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

Contact Information:
bruce@cs.umanitoba.ca
www.cs.umanitoba.ca/~bruce

The Human Ability to Attend

- Our focus: visual saliency

FUNDAMENTALS AND IMPORTANT MODELS

Neil Bruce
Department of Computer
University of Manitoba
Winnipeg, MB, Canada

The Human Ability to Attend

- The Computational Argument for Attention

Important Theories and Models

- Coding based approaches
- Graph-based strategies
- Bayesian approaches
- Spectral domain

Overview

- The Human Ability to Attend
- The Computational Argument for Attention
- Important Theories and Models

The Classic Cocktail Party Effect

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

Shadowing Task


- 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

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

Early Selection Model

Deutsch/Norman/Moray/MacKay 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


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

Vestibulo-Ocular Reflex: stabilization visual image on retina by causing compensatory changes in eye position on head moves

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

Optokinetic Nystagmus: stabilization gaze during sustained, low-frequency rotations at constant velocity

Smooth Pursuit: voluntary tracking of moving stimulus

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

Torsion: coordinated rotation of the eyes around optical axis, dependent on head tilt and eye elevation

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 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


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.

Covert fixations is strongly linked to awareness of covert attention. Overt orienting, whether of the eyes or the head as built-in related movement of covert fixation shifts are called engagement.

Engagement of gaze direction is controlled reflexively by external stimulation while engagement of gaze is controlled by internally generated signals.

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

Visual Search: Attentional Cueing


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

(Barnett-D, J. (1970). The role of visual and semantic codes in object naming)
**Visual Search**


**Feature Search**
- 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.

**Conjunction Search**

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

**Feature Integration Theory (FIT)**


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 Wolfe 1998

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

**Guided Search 1989**


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

**Change Blindness**


Precursors: visual memory - Observers were found to be poor at detecting change if old and new displays were separated by an IS 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?**


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
- Texture
- Scale
- Vernier Offset
- Size, Spatial Frequency, and Scale
- Onset/Offset
- Shape
- Closure/Overlap
- Pictorial Depth Cues
- Stereoscopic Depth

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

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

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

**Superior Colliculus**

**Parietal Cortex**

**Frontal Eye Fields**

**V1 and V2**

**LGN**

**Pulvinar**

**Visual Saliency Map Locus**
- J. Tsotsos, S.E. Petersen, Bounded Visual Search has time complexity linear in the number of test points, Dept. of Computer Science, University of British Columbia, Vancouver, Canada, UBC CS Technical Report 89-22 (September 1989).

For Further Reading...

**NEUROBIOLOGY OF ATTENTION**
- Laurent Tsotsos, University of Southern California
- Geraint Rees, University College London
- John Tsotsos, York University

ISBN: 0-12-375731-2
Pages: 744
www.elsevier.com
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.


Borji & Itti, IEEE PAMI 1998

Why so much interest?

- Lots of applications — even if it doesn't necessarily “solve” attention, there's a lot that can be done
- An important part of the overall process
KL-divergence across scale space representation
Peaks provide a sense of scale of interest

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

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)

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

Spatiotemporal Cells

Examples

Incremental Coding Length

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

$$w_k = \sum_{i \in S} d_i w_i x^k$$
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

Probability/Clutter

  - 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

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

\[ K(x_i, x_j) = \frac{1}{\sqrt{h^2}} \exp \left\{ -\frac{(x_i - x_j)^2}{2h^2} \right\}, \quad C_i \in \mathbb{R}^{2 \times 2} \]

\[ S_i = \frac{1}{\sum_{j=1}^{N} \exp \left( -\frac{1}{\sigma^2} \|F_i - F_j\|^2 \right) } \]
Self-resemblance

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

\[
S_I = \sum_k S_R_{kl} = -\sum_k \left( \pi_k \sum_{l} P_{kl} \log P_{kl} \right)
\]

\( \pi_k \) - Stationary distribution term (frequency with which random walker visits node \( k \) / frequency with which node \( k \) communicates with other nodes)

Random walks: See also Achanta et al. 2009

Information Gain

- Najemnik and Geisler (Nature 2005)
- 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

\[
R_t = \frac{\text{prior}(t)\exp(d_{t}^{2}W_t)}{\sum_{j \neq t} \text{prior}(j)\exp(d_{j}^{2}W_j)}
\]
Information Gain

- Butko and Movellan, ICDL 2008, IEEE TAMD 2010

Discriminant / Decision Theoretic Saliency

- Spatial definition for “c”
  \[ s(d) = \frac{I(X;Y)}{\sum_{x \in X} \sum_{y \in Y} \frac{P(X|x,y) \log \frac{P(X|x,y)}{P(X|x)P(Y|y)}}{P(X|x)P(Y|y)}} \]

- 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

Decision Theoretic Saliency

- Diagrams images
  \[ p_c(x) = \frac{1}{2\pi \sigma^2} \exp \left\{ -\frac{1}{2\sigma^2} \right\} \]

- Spatial saliency
  \[ s(d) = \frac{I(X;Y)}{\sum_{x \in X} \sum_{y \in Y} \frac{P(X|x,y) \log \frac{P(X|x,y)}{P(X|x)P(Y|y)}}{P(X|x)P(Y|y)}} \]

- 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

Suspicious coincidences

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

Bayesian Approaches
Probabilistic and Bayesian models

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

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

\[ s_z = \frac{p(C = 1 | F = f_z, L = l_z)}{p(F = f_z, L = l_z) \cdot p(C = 1)} \]

\[ \log s_z = -\log p(F = f_z) + \log p(F = f_z | C = 1) + \log p(C = 1 | L = l_z) \]

Self-information: Bottom-up saliency

Log likelihood: Top-down knowledge of appearance

Location prior: Top-down knowledge of target's location

Probabilistic and Bayesian models


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
    \[ d(i, j) = \log \frac{M(i, j)}{M(p, q)} \]
    \[ w(i, j, p, q) = d(i, j) \cdot F(i - p, j - q), \text{ where} \]
    \[ F(a, b) = \exp \left( -\frac{a^2 + b^2}{2\sigma^2} \right). \]

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

- **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)**
  - Labels are binary random variables:
    \[ p(l_i | l_j) = \frac{p(l_i | l_j)}{\sum_{l_i | l_j} p(l_i | l_j)} \]
  - Marginalization considering most likely assignments:
    \[ p(l) = \frac{p(l)}{\sum_{l} p(l)} \]
    The saliencies are then:
    \[ p_{l} = \sum_{l} p(l) \cdot t_l \]

E-Saliency

- **Dependency on parent nodes for label**
  \[ p(l_i) = \prod_{l_j \in \text{parent}(l_i)} p(l_j) \]

- **Marginalization considering most likely assignments:**
  \[ p(l) = \frac{p(l)}{\sum_{l} p(l)} \]
  The saliencies are then:
  \[ p_{l} = \sum_{l} p(l) \cdot t_l \]
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