Towards a Closed Form Second-Order Natural Scene Statistics Model

Zeina Sinno, *Student Member, IEEE*, Constantine Caramanis, *Member, IEEE*, and Alan C. Bovik, *Fellow, IEEE*

***Abstract*— Previous work on natural scene statistics (NSS)-based image models has focused primarily on characterizing the univariate bandpass statistics of single pixels. These models have proven to be powerful tools driving a variety of computer vision and image/video processing applications, including depth estimation, image quality assessment, and image denoising, among others. Multivariate NSS models descriptive of the joint distributions of spatially separated bandpass image samples have, however, received relatively little attention. Here, we develop a closed form bivariate spatial correlation model of bandpass and normalized image samples that completes an existing 2D joint generalized Gaussian distribution model of adjacent bandpass pixels. Our model is built using a set of diverse, high-quality naturalistic photographs, and as a control, we study the model properties on white noise. We also study the way the model fits are affected when the images are modified by common distortions.**

***Index Terms*— Natural scene statistics, bivariate correlation models, bandpass images, 1/*f* noise models.**

I. Introduction

In THE early 1990s, Ruderman and Bialek [2] observed that images processed by a simple bandpass operation followed by a local divisive nonlinearity become gaussian distributed with a remarkably high degree of regularity. This observation has led to deeper parametric models of the statistics of visual data, that are commonly referred to as Natural Scene Statistics (NSS) models, although they are applicable to photographic images of the world at large, including of human-made objects.

A. Overview

NSS models are useful probes of the visual brain, and of how it has evolved to efficiently process gigantic amounts of visual data [3]. The parameters of univariate NSS models samples of bandpass images have been used as fundamental low-level picture descriptors to successfully solve image and video processing and analysis tasks such as image interpolation [4], texture modeling [5], [6], full reference and blind image quality prediction [7]–[11], and color depth modeling [12]. Extending NSS models to characterize the bivariate behavior of images could help advance improved solutions to a wide variety of applications. However to date, little effort has been applied towards modeling the bivariate NSS of bandpass image samples. Simoncelli [13] studied the problem, and developed a parametric model of the conditional distributions of neighboring wavelet coefficients that represents and synthesizes textures. Lee *et al.* [14] studied the relationship between spatially separated bandpass (wavelet) image coefficients and found that the spatial covariance functions follows a reciprocal power law.

Prior efforts on the bivariate NSS have not produced closed form representations. The first attempt to do so was reported in Su *et al.* [1]; but their model was incomplete. Here we advance this problem by extending their closed form model to account for spatial separations between the bandpass samples. Specifically, our model captures the correlations that exist between bandpass divisively-normalized image samples, as a function of the spatial separations and the relative orientations between the samples and the bandpass filter tuning. We demonstrate that for any image, the bivariate NSS model correlation can be expressed using 6 parameters, per spatial orientation. We also study the bivariate NSS of distorted images. We find that our model is capable of representing the correlations between distorted image samples. The observed changes in the bivariate NSS model parameters when distortions are introduced are found to be systematic, suggesting their usefulness in image distortion analysis and future image quality models.

B. Background

Here, we review existing relevant models of the second order statistics of natural scenes. Early on, Simoncelli [13] and Liu and Moulin [15] observed that the coefficients of orthonormal wavelet (i.e; bandpass) decompositions of natural images tend to be much less spatially correlated than the source images, yet they exhibit strong intra and inter scale dependencies between bands [16]. These observations formed the basis of an image texture model [5], where a set of parametric constraints imposed on pairs of complex wavelet coefficients occupying adjacent spatial locations, orientations and scales were used to represent and synthesize textures. Po and Do [17] developed a natural image model using a hidden Markov tree, a Gaussian mixture model and two dimensional contourlet features that capture interlocation, interscale and interdirection dependencies. Mumford and Gidas [18] proposed an infinitely divisible statistical bandpass image model that assumes natural segmentations of images into

Manuscript received April 25, 2017; revised October 1, 2017 and January 6, 2018; accepted March 13, 2018. Date of publication March 21, 2018; date of current version April 6, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Keigo Hirakawa. *(Corresponding author: Zeina Sinno.)*

The authors are with The University of Texas at Austin, Austin, TX 78712 USA (e-mail: zeina@utexas.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2018.2817740

high-information objects, cast against an ergodic field of low-information regions. However, their model does not capture the two-dimensional dependencies that occur within bandpass images. Lee *et al.* [14] found that the power law dominates the short spatial covariance function of pairs of bandpass image samples obeying a reciprocal power law over short distances. Su *et al.* [1] developed a closed-form model of the correlation between adjacent oriented bandpass image samples, and used it to develop a bivariate generalized Gaussian distribution model bandpass images and range maps. The authors found it to be useful for tasks such as color depth and range modeling [19] and stereopair quality evaluation [20]. In [21] and [22], we extended their work by studying the bivariate distributions of the responses of horizontally related, oriented bandpass image samples separated by distances of up to 10 pixels. Here, we further generalize this work, by diversifying the model across spatial orientations and by extending the studied distances to at least 25 pixels (up to 35 pixels for some of the cardinal orientations) and by studying the behavior of the model when the images are subject to various types of distortions.

The remainder of the paper is organized as follows; in Section II we review $1/f$ image models (which is a key concept for understanding our model). In Section III we present the details of our model, then in Section IV we study the model in the context of image distortions. Source code and supplementary information regarding the model can be found at: http://live.ece.utexas.edu/research/bivariateNSS/index.html.

II. $1/f$ Processes

It is well known that the power spectra of natural photographic images tend to follow a reciprocal power law [23]:

$$S(f) \propto \frac{k}{f^{\alpha}}, \tag{1}$$

where $\alpha > 0$ determines the rate of spectral fall-off of the process.

Other phenomena that can be described by this law include the extreme case of white processes ($\alpha = 0$) which exhibit no correlation over time or space, and random walks (e.g., Brownian motion where $\alpha = 2$, which is the integral of white noise). Johnson [24] first observed a so called "$1/f$" phenomenon while studying shot noise in vacuum tubes. Processes that can be accurately described as "$1/f$" arise in such widely-varying disciplines as biological evolution [25], animal population studies [26], economics [27], personal growth and development [28], and musical loudness and pitch [29], among many others. The wide range of occurrences of the $1/f$ phenomenon may be attributed to deep natural laws that reflect the self-similarities of certain signal measurements over scales and the behavior of equilibrium systems. Formal mathematical frameworks such as fractional Brownian motion models [30], [31], fractals [32], and iterated function systems [33] have been deeply developed, yet the physical origins of $1/f$ phenomena are often poorly understood. For example, although images of natural scenes are enormously diverse, their power spectra can be reliably described as $1/f$ [3], [23], [34], reflecting statistical regularities underlying their correlation structure.

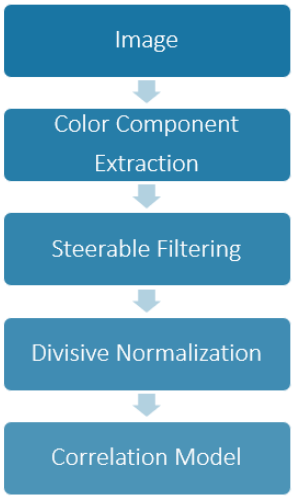


Fig. 1. Image pre-processing used in the NSS correlation model.

Here we are primarily interested in the $1/f$ image model in regards to its implications regarding the correlation structure of bandpass natural images. Our interest in this topic is motivated by the successes that have been obtained on perception-driven image analysis problems using spatial NSS models, and since these solutions might be furthered by expanding these models. This may also lead to insights on how natural correlations may drive spatial interactions between visual cortical neurons [3], [35]–[37]. Keshner [38] derived models of the stationary autocorrelation functions of one-dimensional $1/f$ processes, arriving at a power law of reciprocal separation. In the following, we develop a similar expression for the peak correlation between bandpass image samples, using a stabilized reciprocal power law.

III. Normalized Bandpass Image Correlation Model

Here we present the details of our model beginning with the preprocessing steps of bandpass decomposition and divisive normalization. A flow diagram of the involved processing is shown in Fig. 1. Along the way, we demonstrate the various processing steps used in the model using high quality images from the pristine subset of LIVE Image Quality Assessment database [39].

A. Steerable Filters

The NSS model that we use and develop is based on luminance images that have been subjected to bandpass processing. While the model appears to hold over a wide range of bandpass operations (Gabor, wavelet, etc.), we use steerable filters [40] in our simulations, owing to their simple, easily manipulated form, their invariance to content translations, and their good fit as a frequently used model of bandpass simple cells in primary visual cortex. A steerable filter at a given frequency tuning orientation θ_1 is defined by:

$$F_{\theta_1}(\boldsymbol{x}) = cos(\theta_1)F_x(\boldsymbol{x}) + sin(\theta_1)F_y(\boldsymbol{x}), \tag{2}$$

where $\boldsymbol{x} = (x, y)$, and F_x and F_y are the gradient components of the two-dimensional unit-energy bivariate isotropic gaussian

function:

$$G(\boldsymbol{x}) = \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}}, \tag{3}$$

having scale parameter σ. Steerable filter based decompositions, such as steerable pyramids [41] yield substantially spatially decorrelated responses when applied to high-quality photographic images.

Modifying the scale parameters σ of the bivariate gaussian derivative functions (F_x and F_y) enables the construction of a multi-scale bandpass image decomposition broadly resembling the responses of populations of simple cells in cortical area V1. Other filter models could be used equally well to obtain bandpass orientation and radial frequency responses, such as Gabor filters, but the steerable filters present advantages of simple definition and efficient computation. The radial frequency bandwidth of the steerable filter (1) is fairly narrow (about 2.6 octaves). In the following development and testing of the bivariate correlation model, each analyzed image is passed through steerable filters of scales $\sigma \in \{1, 2, 3, \ldots, 15\}$ and over 15 frequency tuning orientations $\theta_1 \in [0, \pi/15, 2\pi/15, \ldots, \pi]$, yielding 225 bandpass responses. The bandpass images were computed on all 29 pristine images from the LIVE Image Quality Assessment database, yielding a total of 6525 bandpass filtered image responses.

B. Divisive Normalization

Divisive normalization was then applied on all of the steerable filter responses. When applied to naturalistic photographic images that have been bandpass filtered, normalization by the energy of the local signal has been observed to gaussianize and further decorrelate the image data [2], [13]. The divisive normalization model used here is:

$$u_j(\boldsymbol{x}) = \frac{w_j(\boldsymbol{x})}{\sqrt{t + \sum_{\boldsymbol{y}} g(j(\boldsymbol{y}), w_j(\boldsymbol{y}))^2}}, \tag{4}$$

where w_j are the steerable filter responses for filters indexed by j, u are the coefficients obtained after divisive normalization, and $t = 10^{-4}$ is a stabilizing saturation constant. The weighted sum in the denominator is computed over a spatial neighborhood of pixels from the same sub-band, where $g(x_i, y_i)$ is a circularly symmetric Gaussian function having unit volume. To match the increase in scale applied at the steerable filtering step (translated by increasing σ), the variance of $g(x_i, y_i)$ is also increased linearly as a function of σ. Furthermore, we note that this step is also a good functional model of the nonlinear adaptive gain control of V1 neuronal responses in the visual cortex [35]. Divisive normalization causes the subband statistics of good quality natural images to become strongly Gaussianized. If the images are distorted, then the bandpass distribution tends away from Gaussian [11].

C. Bivariate Density Model

Following Su *et al.* [19], we use a multivariate generalized Gaussian distribution (MGGD) to model the joint histogram of a pair of divisively normalized bandpass image samples located at different spatial (pixel) locations. Methods for estimating the parameters of MGGD model fits to multi-dimensional image histograms are studied by Pascal *et al.* [42]. The probability density function of the MGGD is:

$$p(\boldsymbol{x}; \boldsymbol{M}, \eta, s) = \frac{1}{|\boldsymbol{M}|^{\frac{1}{2}}} g_{\eta,s}(\boldsymbol{x}^{\boldsymbol{T}} \boldsymbol{M}^{-\mathbf{1}} \boldsymbol{x}), \tag{5}$$

where $\boldsymbol{x} \in \mathbb{R}^N$, $\boldsymbol{M}$ is an $N \times N$ scatter matrix, η and s are scale and shape parameters respectively, and $g_{\eta,s}(.)$ is the density generator:

$$g_{\eta,s}(y) = \frac{s\Gamma(\frac{N}{2})}{(2^{\frac{1}{s}}\pi\eta)^{\frac{N}{2}}\Gamma(\frac{N}{2s})} e^{-\frac{1}{2}(\frac{y}{\eta})^s}, \tag{6}$$

where Γ is the digamma function and $y \in \mathbb{R}^+$. Note that when $s = 0.5$, (6) becomes a multivariate Laplacian density function, and when $s = 1$, it becomes multivariate Gaussian density. Here we fix $s = 1$, where η controls the spread of the density function.

