PREDICTING VISUAL ATTENTION USING GAMMA KERNELS

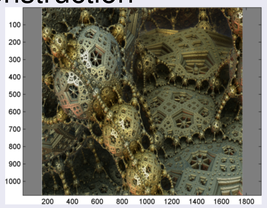
Ryan Burt, Eder Santana, Jose Principe, Nina Thigpen, Andreas Keil
University of Florida, Gainesville, Florida, USA.

Motivation

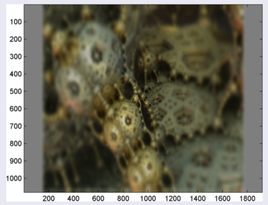
- Saliency measures are used to predict visual attention, but the human vision system does not process an entire scene in full resolution - instead it fixates on small areas successively to build an overall representation of a scene. We introduce a new saliency measure that works in these naturalistic conditions and calculates saliency maps in real time.

Foveated Images

- To mimic the human vision system, we focus on a small area and blur the periphery of an image
- To create the foveated images, we use the visual field simulator created by Geisler (2002)
- Images are downsampled and blurred using a Gaussian pyramid, with the fixation having a higher resolution than surrounding areas after reconstruction

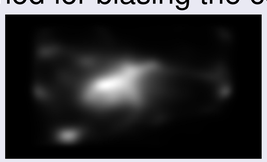


(a) Original image

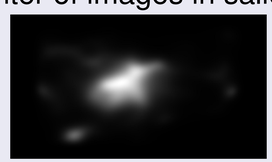


(b) Foveated

- Rather than artificially adjusting saliency maps for center bias, foveation provides a bio-inspired method for biasing the center of images in saliency



(a) Saliency on original image



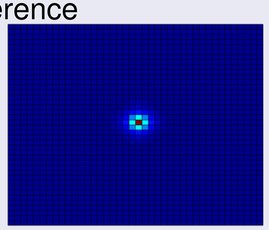
(b) Saliency on foveated image

Gamma Saliency

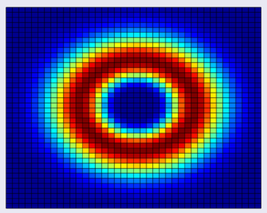
- Gamma Saliency is a center surround method that makes use of the unique shape of the 2D gamma kernel

$$g_{k,\mu}(n_1, n_2) = \frac{\mu^{k+1}}{2\pi k!}\sqrt{n_1^2 + n_2^2}^{k-1} e^{-\mu\sqrt{n_1^2+n_2^2}} \quad (1)$$

- The gamma kernel has two parameters (k, μ) that control the size, shape, and location of the kernel
- By taking a centered kernel and subtracting a kernel that surrounds it, we have a center-surround difference



(a) Center kernel



(b) Surround kernel

- By combining many of these at different scales, we have a multi-scale metric

$$g_{total} = \sum_{m=0}^{M-1} = (-1^m) g_m(k_m, \mu_m) \quad (2)$$

- The test image is converted to LAB color space, then each color is convolved with the combined gamma kernel as follows

$$S = \frac{|g \bullet L| + |g \bullet a| + |g \bullet b|}{3} \quad (3)$$

- Finally, the image is post-processed to accentuate standouts in the map, as well as blur it to match the histogram of the eye-tracking data

$$S = (S^{\alpha} G(\sigma^2)) \bullet G(.5) \quad (4)$$

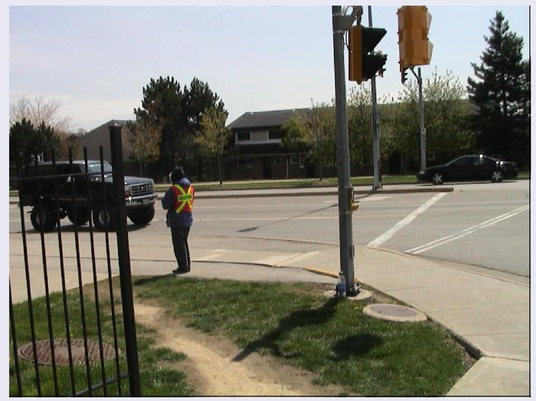
- With predefined scales, this operation consists of simple convolutions and addition, making it extremely fast

Results

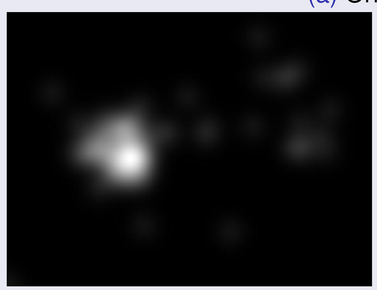
- Compared Gamma Saliency with 7 other common algorithms on Toronto and CAT2000 databases with images both normal and foveated
- Gamma Saliency computes maps faster than all other algorithms and performs the best in most metrics

Method	ROC (Judd)	ROC (Borji)	Similarity	Correlation	NSS	Time
Itti	.737	.597	.403	.314	.369	.25
AIM	.794	.657	.433	.458	.561	1.04
Torralba	.784	.650	.433	.469	.539	1.20
GBVS	.839	.664	.502	.603	**.594**	1.05
FES	.846	.571	.487	.536	.403	.29
RARE2012	.841	.656	.525	.632	.591	1.37
RCS	.819	.629	.517	.595	.517	14.91
Gamma	**.858**	**.684**	**.607**	**.649**	.483	**.21**

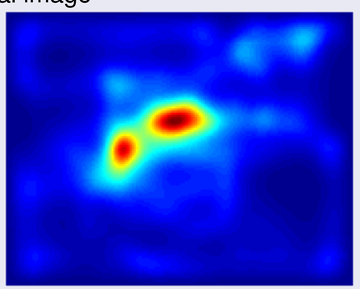
Table : Attention Prediction Results on the Foveated Toronto Database



(a) Original image



(b) Eye-tracking histogram



(c) Gamma saliency

Computational NeuroEngineering Lab
University of Florida