Saliency Detection

显著性检测

Zhu Yafei

CVBIOUC
http://vision.ouc.edu.cn/~zhenghaiyong

March 24, 2015

Contents

1 Background

2 Development

3 Features

4 Priors

5 Aggregation

What?

Saliency(显著)

- 明显、引入注目

What?

Saliency(显著) $\Longleftarrow$

- 明显、引入注目

What?

Saliency(显著) $\Longleftarrow$ Visual Attention

- 明显、引入注目

What?

A rich stream of visual data enters our eyes every second.[1]

[1]K. Koch *et al.*, “How much the eye tells the brain”, Current Biology, 2006.

What?

A rich stream of visual data enters our eyes every second.[1]

$$10^8 \sim 10^9 bits$$

[1] K. Koch *et al.*, "How much the eye tells the brain", Current Biology, 2006.

What?

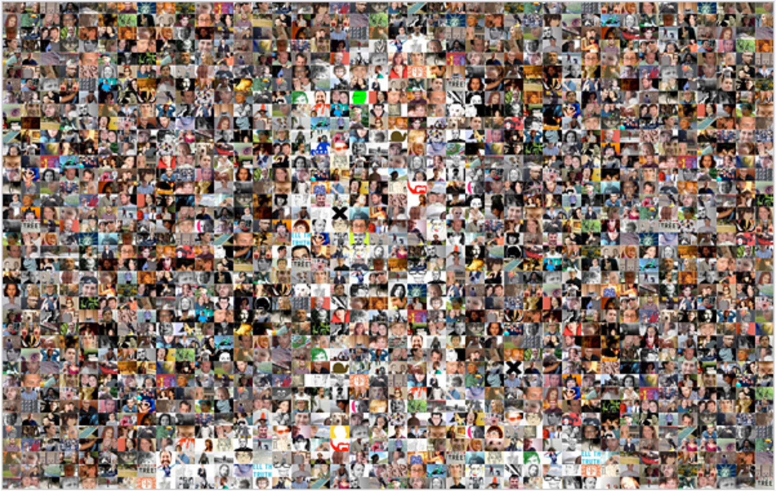
分辨率:

500*375

What?

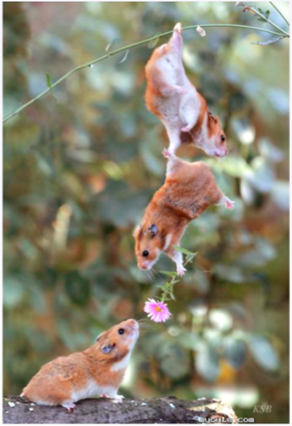
$$500 \times 375 \times 3 \times 8 = 4500000\,bits$$

What?



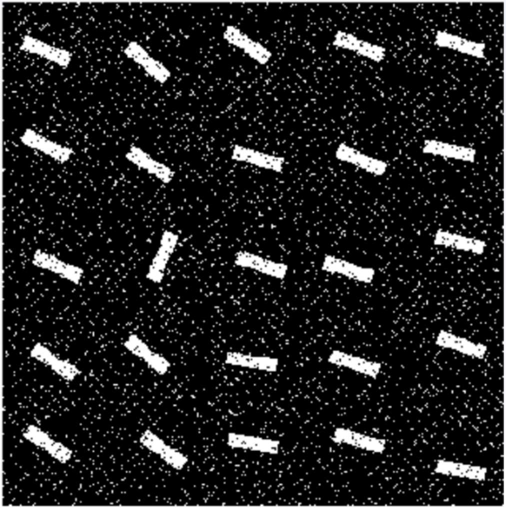
Vision as data reduction

- Raw feed from camera/eyes:
  - 10^{7-9} Bytes/s
- Extraction of edges and salient features
  - 10^{3-4} Bytes/s
- High-level interpretation of scene
  - 10^{1-2} Bytes/s



