Relating Spatial and Spectral Models of Oriented Bandpass Natural Images

Zeina Sinno and Alan C. Bovik
Laboratory for Image and Video Engineering
Department of Electrical and Computer Engineering
The University of Texas at Austin, Austin, TX, USA

***Abstract*—Univariate models of the Natural Scene Statistics (NSS) of perceived digital pictures have been deployed in a wide variety of image and video processing applications. However, much less effort has been made towards understanding, modeling, and using bivariate image NSS. Towards filling this gap, Su *et al.* developed a closed form correlation model of oriented bandpass natural images applicable to adjacent pixels. We later extended this model to account for pixels separated by larger spatial distance. Here, we expand our previous work further to model the bivariate responses of bandpass spatial filters covering a wider range of bandwidths. Furthermore, we study the relationship between the parameters of the closed form correlation model and image spectral models.**

***Index Terms*—Natural Scene Statistics; Bivariate Correlation Models; Bandpass Natural Images; 1/*f* Noise Models.**

INTRODUCTION

Understanding the perception of visual images has proven to be highly useful when designing image and video processing algorithms. By combining visual models with dual models of univariate Natural Scene Statistics (NSS), tremendous success has been obtained [1] on tasks as image interpolation [2], texture modeling [3], [4], full reference and blind image/video quality prediction (e.g. and MOVIE) [5], [6], and image defocus algorithms [7]. Very recent bivariate NSS models have proven quite difficult for improving color depth and range modeling [8] and stereopair quality evaluation [9]. Deepening our understanding of the statistical relationships between neighboring pixels in NSS is a desirable goal.

The main objectives of this work are 1) to extend our previous work [10] across more scales and distances, 2) to uncover the relationship behind the parameters of the correlation model as a function of scale and 3) to understand the relationship between the classical 1/*f* image model [11] and the latter parameters. The paper is organized as follows: in Section I, we review relevant concepts from the literature regarding 1/*f* processes and past work on bivariate natural scene statistics, in Section II we briefly present our model, in Section III we study the relationship between our extended model 1/*f* model [11] and lastly validate our model in Section IV.

I. PREVIOUS WORK AND ESTABLISHED OBSERVATIONS

We begin by reviewing the 1/*f* noise process model and its connection to our correlation model.

A. 1/f Image Model

Models of 1/*f* nonstationary random processes have been applied to noise in vacuum tubes [12], biological evolution [13], animal populations [14] , the development of economic systems [15], and personal growth and development [16].

The wide range of applicability of the 1/*f* model is explained by a deep law of nature that applies to nonequilibirum systems. Using the tools of linear system analysis, Keshner [11] derived a 1D nonstationary autocorrelation function 1/*f* process model.

Regarding images, Carlson *et al.* [17] noted that although natural scenes have immense diversity, the amplitude spectra of images follow the 1/*f* model. Field *et al.* [1] and Tolhurst *et. al.* [18] later confirmed this observation. In the temporal domain, the autocorrelation of 1/*f* processes follows a long tailed form inversely proportional to a power of time. Here we study the correlations of bandpass filtered images over different spatial separations. Our interest in this is heightened by a desire to understand how these correlations may relate to the bivariate behavior of bandpass neurons in cortical area V1 [1], [19], [20], [21], [22]. The basis behind this formidable encoding schema is that the receptive fields of area V1 simple cells are well modeled as localized, oriented, bandpass filters making sparse spatial representations of natural visual stimuli possible [21].

B. Bivariate NSS Models

Here we review early work on bivariate natural scene statistics and summarize recent open-form and closed-form bivariate NSS models.

The coefficients of orthonormal wavelet coefficients of natural images tend to be decorrelated [23] and to exhibit intra and inter scales dependencies [24]. This can be modelled using a circularly symmetric bivariate distribution to account for the spatial dependencies between image wavelet coefficients and their parents (at coarser scale locations); as in Sendur *et al.*'s model [25].

