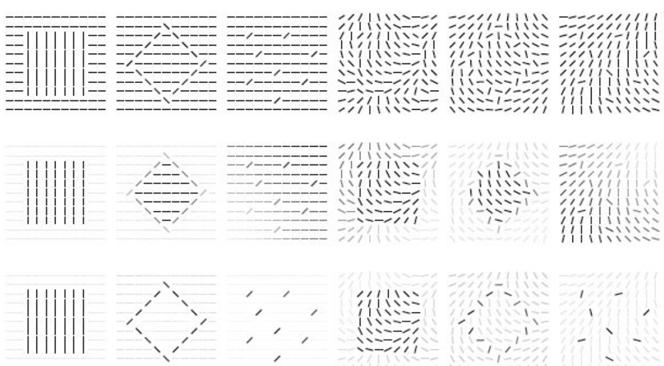
Object detection/part correspondences in grouping, Cour&Shi'04

Visual Popout:



Visual Popout:

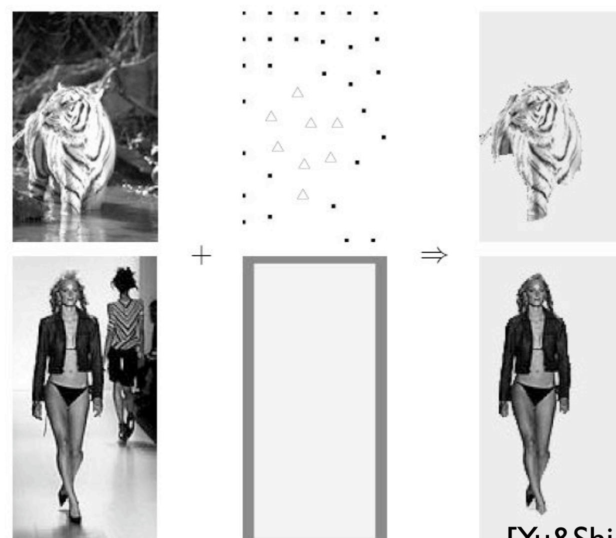
positive links only



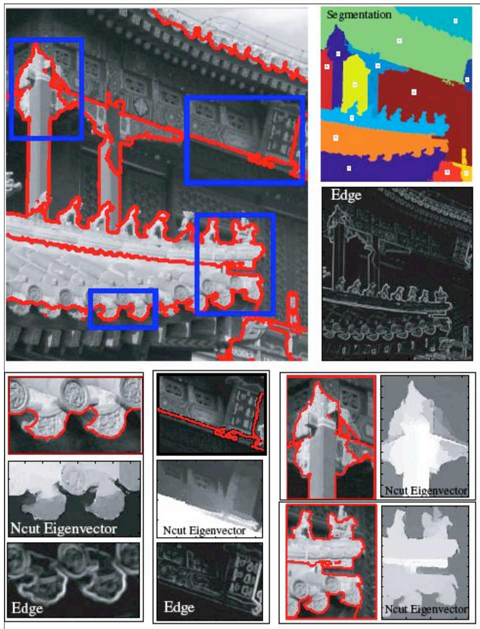
$$W = W_+ - W_- = A - R,$$

[Yu&Shi,CVPR01]

Segmentation with Attention

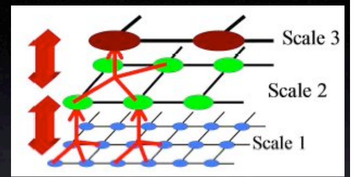


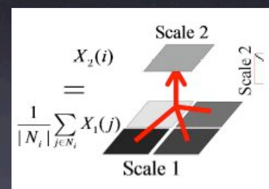
[Yu&Shi,03PAMI]



Multiscale Segmentation

Linear running time





[Cour,Benezit,Shi, CVPR05]

Saliency Region Correspondences

[Toshev,Shi,Daniilidis,CVPR07]

Spectral Graph Segmentation

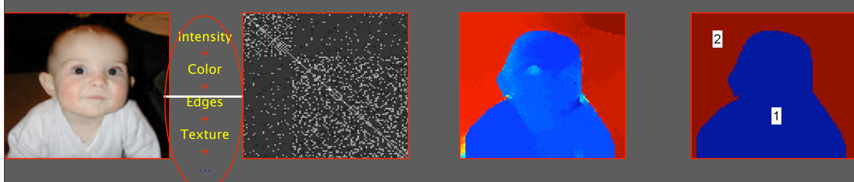


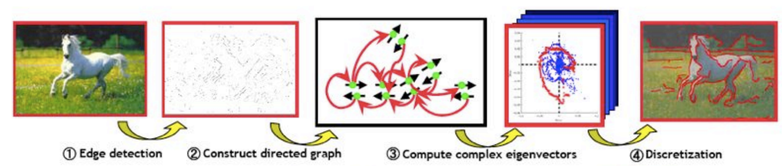
Image I → Graph Affinities W=W(I,Θ) → Eigenvector X (W)

