Markov Models in Computer Vision

Andrew Blake, Microsoft Research Cambridge

Modern probabilistic modelling has revolutionized the design and implementation of machine vision systems. There are now numerous instances of systems that can see stereoscopically in depth, or separate foreground from background, or accurately pinpoint objects of a particular class, all in real time. Each of those three vision functionalities will be demonstrated in the lecture. The underlying advances owe a lot to probabilistic frameworks for inference in images. In particular, the Markov Random Field (MRF), borrowed originally from statistical physics, first appeared in image processing in the 70s. It has staged a resounding comeback in the last decade, for very interesting reasons.

Seeing as an intelligent behaviour

Sony "Qrio" humanoid

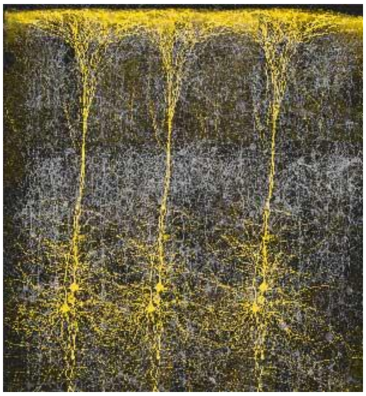
Robot car

Visual cruise control

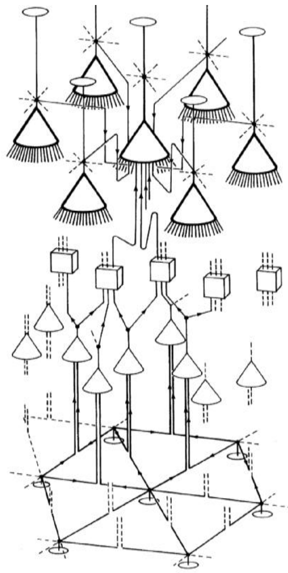
Live Object

Xbox Natal

Seeing: hardware or software?



Hypercolumns in visual cortex
(Mountcastle - 1957, Hubel & Wiesel - 1962)



An account of how vision might work

- Having the ability to test hypotheses
- Dealing with the ambiguity of the visual world
- Having the ability to "fuse" information
- Having the ability to learn
    - <u>Reasoning with probabilities</u>

Machines that see

... perceptions are predictive, never entirely certain, hypotheses *of what may be out there.*

R.L. Gregory, psychologist, 1966

... the essential problem of perception ... is how reliable knowledge of the world around us is extracted from a mass of noisy and potentially misleading sensory messages.

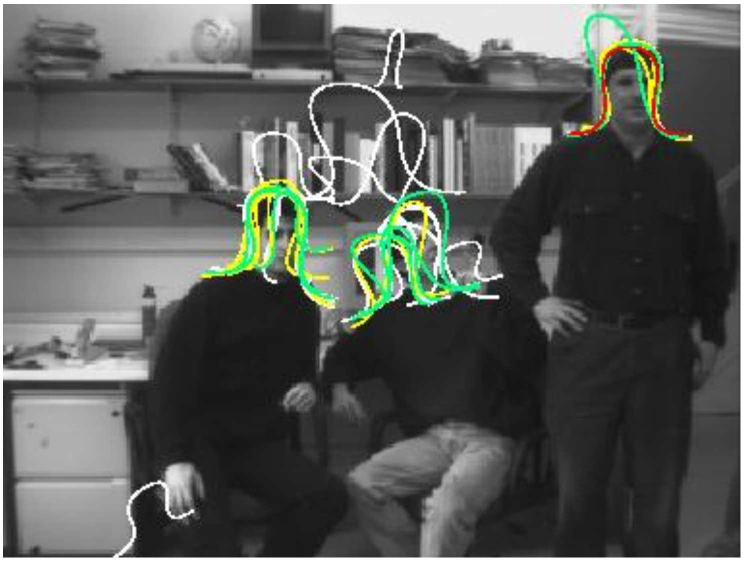
What I am suggesting is that the statistical step of extracting knowledge is often solved before we consciously perceive anything at all, and that is why our perceptions are usually reliable.

HB Barlow, neurophysiologist, 1980.

As Grenander has emphasized, a very useful test for a class of models is to synthesize from it, i.e. choose random samples according to this probability measure and to see how well they resemble the signals.

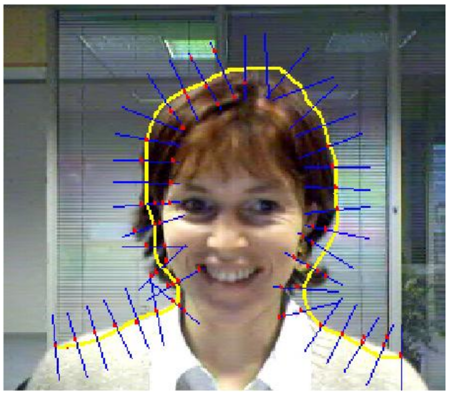
D. Mumford, mathematician, 2002.

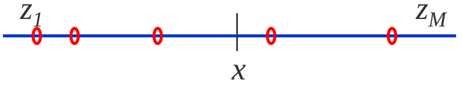
"Particles" for head location



Tracking against clutter 2

Visual ambiguity in the location of a head

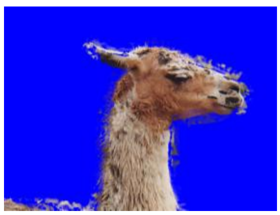




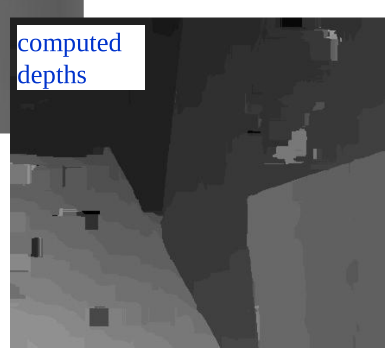
.. What is hard about vision -- ambiguity

-- what you want from an interactive segmentation:

-- what you get from local analysis of colour and texture:



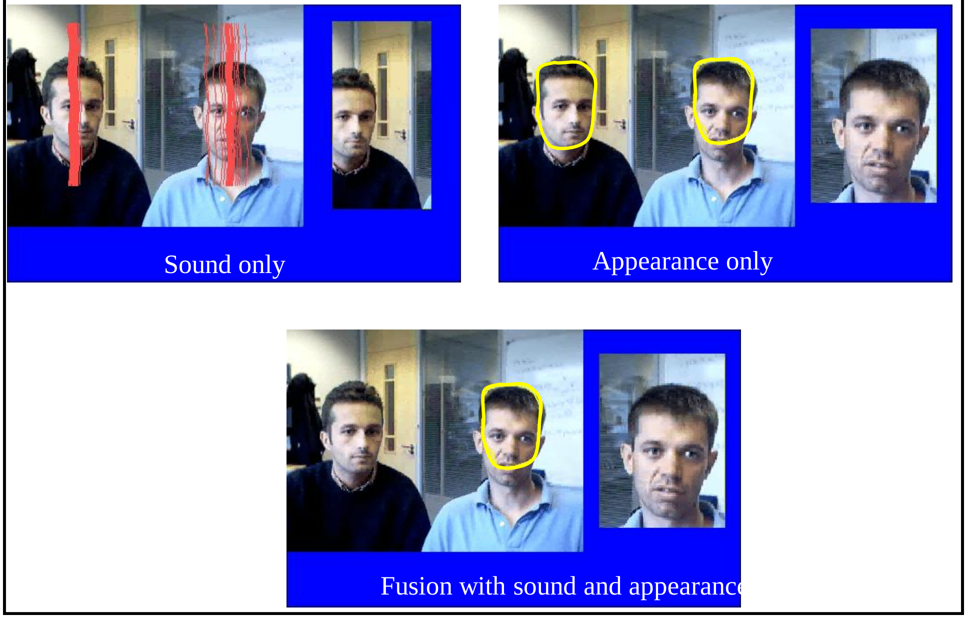
... more ambiguity



Stereo Vision – local analysis

(Scharstein & Szeliski 2002)

Fusion of cues: sound and visual appearance



Learning by example

An account of how vision might work

- Having the ability to test hypotheses
- Dealing with the ambiguity of the visual world
- Having the ability to “fuse” information
- Having the ability to learn
  - Reasoning with probabilities

How vision might work: digging deeper

Remainder of lectures focus on one or two key problems in vision:

-- principally the problem of segmentation
-- considered in probabilistic/optimization framework
-- also look briefly at stereo vision
-- all peppered liberally with some mainstream apps.

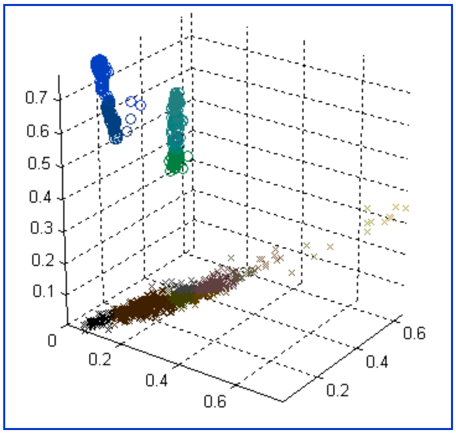
Segmentation (Berkeley)

... Better defined problem: foreground segmentation

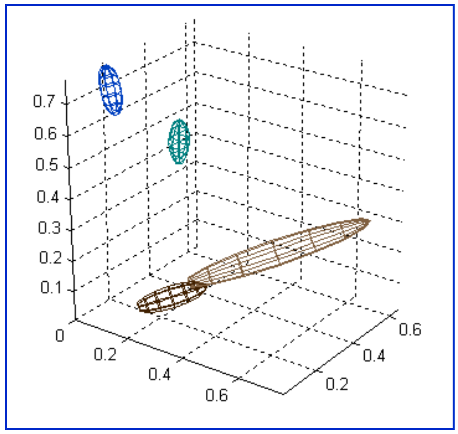
Bayesian Matting (Chuang et al. 01)

Mixture of Gaussian colour priors

Training data in Colour Space



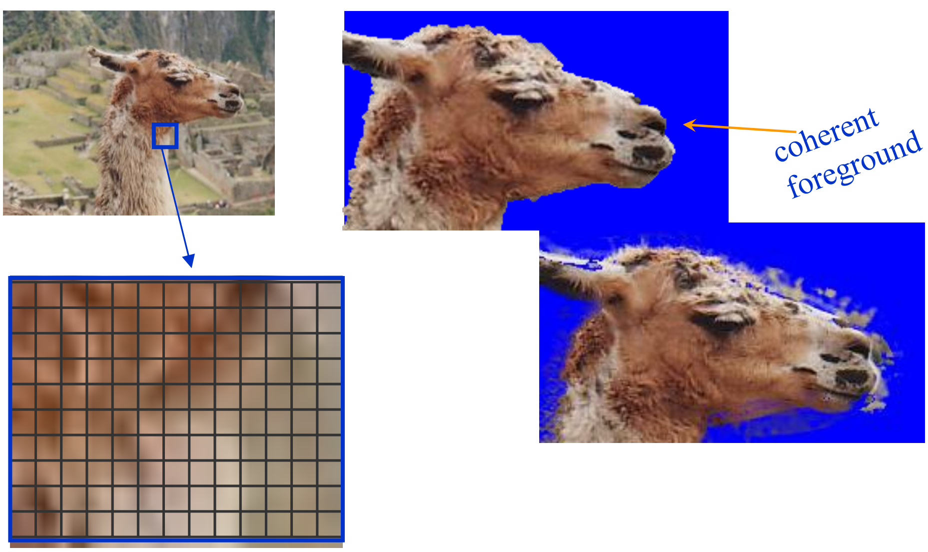
Learned mixture of Gaussian Models



MAP inference of $\alpha, f | z$

(Ruzon and Tomasi 99; Chuang, Curless, Salesin, Szeliski 01)

