ICIP 2015

VISUAL SALIENCY: FUNDAMENTALS, APPLICATIONS, AND RECENT PROGRESS

FUNDAMENTALS AND IMPORTANT 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

NSERC CRSNG

Overview

- The Human Ability to Attend
- The Computational Argument for Attention
- Important Theories and Models
  - Coding based approaches
  - Graph-based strategies
  - Bayesian approaches
  - Spectral domain

Credits: Some slides provided by John K. Tsotsos, and Laurent Itti

The Classic Cocktail Party Effect

Read the red print.
What do you remember from the regular print text?

Somewhere Among hidden the in most the spectacular Rocky Mountains cognitive near abilities Central City is Colorado the an ability old to miner select hid one a message box from of another gold. We Although do several this hundred by people focusing have our looked attention for on it, certain they cues have such not as found type it style.

Shadowing Task

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with 2 Ears, *Journal of the Acoustical Society of America 25(5)*, 975–979.
Poulton, E. C. (1953). Two Channel Listening, *J. of Experimental Psychology 46*, 91–96.

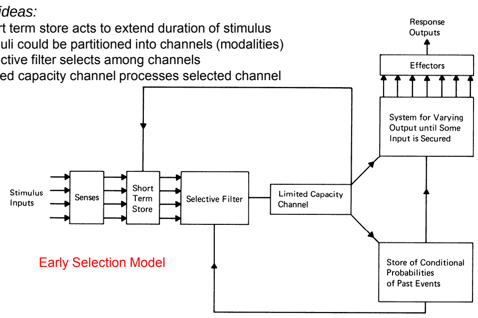
- exposed subjects to two or more verbal messages simultaneously by presentation to different ears.
- subjects were instructed to attend to one particular characteristic (gender, content, language, etc.) or were given no instructions at all, and then were asked questions about the messages.
- some experiments involved “shadowing”, repeating verbal stimuli as they were received

- Subjects showed surprisingly little awareness for the content or even characteristics of unattended stimuli, suggesting that unattended stimuli were rejected from further processing.

Broadbent 1958

Broadbent, D. (1958). **Perception and communication**, Pergamon Press, NY.

Key ideas:
- short term store acts to extend duration of stimulus
- stimuli could be partitioned into channels (modalities)
- selective filter selects among channels
- limited capacity channel processes selected channel



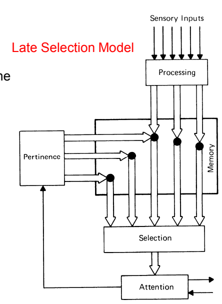
Early Selection Model

Deutsch/Norman/Moray/MacKay Model

Deutsch, J., Deutsch, D. (1963). Attention: Some theoretical considerations, *Psych. Review 70*, 80-90.
Norman, D. (1968). Toward a theory of memory and attention, *Psych. Review 75*, 522-536.
Moray, N. (1969). **Attention: Selective Processes in Vision and Hearing**, Hutchinson, London.
MacKay, D., (1973). Aspects of the Theory of Comprehension, Memory and Attention, *Quarterly J. Exp. Psych. 25*, 22-40.

Late Selection Model

Key ideas:
- all information is recognized before it receives the attention of a limited capacity processor
- recognition can occur in parallel
- stimulus relevance determines what is attended



Treisman 1964

Treisman, A. (1964). The effect of irrelevant material on the efficiency of selective listening, *American J. Psychology 77*, 533-546.

Key ideas:
- filter attenuates (is not binary) unattended signals causing them to be incompletely analyzed
- filter can operate at different levels - signal or meaning - so attention is hierarchical

Overt Attention: Kinds of Eye Movements

Saccade: voluntary jump-like movements

Vestibular-Ocular Reflex: stabilize visual image on retina by causing compensatory changes in eye position as head moves *stimulus:* semicircular canal hair cells (acceleration)

Nystagmus: compensatory eye movements can reach limits of the orbit and must be reset by a primitive saccade.

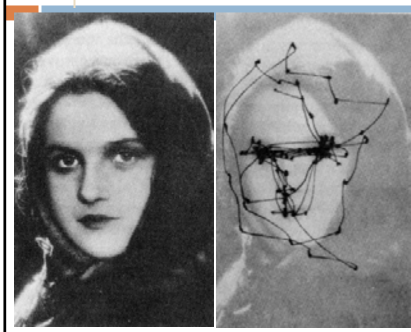
Optokinetic Nystagmus: stabilizes gaze during sustained, low-frequency rotations at constant velocity. *stimulus:* large field motion in retina projected to pretectum

Smooth Pursuit: voluntary tracking of moving stimuli

Vergence: coordinated movement of both eyes , converging for objects moving towards and diverging for objects moving away from the eyes. *stimulus:* stereopsis

Torsion: coordinated rotation of the eyes around optical axis, dependent on head tilt and eye elevation. *stimulus:* vergence away from horizontal head axis

Typical Scanpath



regardless of the claims of biological plausibility or realism, none of the attention models can replicate such scanpaths

Task and Eye Movements

Yarbus, A. L. (1967). **Eye Movements and Vision**. New York: Plenum.