Learn graph parameters Θ, and Graph structure itself

Target segmentation

Reverse pipeline

Untangling Cycles for Contour Grouping. Qihui Zhu, Gang Song, Jianbo Shi, ICCV 2007



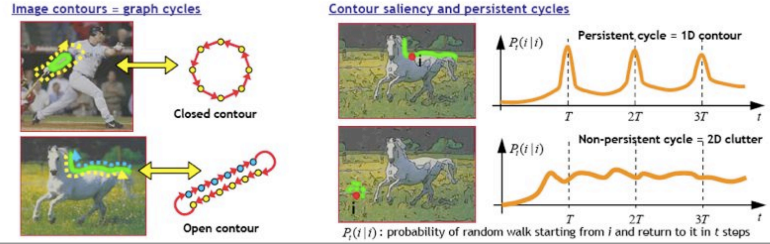
Goal: detect and group salient topologically 1D structures robust to 2D clutter and gap.

Approach: We construct a directed contour graph from image edgels and define a random walk on it. Extracting image contours, either closed or open, boils down to finding persistent graph random walk cycles. The 'peakness' of the returning probability $P_t(i|i)$ indicates 1D contour saliency, shown on the right. We show that this measure is determined by the complex eigenvectors of the random walk matrix, instead of the real eigenvectors. We derive a computational solution by tracing large cycles in complex eigenvector embedding. Detected contours in various datasets successfully capture salient 1D image structures, as shown at the bottom.

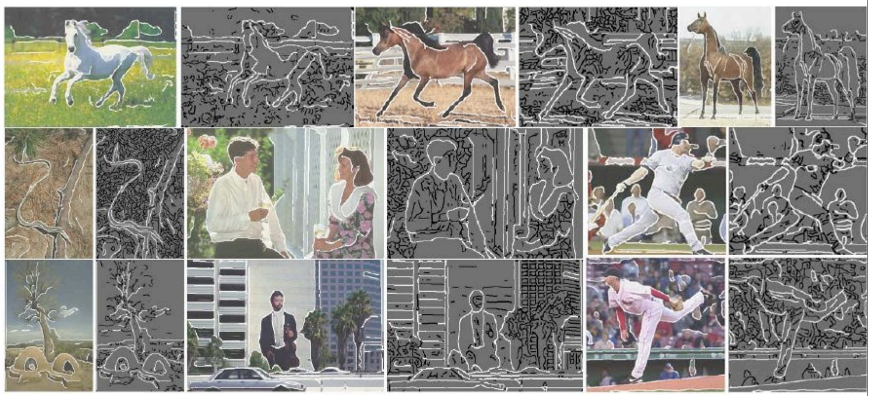
Challenge: 2D clutter and gap

Image contours = graph cycles

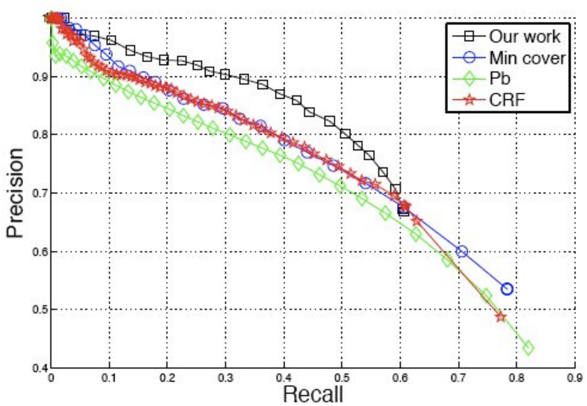
Contour saliency and persistent cycles



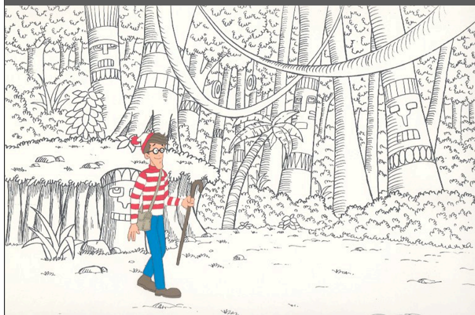
$P_t(i|i)$: probability of random walk starting from i and return to it in t steps



Precision-Recall on Berkeley Segmentation Benchmark



Where is Waldo ?