While pairs of Gaussian random variables are not necessarily jointly Gaussian, pairs of image samples that have been subjected to bandpass processing followed by divisive normalization are observed to be reliably jointly Gaussian. The reason for the Gaussianity of images processed in this perceptually relevant manner remains elusive. It cannot be explained as a consequence of the Central Limit Theorem (CLT), since the only additive process (linear filtering) is on strongly correlated, raw image samples rather-than on uncorrelated or weakly correlated variances, as required by the CLT. Moreover, the outcome of the linear filtering is decidedly non-Gaussian, and instead is distributed with much heavier tails, typically described as leptokurtic generalized Gaussian [2]. The shape of these empirical non-Gaussian "sparsity" densities is typically attributed to the imaging projection of a world that is smooth nearly everywhere (yielding heavily massed bandpass samples near or at zero), except where (blurred) singularities occur (resulting in large responses defining the heavy tails). Gaussianity finally arises as a consequence of a process of local divisive normalization by neighboring bandpass image energy [2], [43]. While this ultimate Gaussianity remains unexplained, there may be connections with theoretical processes defined as quotients of highly correlated quantities, such as the Fisz transform [44], [45].

The bivariate empirical histograms of the sub-band coefficients of natural images are thus modeled here as following a bivariate generalized Gaussian distribution (BGGD), by setting $N = 2$. This also presumes that the images have not been distorted, which may change their statistics. In all of the following, the parameters of the BGDD were estimated using the efficient maximum likelihood estimation method of [42]. We systematically applied this modeling process to all of the bandpass normalized images.

To remove any undesirable border filter effects, we cropped 10 pixels from each image's four borders, defined a window at a fixed position within the cropped image (Window 1) and another sliding window of the same dimensions (Window 2). Denote the distance between the center of the two windows of bandpass, normalized image samples of interest by d, and

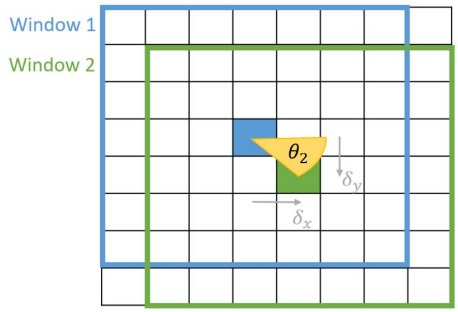


Fig. 2. An illustration of an image after the divisive normalization and steerable filtering (of fixed σ and θ_1 values) are applied, with the two sliding windows, and how θ_2 is computed.

the angle between them by θ_2, as illustrated in Figure 2. Next, define the relative angle $\theta_2 - \theta_1$, where θ_1 is the sub-band tuning of the bandpass filter orientation relative to the horizontal axis. The bivariate histogram takes predictable shapes. For example, when the relative angle $\theta_2 - \theta_1 = 0$, the bivariate joint histogram takes a highly eccentric elliptical shape indicating a strong degree of a correlation, whereas when the relative angle is increased, the bivariate histogram becomes more circular. Figure 3 plots bivariate histograms as intensity for the case of $d = 1$ and $\theta_2 = \pi/2$. As we discuss further, we observe similar histogram shape trends for longer separations d and for all other spatial angles θ_2.

The bivariate model (5) is a closed form, except for the elements of the scatter matrix $\boldsymbol{M}$. The scatter matrix defines the covariance matrix of the bivariate model. To complete the closed form model, we studied the Pearson correlation function between the two windows. We obtain the correlation as the gradient of the covariance of the two entities and the product of their standard deviations. The two windows were separated by horizontal and vertical separations δ_x and δ_y which we varied over the integer range from 1 and 25, i.e. distances of $\sqrt{\delta_x^2 + \delta_y^2}$ at spatial orientations $\theta_2 = \arctan(\frac{\delta_y}{\delta_x})$ (relative to the horizontal axis). We limited the range of θ_2 to $[0, \pi[$ since the quantities being measured are symmetrically defined and are π periodic.

The tuning orientation θ_1 is the frequency tuning orientation of the steerable filter. We used a discrete set of 15 sub-band orientations $\{0, \frac{\pi}{15}, \frac{2\pi}{15}, \ldots, \frac{14\pi}{15}\}$ to build our model.

The correlation function model expresses a periodic behavior in the relative angle $\theta_2 - \theta_1$, which can be well modeled as:

$$\rho(d, \sigma, \theta_2) = A(d, \sigma, \theta_2) cos(2(\theta_2 - \theta_1)) + c(d, \sigma, \theta_2) \quad (7)$$

where $A(d, \sigma, \theta_2)$ is the amplitude, $c(d, \sigma, \theta_2)$ is an offset, d is the spatial separation between the target pixels, σ is the steerable filter spread parameter, and θ_2 is as before. Generally, the shapes of ρ, A, and c vary in a consistent way with d, σ and θ_2, as we shall see.

Figure 4 plots the average correlation function of several processed images from the set of LIVE reference images, as a function of $\theta_2 - \theta_1$, over 4 scales for $\theta_2 = \pi/2\ rad$ and $d = 1$. From this plot, it may be observed that the maximum correlation value $P = \max(\rho)$ that is attained, occurs (as expected) when $\theta_2 - \theta_1 = 0$, falling monotonically from this maximum value as the absolute relative angle is increased to $\pi/2$. Figure 4 also shows that the correlation increases with the scale factor σ, which we have observed over all studied spatial orientations θ_2 and spatial separations d. This is to be expected, since as σ is increased, the filter bandwidths decrease, which tends to increase in-band correlations.

As the spatial separation d is increased, the correlation also drops, as shown in Fig. 5, where the empirical correlations are plotted for a fixed scale σ and spatial orientations θ_2, over several values of the spatial separation d.

As a further illustration of the correlation function's behavior, Figure 6 plots the correlations of adjacent samples measured at the same scale ($\sigma = 2$) and spatial separations but different spatial orientations $\theta_2 \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$. Note that the sample separation takes two values: $d = 1$ for $\theta_2 \in \{0, \frac{\pi}{2}\}$ and $d = \sqrt{2}$ for $\theta_2 \in \{\frac{\pi}{4}, \frac{3\pi}{4}\}$. From the plot, it may be seen that horizontally and vertically related pixels ($\theta_2 = \pi/2\ rad$ and $\theta_2 = 0\ rad$) are more correlated than diagonally related pixels ($\theta_2 = \pi/4\ rad$ and $\theta_2 = 3\pi/4\ rad$), which is also expected owing to the different spatial separations. However, the correlation also likely increases along the cardinal directions because of the preponderance of horizontal and vertical structures in real-world images [46].

In order to better understand and to complete our model of the correlation function ρ in (7), we also model the amplitude and offset functions A and c. To do so, we define the peak correlation function:

$$P = \max(\rho) = A + c. \quad (8)$$

wherein we may rewrite (7) as:

$$\rho(d, \sigma, \theta_2) = A(d, \sigma, \theta_2) cos(2(\theta_2 - \theta_1)) + [P(d, \sigma, \theta_2) - A(d, \sigma, \theta_2)] \quad (9)$$

We did not impose any constraints on the values of A and P when fitting ρ. We have observed the values of A to be positive except in a few instances where the correlation is very small (at large spatial separations) or large and flat (at small separations and large scales). In those cases, A took slightly negative values (10^{-3}).

As mentioned earlier, Lee *et al.* [14] systematically observed that the sample covariances of bandpass image pixels follow an approximate reciprocal power law, of the form $\frac{1}{|d|^b}$, which, like white processes, cannot be realized. Similarly, Keshner [38] remarks on the fact that the nonstationary autocorrelation function of $1/f$ processes take a reciprocal form, and that a practical stationary model might be obtained by modifying the autocorrelation model near the origin. Here, we take a different approach, whereby we model the peak correlation function as having a general version of the form $\frac{1}{|d|^\beta + 1}$.

Figure 7 plots the empirical peak correlation function P against the sample separation d for a few values of σ and θ_2. As expected, the measured correlations decrease rapidly from a peak value of 1 as the spatial separation d increases; which is natural since one should expect reduced correlations between

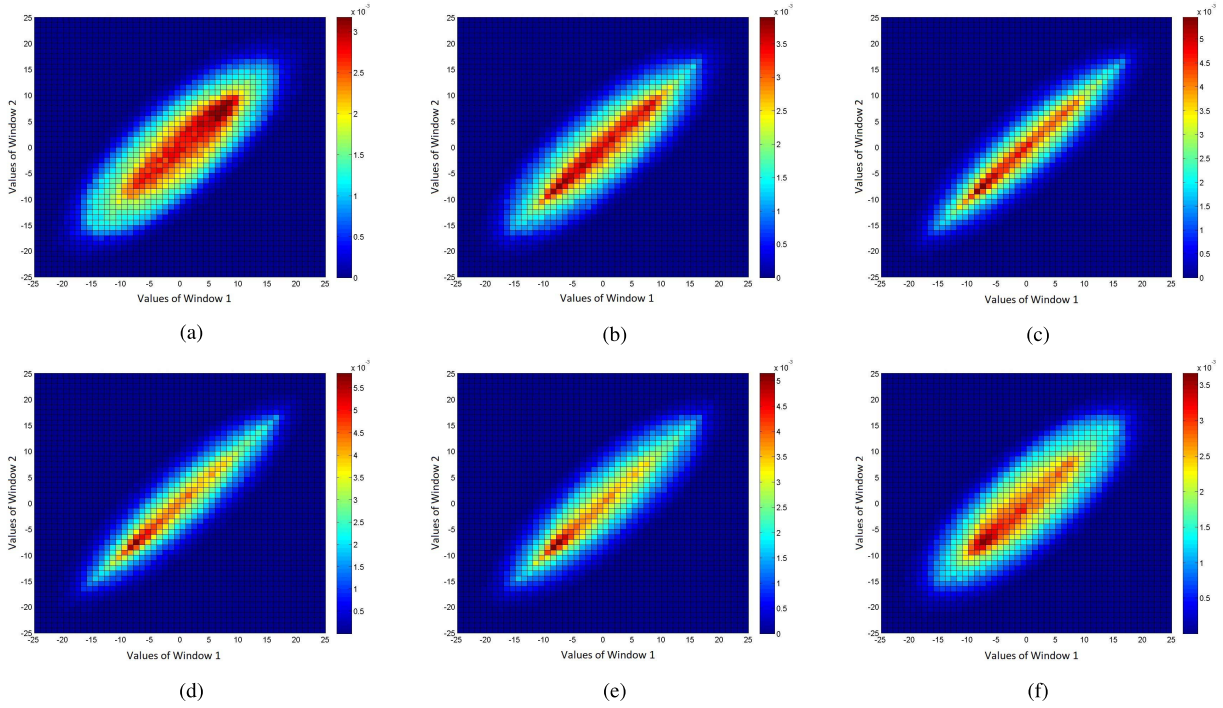


Fig. 3. Bivariate joint histograms of a steerable filter response at distance $d = 1$, scale $\sigma = 2$, tuned to spatial orientation $\theta_2 = \pi/2$ for various spatial angular differences θ_1. Each plot presents the probability of the values that two pixels separated by d and θ_2 will take. (a) $\theta_1 = 0$, and $\theta_2 - \theta_1 = -1.57$. (b) $\theta_1 = \frac{3}{15\pi}$, and $\theta_2 - \theta_1 = -0.94$. (c) $\theta_1 = \frac{6}{15\pi}$, and $\theta_2 - \theta_1 = -0.31$. (d) $\theta_1 = \frac{9}{15\pi}$, and $\theta_2 - \theta_1 = 0.31$. (e) $\theta_1 = \frac{11}{15\pi}$, and $\theta_2 - \theta_1 = 0.73$. (f) $\theta_1 = \frac{14}{15\pi}$, and $\theta_2 - \theta_1 = 1.36$.