Yarbus demonstrated how eye movements changed depending on the question asked of the subject:

1. No question asked
2. Judge economic status
3. "What were they doing before the visitor arrived?"
4. "What clothes are they wearing?"
5. "Where are they?"
6. "How long is it since the visitor has seen the family?"
7. Estimate how long the "unexpected visitor" had been away from the family

Integrating Overt and Covert Attention

Posner, M. I. (1980). 'Orienting of Attention', *Quarterly Journal of Experimental Psychology* **32**, 1, 3–25.

Proposed that attention had three major functions:
- provided the ability to process high priority signals or *alerting*.
- permitted *orienting* and overt foveation of a stimulus.
- allowed *search* to detect targets in cluttered scenes.

Orienting improves efficiency of target processing in terms of acuity, permitting events at the foveated location to be reported more rapidly as well as at lower threshold.

Overt foveation is strongly linked to movement of covert attention. Overt orienting, whether of the eyes or the head or both, is termed *exogenous* while covert fixation shifts are called *endogenous.*

Exogenous control of gaze direction is controlled reflexively by external stimulation while endogenous gaze is controlled by internally generated signals.

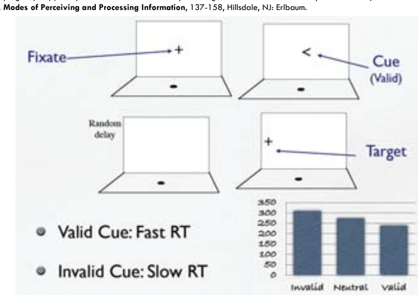
Covert fixations are not observable and thus must be inferred from performance of some task.

Visual Search: Attentional Cueing

Posner, M. I. , Nissen, M., Ogden, W., (1978). Attended and unattended processing modes: The role of set for spatial locations, in Pick & Saltzmann, eds., **Modes of Perceiving and Processing Information**, 137-158, Hillsdale, NJ: Erlbaum.

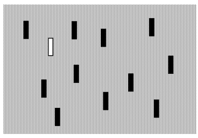
Priming Paradigm - categorization is facilitated if same image presented a second time

(Bartram D., (1974). The role of visual and semantic codes in object naming, *Cognitive Psychology 6(3)* 325-356)

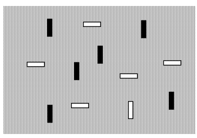


Visual Search

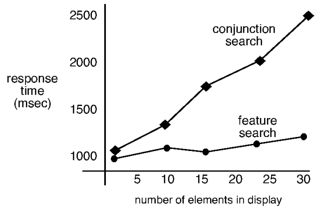
for review, see Wolfe, J. (1998). Visual Search, in **Attention** (ed. Pashler, H.), 13–74, University College London, London.



Feature Search



Conjunction Search

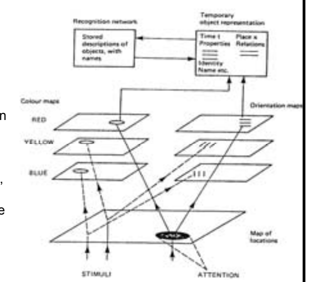


Feature Integration Theory (FIT)

Treisman, A., Gelade, G. (1980). A feature integration theory of attention, *Cognitive Psychology* 12: 97-136.

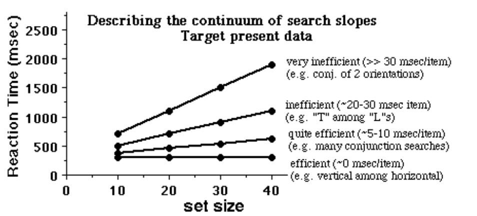
Key ideas:

- we can detect and identify separable features in parallel across a display (within the limits set by acuity, discriminability, and lateral interference)
- this early, parallel, process of feature registration mediates texture segregation and figure ground grouping;
- that locating any individual feature requires an additional operation;
- that if attention is diverted or overloaded, illusory conjunctions may occur;
- conjunctions, require focal attention to be directed serially to each relevant location;
- they do not mediate texture segregation, and they cannot be identified without also being spatially localized.



from Treisman & Sato 1990

But as time moved on....



from Wolfe 1998

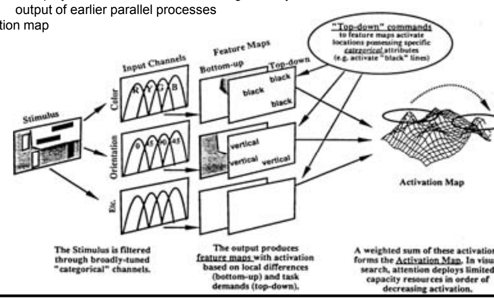
inferring mechanism from search slopes is not easy!
- There is NO serial/parallel dichotomy

Guided Search 1989

Wolfe, J., Cave, K., Franzel, S. (1989). Guided search: An alternative to the feature integration model for visual search, *J. Exp. Psychology: Human Perception and Performance 15*, 419-433.

Key ideas:

- attentional deployment of limited resources is guided by output of earlier parallel processes
- activation map



Change Blindness

Rensink, R., O'Regan, K., Clark, J., (1997). To See or Not to See: The Need for Attention to Perceive Changes in Scenes, *Psychological Science*, 8, 368-373.