Do you use Edges cues ?
Color cues ?
Texture cues ?

–That's not enough, you need
Shape cues
High-level object priors

Image segmentation to Object recognition

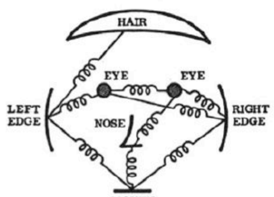
1) Graph based image segmentation

2) Bottom-up and Top-down Recognition

Part-based models

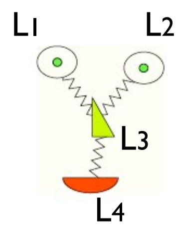
- Combination of appearance-based and geometrical models
  - Each part represents local visual properties
  - Spatial configuration captured by statistical model or spring-like connections
- Pictorial structures, Constellation of parts

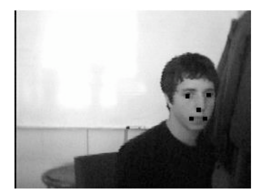
History goes back to
Fischler and Elschlager, 1973



Part-based Object Representation

- Object with n parts labeled 1 through n





- Object configuration given by: $L = (l_1, \ldots, l_n)$
  - Location of each part

(L1, L2, L3, L4) =((300,200), (300,250), (330,230),(360,230))

Part-based Object Representation

Geometrical model: $P(L)$

measuring "goodness" of the part configuration

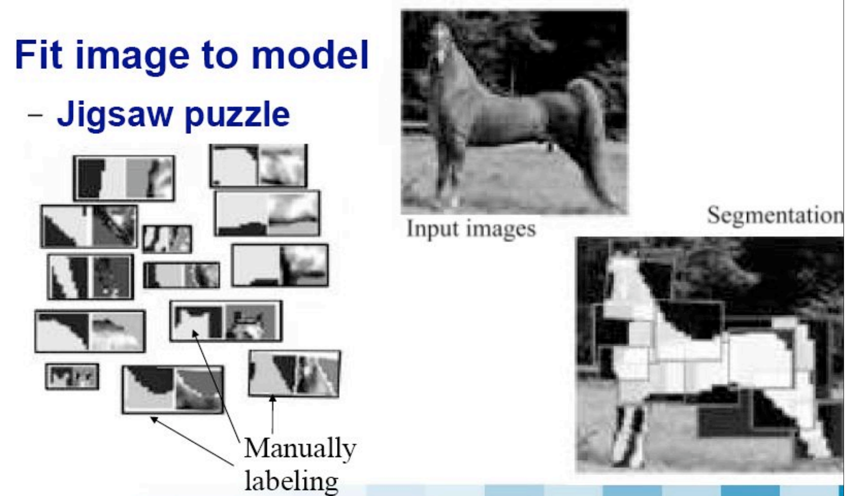
Appearance model: $P(I|L) \propto \prod g_i(I, l_i)$

image ↗ ↖ Label

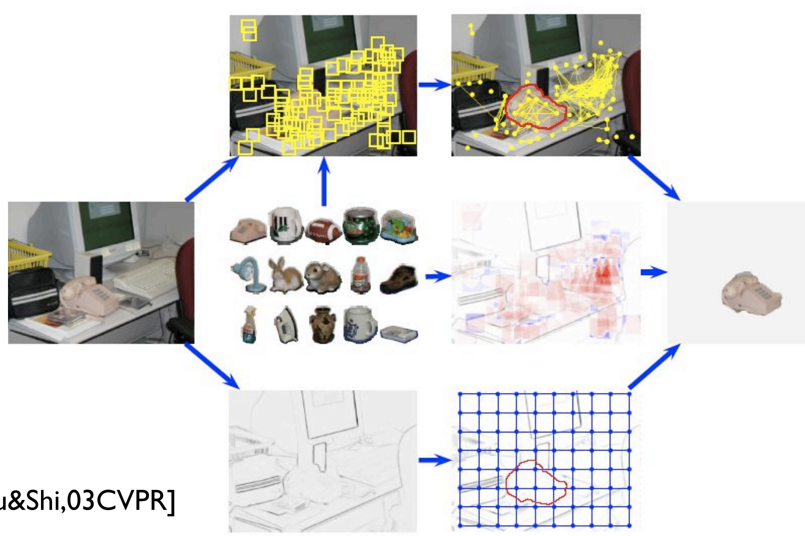
measuring "goodness" of the part appearance

