Saliency Detection

Zhu Yafei

CVBIOUC
Ocean University of China
http://vision.ouc.edu.cn/~zhenghaiyong

March 20, 2015

Contents

What?

A rich stream of visual data ($10^8 - 10^9$ bits) enters our eyes every second.[1]

[1] K. Koch and J. McLean, "How much the eye tells the brain", Current Biology, 2006.

What?

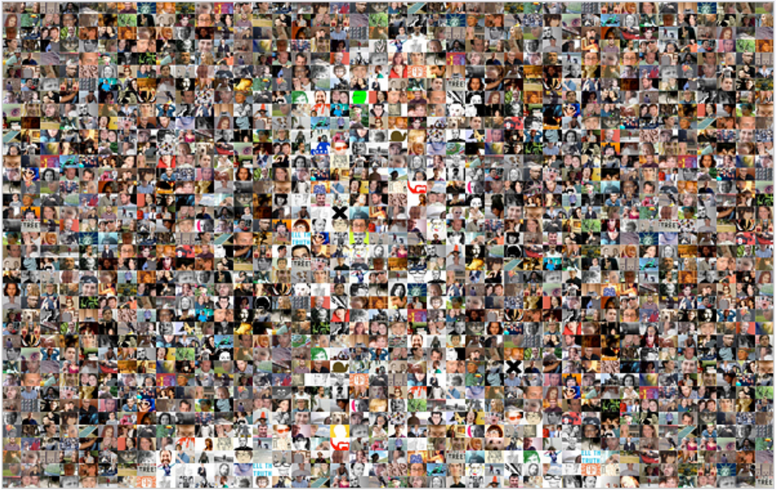
分辨率:

500*375

What?

$$500 \times 375 \times 3 \times 8 = 4500000bit$$

What?

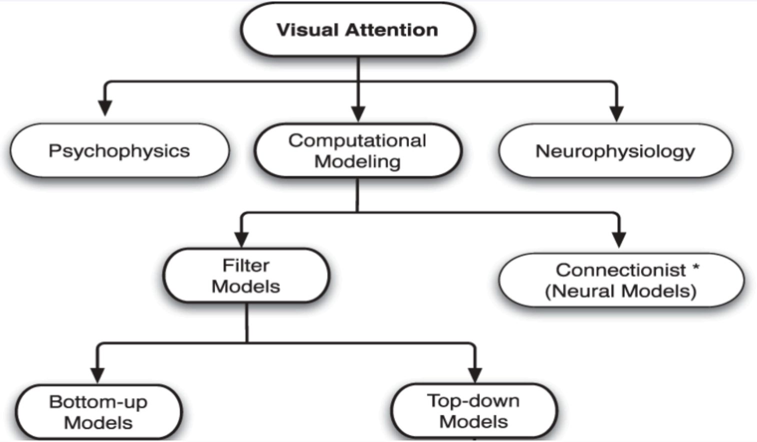


What?

Bottom-Up

Top-Down

Visual Attention



Origins

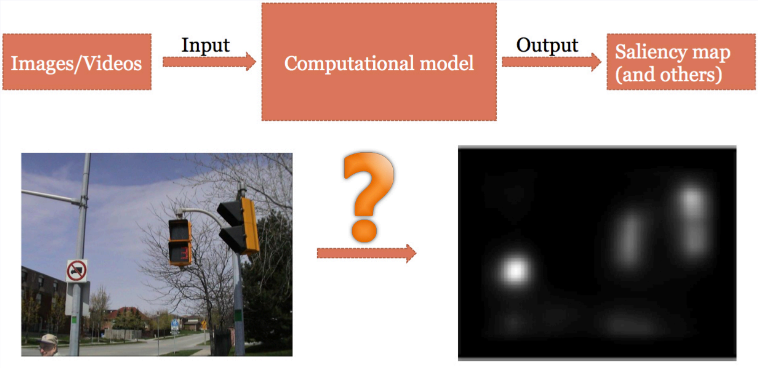
- “Feature Integration Theory”, Treisman & Gelade, 1980.
- “Guided Search Theory”, Wolfe & Cave, 1989.
- “Integrated Competition Theory”, Desimone & Duncan, 1995.