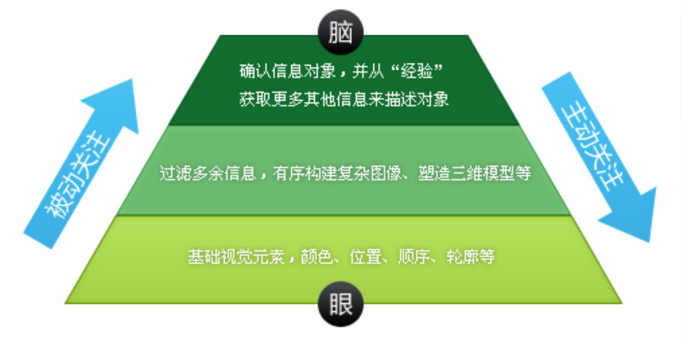
Visual Attention

Visual Attention

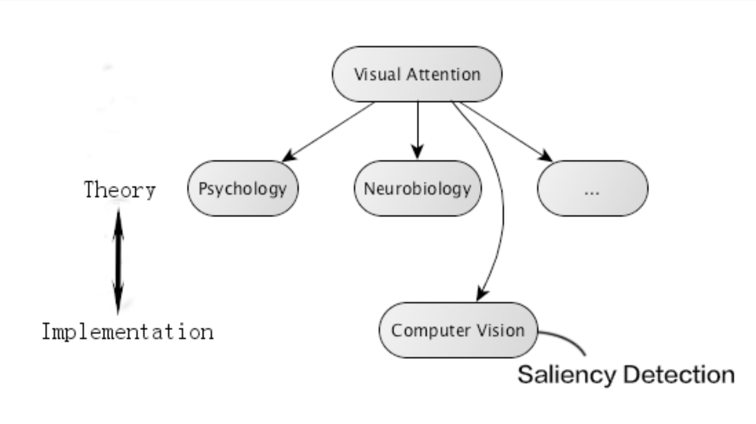


Visual Attention

Bottom-up vs Top-down



Visual Attention



Theory

- "*Feature Integration Theory*", Treisman & Gelade, 1980.
- "*Guided Search Theory*", Wolfe & Cave, 1989.
- "*Integrated Competition Theory*", Desimone & Duncan, 1995.

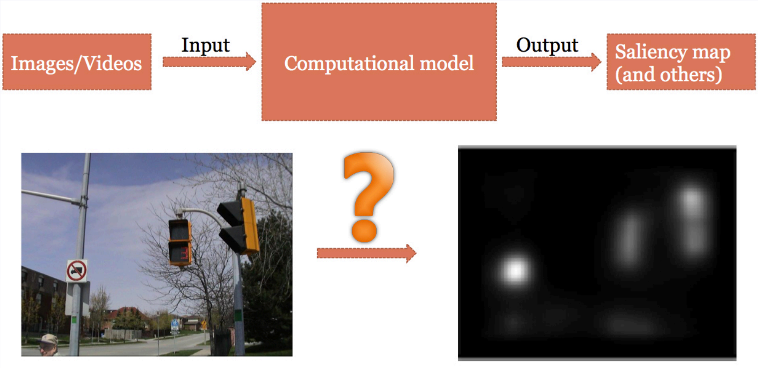
Framework

- Koch and Ullman[2] proposed a feed-forward model to combine features and introduced the concept of a *saliency map* which is a topographic map that represents conspicuousness of scene locations.
- Zoom Lens Model[3]

[2] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry", Human Neurobiology, 1985.

[3] Eriksen, C. W. and J. D. St. James, "Visual attention within and around the field of focal attention: a zoom lens model", Perception and Psychophysics, 1986.

