CVPR 2013 TUTORIAL:
A CRASH COURSE ON VISUAL SALIENCY MODELING:
BEHAVIORAL FINDINGS AND COMPUTATIONAL MODELS

SALIENCY: INFORMATION THEORETIC, BAYESIAN, AND GRAPHICAL MODELS

Neil Bruce

Department of Computer
University of Manitoba
Winnipeg, MB, Canada

Contact Information:

bruce@cs.umanitoba.ca
www.cs.umanitoba.ca/~bruce

UNIVERSITY OF MANITOBA

Overview

- Motivating ideas: A summary
- Information theoretic approaches
  - Some of my work (AIM)
- Coding based approaches
- Graph-based strategies
- Bayesian approaches
- Summary and Discussion

Some basic background

- Entropy

$$H(X) = \mathbb{E}_X[I(x)] = -\sum_{x \in \mathbb{X}} p(x) \log p(x).$$

- Joint entropy

$$H(X,Y) = \mathbb{E}_{X,Y}[-\log p(x,y)] = -\sum_{x,y} p(x,y) \log p(x,y)$$

- Conditional entropy

$$H(X|Y) = \mathbb{E}_Y[H(X|y)] = -\sum_{y \in Y} p(y) \sum_{x \in X} p(x|y) \log p(x|y) = -\sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(y)}.$$

- Mutual information

$$I(X;Y) = \mathbb{E}_{X,Y}[SI(x,y)] = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)}$$

- Can all be expressed in terms of KL-divergence

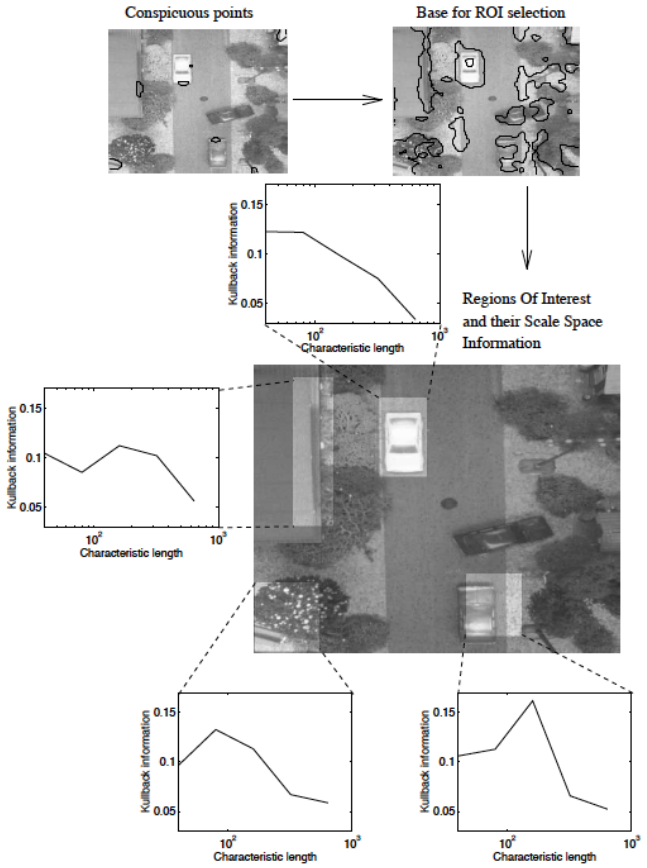
Some basic background

- Other important ideas
  - Suspicious Coincidences (Barlow)
    - One goal of the brain is detecting associations
    - Find suspicious coincidences, and anticipate them
  - Coding theory
    - Rate/Distortion
    - Data compression
    - Redundancy
  - Bayes' Theorem

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}.$$

Computer Vision Classics

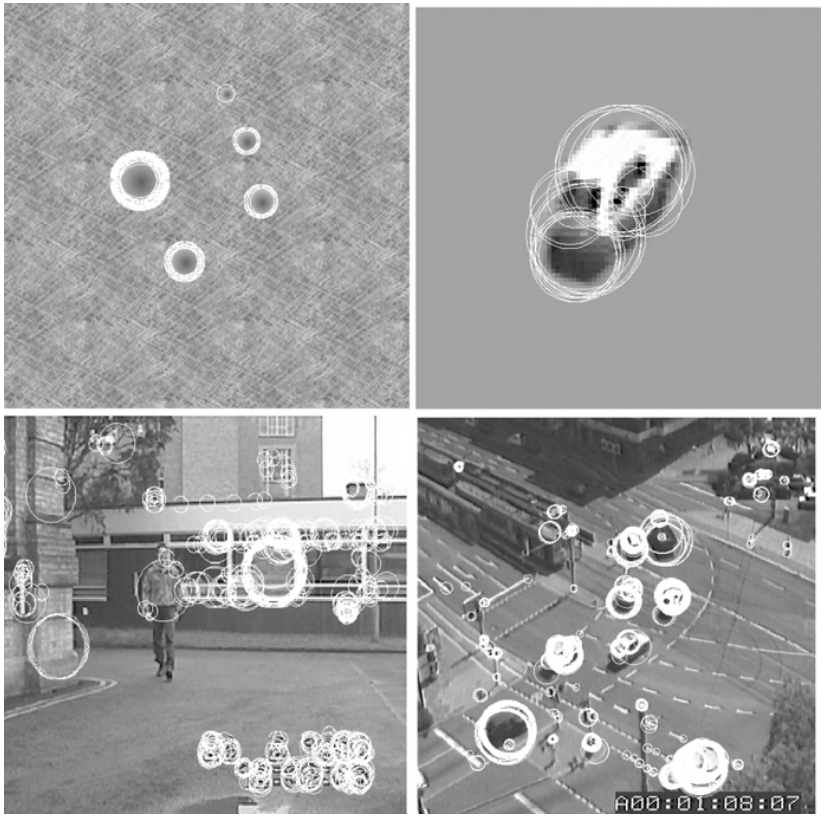
- Jagersand (ICCV 1995)



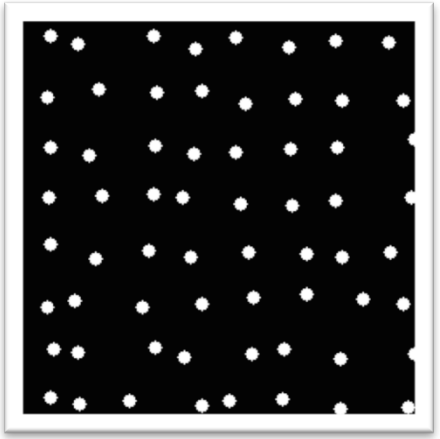
- KL-divergence across scale space representation
- Peaks provide a sense of scale of interest

Computer Vision Classics

- Kadir and Brady (IJCV 2001)
- Again appealing to peaks in scale space subject to local entropy
- Some extra steps (region clustering, etc…)
- Area has been further explored in detail by interest point/descriptor research community



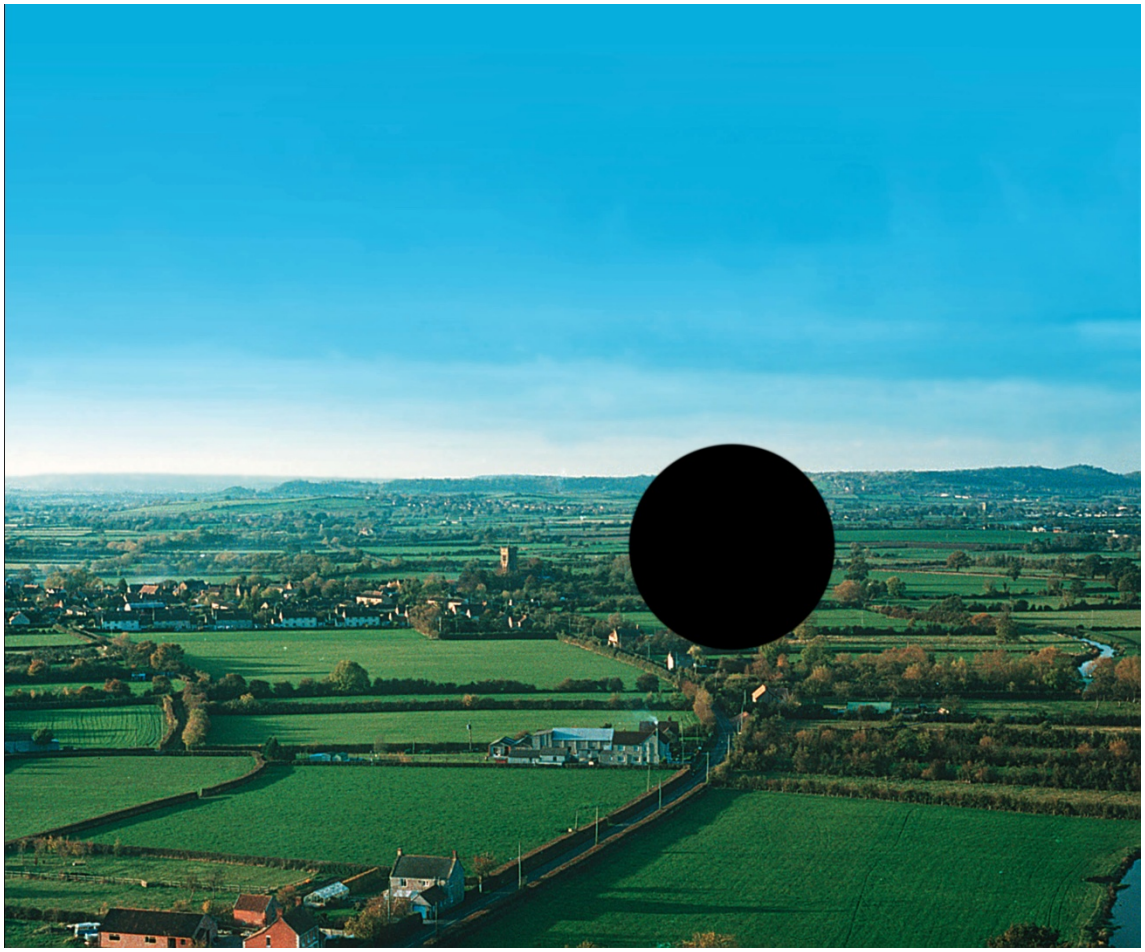
Saliency



Visual content and expectation

Visual content and expectation

Visual content and expectation



Visual content and expectation

Information Theoretic Saliency

A note on “categorization”

- Information theory, coding, Bayesian inference, Graphical models aren’t easily separated
- Grouped thematically, but several may be present within any single model

Source: Unknown

AIM (Attention by Information Maximization)

- Appeals to role of coding, and information theory
- Key points:
  - Independent (sparse) coding
  - Want to quantify likelihood of observing local patch/ region of image
- Likelihood related to self-information via –log(p(x))

Computational Constraints

- $p(X = x_1, x_2, \ldots, x_n)$
- Problem of dimensionality: More general problem

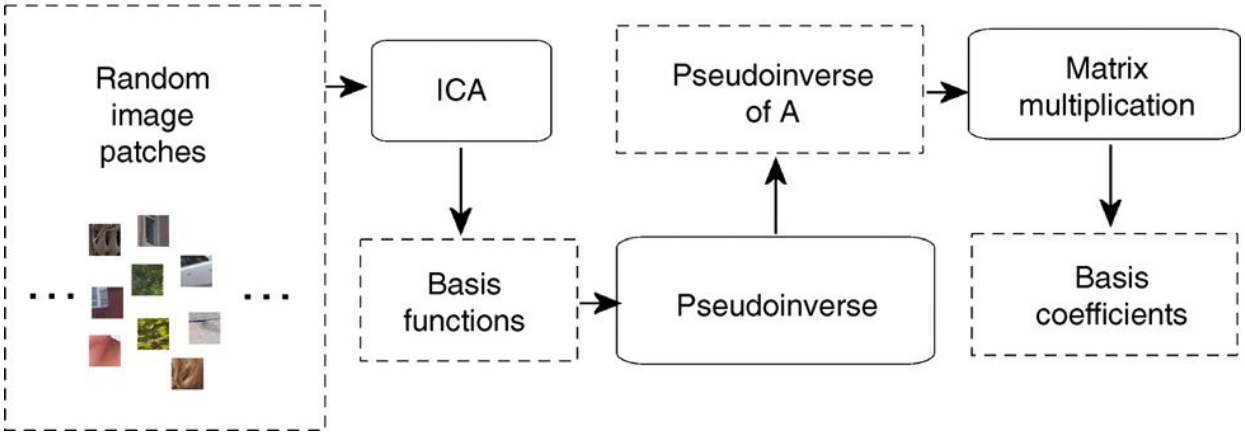
x_1	x_2	x_3	x_4	x_5
x_6	x_7	x_8	x_9	x_{10}
x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
x_{16}	x_{17}	x_{18}	x_{19}	x_{20}
x_{21}	x_{22}	x_{23}	x_{24}	x_{25}

- Seek a representation that makes the computation of $p(X)$ tractable
- Content is not random but highly structured and exists in a lower dimensional manifold
- Solution: Take advantage of structure in the data - Sparse representation as one approach
- Information appeals to Shannon self-information

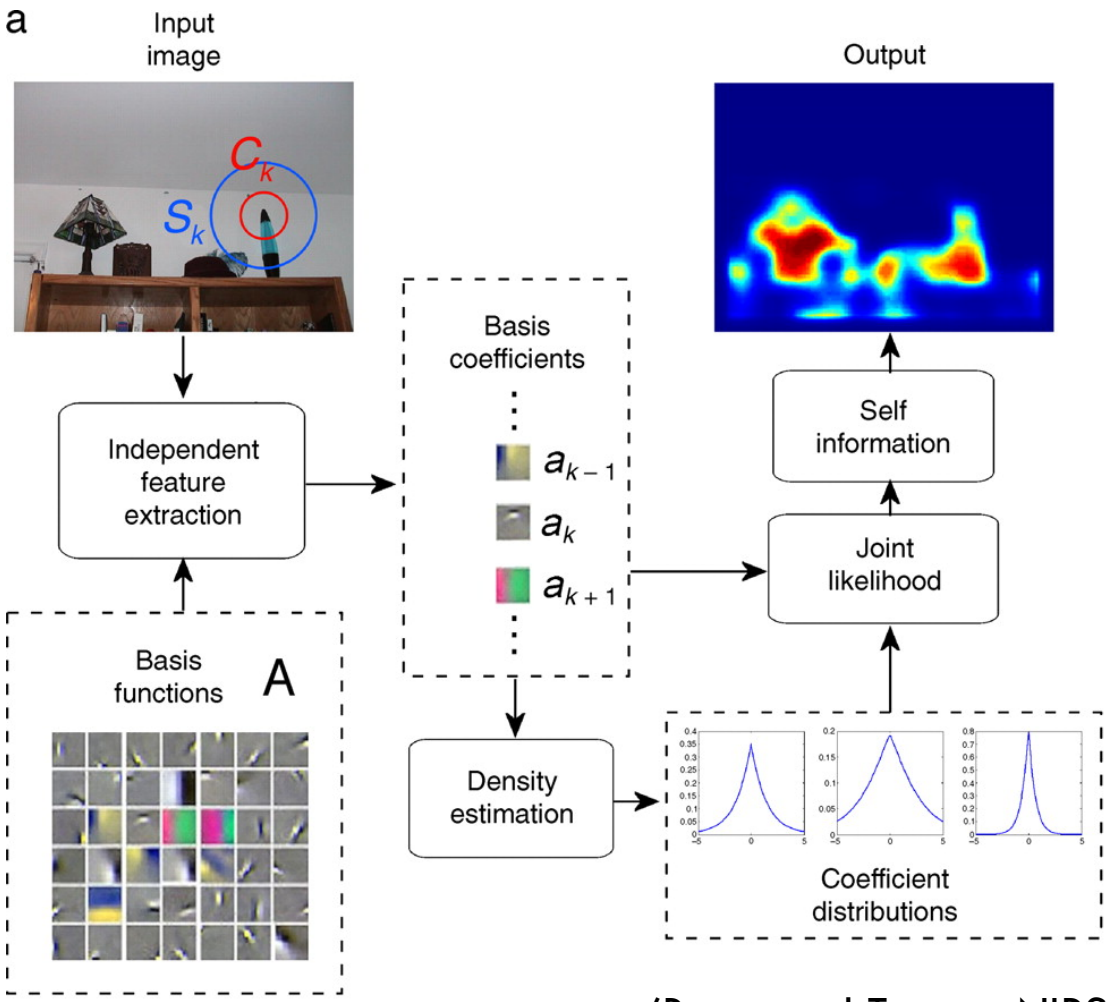
Computational Constraints

$$p(x_1, x_2, \ldots, x_n) = \prod_i p(x_i)?$$

- Ideally $x_1, x_2, \ldots, x_n$ should be independent variables



The Model (AIM)

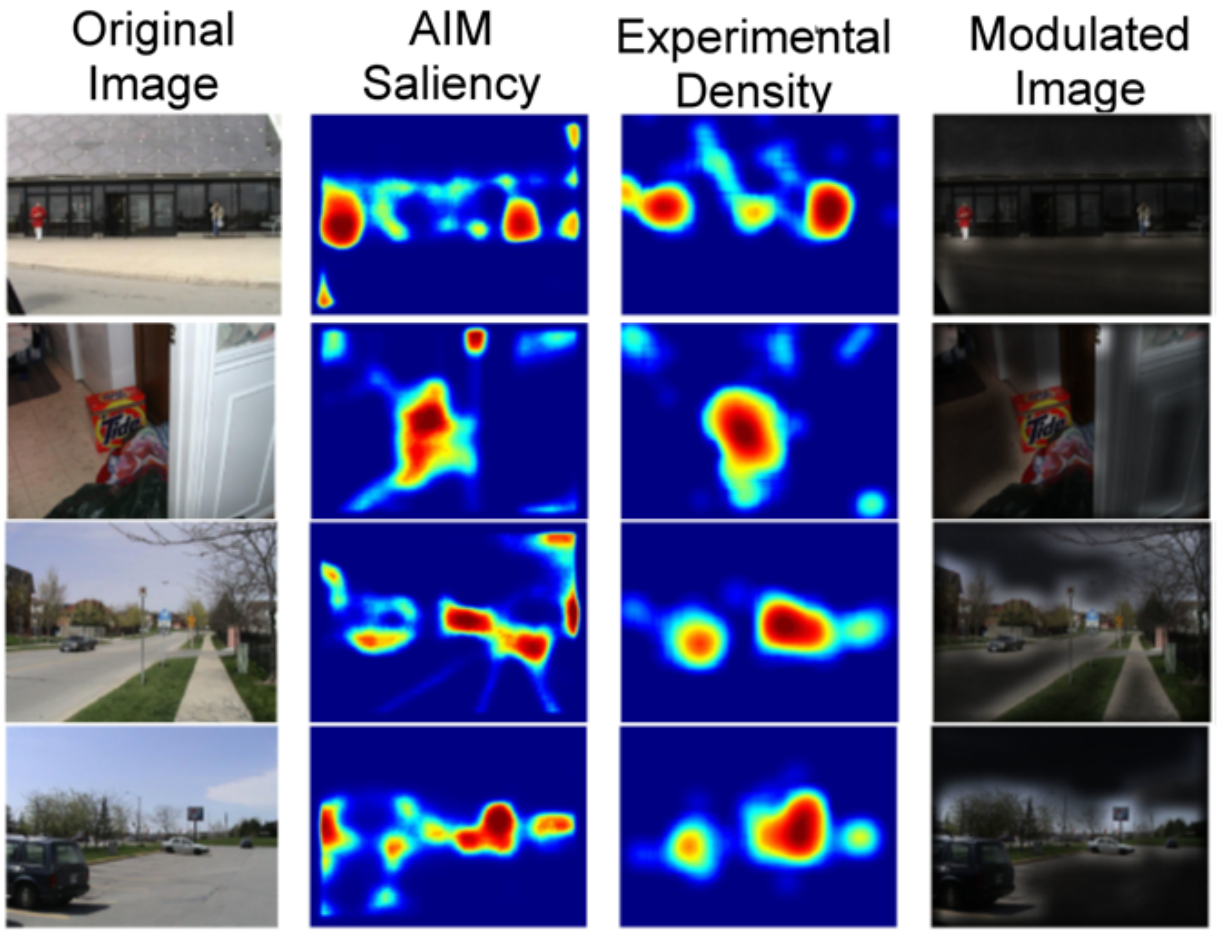


(Bruce and Tsotsos, NIPS 2005, JoV 2009)

