A Computational Model for Saliency Detection based on Probability Distributions

universitätbonn

Dominik A. Klein, Germán M. García and Simone Frintrop
Intelligent Vision Systems Group, Dept. Computer Science III, University of Bonn

IVS intelligent vision systems

Motivation

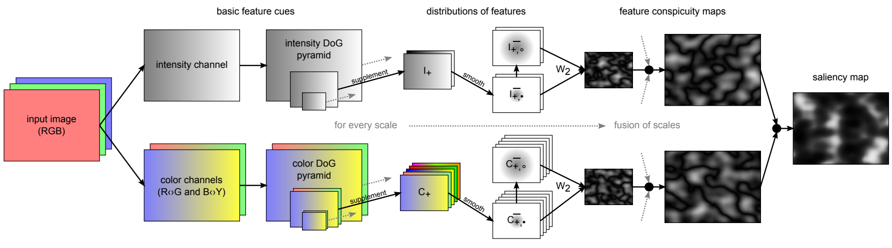
Task: detect salient objects

We present a computational attention system that quickly detects salient objects in web images. It can deal with real-world images and different object sizes. Details in [1].

Approach:
- Use standard structure of visual attention systems [2]: separate feature channels, different scales, center-surround
- Represent feature statistics by multivariate normal distributions
- Compare distributions with the Wasserstein metric based on Euclidean norm.

Obtains good results on psychophysical data and outperforms 9 state-of-the-art saliency systems on the MSRA benchmark [3,4].

CoDi-Saliency Model



CoDi-Saliency: Continuous Distribution Saliency
- use two basic feature channels: *intensity and color (red-green + blue-yellow opponents)*
- compute basic features on 12 scales
- compute center-surround contrast of basic feature distributions based on W_2
- first fuse information from scales, then combine channels into a single saliency map

Method

Compute basic feature cues for intensity: and color:

$$I(x,y) = \left(\frac{R+G+B}{3}\right)_{(x,y)} \qquad C(x,y) = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} R-G \\ B-\frac{R+G}{2} \end{pmatrix}_{(x,y)}$$

Collect feature statistics by multivariate normal distributions represented by ML-estimates:

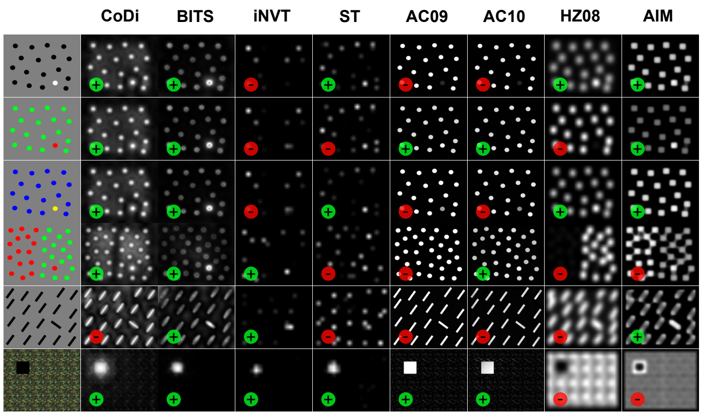
$$\hat{\mu}_{\mathrm{I}} = \bar{i} \qquad \hat{\mu}_{\mathrm{C}} = \begin{pmatrix} \bar{c}_1 \\ \bar{c}_2 \end{pmatrix}$$

$$\hat{\sigma}_{\mathrm{I}} = \overline{i^2} - \bar{i}^2 \qquad \hat{\Sigma}_{\mathrm{C}} = \begin{pmatrix} \overline{c_1^2} - \bar{c}_1^2 & \overline{c_1 c_2} - \bar{c}_1\bar{c}_2 \\ \overline{c_1 c_2} - \bar{c}_1\bar{c}_2 & \overline{c_2^2} - \bar{c}_2^2 \end{pmatrix}$$

Sophisticated optimizations enable efficient computing. See [1] for details. Saliency is computed as difference between center and surround distributions by the Wasserstein distance W_2 based on the Euclidean norm:

$$W_2(P_\bullet, P_\odot) = \sqrt{\|\mu_\bullet - \mu_\odot\|_2^2 + tr(\Sigma_\bullet) + tr(\Sigma_\odot) - 2tr\left(\sqrt{\sqrt{\Sigma_\bullet}\Sigma_\odot\sqrt{\Sigma_\bullet}}\right)}$$

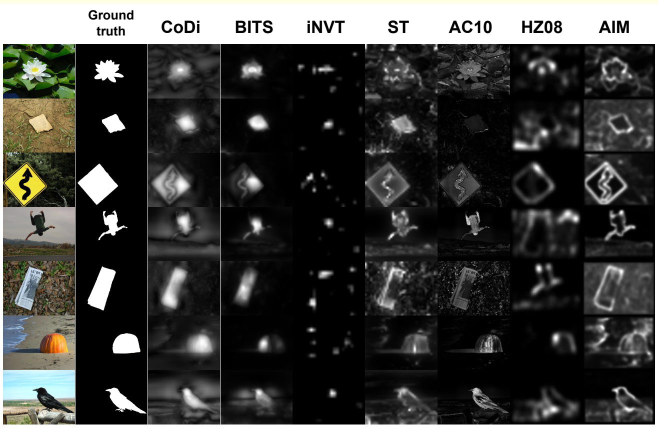
Results on Psychophysical Data



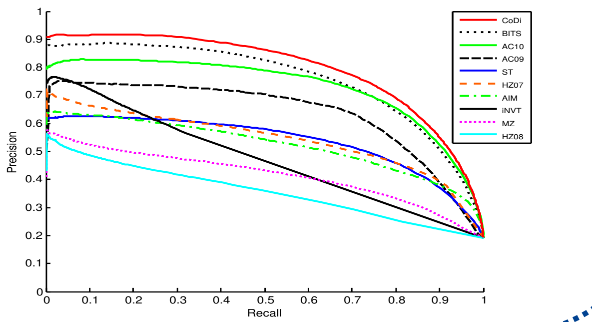
Outlier detected with 1st FOA — Outlier not detected with 1st FOA

BITS (Bonn Information-Theoretic Saliency): Klein/Frintrop
iNVT (Neuromorphic Vision Toolkit): Itti et al.
ST (Saliency Toolbox): Walther et al.
AC09/AC10: Achanta et al.
HZ08: Hou/Zhang
AIM (Attention based on Information Maximation): Bruce/Tsotsos

Results on MSRA Dataset [3]



We use the Achanta benchmark [3] which contains 1000 web images with salient objects and the corresponding ground truth (user-drawn binary maps). It is a subset of the MSRA salient object database [2].



References:
[1] Klein, Frintrop: Salient Pattern Detection using W_2 on Multivariate Normal Distributions, DAGM 2012
[2] Frintrop: Computational Visual Attention, in Computer Analysis of Human Behavior, Springer, 2011
[3] Liu et al.: Learning to detect a salient object, Trans. on PAMI, 2009
[4] Achanta et al.: Frequency-tuned salient region detection, CVPR 2009