FROM ATTENTION TO OBJECT PROPOSALS
PhD. Depth exam

PhD. student: Ivan Bogun
Advisor: Dr. Eraldo Ribeiro

ibogun2010@my.fit.edu

Thursday 5th March, 2015

Computer vision Lab
Department of Computer Science
Florida Institute of Technology
Melbourne, FL 32901, USA
TABLE OF CONTENTS

Introduction
  What is attention

Applications

Physiologically-inspired saliency detection
  Feature-based saliency
  Information maximization
  Spectral methods

Segmentation-based methods
  Global contrast methods
  Multi-cue contrast methods for object proposals

Data-driven methods
  SVMs for saliency detection
  Low-rank factorization
  Deep learning
INTRODUCTION
Attention

Salience - the state or quality by which it stands out relative to its neighbors.

Salient can be:
- object
- person
- pixel
Attention

- the act or faculty of attending, especially by directing the mind to an object. 

Attention
- the act or faculty of attending, especially by directing the mind to an object. ¹

Saliency

¹http://dictionary.reference.com/browse/attention
Attention
- the act or faculty of attending, especially by directing the mind to an object. ¹

Saliency
- is the state or quality by which it stands out relative to its neighbors. ²

¹http://dictionary.reference.com/browse/attention
²http://en.wikipedia.org/wiki/Salience_(neuroscience)
Attention

- the act or faculty of attending, especially by directing the mind to an object. ¹

Saliency

- is the state or quality by which it stands out relative to its neighbors. ² Salient can be
  · object
  · person
  · pixel

¹http://dictionary.reference.com/browse/attention
²http://en.wikipedia.org/wiki/Salience_(neuroscience)
PROPERTIES OF THE BRAIN

- Limited neural capacity
- Energy consumption is dominated by neural activity (Attwell and Laughlin, 2001)
- Energy consumption is constant over time (Sokoloff, 1989)

Brain optimizations
- Efficient representation (e.g., sparse codes Barlow (2009))
- Resources are dynamically allocated according to processing demands.
Physiological facts about our brain

- Limited neural capacity
- Energy consumption is dominated by neural activity (Attwell and Laughlin, 2001)
- Energy consumption is constant over time (Sokoloff, 1989)
- Efficient representation (e.g., sparse codes Barlow, 2009)
- Resources are dynamically allocated according to processing demands.
Physiological facts about our brain

- Limited neural capacity
Physiological facts about our brain

- Limited neural capacity
- Energy consumption is dominated by neural activity (Attwell and Laughlin, 2001)
Physiological facts about our brain

- Limited neural capacity
- Energy consumption is dominated by neural activity (Attwell and Laughlin, 2001)
- Energy consumption is constant over time (Sokoloff, 1989)
Physiological facts about our brain

- Limited neural capacity
- Energy consumption is dominated by neural activity (Attwell and Laughlin, 2001)
- Energy consumption is constant over time (Sokoloff, 1989)

Brain optimizations

- Efficient representation (e.g., sparse codes Barlow (2009))
Physiological facts about our brain

- Limited neural capacity
- Energy consumption is dominated by neural activity (Attwell and Laughlin, 2001)
- Energy consumption is constant over time (Sokoloff, 1989)

Brain optimizations

- Efficient representation (e.g., sparse codes Barlow (2009))
- Resources are dynamically allocated according to processing demands.
Biased-competition hypothesis
Biased-competition hypothesis

Stimuli in the visual field activate neurons which engage in competitive interaction. When observer attends a given location its neurons become more active, while all others are suppressed (Desimone and Duncan, 1995).
Biased-competition hypothesis

Stimuli in the visual field activate neurons which engage in competitive interaction. When observer attends a given location its neurons become more active, while all others are suppressed (Desimone and Duncan, 1995).