Precursors: visual memory - Observers were found to be poor at detecting change if old and new displays were separated by an ISI of more than 60-70 ms.
saccades - observers were found to be poor at detecting change, with detection good only for a change in the saccade target

Two conclusions:

- observers never form a complete, detailed representation of their surroundings.
- attention is required to perceive change, and that in the absence of localized motion signals it is guided on the basis of high-level "interest".

http://www.psych.ubc.ca/~rensink/flicker/download/

Salience: What Attracts Attention?

for a nice summary, see Wolfe, J. (1998). Visual Search, in **Attention** (ed. Pashler, H.), 13–74, University College London, London.

Just about everything someone may have studied can be considered a feature or can capture attention

Wolfe presents the kinds of features that humans can detect 'efficiently':

- Color
- Orientation
- Curvature
- Texture
- Scale
- Vernier Offset
- Size, Spatial Frequency, and Scale
- Motion
- Shape
- Onset/Offset
- Pictorial Depth Cues
- Stereoscopic Depth

For most, subjects can 'select' feature or feature values to attend in advance

Saliency Map Locus

The neural correlate of the saliency map (if it exists at all) remains an open question:

Superior Colliculus
- A.A. Kustov, D.L. Robinson, Shared neural control of attentional shifts and eye movements, Nature 384 (1996) 74–77.
- R.M. McPeek, E.L. Keller, Saccade target selection in the superior colliculus during a visual search task, J. Neurophysiol. 88 (2002) 2019–2034.
- G.D. Horwitz, W.T. Newsome, Separate signals for target selection and movement specification in the superior colliculus, Science 284 (1999) 1158–1161.

LGN
- C. Koch, A theoretical analysis of the electrical properties of an X-cell in the cat LGN: does the spine-triad circuit subserve selective visual attention? AI Memo 787, MIT, February, 1984.
- S.M. Sherman, C. Koch, The control of retinogeniculate transmission in the mammalian lateral geniculate nucleus, Exp. Brain Res. 63 (1986) 1–20.

V1
- Z. Li, A saliency map in primary visual cortex, Trends Cog. Sci. 6 (1) (2002) 9–16.

V1 and V2
- D.K. Lee, L. Itti, C. Koch, J. Braun, Attention activates winner-take-all competition among visual filters, Nat. Neurosci. 2 (4) (1999) 375–381.

Pulvinar
- S.E. Petersen, D.L. Robinson, J.D. Morris, Contributions of the pulvinar to visual spatial attention, Neuropsychologia 25 (1987) 97–105.
- M.I. Posner, S.E. Petersen, The attention system of the human brain, Annu. Rev Neurosci. 13 (1990) 25–42.

Frontal Eye Fields
- D.L. Robinson, S.E. Petersen, The pulvinar and visual salience, Trends Neurosci.15 (4) (1992) 127– 132.

Parietal Cortex
- K.G. Thompson, N.P. Bichot, J.D. Schall, Dissociation of visual discrimination from saccade programming in macaque frontal eye field, J. Neurophysiol. 77 (1997) 1046–1050.
- J. Gottlieb, M. Kusunoki, M.E. Goldberg, The representation of visual salience in monkey posterior parietal cortex, Nature 391 (1998) 481–484.

For Further Reading...

NEUROBIOLOGY OF ATTENTION
2005

- Laurent Itti
  University of Southern California
- Geraint Rees
  University College London
- John Tsotsos
  York University

ISBN: 0-12-375731-2
Pages: 744

www.elsevier.com

Computational Argument: Two Kinds of Visual Search

Tsotsos, J. (1989). The Complexity of Perceptual Search Tasks, Proc. Int. Joint Conference on Artificial Intelligence, Detroit, August, 1989, pp1571 - 1577.
Rensink, R. (1989). A New Proof of The NP-Completeness of Visual Match, Dept. of Computer Science, University of British Columbia, Vancouver, Canada, UBC CS Technical Report 89-22 (September 1989).

Unbounded Visual Search

Recognition where no task guidance to optimize search is permitted.

Corresponds to recognition with all top-down connections in visual processing hierarchy removed. Pure data-directed vision.

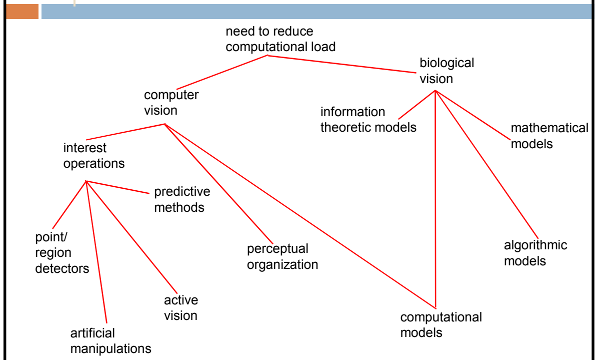
Theorem 1: Unbounded Visual Search is NP-Complete.

Bounded Visual Search

Recognition with knowledge of a target and task in advance, and that knowledge is used to optimize the process.

Theorem2: Bounded Visual Search has time complexity linear in the number of test image pixel locations.

A Road Map....

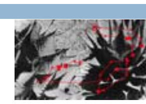


Finding "interesting" information

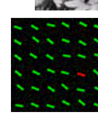
- In principle, very complex task:
  - Need to attend to all objects in scene?
  - Then recognize each attended object?
  - Finally evaluate set of recognized objects against behavioral goals?
- In practice, survival depends on ability to quickly locate and identify important information.
- Need to develop simple **heuristics** or approximations:
  - **bottom-up** guidance towards salient locations
  - **top-down** guidance towards task-relevant locations
  - applications?

