Salience Of Orientation-Filter Responses As Suspicious Coincidence In Natural Images

Subramonia Sarma and Yoonsuck Choe [a]

[a] *Department of Computer Science*
Texas A&M University
College Station, TX 77843-3112

Abstract

Visual cortical neurons have receptive fields resembling oriented bandpass filters, and their response distributions on natural images are non-Gaussian. Inspired by this, we previously showed that comparing the response distribution to a normal distribution with the same variance gives a good thresholding criterion for detecting salient levels of edginess in images. However, the question why the comparison to a normal distribution can be so effective was not fully answered. In this paper, we approach this issue under the general framework of *suspicious coincidence* proposed by Barlow. It turns out that salience defined our way can be understood as a deviation from the unsuspicious baseline. Further, we show that the response threshold directly calculated from the white-noise based distribution closely matches that of humans, providing further support for the analysis.

Key words: Salient Contours, Gabor Filter, Suspicious Coincidence, Orientation Energy

1 Introduction

Oriented Gabor filters have been used successfully to model visual cortical neuron response [3], and as it turns out, they are biologically grounded, i.e., the shape of the Gabor filters closely resemble experimentally measured receptive fields [7]. An interesting property of such filters is that when applied to natural images, the response histogram shows a characteristic non-Gaussian shape with a sharp peak at zero. Thus, when compared to a normal distribution with the same variance, the response distribution has a heavy tail [5, 11]. Such a response property has been found to be useful in tasks such as denoising [11] and salient contour detection through thresholding [8]. For instance, Lee and Choe [8] showed that a simple thresholding criterion based on the comparison of the filter response distribution to a normal distribution of the same variance can accurately predict the salience level perceived by humans. Although the method was effective, it was not shown why a

[1] This research was supported in part by the Texas Higher Education Coordinating Board (ATP000512-0217-2001)

normal distribution serves so well as a baseline in such a comparison. In this paper, we frame the problem in terms of the concept of *suspicious coincidence*, proposed by Barlow [2]: Two statistical events A and B are said to be a suspicious coincidence if they occur more often together than can be expected from their individual probabilities. In other words, the two events should be statistically non-independent in order for them to be deemed a suspicious coincidence: $P(A, B) > P(A)P(B)$. This approach is easily extended into the problem of image analysis, where suspiciousness can be determined by testing the inequality shown above where events correspond to pixels from different locations in the image [1]. Suspiciousness is directly linked to salience, i.e., more suspicious events may be seen as more salient to a perceptual system. An image where each pixel is independent from each other would be seen as containing no suspicious coincidence between any pair of pixels, i.e. a white-noise image does not contain any visible structure.

2 Calculation of Oriented Filter Response

To find the orientation response (or energy) distribution, we follow the procedure described in [6]. The method uses a sequence of convolutions: first the difference of Gaussian (DoG), and then the oriented Gabor filters to calculate the orientation filter response. The DoG filter uses the difference of two Gaussian functions whose widths differ by a factor of 0.5, as $D(x, y) = G_{(\sigma/2)^2}(x, y) - G_{\sigma^2}(x, y)$, where $G_{\sigma^2}(\cdot)$ is a Gaussian function with variance σ^2.

The gray level intensity matrix I of the input image is convolved with the DoG filter to obtain the resultant matrix I_d ($I_d = I * D$). We used a DoG filter of size 7×7 for all of our experiments. The filtered image is then convolved with oriented Gabor functions [4] of both even and odd phases with orientation θ, phase ϕ, and width σ to obtain the orientation energy matrix E_θ. The spatial frequency and aspect ratio parameters of the Gabor filters were set to 1 each, and the convolution kernels were sized 7×7 as usual. The orientation energy matrix for a single orientation θ is

$$E_\theta = ((e^{-\frac{x'^2+y'^2}{2\sigma^2}} \cos(2\pi x')) * I_d)^2 + ((e^{-\frac{x'^2+y'^2}{2\sigma^2}} \cos(2\pi x' + \frac{\phi}{2})) * I_d)^2, \quad (1)$$

where $x' = x\cos(\theta) + y\sin(\theta)$, $y' = -x\sin(\theta) + y\cos(\theta)$, and (x, y) is the pixel location as above.