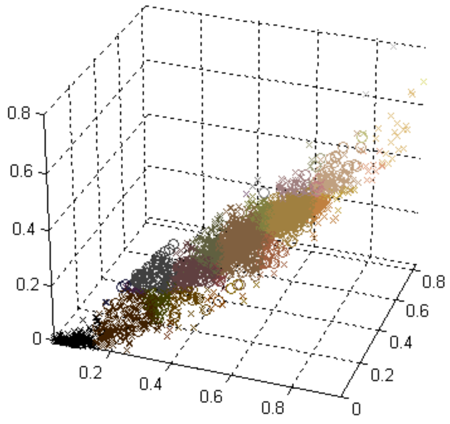
... Segmentation in camouflage

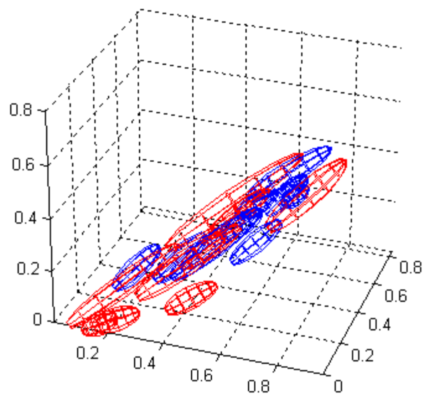


Need for spatial priors

Data in Colour Space

Mixture of Gaussian Models

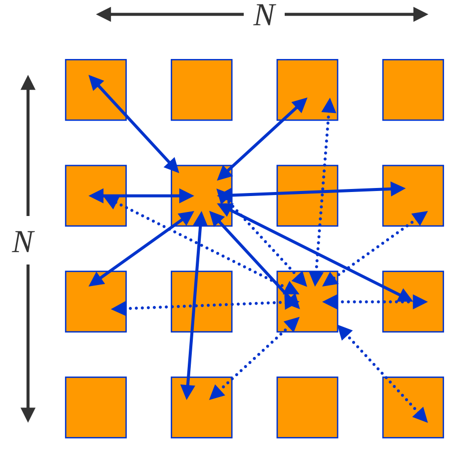




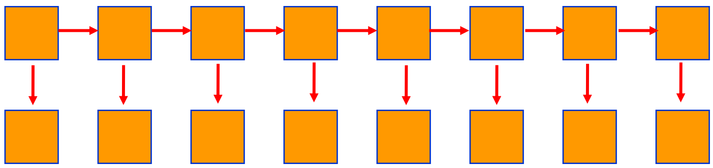
-- f/b distributions intertwined

Connecting up pixels

How can you express coherence in a way that is practical?

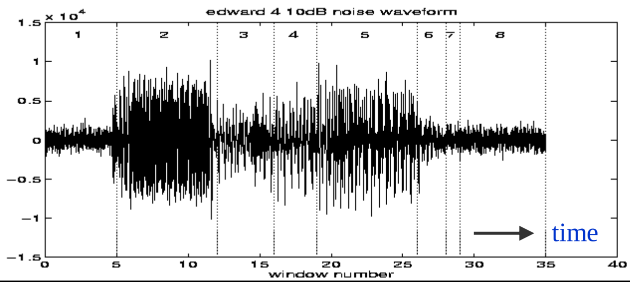


1st order with stochastic observations – Hidden Markov Model

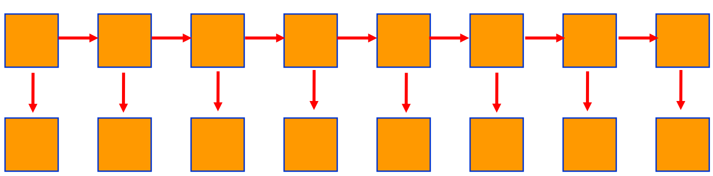


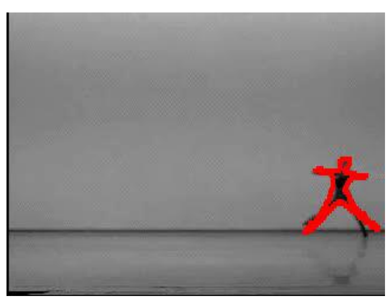
HMMs – ubiquitous in speech recognition
(Rabiner, 89; Jelinek 98) HTK (Young,Woodland et al. 97)

→ tractable –
Dynamic Programming
etc.



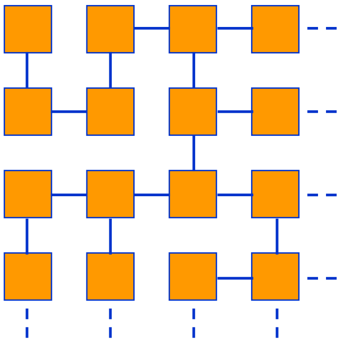
Temporal HMMs in vision



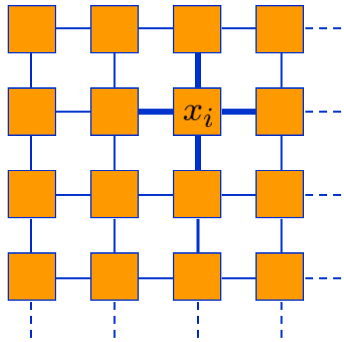


2D Markov model?

Tree of connected pixels
(Veksler 2005)