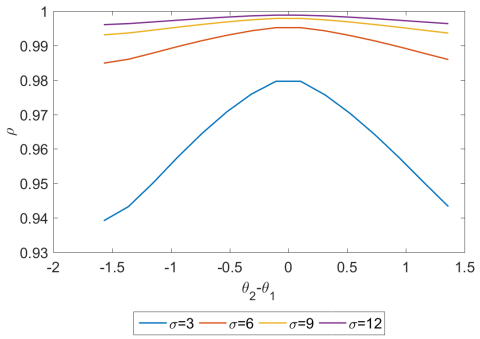


Fig. 4. Average correlation function of the luminance components of natural images plotted against relative angle $\theta_2 - \theta_1$, for $\theta_2 = \pi/2\ rad$, $d = 1$, and $\sigma = 3,\ 6,\ 9$, and 12.

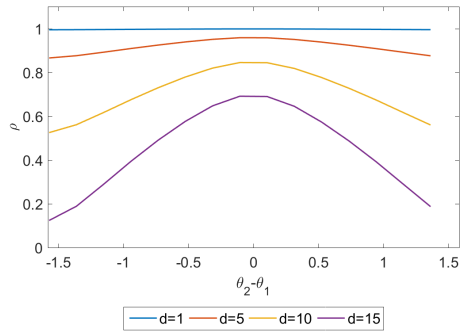


Fig. 5. Average correlation function of the luminance components of natural images plotted against relative angle $\theta_2 - \theta_1$, for $\theta_2 = \pi/2\ rad$, $\sigma = 10$ and $d = 1,\ 5,\ 10$, and 15.

pixels as the spatial separation increases. There is a slight observed undershoot, especially for small σ values, which is likely a consequence of unsmoothness of the applied filter, but this is small and difficult to model, hence we neglect this minor behavior.

The general form of our stabilized peak correlation model is as follows: given a fixed spatial orientation θ_2 and a scale σ, define

$$\hat{P}(d, \sigma, \theta_2) = \frac{1}{(\frac{d}{\alpha_0(\theta_2) * \sigma})^{\beta_0} + 1} \tag{10}$$

where $\{\alpha_0, \beta_0\}$ are parameters that control the shape and fall-off of the peak correlation function, and which depend on the spatial orientation θ_2.

We discuss the validation and application of our model (10) further along, but first we will look at the other function comprising the correlation model (9).

Figure 8 plots the amplitude function $A(d, \sigma, \theta_2)$ against d for few scales σ and spatial orientations θ_2. The graph of A rises from the value 0 at $d = 0$, then decreases with increasing separation. Given the similarity of the graph of A to the difference of two functions of the same form but different scales, and the close relationship between A and P, we model A as the difference of two functions of the form (10):

$$\hat{A}(d, \sigma, \theta_2) = \frac{1}{(\frac{d}{\alpha_1(\theta_2) * \sigma})^{\beta_1(\theta_2)} + 1} - \frac{1}{(\frac{d}{\alpha_2(\theta_2) * \sigma})^{\beta_2(\theta_2)} + 1} \tag{11}$$

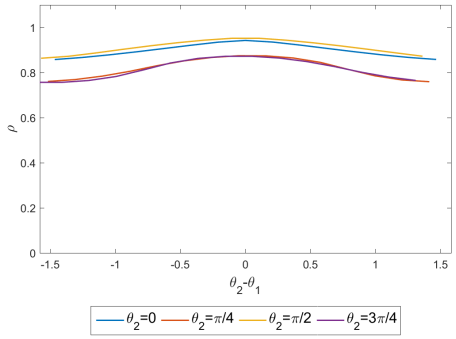


Fig. 6. Average correlation function of the luminance components of bandpass, divisively normalized natural images for the case of adjacent pixels (horizontal, vertical, diagonal) plotted against relative angle, for $\sigma = 2$ for $\theta_2 - \theta_1$ for $\theta_2 = 0$, $\pi/4$, $\pi/2$, and $3\pi/4$.

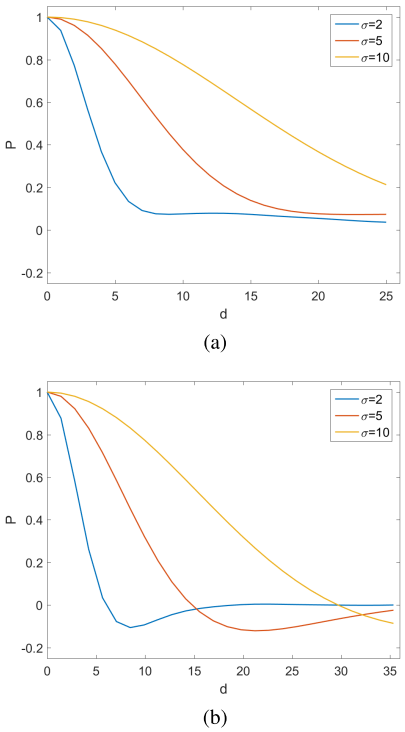


Fig. 7. Peak function $P(d, \sigma, \theta_2)$ plotted against pixel separation d for $\sigma = 2$, 5, and 10 for (a) $\theta_2 = 0$ and (b) $\theta_2 = \frac{\pi}{4}$ (rad).

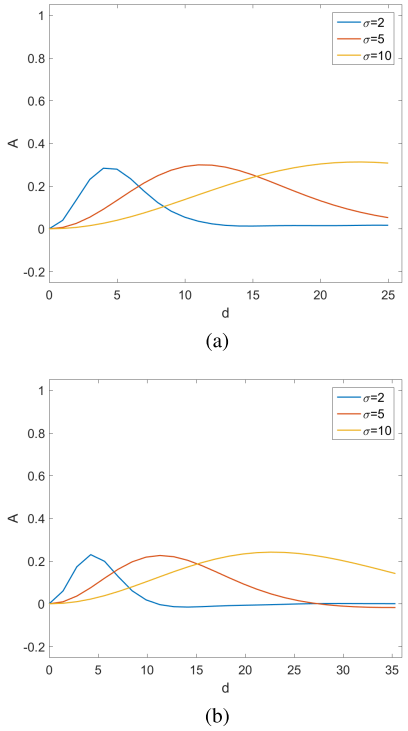


Fig. 8. Amplitude function $A(d, \sigma, \theta_2)$ plotted against pixel separation d for $\sigma = 2$, 5, and 10 for (a) $\theta_2 = 0$ and (b) $\theta_2 = \frac{\pi}{4}$ (rad).

TABLE I

OPTIMAL VALUES OF $\alpha_0, \beta_0, b_0, \alpha_1, \beta_1, b_1, \alpha_2, \beta_2$, AND b_2 FOR THE 8 MOST FREQUENTLY OCCURRING VALUES OF θ_2 ON THE LIVE IQA REFERENCE LUMINANCE IMAGES

θ_2	α_0	β_0	α_1	β_1	α_2	β_2
0.000	1.623	2.628	3.197	3.336	2.092	2.274
0.464	1.519	3.325	3.307	3.501	2.383	2.204
0.785	1.528	3.771	3.302	3.450	2.456	2.252
1.107	1.724	3.175	3.357	3.542	2.370	2.173
1.571	2.210	2.181	3.267	3.171	1.978	2.313
2.034	1.718	3.207	3.358	3.545	2.371	2.161
2.356	1.522	3.767	3.293	3.456	2.448	2.253
2.678	1.506	3.351	3.288	3.471	2.382	2.203

where $\{\alpha_1, \beta_1, \alpha_2, \beta_2\}$ are parameters that are functions of θ_2 that control the shape of A.

Our goal next is then to find, for a fixed spatial orientation θ_2, the values of the parameters $\{\alpha_0, \beta_0\}$ that produce the best fit to (10), and the parameters $\{\alpha_1, \beta_1, \alpha_2, \beta_2\}$, that yield the best fit to (11), in the least mean squared error sense. We form two optimization systems for P and A that account for scale to find those optimal values. The optimization systems minimize the summed squared errors of the peak and amplitude. To accomplish this, we apply unconstrained nonlinear regression using the quasi newton method [47]. We restrict our modeling of the correlation to a span of dimensions 25×25 so that $d \in [0, \sqrt{1250}]$, since the peak correlation becomes negligible if d is increased further. The four functions $P(d, \sigma, \theta_2)$, $A(d, \sigma, \theta_2)$, $\hat{P}(d, \sigma, \theta_2)$, and $\hat{A}(d, \sigma, \theta_2)$ form vectors of size $m \times 1$, where m is the number of occurrences of θ_2 inside the span of interest. Denote by D the set of distances for a given spatial orientation θ_2. For the case $\theta_2 = 0$ or $\pi/2$, $D = \{0, 1, 2, 3, \ldots, 24, 25\}$. For the case $\theta_2 = \pi/4$ or $3\pi/4$, $D = \{0, \sqrt{2}, \sqrt{8}, \sqrt{18}, \ldots, \sqrt{1152}, \sqrt{1250}\}$. Our optimization systems are then expressed as:

$$\min_{\alpha_0, b_0} \quad \sum_{d \in D} \sum_{\sigma=2}^{15} (P(d, \sigma, \theta_2) - \hat{P}(d, \sigma, \theta_2))^2 \qquad (12)$$

and

$$\min_{\alpha_1, b_1, \alpha_2, \beta_2, b_2} \quad \sum_{d \in D} \sum_{\sigma=2}^{15} (A(d, \sigma, \theta_2) - \hat{A}(d, \sigma, \theta_2))^2 \qquad (13)$$

Table I gives the optimal parameters yielding the best average correlation fit to (12) and (13) over all of the (luminance) images in the LIVE reference image set over the 8 most frequently occurring spatial orientations θ_2. It may be observed that the fitting parameters fall within narrow

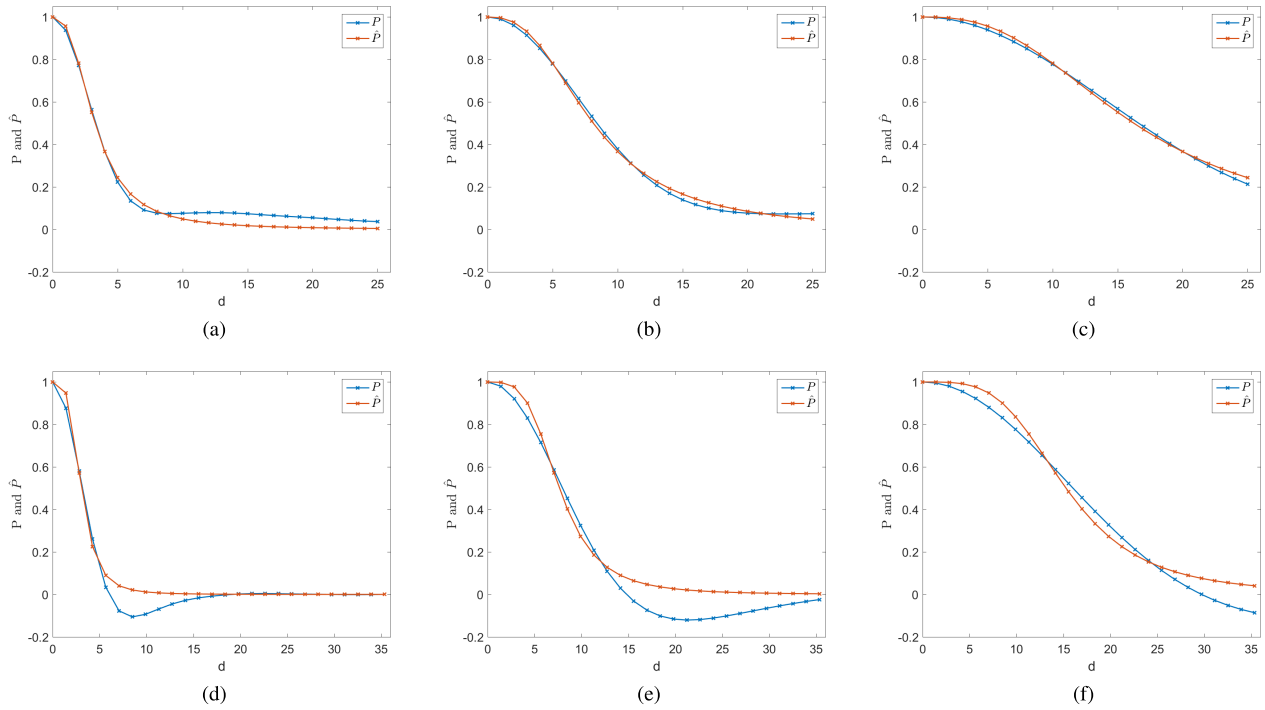


Fig. 9. Examples of best-fitting peak correlation model $\hat{P}$ to the peak correlation P of the average empirical correlation P. (a) $\sigma = 2, \theta_2 = 0$. (b) $\sigma = 5, \theta_2 = 0$. (c) $\sigma = 10, \theta_2 = 0$. (d) $\sigma = 2, \theta_2 = \pi/4$. (e) $\sigma = 5, \theta_2 = \pi/4$. (f) $\sigma = 5, \theta_2 = \pi/4$.