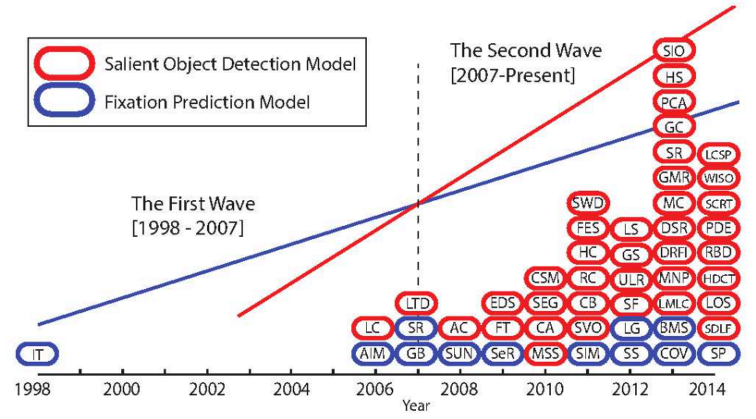
Basic Structure of Computational Models



Two Waves[4]

Saliency Detection

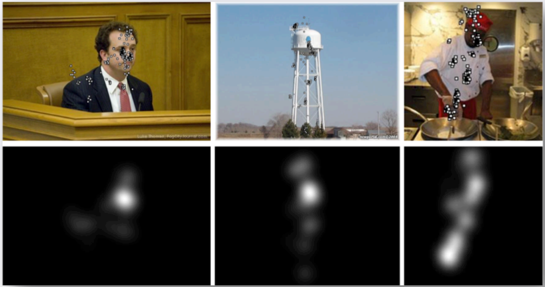
- Background
- **Development**
- Features
- Priors
- Aggregation



[4] Ali Borji *et al.*, "Salient object detection: a survey", 2014.

Fixation Prediction

Fixation prediction models are constructed originally to understand human visual attention and eye movement prediction.



Fixation Prediction: example

Fixation Prediction: example

Fixation Prediction： example

Fixation Prediction: example

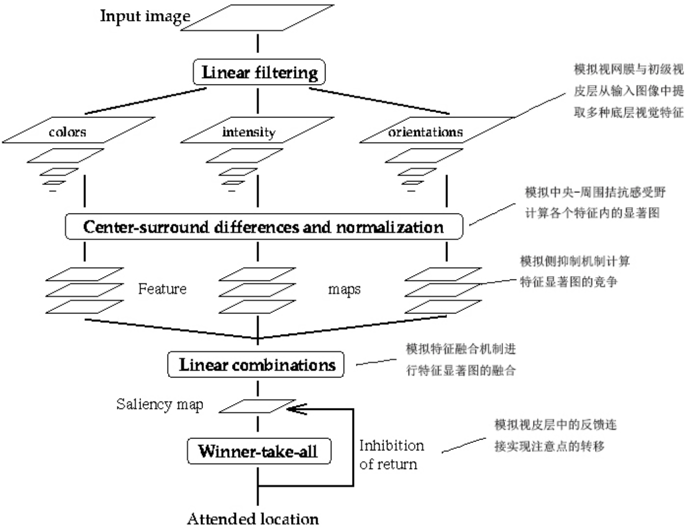
Fixation Prediction: example

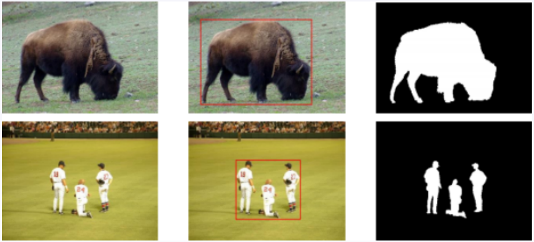
Fixation Prediction： example

Representative Work

- A model of saliency-based visual attention for rapid scene analysis. PAMI 1998, Itti *et al.*.

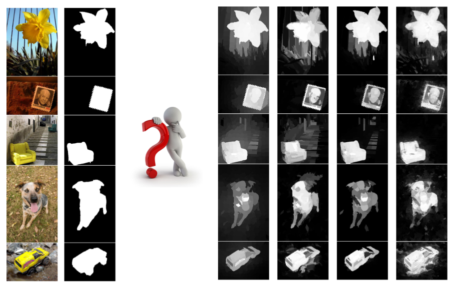
Architecture





- Learning to detect a salient object. CVPR 2007, Tie Liu *et al.*.
- Frequency-tuned salient region detection. CVPR 2009, Achanta *et al.*.