Evidence supporting the hypothesis was found in:

- Electrophysiological studies (Beck and Kastner, 2009)
- Neuroimaging (Reynolds and Chelazzi, 2004)
- Behavioral research (Desimone and Duncan, 1995)
Attention for “active vision”

Artificial researchers used attention as a preprocessing step in the “active vision” of their robots: framework where the robots would mimic human attention processes.
Attention for “active vision”

Artificial researchers used attention as a preprocessing step in the “active vision” of their robots: framework where the robots would mimic human attention processes.

Application-specific saliency

Applications based on attention were found in computer vision, such as:

- Content-aware resizing
- Image collection summarizing
- Thumb-nailing
Problem

Resize the image without changing it’s content.
PHYSIOLOGICALLY-INSPIRED
SALIENCY DETECTION
Introduction

What is attention

Applications

Physiologically-inspired saliency detection

  Feature-based saliency
  Information maximization

Spectral methods

Segmentation-based methods

  Global contrast methods
  Multi-cue contrast methods for object proposals

Data-driven methods

  SVMs for saliency detection
  Low-rank factorization

Deep learning
FEATURE-BASED SALIENCY

Feature-integration theory
Feature-integration theory

Feature-integration theory


Empirical evidence

Attention can select feature values within dimension (Carrasco, 2009; Müller et al., 2006), e.g.,

- vertical vs horizontal orientation
- upward vs downward motion
- color

Properties

- color and line orientation as features
- saliency was defined as a linear combination of feature maps
- each feature map could be amplified or inhibited via amplification parameters
- amplification parameters were set depending on the context
Figure 1: Diagram of the method in James and Ferrier (1988)
Definition

*Visual acuity* is the capacity to resolve two nearby objects.
Definition

*Visual acuity* is the capacity to resolve two nearby objects.

\[
\frac{1}{\theta} \text{ law}
\]

Human-vision acuity is inversely proportional to the viewing angle, thus locations closer to the center of focus are more clear while the ones far away are more blurry.

Corrolary

Sharpness of the object does not depend on the distance to it, only on the viewing angle.
Figure 2: Visualization of the $\frac{1}{\theta}$ attention law.
Gaussian pyramids

Burt (1988) used multi-scale Gaussian pyramids to approximate $\frac{1}{\theta}$ acuity distribution for attention.
Gaussian pyramids

Burt (1988) used multi-scale Gaussian pyramids to approximate $\frac{1}{\theta}$ acuity distribution for attention.

Why $\frac{1}{\theta}$?

Burt (1988) showed that $\frac{1}{\theta}$ provides a fine balance between obtaining a wide field view and high acuity, which are subject to constraint of the size of the optic nerve.
Model by Koch and Ullman (1985)

The model was build on the following principles:

- Number of feature maps computed in parallel
- Feature maps are selectively combined into a central saliency map
- Attended location is chosen as a winner-take-all network
- Visited locations are inhibited from further attention.

- Low-pass filter on 9 different scales
- Intensity feature map from each of 3 color channels
- Orientation feature were computed using Gabor filters
- Contrast maps were calculated as a difference between finer and coarse scales
- After normalization, each of the features was combined into a saliency map
- Winner-take-all network served as a scheduling mechanism for attending consecutive locations
Figure 3: Example of the implementation (Itti et al., 1998).
Alternative point of view: Lee and Yu (1999)

- Attention as a way to maximize information from the image
  - Attend location first with largest uncertainty (e.g., entropy)
  - Entropy is calculated over neurons from visual cortex
  - Discontinue attended locations with mutual information

Properties

- "Inhibition-of-return" is satisfied implicitly (e.g., little information gain over visiting same place twice)
- Attention as a way to maximize information
Alternative point of view Lee and Yu (1999)

- Attention as a way to maximize information from the image
- Attend location first with largest uncertainty (e.g., entropy)
- Entropy is calculated over neurons from visual cortex
- Discontinue attended locations with mutual information
Alternative point of view Lee and Yu (1999)

- Attention as a way to maximize information from the image
- Attend location first with largest uncertainty (e.g., entropy)
- Entropy is calculated over neurons from visual cortex
- Discontinue attended locations with mutual information