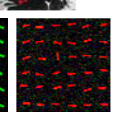
Important Factors

- Local image statistics
  - E.g., Barth et al. '98; Reinagel & Zador '99;
  - Privitera & Stark '00; Parkhurst & Niebur '03;
  - Einhauser et al. `06; Tatler et al. `07

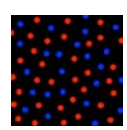


- Spatial outliers – Saliency
  - E.g., Treisman & Gelade '80; Koch & Ullman '85;
  - Tsotsos et al. '95; Li, '98; Itti, Koch & Niebur `98;
  - Bruce & Tsotsos `06; Gao & Vasconcelos `07;
  - Zhang et al. `07





- Temporal outliers – Novelty
  - E.g., Mueller et al. '99; Markou & Singh '01;
  - Theeuwes '95; Fecteau & Munoz '04

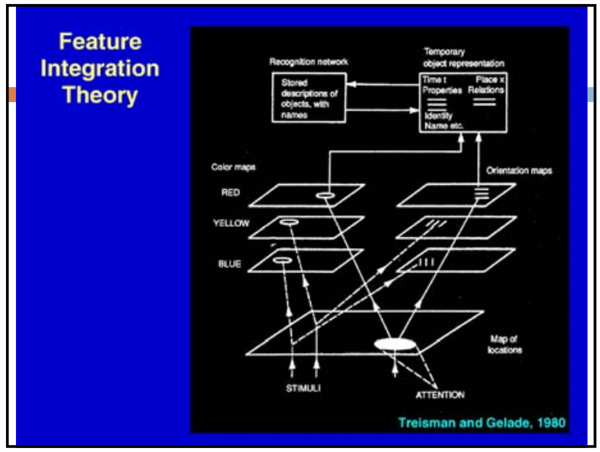


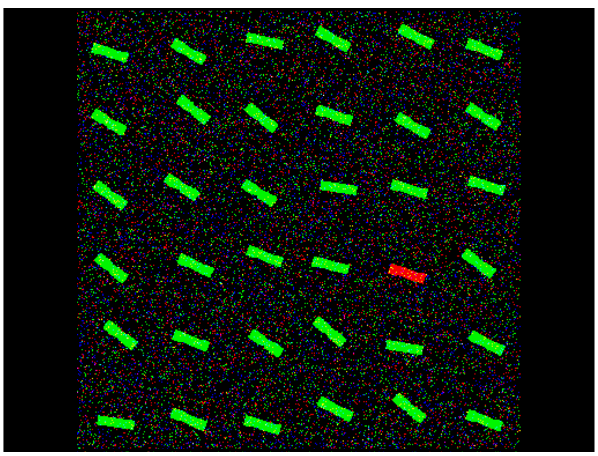
A few definitions

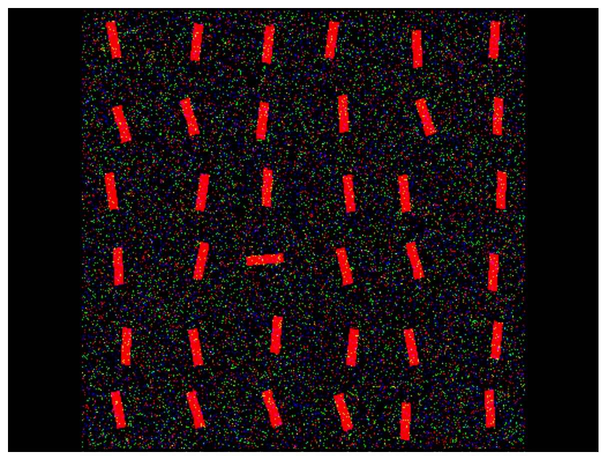
Attention and eye movements:
- overt attention (with eye movements)
- covert attention (without eye movements)

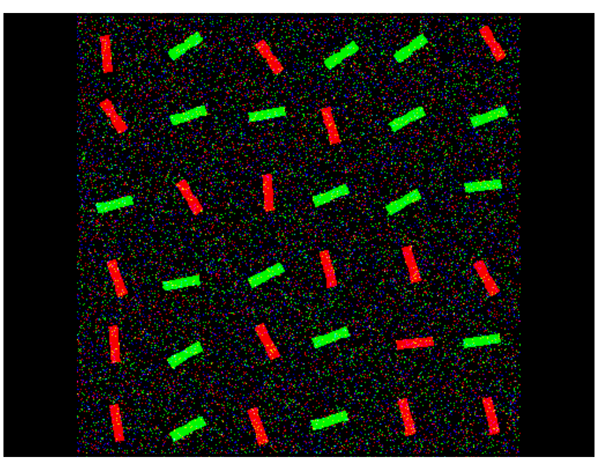
Bottom-up and top-down control:
- bottom-up control
    - based on image features
    - very fast (up to 20 shifts/s)
    - involuntary / automatic
- top-down control
    - may target inconspicuous locations in visual scene
    - slower (5 shifts/s or fewer; like eye movements)
    - volitional