Origins

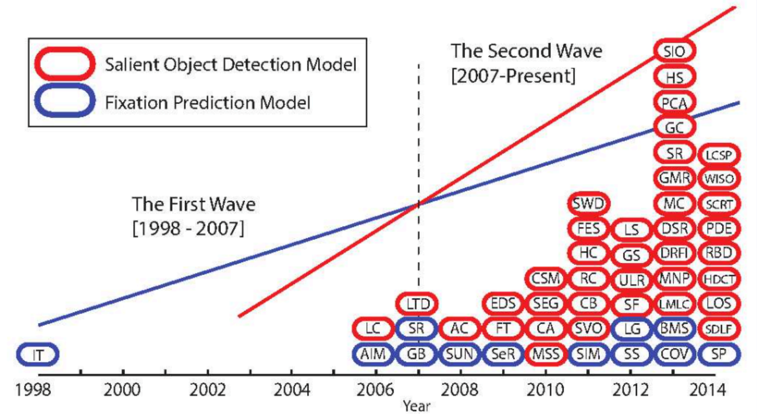
Koch and Ullman[1] proposed a feed-forward model to combine features and introduced the concept of a saliency map which is a topographic map that represents conspicuousness of scene locations.

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

Basic Structure of Computational Models

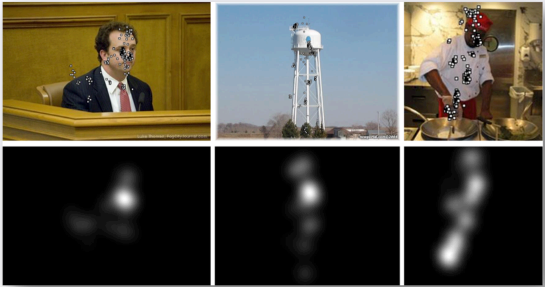


Two Waves


Salient Object Detection Model
Fixation Prediction Model
The Second Wave
[2007-Present]
The First Wave
[1998 - 2007]
SIO
HS
PCA
GC
SR
LCSP
GMR
WISO
SWD
MC
SCRT
FES
LS
DSR
PDE
HC
GS
DRFI
RBD
CSM
RC
ULR
MNP
HDCT
LTD
EDS
SEG
CB
SF
LMLC
LOS
LC
SR
AC
FT
CA
SVO
LG
BMS
SDLF
IT
AIM
GB
SUN
SeR
MSS
SIM
SS
COV
SP
1998
2000
2002
2004
2006
2008
2010
2012
2014
Year