Properties

- ”Inhibition-of-return” is satisfied implicitly (e.g., little information gain over visiting same place twice)
- Attention as a way to maximize information
Silhouettes

Attention from silhouettes (e.g., binary shapes) (Renninger et al., 2005)

- Representation with set of points and tangent directions (e.g., edgelets)
- Feature vector given as histograms of edges
- Location selection via entropy maximization
SALIENCY AS NON-REDUNDANCY

Barlow (1961)

- Visual system reduces redundant visual information from processing
- Input image decomposed into novel (e.g., salient) and redundant parts
Barlow (1961)

- Visual system reduces redundant visual information from processing
- Input image decomposed into novel (e.g., salient) and redundant parts

Spectral residual Hou and Zhang (2007)

- Same distribution of log intensity vs frequency, independent of the image
- Distribution can be approximated by a linear function
- Salient are locations causing discontinuities on the approximation
SEGMENTATION-BASED METHODS
Introduction
  What is attention
Applications
Physiologically-inspired saliency detection
  Feature-based saliency
  Information maximization
Spectral methods
Segmentation-based methods
  Global contrast methods
  Multi-cue contrast methods for object proposals
Data-driven methods
  SVMs for saliency detection
  Low-rank factorization
Deep learning
Figure 2: A street scene (320 × 240 color image), and the segmentation results produced by our algorithm (σ = 0.8, k = 300).

Figure 3: A baseball scene (432 × 294 grey image), and the segmentation results produced by our algorithm (σ = 0.8, k = 300).

Figure 4: An indoor scene (image 320 × 240, color), and the segmentation results produced by our algorithm (σ = 0.8, k = 300).

Segmentation example (Felzenszwalb and Huttenlocher, 2004).

- Flexible representation
- Automatic redundancy removal
- Segmentation is object-independent
Figure 2: A street scene (320 × 240 color image), and the segmentation results produced by our algorithm ($\sigma = 0.8, k = ...$

Figure 3: A baseball scene (432 × 294 grey image), and the segmentation results produced by our algorithm ($\sigma = 0.8, k = 300$).

Figure 4: An indoor scene (image 320 × 240, color), and the segmentation results produced by our algorithm ($\sigma = 0.8, k = 300$).

Figure 4: Segmentation example (Felzenszwalb and Huttenlocher, 2004).
Figure 2: A street scene (320 × 240 color image), and the segmentation results produced by our algorithm (σ = 0.8, k = 300).

Figure 3: A baseball scene (432 × 294 grey image), and the segmentation results produced by our algorithm (σ = 0.8, k = 300).

Figure 4: An indoor scene (image 320 × 240, color), and the segmentation results produced by our algorithm (σ = 0.8, k = 300).

Figure 4: Segmentation example (Felzenszwalb and Huttenlocher, 2004).

- Flexible representation
- Automatic redundancy removal
- Segmentation is object-independent
Define region salient if it is different from all others in the image.

- Cheng and Zhang (2011) used color histograms, weighted by Gaussians depending on a spatial location
- Perazzi et al. (2012) used Euclidean distance on color intensities, but used efficient implementation whose complexity was linear in terms of the regions
Feng et al. (2011)

- Region is salient if it is hard to compose it from other regions in the image
Feng et al. (2011)

- Region is salient if it is hard to compose it from other regions in the image

**Metric**

Feng et al. (2011) defined saliency metric as the cost of representing the region as a linear combination of appearance of other regions weighted with their spatial proximity.
Object proposals

Object proposals is a problem of locating a subset of all possible bounding boxes which are likely to contain the object.

Example

For an image of size 640x480, 1000 of the bounding boxes is 10%.

\[ O(N^2M^2) \]

Evaluate classifier on every bounding box in the image (Dalal and Triggs, 2004).