Eran Borenstein, Shimon Ullman:
Class-Specific, Top-Down Segmentation. ECCV (2) 2002

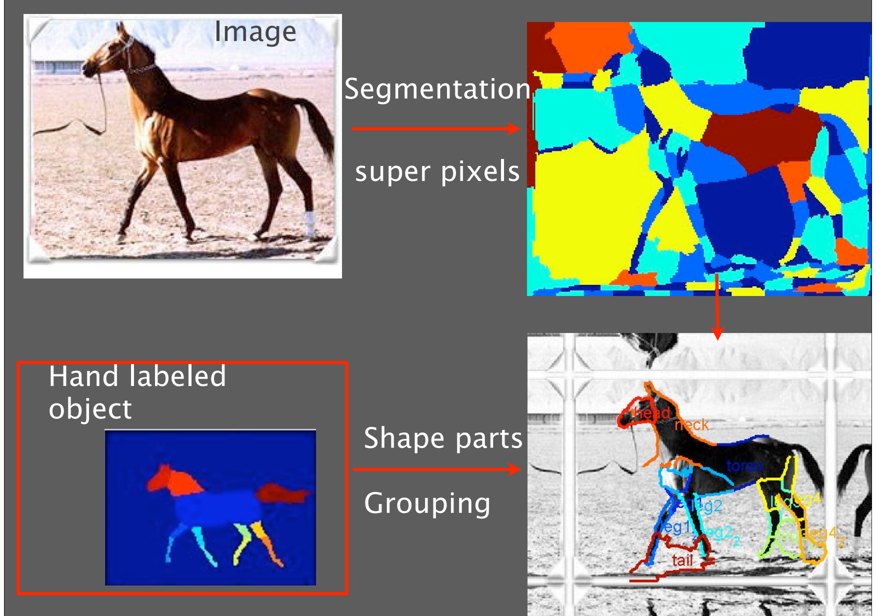
- **Fit image to model**
  - **Jigsaw puzzle**

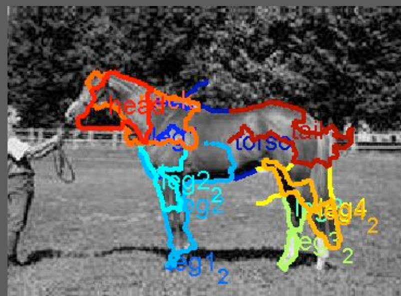


Object Specific Segmentation



[Yu&Shi,03CVPR]





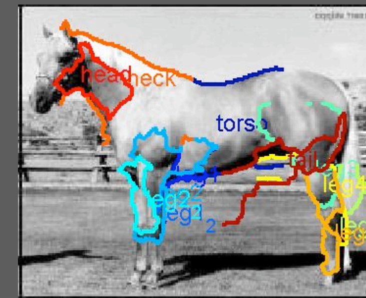
$$E(L = l_1, ..., l_k | I) = \sum_i Shape(l_i) + \sum_{i,j} Config(l_i, l_j)$$

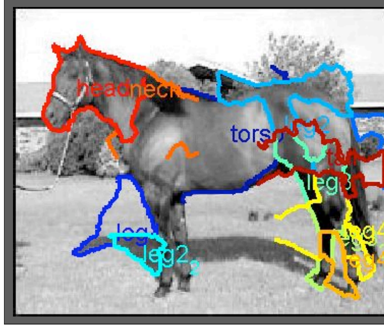
There is efficient computation procedures for this.

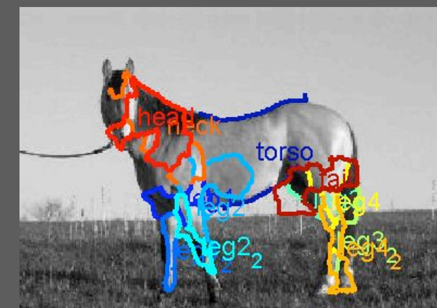
BUT: Results are Distorted:

Shapes are not additive

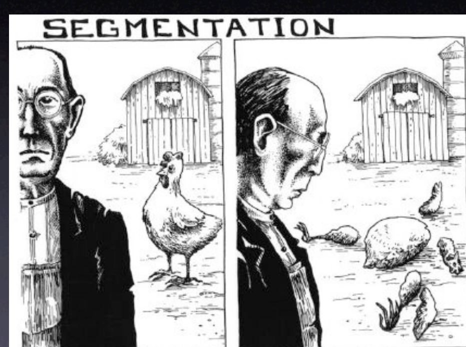
Whole is not sum of its parts







the whole is more than sum of ...