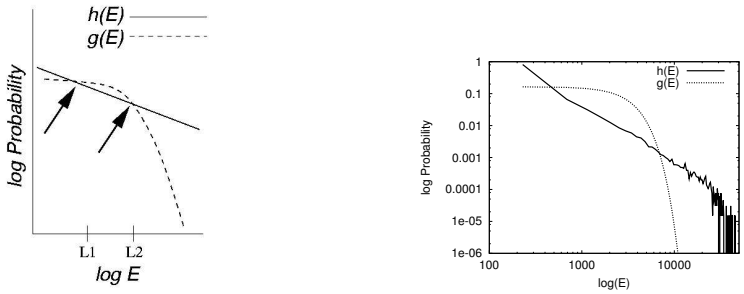
For each location (x, y), we obtained the vector sum of six $(\theta, E_\theta(x, y))$ pairs in polar coordinates ($\theta = 0, \frac{\pi}{6}, \frac{2\pi}{6}, \frac{3\pi}{6}, \frac{4\pi}{6}, \frac{5\pi}{6}$) to find the combined orientation energy which gives the estimated orientation $\theta^*(x, y)$ and the associated orientation energy value $E^*(x, y)$ at that location. The orientation energy distribution is then estimated from the E^* responses using a histogram of bin size 100, followed by normalization, as $h(E) = \frac{f(E)}{\sum_{x \in B_h} f(x)}$ where $f(E)$ is the frequency of energy value E in the histogram, B_h is the set of histogram bin locations, and $h(E)$ is the resulting probability mass function which specifies the orientation energy distribution for the filtered image.

One way to detect salient levels of orientation energy is by comparing the orientation energy distribution for the input image with a normal distribution of the same variance as proposed in [8], so that unusually high levels of orientation energy show up as salient. We calculate the raw second moment of the E distribution (i.e., the expected value of E^2) for the input image as $\sigma_h^2 = \sum_{x \in B_h} x^2 h(x)$. We use this calculated σ_h^2 to find the matching continuous normal probability density function $N(x; 0, \sigma_h^2)$ with mean 0, variance σ_h^2 for all $E \in B_h$ and normalize it to find the discretized normal probability mass function ***g(E)*** of the orientation energy E:

$$g(E) = \frac{N(E; 0, \sigma_h^2)}{\sum_{x \in B_h} N(x; 0, \sigma_h^2)}. \quad (2)$$

3 Experiments and Results

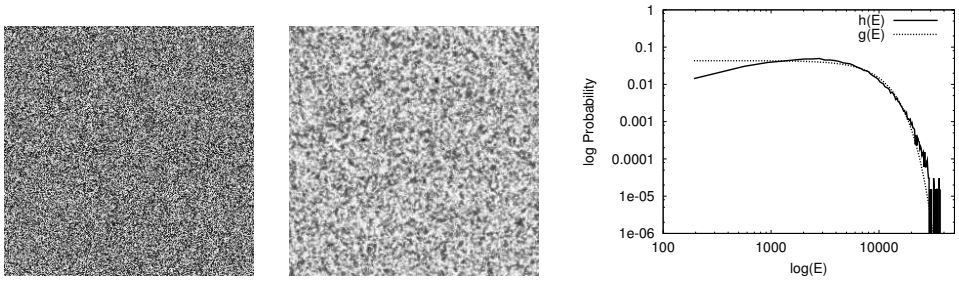
To find out how $h(E)$ relates to suspicious coincidence, we used 42 natural images (resized to a size of 256×256) from a set of publicly available stock photos from Kodak (the same source as in [6]). For each image, the orientation energy E was calculated, and to avoid any orientation bias, only the circular central area of the energy matrix was used for the calculation of its distribution $h(E)$ and the matching normal distribution $g(E)$, following [6]. The orientation energy distribution of the images follows approximately a power law (i.e., $p(x) = 1/x^a$ where ***a*** is the fractal exponent), and as such, it has a heavy-tail where extreme values have higher probability of occurrence compared to a normal distribution of the same variance. For example, Fig. 1(a) shows a theoretical power-law distribution compared with a normal distribution of the same variance, and fig. 1(b) shows the orientation energy distribution calculated from a natural image $h(E)$ compared to its matching normal distribution $g(E)$. The straight declining slope characteristic of a power law distribution is evident in $h(E)$. The two curves intersect at two points, near $E \sim 500$ (i.e., L_1) and $E \sim 7,000$ (L_2). Beyond L_2, $g(E)$ plummets, but $h(E)$ remains high.