The latter observations were exploited in Portilla and Simoncelli [26], who tackled the problem of natural image texture modeling by imposing a set of parametric constraints on pairs of complex wavelet coefficients at adjacent spatial locations, orientations and scales within a non-Gaussian Markov Random Field context. Their method of selecting statistical constraints relied on structural and statistical observations of the early Human Visual System (HVS). Other open-form

models include Po *et al.*'s model [27] which represents natural images using a hidden Markov tree, a Gaussian mixture model, and two dimensional contourlets to capture interlocation, interscale, and interdirection dependencies; and Mumford *et al.* [28] who proposed an infinitely divisible model of generic image statistics by segmenting the environment into objects cast against an ergodic field, containing regions with little information (e.g. blue sky). However this model does not capture the 2D dependencies between (bandpass) image luminances.

The first attempt to create a closed form bivariate correlation model was described in [29], which modeled the responses of adjacent oriented bandpass natural image pixels. This model proved useful in two applications; color depth and range modeling [8] and stereopair quality evaluation [9]. This model of "perceptually transformed" bandpass and normalized images could supply powerful priors for a more extensive set of visual processing problems. In [10], we broadened this model to account for non-adjacent distances, while also exploring simplifications of the model. However, we did not explore the effect of scale on the model.

II. The Model

Next we unveil the processing steps used in our model, while relating each step to relevant processing stages in visual cortex. Image processing was accomplished on the 29 high quality pristine images from the LIVE IQA database [30]. The followed stages are summarized in Fig.1.

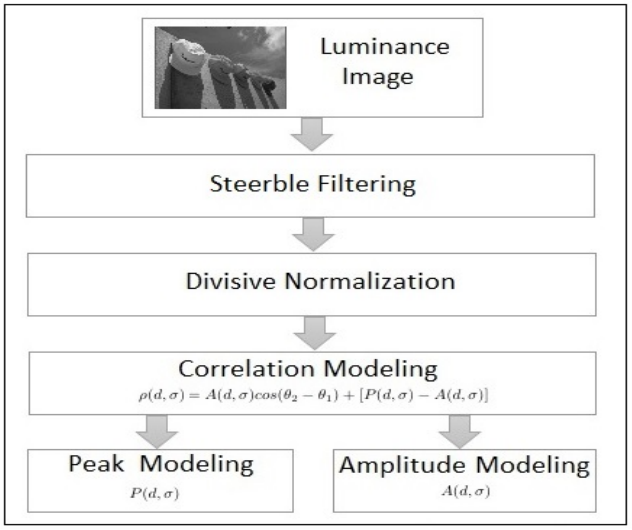

Fig. 1. Flow Chart of the Model

A. Steerable Filters

We only study NSS modeling of bandpass luminance images although chromatic NSS models are also of high interest. We use steerable filters [31] as a simple and easily manipulated model of bandpass simple cells in primary visual cortex. A steerable filter at a given frequency tuning orientation θ_1 is defined by:

$$F(\theta_1) = cos(\theta_1)F_x + sin(\theta_1)F_y \tag{1}$$

where F_x and F_y are the gradient components of the two dimensional bivariate gaussian function, where F_x and F_y have unit energy. Image decompositions using steerable filters yield decorrelated representations over scale and orientation which resemble spatial cortical responses. Altering the variance σ of the bivariate gaussian function (differentiated to obtain F_x and F_y) accounts for the multi-scale decomposition computed by simple cells in area V1. In [29], the variance σ spanned values ranging between 1 to 4.

We deployed many more scales, by varying σ between 1 and 15. The (half-peak) octave bandwidth of the steerable filter (1) is about 2.6 octaves. We computed responses on all images over 15 frequency tuning orientations θ_1 ranging over $[0, \pi/15, 2\pi/15, ..., \pi]$.