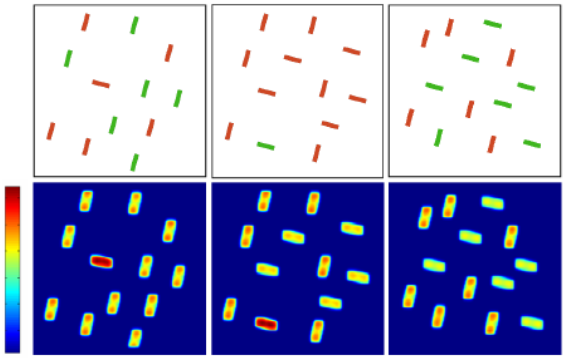
Quantifying Performance

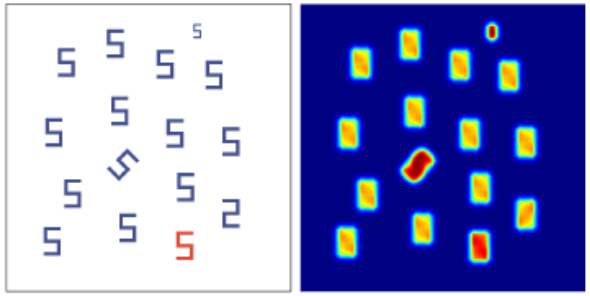
- There now exist a number of datasets, benchmarks, performance metrics, etc.
  - Benchmarking will be discussed later ☺
- Different data sets, methodology and parameters
- Seems to hold up well in benchmarks, despite remaining largely untouched for 8 years…

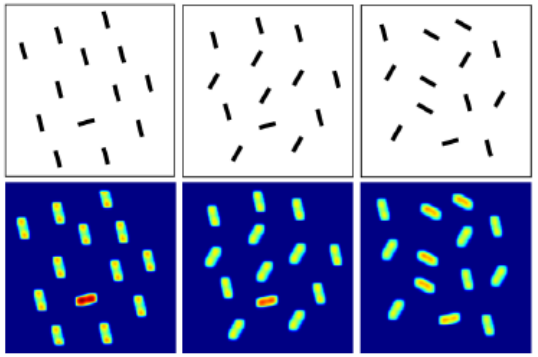
Prediction of fixation patterns

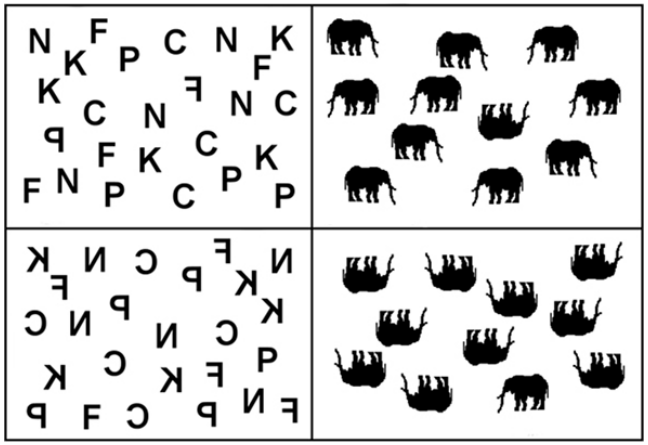


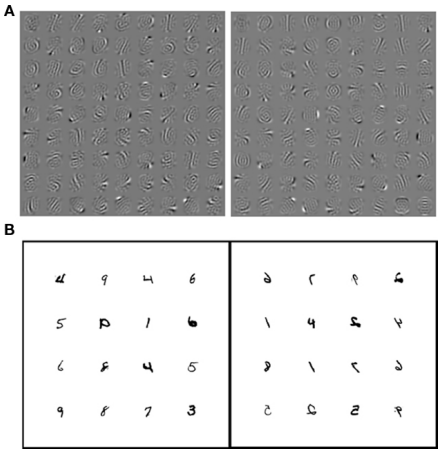
Behavioral phenomena

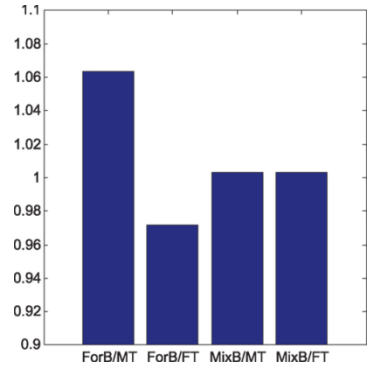






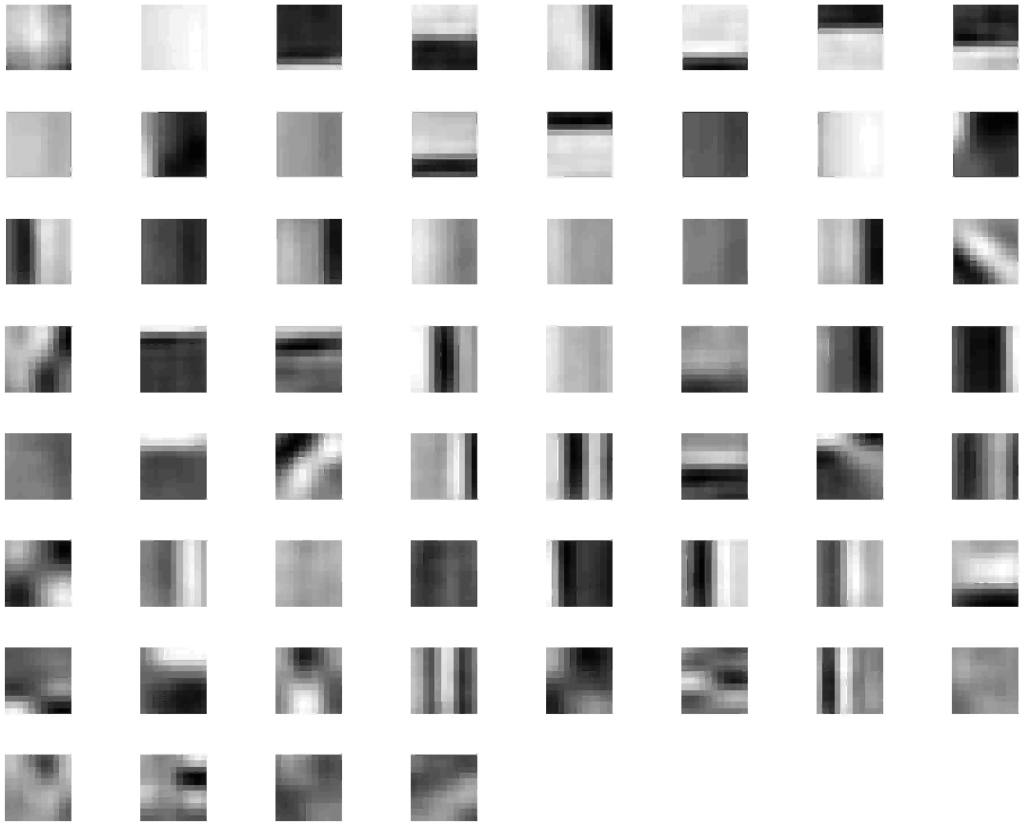




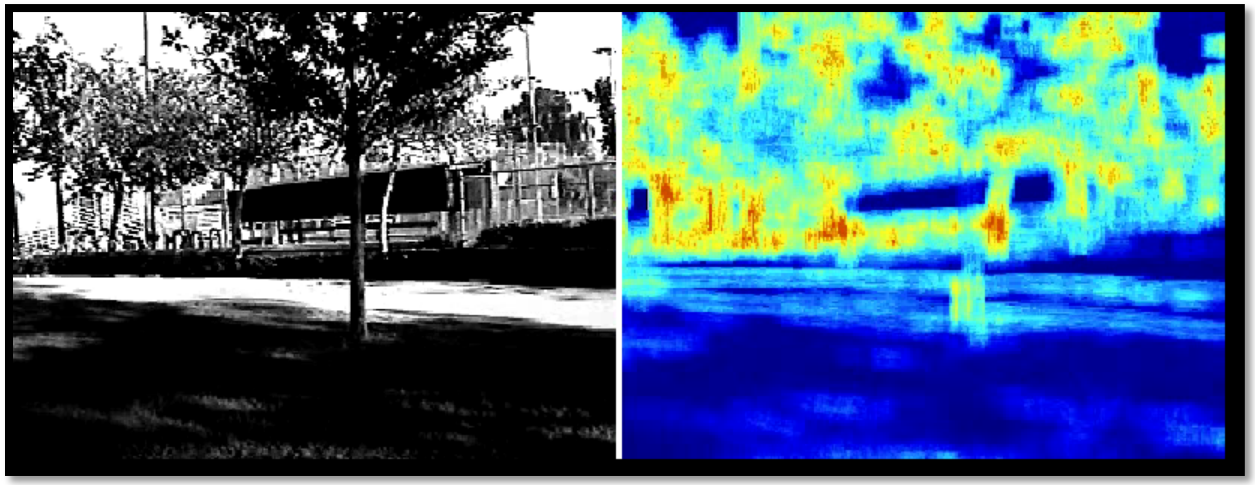


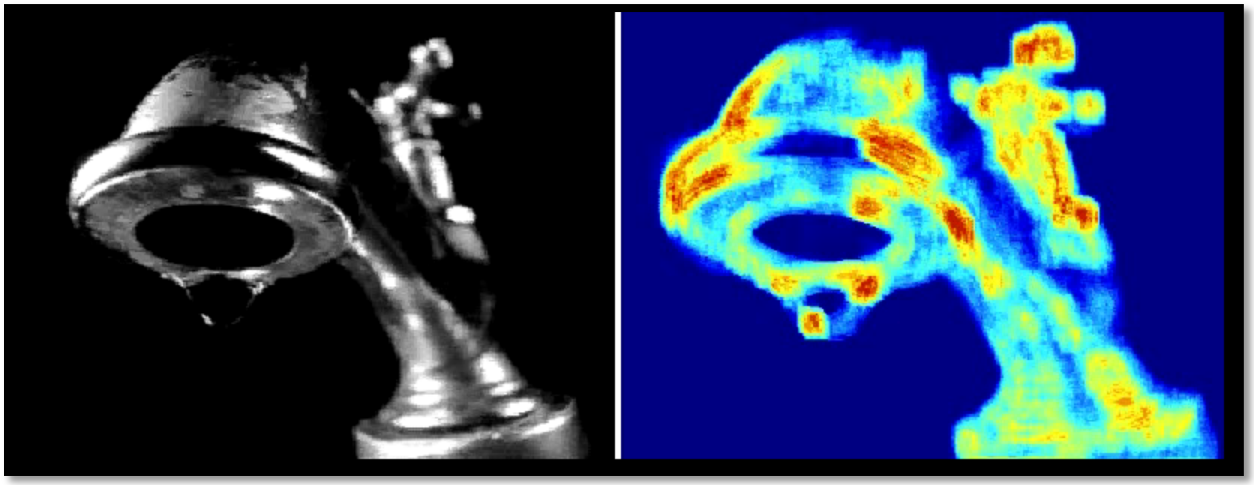
(Bruce and Tsotsos, 2009,
Bruce and Tsotsos 2011)

Spatiotemporal Cells



Examples





Many facets of AIM

- Role of feature representation (performance and neuroscience)

JoV 2009, CRV 2011, Front. Comp. Neurosci. 2011

- Role of context in defining algorithm output

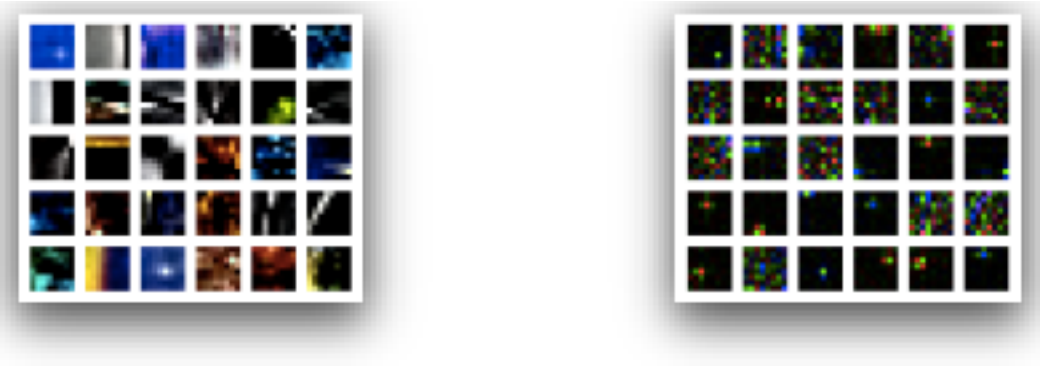
ICIP 2009, ICPR 2004

- Recurrence/hierarchical processing

(CRV 2012, ICIAR 2012, CVPR Workshop Biol. Consistent Vision, 2011)

Incremental Coding Length

- Hou and Zhang (NIPS 2008)
  - Measure entropy gain of each feature
  - Maximize entropy across sample features
  - Select features with large coding length increment



A W

$$\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{192}]^\top$$

Incremental Coding Length

- Activity ratio p_i for i^{th} feature: $p_i = \frac{\sum_k | \mathbf{w}_i \mathbf{x}^k |}{\sum_i \sum_k | \mathbf{w}_i \mathbf{x}^k |}$

  with $\mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k, \ldots]$

- Most efficient strategy is to make equal use of all features – i.e. maximize H(p)

- New excitation of feature i: $\hat{p}_j = \begin{cases} \frac{p_j + \varepsilon}{1 + \varepsilon}, & j = i \\ \frac{p_j}{1 + \varepsilon}, & j \neq i \end{cases}$

Incremental Coding Length

- Changes entropy of feature activities:

$$\frac{\partial H(\mathbf{p})}{\partial p_i} = -\frac{\partial p_i \log p_i}{\partial p_i} - \frac{\partial \sum_{j \neq i} p_j \log p_j}{\partial p_i}$$

- Which boils down to:

$$\mathrm{ICL}(p_i) = \frac{\partial H(\mathbf{p})}{\partial p_i} = -H(\mathbf{p}) - p_i - \log p_i - p_i \log p_i$$

Incremental Coding Length

- Salient feature set:

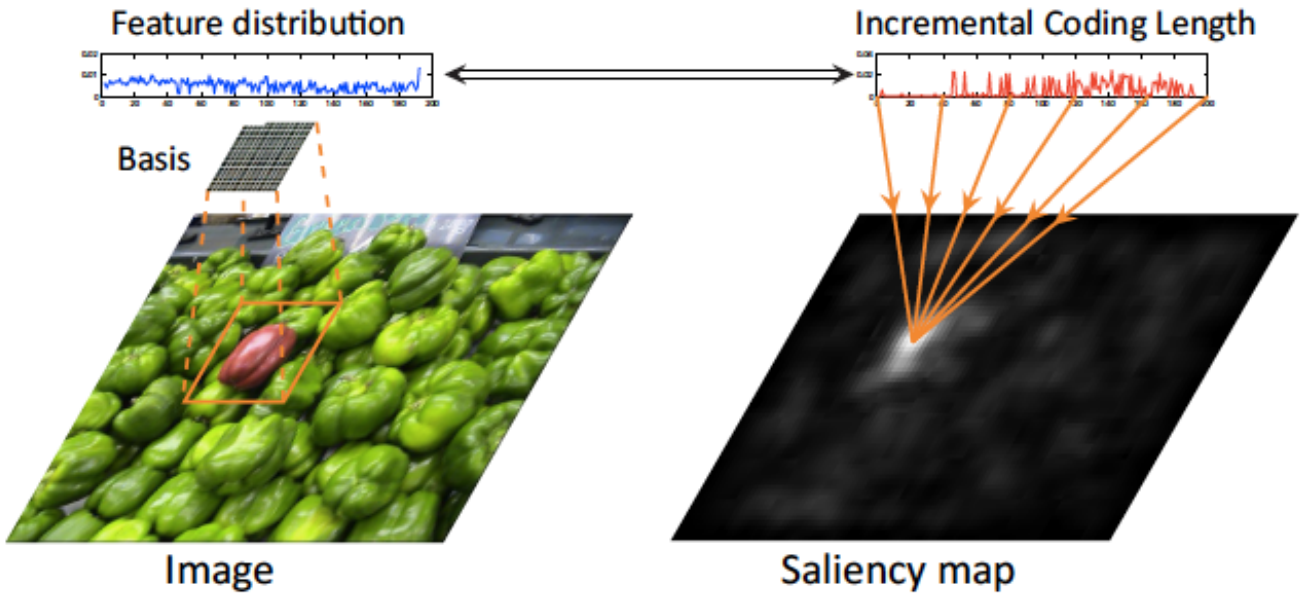
$$\mathcal{S} = \{i \mid \mathrm{ICL}(p_i) > 0\}$$

  i.e. Do subsequent activations of feature i increase entropy of the system?

