Application of Conditional Random Fields in Salient Object Detection within an Image, using Local, Regional, and Global Features

Jimmy Lin
Chris Claoue Long

Dr. Stephen Gould

College of Engineering and Computer Science
Australian National University

May 21, 2013

Introduction and Motivation

Images from MSRA dataset B

Saliency is the prominence of an object in an image.

Often detected by its high contrast to its boundary with the background, its unique colour distribution compared to its surrounds, and the break in spatial continuity of colour that it represents in the image.

Salient object detection is useful in numerous areas!

Related Works

Formulation

Given an image I, we want to compute the location of a salient object.

Binary labelling task – for each pixel x, indicate whether it belongs to the salient object (1) or not (0)

Build up a probabilistic model $P(A|I) = \frac{1}{Z}e^{-E(A|I)}$, where $\frac{1}{Z}$ is the normalising factor, and $E(A|I)$ is the energy function incorporating both unary/static and pairwise potentials between pixels.

Energy function in more detail

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

where λ_k is the weight of the k^th feature and x, x' represent two adjacent, or neighbouring, pixels

Energy function in more detail

The unary/static feature $F_k(a_x, I)$ is a normalised feature map $f_k(x, I) \in [0, 1]$ for each pixel:

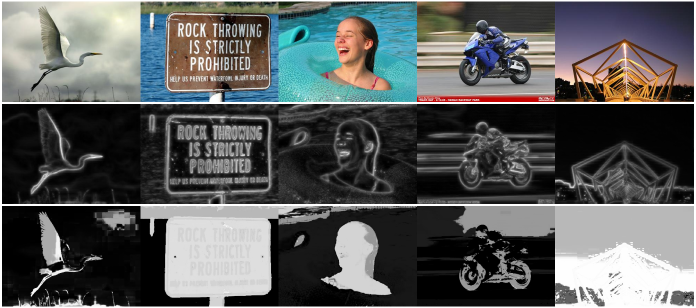
$$F_k(a_x, I) = \begin{cases} f_k(x, I), & a_x = 0 \\ 1 - f_k(x, I), & a_x = 1 \end{cases}$$

The pairwise feature $S(a_x, a_{x'}, I)$ exploits the spatial relationship between two adjacent pixels. It can be viewed as a "penalty" for labelling adjacent pixels the same or differently.

$$S(a_x, a_{x'}, I) = |a_x - a_{x'}| \cdot e^{-\beta d_{x,x'}}$$

where d_x, x' is the L2-norm (standard norm) representing the colour difference between the two pixels, and $\beta = (2\langle||I_x - I_{x'}||^2\rangle)^{-1}$ is a robust parameter to weight the colour contrast.

Feature Extraction



Local: Multiscale Contrast

Create a contrast map from the linear combination of image contrast at all levels of an N-level gaussian image pyramid, using the pixels x in the image I:

$$f_c(x, I) = \sum_{n=1}^{N} \sum_{x' \in W(x)} ||I^n(x) - I^n(x')||^2$$

where W(x) is a window that delineates which area to consider for neighbouring pixels to compare contrast values.

Regional: Center-Surround Histogram

Given a rectangle $R_s(x)$around a salient region, create a frame $R(x)$ around it so that the area of the frame is equal to that of the rectangle (this is displaced as needed to fit into the image dimensions), at a suitable aspect ratio.

Create a colour RGB histogram for both the rectangle and the surrounding frame with a certain resolution (number of "bins" for each colour), and calculate how many pixels fall into each colour's bins in the frame and the rectangle.

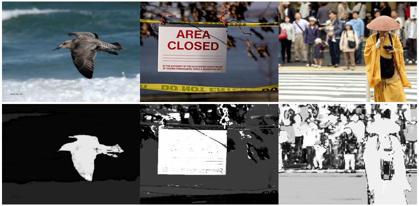
Finally, calculate the χ^2 value between the two histograms to obtain the differences between the rectangle and the surrounding frame. Do this for multiple aspect ratios, and return the largest χ^2 value and the frame that formed it:

$$f_s(x, I) = \arg\max_{R(x)} \chi^2(R(x), R_s(x)) = \arg\max_{R(x)} \frac{1}{2} \cdot \sum_{i \in bins} \frac{(hist_{R(x)_i} - hist_{R_s(x)_i})^2}{hist_{R(x)_i} + hist_{R_s(x)_i}}$$

Global: Color Spatial Distribution

Create a Gaussian Mixture Model to compute the spatial variance of colour in an image; this model is not created from every pixel but a subset in order to improve computational time without sacrificing much accuracy.

Average all the values, sum them up to get the overall covariance of the image, then assign a normalised spatial feature to each pixel based on its colour.



Learning

The values from the before-mentioned features are accumulated in a CRF model, which the dataset then trains on. MORE DETAIL HERE

Inference

CRF INFERENCE HERE

References

- Liu, Tie, et al. "Learning to detect a salient object." *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. IEEE, 2007.*
- Liu, Tie, et al. "Learning to detect a salient object." *Pattern Analysis and Machine Intelligence, IEEE Transactions on 33.2 (2011): 353-367.*
- Itti, Laurent, Christof Koch, and Ernst Niebur. "A model of saliency-based visual attention for rapid scene analysis." *Pattern Analysis and Machine Intelligence, IEEE Transactions on 20.11 (1998): 1254-1259.*
- Ma, Yu-Fei, and Hong-Jiang Zhang. "Contrast-based image attention analysis by using fuzzy growing." *Proceedings of the eleventh ACM international conference on Multimedia. ACM, 2003.*
- Stephen Gould, "DARWIN: A Framework for Machine Learning and Computer Vision Research and Development", *Journal of Machine Learning Research (JMLR), 13(Dec):35333537, 2012.*