Markov Network (1st order)
(Geman & Geman 84; Besag, 1974, 1986)



neighbours of i

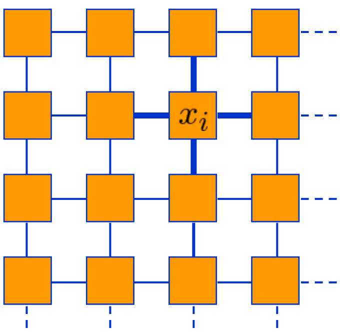
Independence:

$$p(x_i \mid x \backslash \{x_i\}) = p(x_i \mid \mathcal{N}_i)$$

where $x = \{x_1, \ldots, x_N\}$

2D Markov Model 1st order Example

Ising Model



Binary variables: $x_i \in \{0, 1\}$

Joint probability distribution:

$$p(x) = \prod_{(i,j)\in\mathcal{N}} G_{i,j}(x_i, x_j)$$

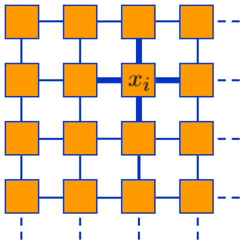
where

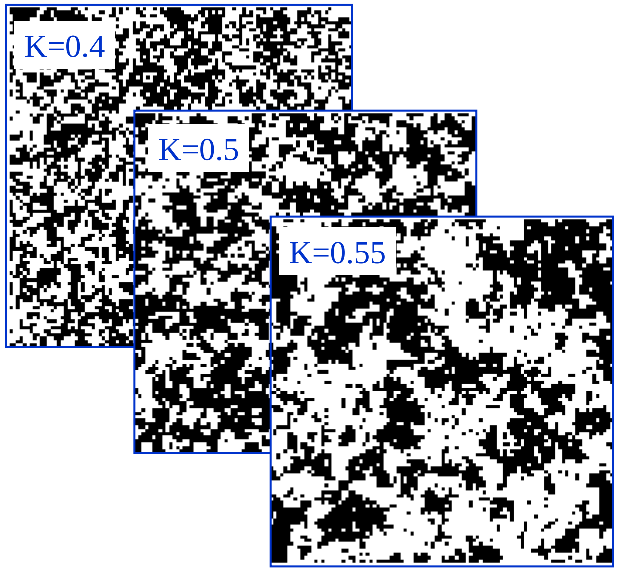
$$G_{i,j}(x_i, x_j) = 1 - K|x_i - x_j|$$

and $1 > K > 0$

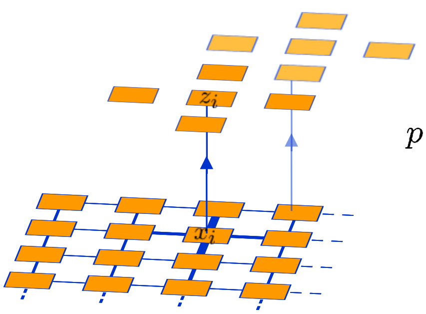
2D Markov simulation (Swensden Wang MCMC)

Ising Model





2D Hidden Markov Model



$$p(x \mid z) \propto \underbrace{p(z \mid x)}_{\text{observation likelihood}} \; \underbrace{p(x)}_{\text{prior}}$$

$$p(z \mid x) = \prod_i F_i(x_i, z_i)$$

$$\prod_{(i,j)\in\mathcal{N}} G_{i,j}(x_i, x_j)$$

1st order model

Simple segmentation --- Ising prior

MRF – expressed as additive energy terms

$p(x \mid z) \propto \exp -E$ where "energy" $E = U + V$

and

$$U(x,z) = \sum_i f_i(z_i \mid x_i) \qquad f = -\log F$$

colour observations

$$V(x) = \sum_{i,j} g_{i,j}(x_i, x_j) \qquad g = -\log G$$

(-ve) log-prior *V(x)*

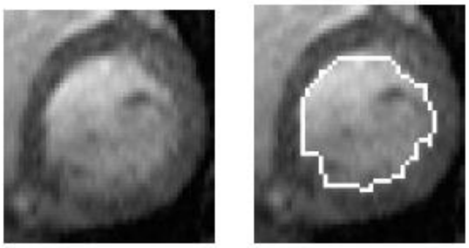
with $g_{i,j}(x_i, x_j) = \gamma |x_i - x_j|$ and $\gamma = -\log(1-K)$

MAP $\longrightarrow$ energy/cost minimization

?? How to compute $\max_{\mathbf{x}} \; p(\mathbf{x} \mid \mathbf{z})$ ie $\min_{\mathbf{x}} \; E(\mathbf{x})$

Segmentation artefacts --- Ising prior

(Boykov and Kolmogorov ICCV 2003)



?? How to overcome artefacts

Boykov-Jolly contrast-sensitive segmentation

(Boykov and Jolly 2001; Rother et al. 2004; Li, Shum et al. 2004)

- Conditional Random Field -- CRF

  (Lafferty et al. 2001; Kumar and Hebert 2003)

$$p(\mathbf{x} \mid \mathbf{z}) \propto \exp -E \qquad \text{with} \qquad E = U + V$$

where now

$$V = \underline{\underline{\gamma \sum_{i,j} |x_i - x_j| \left(\exp - \frac{\|z_i - z_j\|^2}{2\sigma^2} \right)}}$$

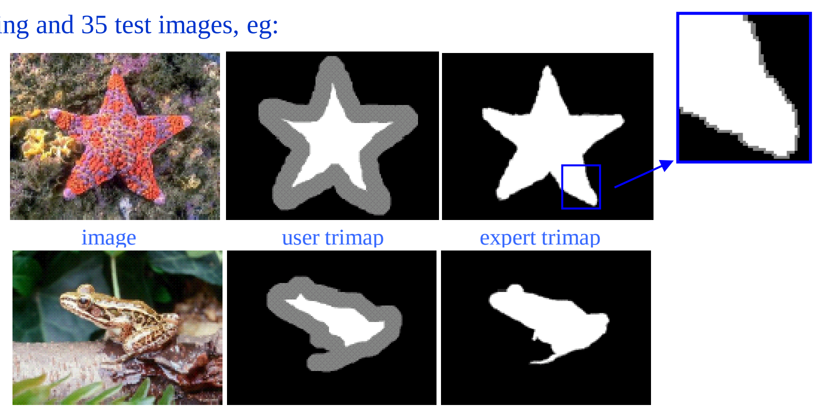
log-"prior" $V(\mathbf{x},\mathbf{z})$

data-dependence

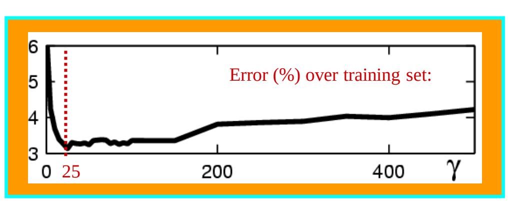
Labelled database

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

15 training and 35 test images, eg:



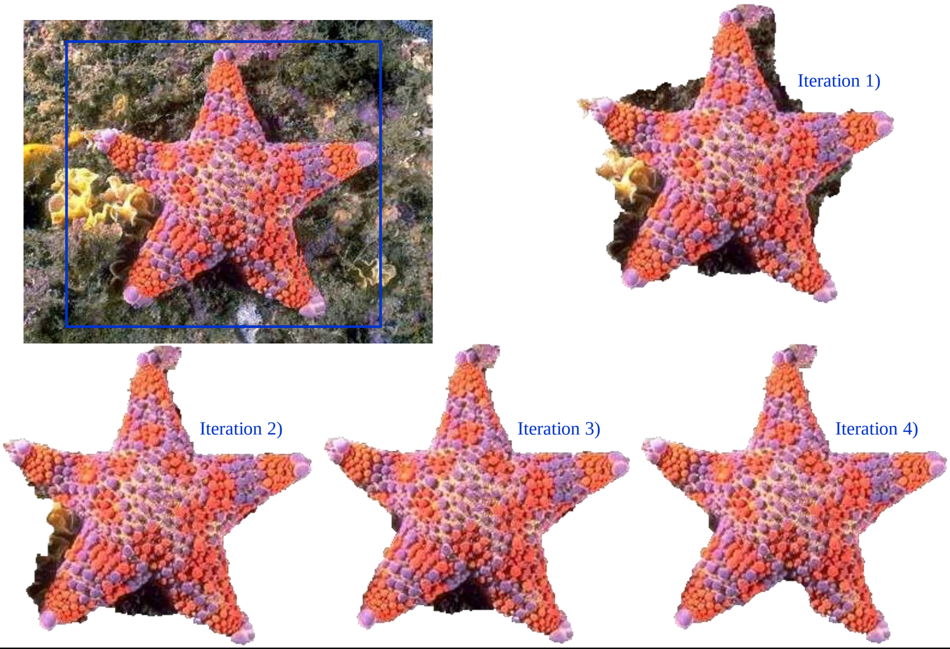
Training error for γ:



(see also Blake et al. ECCV 2004)

GrabCut: partially supervised inference

(Rother, Kolmogorov and Blake, Siggraph 2004;)



Evaluation of Grabcut over database

Input Image	Ground Truth	Trimap Boykov and Jolly	Bimap GrabCut
		Error Rate: 1.36%	Error Rate: 2.13%

- User Interactions - considerably reduced
- Error rate: modest increase

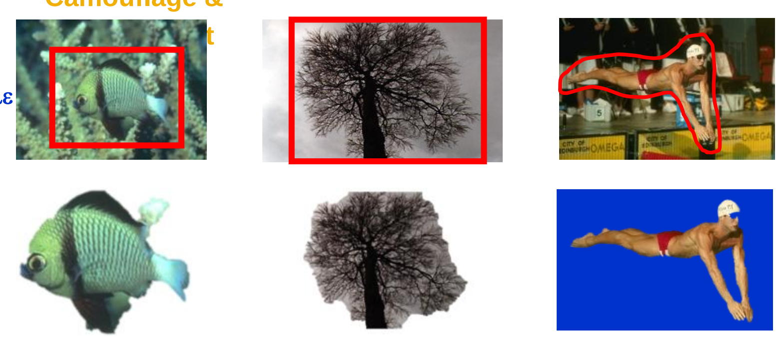
Straightforward examples

Difficult Examples

Camouflage & t | Fine structure | No telepathy

Σελεχτεδ Ρεχτανγλε

Ρεσυλτ



MAP estimation for Markov Random Fields – Energy Minimization

Generally NP-hard, so approximate:

Simulated annealing [Metropolis, Rosenbluth, Rosenbluth, Teller and Teller, 1953]

Gibbs sampling [Geman and Geman 1984]

Iterated conditional Modes [Besag 1986]

Approximate variational extremum [Mumford and Shah 1985,9]

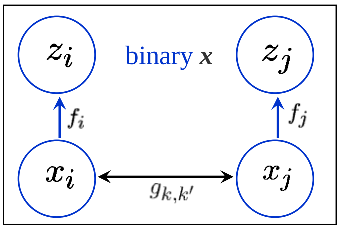
Graduated nonconvexity [Blake and Zisserman 1987]

Graph cut [Greig, Porteous and Seheult, 1989]

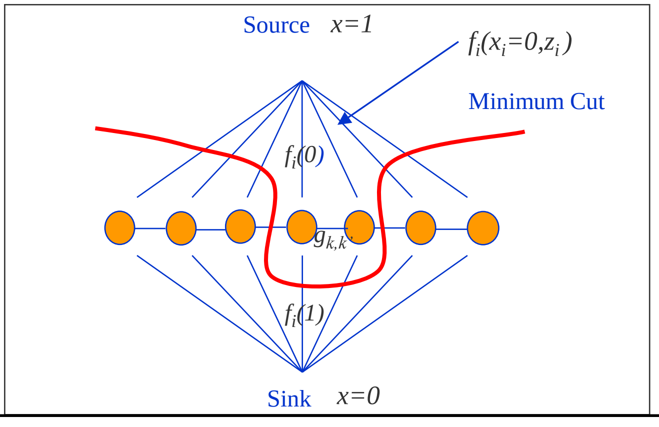
Loopy Belief Propagation [Freeman and Pasztor, 1999]

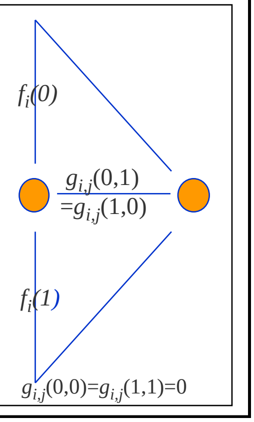
"Modern" graph cut [Boykov, Veksler and Zabih, 2001]