State-of-the-art



From left to right: PBS[5], DRFI[6], RBD[7], HDCT[8]

[5]Chuan Yang *et al.*, "Graph-regularized saliency detection with convex-hull-based center prior, Signal Processing Letters, 2013.

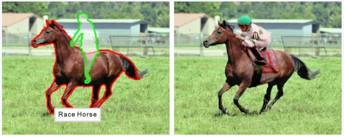
[6]Huaizu Jiang *et al.*, "Salient object detection: a discriminative regional feature integration approach", in CVPR, 2013.

[7]Wangjiang Zhu *et al.*, "Saliency optimization from robust background detection", in CVPR, 2014.

[8]Jiwhan Kim *et al.*, "Salient region detection via high-dimensional color transform", in CVPR, 2014.

Why?

The emergence of salient object detection models is driven by the requirement of saliency-based applications.



Content aware resizing

Object maniputation

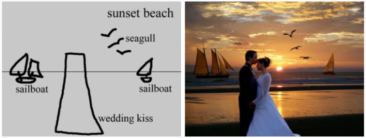
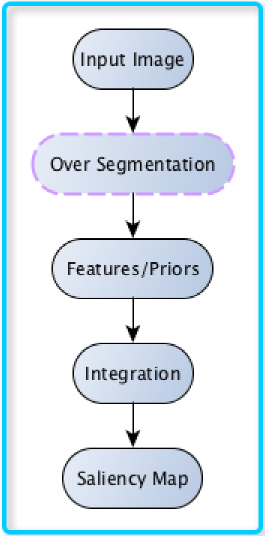


Image montage

Image collage

What can saliency not do?

General Procedure



Features

Salient object detection: A discriminative regional feature integration approach, CVPR 2013

- 26 维颜色纹理等对比度特征

Color and texture features			Differences of features		Contrast	Backgroundness
	features	dim	definition	dim		
$\mathbf{a}_1$	the average RGB values	3	$d(\mathbf{a}_1^R, \mathbf{a}_1^S)$	3	$c_1 \sim c_3$	$b_1 \sim b_3$
$\mathbf{a}_2$	the average L*a*b* values	3	$d(\mathbf{a}_2^R, \mathbf{a}_2^S)$	3	$c_4 \sim c_6$	$b_4 \sim b_6$
$\mathbf{r}$	the absolute response of LM filters	15	$d(\mathbf{r}^R, \mathbf{r}^S)$	15	$c_7 \sim c_{21}$	$b_7 \sim b_{21}$
r	the max response among the LM filters	1	$d(r^R, r^S)$	1	c_{22}	b_{22}
$\mathbf{h}_1$	the L*a*b* histogram	$8 \times 16 \times 16$	$\chi^2(\mathbf{h}_1^R, \mathbf{h}_1^S)$	1	c_{23}	b_{23}
$\mathbf{h}_2$	the hue histogram	8	$\chi^2(\mathbf{h}_2^R, \mathbf{h}_2^S)$	1	c_{24}	b_{24}
$\mathbf{h}_3$	the saturation histogram	8	$\chi^2(\mathbf{h}_3^R, \mathbf{h}_3^S)$	1	c_{25}	b_{25}
$\mathbf{h}_4$	the texton histogram	65	$\chi^2(\mathbf{h}_4^R, \mathbf{h}_4^S)$	1	c_{26}	b_{26}

- 34 维的区域特征

description	notation	dim	description	notation	dim
the average normalized x coordinates	p_1	1	the average normalized y coordinates	p_2	1
the normalized perimeter	p_7	1	the 10th percentile of the normalized x coordinates	p_3	1
the aspect ratio of the bounding box	p_8	1	the 10th percentile of the normalized y coordinates	p_4	1
the variances of the RGB values	$p_9 \sim p_{11}$	3	the 90th percentile of the normalized x coordinates	p_5	1
the variances of the L*a*b* values	$p_{12} \sim p_{14}$	3	the 90th percentile of the normalized y coordinates	p_6	1
the variances of the HSV values	$p_{15} \sim p_{17}$	3	the variance of the response of the LM filters	$p_{18} \sim p_{32}$	15
the normalized area	p_{33}	1	the normalized area of the neighbor regions	p_{34}	1

