Kurtosis and the Phase Structure of Images

Joshua Gluckman
jgluckma@poly.edu
Dept. of Computer Science
Polytechnic University
6 Metrotech Center
Brooklyn, NY, 11201

Abstract

Many studies have shown that natural imagery exhibit certain statistical regularities such as scale invariance and non-Gaussian behavior. In particular, it has been shown that the marginal statistics (histograms) of filtered images have distributions with large kurtosis. Furthermore, the joint statistics of two or more filter responses are non-Gaussian often with circular and "diamond" shaped distributions. In this paper, we examine the relationship between the phase angle of a signal and these non-Gaussian statistics. It is shown that both local and global correlations in the phase angle lead to many of the statistical regularities that have been observed in natural images. In addition, the discriminative power of marginal and joint statistical representations is studied. From this study we are able to show that these representations have difficulty differentiating changes in the local correlations of phase angle. Yet, these small changes can manifest as global and visually distinct differences.

1 Introduction

Over the last decade statistical representations have emerged as one of the leading image models and have led to successful applications in denoising, compression, retrieval and texture analysis/synthesis [1] [2] [3] [4] [5] [6] [7] (to name a few). These models have been motivated by the empirical discovery of certain regularities present in a wide range of natural imagery.

One of the first regularities discovered was the similarity of power spectra across a variety of natural stimuli. The power spectrum records energy as a function of spatial frequency and for natural image spectra appears to behave according to $a^{-\gamma}$, where a is frequency and γ is a parameter that varies from image to image but is typically small [8]. This power-law distribution implies that second-order image statistics are independent of scale over suitable bandwidths. It has also been observed that some higher order statistics are scale-invariant. Several studies have shown that histograms of derivative images ,when averaged over a large ensemble of images, are approximately invariant to scale [8] [9] [10]. These scale-invariant observations point to the "fractal" nature of images and have been attributed to the ubiquity of edges and the distribution of object sizes in natural images. Indeed, it has been shown that images synthesized from randomly placed occluding objects with sizes r distributed according to $\frac{1}{r^3}$ have scale invariant properties [11] [12]. Yet a third instance of scale-invariance has been uncovered in the second spectrum, the power spectrum of the image squared [13].

Many researchers have also demonstrated the non-Gaussian statistics of natural images (for a review see [14]). In particular, it has been shown that the histograms (marginal statistics) of filtered images typically have single modes with heavy tails and cusps at zero, characteristic of highly kurtotic or non-Gaussian distributions. This statistical regularity has been observed for derivative filters, Gabor filters, Laplacian filters, wavelets and even small random kernel filters [15] [16] [9] [6] [17] [18].

Studies of joint histograms of two or more filter responses have also shown regularities and dependencies across scale and orientation. Pairs of even/odd Gabor filters exhibit circular symmetry [19]. Wavelets at the same orientation but adjacent scale are highly dependent [20], which has been shown to be the result of highly kurtotic but circularly symmetric joint histograms [18]. The same study also pointed out that joint histograms of wavelets applied at the same scale but different orientation lead to kurtotic distributions with polyhedral shapes.

Although these observations provide insight into the space of images they are somewhat disjoint and the observations do not necessarily imply each other. Therefore, it is not clear which model to use or how to combine them for a particular application. In this paper, we attempt to bridge this gap by examining the relationship between the phase structure of images and these observed statistical regularities. As we will show, correlations in the phase angle of an image can "explain" the non-Gaussian statistics of marginal and joint histograms of filtered images. We note that "phase correlations" are not the physical cause of non-Gaussian statistics (clearly, edges and objects are). Thus, it would seem that we are merely replacing one descriptive model (non-Gaussian marginal and joint statistics) with another (correlations of phase angle). However, the study of phase structure points to a potentially unifying framework for the observed statistical regularities of images. In addition, it will also shed light on the descriptive/discriminative power of some commonly used statistical representations for textures and visual patterns.

2 Image Statistics and Phase Structure

In a now classic experiment, Oppenheim and Lim demonstrated the perceptual importance of the information encoded in the phase of a signal [21]. Figure 1 shows an example of an image whose phase structure has been removed (middle column) next to an image whose power spectrum (right column) has been removed. For an image f with Fourier transform F the phase structure is removed by setting $F' = |F|e^{-2\pi i\rho}$ where ρ is drawn from a uniform random distribution, while the power spectrum is removed by setting $F'' = \frac{F}{|F|}$. After taking the inverse Fourier transform we

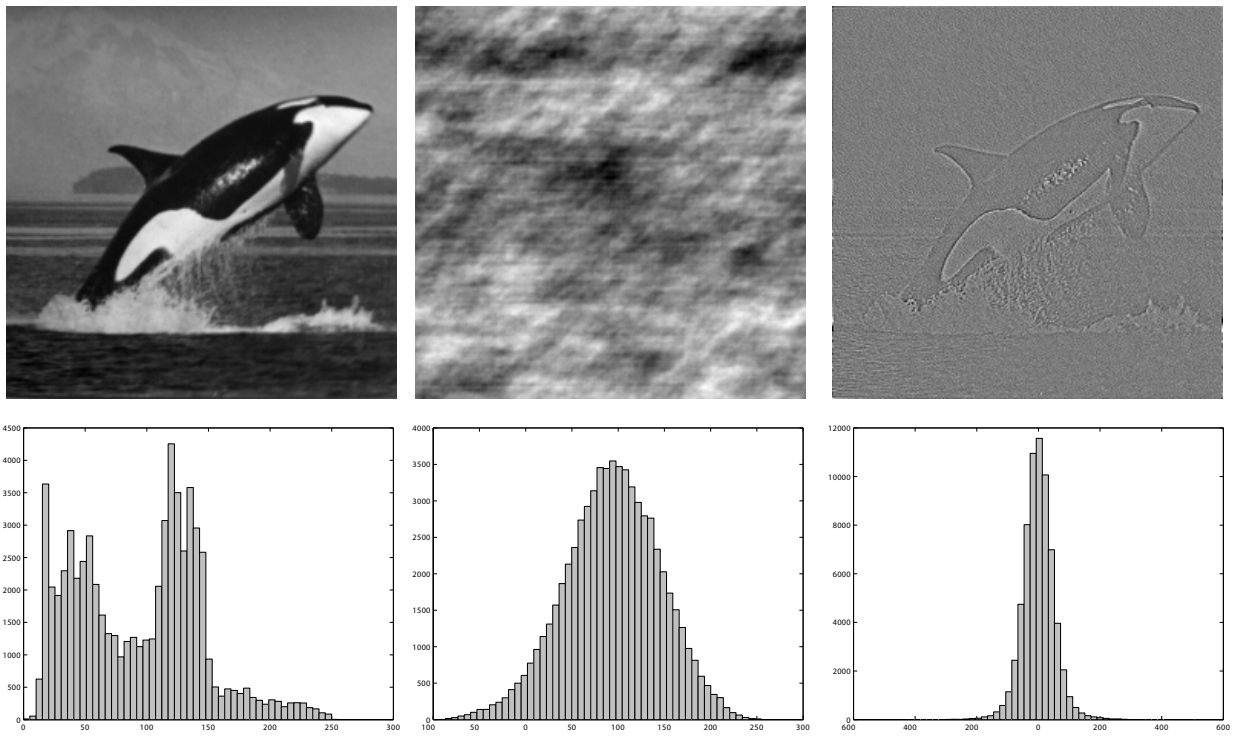


Figure 1: The importance of phase. On the left is an image with its histogram. The center image has an identical power spectrum but randomized phase, resulting in a Gaussian histogram. The right image shares the same phase but with a flattened power spectrum, resulting in a non-Gaussian (kurtotic) histogram.

have images f' and f'' whose phase structure and power spectrum have been removed, respectively.

In the signal processing literature removing or flattening the power spectrum is known as "whitening" a signal. What is immediately apparent is that the perceptually salient features (edges) are present in the "whitened" image but not in the power spectrum image. Below each image in figure 1 is the histogram formed from the image intensities. Because f and f' share the same power spectrum their histograms have identical variance ($\mu_2(f') = \mu_2(f)$). For f' the distribution is clearly Gaussian which results from the entropy maximizing operation of randomizing the phase. For f'', the whitened image, the variance is arbitrary in the sense that the power spectrum remains flat ("white") if the intensity is scaled; thus we set $\mu_2(f'') = \mu_2(f')$ in the histogram. Given the identical variance of the histograms it is clear that the histogram of f'' is non-Gaussian.