Sliding window
Sliding window

- Evaluate classifier on every bounding box in the image (Dalal and Triggs, 2004)
- For an image of size $N \times M$ there are $O(N^2M^2)$ possible bounding boxes
Sliding window

- Evaluate classifier on every bounding box in the image (Dalal and Triggs, 2004)
- For an image of size $N \times M$ there are $O(N^2M^2)$ possible bounding boxes

Object proposals

*Object proposals* is a problem of locating a subset of all possible bounding boxes which are likely to contain the object.
Sliding window

- Evaluate classifier on every bounding box in the image (Dalal and Triggs, 2004)
- For an image of size $N \times M$ there are $O(N^2M^2)$ possible bounding boxes

Object proposals

*Object proposals* is a problem of locating a subset of all possible bounding boxes which are likely to contain the object.

Example

For an image of size $640 \times 480$ 1000 of the bounding boxes is $10^{-6}$%
To generate object proposals Uijlings and van de Sande (2013) used hierarchical segmentation where the merging metric depended on

- color
- texture
- size
- object’s fill
To generate object proposals Uijlings and van de Sande (2013) used hierarchical segmentation where the merging metric depended on

- color
- texture
- size
- object’s fill

To diversify proposals same procedure was applied in different color spaces (e.g., RGB, HSV, LAB)
Alexe et al. (2010, 2012) pointed out that salient object also has to have well-defined closed boundary.

- **Edge density**: Calculate the fraction of the edges on the boundary of the bounding box, compared to the slightly enlarged bounding box.

- **Superpixel straddling**: Bounding box with the object is likely to occupy fewer segmentation regions (e.g., superpixels) and cut them into big parts.
Alexe et al. (2010, 2012) pointed out that salient object also has to have well-defined closed boundary.
Alexe et al. (2010, 2012) pointed out that salient object also has to have well-defined closed boundary.

**Edge density**

Calculate the fraction of the edges on the boundary of the bounding box, compared to the slightly enlarged bounding box.
Alexe et al. (2010, 2012) pointed out that salient object also has to have well-defined closed boundary.

**Edge density**

Calculate the fraction of the edges on the boundary of the bounding box, compared to the slightly enlarged bounding box.

**Superpixel straddling**

Bounding box with the object is likely to occupy fewer segmentation regions (e.g., superpixels) and cut them into big parts.
Figure 5: Edge density visualization (Alexe et al., 2012).
Figure 6: Superpixel straddling visualization (Alexe et al., 2012).
Critique

Although the method Alexe et al. (2012) achieved state-of-the art on multiple benchmarks, it was rather slow.
Critique

Although the method Alexe et al. (2012) achieved state-of-the art on multiple benchmarks, it was rather slow.

Cascades

Rahtu et al. (2011) used cascades to prune bad locations earlier

Randomization

Manen et al. (2013) used randomized Prim’s algorithm to “grow” object proposal from segmentation regions. Each proposal was created as a partial spanning tree with high sum of edge weights.
Which proposal is better than other?

- Define “objectness” metric
- Rank each bounding box according to the objectness metric
Which proposal is better than other?

- Define “objectness” metric
- Rank each bounding box according to the objectness metric

- Hoiem et al. (2011); Endres et al. (2014) used region features based on color, texture, and boundary strength to learn boosted decision-tree classifier and ranked structured support vector machine for ranking

- Carreira and Sminchisescu (2012) cast segmentation as a max-flow problem where ranking is solved as a regression problem solved by random forests
DATA-DRIVEN METHODS
Introduction

What is attention

Applications

Physiologically-inspired saliency detection

Feature-based saliency

Information maximization

Spectral methods

Segmentation-based methods

Global contrast methods

Multi-cue contrast methods for object proposals

Data-driven methods

SVMs for saliency detection

Low-rank factorization

Deep learning
Psychologically-inspired methods critique

- Manually engineered feature maps
- Choice of weights in the linear combination of the saliency map
Psychologically-inspired methods critique

- Manually engineered feature maps
- Choice of weights in the linear combination of the saliency map

SVMs Kienzle et al. (2006)

- Data-driven way of feature maps (images themselves)
- Automatic way of the weights (support vector coefficients)
- Linear combination of non-linear maps as a decision rule
Objectness