Features

- Pixel-based
  - Contrast: $U_p(x) = \sum_{x' \in I \setminus \{x\}} D(x', x)$
  - Property: Geometric features
- Patch/Region/Superpixel-based
  - Contrast: $U_r(A_i) = \sum_{1 \leqslant k \leqslant N, k \neq i} |\Lambda| D(\Lambda_k, \Lambda_i)$
  - Property: Geometric features/Appearance features
- Multi-scale based

① Contrast

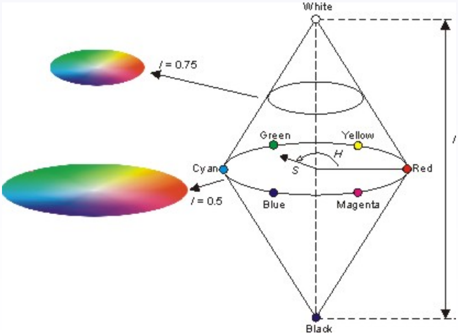
- Color
- Texture

Color

- *Color* is the visual perceptual property corresponding in humans to the categories called red, blue, yellow and others.
- *Color Space* is defined to identify colors numerically by their coordinates.

Three Elements of Color

- Hue: What we think as "color"–yellow, orange, cyan and magenta are examples of different hues
- Saturation: Saturation refers to the relative purity or the amount of white light mixed with a hue
- Intensity



Coding methods for humans

- **RGB** is an additive system (add colors to black) used for displays.
- **CMY** is a subtractive system for printing.
- **HSI** is a good perceptual space for art, psychology, and recognition.
- **YIQ** is used for TV and is good for compression.

Color Space

- RGB
- Lab
- HSV

Color Histogram[9] (**Demo**)

- Quantize each color channel to have 16 different values
- Compute matrix Q_rgb
- Superpixel segmentation
- Compute color histogram of each superpixel
- Compute color histogram contrast feature of each super pixel

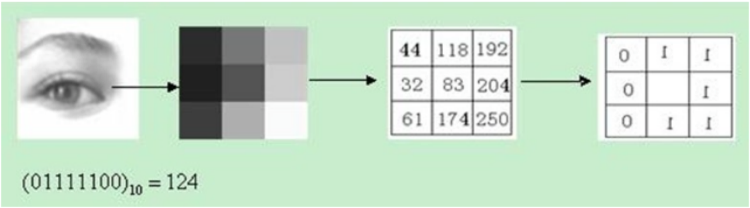
[9] Huaizu Jiang *et al.*, "Salient object detection: a discriminative regional feature integration approach", in CVPR, 2013.

What is texture?

Texture is actually a very nebulous concept, often attributed to human perception, as either the feel or the appearance of (woven) fabric[10].

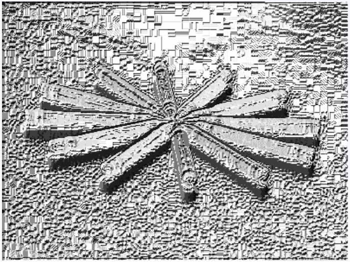
[10] Mark S.Nixon and Alberto S.Aguado, “Feature Extraction & Image Processing for Computer Vision”, Publishing House of Electronics Industry, 2013.

LBP (Local Binary Pattern)



Saliency Detection

Background
Development
Features
Priors
Aggregation



② Regional property descriptor

- Geometric feature descriptor
  - size
  - perimeter
  - position
  - area of the neighbor regions
- Appearance feature descriptor
  - the variances of the RGB values
  - the variances of the response of the LM filters

Priors

1. Center Prior/Location Prior
2. Backgroundness Prior
3. Boundary Connectivity Prior
4. Color Prior
5. Objectness Prior
6. Smoothness Prior

① Center Prior/Location Prior (**Demo**)

Objects near the image center are more attractive to people[11].