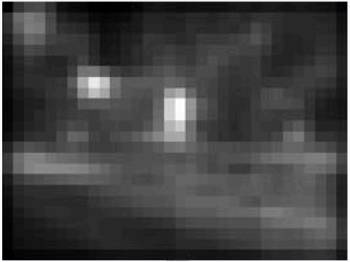
Fixation Prediction

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

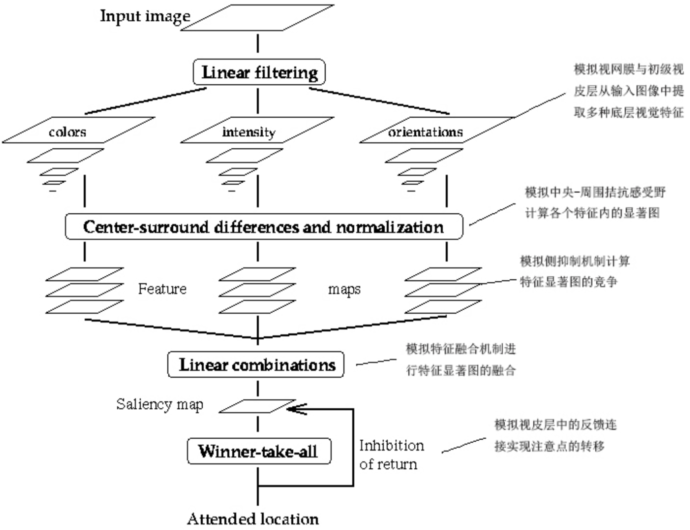


Representative Method

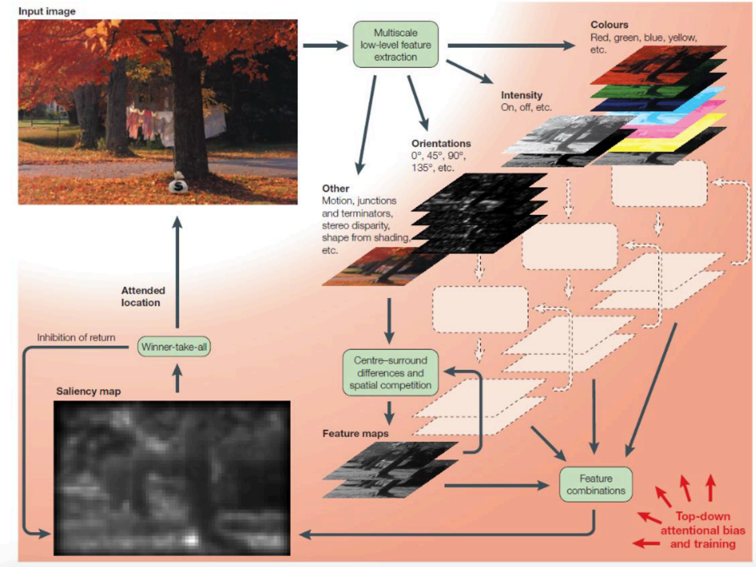
- A model of saliency-based visual attention for rapid scene analysis. PAMI 1998, Itti et al.



Architecture



General Procedure of fixation prediction models



How to evaluate fixation prediction models?

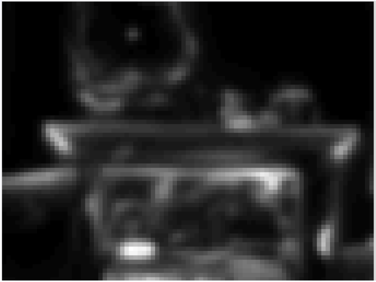
Saliency Detection

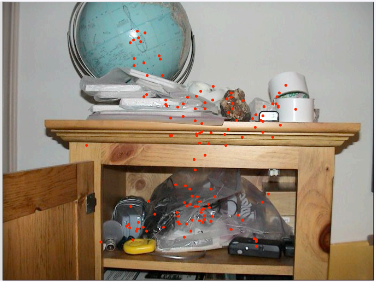
Introduction

Features

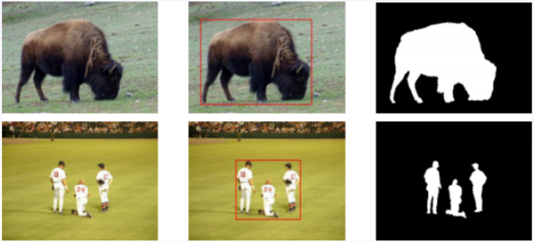
Priors

Aggregation





Salient Object Detection

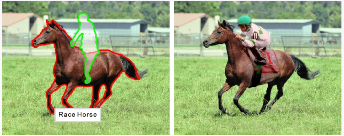


- Learning to detect a salient object. CVPR 2007, Tie Liu et al.
- Frequency-tuned salient region detection. CVPR 2009, Achanta et al.

Why?

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

Content aware resizing



Object maniputation

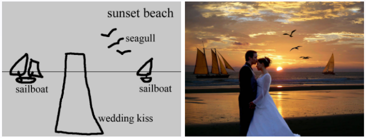
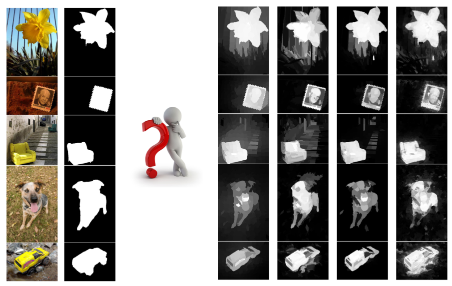


Image montage

Image collage

What can saliency not do?

Latest research achievements



From left to right: PBS[2]、DRFI[3]、RBD[4]、HDCT[5]

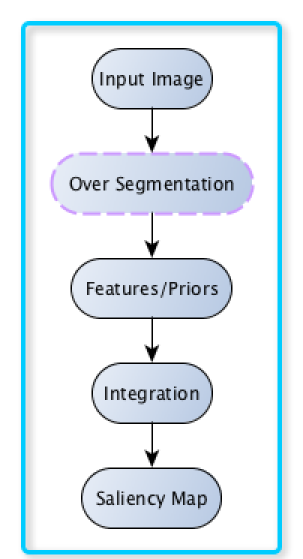
[2] Chuan Yang et al., "Graph-regularized saliency detection with convex-hull-based center prior, IEEE Signal Processing Letters, 2013.