Graph Cut engine for Markov segmentation

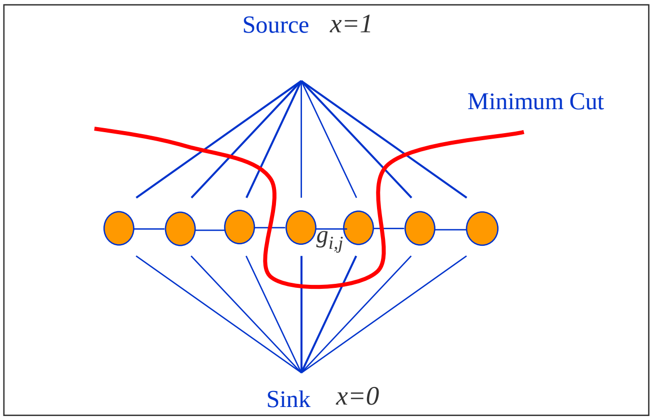


$$E(\mathbf{x},\mathbf{z}) = \sum_k \sum_{k'} g_{k,k'}(x_k, x_{k'}) + \sum_k f_k(x_k, z_k)$$





Ford-Fulkerson Min-cut/Max Flow



* Max flow you can push through the network = min cut

-- ie capacity of cut with smallest total capacity

* Links saturated by max flow = links separated by min cut

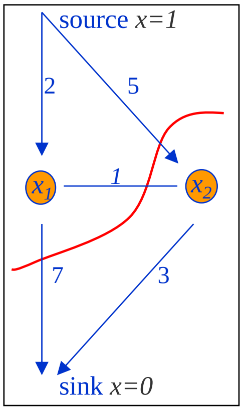
Example: optimization as graph cut problem

Problem: $\min_{\mathbf{x}} \; E = f_1(x_1) + f_2(x_2) + g(x_1,x_2)$

where: $f_1(0)=2; f_1(1)=7; \; f_2(0)=5; f_2(1)=3;$ [all weights +ve]

$g(0,1)=g(1,0)=1; \; g(0,0)=g(1,1)=0;$ [canonical form]

Graph:

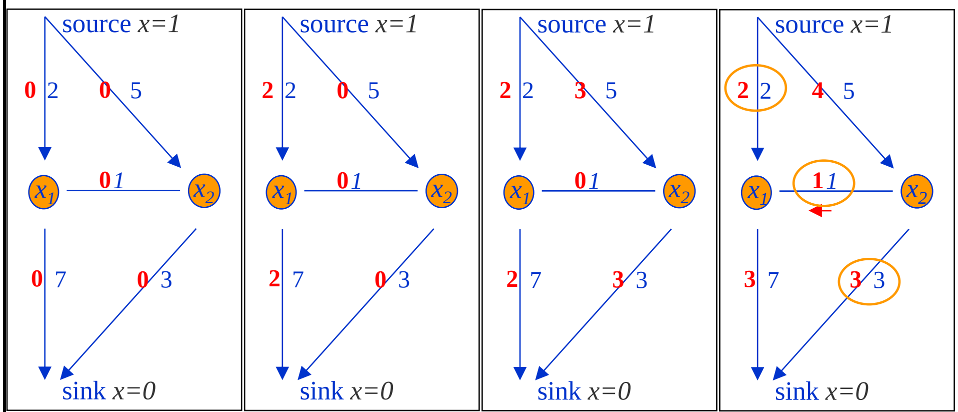


Trial Solution $x_1=1, x_2=0$

Energy: $E(x_1, x_2) = 7+1+5 = 13$

Example: graph cut optimization – min cut/max flow

Solve (exactly) by augmenting flow:



Solution from saturated paths

$x_1=0, x_2=1 \quad E(x_1, x_2) = 2+1+3 = 6$

Augmenting flow

[Boykov and Kolmogorov, PAMI 2004]

-- Breadth-first search for augmenting paths

-- Augment flow along non-saturated paths

-- Termination if weights all positive

-- Typical complexity $O(n^3)$ [Dinic 1970]

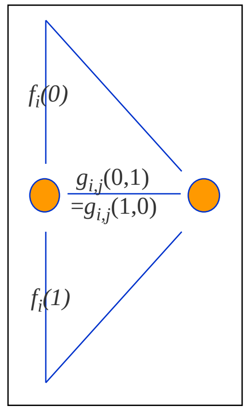
-- Special algorithms for vision problems – wide, shallow graphs

http://www.adastral.ucl.ac.uk/~vladkolm/software.html

Graph cut – submodularity

What energy functions E can be minimized by graph cut?

-- Augmenting paths terminates finitely if all costs +ve.



-- WLOG reduce to canonical form

define $D=g_{i,j}(0,1)+g_{i,j}(1,0) - g_{i,j}(0,0)-g_{i,j}(1,1)$

canonical form: $g_{i,j}(0,0)=g_{i,j}(1,1) \rightarrow 0;$

$g_{i,j}(0,1)=g_{i,j}(1,0) \rightarrow D/2;$

$f_i(x_i) \geq 0$

-- Solvability if $D \geq 0$ [invariant to transfms of problem]

Submodularity [Kolmogorov and Zabih 2004]

Regularity: achieving canonical form

[Kolmogorov and Zabih PAMI 2004]

Simplifying unary $f_i(x_i)$:

$m = \min(f_i(0), f_i(1))$

if $m<0$

$f_i(0) := f_i(0) - m$

$f_i(1) := f_i(1) - m$

-- reparameterization

??? Simplifying pairwise $g_{i,j}(x_i, x_j)$:

... Regularity: achieving canonical form

[Kolmogorov and Zabih 2004]

Simplifying pairwise $g_{i,j}(x_i, x_j)$:

$g(0,0)$	$g(0,1)$
$g(1,0)$	$g(1,1)$

→

0	$g(0,1) - g(0,0)$
$g(1,0) - g(0,0)$	$g(1,1) - g(0,0)$

→

0	0
$g(1,0) - g(0,0)$	$g(1,1) - g(0,1)$

→

0	0
$g(1,0) - g(0,0)$ $-g(1,1) + g(0,1)$	0

D

-- reparameterization

Graph cut – submodularity

Do they satisfy submodularity?