This prior can be simply and effectively modeled as a Gaussian map.

$$S_D(\mathbf{x}) = exp\left(-\frac{||\mathbf{x}-\mathbf{c}||_2^2}{\sigma_D^2}\right) \tag{1}$$

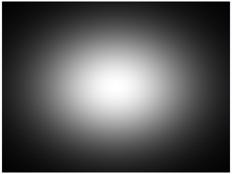
[11] Lin Zhang *et al.*, “SDSP: A novel saliency detection method by combining simple priors”, in ICIP, 2013.

① Center Prior/Location Prior

The salient object in an image is most probably placed near the center of the image[12].

Gaussian falloff weight:

$$w_i^{(n)} = exp\left(-9\frac{(dx_i^{(n)})^2}{w^2} - 9\frac{(dy_i^{(n)})^2}{h^2}\right) \tag{2}$$



[12]Huaizu Jiang *et al.*, "Automatic salient object segmentation based on context and shape prior", in BMVC, 2011

① Center Prior/Location Prior

Assigning higher saliency to the image elements near the image center becomes invalid when the objects are placed far off the image center[13].

- Compute a convex hull enclosing interesting points to estimate the location of salient region.
- Use the centroid of the convex hull as the center to get the convex-hull-based center prior map.

[13] Chuan Yang *et al.*, "Graph-regularized saliency detection with convex-hull-based center prior", Signal Processing Letters, 2013.

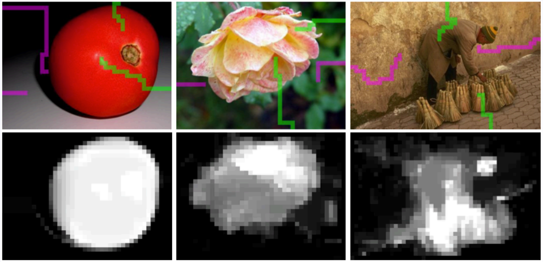
② Backgroundness Prior

Backgroundness prior is more general than center prior because salient objects can be placed off the center, but they seldom touch the image boundary[14].

- Assuming that a narrow border of the image is background region, regional saliency can be computed as the contrast versus "background".

[14] Yichen Wei *et al.*, "Geodesic saliency using background priors", in ECCV, 2012.

② Backgroundness Prior (Demo)

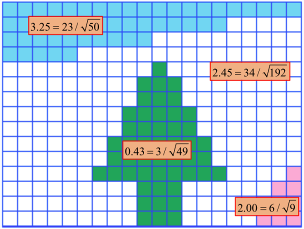


$$saliency(P) = \min_{P_1=P, P_2, \ldots, P_n=B} \sum_{i=1}^{n-1} weight(P_i, P_{i+1}), \qquad (3)$$

$$s.t. (P_i, P_{i+1}) \in \varepsilon$$

③ Boundary Connectivity Prior

Object regions are much less connected to image boundaries than background ones[15].



Boundary connectivity is defined to quantify how heavily a region R is connected to the image boundaries.

$$BndCon(R) = \frac{|\{p|p \in R, p \in Bnd\}|}{\sqrt{|\{p|p \in R\}|}} \tag{4}$$

[15] Wangjiang Zhu *et al.*, "Saliency optimization from robust background detection", in CVPR, 2014.

④ Color Prior (**Demo**)

Saliency Detection

Background

Development

Features

Priors

Aggregation

Warm colors, such as red and yellow, are more pronounced to the human visual system than cold colors, such as green and blue[16].

$$f_{an}(\mathbf{x}) = \frac{f_a(x) - min(a)}{max(a) - min(a)}, f_{bn}(\mathbf{x}) = \frac{f_b(x) - min(b)}{max(b) - min(b)} \tag{5}$$

$$S_c(\mathbf{x}) = 1 - exp\left(-\frac{f_{an}^2(\mathbf{x}) + f_{bn}^2(\mathbf{x})}{\sigma_c^2}\right) \tag{6}$$

- a^*-channel represents green-red information
- b^*-channel represents blue-yellow information

[16] Lin Zhang *et al.*, "SDSP: A novel saliency detection method by combining simple priors", in ICIP, 2013.

⑤ Objectness Prior

Objectness is defined as the probability of there being a complete object in a local window centered on each pixel[17].