- Allows redistribution of energy among salient features:

$$d_i = \frac{\mathrm{ICL}(p_i)}{\sum_{j \in \mathcal{S}} \mathrm{ICL}(p_j)}$$

Incremental Coding Length



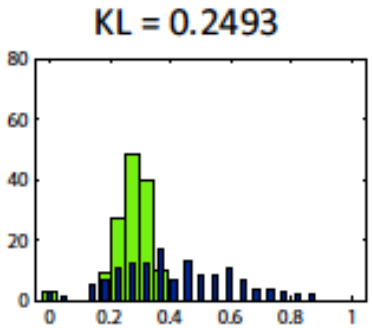
Finally, salience may be computed, with $M=[m_1,m_2,\ldots,m_n]$

$$\mathbf{m}_k = \sum_{i \in \mathcal{S}} d_i \mathbf{w}_i \mathbf{x}^k$$

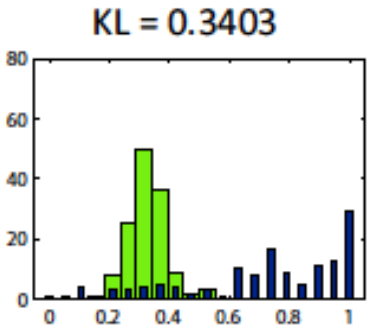
Incremental Coding Length

$$\mathbf{p}^t = \frac{1}{Z} \sum_{\tau=0}^{t-1} \exp\left(\frac{\tau - t}{\lambda}\right) \cdot \hat{\mathbf{p}}^\tau$$

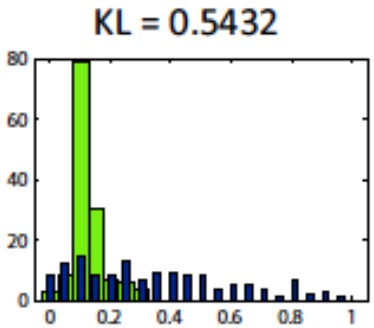
A: input sample



B: model in [7]



C: model in [5]



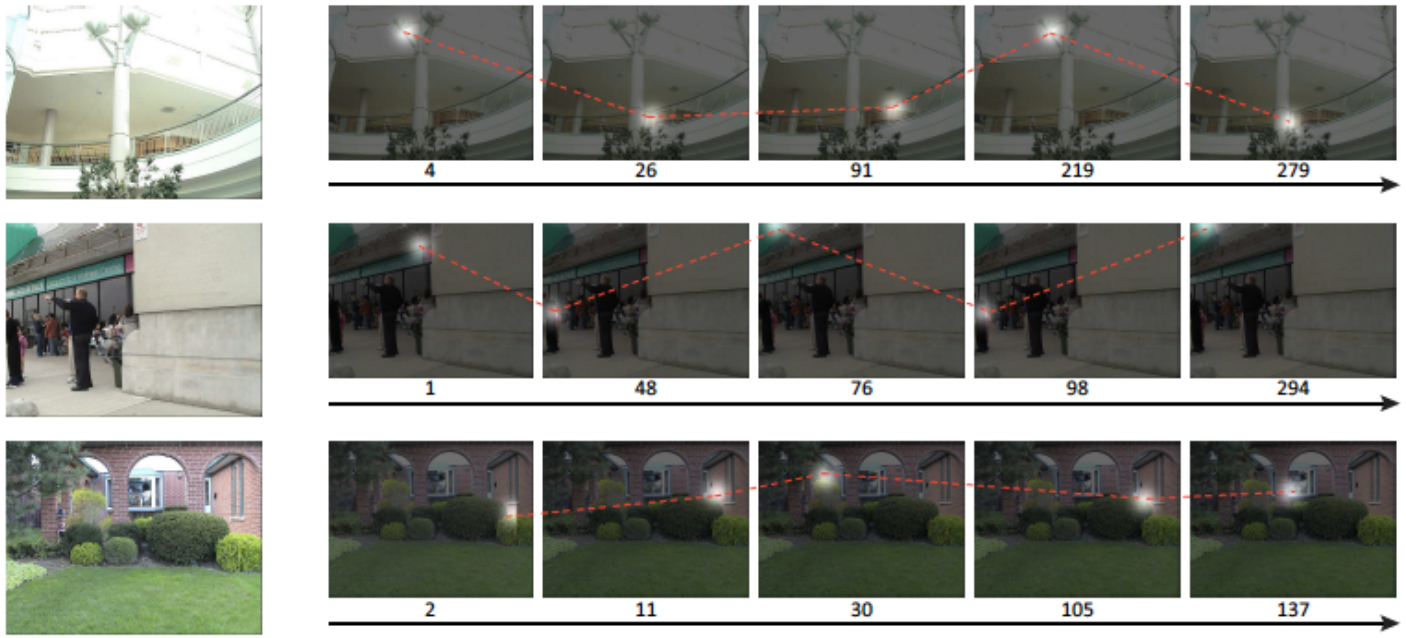
D: our model

- Spatiotemporal data

Incremental Coding Length

Dynamic Visual Attention

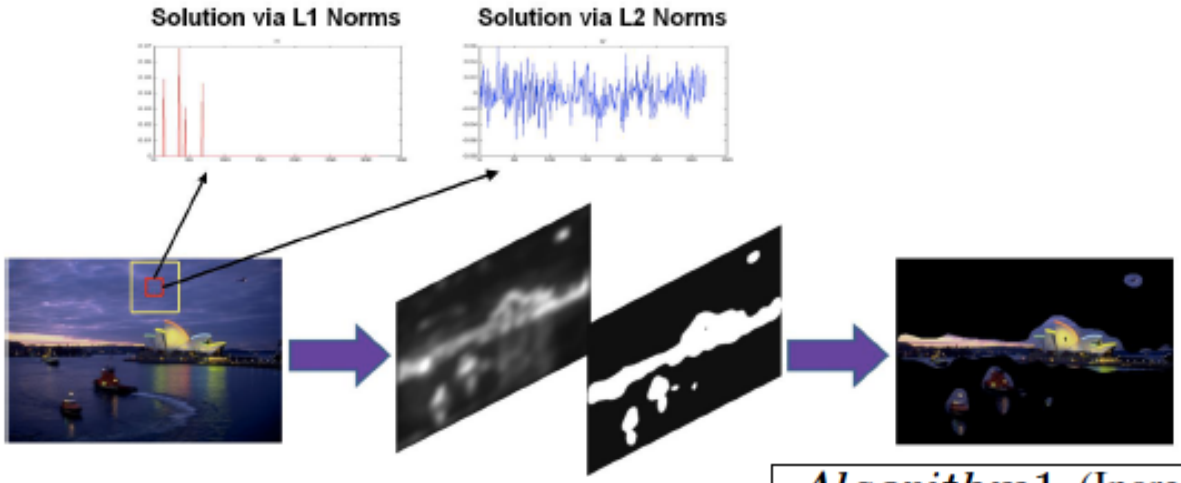
1. At time t, calculate feature ICL based on $\mathbf{p}^t$
2. Given current eye fixation, generate a saliency map with foveal bias.
3. By a saccade, move eye to the global maximum of the saliency map.
4. Sample top N "informative" (largest ICL) features in fixation neighborhood. (In our experiment, $N = 10$)
5. Calculate $\hat{\mathbf{p}}^t$, update $\mathbf{p}^{t+1}$, and go to Step. 1.



Conditional Entropy

- Li, Zhou, Yan and Yang (ACCV 2009)
  - Saliency based on conditional entropy
  - Minimum uncertainty of local region given surround
  - Conditional entropy given by coding length (assuming lossy distortion) modeled as multivariate Gaussian data
  - Segmentation to detect proto-objects
  - Extended by Yan et al. to multi-resolution

Conditional Entropy



Algorithm1 (Incremental Sparse Saliency)

1.***Input*** : given image I

2.***for*** each patch c of the image I, calculate $\boldsymbol{x} = Fc$ and take patches from its surroundings to form $\boldsymbol{S}$

- solve the optimization problem $min\ \lambda\|\boldsymbol{w}\|_1 + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{S}\boldsymbol{w}\|_2^2$
- given the sparse solution $\boldsymbol{w}$, calculate the patch saliency $Sa(c)$ by $Sa(c) = \|w\|_0$, and accumulate the saliency by pixels

3.***end***

4.***Output*** : the saliency map of I

Probability/Clutter

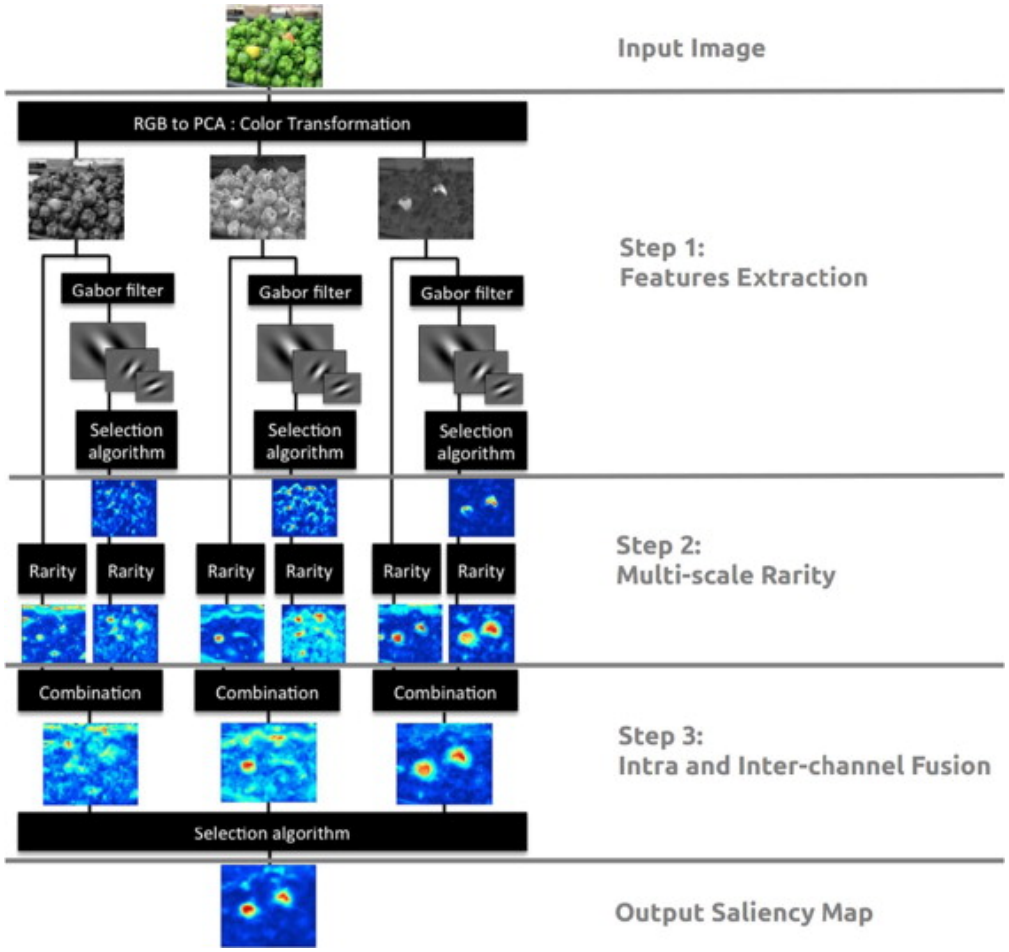
- Rosenholtz (Vis. Res. 1999, JoV 2007, ACM TAP 2011)
  - Distribution of features determined (e.g. in color-space)
  - Mean and covariance of distractors computed
  - Target saliency given by Mahalanobis distance given target, and mean/covariance of distractor distribution
  - Later versions also account for role of "clutter"

Rarity Based Saliency

- Mancas (2007)
  - Considers rarity of features (both local and global, including subject to self-information)
  - Multi-scale approach reminiscent of Itti et al.
  - Also consider many applications
- Mancas (2012) (RARE)
  - Normalization/Whitening across color inputs and across scale, weighted combination/fusion

$$Attention(I_j) = -\log\left(\frac{1}{S * |I_j|}\sum_{i=1}^{S} n_i\right)$$

RARE 2012


Input Image
RGB to PCA : Color Transformation
Gabor filter
Gabor filter
Gabor filter
Selection algorithm
Selection algorithm
Selection algorithm
Step 1:
Features Extraction
Rarity
Rarity
Rarity
Rarity
Rarity
Rarity
Step 2:
Multi-scale Rarity
Combination
Combination
Combination
Step 3:
Intra and Inter-channel Fusion
Selection algorithm
Output Saliency Map

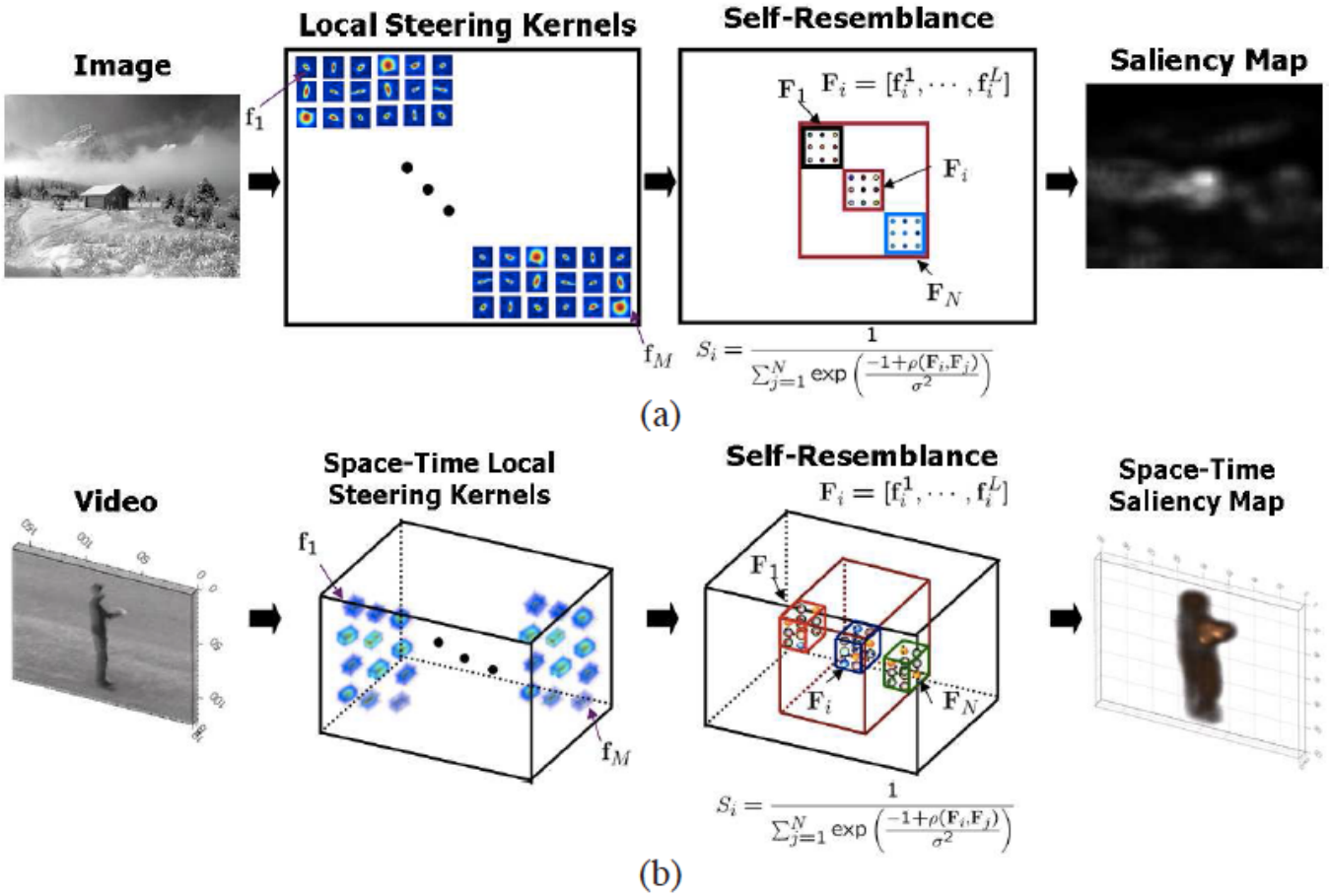
Self-resemblance

- Seo and Milanfar (Journal of Vision, 2009)
  - Local structure represented by matrix of local descriptors (steering kernels robust to noise/image distortions)
  - Matrix cosine similarity forms a metric for resemblance at pixel to surround
  - Amounts to an estimate of likelihood of local feature matrix given feature matrix of pixels in surround

$$K(\mathbf{x}_l - \mathbf{x}_i) = \frac{\sqrt{\det(\mathbf{C}_l)}}{h^2} \exp\left\{\frac{(\mathbf{x}_l - \mathbf{x}_i)^T \mathbf{C}_l (\mathbf{x}_l - \mathbf{x}_i)}{-2h^2}\right\}, \quad \mathbf{C}_l \in \mathbb{R}^{2\times 2}$$

