Salient Region Detection

NINGYI DENG 邓凝旖

2014.9
Ice Breaking

KEY PERSON IN THIS PRESENTATION
An algorithm
- Log spectrum representation of image
- Five-line source code
- [07CVPR] Saliency Detection: A Spectral Residual Approach

A dataset
- PASCAL-S
- [14CVPR] The Secrets of Salient Object Segmentation

A joke
- Two reviews from CVPR reviewer for Xiaodi Hou
  - 5 - Definitely reject This paper is very interesting, and it should definitely be published, but not at NIPS.
  - 3 - References missing Some important references are missing, the followings list a few: Xiaodi Hou, Liqing Zhang: Saliency Detection: A Spectral Residual Approach. CVPR 2007
Outline

Motivation
Application
Two Different Worlds of Saliency
Perceptual Feature
Combining Two Tasks
Problems in Previous Work
Future Challenges
Future Directions
Motivation
How Men And Women Look At Advertisement Differently
How Men And Women Look At Advertisement Differently

Visual Fixation Order

[Diagram showing visual fixation order for men and women]
How Men And Women Look At Advertisement Differently

Visual Attention Pattern
Eye Tracking
Application

CONTENT-AWARE IMAGE/VIDEO RESIZING
OBJECT SEGMENTATION
Content-aware Image/video Resizing
Object Segmentation ★
Object Segmentation★
Two Different Tasks of Saliency

FIXATION PREDICTION

SALIENT OBJECT SEGMENTATION ★
Fixation Prediction

Purpose
- To compute a probabilistic map of an image to predict the actual human eye gaze patterns

Result

STATE-OF-THE-ART
- ITTI [Itti et al. PAMI 98]
- AIM [Bruce et al. NIPS 06]
- GBVS [Harel et al. NIPS 07]
- DVA [Hou et al. NIPS 08]
- SUN [Zhang et al. JOV 08]
- SIG [Hou et al. PAMI 12]
Saliency Detection: A Spectral Residual Approach

Log Spectrum Representation

- The amplitude $A(f)$ of the averaged Fourier spectrum of the ensemble of natural images obeys a distribution:

$$A(f) = \Re \left( \hat{S} [I(x)] \right), \quad \mathcal{P}(f) = \Im \left( \hat{S} [I(x)] \right),$$

$$\mathcal{L}(f) = \log (A(f)), \quad \mathcal{R}(f) = \mathcal{L}(f) - h_n(f) * \mathcal{L}(f),$$

$$\mathcal{S}(x) = g(x) * \hat{S}^{-1} \left[ \exp (\mathcal{R}(f) + \mathcal{P}(f)) \right]^2.$$
Saliency Detection: A Spectral Residual Approach

Result

Figure 7. An example of attention in different scales.
Salient Object Segmentation

**Purpose**
- To generate masks that matches the annotated silhouettes of salient objects

**Ground-truth: mask**

**Approaches**
- Low-level feature (+ segmentation)
- Popular algorithm
- Perceptual feature
Low-level feature + Segmentation

[09CVPR] Frequency-tuned Salient Region Detection

- First using saliency for outstanding region detection
- Build up first dataset with mask of objects

Groundtruth of [6]
- User drawn rectangles
- Inaccurate and clubs multiple objects into one.

Our groundtruth
- Boundary based segmentation
- Accurate and multiple objects separated.
FT- Low-level feature

Color & brightness

\[
I_\mu = \begin{bmatrix} L_\mu \\ a_\mu \\ b_\mu \end{bmatrix}
\]

\[
S(x, y) = \| I_\mu - I_{\omega hc}(x, y) \|
\]

\[
I_{\omega hc}(x, y) = \begin{bmatrix} L_{\omega hc} \\ a_{\omega hc} \\ b_{\omega hc} \end{bmatrix}
\]
FT-Saliency + Segmentation

Combine with mean-shift
FT-Result

Real image for test

Result
Popular algorithm


- Low rank matrix
- Higher-level prior integration

\[ F = \text{Low rank matrix} \ L + \text{Sparse noises} \ S \]
LR-Simplifying Low-level Matrix by Pre-Processing

\[(L^*, S^*) = \arg\min_{L,S} (\text{rank}(L) + \lambda ||S||_0) \quad \text{s.t.} \quad F = L + S\]

Unified Model \[(L^*, S^*) = \arg\min_{L,S} (||L||_* + \lambda ||S||_1) \quad \text{s.t.} \quad TFP = L + S\]

T : a learned linear metric
P : a higher-level prior map
LR-Result

Real image for test

Result
Perceptual Feature
Figure/Ground Organization in visual system

- Surroundedness
- Size
- Contrast
- Horizontal/vertical orientation
- Symmetry
- Convexity
- Parallelism
Figure/Ground Organization in visual system

\[
\text{convex}(r_i) = 1 - \frac{\text{size}(CH_i) - \text{size}(r_i)}{\text{size}(CH_i)}
\]

\[
\text{surrou nd}(r_i) = \max_{r_j \in S_i} \left( \frac{C(r_i) \cap C(r_j)}{C(r_i)} \right)
\]

\[
\text{surrou nd}(r_i) = \max_{r_j \in S_i} \left( \frac{\text{size}(CH_i) \cap \text{size}(CH_j)}{\text{size}(CH_i)} \right)
\]
Figure/Ground Organization in visual system