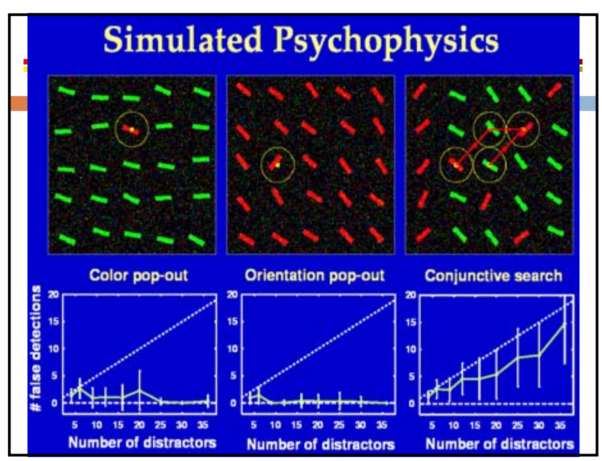
Control and modulation:
- direct attention towards specific visual locations
- attention modulates early visual processing at attended location









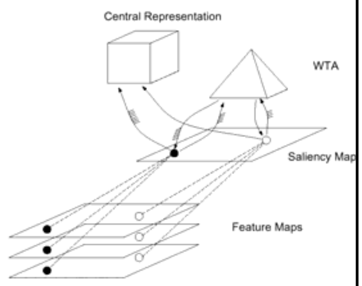


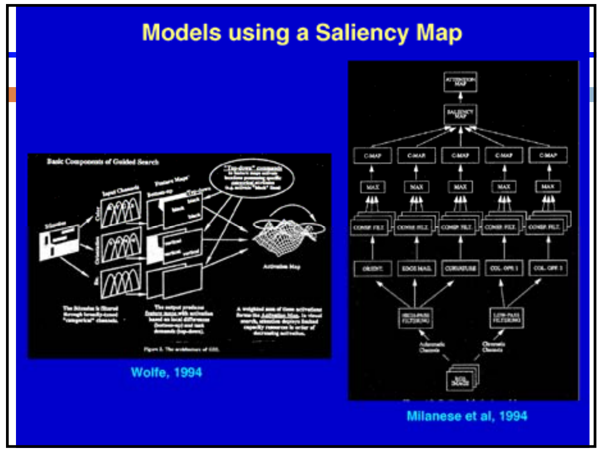
First computational model

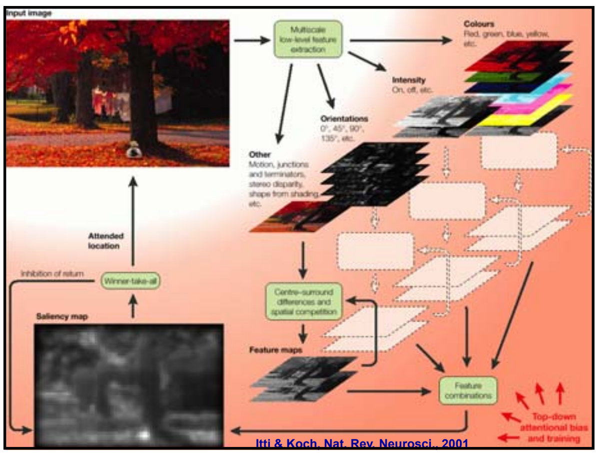
Koch & Ullman,
Hum. Neurobiol., 1895

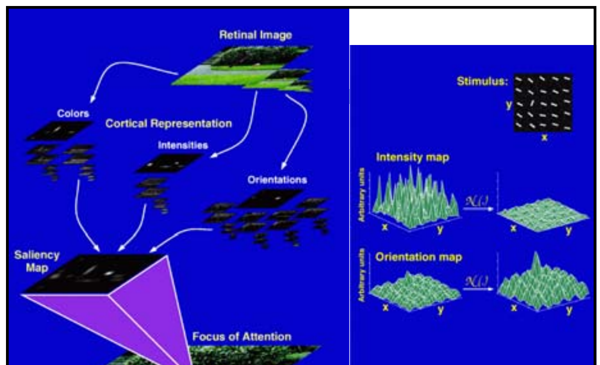
Introduce concept of a single topographic saliency map.

Most salient location selected by a winner-take-all network.

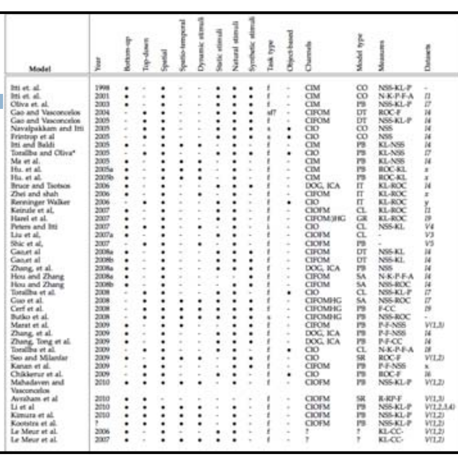








Itti, Koch & Niebur,
IEEE PAMI 1998



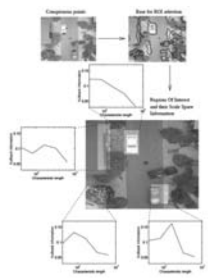
Borji & Itti

Why so much interest?