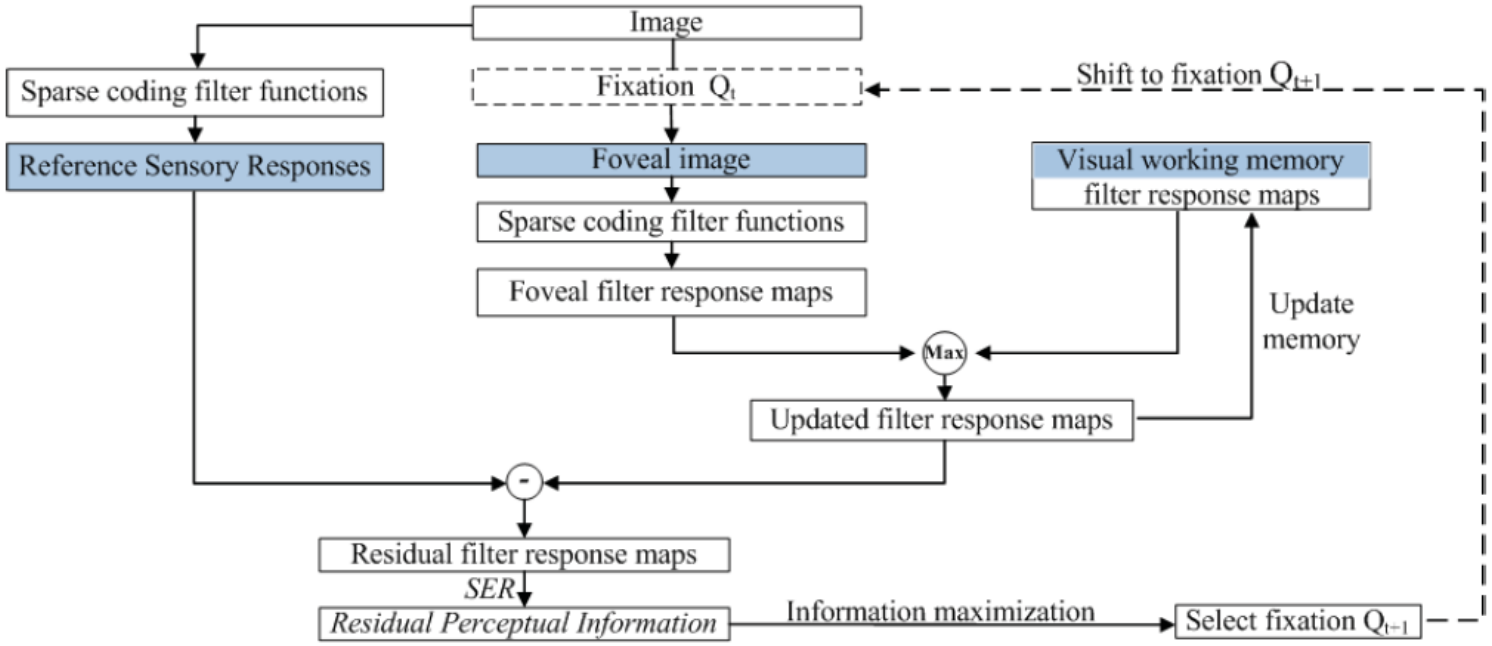
$$S_i = \frac{1}{\sum_{j=1}^{N} \exp\left(\frac{-1+\rho(\mathbf{F}_i, \mathbf{F}_j)}{\sigma^2}\right)}$$

Self-resemblance



Site-Entropy Rate

- Wang et al. CVPR 2011 (following Wang et al. CVPR 2010)



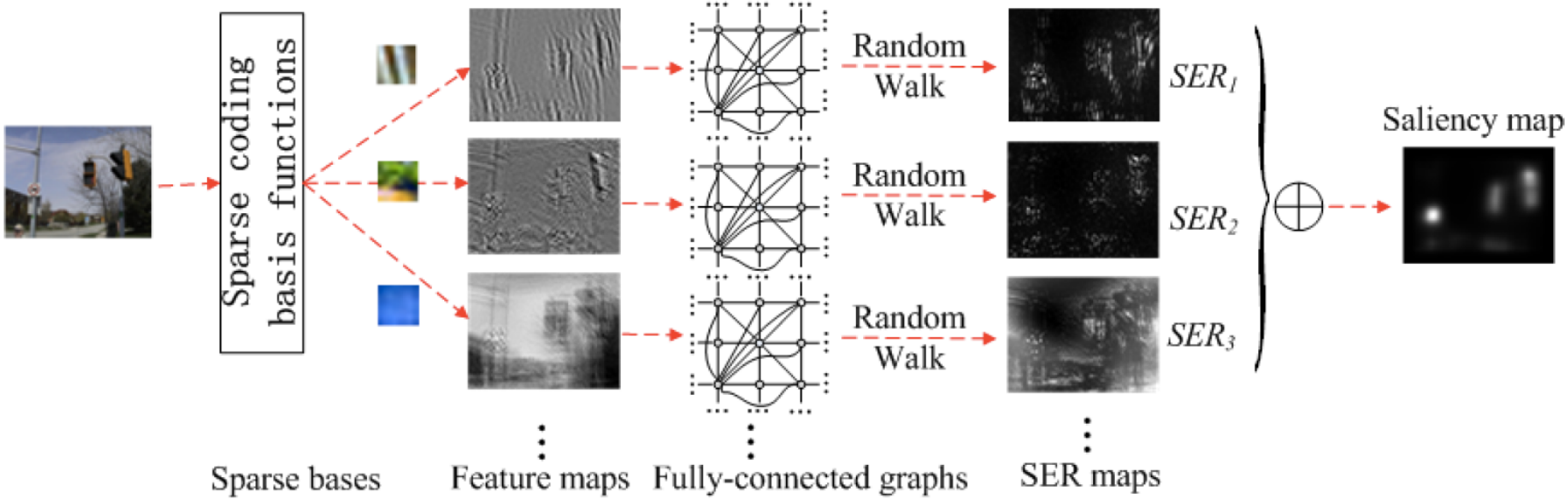
Site-Entropy Rate

- Wang et al. CVPR 2011 (following Wang et al. CVPR 2010)
- Average total information transmitted from location l to other nearby locations

$$S_i = \sum_k SER_{ki} = -\sum_k (\pi_{ki} \sum_j P_{kij} \log P_{kij})$$

π_i - Stationary distribution term (frequency with which random walker visits node i / frequency with which node i communicates with other nodes)

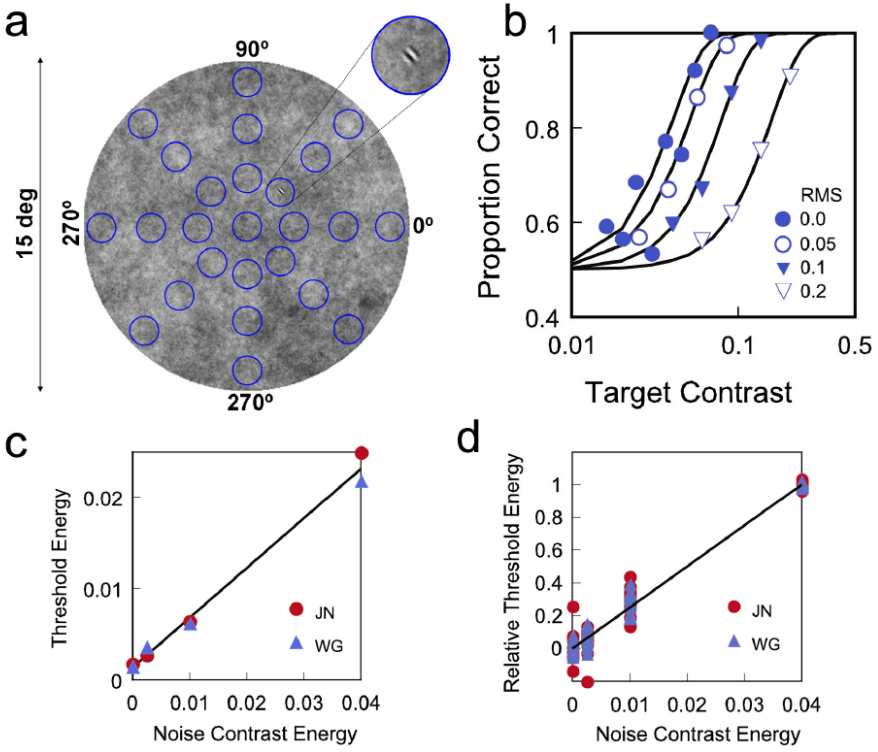
Site-Entropy Rate



- Random walks: See also Achanta et al. 2009

Information Gain

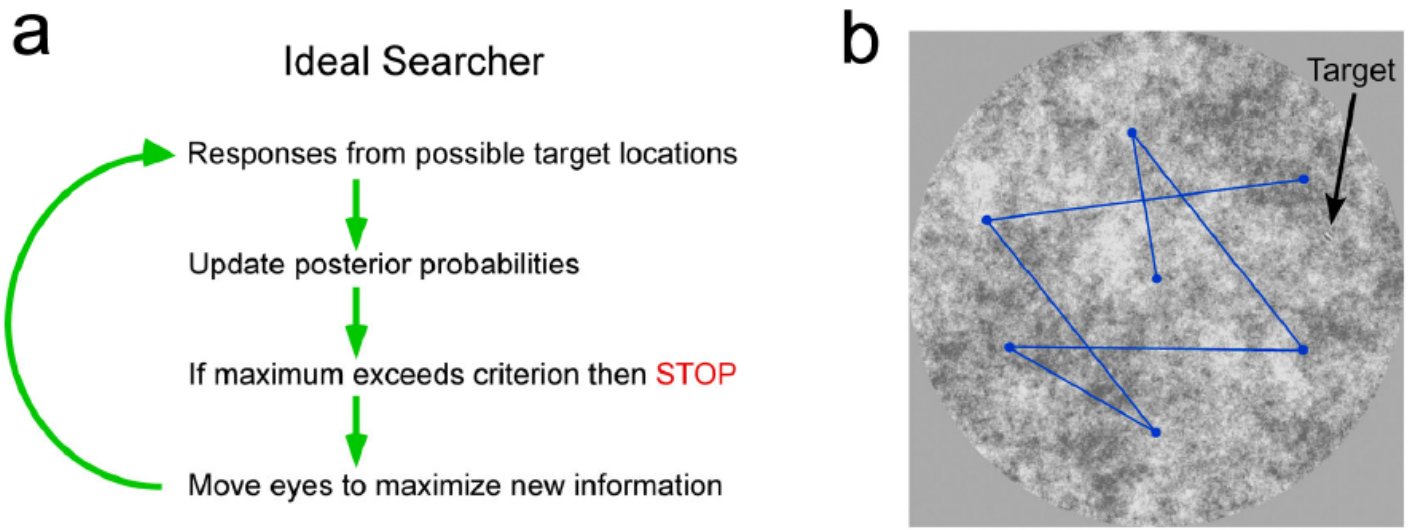
- Najemnik and Geisler



- The "ideal observer"
  - Subject to simulated constraints/uncertainty on perception
  - Wish to maximize the information gain, or minimize uncertainty with respect to defined target location in making a saccade

Information Gain

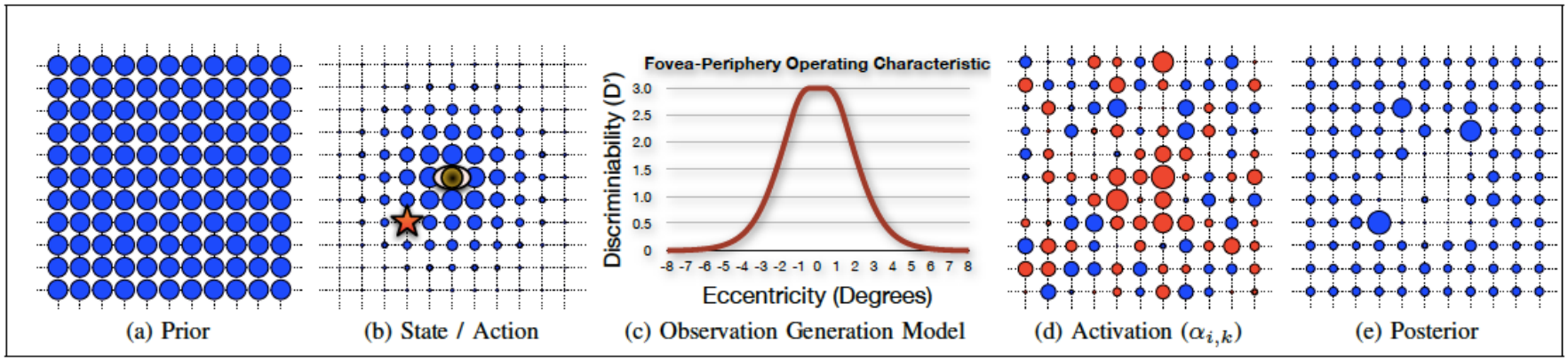
- Najemnik and Geisler (Nature 2005)

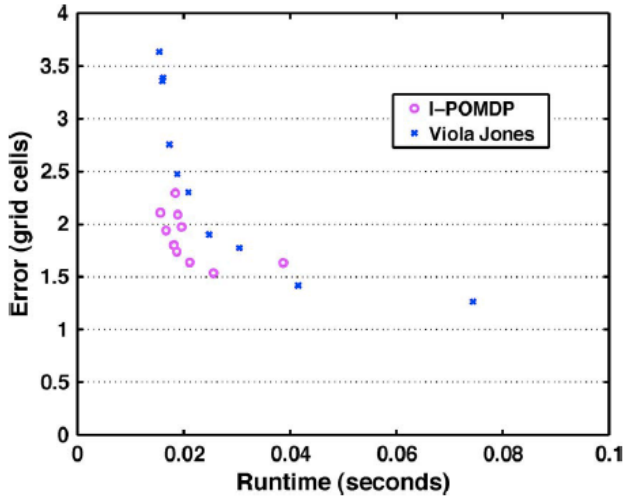


$$p_i = \frac{prior(i)\exp\left(d_i'^2 W_i\right)}{\sum_{j=1}^{n} prior(j)\exp\left(d_j'^2 W_j\right)}$$

Information Gain

- Butko and Movellan, ICDL 2008, IEEE TAMD 2010

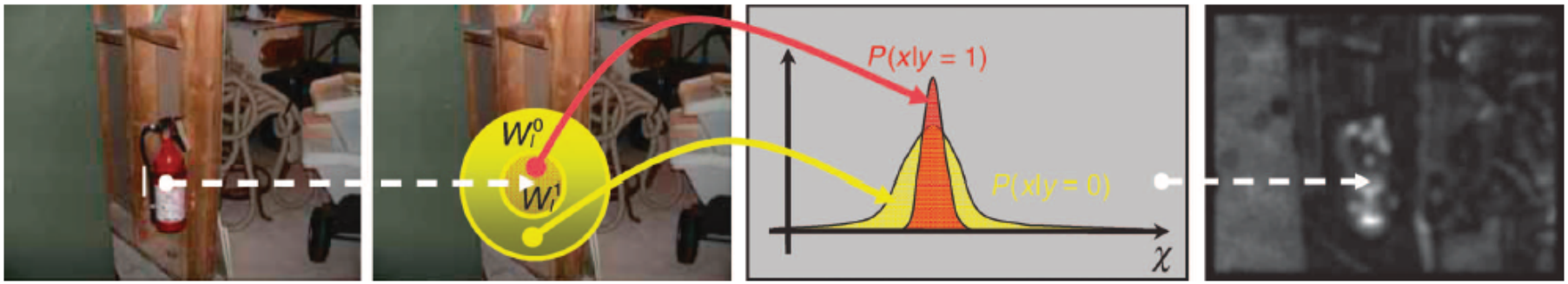




Discriminant / Decision Theoretic Saliency

- Spatial definition for "c"

$$S(l) = I_l(\mathbf{X}; Y) = \sum_c \int p_{\mathbf{X}(l),Y(l)}(\mathbf{x}, c) \log \frac{p_{\mathbf{X}(l),Y(l)}(\mathbf{x}, c)}{p_{\mathbf{X}(l)}(\mathbf{x}) p_{Y(l)}(c)} d\mathbf{x}.$$



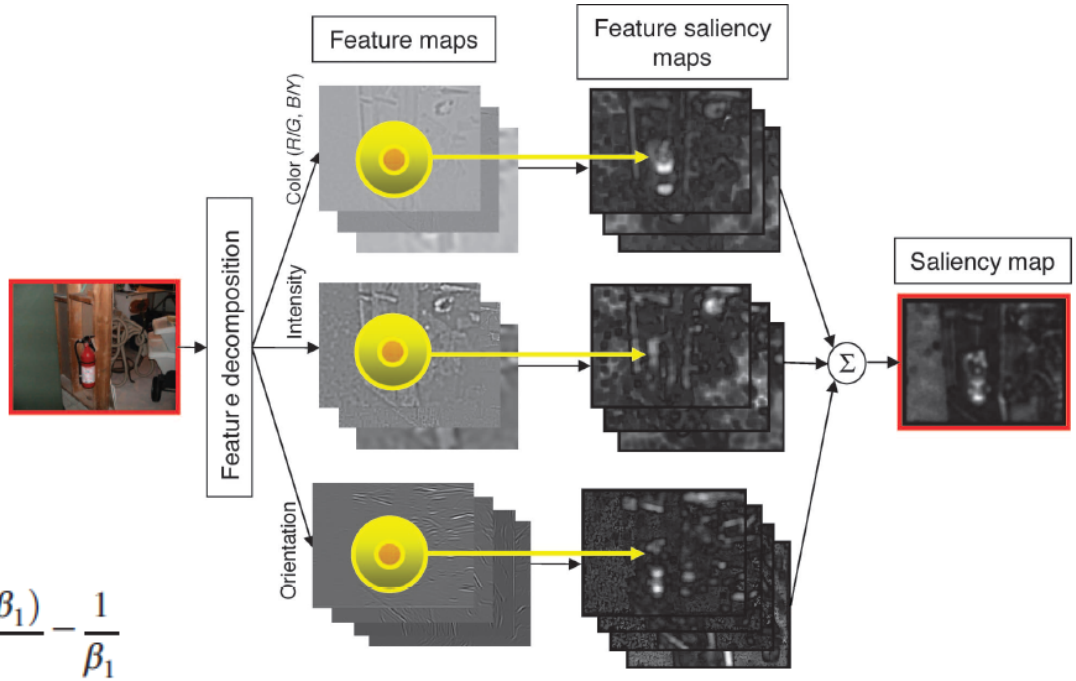
Decision Theoretic Saliency

- Diagrams images

$$P_X(x;\alpha,\beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)}\exp\left\{-\left(\frac{|x|}{\alpha}\right)^{\beta}\right\}$$

$$I(X;Y) = \sum_c P_Y(c)KL\left[P_{X|Y}(x|c)\,||P_X(x)\right],$$

$$\mathrm{KL}[P_X(x;\alpha_1,\beta_1)\,||P_X(x;\alpha_2,\beta_2)]$$
$$= \log\left(\frac{\beta_1\alpha_2\Gamma(1/\beta_2)}{\beta_2\alpha_1\Gamma(1/\beta_1)}\right) + \left(\frac{\alpha_1}{\alpha_2}\right)^{\beta_2}\frac{\Gamma((\beta_2+1)/\beta_1)}{\Gamma(1/\beta_1)} - \frac{1}{\beta_1}$$