ranges, the exceptions being the peak correlation parameters (α_0, β_0) which deviate a little more along the cardinal orientations, and to a lesser degree, along the diagonal orientations. This is not unexpected given the well-known prevalence of horizontal, diagonal, and vertical oriented structures in the visual environment [46]. What is perhaps surprising is the high degree of uniformity of the other parameters against orientation, particularly those of the amplitude function (11). We also computed these parameters over the larger set of values $\theta_2 = \{0.000, 0.785, 1.571, 2.356, 0.464, 1.107, 2.034, 2.678, 0.322, 0.588, 0.983, 1.249, 1.893, 2.159, 2.554, 2.820, 0.245, 0.644, 0.927, 1.326, 1.816, 2.214, 2.498, 2.897\}$. These values occur at least 5 times in the area of interest. Values of θ_2 where there was insufficient data (viz., pairs of pixels at those orientations) are left out to conduct the optimization. We computed the optimal parameters $\alpha_0, \beta_0, \alpha_1, \beta_1, \alpha_2$, and β_2 for this set of θ_2 values for each $\sigma \in \{1, 2, 3, \ldots, 15\}$. Since this is a sizeable amount of tabulated data, we make it available at the following link: http://live.ece.utexas.edu/research/bivariateNSS/index.html.

D. Model Validation

Next, we validate our model by examining the closeness of fit of the models $\hat{P}$, $\hat{A}$ and $\hat{\rho}$ to the empirical functions P, A and ρ.

1) Validation of A and P: We computed the mean squared error (MSE) between the reconstructed peak and amplitude correlation functions $\hat{P}$ and $\hat{A}$, relative to the empirical average functions P and A that were computed and measured, respectively, on the LIVE Image Quality Assessment Database [39] luminance images across integer scales $\sigma \in \{2, 3, \ldots, 15\}$. The MSE between P and $\hat{P}$ for a fixed scale σ and orientation θ_2 is defined as:

$$MSE_P = \sum_{d \in D} \frac{(P(d, \sigma, \theta_2) - \hat{P}(d, \sigma, \theta_2))^2}{|D|}, \tag{14}$$

where $|D|$ is the cardinality of D. Similarly, for a fixed scale σ and orientation θ_2, the MSE of between A and $\hat{A}$ is defined as:

$$MSE_A = \sum_{d \in D} \frac{(A(d, \sigma, \theta_2) - \hat{A}(d, \sigma, \theta_2))^2}{|D|}. \tag{15}$$

The largest errors between P and $\hat{P}$ and A and $\hat{A}$ over all pairs (σ, θ_2) were on the order of 10^{-3}. The results for the considered (σ, θ_2) pairs can be found at the same link as above. Examples of the empirical functions P and A are shown in Fig. 9 and Fig. 10, which visually illustrate the goodness of our model in capturing P and A. It is worth remarking that the results obtained by finding the best-fitting $\hat{P}$ and $\hat{A}$ on the average empirical correlation data, were as good as those obtained by finding the best fits on the empirical correlations from each of the naturalistic images in the LIVE Image Quality Assessment [39] database and their corresponding best fits $\hat{P}$ and $\hat{A}$.

Furthermore, for each fixed scale σ and spatial orientation θ_2, we computed the χ^2 test statistic:

$$\chi^2_A = \sum_{d \in D} \sum_{i=1}^{29} \frac{(A_i(d, \sigma, \theta_2) - \hat{A}(d, \sigma, \theta_2))^2}{\hat{A}(d, \sigma, \theta_2)} \tag{16}$$

where $A_i(d, \sigma, \theta_2)$ is the empirical amplitude function of the i^{th} naturalistic image from the LIVE Image Quality

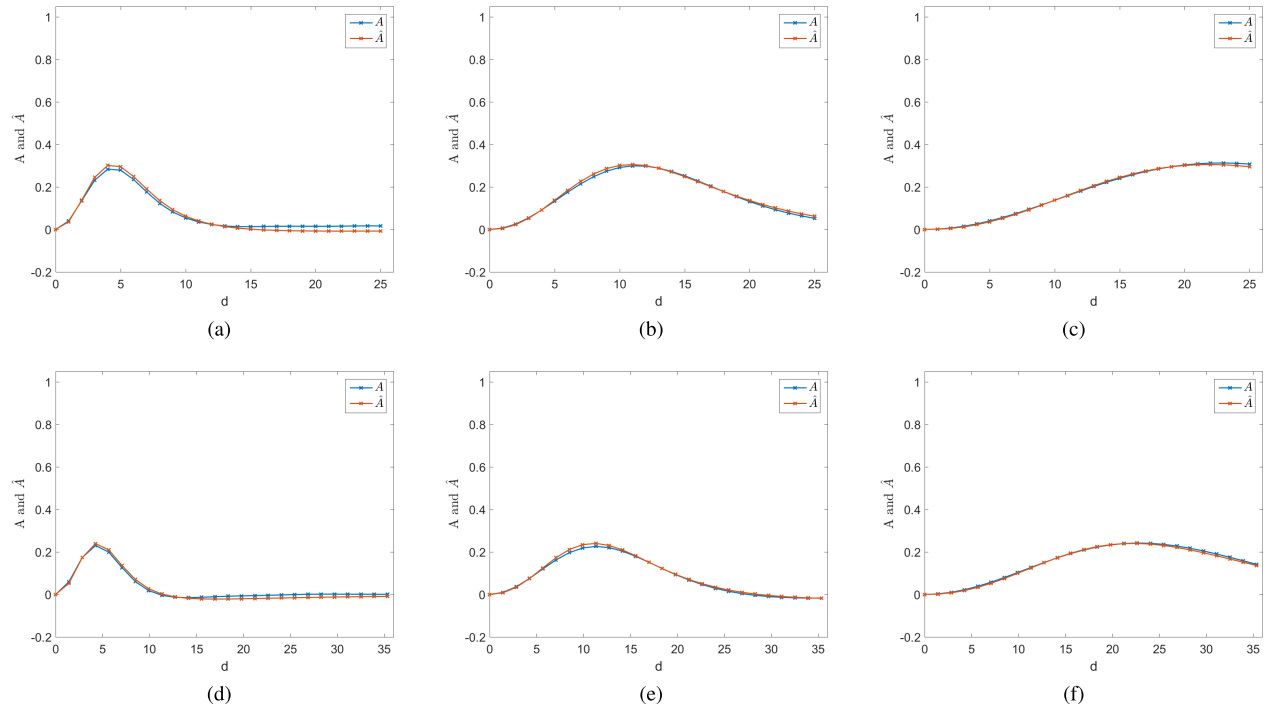


Fig. 10. Examples of best-fitting amplitude correlation model $\hat{A}$ to the peak correlation A of the average empirical correlation A. (a) $\sigma = 2, \theta_2 = 0$. (b) $\sigma = 5, \theta_2 = 0$. (c) $\sigma = 10, \theta_2 = 0$. (d) $\sigma = 2, \theta_2 = \pi/4$. (e) $\sigma = 5, \theta_2 = \pi/4$. (f) $\sigma = 5, \theta_2 = \pi/4$.

TABLE II

χ_P^2 RESULTS FOR THE 8 MOST FREQUENTLY OCCURRING θ_2 ON THE LIVE IMAGE QUALITY ASSESSMENT REFERENCE LUMINANCE IMAGES

	$\theta_2 = 0$	$\theta_2 = 0.464$	$\theta_2 = 0.785$	$\theta_2 = 1.107$	$\theta_2 = 1.571$	$\theta_2 = 2.034$	$\theta_2 = 2.356$	$\theta_2 = 2.678$
σ =1	667.17	177.88	2063.83	166.83	321.36	141.25	2051.45	157.31
σ =2	309.70	126.07	670.77	105.02	162.62	112.74	657.30	115.90
σ =3	170.18	125.16	455.59	124.99	101.75	123.38	382.69	116.03
σ =4	106.40	108.16	476.94	93.09	73.82	92.76	465.55	116.56
σ =5	79.20	86.79	521.75	57.49	60.19	59.49	549.10	102.24
σ =6	65.15	64.17	490.16	33.70	50.72	35.21	503.69	78.63
σ =7	54.00	45.57	394.72	20.73	41.62	21.76	378.95	54.75
σ =8	43.89	31.92	277.88	13.49	32.88	14.37	247.53	35.99
σ =9	35.02	23.14	180.06	9.62	25.59	10.24	149.55	23.99
σ =10	27.54	17.45	110.65	7.09	19.50	7.69	85.80	16.64
σ =11	21.50	13.56	69.57	5.30	14.71	5.92	51.05	12.16
σ =12	16.71	10.59	45.95	4.06	11.13	4.63	32.55	9.22
σ =13	12.91	8.23	32.28	3.18	8.45	3.69	22.91	7.18
σ =14	10.05	6.31	24.16	2.50	6.54	2.93	17.54	5.61
σ =15	7.83	4.79	18.67	2.00	5.15	2.36	13.98	4.39

Assessment [39] for fixed σ and θ_2 values, and $\hat{A}(d, \sigma, \theta_2)$ was obtained by finding the best fit to the amplitude function of the average empirical correlation. Likewise the χ^2 statistic for the peak correlation was also computed:

$$\chi_P^2 = \sum_{d \in D} \sum_{i=1}^{29} \frac{(P_i(d, \sigma, \theta_2) - \hat{P}(d, \sigma, \theta_2))^2}{\hat{P}(d, \sigma, \theta_2)}, \qquad (17)$$

where $P_i(d, \sigma, \theta_2)$ is the empirical peak correlation function of the i^{th} naturalistic image from the LIVE Image Quality Assessment [39] for σ and θ_2 fixed, and $\hat{P}(d, \sigma, \theta_2)$ was obtained by finding the best fit to the peak of the average correlation.

The results of the χ_P^2 and χ_A^2 tests for the 8 most frequently occurring values of θ_2 are presented in Tables II and III respectively. The values are in general small (on the order of 10) except at the smallest scales of σ. This is not unexpected, as highly localized (less smoothed) measurements of the correlation will be less certain. However, we have observed the functional fits to be reasonably good, even when $\sigma = 1$. The somewhat less consistent behavior of the results when $\sigma = 1$ is likely due to two reasons: first, the spatial implementation of the steerable filters begins to become degenerate at that scale, leading to poorer localization properties than afforded by larger Gaussian envelopes, and second, the presence of high frequency noise, including quantization, present even in high-quality pictures, may affect the steerable responses as well.