- Lots of applications – even if it doesn't necessarily "solve" attention, theres a lot that can be done
- An important part of the overall process

Computer Vision Classics

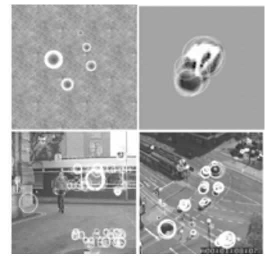
- Jagersand (ICCV 1995)



- 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



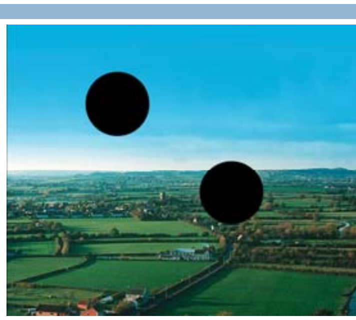
Some basic background

- 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

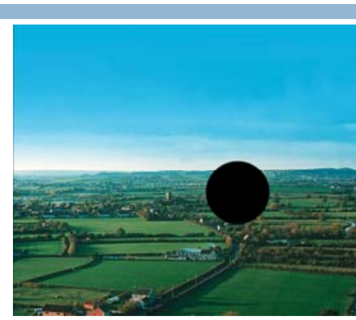
$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}.$$

Visual content and expectation

Visual content and expectation



Visual content and expectation



Visual content and expectation

A note on "categorization"

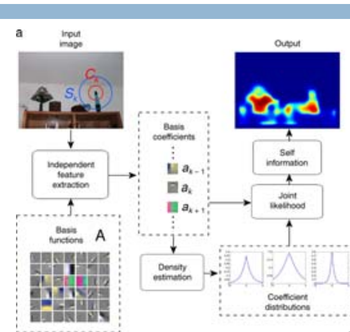
- Information theory, coding, Bayesian inference, Graphical models aren't easily separated
- Grouped thematically, but several may be present within any single model

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 (in a general sense)
- Likelihood related to self-information via –log(p(x))

The Model (AIM)

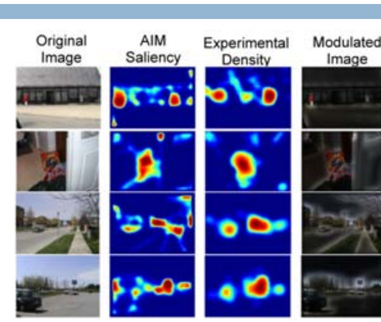


(Bruce and Tsotsos, NIPS 2005, JoV 2009)

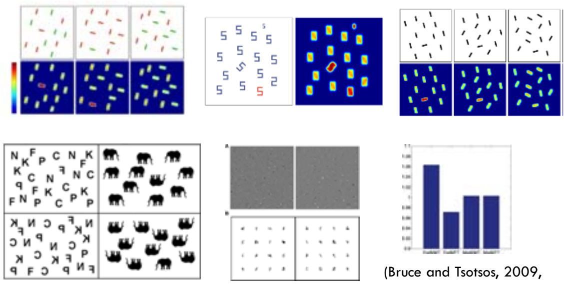
Quantifying Performance

- There now exist a number of datasets, benchmarks, performance metrics, etc.
  - Benchmarking will be discussed later!!
- Different data sets, methodology and parameters
- There are also distinct problems that get called "saliency"

Prediction of fixation patterns

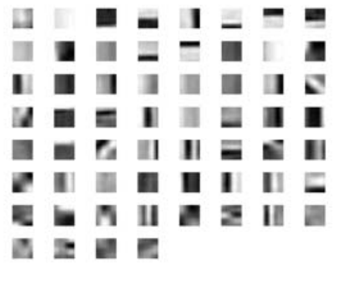


Behavioral phenomena

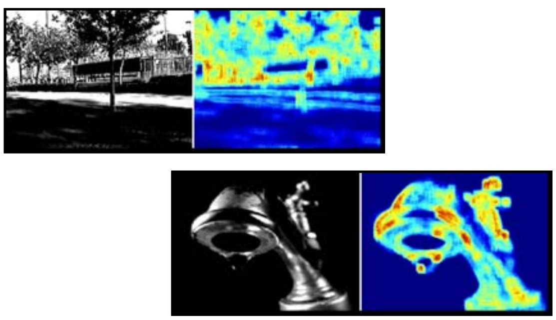


(Bruce and Tsotsos, 2009, Bruce and Tsotsos 2011)

Spatiotemporal Cells

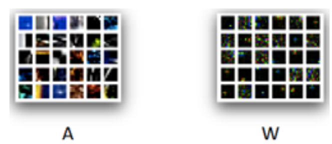


Examples



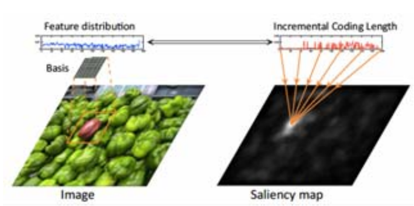
Incremental Coding Length

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



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

Incremental Coding Length



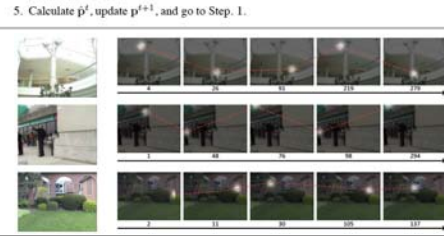
Finally, salience may be computed, with $M=[m_1,m_2,\ldots,m_n]$