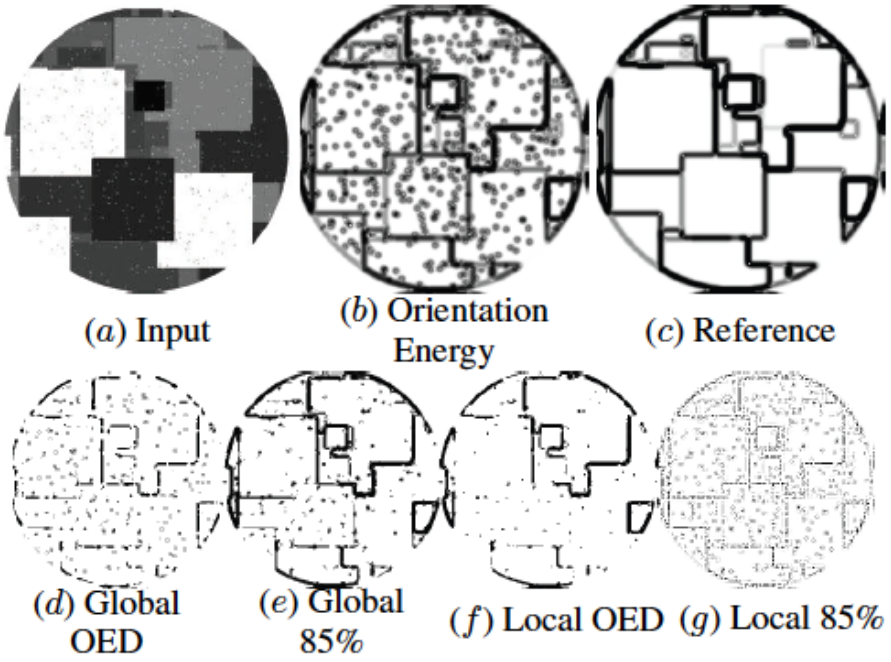


Discriminant / Decision Theoretic Saliency

- Derived explicitly from a minimum Bayes error definition
- "c" applicable to centre/surround, but also other classes (e.g. face vs. null hypothesis)
- Specific mathematical relationship can be shown to:
  - Suspicious coincidences, decision theory, neural computation/complex cells/circuitry, tracking
- See: Han and Vasconcelos Vis. Res. 2010, Mahadevan and Vasconcelos, TPAMI 2010, Gao et al. IEEE TPAMI 2009, Gao and Vasconcelos Neur. Comp 2009, Gao, Mahadevan and Vasconcelos, 2007, Gao and Vasconcelos ICCV 2007, Gao, Mahadevan and Vasconcelos NIPS 2007

Suspicious coincidences

- See also: Choe and Sarma AAAI 2006
  (On relation between orientation filter responses and natural image statistics)

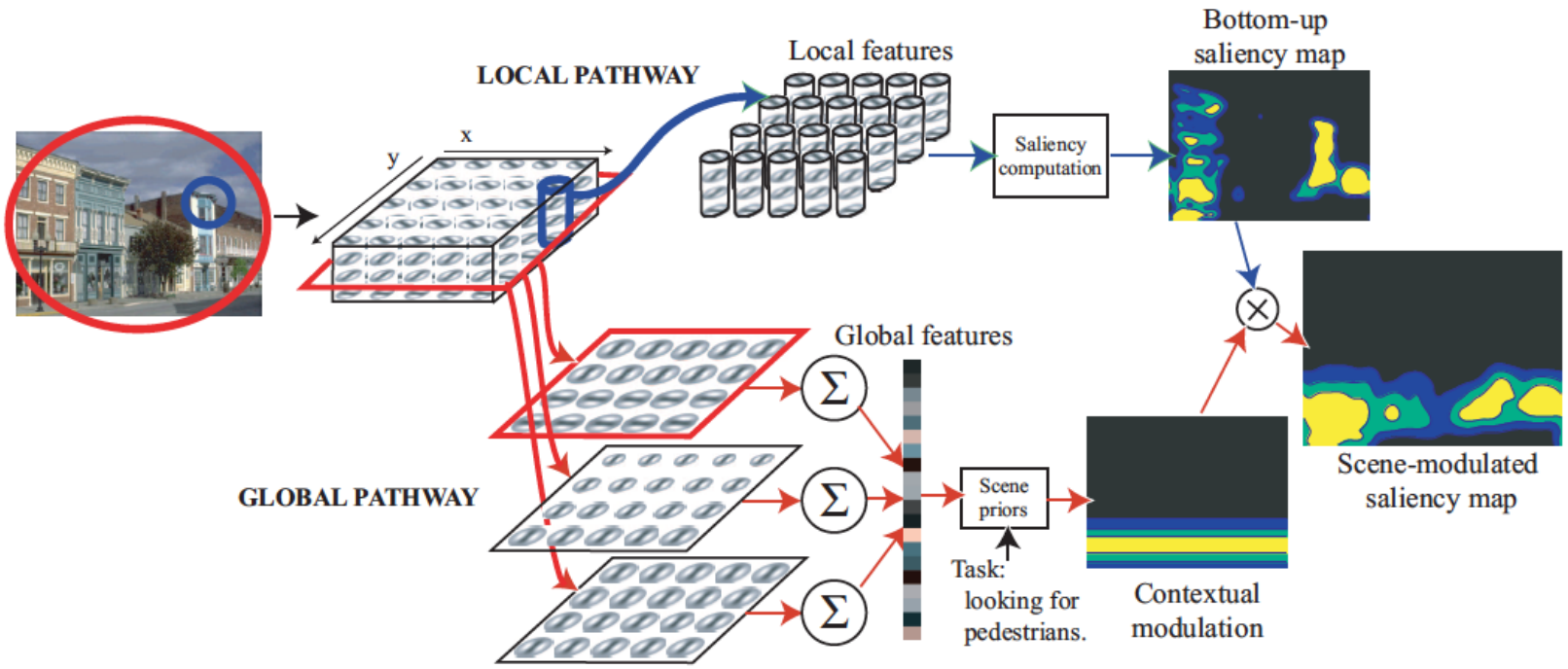


$$g(E) = \frac{\mathcal{N}(E; 0, \sigma_h^2)}{\sum_{x \in B_h} \mathcal{N}(x; 0, \sigma_h^2)}$$

Bayesian Approaches

Probabilistic and Bayesian models

- Torralba, Oliva, Castelhano and Henderson, Psych. Rev. 2006



$$p(O = 1, X|L, G)$$

$$= \frac{1}{p(L|G)} p(L|O = 1, X, G) p(X|O = 1, G) p(O = 1|G)$$

Probabilistic and Bayesian models

- This builds on several prior efforts, followed by some additional targeted efforts:
  - Context/Contextual priors:
    - Hidalgo-Sotelo, Oliva and Torralba, CVPR 2005
    - Torralba, NIPS 2001
    - and others…
  - Top-down control:
    - Oliva, Torralba, Castelhano and Henderson, ICIP 2003
    - Ehinger, Hidalgo-Sotelo, Torralba, Oliva, 2009
    - Oliva and Torralba, TICS 2007

Probabilistic and Bayesian models

- Zhang et al., J. of Vision, 2008

- SUN $s_z = p(C=1 \mid F=f_z, L=l_z)$

$$= \frac{p(F=f_z, L=l_z \mid C=1)\,p(C=1)}{p(F=f_z, L=l_z)}$$

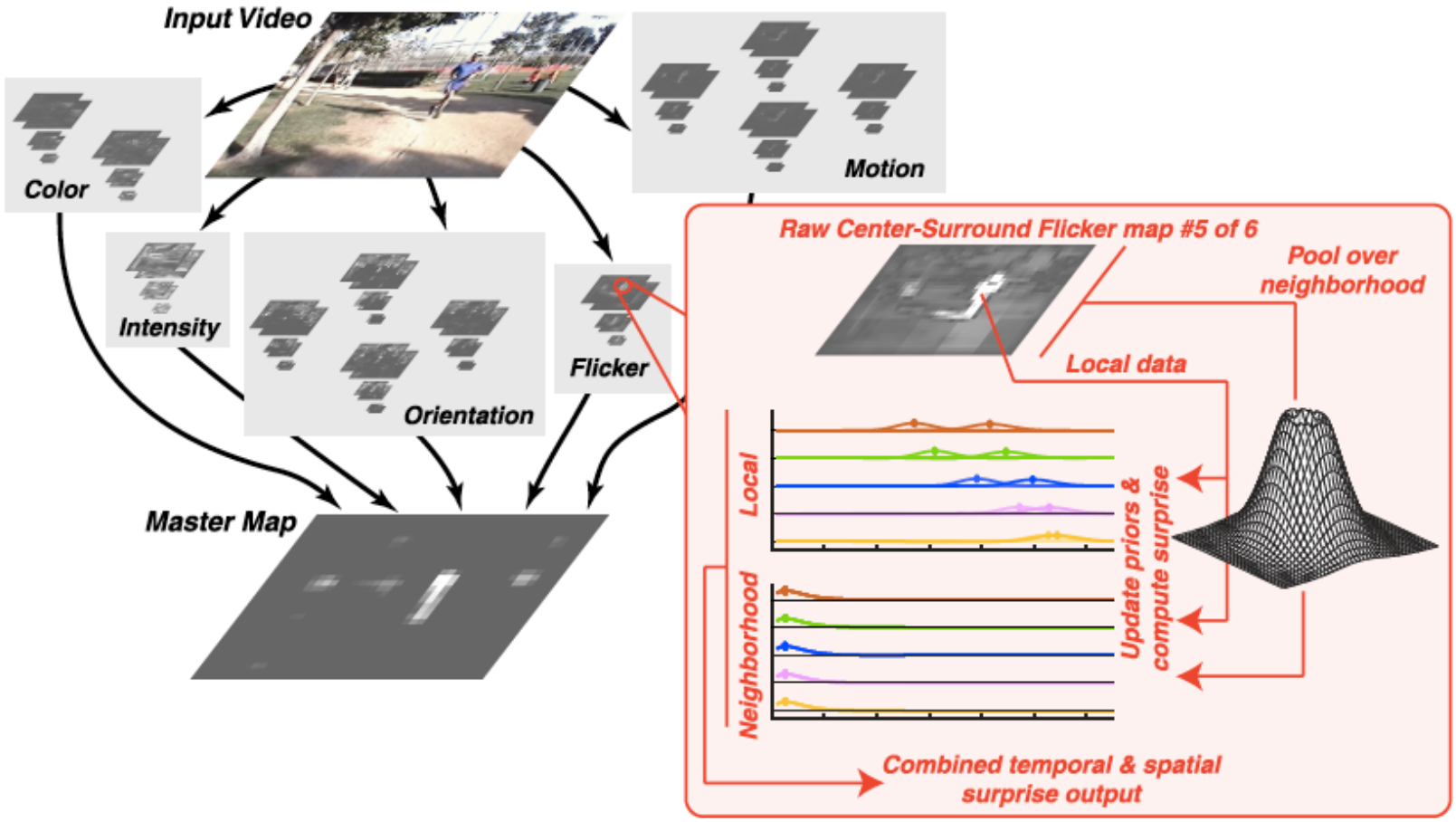
$$\log s_z = \underbrace{-\log p(F=f_z)}_{\text{Self-information: Bottom-up saliency}} + \underbrace{\log p(F=f_z \mid C=1)}_{\text{Log likelihood: Top-down knowledge of appearance}} + \underbrace{\log p(C=1 \mid L=l_z)}_{\text{Location prior: Top-down knowledge of target's location}}$$

Probabilistic and Bayesian models

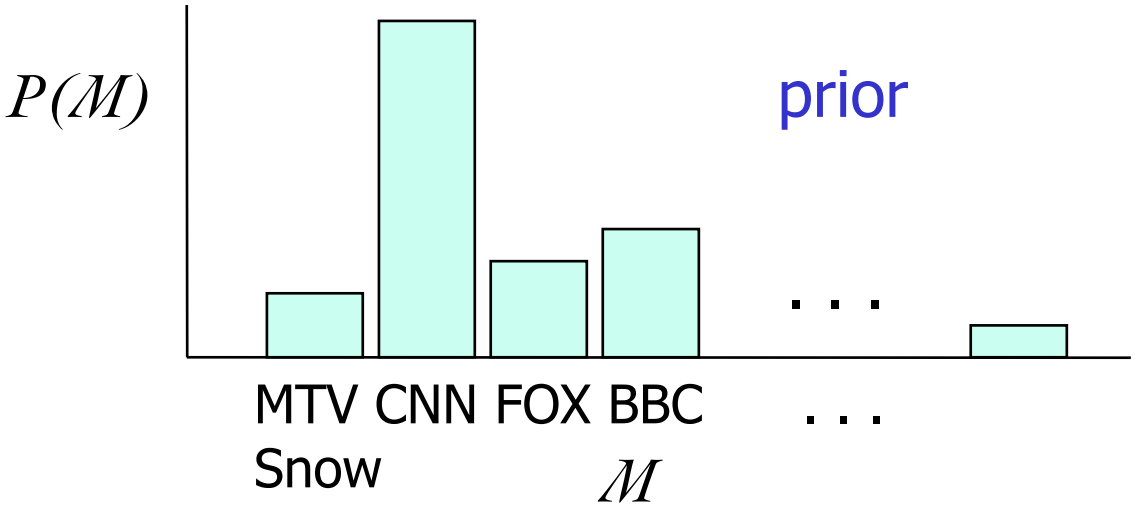
- Zhang et al. Proc. Cog. Sci. Soc. 2009,
  - SUNDAy, Dynamic analysis of scenes
- Kanan et al. 2009, Visual Cognition
  - Top down saliency
- Barrington et al. J. of Vision, 2008
  - NIMBLE: Saccade based visual memory
- Static model of natural image statistics, modeled as GGD lends itself to a very fast computational framework

Probabilistic and Bayesian models

- Itti and Baldi, NIPS 2006, Vis. Res. 2009, Neural Netw, 2010



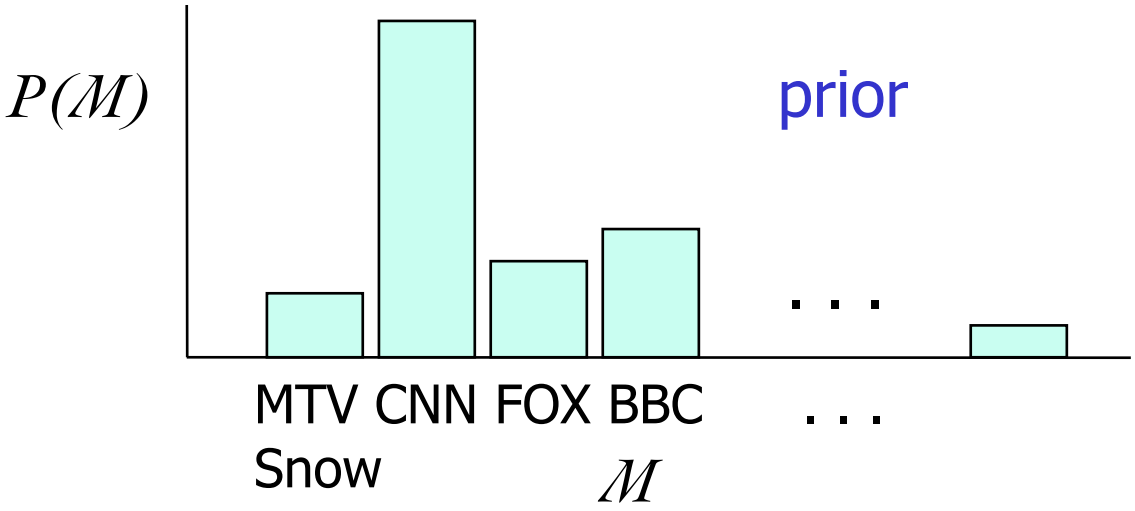
Bayesian Surprise



- Family $\mathcal{M}$ of observer-dependent models or hypotheses about the world
- Observer beliefs: $\{P(M)\}_{M \in \mathcal{M}}$

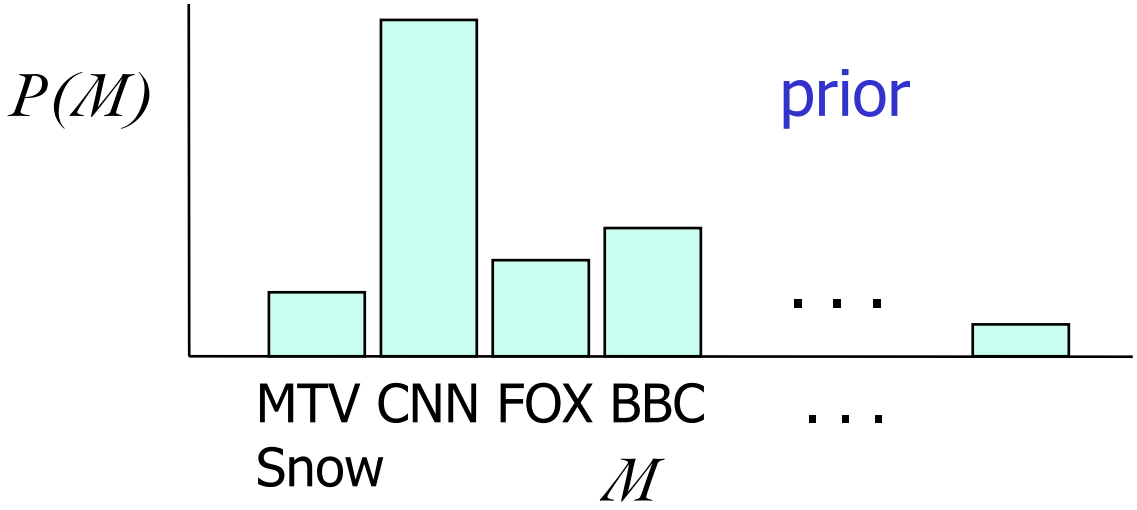
Bayesian Surprise

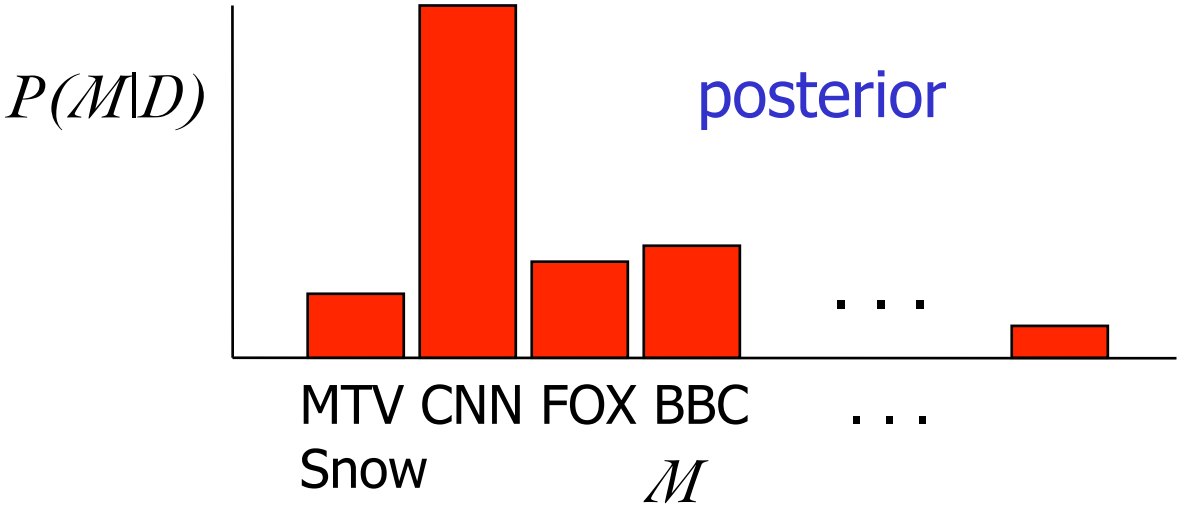
- Bayesian foundation of probability: data is what changes a prior into a posterior:



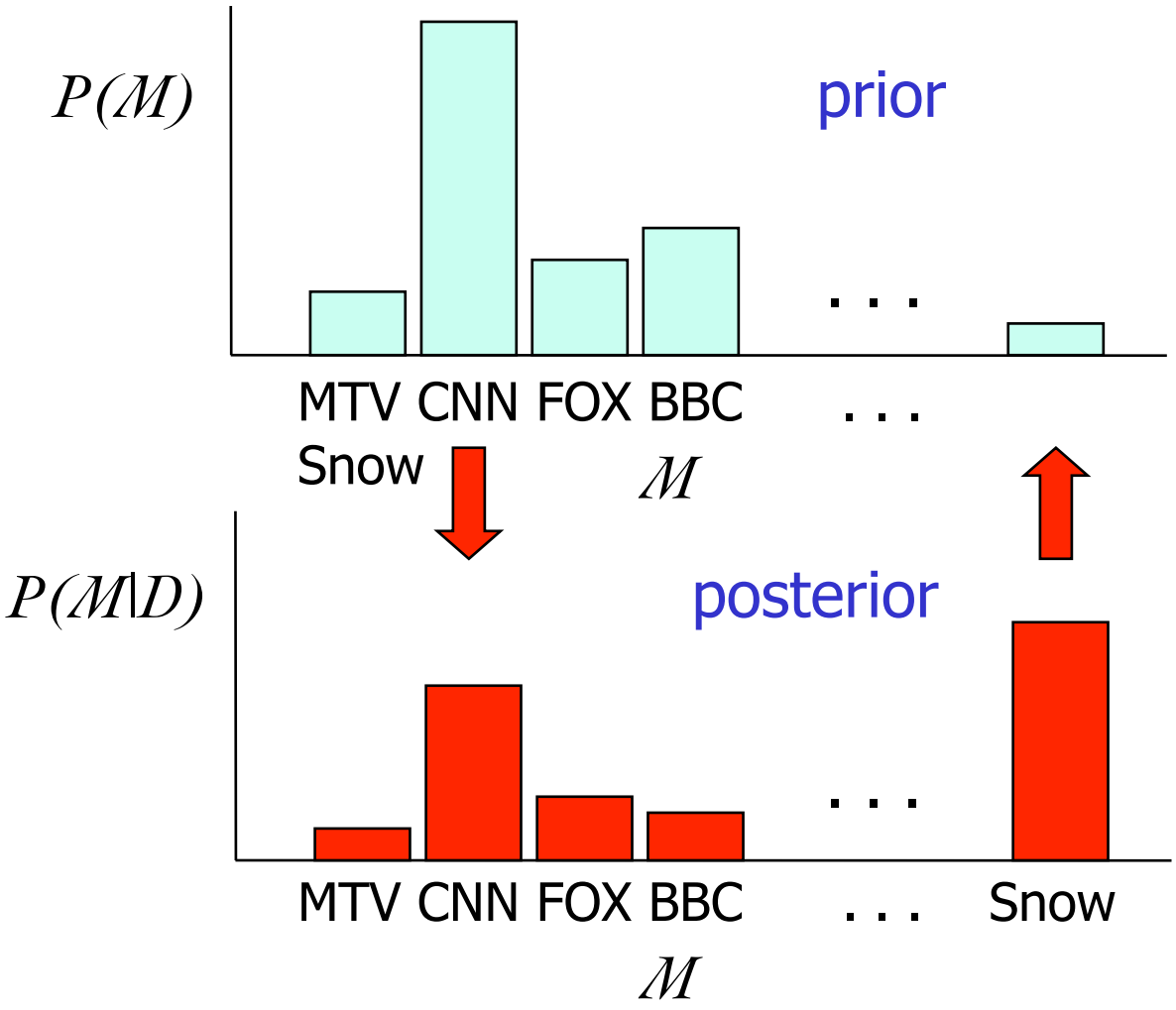
$$P(M \mid D) = \frac{P(D \mid M)}{P(D)} P(M)$$

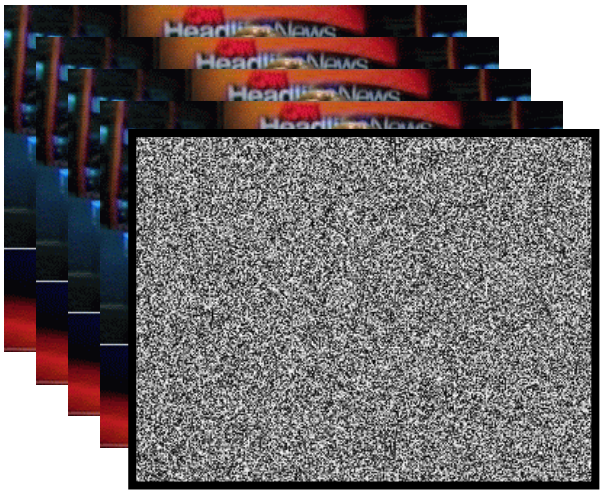
Bayesian Surprise





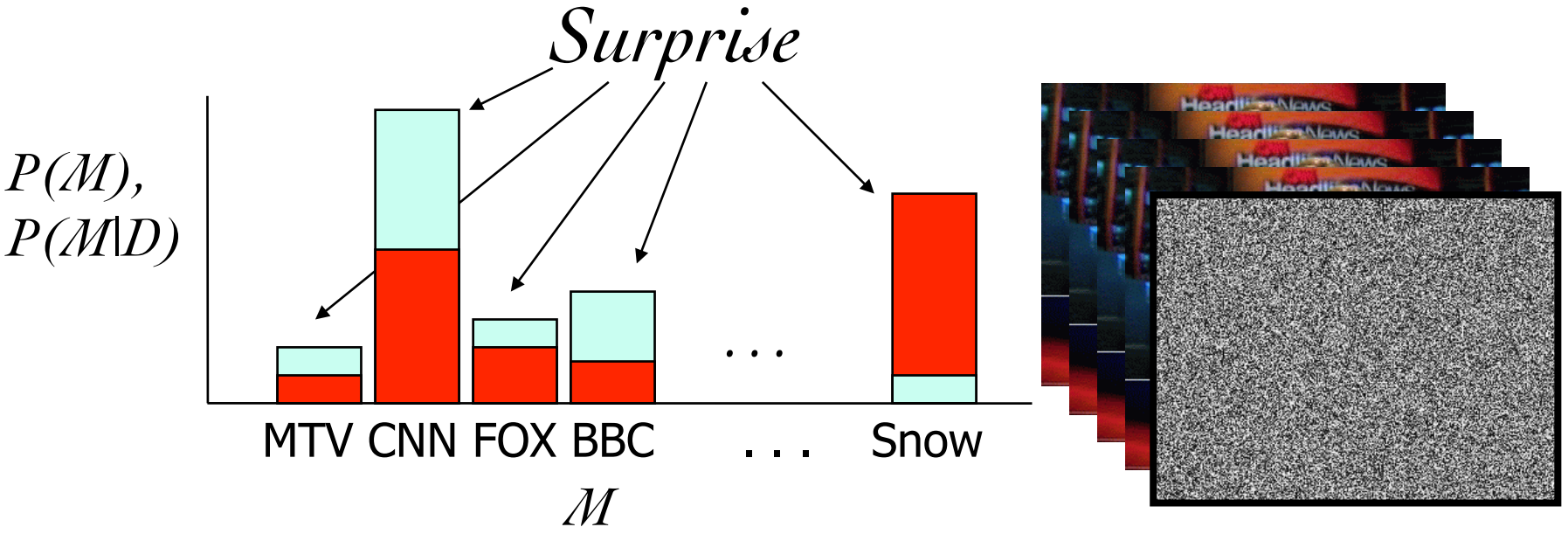
Bayesian Surprise





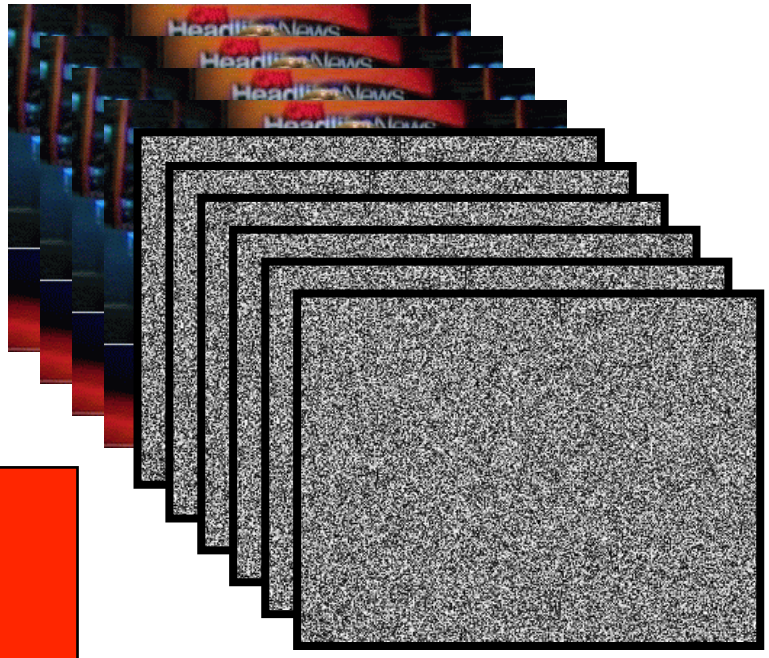
Bayesian Surprise

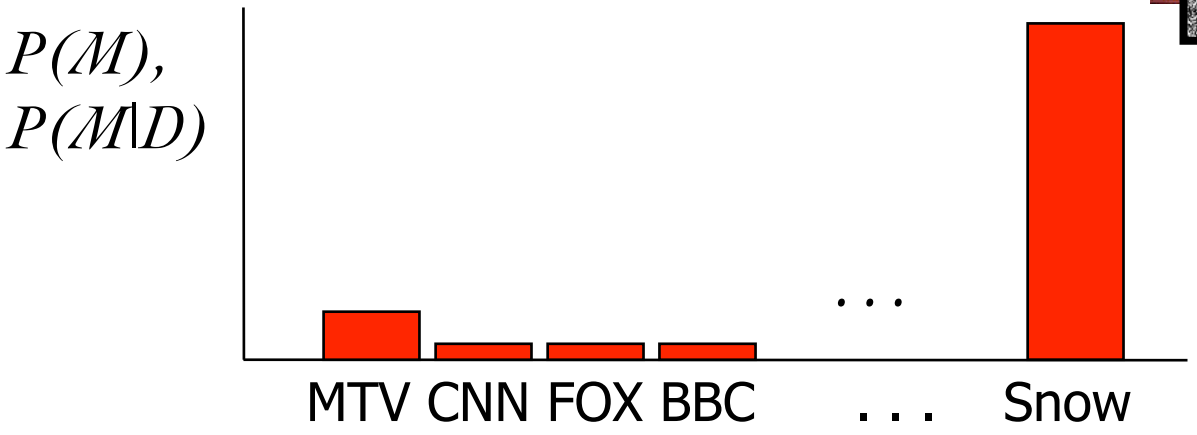
$$Surprise = d\left[P(M \mid D), P(M)\right]_{M \in \mathcal{M}}$$



After a moment...

- Beliefs stabilize, prior and posterior become identical,
- and additional snow frames carry no surprise.





Unit of surprise

- Shannon: $I(D) = -\log P(D)$
  - 1 bit observed when outcome has probability 0.5
  - E.g., toss a fair coin and observe "heads"

- Surprise: $S(M) = \log \frac{P(M \mid D)}{P(M)}$
  - 1 wow experienced when belief in M changes by factor 2
  - E.g., from 20% belief that a coin is fair to 40%

Computing surprise

- Discrete Case (Binomial/Dirichlet)

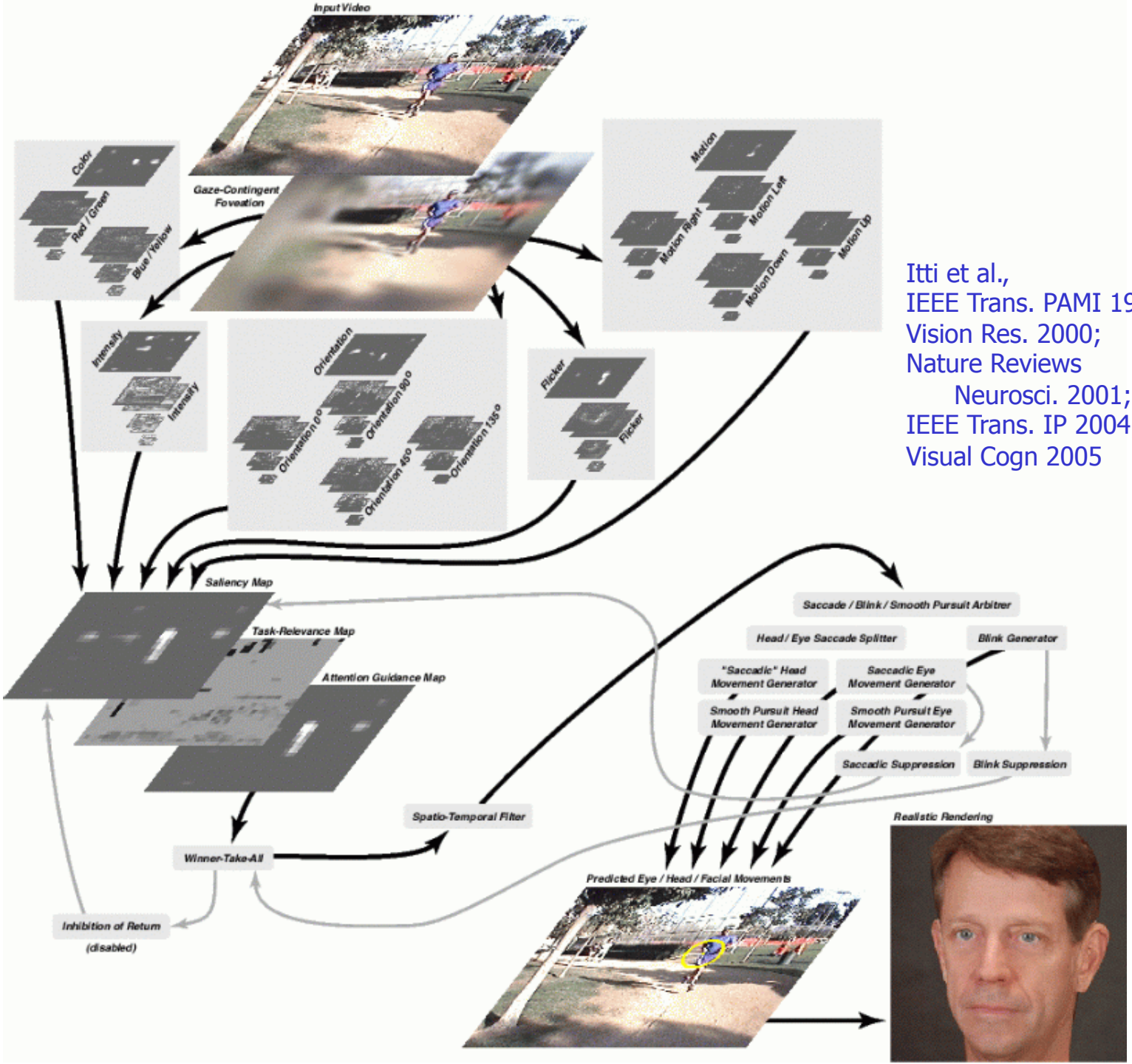
$$S(D,\mathcal{M}) \approx N \times KL\left[p, \frac{a_1}{a_1 + b_1}\right]$$

p =observed frequency over N binary data samples

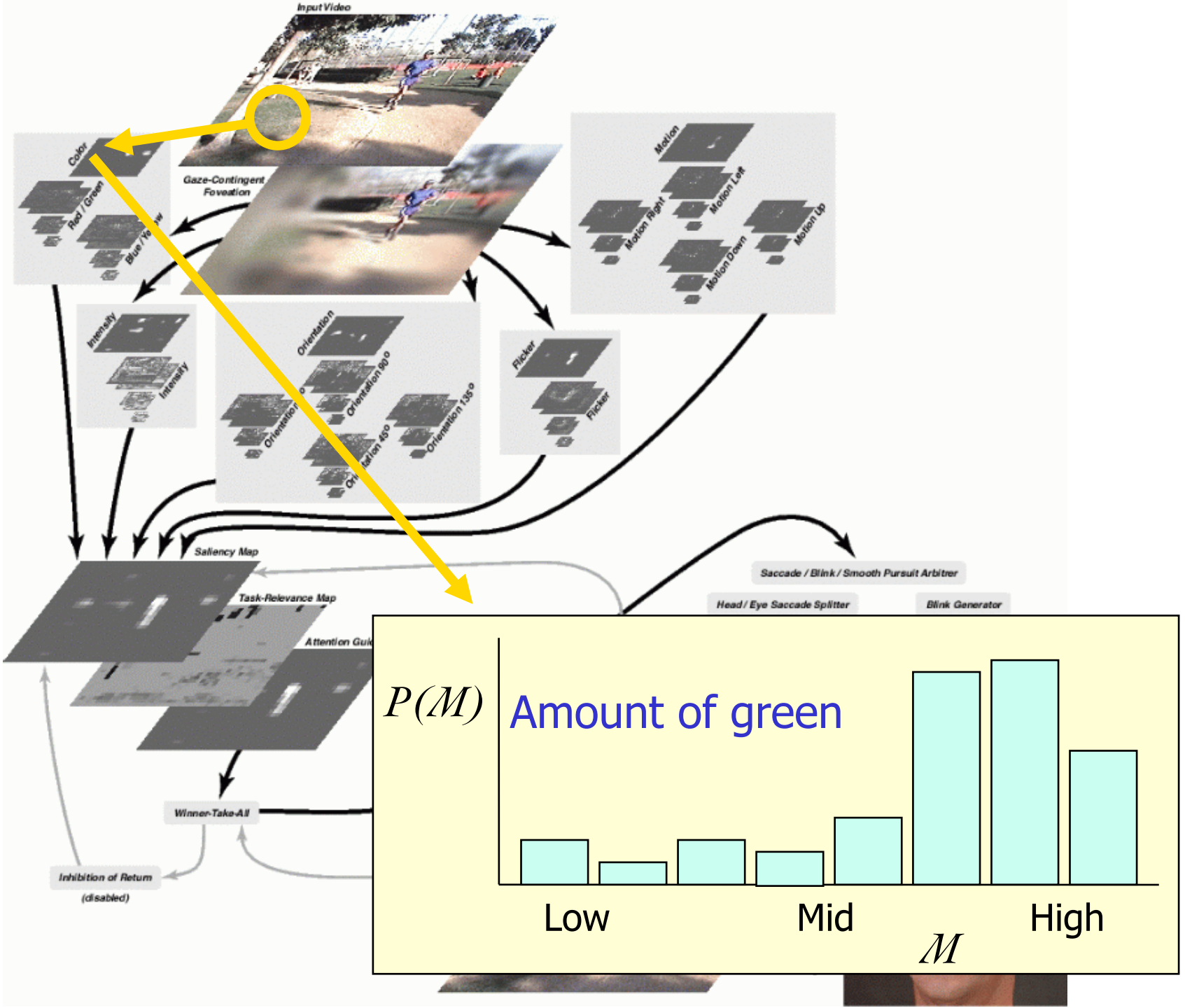
a1 /(a1+b1) = expected frequency (Dirichlet)