Thomson has studied the statistics of a large database of natural images after "whitening" the signals and found a non-Gaussian distribution in the histograms [13]. Because the power specturm is removed the non-Gaussian statistics imply the presence of non-random structure in the

phase angle of natural images. While the histogram of an image is independent of spatial image structure, histograms formed from filtered images are dependent. These marginal statistics have been successfully used for texture synthesis and classification [22] [16] [23] [24] [25]. Images that share the same power spectrum not only share the same variance but have the same variance after convolution with a filter. For an arbitrary filter g we have $\mu_2(g*f) = \mu_2(g*f')$. While the variance of a filtered image is independent of phase structure, higher order moments (such as kurtosis) are not. Thus, the non-Gaussian statistics of filtered images are further evidence of information in the phase of an image signal. The phase dependency of moments beyond two has motivated the development of higher order spectral analysis [26]. Using the van Hateran image database, Thomson studied the higher order statistics of natural images and found a power-law like behavior of the phase only second spectrum, which for a "whitened" image f is the power spectrum of f^2. Taking the square of the image function introduces phase information into the power spectrum. It is this connection between signal phase and the non-Gaussian statistics of images that motivates us to study image kurtosis. In addition, it has been shown that the histograms of filtered images can be modeled by two parameter density functions such as the generalized Laplacian, Bessel K forms and Weibull distributions [15] [6] [25] [27]. The parameters of these densities are functions of variance and kurtosis.

Kurtosis is the fourth moment normalized by the second moment which is defined as

$$\kappa = \frac{\mu_4}{\mu_2^2} - 3, \tag{1}$$

where μ_4 and μ_2 are the fourth and second central moments. For a Gaussian distribution $\frac{\mu_4}{\mu_2^2} = 3$, therefore a non-Gaussian distribution has a non-zero kurtosis. The purpose of this to examine the relationship between image kurtosis and phase structure in order to develop a model of phase structure that follows from the observed non-Gaussian statistics of natural images. In addition, we study the ability of marginal and joint statistics to discriminate variations in phase structure.

3 Kurtosis and Phase Structure

Let $f(x, y)$ be a bandlimited function defined on the continuous image domain $\Omega \subset \Re^2$. We represent the image function by a finite Fourier series

$$f(x, y) = \sum_{i=1}^{n} m_i \cos(u_i x + v_i y + \phi_i),$$

where m_i and ϕ_i are the magnitude and phase angle associated with integer frequency (u_i, v_i). Note that $\cos(ux + vy) = \cos(-ux - vy)$, so to force a unique representation we assume that $u_i \in \mathbb{Z}^+$ and $v_i \in \mathbb{Z}$, which also implies that f is zero-mean. The k^{th} central moment of the marginal distribution (normalized histogram) of the zero-mean function f is

$$\mu_k(f) = \frac{1}{|\Omega|} \int_\Omega f(x,y)^k d\Omega.$$

From the orthogonal property of a cosine function it follows that the second moment reduces to

$$\mu_2 = \frac{1}{|\Omega|} \int_\Omega \sum_{i=1}^{n} m_i^2 \cos^2(u_i x + v_i y + \phi_i) = \frac{1}{2} \sum_{i=1}^{n} m_i^2,$$

which is independent of the phase angle. Note that this expression is Parsevels theorem; energy in the spatial domain is equivalent to energy in the frequency domain. Unfortunately, such a simple relationship does not exist for the fourth moment, which is defined as

$$\mu_4 = \frac{1}{|\Omega|} \int_\Omega \left(\sum_{i=1}^{n} m_i \cos(u_i x + v_i y + \phi_i) \right)^4 d\Omega. \tag{2}$$

Dependency of moments beyond two on phase is due to the fact that while cosine pairs are orthogonal, products of three or more cosine functions do not imply orthogonality. In order to examine the relationship between phase angle and the fourth moment independent of the power spectrum we assume the magnitudes are unity. To simplify the analysis we also assume the phase of each frequency is either 0 or π, and represented by the signed magnitudes $s_i \in \{-1, 1\}$. This is equivalent to restricting the analysis to functions whose Fourier transform is real (symmetric functions). Note that in terms of representing "interesting" images this is not particularly restrictive, because a symmetric image can always be created by reflection.

The terms in (2) can be grouped according to the number and power of the cosine functions,

$$\mu_4 = \sum_i M_1(i) + \sum_{i,j} M_2(i,j) + \sum_{i,j} M_3(i,j) + \sum_{i,j,k} M_4(i,j,k) + \sum_{i,j,k,l} M_5(i,j,k,l), \tag{3}$$

where

$$\begin{aligned}
M_1 &= \frac{1}{|\Omega|} \int_\Omega \cos^4(z_i) d\Omega \\
M_2 &= \frac{1}{|\Omega|} \int_\Omega 6 \cos^2(z_i) \cos^2(z_j) d\Omega \\
M_3 &= \frac{1}{|\Omega|} \int_\Omega 4 s_i^3 s_j \cos^3(z_i) \cos(z_j) d\Omega \\
M_4 &= \frac{1}{|\Omega|} \int_\Omega 12 s_j s_k \cos^2(z_i) \cos(z_j) \cos(z_k) d\Omega \\
M_5 &= \frac{1}{|\Omega|} \int_\Omega 24 s_i s_j s_k s_l \cos(z_i) \cos(z_j) \cos(z_k) \cos(z_l) d\Omega.
\end{aligned}$$

For integer frequencies, $u_i \in \mathbb{Z}^+$ and $v_i \in \mathbb{Z}$ the integrals can be solved and we have,

$$\begin{aligned}
M_1 &= \frac{3}{8} \\
M_2 &= \frac{3}{2} \\
M_3 &= \begin{cases} s_i^3 s_j \frac{1}{2} & \text{if } 3(u_i, v_i) = (u_j, v_j) \\ 0 & \text{otherwise} \end{cases} \\
M_4 &= \begin{cases} s_j s_k \frac{3}{2} & \text{if } 2(u_i, v_i) = (u_j, v_j) \pm (u_k, v_k) \\ 0 & \text{otherwise} \end{cases} \\
M_5 &= \begin{cases} s_i s_j s_k s_l 3 & \text{if } (u_i, v_i) - (u_j, v_j) = (u_k, v_k) \pm (u_l, v_l) \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

The M_1 and M_2 terms are independent of both the location and phase of the frequencies, therefore

$$\sum_{i=1}^{n} M_1(i) = \sum_{i=1}^{n} \frac{3}{8} m_i^4 = \frac{3}{8} n$$

and

$$\sum_{i,j} M_2(i,j) = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{3}{2} m_i^2 m_j^2 = \binom{n}{2} \frac{3}{2}.$$

This is a consequence of the fact that they are composed of even powers of cosine functions. In contrast, the other groups of terms M_3, M_4, and M_5 depend on the spatial arrangement of the frequencies, in the (u, v) frequency plane, and the phase of the individual frequencies. The spatial arrangement determines if the k-tuple of frequencies contribute to one of the groups and the phases determine if they contribute positively or negatively. Figure 2 shows the geometric arrangements of frequency locations that contribute to these terms. Frequency pairs that contribute to M_3 must lie on a line through the origin of frequency space. Frequency triples that contribute to M_4 come in two flavors. Type I is composed of three equidistant points on a line in frequency space. In type II configurations the three points and the origin form a trapezoid where one of the parallel sides is twice the length of the other. Frequency quadruples contribute to M_5 and also have two configurations. Type I is composed of four points that form a parallelogram this includes the degenerate case where the four points lie on a line with equidistance between each pair. A type II configuration of four points is a parallelogram formed by two of the points, the origin and the sum of the other two points. Type II configurations are analogous to what are referred to as sum beats in the literature on higher order statistical analysis and type I configurations are analogous to difference beats.

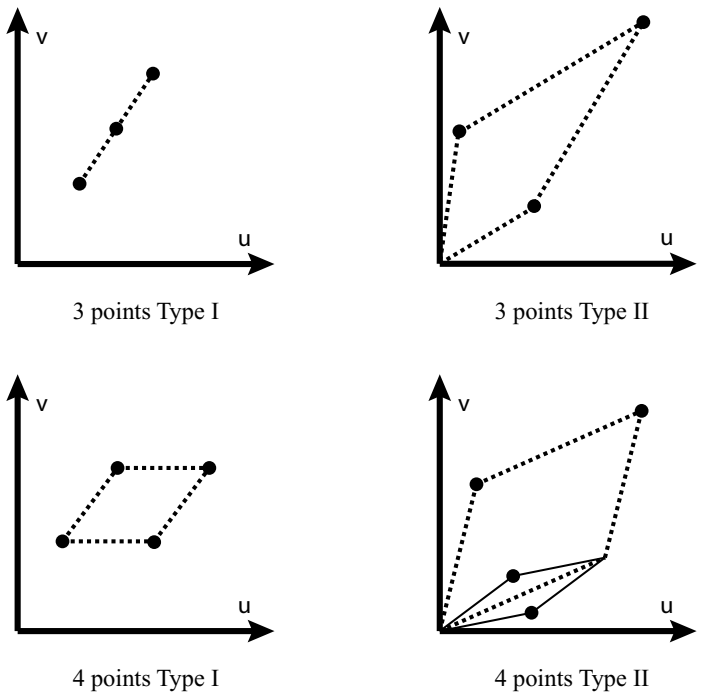


Figure 2: Geometric arrangements of frequencies that contribute to the fourth moment. Type I geometries are independent of the origin while type II involve the origin.