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

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

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

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、Frequency
- 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

Contrast

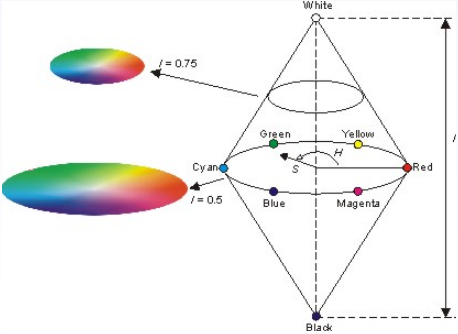
- Color
- Texture

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

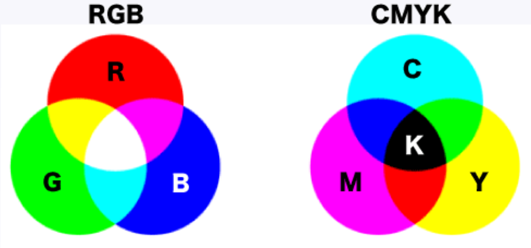


Coding methods for humans

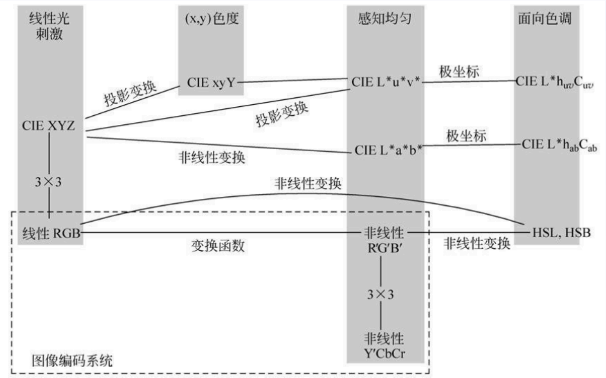
- **RGB** is an additive system (add colors to black) used for displays.
- **CMY** is a subtractive system for printing.
- **HSI** is a good perceptual space for art, psychology, and recognition.
- **YIQ** is used for TV and is good for compression.

Device-dependent spaces

- RGB、CMYK



Almost all the color spaces can be transformed from RGB color space.



Color Histogram

- 对图像的颜色空间进行量化
- 计算矩阵 Q
- 超像素分割
- 对每个超像素区域，依据其中像素点的坐标值，求出对应的矩阵 Q 中的值
- 求区域之间的直方图的对比度

What is texture?

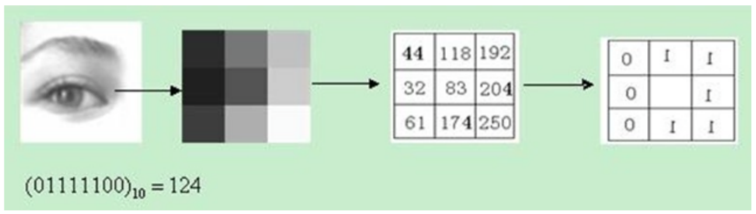
Texture is actually a very nebulous concept, often attributed to human perception, as either the feel or the appearance of (woven) fabric.[6]

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

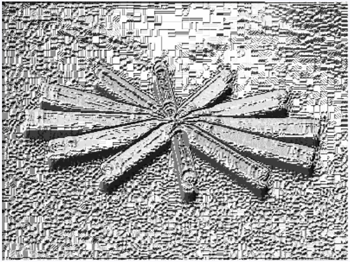
Texture Descriptor

- Structural approaches
- Statistical approaches
- Combination approaches
- Local binary patterns

LBP(Local Binary Pattern，局部二值模式)



LBP



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

- Center Prior/Location Prior
- Backgroundness Prior
- Boundary Connectivity Prior
- Color Prior
- Objectness Prior
- Smoothness Prior

Center Prior/Location Prior

Objects near the image center are more attractive to people.[7]

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

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

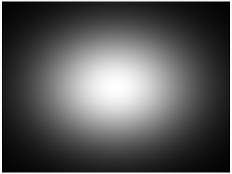
[7] Zhang, Lin and Gu, Zhongyi and Li, Hongyu, "SDSP: A novel saliency detection method by combining simple priors", in ICIP, 2013.

Center Prior/Location Prior

The salient object in an image is most probably placed near the center of the image.[8]

Gaussian falloff weight:

$$w_i^{(n)} = exp(-9(dx_i^{(n)})^2/w^2 - 9(dy_i^{(n)})^2/h^2 \qquad (2)$$



[8] Jiang, Huaizu and Wang, Jingdong, "Automatic salient object segmentation based on context and shape prior", in BMVC, 2011

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[9].