Cheng et al. (2014) build approach which would satisfy the following
Objectness

Cheng et al. (2014) build approach which would satisfy the following

- Achieve high detection rate
- Generate small number of object proposals
- Be fast to evaluate
- Generalize to unseen objects
Objectness

Cheng et al. (2014) build approach which would satisfy the following:

- Achieve high detection rate
- Generate small number of object proposals
- Be fast to evaluate
- Generalize to unseen objects

To achieve this Cheng et al. (2014) used binarized normed (BING) gradients with linear SVM.
Objectness

Cheng et al. (2014) build approach which would satisfy the following

- Achieve high detection rate
- Generate small number of object proposals
- Be fast to evaluate
- Generalize to unseen objects

To achieve this Cheng et al. (2014) used binarized normed (BING) gradients with linear SVM.

Speed

0.003 seconds/image.
To find generic objects within an image, we scan over a predefined quantized window sizes (scales and aspect ratios). As illustrated in Fig. 1(d), the learned linear model $w$ (see Sec. 4 for experimental settings), looks similar to the input image to 36 sizes so that the learned NG features correspond to target windows. These features are very useful for detecting objects of arbitrary categories. In other words, NG features are insensitive to change of translation, scale and aspect ratio, which will not affect the existence of object instances, we firstly resize the input image to different sizes so that NG features are less likely than others to contain an object instance.

Our NG feature, as a dense and compact objectness feature of its corresponding window. A 64D region of these resized normed gradients maps are defined as the gradients of each resized image. The values in an image to different sizes can represent different aspects of the image.

To utilize this observation for efficiently predicting object boundaries and centers, we learn a single $64 \times 64$ linear model for selecting filter score as 1D features, and check their labeling using training samples. Their filter scores are used as proposals as training samples, their filter scores are separately learnt coefficient and bias using linear SVM.

In Sec. 3.2, we evaluate the NG feature as a dense and compact representation of the NG feature makes it great potential to be involved in real-time applications.

Figure 7: Visualization of the BING feature (Cheng et al., 2014).
Intuition

- Method to decompose a matrix into sum of sparse plus low-rank matrices (Candès and Recht, 2009)
Intuition

- Method to decompose a matrix into sum of sparse plus low-rank matrices (Candès and Recht, 2009)
- Decompose video into moving/non-moving parts
Intuition

- Method to decompose a matrix into sum of sparse plus low-rank matrices (Candès and Recht, 2009)
- Decompose video into moving/non-moving parts
- Remove noise from the image
Intuition

- Method to decompose a matrix into sum of sparse plus low-rank matrices (Candès and Recht, 2009)
- Decompose video into moving/non-moving parts
- Remove noise from the image

Factorization for saliency detection (Yan et al., 2010)

- Learn overcomplete sparse-code dictionary (Mairal et al., 2010)
- Perform decomposition in the new feature space
Overview

- State-of-the-art method for object recognition (Krizhevsky et al., 2012)
- Powerful representation capabilities
Overview

- State-of-the-art method for object recognition (Krizhevsky et al., 2012)
- Powerful representation capabilities

Saliency as regression

Erhan et al. (2014) cast saliency detection as a regression problem
Overview

- State-of-the-art method for object recognition (Krizhevsky et al., 2012)
- Powerful representation capabilities

Saliency as regression

Erhan et al. (2014) cast saliency detection as a regression problem

Joint segmentation and saliency detection (Wang et al., 2014)

- Trained two convnets for saliency detection and segmentation
- Introduced novel optimization with latent variables
Evaluation criteria

- Independence of the application
- Evaluation protocol for object-generic evaluation
Evaluation criteria

- Independence of the application
- Evaluation protocol for object-generic evaluation

Robustness

- Experiment to see how saliency changes when the image is perturbed with noise
- Repeatability
Possible applications

- Object proposals can improve speed, can they also improve the accuracy?
- Objectness priors in semi-supervised vision tasks (e.g., model-free tracking, weakly supervised action or interaction recognition)
- Extension to the temporal domain in non-trivial manner
video
Questions?