- Randomly sample N windows over the image
- Assign each window w a probability score $P(w)$ to indicate its objectness
- Sum all the probability scores in windows that contains pixel x

$$O_p(x) = \sum_{w \in W \text{ and } x \in w} P(W_x) \tag{7}$$

[17] Peng Jiang *et al.*, "Salient region detection by ufo: uniqueness, focusness and objectness", in ICCV, 2013.

⑥ Smoothness Prior

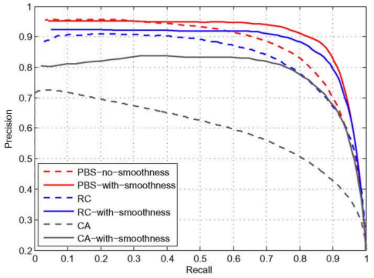
Smoothness constraint is often encoded by adding a pair-wise potential to the energy function which encourages neighboring pixels in the image to take the same label[18].

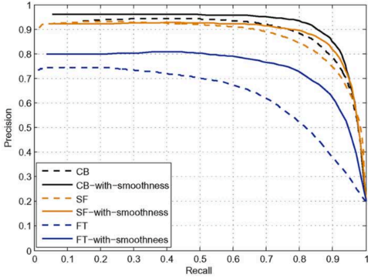
$$w_{ij} = exp\left(-\frac{||c_i - c_j||}{2\sigma_w^2}\right) \tag{8}$$

$$E(S) = \sum_i (S(i) - S_{in}(i))^2 + \lambda \sum_{i,j} w_{ij}(S(i) - S(j))^2 \tag{9}$$

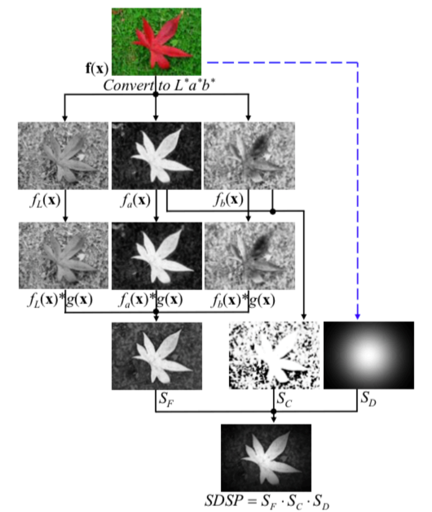
[18] Chuan Yang *et al.*, "Graph-regularized saliency detection with convex-hull-based center prior", Signal Processing Letters, 2013.

⑥ Smoothness Prior





Example: SDSP[19] (**Demo**)



[19] Lin Zhang *et al.*, "SDSP: A novel saliency detection method by combining simple priors", in ICIP, 2013.

① Non-linear[20]

Combining the focusness, the objectness and the uniqueness maps:

$$S = exp(F + U) \times O \tag{10}$$

[20] Peng Jiang *et al.*, "Salient region detection by ufo: uniqueness, focusness and objectness", in ICCV, 2013.

② Energy minization[21]

The ideal output of salient object detection is a clean binary object/background segmentation.

$$\underbrace{\sum_{i=1}^{N} w_i^{bg} s_i^2}_{background} + \underbrace{\sum_{i=1}^{N} w_i^{fg} (s_i - 1)^2}_{foreground} + \underbrace{\sum_{i,j} w_{ij} (s_i - s_j)^2}_{smoothness} \tag{11}$$

[21] Wangjiang Zhu *et al.*, "Saliency optimization from robust background detection", in CVPR, 2014.

③ Machine Learning[22]

Saliency Detection

Background

Development

Features

Priors

Aggregation

Training examples:

- a set of confident regions $R = \{R_1, R_2, \ldots, R_Q\}$
- the responding saliency scores $A = \{a_1, a_2, \ldots, a_Q\}$

[22] Huaizu Jiang *et al.*, "Salient object detection: a discriminative regional feature integration approach", in CVPR, 2013.

Thanks

Yafei Zhu
Ocean University of China
2015.03