- 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.

[9] Yang, Chuan and Zhang, Lihe and Lu, Huchuan, "Graph-regularized saliency detection with convex-hull-based center prior", in Signal Processing Letters, IEEE, 2013

Backgroundness Prior

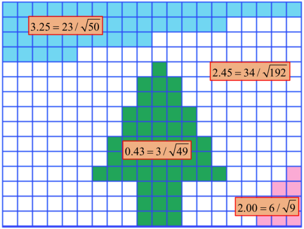
Backgroundness prior is more general than center prior because salient objects can be placed off the center, but they seldom touch the image boundary.[10]

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

[10] Wei, Yichen and Wen, Fang and Zhu, Wangjiang, "Geodesic saliency using background priors", in ECCV, 2012.

Boundary Connectivity Prior

Object regions are much less connected to image boundaries than background ones.[11]



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

$$BndCon(R) = \frac{|\{p|p \in R, p \in Bnd\}|}{\sqrt{|\{p|p \in R\}|}} \tag{3}$$

[11] Zhu, Wangjiang and Liang, Shuang, "Saliency optimization from robust background detection", in CVPR, 2014.

Color Prior

Warm colors, such as red and yellow, are more pronounced to the human visual system than cold colors, such as green and blue.[12]

$$f_{an}(\mathbf{x}) = \frac{f_a(x) - mina}{maxa - mina}, f_{bn}(\mathbf{x}) = \frac{f_b(x) - minb}{maxb - minb} \tag{4}$$

$$S_c(\mathbf{x}) = 1 - exp\left(-\frac{f_{an}^2(\mathbf{x}) + f_{bn}^2(\mathbf{x})}{\sigma_c^2}\right) \tag{5}$$

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

[12] Zhang, Lin and Gu, Zhongyi and Li, Hongyu, "SDSP: A novel saliency detection method by combining simple priors", in ICIP, 2013.

Objectness Prior

Objectness is defined as the probability of there being a complete object in a local window centered on each pixel.[13]

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

$$O_p(x) = \sum_{w \in W \text{ and } x \in w} P(W_x) \tag{6}$$

[13] Jiang, Peng and Ling, Haibin, "Salient region detection by ufo: uniqueness, focusness and objectness", in ICCV, 2013

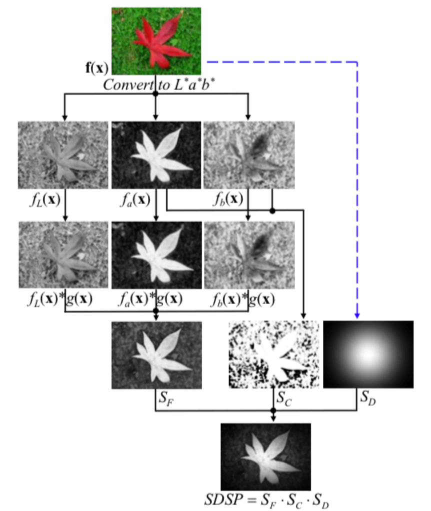
Smoothness constraint is often encoded by adding a pair-wise potential to the energy function which encourages neighboring pixels in the image to take the same label.[14]

$$w_{ij} = exp\left(-\frac{||c_i - c_j||}{2\sigma_w^2}\right) \tag{7}$$

$$E(S) = \sum_i (S(i) - S_in(i))^2 + \lambda \sum_{i,j} w_{ij}(S(i) - S(j))^2 \tag{8}$$

[14] Yang, Chuan and Zhang, Lihe and Lu, Huchuan, "Graph-regularized saliency detection with convex-hull-based center prior", in Signal Processing Letters, IEEE, 2013.

Example: SDSP[15]



[15] Zhang, Lin and Gu, Zhongyi and Li, Hongyu, "SDSP: A novel saliency detection method by combining simple priors", in ICIP, 2013.

Non-linear[16]

Combining the focusness, the objectness and the uniqueness maps:

$$S = exp(F + U) \times O \tag{9}$$

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

Energy minization[17]

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

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

[17] Wangjiang Zhu and Shuang Liang, “Saliency Optimization from Robust Background Detection ” , in CVPR, 2014.

Machine Learning[18]

Training examples:

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

[18] 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