B. Divisive Normalization

Divisive normalization was applied on all the steerable filter responses. This step models the nonlinear adaptive gain control of V1 neuronal responses in visual cortex [19]. This process also gaussianizes and further decorrelates the image data [32], [23]. The divisive normalization model used here is:

$$u(x_i, y_i) = \frac{w(x_i, y_i)}{\sqrt{s + \sum_j g(x_j, y_j) w(x_j, y_j)^2}} \tag{2}$$

where (x_i, y_i) are spatial coordinates, w are the wavelet coefficients, u are the coefficients obtained after divisive normalization, and $s = 10^{-4}$ is a semi saturation constant. The weighted sum is computed over a spatial neighborhood of pixels in the same sub-band indexed by j (assuming a window of dimensions 3×3 hence $j = 9$). The Gaussian weighting function, $g(x_i, y_i)$, is circularly symmetric and has unit volume.

C. Modeling the Correlation Function

We will start by giving the intuition behind computing the correlation function. We computed the bivariate joint distribution of our model and used a multivariate Bivariate Generalized Gaussian Distribution (BGGD) to model it:

$$p(\mathbf{x}; \mathbf{M}, \alpha, \beta) = \frac{1}{|\mathbf{M}|^{\frac{1}{2}}} g_{\alpha,\beta}(\mathbf{x^T M^{-1} x}) \tag{3}$$

where $\mathbf{x} \in \boldsymbol{R}^2$, $\boldsymbol{M}$ is an 2×2 scatter matrix, α and β are scale and shape parameters respectively, and $g_{\alpha,\beta}(.)$ is the density generator. In order to understand the behavior of the scatter matrix, we computed the correlation function between a window of the image and a shifted version of it. The distance between the two shifted windows will be denoted by d while the angle between them will be denoted by θ_2. The correlation function model exhibits a periodic behavior in the relative angle $\theta_2 - \theta_1$, where θ_1 and θ_2 are the sub-band tuning and spatial orientations respectively. The spatial orientation is $\theta_2 = \arctan(\frac{\delta_y}{\delta_x})$ where δ_x and δ_y are the row

and column differences between coordinates of the responses after divisive normalization. Similarly to [10], δ_x and δ_y span absolute integer distances between 1 and 10, although sub-pixel distances are of interest. The tuning orientation θ_1 is the frequency tuning orientation of the steerable filter. We deployed 15 sub-band orientations $\{0, \frac{\pi}{15}, \frac{2\pi}{15}, ..., , \frac{14\pi}{15}\}$ rad when building our model.

The correlation is modeled as:

$$\rho(d,\sigma) = A(d,\sigma)cos(2(\theta_2 - \theta_1)) + c(d,\sigma) \quad (4)$$

where $A > 0$ is the amplitude, c is an offset, d is the spatial separation between the target pixels and σ is the steerable filter variance.

This modeling process was performed on all the 100 pristine images for each θ_1, θ_2 and scale. Model fits were applied on the average correlation values of the 100 images from the two databases. We used non-linear regression to find the best fit.

By examining the correlation coefficient plots we observe that the maximal correlation $P = max(\rho)$ is obtained when $\theta_2 - \theta_1$ is equal to 0, as in Fig. 2. Generally, the maximal correlation falls as the relative distance between the bandpass samples increases. This is depicted in Fig. 2.

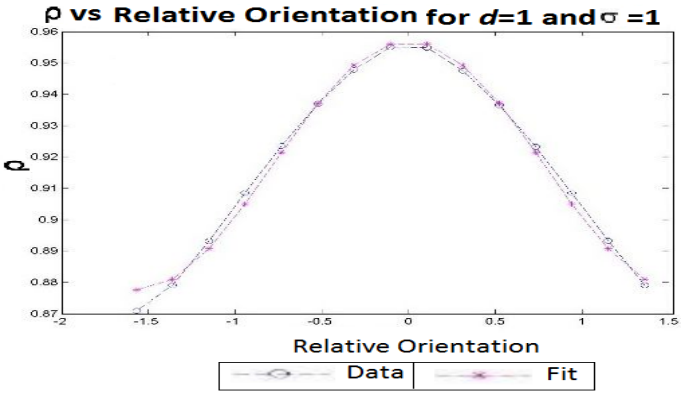

Fig. 2. ρ versus the relative orientation at $\theta_2 = \pi/2$, for $d = 1$ and $\sigma = 1$

