Focal attention models driven by image statistics

Jan-Mark Geusebroek[1], Victoria Yanulevskaya[1], Jan Bernard Marsman[2], and Frans Cornelissen[2]
[1]*Informatics Institute, University of Amsterdam, The Netherlands, mark@science.uva.nl*
[2]*School of Behavioural and Cognitive Neurosciences, University Medical Centre Groningen, The Netherlands*

The visual system evolved to provide fast and adequate, and not necessarily the best, estimates of physical properties. One way in which the visual analysis has been speeded up is by analyzing only part of the scene in detail, with our foveae. The trade-off is that we must move our eyes to areas of interest for detailed analysis. Deciding where to move requires a fast selection process, ideally based on a moment-to-moment representation of potentially "important" regions in the natural scene. We propose the visual system does so by determining areas that deviate in a statistical sense from the rest of the visual scene. This information constitutes a statistical salience map that will drive gaze and attentional shifts.

The concept of the salience map is central in the study of attention and eye movements. A number of neurally plausible computational schemes have been proposed for the creation of salience maps, most notably Li's [1] work on V1, on which our hypothesis is based, but see also [3,4]. The important new aspect that this study brings is that, unlike previous imaging and most behavioral studies, we will start from a theoretical understanding of the statistical characteristics of natural images [5,6]. Statistical analysis of natural scenes and measurement-theoretical assumptions can tell us what information can be retrieved from the visual environment, with initial experiments shown in [7].

In computer vision, the selection of salient regions in an image is a well studied topic, for a review see [8], being the first step in many computer vision algorithms. Combining the insights from eye tracking studies with theoretical models of salient regions detection, we aim to arrive at better prediction of regional image saliency. Natural image statistics provide a means to quantize saliency, by comparing the common visual structures of the world around us, in contrast to salient visual structures which stand out by their sparse occurrence in the visual field. Particularly, we consider the integrated Weibull distribution as a parameterized model, which provides a good fit to the statistics of natural images [5]. We show how distinct regimes of the integrated Weibull distribution leads to various local saliency mechanisms. With model selection techniques from information theory, we can determine the probability for every distinct regime, to explain the statistical properties of local image content. These lead to different mechanisms for saliency determination, see Figure 1 for an example.

Measurement of saliency from natural image statistics

The contrasts in a large range of natural scenes conform to a Weibull type of distribution [5]. The integrated Weibull distribution includes the class of symmetric exponential distributions. Its shape indicates different regimes of the distribution (see Figure 2), thereby roughly categorizing image content [9]. When the contrasts in the image constitute a power-law, this indicates a contrasting foreground object against a background. When more and more clutter occurs, the shape progresses from power-law through exponential to become Gaussian for high frequent textures. Images with a regular pattern can not be described well with the integrated Weibull distribution. Note that parameters can be estimated with Maximum Likelihood Estimation (MLE) technique. It is possible to distinguish tree types of images according to the behavior of integrated Weibull distribution, or conclude that integrated Weibull distribution does not describe data well. For the first case, we use Akaike's information criterion (AIC) for appropriate model selection [10]. AIC estimates expected Kullback-Leibler information, based on the log-likelihood function at its maximum point. Hence, we do not need to assume that the "true model" is in the set of candidates. Regarding the latter case, we use Anderson-Darling goodness of fit test at a 5% confidence level. The Anderson-Darling test is a generalization of the Kolmogorov-Smirnov test, which is more sensitive to deviations in the tails of the distribution. This is a relevant characteristic of the test, as in our case the tails capture the strong, hence important contrasts present in the scene. We applied the approach to 50 by 50 patches from a hundred natural images (1 megapixels) taken from the National Geographic website. Of the patches, 72% were Weibull distributed of which 22% power-law, 42% exponential, and 8% Gaussian, respectively. The remaining 28% of the patches was rejected by the Anderson-Darling test, and constitutes a mixture of Weibulls [11], or regular patterns [6], experimentally evaluated in Figure 3. Each of these distributions indicates a different mechanism for regional saliency.