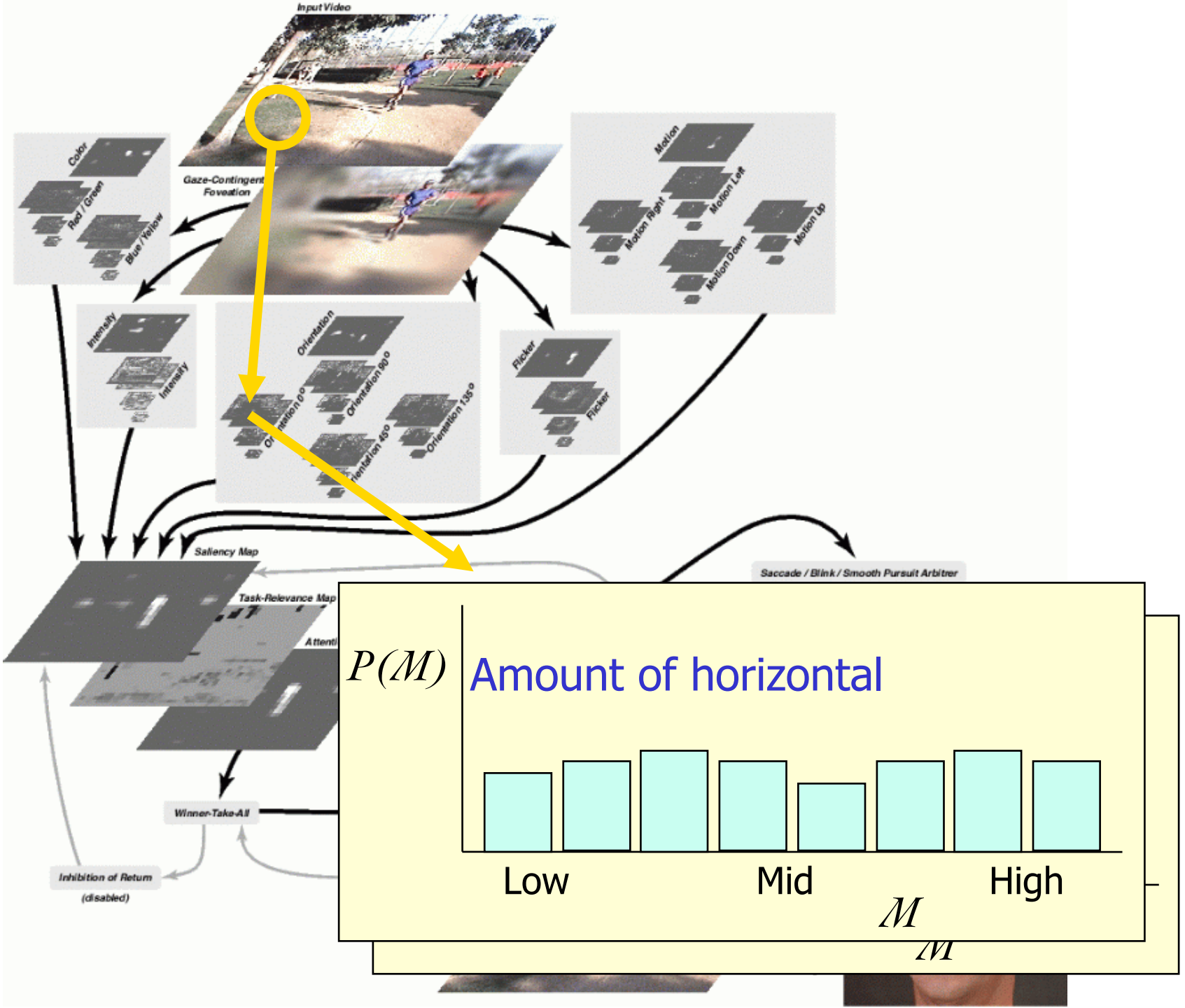
$$S(D,\mathcal{M}) = \frac{N}{2}\left[\frac{1}{\kappa_1} + \frac{\overline{\sigma}^2}{s_1^2} + \log\frac{\upsilon_1 s_1^2}{2\overline{\sigma}^2} - \Psi\left(\frac{\upsilon_1}{2}\right) + \frac{(\overline{m} - \mu_1)^2}{s_1^2}\right]$$

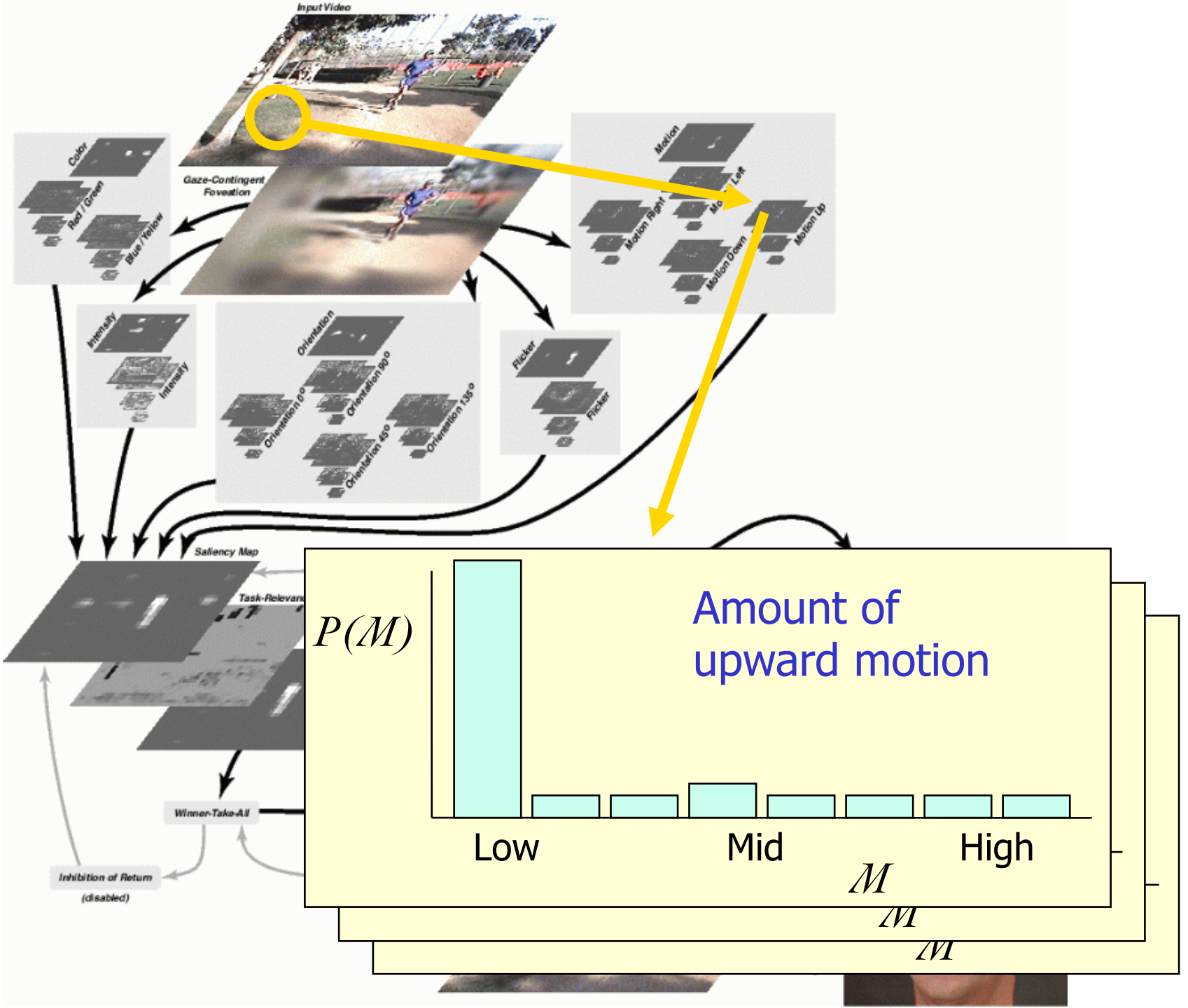
- Continuous Case (Gaussian, unknown mean and variance)

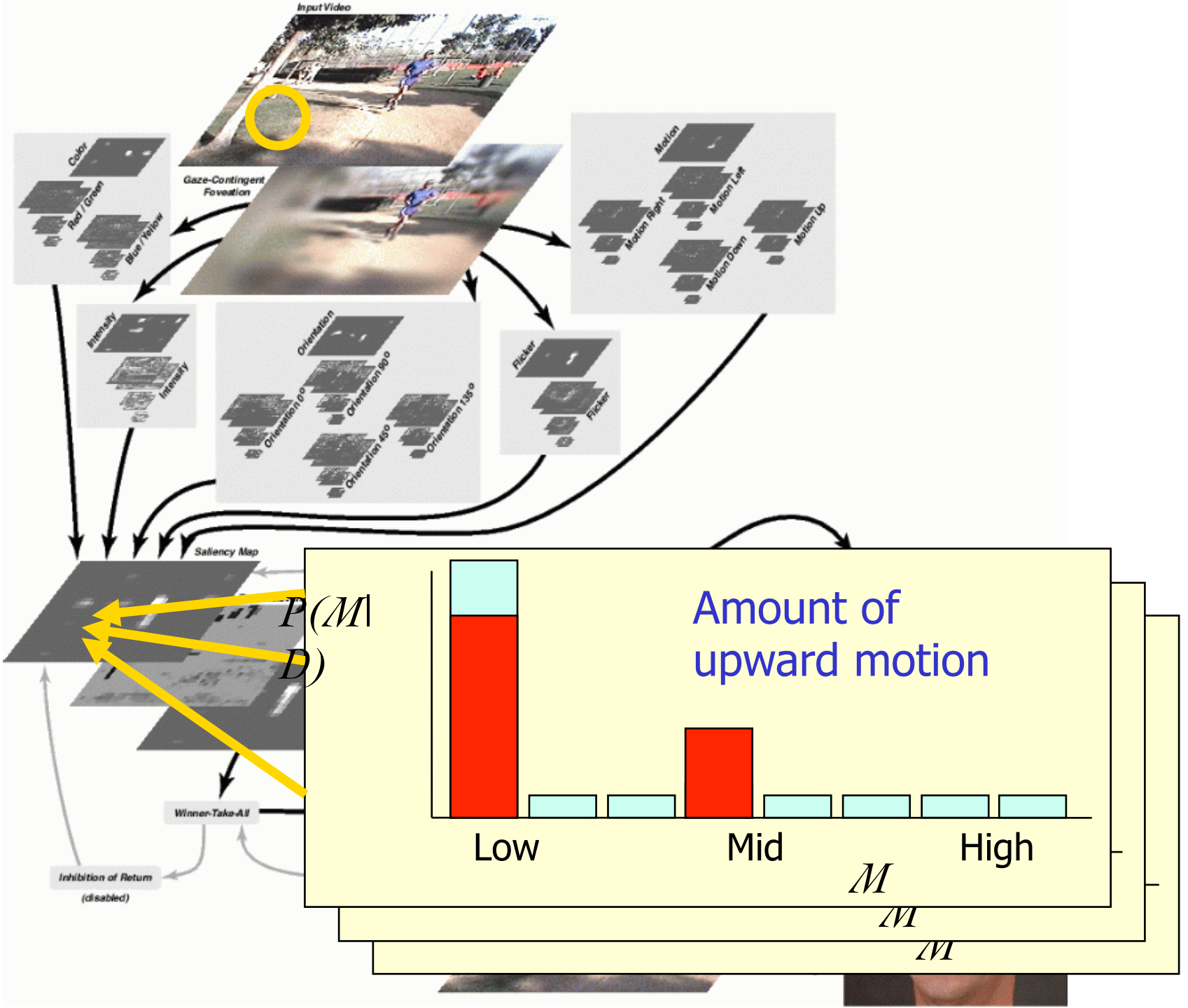


Itti et al.,
IEEE Trans. PAMI 1998;
Vision Res. 2000;
Nature Reviews Neurosci. 2001;
IEEE Trans. IP 2004;
Visual Cogn 2005


Input Video
Gaze-Contingent Foveation
Color
Red / Green
Motion
Motion Right
Motion Left
Motion Down
Motion Up
Intensity
Orientation
Orientation 90°
Orientation 135°
Flicker
Saliency Map
Task-Relevance Map
Winner-Take-All
Inhibition of Return
(disabled)
Saccade / Blink / Smooth Pursuit Arbitrer
Head / Eye Saccade Splitter
Blink Generator
P(M)
Amount of green
Low
Mid
High
M

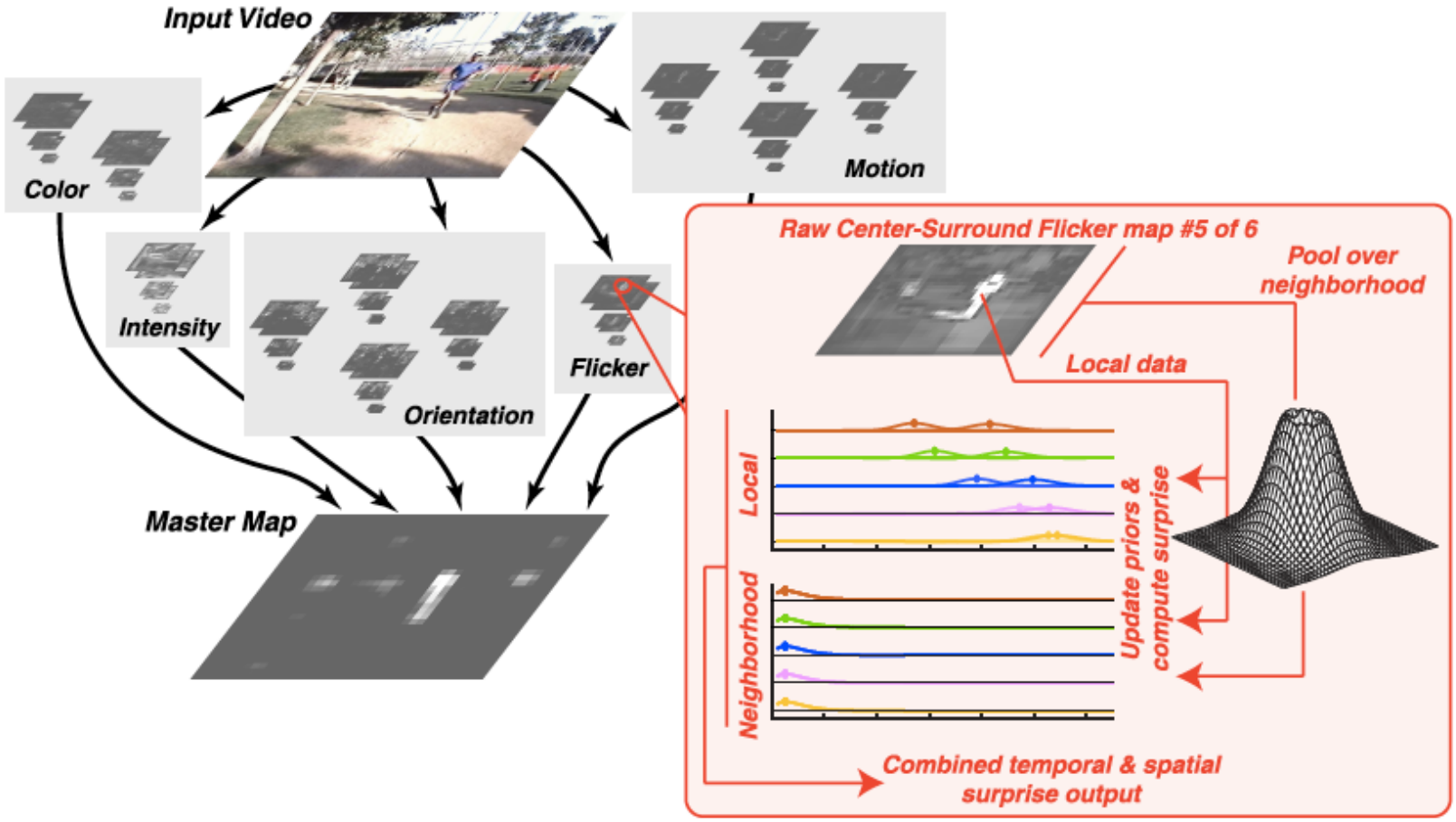

Input Video
Gaze-Contingent Foveation
Color
Red / Green
Blue / Yellow
Motion
Motion Right
Motion Left
Motion Down
Motion Up
Intensity
Intensity
Orientation
Orientation 90°
Orientation 135°
Orientation 45°
Flicker
Flicker
Saliency Map
Task-Relevance Map
Saccade / Blink / Smooth Pursuit Arbitrer
Winner-Take-All
Inhibition of Return
(disabled)
$P(M)$
Amount of horizontal
Low
Mid
High
M


Input Video
Gaze-Contingent Foveation
Color
Red / Green
Blue / Yellow
Intensity
Intensity
Orientation
Orientation 0°
Orientation 90°
Orientation 45°
Orientation 135°
Flicker
Flicker
Motion
Motion Right
Motion Down
Motion Up
Saliency Map
Winner-Take-All
Inhibition of Return
(disabled)
P(M)
Amount of upward motion
Low
Mid
High
M