III. Relation of the Model to the 1/*f* Processes

The correlation function of 1/*f* processes follows a long tailed distribution [11] as a function of time. We would like to understand this relationship for images: we consider the correlation function, ρ as a function of spatial separation. Here, we define the peak as $P = max(\rho) = A + c$, therefore for convenience we will rewrite (3) as:

$$\rho(d,\sigma) = A(d,\sigma)cos(2(\theta_2 - \theta_1)) + [P(d,\sigma) - A(d,\sigma)] \quad (5)$$

Looking at $P(d)$, the value decreases from $P(0) = 1$ as the spatial separation increases; which is natural because one should expect less correlation between pixels as the spatial separation increases. We decided to model the peak in the form of $\frac{K}{\frac{d}{a_0}^{b_0}+K}$, which is similar to the model arrived at in [11] but using a different way of stabilizing near $d = 0$. For $\theta_2 = \pi/2$, P is expressed as:

$$P(d,\sigma) = \frac{3}{(\frac{d}{a_0(\sigma)})^{b_0(\sigma)} + 3} \quad (6)$$

The value $K = 3$ in the denominator guarantees the stability of the system when d approaches 0 while giving a good fit to the empirical data. This results in the presence of the term 3 in the nominator so that the peak value does not exceed 1 because the correlation cannot exceed a value of 1. The parameters a_0 and b_0 exhibit a nearly ideal linear behavior against scale:

$$a_0 = 0.777\sigma + 0.487 \quad (7)$$

$$b_0 = -0.059\sigma + 2.514 \quad (8)$$

We model A as the difference of two functions of the form (6) scaled differently. For $\theta_2 = \pi/2$, the following good fit is obtained:

$$A(d,\sigma) = \frac{3}{(\frac{d}{a_1(\sigma)})^{b_1(\sigma)} + 3} - \frac{3}{(\frac{d}{a_2(\sigma)})^{b_2(\sigma)} + 3} \quad (9)$$

Similarly to a_0 and b_0, a_1, b_1, a_2 and b_2 depict a linear behavior:

$$a_1 = 1.907\sigma + 0.511 \quad (10)$$

$$b_1 = 0.048\sigma + 3.464 \quad (11)$$

$$a_2 = 0.923\sigma + 0.440 \quad (12)$$

$$b_2 = -0.090\sigma + 2.417 \quad (13)$$

We fitted P and A using a two steps nonlinear regression. First, we fitted P and A to (6) and (8) respectively. Then we fixed the values a_0, a_1 and a_2 and reused (6) and (8) to find the values for b_0, b_1 and b_2.

Fig. 3 presents examples of the peak and amplitude versus the spatial separation, d, for multiple values of σ at $\theta_2 = \pi/2$.

IV. Validation of the Models of P and A

We validated the models of P and A by taking the Mean Squared Error (MSE) and Mean Absolute Error (MAE) between P obtained by fitting the data and the estimated P using (6). We repeated the same operations for A and (9). The results for a few scales are summarized in Table I.

