Learning to Detect A Salient Object

Tie Liu, Jian Sun, Nan-Ning Zheng, Xiaoou Tang, and Heung-Yeung Shum

Presenter: Che-Chun Su
2012/10/26
Outline

• Introduction
• Image Database
• Salient Object Detection
  – CRF Learning
  – Salient Object Features
• Evaluation
• Discussion and Conclusion
Introduction

- Study visual attention by detecting a salient object in an input image.
- People naturally pay more attention to salient objects.
  - A person, a face, a car, an animal, a road sign, etc.
- Formulate salient object detection as image segmentation problem.
  - Separate the salient object from the image background.
Introduction

• Applications for visual attention
  – Automatic image cropping, adaptive image display, image/video compression, advertising design, etc.

• Existing visual attention approaches
  – Bottom-up computational framework
Introduction

• **Difficulty**
  - Although existing approaches work well in finding a few fixation locations, they are not able to accurately detect where visual attention should be.
Introduction

• Contributions
  – The first large image database available for quantitative evaluation
  – High-level concept of salient object for visual attention computation
  – CRF learning framework with a set of novel local, regional, and global features to define a generic salient object
Outline

• Introduction
• Image Database
• Salient Object Detection
  – CRF Learning
  – Salient Object Features
• Evaluation
• Discussion and Conclusion
Different people have different ideas about what a salient object in an image is.

- Voting strategy by multiple users.
Image Database

• Salient object representation
  – A binary mask
    \[ A = \{ a_x \}, \text{ for each pixel } x, a_x \in \{1, 0\} \]

• Image source
  – 130,099 high quality images from a variety of sources
  – 60,000+ images with a salient object or a distinctive foreground object
  – 20,840 images for labeling

• Two-stage labeling process
  – Ask the user to draw a rectangle which encloses the most salient object in the image.
  – Reduce labeling inconsistency with voting.
Image Database

• The first stage
  – 3 users label all 20,840 images.
  – Saliency probability map
    \[ g_x = \frac{1}{M} \sum_{m=1}^{M} a^m_x \]
    
    \( M \): the number of users
    \( A^m = \{ a^m_x \} \): the binary mask labeled by the \( m \)th user
  – Image set \( A \)
  – Labeling consistency
    \[ C_t = \frac{\sum_{x \in \{g_x > t\}} g_x}{\sum_x g_x} \]
Image Database

• The second stage
  – Randomly selected 5000 highly consistent images from the image set $A$ (i.e., $C_{0.9} > 0.8$)
  – 9 users label the salient object rectangle.
  – Image set $B$

• After the two-stage labeling process, the salient object is defined based on the majority agreement of users and represented as a saliency probability map.
Image Database

$C_{0.9}$

Image set $A$

$C_{0.5}$

Image set $B$
Outline

• Introduction
• Image Database
• Salient Object Detection
  – CRF Learning
  – Salient Object Features
• Evaluation
• Discussion and Conclusion
Salient Object Detection

• Formulated as binary labeling problem
• Conditional Random Field (CRF) framework
  – The probability of the label $A = \{a_x\}$ given the image $I$ is modeled as a conditional distribution:

$$P(A|I) = \frac{1}{Z} e^{-E(A|I)}$$

$$E(A|I) = \sum_x \sum_{k=1}^{K} \lambda_k F_k(a_x, I) + \sum_{x,x'} S(a_x, a_{x'}, I)$$
Salient Object Detection

- Conditional Random Field (CRF) framework
  - Get an optimal linear combination of features by estimating the linear weights under the Maximized Likelihood (ML) criteria:

\[
\lambda^* = \arg \max \sum_n \log P(A^n | I^n; \lambda), \quad \lambda = \{\lambda_k\}_{k=1}^K
\]

- Advantages over Markov Random Field (MRF)
  - Arbitrary low-level or high-level features can be used.
  - Provide an elegant framework to combine multiple features with effective learning.
Salient Object Features

• Multi-scale contrast
  – Contrast is the most commonly used local feature because the contrast operator simulates the human visual receptive fields.
  – A linear combination of contrasts in the Gaussian image pyramid:

\[
f_c(x, I) = \sum_{l=1}^{L} \sum_{x' \in N(x)} ||I_l(x) - I_l(x')||^2
\]
Salient Object Features

- Center-surround histogram
  - Salient objects usually have a larger extent than local contrast and can be distinguished from its surrounding context.
  - Measure how distinct the salient object is with respect to its surrounding area, using the distance between color histograms.
Salient Object Features

• Center-surround histogram
  – Sum of spatially weighted distances:
Salient Object Features

• Center-surround histogram

Non-rectangular shape of salient object?
Other visual cues?
Salient Object Features

• Color spatial distribution
  – The wider a color is distributed in the image, the less possible a salient object contains this color.
  – Spatial variance of color, horizontal and vertical:

\[
p(c|I_x) = \frac{w_cN(I_x|\mu_c, \Sigma_c)}{\sum_c w_cN(I_x|\mu_c, \Sigma_c)} , \quad N : \text{Gaussian Mixture Model}
\]
Salient Object Features

• Color spatial distribution
  – The spatial variance of color at image corners or boundaries may also be small because the image is cropped from the whole scene.
  – Center-weighted, spatial-variance color feature:

\[ f_s(x, I) \propto \sum_c p(c|I_x)(1 - V(c))(1 - D(c)) \]
Salient Object Features

• Color spatial distribution

Non-centered salient object?
Outline

• Introduction
• Image Database
• Salient Object Detection
  – CRF Learning
  – Salient Object Features
• Evaluation
• Discussion and Conclusion
Evaluation

- Effectiveness of features and CRF learning

1. multi-scale contrast, 2. center-surround histogram, 3. color spatial distribution, 4. combination

\[
\text{Precision} = \frac{\sum_x g_x a_x}{\sum_x a_x}, \quad \text{Recall} = \frac{\sum_x g_x a_x}{\sum_x g_x}, \quad \text{F-measure} = \frac{(1 + \alpha) \times \text{Precision} \times \text{Recall}}{\alpha \times \text{Precision} + \text{Recall}}
\]
Evaluation

- Effectiveness of features and CRF learning

Contribution of contrast?
Evaluation

• Comparison with other approaches
  – Recall rate is not much of a useful measure in visual attention.
Evaluation

• Comparison with other approaches
  – Recall rate is not much of a useful measure in visual attention.

(c) BDE, image set $A$
(d) BDE, image set $B$
BDE: boundary displacement error
Evaluation

• Comparison with other approaches
  – The real challenge: high precision on small salient objects
    • Object/image ratio in the range \([0, 0.25]\)
Outline

• Introduction

• Image Database

• Salient Object Detection
  – CRF Learning
  – Salient Object Features

• Evaluation

• Discussion and Conclusion
Discussion and Conclusion

• Present a supervised approach for salient object detection formulated as an image segmentation problem using a set of local, regional, and global salient object features.

• Salient object detection has wider applications.
  – Content-based image retrieval
  – Automatic collecting and labeling of image data

• Future work
  – Non-rectangular shapes of salient objects
  – Non-linear combination of features
  – More sophisticated visual features
  – Multiple salient object detection
Discussion and Conclusion

• Multiple salient object detection
Discussion and Conclusion

• Failure cases and challenges
  – Hierarchical salient object detection
Thank You!