For both three and four points type I configurations are formed locally in frequency space while type II configurations are tied to the origin and therefore non-local. Naturally one would expect that local alignment of phases in frequency space to dominate and indeed, Thomson found that the kurtosis of natural images is almost entirely accounted for by difference beats (type I in our case) [13]. Therefore, in the analysis we ignore type II configurations.

For a k-tuple of frequencies that satisfy the geometric constraints of M_3, M_4, or M_5 we refer to the k-tuple as "in phase" if M_i is positive and "out of phase" if negative. For example, consider a 4-tuple of frequencies that form a parallelogram. If $s_i s_j s_k s_l$ is positive then the 4-tuple is "in phase." Assuming the magnitudes m_i are unity then we have the following quantities:

$$\sum_{i,j} M_3(i,j) = \frac{1}{2}W_3, \qquad \sum_{i,j,k} M_4(i,j,k) = \frac{3}{2}W_4, \qquad \text{and} \qquad \sum_{i,j,k,l} M_5(i,j,k,l) = 3W_5,$$

where W_3 is the difference between the number of frequency pairs that are in phase and those that are out of phase. Likewise, W_4 and W_5 are the difference between in and out of phase frequency triples and quadruples.

Now, we can define kurtosis in terms of phase structure, which is

$$\kappa = \frac{\mu_4}{\mu_2^2} - 3 = \frac{3n^2 - \frac{3}{2}n + 2W_3 + 6W_4 + 12W_5}{n^2} - 3 \approx \frac{2W_3 + 6W_4 + 12W_5}{n^2}. \tag{4}$$

For a sampled image the number of frequencies is equivalent to the number of pixels. Because the n pixels (frequencies) are defined on a two dimensional lattice, the number of geometric configurations of frequencies that contribute to W_3, W_4, and W_5 is bounded by

$$W_3 = O(n), \quad W_4 = O(n^2), \quad \text{and} \quad W_5 = O(n^3). \tag{5}$$

This is a consequence of the fact that for configurations of 2, 3, and 4 points we can choose any 1,2 or 3 points on the integer lattice in frequency space and solve for the remaining point that satisfies the geometry (a finite number of solutions will exist). For a total of n points there are n choose 1, n choose 2, and n choose 3 possible configurations for W_3, W_4, and W_5 respectively. For images n is large and the kurtosis will be dominated by W_5 reducing (4) to

$$\kappa \approx \frac{12W_5}{n^2}. \tag{6}$$

Thus, we have arrived at a simple relationship between the phase structure of images and kurtosis. To summarize, if we consider all possible parallelograms formed by the frequency space lattice then kurtosis is dependent on the difference between the number of in phase and out of phase parallelograms. In order for the alignment of phase to effect image kurtosis in a meaningful way, W_5 must have a combinatorial dependence on n.

Next, some simple models are considered of phase structure of the form $I = P \cdot N$, where the image I is formed by modulating Gaussian noise N with a pattern P. Modulation in the spatial domain is convolution in the frequency domain, which correlates the phase angles of I. The strength of the correlations are dependent on the pattern P. Such signals, called contrast modulations, are often used to study second-order visual processing [28][29].

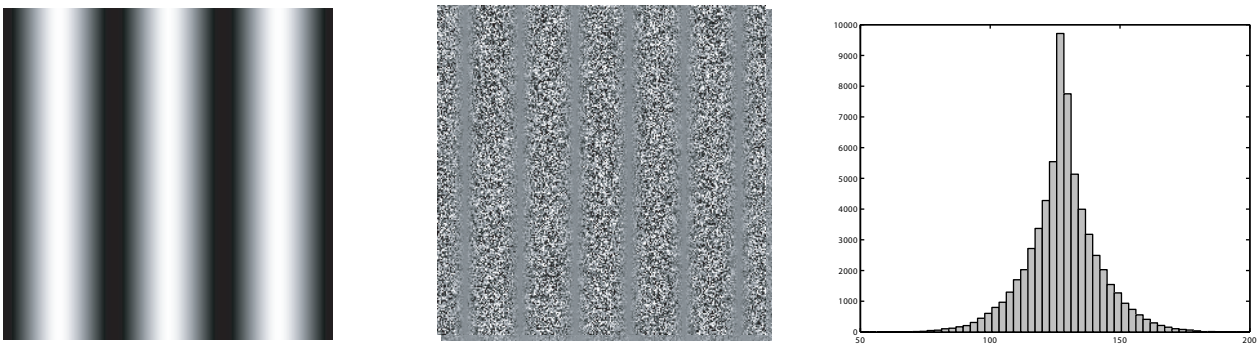
The simplest model of phase structure is $P = \cos(ax)$, whose Fourier transform is the pair of points $\{(a, 0), (-a, 0)\}$. Therefore, frequency pairs at the same offset, (u, v) and $(u + 2a, v)$, in I will have correlated phases. For an image with n frequencies there are $\frac{n}{2}$ frequency pairs with this offset. If each pair is in phase then a parallelogram formed by any *two* pairs will also be in phase which leads to,

$$W_5 = \binom{\frac{n}{2}}{2}, \quad \text{and} \quad \kappa \approx 1.5. \tag{7}$$

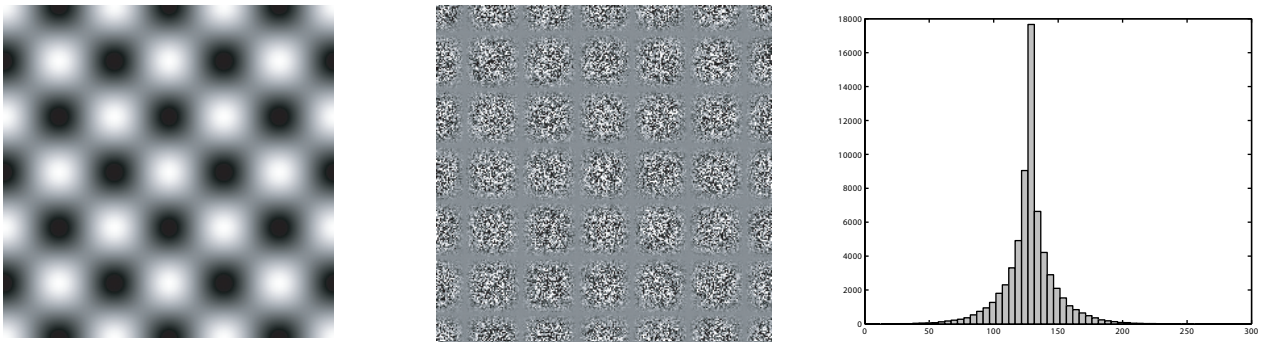
The top panel of figure 3 shows an example of an image generated with this model.

Another model of phase structure is one where the phase of parallelograms in frequency space are aligned. From each pair of parallelograms, 10 phase aligned parallelograms can be formed by choosing two points from each parallelogram. For n frequencies there are $\frac{n}{4}$ parallelograms therefore,

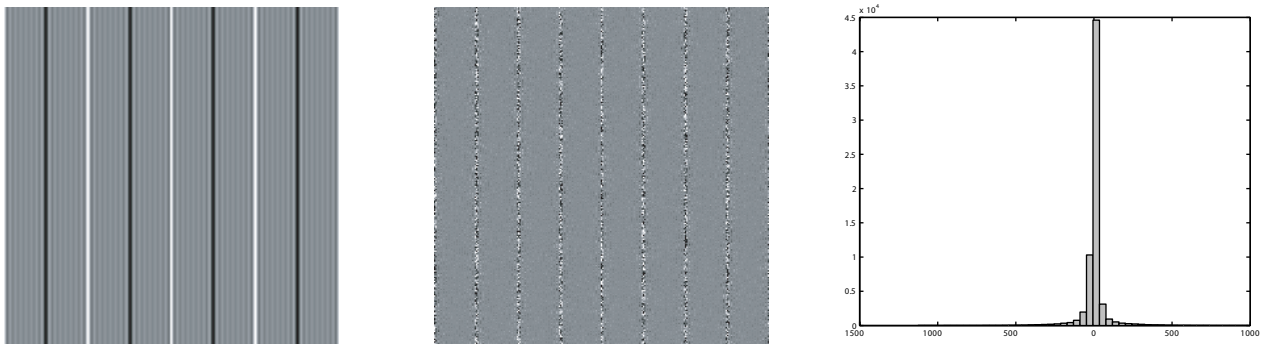
$$W_5 = 10\binom{\frac{n}{4}}{2} \quad \text{and} \quad \kappa \approx 3.75. \tag{8}$$



(a)



(b)



(c)

Figure 3: Examples of phase alignment by modulating Gaussian noise with sinusoidal patterns. The modulation patterns are shown on the left column, the resulting modulated noise is shown in the middle column and the histograms are on the right. The alignment of phase leads to a positive kurtosis κ. (a) A single cosine function. (b) A two dimensional sinusoid ($\cos(a_1 x + b_1 y)\cos(a_2 x + b_2 y)$). (c) A sum of K one dimensional cosine functions at equidistant points in frequency space ($\sum_{k=1}^{K}\cos(akx)$).