TABLE III

χ^2_A RESULTS FOR THE 8 MOST FREQUENTLY OCCURRING θ_2 ON THE LIVE IMAGE QUALITY ASSESSMENT REFERENCE LUMINANCE IMAGES

	$\theta_2 = 0$	$\theta_2 = 0.464$	$\theta_2 = 0.785$	$\theta_2 = 1.107$	$\theta_2 = 1.571$	$\theta_2 = 2.034$	$\theta_2 = 2.356$	$\theta_2 = 2.678$
σ =1	-81.17	-2.93	0.05	-3.52	-206.53	-2.73	0.42	-2.45
σ =2	-89.16	1.90	1.28	0.64	522.65	1.74	2.28	0.93
σ =3	-51.07	-3.27	3.41	-6.45	20.09	-2.54	4.56	-2.85
σ =4	12.66	5.20	7.90	2.21	8.68	3.16	6.69	4.58
σ =5	8.24	5.05	9.25	5.57	6.34	4.02	8.42	5.95
σ =6	6.40	2.72	13.13	2.68	5.21	2.33	13.18	2.97
σ =7	5.48	2.33	9.17	1.99	4.60	2.17	8.20	2.38
σ =8	4.91	2.23	7.33	1.71	4.16	2.23	7.03	2.22
σ =9	4.44	2.20	6.59	1.54	3.80	2.29	6.76	2.15
σ =10	3.81	2.17	6.22	1.40	3.42	2.31	6.72	2.07
σ =11	3.28	2.11	5.85	1.28	3.13	2.26	6.58	1.97
σ =12	2.99	2.06	5.48	1.18	2.90	2.15	6.40	1.91
σ =13	2.83	1.98	5.11	1.10	2.64	2.00	6.21	1.88
σ =14	2.93	1.90	4.77	1.05	2.55	1.88	5.95	1.89
σ =15	3.17	1.84	4.48	1.06	2.56	1.79	5.69	1.91

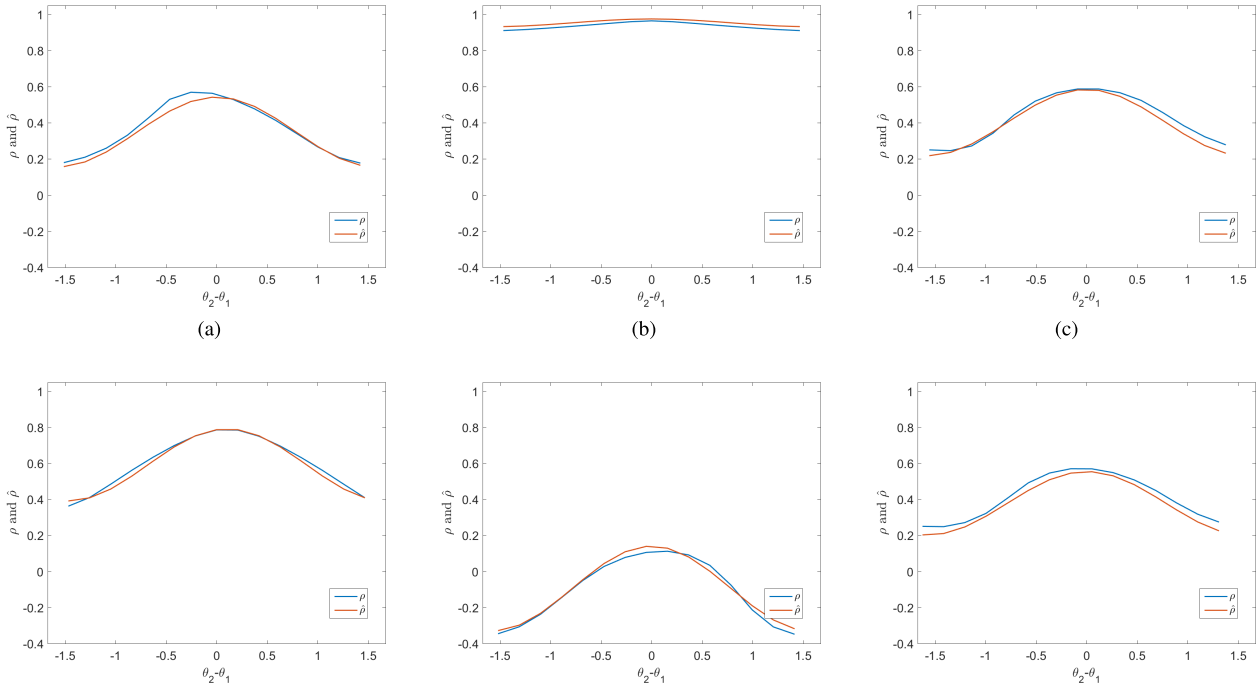


Fig. 11. Graphs of the model and empirical correlation functions ρ and $\hat{\rho}$ plotted against $\theta_2 - \theta_1$ for various values d, θ_2 and σ values. (a) $d = 20.13$, $\sigma = 14, \theta_2 = 2.678$. (b) $d = 2, \sigma = 5, \theta_2 = 0$. (c) $d = 10, \sigma = 7, \theta_2 = 2.214$ (d) $d = 12, \sigma = 10, \theta_2 = 1.571$. (e) $d = 29.70, \sigma = 12, \theta_2 = 0.7854$. (f) $d = 4.24, \sigma = 3, \theta_2 = 2.356$.

2) Validation of ρ: To validate the correlation model, ρ, we followed a similar approach. The MSE for a fixed scale σ, distance d, and orientation θ_2, is defined as:

$$MSE_\rho = \sum_{\theta_1=0,\frac{\pi}{15}}^{\frac{14\pi}{15}} \frac{(\rho(\theta_2-\theta_1) - \hat{\rho}(\theta_2-\theta_1))^2}{15}. \tag{18}$$

We computed the MSE values between the model correlation function $\hat{\rho}$ and the average empirical correlation function ρ on the luminance components of the LIVE IQA dataset. Again, the largest error was on the order of 10^{-3} for $\theta_2 = 0$. We observed similar results across other θ_2 values, which are not included here for lack of space, but could be found on http://live.ece.utexas.edu/research/bivariateNSS/index.html. Figure 11 plots the best-fitting model correlation $\hat{\rho}$ along with the empirical correlation function for a variety of randomly selected values of d, θ_2 and σ.

We also performed the χ^2 test for θ_2 and d fixed. The statistic is computed as:

$$\chi^2_\rho = \sum_{\theta_1=0,\frac{\pi}{15}}^{\frac{14\pi}{15}} \sum_{i=1}^{29} \frac{(\rho_i(d,\sigma,\theta_2) - \hat{\rho}(d,\sigma,\theta_2))^2}{\hat{\rho}(d,\sigma,\theta_2)}, \tag{19}$$

where $\rho_i(d,\sigma,\theta_2)$ is the correlation of the i^{th} naturalistic image from the LIVE Image Quality Assessment reference set for given values of d, σ and θ_2, and $\hat{\rho}(d,\sigma,\theta_2)$ is the best fit

TABLE IV

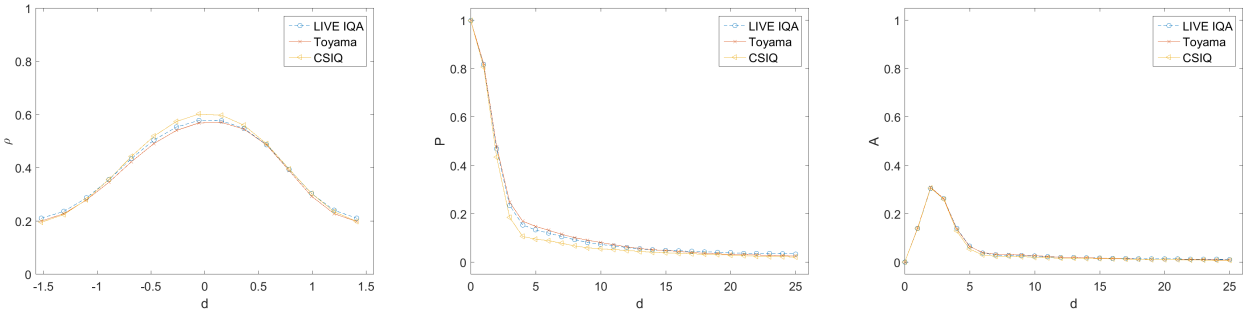
MEDIAN χ^2_ρ WITH RESPECT TO THE AVERAGE LUMINANCE CORRELATION FOR $\theta_2 = 0$ ON THE LIVE IQA REFERENCE IMAGES AS A FUNCTION OF THE SCALE PARAMETER σ

$\sigma=1$	$\sigma=2$	$\sigma=3$	$\sigma=4$	$\sigma=5$	$\sigma=6$	$\sigma=7$	$\sigma=8$	$\sigma=9$	$\sigma=10$	$\sigma=11$	$\sigma=12$	$\sigma=13$	$\sigma=14$	$\sigma=15$
159.19	116.48	10.69	0.39	0.62	0.15	0.96	2.90	2.26	2.90	2.41	2.74	2.19	3.20	1.54

TABLE V

MEDIAN χ^2_ρ WITH RESPECT TO THE AVERAGE LUMINANCE CORRELATION FOR $\theta_2 = 0$ ON THE LIVE IQA REFERENCE IMAGES AS A FUNCTION OF THE SPATIAL SEPARATION d

$d=1$	$d=2$	$d=3$	$d=4$	$d=5$	$d=6$	$d=7$	$d=8$	$d=9$	$d=10$	$d=11$	$d=12$	$d=13$	$d=14$	$d=15$	$d=16$	$d=17$	$d=18$	$d=19$	$d=20$	$d=21$	$d=22$	$d=23$	$d=24$	$d=25$
0.01	0.06	0.15	0.31	0.53	0.81	0.97	1.41	1.68	3.18	2.74	7.39	5.35	17.61	7.52	10.69	9.91	34.62	17.71	37.60	27.36	27.42	86.23	18.57	44.70



(a) (b) (c)

Fig. 12. Comparison of the behavior of the model over LIVE IQA, Toyama and CSIQ. (a) $\rho(d = 1, \sigma = 1, \theta_2 = \pi/2)$ vs $\theta_2 - \theta_1$. (b) $P(d, \sigma = 1, \theta_2 = \pi/2)$ vs d. (c) $A(d, \sigma = 1, \theta_2 = \pi/2)$ vs d.

of the average correlation. Due to the lack of space, we only present the median results for $\rho_i(d, \sigma, \theta_2)$ as a function of σ and d for the case of $\theta_2 = 0$, in Tables IV and V respectively. Results for other angles can be found at the same web link given earlier.

The very low MSE values, the low values of χ^2_ρ, and the apparent good functional fits shown in the plots validates the accuracy of our model. In a few instances, the values of χ^2_ρ took larger values, as a byproduct of numerical instability when computing (19): the appearance of small values in the denominator of (19) resulted in larger values of χ^2_ρ. However, even in those cases, we still observed excellent alignment between the empirical data and the functional fits.

We also found that the model correlations computed on the individual pristine LIVE reference luminance images yielded similar measurements of goodness of fit.

E. Validation on Other Databases

As an additional way to validate our model, we studied its behavior of on other databases; first on the CSIQ database [48] which contains 30 pristine images and second on the Toyama Database [49] which contains 14 pristine images. As depicted by the example in Fig. 12 we obtained a great overlap between the average correlation, amplitude in peak of the different databases. Also we observed small χ^2_ρ, χ^2_P and χ^2_A values between the mean case from the LIVE IQA database and the images from the other databases.

F. Scale Invariance

Several aspects in the environment are statistically self-similar, meaning that their structure is invariant over multiple scales. An observed property of natural images is the invariance of their statistics with respect to the scale at which the image is observed. For example, the power spectrum of images is invariant to scaling [50], which implies a similar correlation scale-invariance property. Also, many natural structures are scale-invariant [32]. Figure 13 plots P and A against the scaled spatial separation for several values of the scale, for the case $\theta_2 = 0$. Excellent alignment of the plots across scales is observed, in agreement with the scale-invariance property, over all θ_2 values. To conserve space, we present the results for only a few scales in Fig. 13, however we have observed invariance to also hold over all the other scales.

G. White Noise Images

Next we study how our correlation model applies to composed images of simulated white noise (i.e, random matrices). We conduct this analysis both as an experimental control and as a way to better understand the properties of our model. The relative correlation structure of bandpass filtered and normalized white noise against that of natural images is of interest. For example, while the perceptually relevant processing used in our model is known to decorrelate natural images, which are otherwise strongly correlated, instead it introduces correlation on white noise images.