TABLE I
MSE and MAE of A and P

	MSE of A	MAE of A	MSE of P	MAE of P
$\sigma = 1$	1.223E-04	8.553E-03	8.679E-04	2.519E-02
$\sigma = 2$	2.962E-05	4.185E-03	7.682E-04	2.083E-02
$\sigma = 3$	3.215E-06	1.560E-03	9.823E-05	8.182E-03
$\sigma = 4$	1.781E-07	3.043E-04	4.598E-05	5.601E-03
$\sigma = 5$	3.992E-08	1.501E-04	3.374E-05	5.003E-03
$\sigma = 6$	2.063E-08	8.568E-05	1.735E-05	3.482E-03

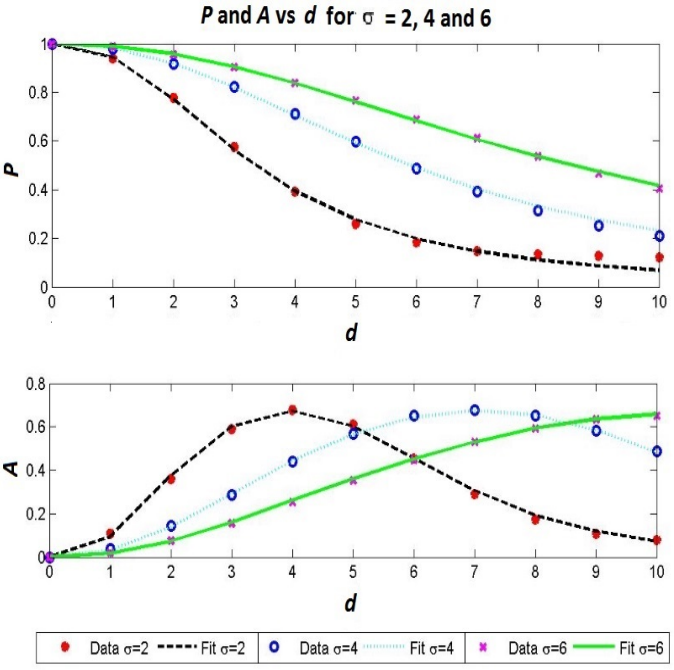

Fig. 3. P and A versus d for $\theta_2 = \pi/2$ for σ = 2, 4 and 6

Conclusion and Future Work

In this paper, we extended the work in [10] over more scales and presented how the correlation of pixels in natural images at different spatial separations follow an intuitive functional form. One advantage of our model is its simplicity. In the future, we will extend this model to all possible spatial angles or θ_2 values, study how the model varies for different modalities of images such as infrared and HDR images, as the bivariate NSS of these two types of images have not been explored yet, and use the model as a building block in image/video quality assessment metrics and in texture modeling.

References

[1] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *JOSA A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[2] A. D. D'Antona, J. S. Perry, and W. S. Geisler, "Humans make efficient use of natural image statistics when performing spatial interpolation," *J. Vision*, vol. 13, no. 14, p. 11, 2013.

[3] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *IEEE Trans. Pattern Anal. Machine Intel.*, vol. 12, no. 1, pp. 55–73, 1990.

[4] M. Clark and A. C. Bovik, "Experiments in segmenting texton patterns using localized spatial filters," *Patt. Recog.*, vol. 22, no. 6, pp. 707–717, 1989.

[5] Z. Wang, A. C. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Imag. Proc.*, vol. 13, no. 4, pp. 600–612, 2004.

[6] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the dct domain," *IEEE Trans. on Imag. Proc.*, vol. 21, no. 8, pp. 3339–3352, 2012.

[7] J. Burge and W. S. Geisler, "Optimal defocus estimation in individual natural images," *Proc. of the Nat. Acad. of Sci.*, vol. 108, no. 40, pp. 16 849–16 854, 2011.

[8] C. C. Su, L. K. Cormack, and A. C. Bovik, "Bivariate statistical modeling of color and range in natural scenes," in *Proc. SPIE, Human Vis. Electron. Imag. XIX*, vol. 9014, Feb. 2014.

[9] C. Su, L. Cormack, and A. Bovik, "Oriented correlation models of distorted natural images with application to natural stereopair quality evaluation," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1685–1699, May 2015.