Contour Context Selection for Object Detection

Contour Context Selection for Object Detection. Qihui Zhu, Liming Wang, Yang Wu, Jianbo Shi, submitted to CVPR 2008

Fig. 1 Fig. 2 Fig. 3

Why selection? Given an object model (Fig. 1), our goal is to detect and match object instances in images via its shape. In real images, target objects are often swamped by background clutter as shown in Fig. 2. To obtain reliable shape measure robust to background clutter, we introduce context selection to point-to-point matching. Without selection, shape configuration around point B in Fig. 2 is totally different from that around point A in the swan model (Fig.1). With selection (foreground as blue lines in Fig.3), shape at point A becomes more similar to point B.

What to select? Our context selection is based on contour (a group of edge points which act as a integral part) instead of isolated edge points. Accidental alignment (green points in Fig. 4) can be pruned by identifying that a long contour (white in Fig. 4) should be selected and matched as a whole. Fig.5 gives a simple example of contour selection.

Fig. 4

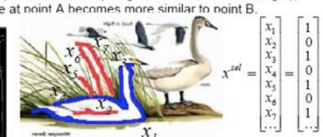
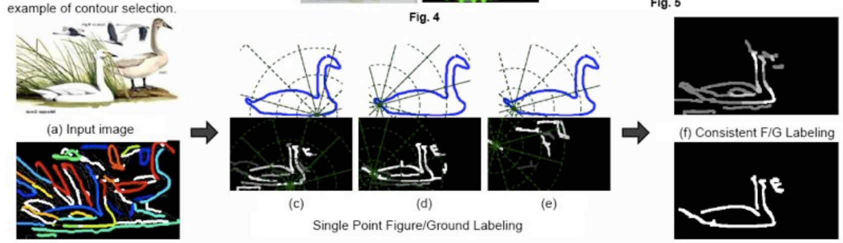


Fig. 5



(a) Input image (b) Contours (c) (d) (e) Single Point Figure/Ground Labeling (f) Consistent F/G Labeling (g) Joint Contour Selection

Fig. 6

How to select and detect? Given an input image (Fig.6-(a)), we extract long salient contours (Fig.6-(b)) from detected edges. To perform point-to-point matching, we select contours whose configuration is most similar to model (Fig.6 (c) and (d) are good matches with selected contours and (e) is a bad match). We accumulate single point matching and selection results to a consistent F/G labeling (Fig.6-(f)). Fig.6-(g) is the selection result on contours from both model and input image. Images below show our selection and detection result with model.

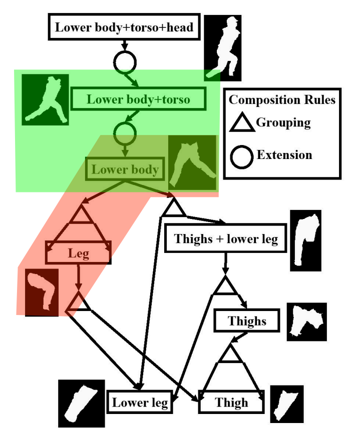
Parsing: rules guide search

- {Lower leg, Thigh} → Leg
- {Thigh, Thigh} → Thighs
- {Thighs, Lower leg} → Thighs+Lower leg
- {Thighs+Lower leg, Lower leg} → Lower body
- {Leg, Leg} → Lower body
- {Lower body} → Lower body+torso
- {Lower body+torso} → Lower body+torso+head

Figure 2. Our parse rules. We write them in reverse format to emphasize the bottom-up nature of the parsing.

Parsing: proposal and evaluation

- Parsing begins at leaves, continues upwards
- Parse rules create proposals for each part (**proposal**)
- Proposals scored according to shape, ranked/pruned (**evaluation**)



Bottom-up Recognition and Parsing of Human Body, [Srinivasan & Shi, CVPR 07]

GRASP

Penn
UNIVERSITY of PENNSYLVANIA

Joint work with, T. Cour, P. Srinivasan, A. Toshev, Q. Zhu, G. Song, F. Benezit, L. Wang, S. Yu, K. Daniilidis

Thank you!