(a) (b)

Fig. 1. **Orientation Energy Distribution of a Natural Image vs. its Matching Normal Distribution.**

Lee and Choe empirically derived the effective threshold for the detection of salient contours, which was linear to the orientation energy corresponding to the second point of intersection (L_2) of the response distribution and its matching normal distribution [8]. However, it was not clear why the simple idea of comparing to a



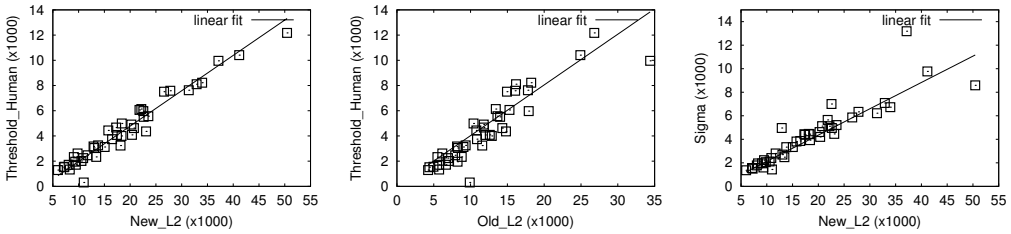
(a) White-Noise (b) Ori. Energy (OE) (c) OE vs. Normal Distribution

Fig. 2. **Orientation Energy Distribution of a White-Noise Image vs. Its Matching Normal Distribution..** (a) A white-noise image shown in gray-scale. (b) The orientation energy E of the image in (a). There is no clear structure visible. (c) The log-log plot shows the orientation energy distribution $h(E)$ for the white-noise image against a normal distribution of the same variance $g(E)$. The two distributions show a close resemblance.

normal distribution has to be so effective. That is, why does a Gaussian distribution form a reasonable baseline for comparison? We observe that the detected salience in our method may correspond to a *suspicious* event in the image, i.e., a suspiciously high degree of edginess. If suspiciousness (as defined by Barlow [2]) is indeed related to salience defined in our way, we can expect that a random image with completely independent statistical features (e.g., a uniformly randomly distributed white-noise image) may not show suspicious (i.e., salient) levels of orientation energy under our criterion. For this to happen in our method, the orientation energy distribution of white-noise images should not have a heavy tail, and in a more strict sense, it should coincide with its matching normal distribution. That is, it should be near-Gaussian.

To test if this is the case, we calculated the orientation energy distribution from a white-noise image and compared it with the matching normal distribution. The white-noise image was a 256×256 intensity matrix of uniformly randomly distributed values between 0 and 255. The orientation energy distribution was then found using the procedure outlined in the previous section. We then compared the orientation energy distribution to the matching normal distribution of the same variance to see if there is any similarity between the two. It turns out that the two distributions closely overlap as expected (Fig. 2; cf. [11]). These results suggests that normal distributions correspond to a baseline where all pixel values are independent (and thus no suspicious coincidence), and any deviation from this baseline can be seen as suspicious, or *salient*. Thus, salience as defined in our work can be understood as a deviation from the unsuspicious baseline.

From this result, we expect that the white-noise based orientation energy distribution can also be used directly in finding the appropriate threshold. To test this, we conducted another experiment in which we generated new L_2 values by comparing the orientation energy distribution with the white-noise based distribution. Since the standard deviation of a random variable scaled by the factor of c is $c \times \sigma$ where σ was the standard deviation before scaling, we multiplied the orientation energy



(a) (b) (c)

Fig. 3. **Comparison of the White-Noise and Gaussian L_2 Values with human-chosen thresholds and σ_h.** (a) The new L_2 values derived from the white-noise based distribution against the human-chosen thresholds for a set of 42 natural images are shown (each square represents one image). The correlation coefficient was $r = 0.98$. (b) The same is shown for the case with the old L_2 values based on the normal-distribution baseline. The correlation coefficient was $r = 0.91$, lower than in (a). (c) The new L_2 values derived from the white-noise based distribution against the raw standard deviation σ_h of the distribution $h(E)$ is shown. (Each square represents one image.) We can see a clear linear relationship ($r = 0.91$). In all three plots, the least-square fit is shown in the background.