[10] Z. Sinno and A. C. Bovik, "Generalizing a closed-form correlation model of oriented bandpass natural images," in *IEEE Glob. Conf. on Sig. and Info. Proc.*, Dec 2015.

[11] M. S. Keshner, "1/f noise," *Proc. of the IEEE*, vol. 70, no. 3, pp. 212–218, 1982.

[12] J. B. Johnson, "Johnson and 1/f noise," *Nature*, vol. 119, p. 50, 1927.

[13] J. M. Halley, "Ecology, evolution and 1f-noise," *Trends in Ecology & Evolution*, vol. 11, no. 1, pp. 33–37, 1996.

[14] A. E. Cohen, L. J. H. Gonzalez, A., O. L. Petchey, D. Wildman, and J. E. Cohen, "A novel experimental apparatus to study the impact of white noise and 1/f noise on animal populations," *Proc. of the Roy. Soc. of London B: Bio. Sci.*, vol. 265, no. 1390, pp. 11–15, 1998.

[15] R. T. Baillie, "Long memory processes and fractional integration in econometrics," *Journal of econometrics*, vol. 73, no. 1, pp. 5–59, 1996.

[16] D. L. Gilden, T. Thornton, and M. W. Mallon, "1/f noise in human cognition," *Science*, vol. 267, no. 5205, pp. 1837–1839, 1995.

[17] C. Carlson, "Thresholds for perceived image sharpness," *Photog. Sci. Engng.*, vol. 22, pp. 69–71, 1982.

[18] D. J. Tolhurst, Y. Tadmor, and T. Chao, "Amplitude spectra of natural images," *Opht. Phys. Opt.*, vol. 12, no. 2, pp. 229–232, 1992.

[19] M. Carandini, D. J. Heeger, and A. J. Movshon, "Linearity and normalization in simple cells of the macaque primary visual cortex," *J. Neurosci.*, vol. 17, no. 21, pp. 8621–8644, 1997.

[20] H. B. Barlow, "The coding of sensory messages," in *Curr. Prob. in Anim. Behav.*, 1961.

[21] B. A. Olshausen and D. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[22] H. Lee, C. Ekanadham, and A. Y. Ng, "Sparse deep belief net model for visual area v2," in *Adv. in Neur. Infor. Proc. Sys.*, 2008, pp. 873–880.

[23] E. P. Simoncelli, "Modeling the joint statistics of images in the wavelet domain," *SPIE Int'l Symp. on Opt. Sci., Eng., and Instrum.*, pp. 188–195, 1999.

[24] J. Liu and P. Moulin, "Information-theoretic analysis of interscale and intrascale dependencies between image wavelet coefficients," *IEEE Trans. Image Process.*, vol. 10, no. 11, pp. 1647–1658, 2001.

[25] L. Sendur and I. W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2744–2756, 2002.

[26] J. Portilla and E. P. Simoncelli, "Texture modeling and synthesis using joint statistics of complex wavelet coefficients," *IEEE workshop on stat. comp. th. vision*, vol. 12, 1999.

[27] D.-Y. Po and M. N. Do, "Directional multiscale modeling of images using the contourlet transform," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1610–1620, 2006.

[28] D. Mumford and B. Gidas, "Stochastic models for generic images," *Quarterly App. Math.*, vol. 59, no. 1, pp. 85–112, 2001.

[29] C. C. Su, L. Cormack, and A. C. Bovik, "Closed-form correlation model of oriented bandpass natural images," *Signal Processing Lett., IEEE*, vol. 22, no. 1, pp. 21–25, Jan 2015.

[30] H. Sheikh, M. Sabir, and A. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Processing*, vol. 15, no. 11, pp. 3440–3451, Nov 2006.

[31] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Machine Intell.*, no. 9, pp. 891–906, 1991.

[32] D. L. Ruderman and W. Bialek, "Statistics of natural images: Scaling in the woods," *Phys. Rev. Lett.*, vol. 73, no. 6, p. 814, 1994.