However, the processed white noise images still exhibit less correlation than processed natural images, as may be observed in Fig. 15 (a). Note that both the correlation and peak correlation functions of the (bandpass, normalized) processed natural images are everywhere higher than for the processed white noise. Overall, we have found the parametric fits (and associated parameters) to natural images and white

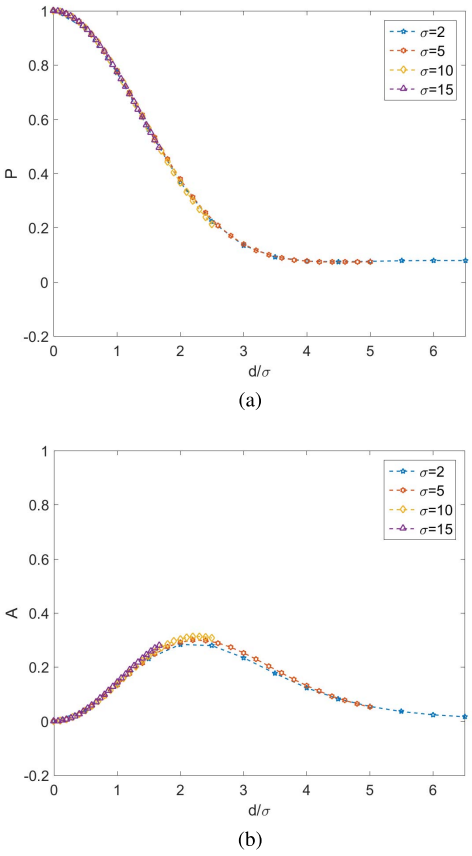


Fig. 13. Plots of (a) peak function and (b) amplitude correlation function for $\theta_2 = 0$ rad, for several values of scale σ, illustrating the scale invariance of these functions.

noise to be quite different and to obey the ordering observed in Fig. 15 (b). This serves not only to validate the unique characteristics of high-quality natural images processed in this manner (like those in the LIVE reference dataset), but also raises the question of how the model applies to distorted images, and how it might be exploited to analyze them. For example, they might be exploited to augment or improve upon existing image quality prediction models and algorithms [7].

IV. Behavior of the Model in the Presence of Distortions

Distortions lead to consistent changes in the behavior of bandpass image fits to univariate NSS models [11]. Next, we examine how our correlative model behaves in the presence of image distortion. To study this, we applied the model to both reference and distorted images taken from the LIVE IQA database [39]. The database contains images impaired by gaussian blur, JPEG compression, JPEG 2000 compression, fast fading channel noise and additive white noise. We begin by using a simple example image from the LIVE IQA database to demonstrate our observations along the way; image "Woman Hat" is shown in Fig. 16(a). While we restrict ourselves to commenting on "Woman Hat," we have observed very similar results on distorted versions of all the other LIVE IQA images.

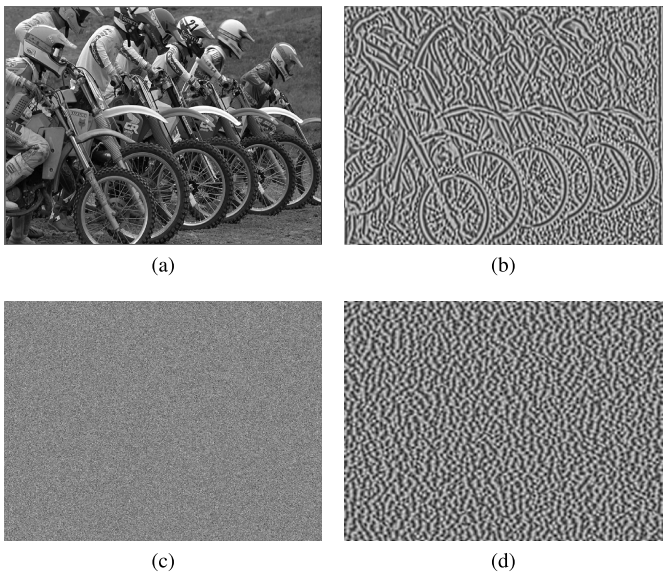


Fig. 14. A sample natural image (a) before and (b) after bandpass filtering and normalization. Similar for white noise image (c) and processed version of it (d).

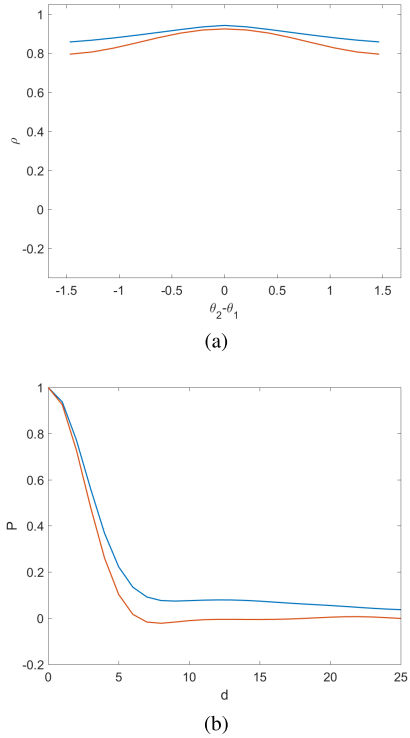
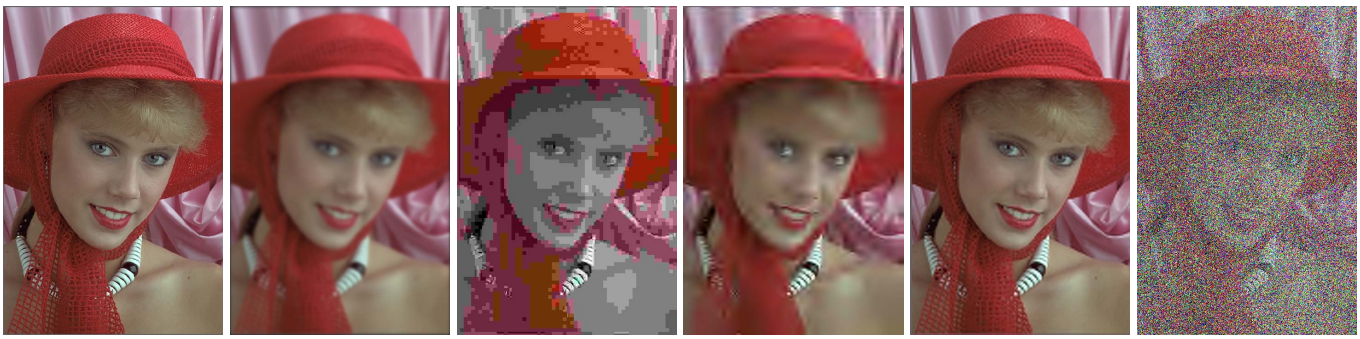


Fig. 15. Graphs of (a) correlation function ρ plotted against $\theta_2 - \theta_1$ for $\sigma = 2$, $d = 1$, and $\theta_2 = 0$ (b) peak correlation function P plotted against d for $\sigma = 2$ and $\theta_2 = 0$. Each plot shows the result of processing natural images (in blue) and white noise (in red).

A. Impact of Distortions on Correlation, Amplitude and Peak

We begin by visualizing the correlation, peak and amplitude functions as they are modified by distortion. Figures 17-19 show plots of ρ, P, and A, respectively, on the images modified by the distortions. The observations that will be drawn from the presented examples are generalized across other scales, σ, and spatial orientation θ_2.



(a) (b) (c) (d) (e) (f)

Fig. 16. Image "Woman Hat" and several distorted versions of it. (a) Distortion Free. (b) Blur. (c) JPEG. (d) JPEG 2000. (e) Fast Fading. (f) White Noise.

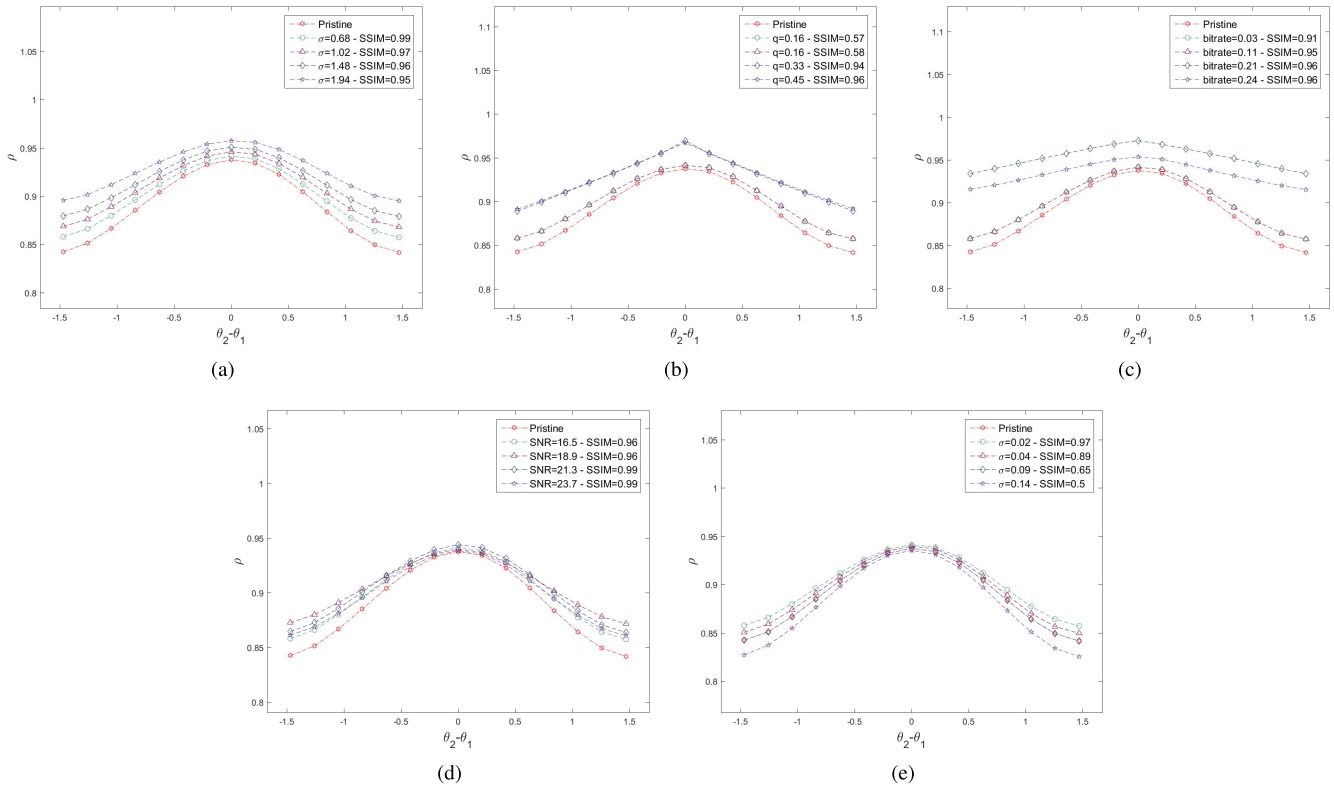


Fig. 17. Plots of the correlation function of image "Woman Hat" subject to (a) blur; (b) JPEG; (c) JPEG 2000; (d) Fast fading; (e) White noise.

1) Blur: Increases in blur were produced by increasing the space constant of the applied gaussian filter σ, which generally leads to worsening degradation of the perceptual image quality, which are reflected in drops in the Structural Similarity Index (SSIM) [8] between the blurred images and the undistorted original values. As expected, blur leads to an increase in the correlation functions of the bandpass normalized images, as can be seen in Fig. 17(a). The reductions of detail and diversity as a consequence of low-pass smoothing (Fig. 16(b)) progressively increases the correlation as the filter bandwidth is decreased. The increase in the correlation is monotonic with the level of blur. Furthermore, at small spatial separations, the values of the peak and amplitude functions increase, as can be seen in Fig. 18(a) and Fig. 19(a), respectively.