matrix of the white-noise image with a constant σ_h/σ_r, where σ_h and σ_r are the standard deviations from a natural image and the white-noise image, respectively. Then the resulting orientation energy matrix has the same variance as the reference distribution calculated from a given natural image. The new L_2 values were then found computationally by comparing the two distributions. These values were compared to the orientation energy thresholds selected by humans. All of the thresholds for the 42 natural images were determined by a single person (SB) in our research group. For each image, the thresholded E at 55% to 95% percentile of $h(E)$ at an interval of 5% was shown to SB and he determined the best threshold according to the following criteria [8]: (1) object contours should be preserved, and (2) noisy background edges must be removed, as much as possible. The results are shown in Fig. 3(a). It is clear that the new white-noise based L_2 values also have a strong linear relationship with the human-selected thresholds, even more so than the old Gaussian-based L_2 values(Fig. 3(b)). (We are currently investigating the cause of this small enhancement.)

4 Discussion

The usefulness of identifying significant values of orientation energy has been studied previously, in applications such as denoising, texture perception and image representation. For example, Malik et al. [10] used peak values of orientation energy to define boundaries of regions of coherent brightness and texture. The non-Gaussian nature of orientation energy (or wavelet response) histograms has also been recognized and utilized for some time now, especially in denoising and compression [11]. Our approach was inspired from Barlow's observation [1] that comparing peaked distributions with high kurtosis and distributions derived from an unsuspicious baseline could be useful. The current work, to our knowledge, is the first systematic study of the relationship between perceived salience in humans and the orientation energy distribution under the framework of suspicious coincidence.

Response histograms in general are also widely used in computational vision. For example, Liu and Wang used what they call the spectral histogram (which is a combination of many different kinds of filter response histograms) to segment and synthesize texture images [9]. It would be interesting to find out whether other response histograms can be analyzed and used in a similar manner as described in this paper for salience detection in different image feature spaces.

5 Conclusion

In this paper, we have shown that the good performance shown in orientation energy thresholding based on the comparison of the orientation energy distribution and its matching normal distribution can be analyzed and understood under the general framework of suspicious coincidence. We directly used a scaled energy distribution from white-noise images to further demonstrate this point through a comparison to human performance. The results suggest that a similar approach can be applied to other sensory tasks where a similar response distribution is found.

References

[1] Barlow, H. (1994). What is the computational goal of the neocortex? In Koch, C., and Davis, J. L., editors, *Large Scale Neuronal Theories of the Brain*, 1–22. Cambridge, MA: MIT Press.

[2] Barlow, H. B. (1989). Unsupervised learning. *Neural Computation*, 1:295–311.

[3] Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, 20:847–856.

[4] Daugman, J. G. (1989). Entropy reduction and decorrelation in visual coding by oriented neural receptive fiends. *IEEE Transactions on Biomedical Engineering*, 36:107–114.

[5] Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4:2379–2394.

[6] Geisler, W. S., Perry, J. S., Super, B. J., and Gallogly, D. P. (2001). Edge Co-occurrence in natural images predicts contour grouping performance. *Vision Research*, 41:711–724.

[7] Jones, J. P., and Palmer, L. A. (1987). An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233–1258.

[8] Lee, H.-C., and Choe, Y. (2003). Detecting salient contours using orientation energy distribution. In *Proceedings of the International Joint Conference on Neural Networks*, 206–211. IEEE.

[9] Liu, X., and Wang, D. (2002). A spectral histogram model for texton modeling and texture discrimination. *Vision Research*, 42:2617–2634.

[10] Malik, J., Belongie, S., Shi, J., and Leung, T. K. (1999). Textons, contours and regions: Cue integration in image segmentation. In *ICCV (2)*, 918–925.

[11] Simoncelli, E. P., and Adelson, E. H. (1996). Noise removal via bayesian wavelet coring. In *Proceedings of IEEE International Conference on Image Processing*, vol. I, 379–382.