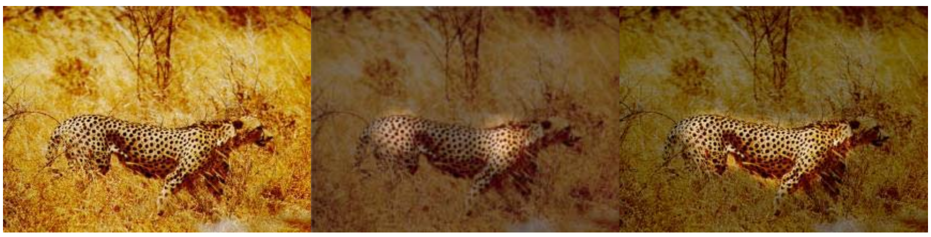

Figure 1. *An example of the use of the Weibull distribution in the determination of salient regions. Middle image shows example fixation points, sparsely sampling the object (original left). On the right, the region which deviates from the common statistics in the scene is highlighted, as determined by testing goodness-of-fit over an extended region.*

References

1. Li, Z. (2002). A salience map in primary visual cortex. *TINS*, **6**, 9-16.
2. Itti L., Koch C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res*. **40**, 1489-1506.
3. Deco G., Rolls E.T. (2004). A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Res*. **44**, 621-642.
4. Geusebroek J.M., Smeulders A.W.M. (2003). Fragmentation in the vision of scenes. *Int. Conf. Comput. Vision*, 130-135.
5. Geusebroek J.M., Smeulders A.W.M. (2005). A six-stimulus theory for stochastic texture. *Int. J. Comput. Vision*, **62**, 7-16.
6. Marsman J.B.C., Yanulevskaya V., Cornelissen F.W., Geusebroek J.M. (2007). Spatial statistics of natural images predict gaze direction, Perception **36** E*CVP Abstract Supplement*.
7. Mikolajczyk K., Schmid C. (2004). Scale & affine invariant interest point detectors. *Int. J. Comput. Vision* **60**, 63–86.
8. van Gemert J.C., Geusebroek J.M., Veenman C.J., Snoek C.G.M., Smeulders A.W.M. (2006). Robust scene categorization by learning image statistics in context, *Proc. CVPR Workshop on SLAM*.
9. Burnham K.P., Anderson D.R. (2002). Model selection and multimodel inference: A practical information-theoretic approach. Springer.
10. Burghouts G.J., Smeulders A.W.M., Geusebroek J.M. (2007). The distribution family of similarity distances. *Proc. Advances Neural Inf. Process. Syst*., 2007.

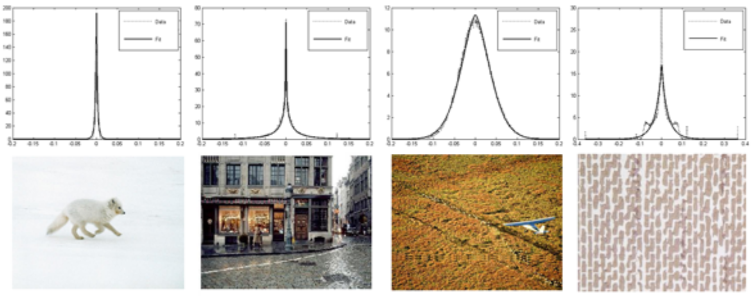

Figure 2. *The integral form of the Weibull distribution fits the statistics of contrasts in natural images very well. The distribution is characterized by a width and an exponent parameter. Its shape is exponential, where the exponent parameter determines shape from power-law like (most left) for single contrasting objects against background; to double exponential (left) for more fragmented scenes; to Gaussian (right) for high frequent textures. If the scene contains a close to regular pattern, the integrated Weibull does not fit (most right). Hence, fitting a Weibull distribution measures various regimes of natural image statistics.*

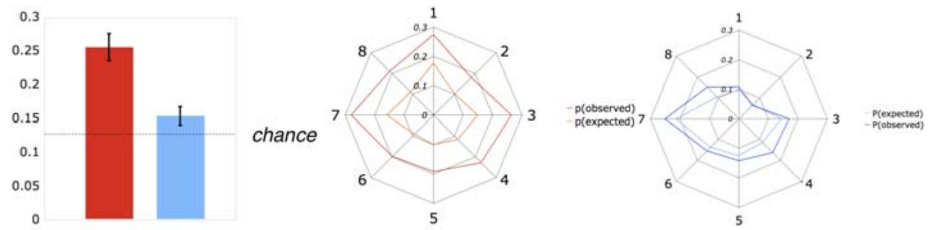

Figure 3. *Results from an eyetracker experiment testing saccade behaviour for non-Weibull versus Weibull textures. Levels indicate the probability of making a saccade towards a non-Weibull texture out of 8 circular arranged textures. Red denotes results from 15 subjects for the CuRet dataset (9 non-Weibull, 52 Weibull); Blue indicates data from 9 subjects for the Alot dataset (34 non-Weibull, 216 Weibull).*