Using the modulation pattern $P = \cos(a_1 x + b_1 y)\cos(a_2 x + b_2 y)$, an image with this phase structure is generated in the middle panel of figure 3.

Next, the general case where k frequencies are correlated is considered. If sets of k frequencies are correlated then there are approximately $\frac{n}{k}$ sets. Assuming the k frequencies are arranged on a grid (could be a line) in frequency space then for each pair of k frequency sets $O(\frac{k^3}{3})$ parallelograms can be formed between them, therefore

$$W_5 = \frac{k^3}{3}\binom{\frac{n}{k}}{2} \quad \text{and} \quad \kappa = O(ck), \tag{9}$$

where c is a constant dependent on the particular arrangement of points. For example it is approximately 2 when the points are on a line and close to 1 when the points form a square grid. Note that when the arrangements of the frequencies deviates greatly from a grid, such as a ring, then the kurtosis is approximately constant. In the bottom panel of the figure is the modulation pattern $P = \sum_{i=1}^{8} \cos(aix)$ whose Fourier transform is $k = 16$ equidistant points on a line.

In the previous examples the same phase structure is correlated across the entire frequency plane. What if only a portion is correlated? Suppose that a particular pattern of phase correlation results in a kurtosis of $\kappa = \hat{\kappa}$. If α is the proportion (or density) of frequency space that is phase correlated, from (6) kurtosis is reduced by $\kappa \approx \hat{\kappa}\alpha^2$ There are two interpretations of α. The density of the phase correlations throughout the frequency plane or the portion of area of the frequency plane that is phase correlated.

So what can we infer about phase structure from the kurtosis of "whitened" natural images? Clearly, phase correlations must vary throughout the frequency plane otherwise only the simplest of patterns could be formed such as those in figure 3. However, as more variations of phase correlation are introduced into the frequency plane the density of each will go proportionaly down. For example, if n equal sized regions of the frequency plane each have kurtosis $\hat{\kappa}$ with independent correlations the kurtosis of the whole image is $\kappa = n\hat{\kappa}\frac{1}{n^2} = \frac{\hat{\kappa}}{n}$. In order to "explain" the high kurtosis of natural images the phase correlations in each region must be strong, meaning a large k in equation (9).

Next, we address the discriminative power of kurtosis. While kurtosis is sensitive to changes in both k and α, a tradeoff between k and α can result in the same amount of kurtosis. In addition, a change in density can not be distinguished from confining the phase correlations to a portion of frequency space. Figure 4 illustrates this by showing some examples of stochastic images with the same kurtosis (and variance) but which are visually distinct. These images are generated using patterns of the form $P = \cos^a(ux)\cos^b(vy)$, for which the Fourier transform is a grid of $k =$

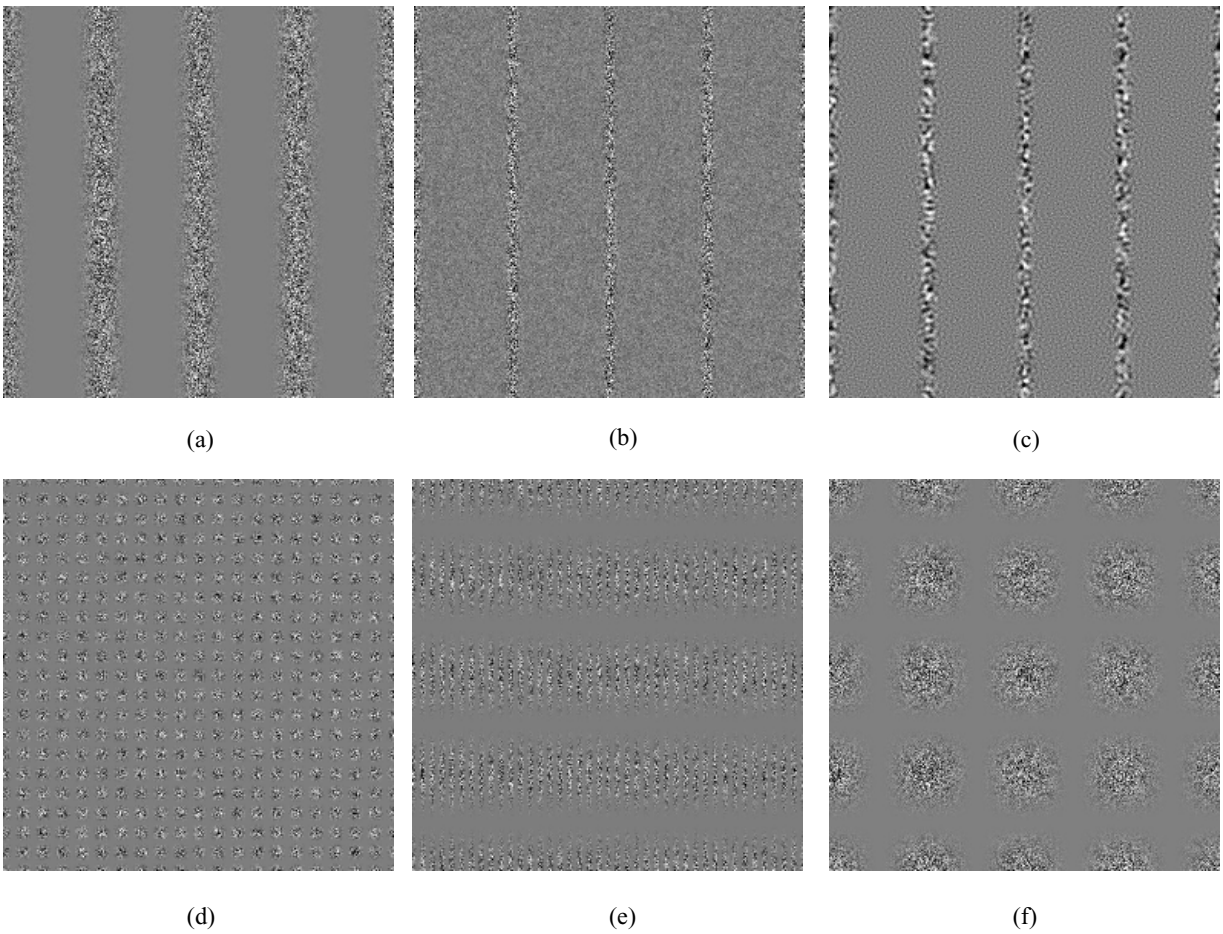


Figure 4: Examples of images with the same kurtosis. By varying the geometry, density and location of the phase structure, visually distinct images can be created that have the same kurtosis and variance.

$(a+1) \times (b+1)$ points. Images (b) and (c) use a larger value for k as compared to (a) but have the same kurtosis due to a reduced density. The image in (b) is generated by $P \cdot N + N$. Adding noise to the pattern reduces the density of the phase correlations. The variance of the noise is chosen so that the kurtosis of (b) matches (a). The image in (c) is generated from $P \cdot N_{\text{low}} + N_{\text{high}}$ where N_{low} and N_{high} are low and high frequency Gaussian noise. The kurtosis of (c) is matched to (b) and (a) by choosing the right cutoff for the high and low frequencies. Even for a fixed k and α kurtosis in invariant to the spacing and orientation of the frequency points that make up k. For $P = \cos^a(ux)\cos^b(vy)$ the vertical and horizontal spacing of the points is determined by u and v. By varying these parameters the images in (d), (e), and (f) were created all of which have the same kurtosis. Of course we should not except a single statistic to be very descriptive of an image however several of these invariants will hold for the statistics of filtered images.

4 Marginal Statistics

Given an image $f(x, y)$ and a set of filters $g = \{g_1, \ldots, g_r\}$ the marginal statistics of an image are the histograms of the subband images f_i formed by the convolution $f_i = g_i * f$. These marginal statistics have been successfully used for many applications in computational vision including texture classification and synthesis, image classification and retrieval and object detection and identification.

Several studies of the marginal statistics of natural images have shown that for small kernel mean-zero filters the resulting histograms have a single mode and are highly kurtotic (non-Gaussian). A variety of filters have been studied, including Gabor functions, wavelets, Gaussian derivative filters, Laplacian filters and random mean-zero filters [15] [16] [9] [6] [17] [18]. Many of these studies have shown that low order (two parameter) parametric models such as the generalized Laplacian and Bessel K forms provide an excellent fit to the histograms [15] [6] [25]. Because the parameters of these models are a function of variance and kurtosis the effectiveness of marginal statistics as a representation is largely governed by the ability of the second and fourth moments to discriminate between various image patterns and textures.

In the previous section, we saw that correlation in the phase angle leads to a positive kurtosis. However, many different geometries of correlation will lead to the same non-Gaussian statistics. By measuring kurtosis after applying a set of filters some of these geometries can be distinguished although as we will show some cannot.