Ising prior $$V(\mathbf{x}) = \gamma \sum_{k,k'} |x_k - x_{k'}|$$?

$g(x,x')$	x = 0	x = 1
x' = 0	0	[illegible]
x' = 1	[illegible]	0

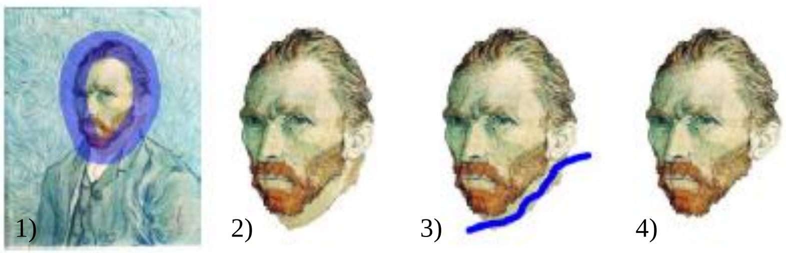
Ising with contrast sensitivity

$$V(\mathbf{x}, \mathbf{z}) = \gamma \sum_{k,k'} |x_k - x_{k'}| \left(\exp - \frac{\|z_k - z_{k'}\|^2}{2\sigma^2} \right)$$

?

Dynamic graph cut – Markov editing.

Boykov, Y., Jolly, M.P.: Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. Proc CVPR 2001.



Dynamic graph cuts

(Boykov and Jolly, 2001; Kohli & Torr ICCV 2005)

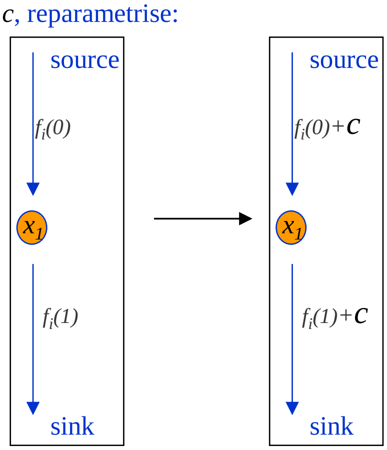
Changing the unaries:

Re-use old flow with new Energy fn ???problem

Where flow exceeds capacity by *c*, reparametrise:

$$f_i(0) := f_i(0) + c$$

$$f_i(1) := f_i(1) + c$$



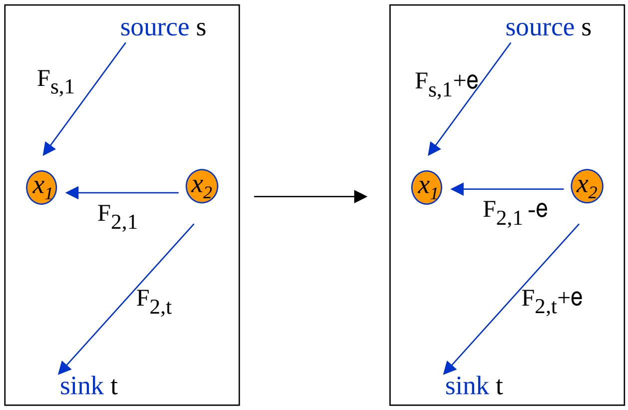
???Changing the pairwise potentials

Dynamic graph cuts

(Boykov and Jolly, 2001; Kohli & Torr ICCV 2005)

Changing the binaries:

Where flow, eg $F_{2,1}$ exceeds capacity by e, reparametrise:

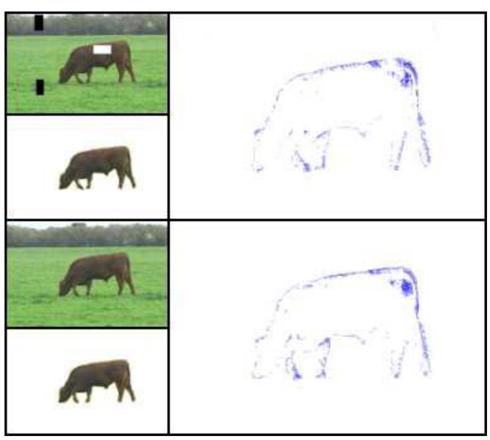


-- then follow procedure for unary flow excess

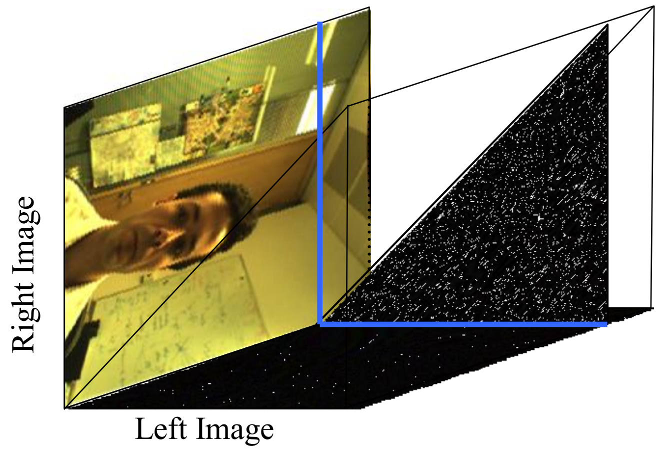
??? what's this good for

Dynamic graph cuts

(Kohli & Torr ICCV 2005)



Stereo matching – full depth



Stereo matching -- cost function

(Ohta & Kanade, 1985; Cox, Hingorani & Rao, 1996; Belhumeur 1996)

Minimize: $\min_{\mathbf{x}} F(\mathbf{x}, \mathbf{z})$

$$F(\mathbf{x}, \mathbf{z}) = \underbrace{\sum_i g_k(x_i, x_{i-1})}_{\text{log-prior } V(\mathbf{x})} + \underbrace{\sum_i f_k(x_i, x_{i-1}, \mathbf{z})}_{\text{log-likelihood } U(\mathbf{x},\mathbf{z})}$$

"Potts"