Input Video
Gaze-Contingent Foveation
Color
Red / Green
Blue / Yellow
Intensity
Intensity
Orientation
Orientation 0°
Orientation 90°
Orientation 45°
Orientation 135°
Flicker
Flicker
Motion
Motion Right
Motion Left
Motion Down
Motion Up
Saliency Map
Winner-Take-All
Inhibition of Return
(disabled)
P(M|D)
Amount of upward motion
Low
Mid
High
M
M
M

Probabilistic and Bayesian models

- Itti and Baldi, NIPS 2006, Vis. Res. 2009, Neural Netw, 2010



Graphical Models

Graph Based techniques

- Harel 2006
  - Scale-space pyramid from intensity, color, orientation
  - Fully connected graph over all grid locations
  - Graph weights proportional to similarity of feature values, and spatial distance

$$d((i,j)||(p,q)) \triangleq \left|\log \frac{M(i,j)}{M(p,q)}\right|$$

$$w_1((i,j),(p,q)) \triangleq d((i,j)||(p,q)) \cdot F(i-p, j-q), \text{ where}$$

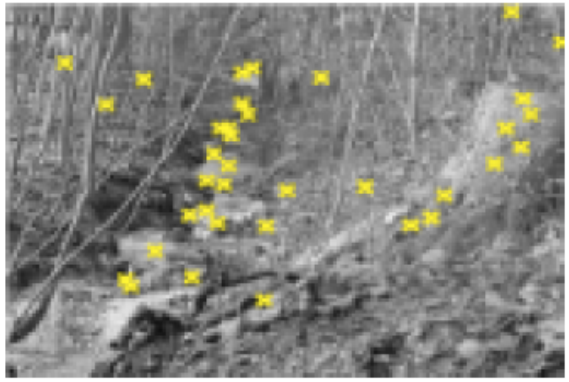
$$F(a,b) \triangleq \exp\left(-\frac{a^2+b^2}{2\sigma^2}\right).$$

Graph Based techniques

- Harel, NIPS 2006
  - Treated as Markov chain that reflects expected time spent by a random walker (walking forever)
  - Weights of outbound edges normalized to 1 with equivalence relation defined between nodes/states and edges/transition probabilities
  - Saliency corresponds to equilibrium distribution

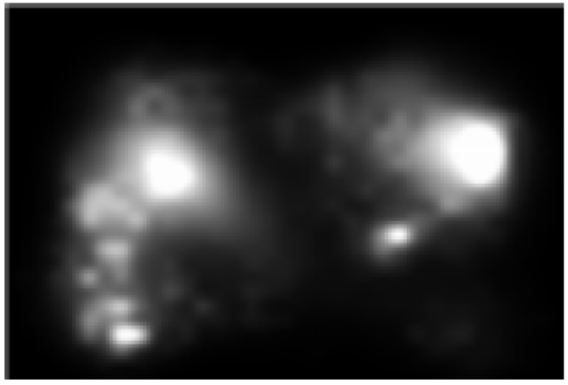
Graph Based techniques

(a) Sample Picture With Fixation



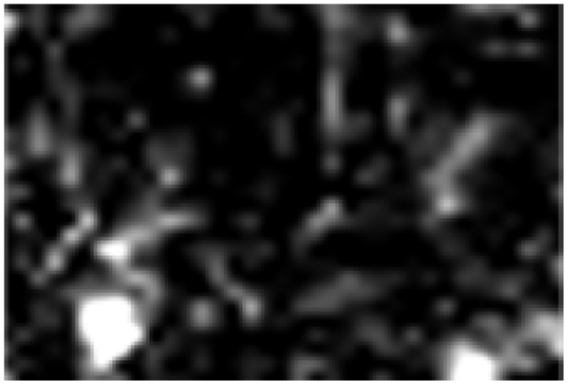
(b) Graph-Based Saliency Map

ROC area = 0.74



(c) Traditional Saliency Map

ROC area = 0.57



Graph Based techniques

- Pang ICME 2008
  - Stochastic model based on signal detection theory
  - Dynamic Bayes net with 4 layers
    - Layer 1: Itti-like saliency determination
    - Layer 2: Gaussian state-space model (stochastic saliency map)
    - Layer 3: Overt shifts determined by HMM
    - Layer 4: Density map predicts positions

Graph Based techniques

- Avraham and Lindenbaum (PAMI 2010)

The Esaliency Algorithm

1) Select candidates using some segmentation process.
2) Use the preference for a small number of expected targets (and possibly other preferences) to set the initial (prior) probability for each candidate to be a target.
3) Measure visual similarity between every two candidates and infer the correlations between the corresponding labels.
4) Represent the label dependencies using a Bayesian network.
5) Find the N most likely joint assignments.
6) Deduce the saliency of each candidate by marginalization.

$$I(l_i, l_j) = \sum_{l_i=0,1} \sum_{l_j=0,1} p(l_i, l_j) \log \frac{p(l_i, l_j)}{p(l_i)p(l_j)}$$

Labels are binary random variables:

$$\begin{aligned}
p(l_i = 1, l_p = 1) &= \gamma(d_{ij})\sqrt{\mu_i(1-\mu_i)\mu_j(1-\mu_j)} + \mu_i\mu_j \\
p(l_i = 1, l_j = 0) &= \mu_i - p(l_i = 1, l_j = 1) \\
p(l_i = 0, l_j = 1) &= \mu_j - p(l_i = 1, l_j = 1) \\
p(l_i = 0, l_j = 0) &= 1 - \mu_i - \mu_j + p(l_i = 1, l_j = 1).
\end{aligned}$$

E-Saliency

- Dependency on parent nodes for label

$$p(\bar{l}) = p(l_r) \prod_{i=1,\ldots,n;i\neq r} p(l_i|l_{\mathrm{par}(i)})$$

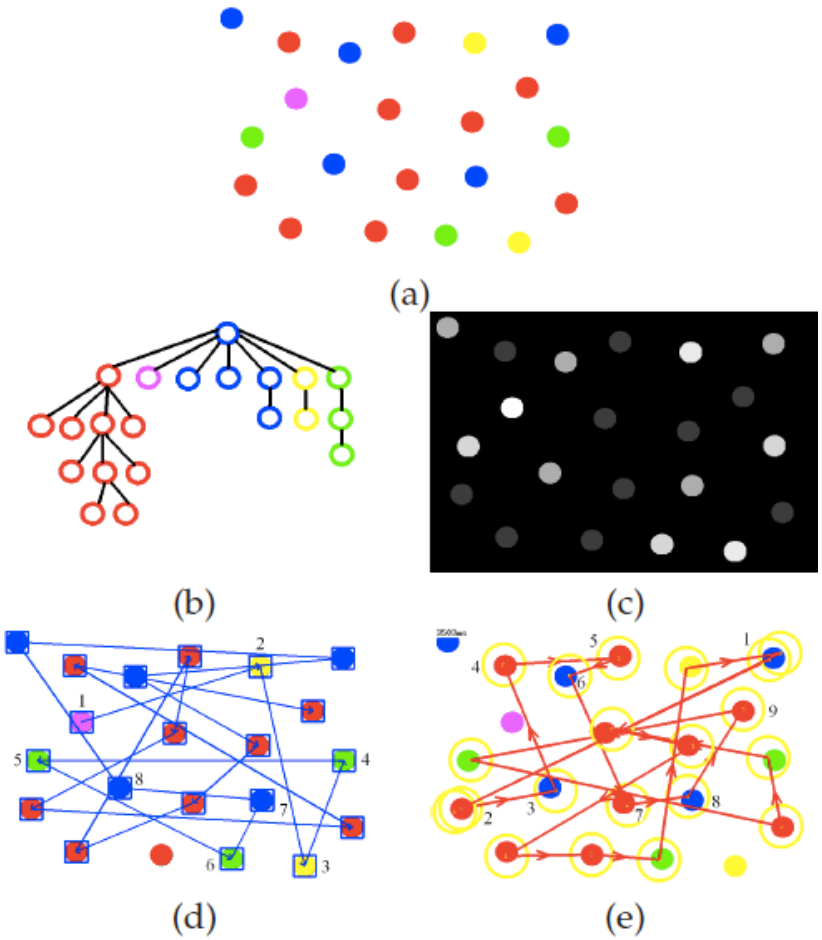
- Marginalization considering most likely assignments:

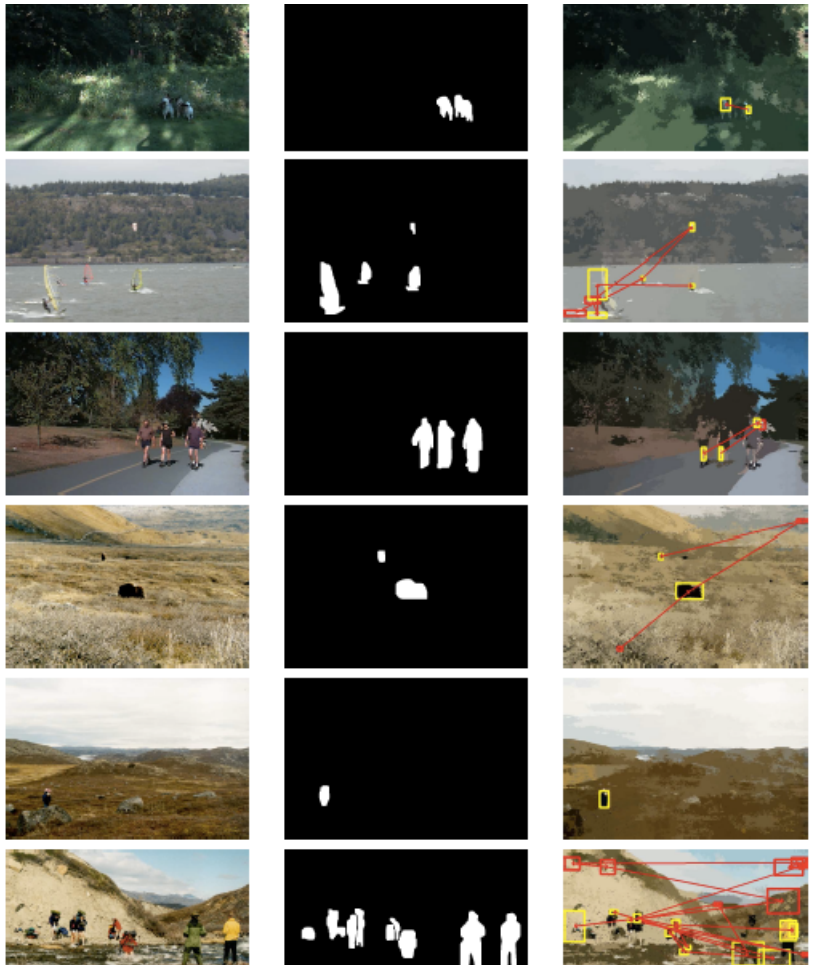
$$p'(\bar{l}) = \frac{p(\bar{l})}{\sum_{j=1}^{N} p(\bar{l^j})}$$

The saliencies are then:

$$p_T(c_i) = \sum_{j=1}^{N} p'(\bar{l^j}) \cdot l_i^j$$

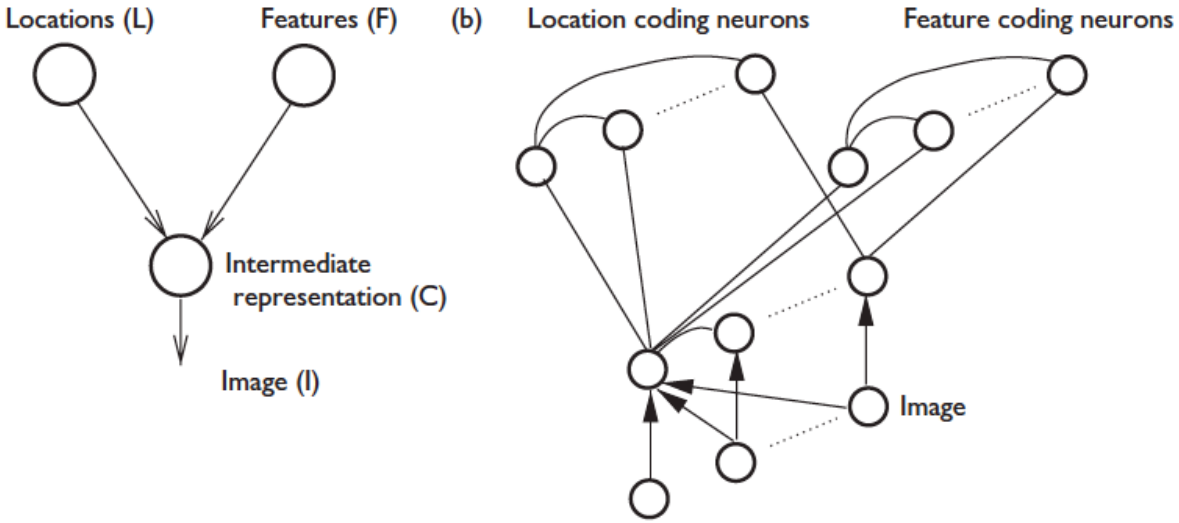
E-saliency





Probabilistic and Bayesian models

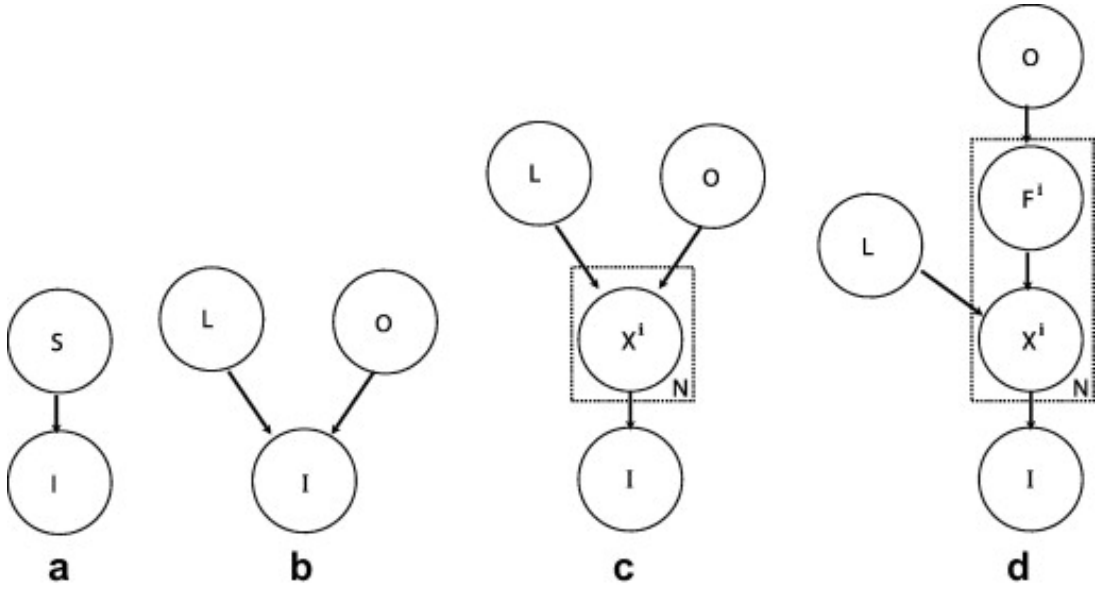
- Rao, NeuroReport 2005



- Bayesian, Integrate and Fire model
- Heavily inspired by biology, brain imaging
- See also Rao and Ballard, Nat. Neurosci. 1999

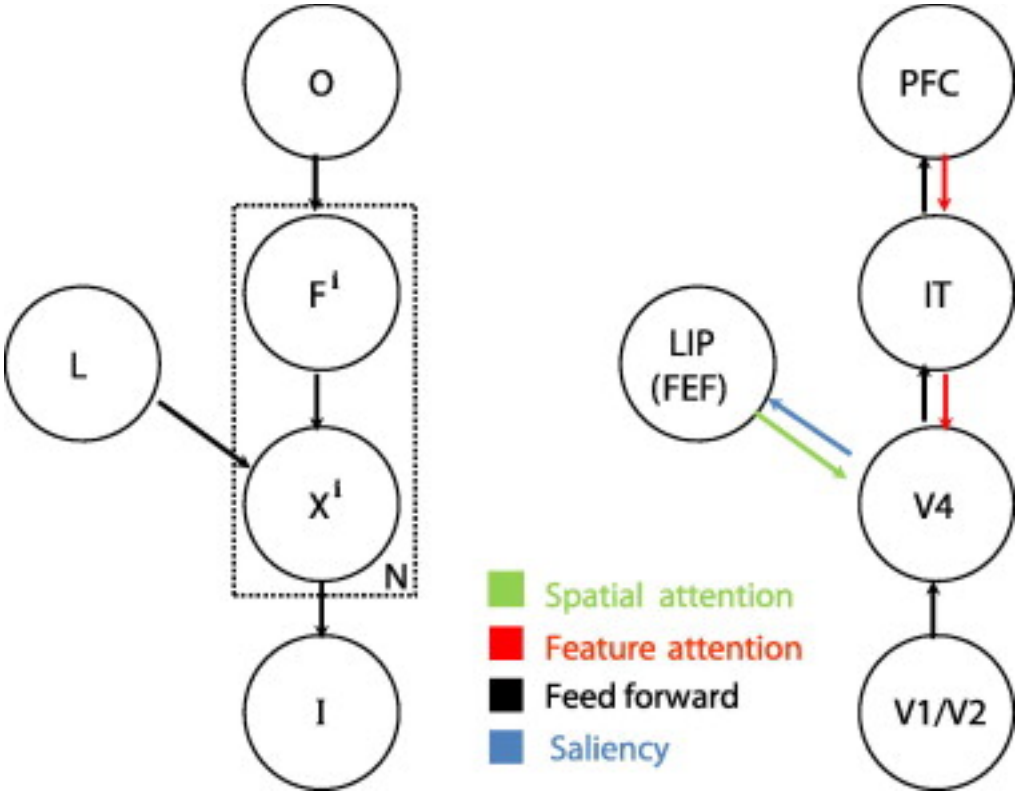
Probabilistic and Bayesian models

- Chikkerur et al., Vis. Res. 2010, MIT Ph.D. Thesis



Graph Based techniques

- Strongly inspired by biology



Learning/Object detection Methods

- See also:
  - What is an object? (Alexe et al. 2010)
  - Deselaers et al. (ECCV 2010)
  - Carreira and Sminchisescu (CVPR 2010)
  - Gu et al. (CVPR 2009)
  - van de Sande et al. (ICCV 2011)
  - and many more…

Some take home points

Take home points…

- Much overlap in fundamental ideas that inspire techniques in this domain
- This isn't surprising (these are all fundamental principles in many efforts – not just saliency)
- Reveals that the details are important (think of the tree)
- There are several benchmarks (which are important), but consider also application
- Saliency is very useful – but won't solve everything