Figure 5 is an example of the effect a filter has on the measured kurtosis of an image. Modulating two Gaussian noise patterns (top left) produces a kurtotic ($\kappa = 6$) distribution (top right). Because Gaussian noise has equal energy at all frequencies the phase correlations are spread over large distances (frequency offsets). Applying a filter (Gaussian derivative in this case) isolates a region of the frequencies, thereby eliminating any long range phase correlations outside of the response region of the filter. The result is a much smaller W_5 and a low kurtosis or Gaussian like distribution (bottom right). In general, the kurtosis of a subband image is a measure of the amount (strength) of phase correlation in the response region of the filter.

The relationship between the kurtosis of subband images and phase structure is very simple. The one characteristic common to all small filters (including random kernels) is a smooth or slowly varying Fourier Transform. Figure 6 shows an example Fourier transform of a random 5×5 filter. Consider a quadruple of frequencies that both lie on a parallelogram in frequency space and are close to each other (a local parallelogram). If prior to applying the filter these frequencies

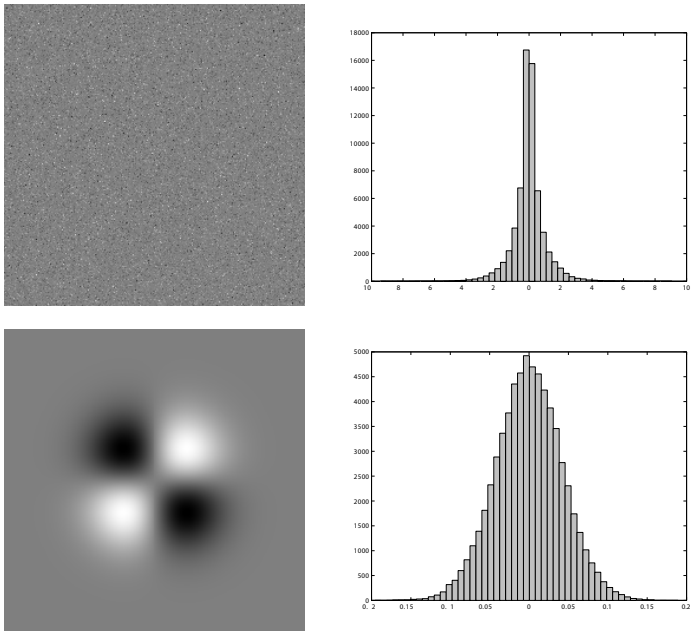


Figure 5: The top image is formed by modulating two Gaussian noise patterns resulting in a noise pattern with l large kurtosis (top right). The bottom image shows the Fourier transform of a derivative filter. Convolution of the top image with this filter restricts the phase correlations to the response region of the filter resulting in the histogram shown on the right, which is close to Gaussian

are in phase then they will remain in phase after applying the filter because the slowly varying Fourier transform of the filter will not significantly alter the phase so long as the quadruple of frequencies are close to each other. Thus, an image model composed of dense, strong and local phase correlations will give rise to the observed marginal statistics of natural images.

The role of a filter is to amplify or isolate regions of the frequency plane and the kurtosis of the filter response is an estimate of the amount of phase correlation in the amplified region. For a particular filter if the phase correlations between a pair of images is measurably different in the response region of the filter then the images are distinguishable by the histogram. By using a large bank of filters many different patterns can be identified. Several of the images in figure 4 have different marginal statistics given a large and diverse filter bank. For example, figure 4(c) has large local phase correlations in the low frequencies therefore a low pass filter would amplify the difference in kurtosis between (a) and (c). The phase correlations in (d) are more distant in frequency space than those of (f). Therefore, a filter that isolates a small region of frequency space can differentiate (d) and (f). However, kurtosis is only a single number and therefore, by varying the geometry, density and location of the phase structure in small regions, visually distinct images

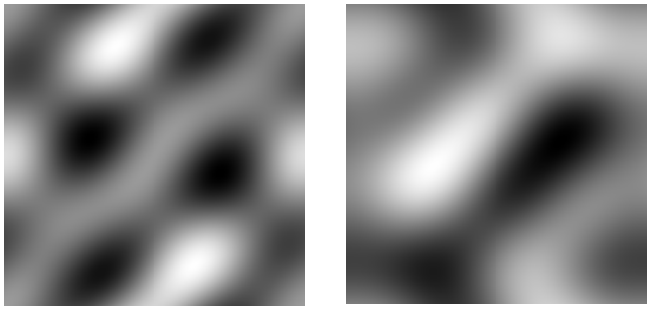
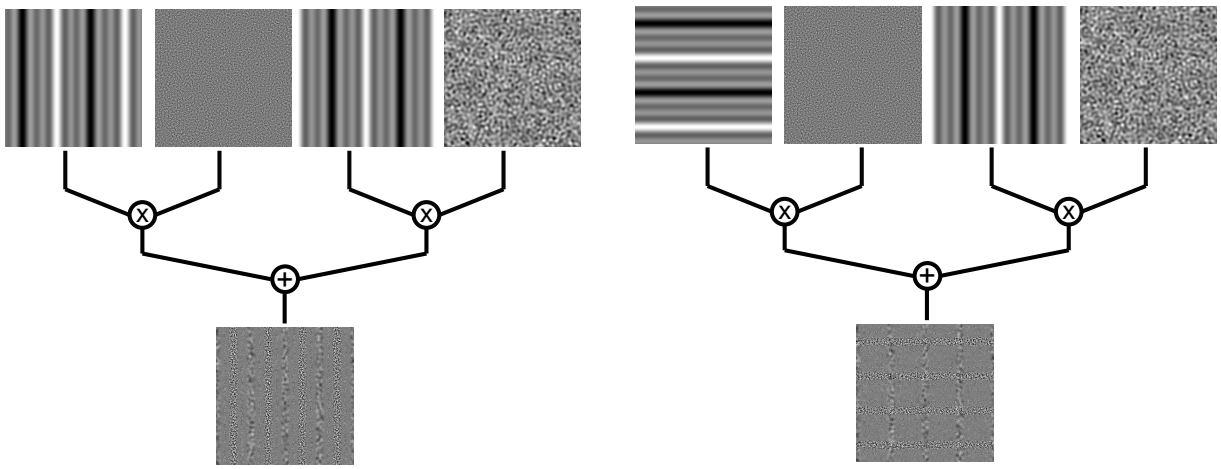


Figure 6: The real and imaginary parts of the Fourier transform of a small random filter. The smoothness is a result of the small size of the filter relative to the size of the image.

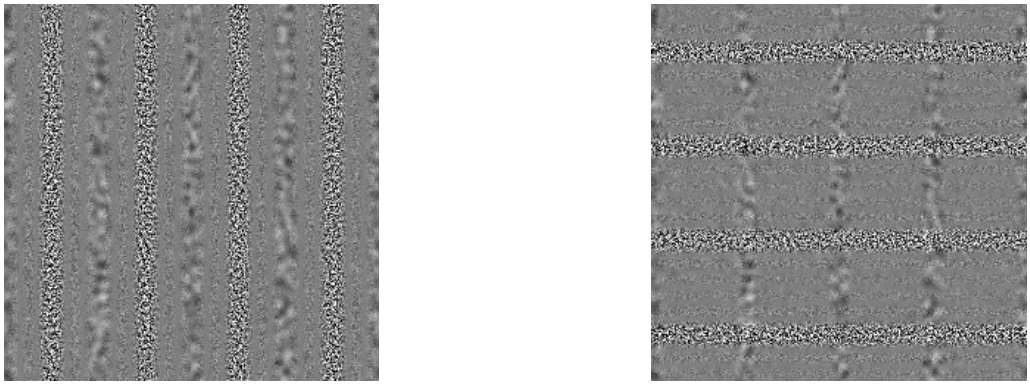
can be created that have the same kurtosis and variance for many commonly used filters. The only way to distinguish such differences from histograms alone is to use filters tuned to very small regions of frequency space. However, such filters would be very large in the spatial domain.

Figure 7 shows a pair of images that have the same kurtosis even after convolution with a large bank of filters. These patters are created from ideal high and low bandpass noise images N_{low} and N_{high}. The noise images are modulated separately by patterns P_1 and P_2, which are chosen to be de-correlated, and added together. A pair of visually distinct images I and I' are created by $I = P_1 \cdot N_{\text{low}} + P_2 \cdot N_{\text{high}}$ and $I' = P_1^{\text{T}} \cdot N_{\text{low}} + P_2 \cdot N_{\text{high}}$. P_1 and P_2 are approximately decorelated by using a translated version of the same sinusoid while P_1^{T} and P_2 are decorrelated because the sinusoids have different orientations. Because these images have the same amount of both local and global phase correlations marginal statistics have difficulty differentiating between them. In fact most commonly used filters result in nearly identical histograms (statistically indistinguishable) when applied to these images.