Note: x now integer valued

$x_i \in \{1, \ldots, , D\}$

max disparity

1D DP: Dijkstra

$\boxed{\text{?? 2D algorithm}}$

Nonlinear image restoration


Image restoration

data

$$p(\mathbf{x} \mid \mathbf{z}) \propto \exp -E$$

with $\quad E = U + V$

data $\qquad\qquad$ coherence

$$V(\mathbf{x}) = \sum_{i,j} g_{i,j}(x_i, x_j) \qquad\qquad U(\mathbf{x},\mathbf{z}) = \sum_i f_i(x_i, z_i)$$

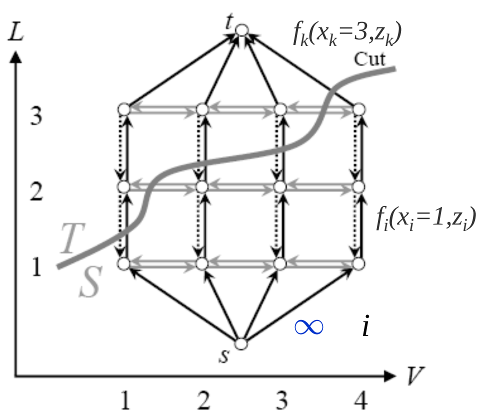
?? choice of *g* $\qquad\qquad$ noise model

Graph cut – expanded graph (Roy and Cox, ICCV 1998)

Problem graph of size *n* nodes, with *k* integer levels:

$$E(\mathbf{x},\mathbf{z}) = \sum_{i,j} g_{i,j}(x_i, x_j) + \sum_i f_k(x_i, z_i)$$

-- optimize by cutting graph of *nk+2* nodes



(Ishikawa, 03)

?? *g*

Convexity constraint (general connectivity)

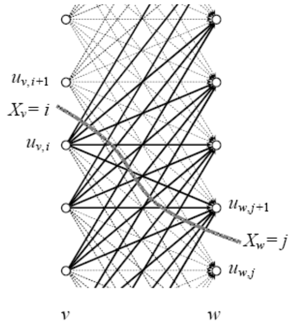
(Ishikawa 03)

$$E(\mathbf{x},\mathbf{z}) = \sum_{i,j} g_{i,j}(x_i, x_j) + \sum_i f_i(x_i, z_i)$$

Suppose:

$$g_{i,j}(x_i, x_j) = h_{i,j}(x_i - x_j)$$

then $h_{i,j}(.)$ must be convex



?? beyond convexity

Graph cut – α-expansion

Alternative optimizer for

$$E(\mathbf{x},\mathbf{z}) = \sum_{i,j} g_{i,j}(x_i, x_j) + \sum_i f_i(x_i, z_i)$$

with $x \in \{1,\ldots,L\}$

– repeated binary optimization:

-- cycle over α in the range of x :

$$\min_{\mathbf{y}} E(\mathbf{x},\mathbf{z}) \text{ where } y_i \in \{0,1\} \text{ and } x_i \to y_i\alpha + (1-y_i)x_i$$

?? escapes limitations on g

Submodularity – α-expansion

(Kolmogorov and Zabih, PAMI 2004)

Back to integer-valued problems: restoration, stereo, Tapestry, obj-rec with parts.

Given: $y_k \in \{0,1\}$ where $x_k \to y_k\alpha + (1-y_k)x_k$

$$g^y(y,y') = g(y\alpha + (1-y)x,\ y'\alpha + (1-y')x')$$

giving

$g^y(y,y')$		$y=0$	$y=1$
y'	0	$g(x,x')$	$g(x,\alpha)$
	1	$g(\alpha,x')$	$g(\alpha,\alpha)$

then regularity requires (diagonal subdominance):

$$g(x,x') + g(\alpha,\alpha) \le g(x,\alpha) + g(\alpha, x')$$

satisfied by any metric g (since: $g \ge 0$; $g(\alpha,\alpha)=0$; triangle ineq)

-- eg Potts

Submodularity – α-β swap

Alternative move – more expensive – more generality in model

Given: $y_k \in \{0,1\}$ where $x_k \in \{\alpha,\beta\} \to y_k\alpha + (1-y_k)\beta$

$$g^y(y_k,y_{k'}) = g(y\alpha + (1-y)\beta,\ y'\alpha + (1-y')\beta)$$

giving

$g^y(y,y')$		$y=0$	$y=1$
y'	0	$g(\beta,\beta)$	$g(\alpha,\beta)$
	1	$g(\beta,\alpha)$	$g(\alpha,\alpha)$

class of g ?

Energy bound for α-expansion

(Boykov, Veksler and Zabih, PAMI 2001)

Problem as usual:

$$E(\mathbf{x},\mathbf{z}) = \sum_{k,k'} g_{k,k'}(x_k, x_{k'}) + \sum_k f_k(x_k, z_k)$$

Global minimum $\mathbf{x}^*$

Local min under α-expansion: $\tilde{\mathbf{x}}$

Energy bound $E(\tilde{\mathbf{x}},\mathbf{z}) \leq 2cE(\mathbf{x}^*,\mathbf{z})$

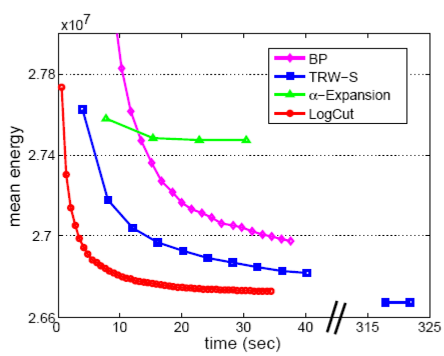
where $c = \max_{k,k'} \left(\frac{\max_{u,v} g_{k,k'}(u,v)}{\min_{u,v} g_{k,k'}(u,v)} \right)$

?? bound for Potts

Energy bound – is it useful??

(Lempitsky, Rother and Blake, ICCV 2007)

Sample restoration problem:



a) image restoration - model 1

$E(\tilde{\mathbf{x}},\mathbf{z}) \leq 2cE(\mathbf{x}^*,\mathbf{z})$

?? useful bound

Results of stereo matching

(Woodford, Torr, Reid, Fitzgibbon 2008)