2) JPEG: Increases in JPEG compression is controlled by decreases in the JPEG coefficient quantization q, which in turn leads to reduction of the SSIM values between compressed and original images. JPEG distortion also leads to an increase in the computed correlation function since it causes both over-smoothing and blocking artifacts, and hence greater degrees of local homogeneity, as may be observed in Fig. 16(c). The increase in the correlation is monotonic, as may be seen by comparing the plots in Fig. 17(b). As a result, the peak and amplitude functions also increase in value, as in Fig. 18(b) and in Fig. 19(b) respectively. This increase is not limited to small spatial separations, unlike the case with blur. The values of A and P remain high even for larger separations when measured on heavily compressed images.

3) JPEG 2000: Ringing and blur are two common artifacts that afflict JPEG 2000 compressed images as may be observed in Fig. 16(c). Generally, the correlation is increased as shown in Fig. 17(c). However this increase is not monotonic

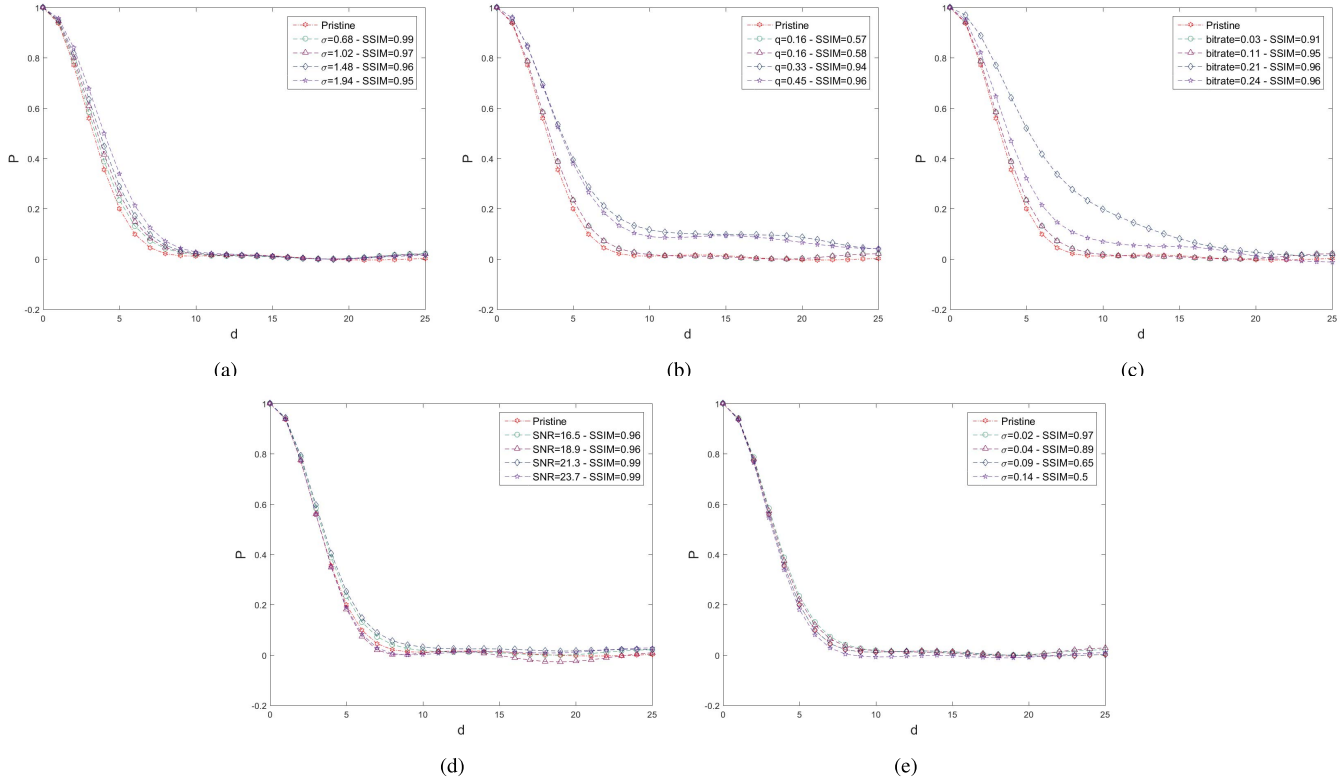


Fig. 18. Plots of the peak function of image "Woman Hat" subject to (a) blur; (b) JPEG; (c) JPEG 2000; (d) Fast fading; (e) White noise.

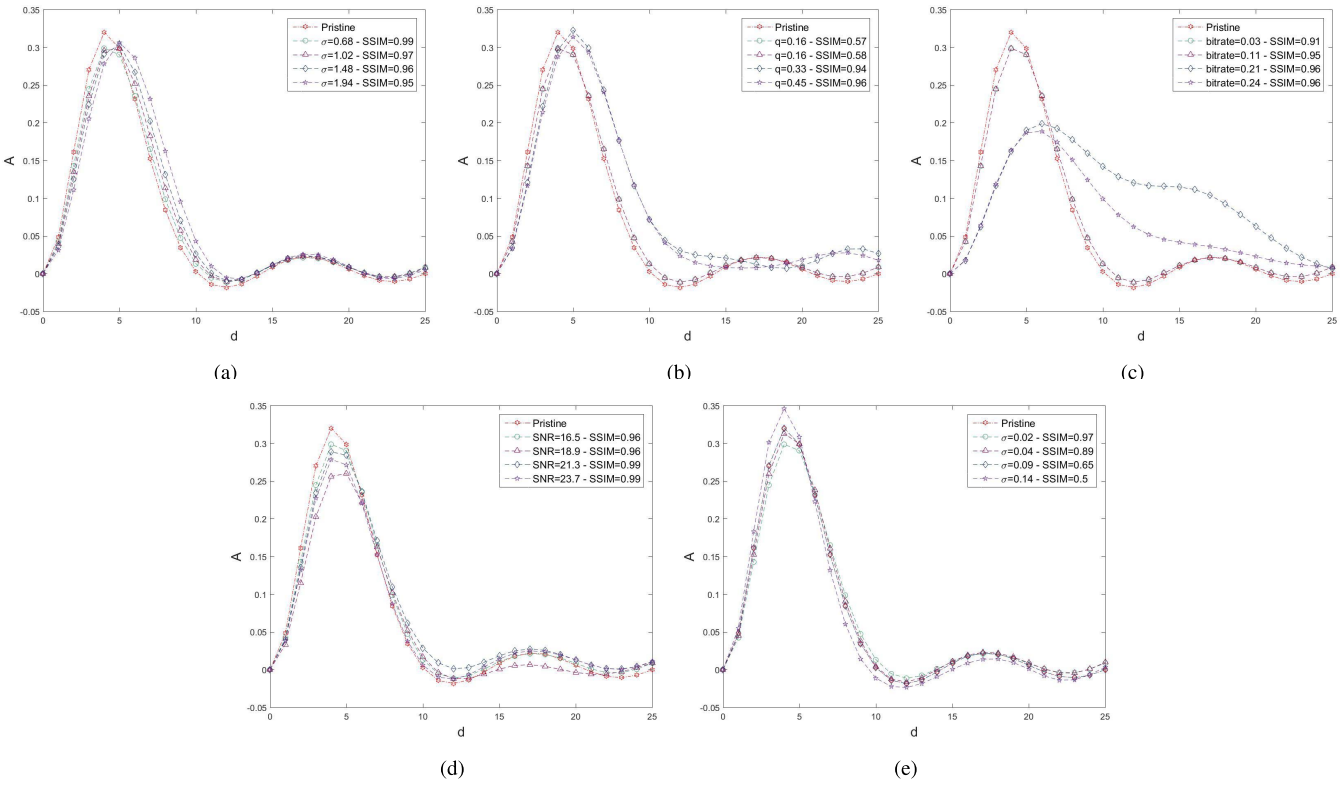


Fig. 19. Plots of the amplitude function of image "Woman Hat" subject to (a) blur; (b) JPEG; (c) JPEG 2000; (d) Fast fading; (e) White noise.

with increased compression. This may be partially explained by the fact that multiple parameters control JPEG 2000, such as the assigned weighting. The LIVE IQA database [39] does not indicate the weighting used on each image. P and A are impacted in ways similar to JPEG, as may seen in Fig. 18(c) and Fig. 19(c), respectively. The peak and amplitude functions

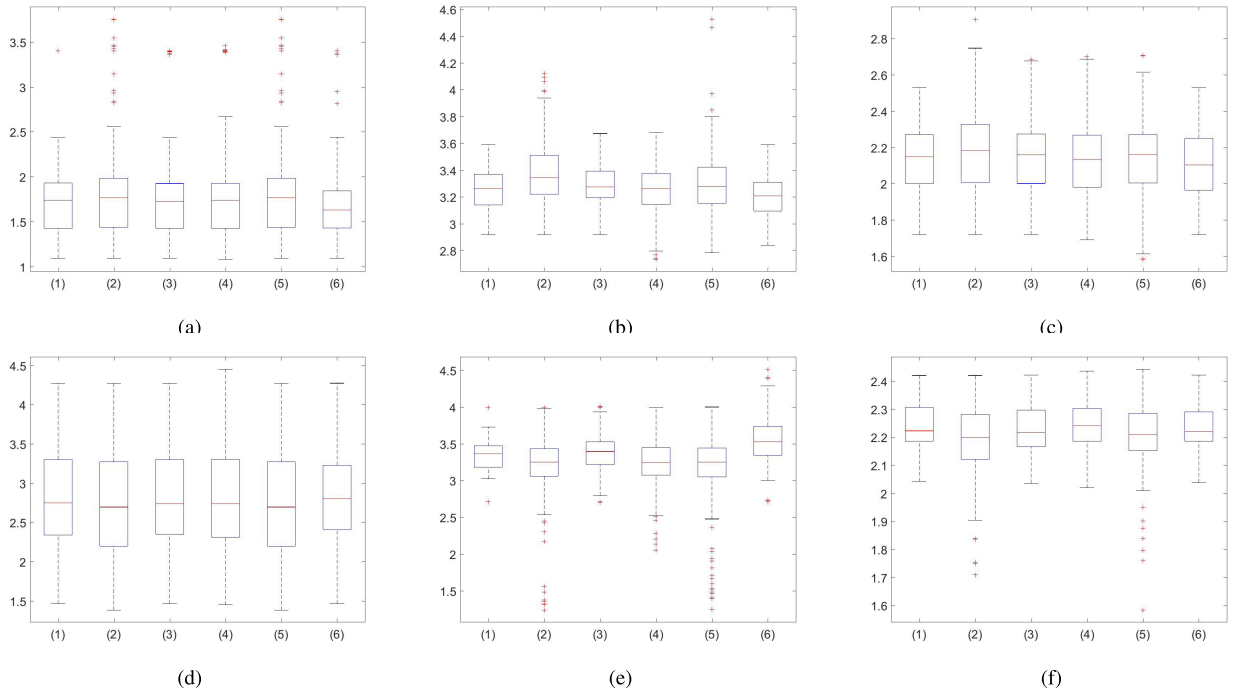


Fig. 20. Boxplots of the different parameters (a) α_0; (b) β_0; (c) α_1; (d) β_1; (e) α_2; and (f) β_2 for the various applied controlled distortions; (1) Undistorted; (2) Blur; (3) JPEG; (4) JPEG 2000; (5) Fast Fading; (6) White Noise.

in these cases exhibit bumps that are possibly caused by the ringing distortions.

4) Fast Fading: The fast fading category in LIVE IQA [39] is a complex, difficult distortion that is modeled as JPEG-2000 compression followed by fast fading bit errors. It also leads to increases in the correlation functions shown in Fig. 17(d). The behavior is not entirely monotonic owing to the complexity of the distortion. Generally, however, there is a resemblance in the correlation plots of JPEG 2000 and fast fading channel noise, since both contain JPEG 2000 compression artifacts (Fig. 16(d)). However, the increases in the peak and amplitude values are less subtle as compared to JPEG 2000, as depicted in Fig. 18(d) and in Fig. 19(d). This is because low compression (2.5 bits per pixel) was used to generate the JPEG 2000 distortion on all of the fast fading data, leading to less harsh ringing or blur artifacts as compared to the pure JPEG 2000 distortions.

5) White Noise: White noise of standard deviation σ was added to the R, G and B components. This leads to a decrease in the SSIM values. As a general trend, white noise leads to a decrease in the correlation functions, as expected. The peak correlation function is not impacted, as shown in Fig. 18(e). The amplitude functions appears to absorb most of the variation, as seen in Fig. 19(e), where the amplitude at small distances is higher. The exception to this general observation occurs at small standard deviations. In this case, the correlation slightly increased.