Vertical derivative filters at all scales have nearly the same variance, kurtosis and histograms. Figure 8(a) shows the log histogram for the largest L^2 norm between the histograms for derivative filters computed over 4 scales. They are nearly the same. In order to differentiate these images from marginal statistics a filter is needed that is finely tuned to the orientation of the phase correlations. An example of such a filter is shown in figure 8(c), which is a highly anisotropic Gabor function. The plot in (d) is the frequency log response of this function. Because the function is tuned to such a narrow band of frequencies the correlations in image I will be destroyed. However, the correlations in I' are in the direction of the filter response and will remain. This is clearly shown in the log histograms of the filtered images displayed in (b). The use of such highly anisotropic filters for visual analysis has been advocated by the ridglet and curvlet transforms [30]. Note that (c) is also similar in appearance to components that have been derived from finding a basis which



(a)



(b)

Figure 7: Visually distinct images with the same subband kurtosis are created by varying the location and geometry of phase correlations. (a) Shows the process used in creating the images in (b), which have the same marginal statistics (variance and kurtosis) for many commonly used filters. Both images have the same phase correlations in the low frequencies, while they differ in the direction of the correlations in the high frequencies. Such differences are difficult to discern from filter response histograms.

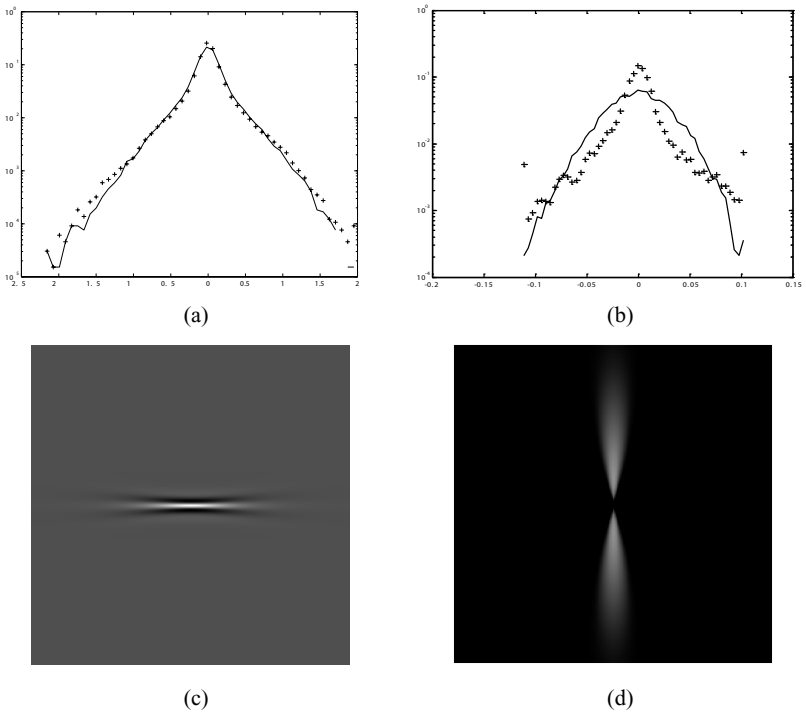


Figure 8: The two images in the preceding figure can not be distinguished by comparing histograms of vertical derivative filter responses. Derivatives at several different scales were applied. The Log histogram (solid line for the left image and '+' for the right) of the filter that provided the largest L^2 difference between the images is displayed in (a). In order to produce a robust separation between the histograms a filter must be specially constructed, such a filter is shown in (c) with the histograms of filter responses in (b). The power spectrum of the filter is shown in (d).

maximizes kurtosis in natural images [31] [32] [33].

Images can be generated that even anisotropic filters such as finely tuned Gabor functions or ridglets have difficulty distinguishing. Figure 9 is an example of such patterns. The two images $I = P_1 \cdot N_{\text{low}} + P_2 \cdot N_{\text{high}}$ and $I' = P_1 \cdot N_{\text{low}} + P_3 \cdot N_{\text{high}}$ are created using three patterns P_1, P_2, and P_3 which are all translated versions of the $\cos^a(ux)\cos^a(vy)$ function shown in (c). After filtering these images with a bank of 120 finely tuned Gabor filters [1], 24 derivative filters and 8 Gaussian filters. The histograms of all subband images were nearly identical. The largest L^2 difference was less than the difference of the raw intensity histograms which are shown in (d). In order to distinguish these images from marginal statistics an extremely anisotropic filter must be carefully designed that is only a one or two frequencies wide and practically the length of the image. Such a function hardly resembles a filter.

To summarize, the marginal statistics of filtered images can be explained by a model of phase

[1]The Gabor filters were constructed to tile the frequency plane with 3 scales and 20 orientations (9° seperations).

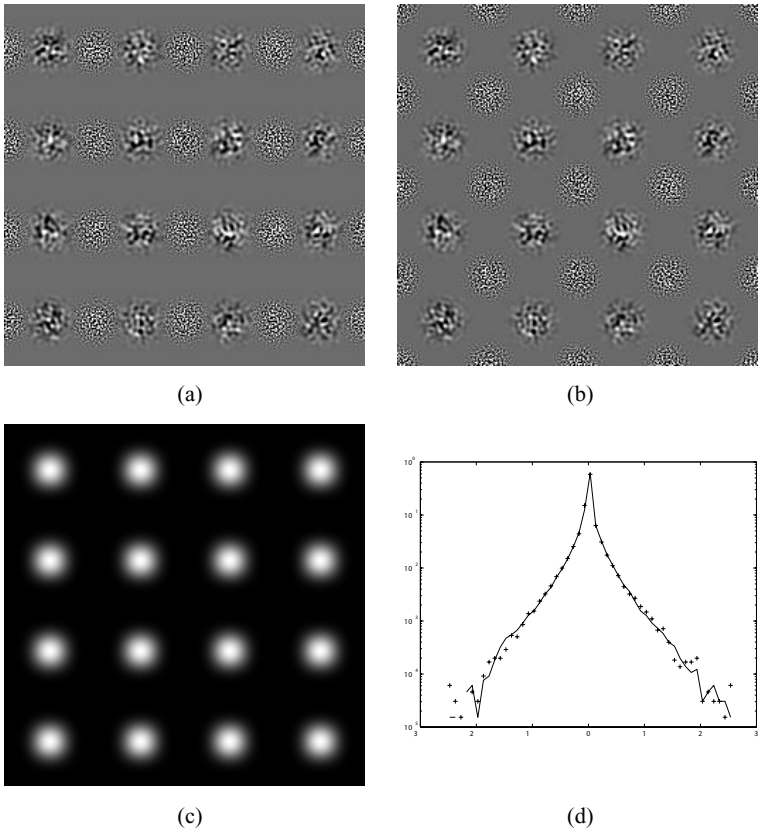


Figure 9: Images with nearly identical statistical properties. The images in (a) and (b) were constructed from high and low frequency noise modulated by translated versions of the pattern in (c). The histograms of a set of 152 filters resulted in L^2 differences all less than the raw intensity histograms (log) shown in (d).

structure which has strong local correlations throughout the frequency plane. We have also demonstrated that histograms of filter responses can easily differentiate variations in the amount of phase structure throughout the frequency plane, but have difficulty differentiating variations in local geometry that makes up the phase structure. However, local variations in the phase structure give rise to global and visually distinct differences in the spatial domain. Next, the relationship between phase structure and joint statistics is examined.

5 Joint Statistics of Images

While marginal statistics consider each subband image in isolation, joint statistics are multi-dimensional histograms constructed from a set of subband images. Several researchers have developed techniques for texture synthesis based on the joint statistics of filter outputs [7] [5]. Recently,

good results on texture classification have been obtained using "textons", which are clusters in the joint statistical space of filter outputs [34] [35] [36]. Joint distributions have also been used for object recognition, image retrieval, and segmentation [37] [2] [3].

Several studies of natural imagery have shown regularities in bi-variate distributions, a joint statistic of two filter outputs. In particular, it has been observed that joint distributions of filters with the same orientation but different scale have kurtotic distributions whose level curves are circular (or elliptic). Such distributions lead to dependencies between the magnitude of filter responses at adjacent scales and have been used for image compression [20]. Similar distributions have been found for pairs of real and imaginary Gabor filters [19]. For filters with different orientations, such as vertical and horizontal wavelets, the joint distributions exhibit polyhedral (diamond) shaped level curves [14].