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

Incremental Coding Length

Dynamic Visual Attention

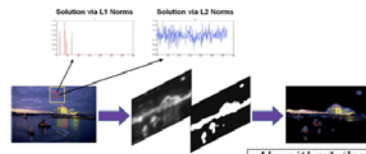
1. At time t, calculate feature ICL based on $\mathbf{p}^t$
2. Given current eye fixation, generate a saliency map with foveal bias.
3. By a saccade, move eye to the global maximum of the saliency map.
4. Sample top N "informative" (largest ICL) features in fixation neighborhood. (In our experiment,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



***Algorithm*1** (Incremental Sparse Saliency)

1. ***Input*** : given image I
2. ***for*** each patch c of the image I, calculate $x = Fc$ and take patches from its surroundings to form S
   - solve the optimization problem $min\ \lambda\|w\|_1 + \frac{1}{2}\|x - Sw\|_2^2$
   - given the sparse solution 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

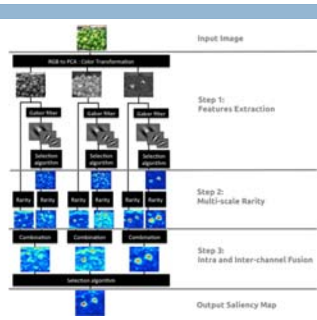
- Rosenholtz (Vis. Res. 1999, JoV 2007, ACM TAP 2011)
  - Distribution of features determined (e.g. in color-space)
  - Mean and covariance of distractors computed
  - Target saliency given by Mahalanobis distance given target, and mean/covariance of distractor distribution
  - Later versions also account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

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

RARE 2012



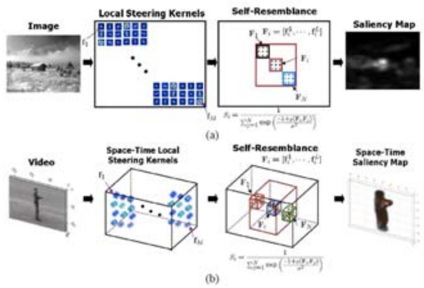
Self-resemblance

- 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

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

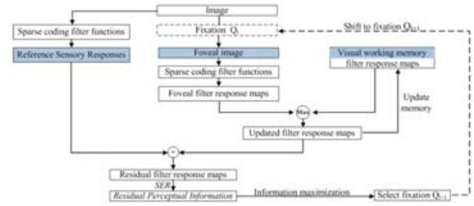
$$S_i = \frac{1}{\sum_{j=1}^{N}\exp\left(\frac{-1+\rho(\mathbf{F}_i,\mathbf{F}_j)}{\sigma^2}\right)}$$

Self-resemblance



Site-Entropy Rate

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



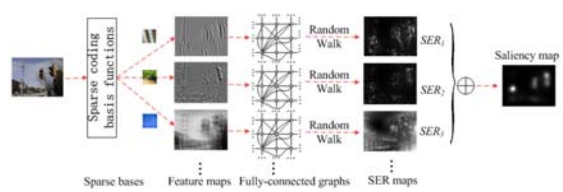
Site-Entropy Rate

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

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

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

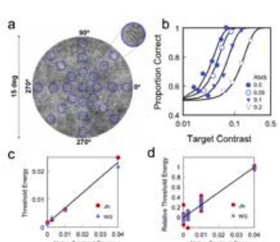
Site-Entropy Rate



- Random walks: See also Achanta et al. 2009

Information Gain

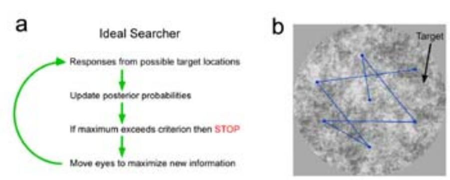
- Najemnik and Geisler



- 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

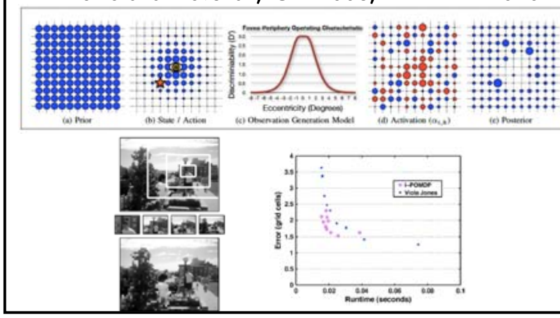
- Najemnik and Geisler (Nature 2005)



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

Information Gain

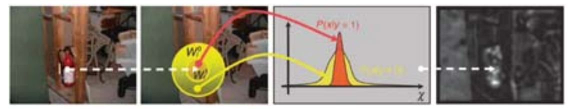
- Butko and Movellan, ICDL 2008, IEEE TAMD 2010



Discriminant / Decision Theoretic Saliency

- Spatial definition for "c"