Meaningfulness
- High-level features
- Top-down info, i.e. Classification

![Images of figure-ground organization](image)
Combining Two Tasks

FIXATION PREDICTION + SEGMENTATION
≈ SALIENT OBJECT SEGMENTATION
FROM THE SECRETS OF SALIENT OBJECT SEGMENTATION
Combination Model
CMPC
Result
Problems in Previous Work

ALSO FROM THE SECRETS OF SALIENT OBJECT SEGMENTATION
Problems

Dataset bias
- Salient object segmentation dataset is heavily biased
- Images in FT dataset usually have a foreground object with discernable boundaries being
- Surrounded by background that have contrastive colors

Center-bias
- Easily applied by simple function, e.g. Gaussian filter
- Do great contribution to saliency (human fixations are heavily biased towards the center of the image)
Future Directions

IMAGE DETECTION
VIDEO DETECTION
COMBINATION WITH OBJECTNESS
Future Challenges

Image detection
- Segmentation
- Precision

Video detection
- Time
- Motion information
Future Directions

Combination with Objectness

- [10CVPR] *What is an object*
- [13IJCV] *Selective Search for Object Recognition*
- [14CVPR] *BING Binarized Normed Gradients for Objectness Estimation at 300fps*
Future Directions

Combination with Objectness
- [10CVPR] What is an object
- [13IJCV] Selective Search for Object Recognition
- [14CVPR] BING Binarized Normed Gradients for Objectness Estimation at 300fps
Future Directions

Combination with Objectness
- [10CVPR] What is an object
- [13IJCV] Selective Search for Object Recognition
- [14CVPR] BING Binarized Normed Gradients for Objectness Estimation
What is an object

**Image cues**

Multi-scale Saliency (MS)

\[
\text{MS}(w, \theta_{\text{MS}}^s) = \sum_{\{p \in w | I_{\text{MS}}^s(p) \geq \theta_s\}} \frac{|\{p \in w | I_{\text{MS}}^s(p) \geq \theta_s\}|}{|w|}
\]

Color Contrast (CC)

\[
\text{CC}(w, \theta_{\text{CC}}) = \chi^2(h(w), h(\text{Surr}(w, \theta_{\text{CC}})))
\]

\[
\frac{|\text{Surr}(w, \theta_{\text{CC}})|}{|w|} = \theta_{\text{CC}}^2 - 1
\]

Edge Density (ED)

\[
\text{ED}(w, \theta_{\text{ED}}) = \frac{\sum_{p \in \text{Inn}(w, \theta_{\text{ED}})} I_{\text{ED}}(p)}{\text{Len}(\text{Inn}(w, \theta_{\text{ED}}))}
\]

\[
\frac{|\text{Inn}(w, \theta_{\text{ED}})|}{|w|} = 1/\theta_{\text{ED}}^2
\]

Superpixels Straddling (SS)

\[
\text{SS}(w, \theta_{\text{SS}}) = 1 - \sum_{s \in \text{S}(\theta_{\text{SS}})} \frac{\min(|s \setminus w|, |s \cap w|)}{|w|}
\]

Saliency Detection: A Spectral Residual Approach
Selective Search for Object Recognition

**Similarity Measures**

**Colour**

\[ s_{\text{colour}}(r_i, r_j) = \sum_{k=1}^{n} \min(c_i^k, c_j^k) \]

**Texture**

\[ s_{\text{texture}}(r_i, r_j) = \sum_{k=1}^{n} \min(t_i^k, t_j^k) \]

**Size**

\[ s_{\text{size}}(r_i, r_j) = 1 - \frac{\text{size}(r_i) + \text{size}(r_j)}{\text{size}(im)} \]

\[ \text{fill}(r_i, r_j) = 1 - \frac{\text{size}(BB_{ij}) - \text{size}(r_i) - \text{size}(r_j)}{\text{size}(im)} \]
For linear model $w$:
\[ w \approx \sum_{j=1}^{N_w} \beta_j a_j \quad \text{with} \quad a_j \in \{-1, 1\}^{64} \]
\[ a_j = a_j^+ - a_j^- \quad \text{and} \quad a_j^+ \in \{0, 1\}^{64} \]
\[ \langle w, b \rangle \approx \sum_{j=1}^{N_w} \beta_j (2\langle a_j^+, b \rangle - |b|) \]

For normed gradients feature $NG$: using the top $N_g$ binary bits of the BYTE values
\[ g_i = \sum_{k=1}^{N_g} 2^{8-k} b_{k,i} \]

For SVM compute:
\[ s_l \approx \sum_{j=1}^{N_w} \beta_j \sum_{k=1}^{N_g} C_{j,k} \]

For window shift:
\[ C_{j,k} = 2^{8-k} (2\langle a_j^+, b_{k,i} \rangle - |b_{k,i}|) \]
Future Directions

Object proposal
- Superpixel feature
- Similarity measure
- Re-ranking

Combining top-down information
- Preprocess image for semantic segmentation or classification
- Get top-down info from these following processing
Thanks!

Q&A
Reference

Research from MRC: How do men and women look at advertisements differently?