We examine the relationship between joint statistics and phase structure through the multi-variate moments. The order $k = \sum_{i=1}^{r} k_i$ moments are defined as

$$\mu_{k_1 \ldots k_r} = \frac{1}{|\Omega|} \int_{\Omega} \prod_{i=1}^{r} f_i(x,y)^{k_i} d\Omega,$$

where $f_i = g_i * f$ for filters $g = \{g_1, \ldots, g_r\}$. For simplicity we consider the bi-variate moments of order 4 which are μ_{40},μ_{04}, μ_{31}, μ_{13}, μ_{22} where

$$\mu_{pq}(f) = \frac{1}{|\Omega|} \int_{\Omega} f_1(x,y)^p f_2(x,y)^q d\Omega.$$

These moments are a function of two subband images f_1 and f_2 of the image function f. Assuming that f is a whitened symmetric signal then

$$f_1 = \sum_{j=1}^{n} m_{1,j} s_j \cos(z_j) \quad \text{and} \quad f_2 = \sum_{j=1}^{n} m_{2,j} s_j \cos(z_j), \tag{10}$$

where $m_1 = \{m_{1,1}, \ldots, m_{1,n}\}$ and $m_2 = \{m_{2,1}, \ldots, m_{2,n}\}$ are the Fourier coefficients of the filters used to construct f_1 and f_2 and $s_j \in \{1, -1\}$ are the phases of f.

As previously discussed the M_5 type I terms are the most significant contributor to the non-Gaussian part of the fourth moment, the same is true for multi-variate moments. For bi-variate moments these terms are of the form

$$\mu_{31} = 3 \sum_{i,j,k,l} (m_{1,i} m_{1,j} m_{1,k} m_{2,l})(s_i s_j s_k s_l)\, T(i,j,k,l) \tag{11}$$

$$\mu_{13} = 3 \sum_{i,j,k,l} (m_{1,i} m_{2,j} m_{2,k} m_{2,l})(s_i s_j s_k s_l)\, T(i,j,k,l) \tag{12}$$

$$\mu_{22} = \mu_{20}\mu_{02} + 3 \sum_{i,j,k,l} (m_{1,i} m_{1,j} m_{2,k} m_{2,l})(s_i s_j s_k s_l)\, T(i,j,k,l), \tag{13}$$

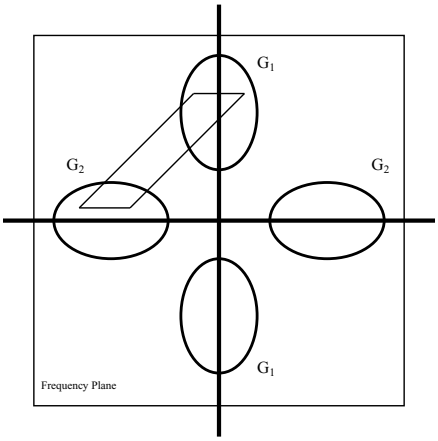


Figure 10: Most filters, such as Gabor functions and Gaussian derivatives, respond to a particular region of frequency space. For bi-variate distributions constructed from filters G_1 and G_2, which have disjoint response regions, the moments $\mu_{13} = 0$ and $\mu_{31} = 0$ because a parallelogram of four frequencies can not be formed where 3 of the vertices are in a single region. However, the μ_{22} moment measures correlation between the phase structure of the two regions because a quadruple of frequencies can be formed where two of the vertices are in each region.

where

$$T(i,j,k,l) = \begin{cases} 1 & \text{if } (u_i, v_i) - (u_j, v_j) = (u_k, v_k) - (u_l, v_l) \\ 0 & \text{otherwise} \end{cases} . \tag{14}$$

$T(i,j,k,l)$ is an indicator function that is 1 if the four frequencies form a parallelogram in frequency space. We ignore the μ_{40} and μ_{04} moments because they are equivalent to the μ_4 moment of f_1 and f_2, respectively.

Analogous to the single variate moment μ_4 only a quadruple of four frequencies forming a parallelogram will contribute to the multi-variate fourth order moments. Independent of the filters, they contribute positively if they are in phase, $s_i s_j s_k s_l = 1$, and negatively otherwise. Just like the marginal statistics the joint moments measure the amount of phase structure. However, they differ in the use of the weights, $\{m_{1,1}, \ldots, m_{1,n}\}$ and $\{m_{2,1}, \ldots, m_{2,n}\}$ which depend on the combination of filters employed. The effect of the filters is to amplify the effect of a particular group of frequencies.

For filters whose response functions do not significantly overlap the odd moments $\mu_{31} = 0$ and $\mu_{13} = 0$ independent of the both the phase structure and the image function f (see figure 10). As previously mentioned the μ_{40} and μ_{04} moments are identical to the single variate μ_4 moments therefore the only new information from the bi-variate distribution is the μ_{22} moment (among the $k = 4$ moments). When a quadruple of frequency locations contributes to this moment a pair of

frequencies must lie in each response region of the filters (see figure 10). Thus the μ_{22} moment responds to the amount of phase correlation between the two regions.

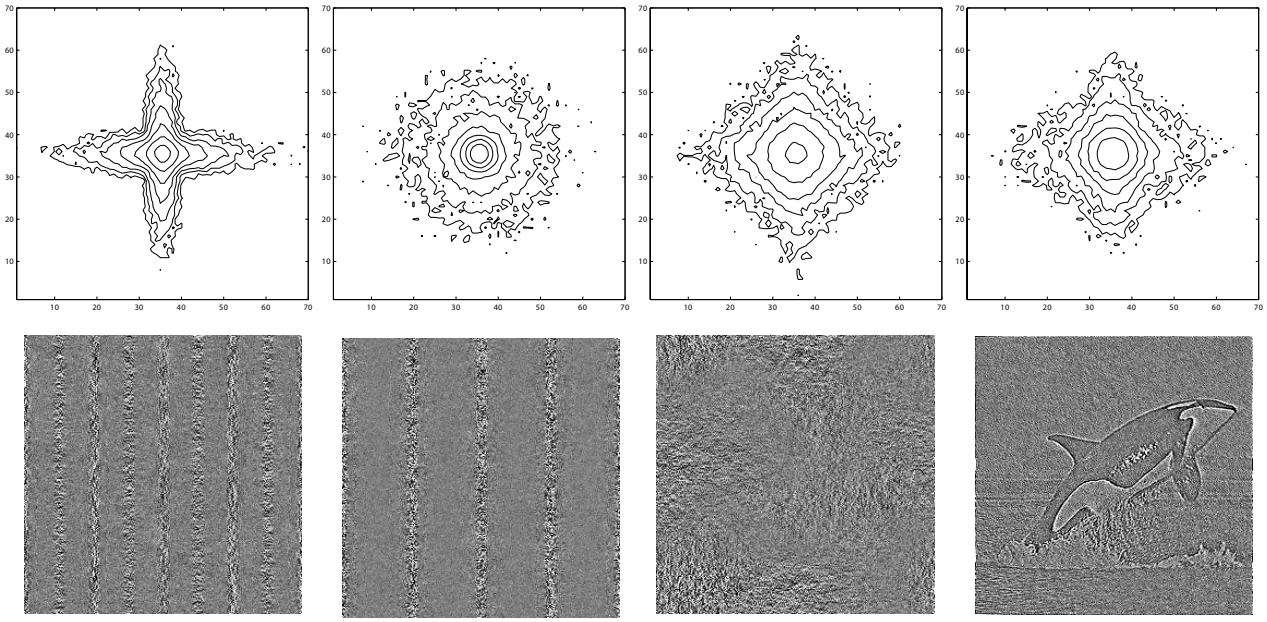
An alternate way to measure the phase correlation between two regions is to use marginal statistics from a filter whose response is just the sum of the two regions. Consider a pair of non-overlapping filters g_1 and g_2 where $f_1 + f_2 = (g_1 + g_2) * f$. The fourth moment is

$$\mu_4(f_1 + f_2) = \mu_4(f_1) + \mu_4(f_2) + 6\mu_{22}(f_1 f_2). \tag{15}$$

Clearly, joint statistics and marginal statistics are just two different ways to measure phase correlations in the frequency domain. In fact there is no additional information in the joint statistics that will aid in the discrimination among visual patterns such as those in figures 7 and 9. The distinctness of these images results from local variations in phase structure and there is no information in equations (11-13) for measuring these differences.

For jointly kurtotic distributions the shape of the level curves is determined by the μ_{22} moment. As this moment increases the shape of the level curves will move from a '+' to a diamond to a circular shape. Figure 11 shows some examples of images that have different shaped level curves in the joint distributions of derivative filters dx and dy. These images were constructed by modulating derivative filtered Gaussian noise, N_x and N_y, with different patterns P_1 and P_2. Image (a) is constructed from $P_1 \cdot N_x + P_2 \cdot N_y$, where P_1 and P_2 are "out of phase" sinusoid patterns. It is important to note that $P_1 \cdot N_x$ and $P_2 \cdot N_y$ produce "in phase" correlations in the horizontal and vertical frequency regions respectively, implying μ_{40} and μ_{04} are large. However, a quadruple of frequencies with two from each of the horizontal and vertical regions will be "out of phase" leading to $\mu_{22} << \mu_{20}\mu_{02}$. Image (b) is generated by the same process except that $P_1 = P_2$ therefore the phase structure is fully correlated between vertical and horizontal regions of the frequency plane, which leads to a circularly symmetric joint distribution with $\mu_{22} >> \mu_{20}\mu_{02}$. For image (c) randomly generated low frequency patterns are used. The use of low frequency patterns produces in phase correlations local to both the horizontal and vertical regions. However P_1 and P_2 are generated independently. Because the phase structure is independent between the vertical and horizontal regions we have $\mu_{22} \approx \mu_{20}\mu_{02}$, which leads to the familiar diamond shaped distributions found in natural images. Figure (c) is an example of such a joint distribution for a "whitened" natural image [2]. The "diamond" shaped distributions commonly observed in the joint statistics of natural images is a result of the independence of the phase correlations between regions in frequency space with different orientations. Meanwhile, the "circular" shaped distributions

[2] A "whitened" signal is used to remove the effects of the power spectrum on the joint distribution.