B. Impact of Distortions on Model Parameters α_0, β_0, α_1, β_1, α_2, and β_2

Understanding how the values of α_0, β_0, α_1, β_1, α_2, and β_2 are impacted as function of the distortions is less straightforward. It is not clear yet how changing trends in the different parameters impact A, P, and ρ. Figure 20 shows these parameters against the various considered distortions. In the presence of distortions, the distributions of the values of these different parameters are modified. Some parameters seem to respond to distortions better than others; by comparing the boxplots of α_2 and β_1, for example, it may be observed that the distribution of values of α_2 changes more drastically than does the distribution of values of β_1. In the near future, we will describe ways to use the parameters α_0, β_0, α_1, β_1, α_2, and β_2 as features to build correlation-based models that are able to automatically assess the perceptual quality of images.

Towards exploring the utility of correlation features for a wide array of possible distortion-sensitive applications (quality assessment, denoising, deblurring, deblocking, etc), we built a system to classify images by distortions. To do this, we focused on the set of distortions that are common to the LIVE IQA database [39] and the TID database [51]: JPEG 2000, JPEG, White Noise and Gaussian Blur. We partitioned the data into random 80%-20% training-testing splits, on which we trained a Support Vector Machine (SVM) [52] to classify images by distortions. We used $\{\alpha_0, \beta_0, \alpha_1, \beta_1\}$ at $\theta_2 = 0$ and $\theta_2 = \frac{\pi}{2}$ as training features (8 features). The experiment was repeated over 100 iterations, yielding a median correct classification rate of 70% on the LIVE IQA database [39] and 71% on the TID database [51]. These are good results using only a sparse set of second-order features. Including simple first-order (univariate) NSS features such as the shape and the variance parameters of the Mean Subtracted Contrast Normalized (MSCN) coefficients [9] improves the classification accuracy. Using the shape and variance parameters at scale 1 and the shape parameter at scales 2

TABLE VI
DISTORTION CLASSIFICATION PERFORMANCE

	JPEG2000	JPEG	White Noise	Gaussian Blur	Overall
Bivariate Model (LIVE IQA)	69%	67%	71%	75%	70%
Univariate + Bivariate Model (LIVE IQA)	85%	78%	94%	88%	85%
Bivariate Model (TID)	75%	50%	90%	74%	71%
Univariate + Bivariate Model (TID)	77%	80%	100%	90%	86%

(as defined in [9]) increased the correct classification rate to 85% on the LIVE IQA database [39] and to 86% on the TID database [51], using a total of 11 features. Details regarding the individual distortion classification performance and the inter-class accuracies are given in Table VI.

V. CONCLUSION

We built a simply, parametric bivariate natural scene statistic model of images and demonstrated its validity on a well-known set of high quality images. Our new model is global and is able to accurately capture the correlation behavior of natural images as well distorted images. As an application, we plan to use the model to derive new methods for predicting the quality of images [53]. We also plan to model color space structures [54], and to study the applicability of the bivariate NSS model to different modalities such as millimeter wave, X-ray and infra red images. These kinds of images have already been shown to nicely satisfy natural scene statistic models [55], [56]. We also plan to apply our new correlation model for the estimation of depth maps from both monocular and binocular images, building on the work in [1], [17], and [57].

REFERENCES

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

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

[3] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 4, no. 12, pp. 2379–2394, 1987.

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

[5] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 49–70, Oct. 2000.

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

[7] A. C. Bovik, "Automatic prediction of perceptual image and video quality," *Proc. IEEE*, vol. 101, no. 9, pp. 2008–2024, Sep. 2013.

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

[9] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[10] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar.2013.

[11] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[12] C.-C. Su, L. K. Cormack, and A. C. Bovik, "Color and depth priors in natural images," *IEEE Trans. Image Process.*, vol. 22, no. 6, pp. 2259–2274, Jun. 2013.

[13] E. P. Simoncelli, "Modeling the joint statistics of images in the wavelet domain," *Proc. SPIE*, vol. 3813, pp. 188–195, Jul. 1999.

[14] A. Lee, D. Mumford, and J. Huang, "Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model," *Int. J. Comput. Vis.*, vol. 41, nos. 1–2, pp. 35–59, 2001.

[15] 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, Nov. 2001.

[16] 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, Nov. 2002.

[17] D. 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, Jun. 2006.

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

[19] C.-C. Su, L. K. Cormack, and A. C. Bovik, "Bivariate statistical modeling of color and range in natural scenes," *Proc. SPIE*, vol. 9014, p. 90141G, Feb. 2014.

[20] C.-C. Su, L. K. Cormack, and A. C. 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.

[21] Z. Sinno and A. C. Bovik, "Generalizing a closed-form correlation model of oriented bandpass natural images," in *IEEE Global Conf. Signal Inf. Process.*, Dec. 2015, pp. 373–377.

[22] Z. Sinno and A. C. Bovik, "Relating spatial and spectral models of oriented bandpass natural images," in *Proc. Southwest Symp. Image Anal. Interpretation*, Mar. 2016, pp. 89–92.

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

[24] J. B. Johnson, "Thermal agitation of electricity in conductorse," *Nature*, vol. 119, pp. 50–51, 1927.

[25] J. M. Halley, "Ecology, evolution and 1f-noise," *Trends Ecol. Evol.*, vol. 11, no. 1, pp. 33–37, 1996.

[26] A. E. Cohen, A. Gonzalez, J. H. Lawton, 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. Roy. Soc. London B, Bio. Sci.*, vol. 265, no. 1390, pp. 11–15, 1998.

[27] R. T. Baillie, "Long memory processes and fractional integration in econometrics," *J. Econometrics*, vol. 73, no. 1, pp. 5–59, Jul. 1996.

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

[29] R. F. Voss and J. Clarke, "'$1/f$ noise' in music: Music from $1/f$ noise," *J. Acoust. Soc. Amer.*, vol. 63, no. 1, pp. 258–263, 1978.

[30] A. H. Tewfik and M. Kim, "Correlation structure of the discrete wavelet coefficients of fractional Brownian motion," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 904–909, Mar. 1992.

[31] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, no. 4, pp. 422–437, Oct. 1968.

[32] B. B. Mandelbrot, *Fractals: Form, Chance and Dimension*. San Francisco, CA, USA: Freeman, 1979, p. 365.

[33] M. A. Riley and G. C. Van Orden, "Tutorials in contemporary nonlinear methods for the behavioral sciences," Nat. Sci. Found., Alexandria, AZ, USA, Tech. Rep., 2005. [Online]. Available: https://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp

[34] C. Carlson, "Thresholds for perceived image sharpness," *Photograph. Sci. Eng.*, vol. 22, no. 2, pp. 69–71, 1982.

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

[36] B. Olshausen and D. Field, "How close are we to understanding V1?" *Neural Comput.*, vol. 17, no. 8, pp. 1665–1699, Aug. 2005.

[37] H. Lee, C. Ekanadham, and A. Y. Ng, "Sparse deep belief net model for visual area V2," in *Proc. Neural Inf. Process. Syst.*, 2008, pp. 873–880.

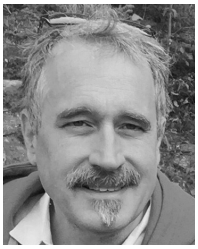
[38] M. S. Keshner, "$1/f$ noise," *Proc. IEEE*, vol. 70, no. 3, pp. 212–218, Mar. 1982.

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

[40] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 9, pp. 891–906, Sep. 1991.
[41] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. IEEE Int. Conf. Imag. Process.*, vol. 3. Oct. 1995, pp. 444–447.
[42] F. Pascal, L. Bombrun, J.-Y. Tourneret, and Y. Berthoumieu, "Parameter estimation for multivariate generalized Gaussian distributions," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5960–5971, Dec. 2013.
[43] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky, "Random cascades on wavelet trees and their use in analyzing and modeling natural images," *App. Comput. Harmonic Anal.*, vol. 11, no. 1, pp. 89–123, 2001.
[44] M. Fisz, "The limiting distribution of a function of two independent random variables and its statistical application," *Colloq. Math.*, vol. 2, no. 3, pp. 138–146, 1955.
[45] P. Fryzlewicz and G. P. Nason, "A Haar-Fisz algorithm for Poisson intensity estimation," *J. Comput. Graph. Statist.*, vol. 13, no. 3, pp. 621–638, 2004.
[46] B. A. Olshausen and D. J. 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.
[47] P. E. Gill and W. Murray, "Quasi-Newton methods for unconstrained optimization," *IMA J. Appl. Math.*, vol. 9, no. 1, pp. 91–108, Feb. 1972.
[48] E. Larson and D. Chandler. *Categorical Image Quality CSIQ Database 2009*. Accessed: Jan. 2, 2018. [Online]. Available: http://vision.okstate.edu/csiq
[49] Y. Horita, K. Shibata, Y. Kawayoke, and Z. M. P. Sazzad. *MICT Image Quality Evaluation Database*. [Online]. Available: http://mict.eng.u-toyama.ac.jp/mictdb.html
[50] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annu. Rev. Neurosci.*, vol. 24, no. 1, pp. 1193–1216, 2001.
[51] N. Ponomarenko *et al.*, "Image database TID2013: Peculiarities, results and perspectives," *Signal Process., Image Commun.*, vol. 30, pp. 57–77, Jan. 2015.
[52] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207–1245, 2000.
[53] Z. Sinno, C. Caramanis, and A. C. Bovik, "Second order natural scene statistics model of blind image quality assessment," in *Proc. IEEE Conf. Acous. Speech Signal Process. (ICASSP)*, Apr. 2018.
[54] Z. Sinno and A. C. Bovik, "On the natural statistics of chromatic images," in *Proc. Southwest Symp. Image Anal. Interpretation*, Apr. 2018.
[55] T. R. Goodall, A. C. Bovik, and N. G. Paulter, "Tasking on natural statistics of infrared images," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 65–79, Jan. 2016.
[56] P. Gupta, J. Glover, P. J. N. G, and A. C. Bovik, "Studying the statistics of natural X-ray pictures," *ASTM J. Test. Eval.*, to be published.
[57] C. C. Su, L. K. Cormack, and A. C. Bovik, "Bayesian depth estimation from monocular natural images," *J. Vis.*, vol. 17, no. 5, pp. 1–29, May 2017.

Zeina Sinno received the B.E. degree (Hons.) in electrical and computer engineering with a minor in mathematics from the American University of Beirut, in 2013. She received the M.S. degree in electrical and computer engineering from The University of Texas at Austin in 2015, where she is currently pursuing the Ph.D. degree with the Laboratory for Image and Video Engineering. Her research focuses on image and video processing and machine learning.

Constantine Caramanis received the A.B. degree in mathematics from Harvard University, and the M.S. and Ph.D. degrees in electrical engineering and computer science from MIT. He joined the Electrical and Computer Engineering Department, The University of Texas at Austin, in 2006, where he is currently an Associate Professor. His research interests center on decision-making in large-scale complex systems, with a focus on learning and computation, robust and adaptable optimization, high-dimensional statistics and machine learning, and applications to large-scale networks, including social networks, wireless networks, transportation networks, and energy networks. He also works on applications of machine learning and optimization to computer-aided design.



Alan C. Bovik (F'95) received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from the University of Illinois in 1980, 1982, and 1984, respectively. He is currently the Cockrell Family Regents Endowed Chair Professor at The University of Texas at Austin. He has authored the books *The Handbook of Image and Video Processing*, *Modern Image Quality Assessment*, *The Essential Guides to Image Processing*, and *The Essential Guides to Video Processing*. He received the 2017 Edwin H. Land Medal from the Optical Society of America, a 2015 Primetime Emmy Award for the Outstanding Achievement in Engineering Development, the 2013 IEEE Signal Processing Society Award, and about 10 journal best paper awards, including the 2017 IEEE Signal Processing Society Sustained Impact Award. He co-founded and was the longest-serving Editor-in-Chief of the IEEE TRANSACTIONS ON IMAGE PROCESSING, and created the IEEE International Conference on Image Processing in Austin, Texas, in 1994.