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



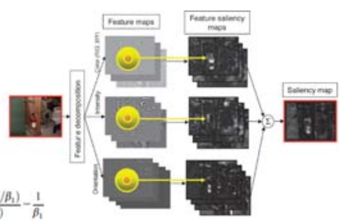
Decision Theoretic Saliency

- Diagrams images

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

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

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

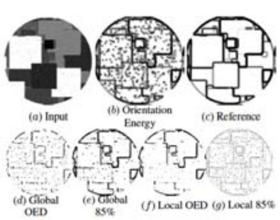


Discriminant / Decision Theoretic Saliency

- 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)

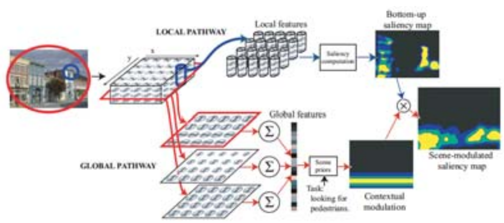


$$g(E) = \frac{\mathcal{N}(E;0,\sigma_h^2)}{\sum_{x\in B_h}\mathcal{N}(x;0,\sigma_h^2)}$$

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

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

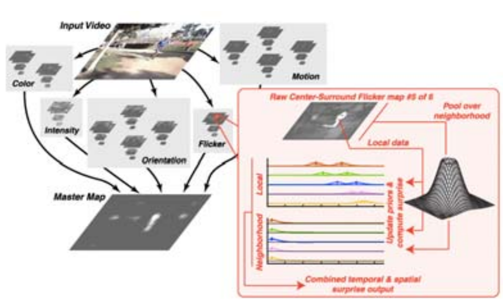
$$\log s_z = \underbrace{-\log p(F=f_z)}_{\substack{\text{Self-information:}\\ \text{Bottom-up saliency}}} + \underbrace{\log p(F=f_z \mid C=1)}_{\substack{\text{Log likelihood:}\\ \text{Top-down knowledge}\\ \text{of appearance}}} + \underbrace{\log p(C=1 \mid L=l_z)}_{\substack{\text{Location prior:}\\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

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



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

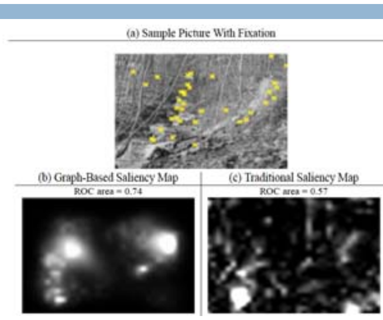
$$d((i,j)\|(p,q)) \triangleq \left|\log \frac{M(i,j)}{M(p,q)}\right|$$

$$\begin{aligned} w_1((i,j),(p,q)) &\triangleq d((i,j)\|(p,q)) \cdot F(i-p, j-q), \text{ where} \\ F(a,b) &\triangleq \exp\left(-\frac{a^2+b^2}{2\sigma^2}\right). \end{aligned}$$

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



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

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

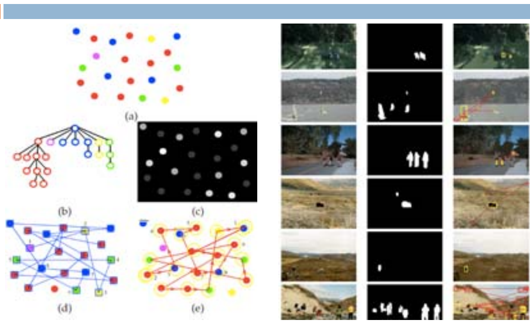
- Marginalization considering most likely assignments:

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

The saliencies are then:

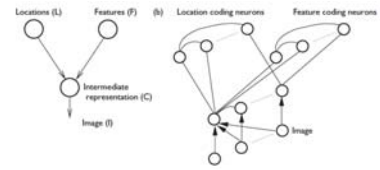
$$p_T(c_i) = \sum_{j=1}^{N} p'(\bar{l}^j) \cdot l_i^j$$

E-saliency



Probabilistic and Bayesian models

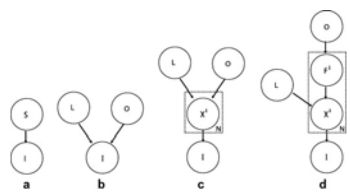
- Rao, NeuroReport 2005



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

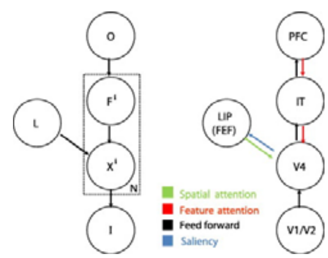
Probabilistic and Bayesian models

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



Graph Based techniques

- Strongly inspired by biology

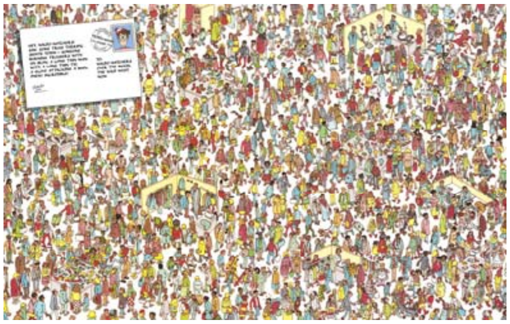


Learning/Object detection Methods

- 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… which you'll hear about

BEYOND SALIENCY?

Beyond saliency



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
- There are several benchmarks (which are important) but can influence research direction
- Saliency is useful for many purposes – but won't solve everything