(a) (b) (c) (d)

Figure 11: Examples of images with different joint distributions of vertical and horizontal filter responses. See the text for an explanation of how these images are generated. (a) Out of phase correlations between the filter responses. (b) In phase correlations. (c) Uncorrelated. (d) A natural image.

commonly found in natural images is a result of the dependence of the phase correlations between regions with the same orientation but different scale (radial distance) in frequency space. Note that while an independent Gaussian distribution leads to an elliptic shaped joint distribution an independent non-Gaussian (kurtotic) distribution leads to a diamond shaped joint distribution.

The preceding discussion was intended to illustrate a simple picture of the phase structure of natural images. The simplest model that explains the bi-variate distributions found in natural imagery can be described by anisotropic correlations of phase angle that are oriented in the radial direction of frequency space. Such correlations in the radial direction lead to inter-scale dependencies of filter responses and elliptical shaped distributions. In contrast, rapidly varying correlations in the angular direction lead to independent or 'diamond' shaped joint distributions for filters tuned to different orientations. Note that any phase correlation will lead to circularly symmetric distributions for pairs of real and imaginary Gabor filters because the repsonse regions of these functions exactly overlap.

6 Conclusion

In this paper we examined the relationship between phase structure and the statistics of images. From this study we were able to show that the observed statistics of natural images can be "explained" by a model of correlations of phase. To summarize the main results:

1. The observed statistics of "whitened" natural images can be explained by strong correlations in the phase angle.
2. The observed marginal statistics of filtered images can be explained by strong *local* correlations in the phase angle.
3. The observed joint distributions of filter responses can be explained by strong anisotropic correlations of phase angle oriented in the radial direction of frequency space.

A basis that is designed to maximize kurtosis would equivalently maximize correlations of phase angle and would lead to highly anisotropic edge like functions similar to those found in ridgelet transforms or through independent component analysis. Therefore, this model of phase structure is also in agreement with these findings [30] [31] [32] [33].

Scale invariance has not been directly addressed in this paper. However, radially oriented phase correlations would lead to scale invariant kurtosis. When combined with a power-law spectral distribution both second and fourth order statistics would be scale invariant.

In addition to providing a unifying view of marginal and joint statistics of natural images, the discriminative power of these statistics has also been investigated. The main result is that marginal and joint histograms of filter responses have difficulty distinguishing changes in the local correlations of phase angle. Yet, these local changes in the frequency domain lead to global and visually distinct changes in the spatial domain.

In this paper we have concentrated on fourth order moments and bi-variate distributions. In a separate study, we have shown the existence of deterministic and visually distinct patterns that have provably identical higher order (beyond 4) moments for marginal distributions of a given filter bank [38]. We suspect that similar results can be derived for joint distributions. In fact, there is nothing to suggest that mulit-variate distributions of any size and moments of any order help in distinguishing differences in local correlations of phase angle. However, some phase correlations can be detected by applying a non-linearity and a second stage of filtering. Such "filter-rectify-filter" processing is often used for modeling complex channels of human vision [39].

References

[1] M. Mandal and T. Aboulnasr, "Image indexing using moments and wavelets," *IEEE Transactions on Consumer Electronics*, vol. 42, no. 3, pp. 557–565, 1996.

[2] B. Schiele and J. Crowley, "Recognition without correspondence using multidimensional receptive field histograms," *International Journal of Computer Vision*, vol. 36, no. 1, 2000.

[3] C. Schmid, "Constructing models for content-based image retrieval," *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2001.

[4] H. Schneiderman and T. Kanade, "A statistical method for 3d object detection applied to faces and cars," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2000.

[5] J. Portilla and E. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *International Journal of Computer Vision*, vol. 40, no. 1, pp. 49–71, 2000.

[6] E. Simoncelli and E. Adelson, "Noise removal via bayesian wavelet coding," in *International Conference on Image Processing*, 1996.

[7] J. De Bonet and P. Viola, "A non-parametric multi-scale statistical model for natural images," *Advances in Neural Information Processing*, vol. 10, 1997.

[8] D.J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America*, vol. 4, no. 12, 1987.

[9] D.L. Ruderman, "The statistics of natural images," *Network*, 1994.

[10] S.C. Zhu and D. Mumford, "Prior learning and gibbs reactionn-diffusion," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 1, 1997.

[11] D.L. Ruderman, "The origins of scaling in natural images," *Vision Research*, vol. 37, 1997.

[12] A. Lee and D. Mumford, "Occlusion models for natural images: A statistical study of scale invariant dead leaves model," *International Journal of Computer Vision*, vol. 41, no. (1,2), 2001.

[13] M.G. Thomson, "Beats, kurtosis and visual coding," *Network: Computation in Neural Systems*, vol. 12, no. 3, 2001.

[14] E.P. Simoncelli A. Srivastava, A.B. Lee and S-C. Zhu, "On advances in statistical modeling of natural images," *Journal of Mathematical Imaging and Vision*, vol. 18, no. 1, 2003.

[15] S.G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1989.

[16] Y. Wu S.C. Zhu and D. Mumford, "Filters, random-fields and maximum-entropy (frame): Towards a unified theory for texture modeling," *International Journal of Computer Vision*, vol. 27, no. 2, pp. 107–126, 1998.

[17] C. Zetzsche, "Polyspectra of natural images," in *Natural Scene Statistics Meeting*, 1997.

[18] J. Huang and D. Mumford, "Statistics of natural images and models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1999.

[19] B. Wegmann and C. Zetzsche, "Statistical dependence between orientation filter outputs used in a human vision based image code," *Proceedings of Visual Communication and Image Processing*, vol. 1360, 1990.

[20] R.W. Buccigrossi and E.P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 8, no. 12, 1999.

[21] A.V. Oppenheim and J.S. Lim, "The importance of phase in signals," *Proceedings of the IEEE*, 1981.

[22] D. Heeger and J Bergen, "Pyramid-based texture analysis/synthesis," in *Proc. of SIGGRAPH*, 1995.

[23] M. Grossberg E. Hadjidemetriou and S. Nayar, "Spatial information in multiresolution histograms," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2001.

[24] O. Faugeras and W. Pratt, "Decorrelation methods of texture feature extraction," *The IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 4, pp. 323–332, 1980.

[25] A. Srivastava, "Universal analytical forms for modelling image probabilities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, 2002.

[26] C. Nikias and A Petropolu, *Higher-Order Spectra Analysis*, Prentice-Hall, 1993.

[27] Jan-Mark Geusebroek and Arnold W. M. Smeulders, "A physical explanation for natural image statistics," in *2nd Int'l Workshop on Texture Analysis and Synthesis*.

[28] G. Sperling A. Sutter and C. Chubb, "Measuring the spatial frequency selectivity of second-order texture mechanisms," *Vision Research*, vol. 35, no. 7.

[29] A. Schofield and M. Georgeson, "Sensitivity to constrast modulation: the spatial frequency dependence of second-order vision," *Vision Research*, vol. 43, pp. 243–259.

[30] D. Donoho and A. Flesia, "Can recent innovations in harmonic analysis explain key findings in natural image statistics," *Network: Computation in Neural Systems*, vol. 12, no. 3, 2001.

[31] A.J. Bell and T.J. Sejnowski, "The independent components of natural scenes are edge filters," *Vision Research*, vol. 37, no. 23, 1997.

[32] 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, 1996.

[33] J.A. van Hateren, "Independent componet filters of natural images compared with simple cells," *Proc. of the Royal Statisticsl Society of London (B)*, vol. 256, 1998.

[34] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *International Journal of Comptuer Vision*, Dec 1999.

[35] K. Dana and S. Nayar, "Histogram model for 3d textures," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 1998.

[36] M. Varma and A. Zisserman, "Classifying images of materials: achieving viewpoint and illumination independence," in *Proceedings of European Conference on Computer Visions*, 2002.

[37] S. Konishi and A. Yuille, "Statistical cues for domain specific image segmentation," *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2000.

[38] Joshua Gluckman, "On the use of marginal statistics of subband images," in *Proceedings of the IEEE Intl. Conference on Computer Vision*, 2003.

[39] J. Beck N. Graham and A. Sutter, "Nonlinear processes in spatial-frequency channel models of preceived texture segregation," *Vision Research*, vol. 32, pp. 719–743.