9-1-2010

Out of the Depths: Image Statistics of Space, Water, and the Minuscule World

Nimit S. Dhulekar
Dartmouth College

OUT OF THE DEPTHS: IMAGE STATISTICS OF SPACE, WATER, AND THE MINUSCULE WORLD

Dartmouth Computer Science Technical Report TR2011-678

A Thesis

Submitted to the Faculty

in partial fulfillment of the requirements for the

degree of

Master of Science

in

Computer Science

by

Nimit S. Dhulekar

DARTMOUTH COLLEGE

Hanover, New Hampshire

September 2010

Examining Committee:

(chair) Richard H. Granger, Jr.

Hany Farid

Peter Winkler

Brian W. Pogue, Ph.D.

Dean of Graduate Studies

Abstract

In images of natural scenes, a consistent relationship exists between spectral power and spatial frequency. The power spectrum falls off with a form $1/f^p$ as spatial frequency f increases, with values of p approximately equal to 2. To quantify the extent to which this statistical characteristic is exhibited by other classes of images, we examined astronomical, underwater, and microscale images. It was found that this property holds for all three categories of images, although the value of p varies in the range 1.76 to 2.37. The second statistical characteristic computed was the angular spread of the power spectrum. This metric is a means to verify whether the image categories investigated tend to display more power in the horizontal and vertical orientations, akin to natural images. It was found that these image categories have primarily isotropic spectral signatures with a much reduced anisotropy as compared to natural images. Along similar lines, we introduce a new measure called the anisotropy index which compares the power in the horizontal and vertical orientations with power in oblique orientations. The statistics thus presented are for the ensemble power spectrum. We also construct 4 classifiers to distinguish between natural images and astronomical, microscale, and underwater images. The k -nearest neighbor classifier with Mahalanobis distance had the best accuracy of 70.5% on the training set and 66.9% on the test set, for correctly identifying natural images. From these classifiers, we can not only view the confusion in classification among the investigated image categories, but also the difference in statistics as compared to natural images. These classifiers also make it possible to verify that the images in a particular class display statistics similar to that of the ensemble image.

Acknowledgments

This work is the result of many minds coming together for a common goal. All of these people have shaped and molded me into not just a better computer scientist, but also a better human being.

I have to start by thanking my advisor, Professor Richard Granger. I have learnt so much from Professor Granger. He taught me the correct way to think and analyze a problem. He taught me that even though a goal may seem difficult to achieve, if you plug away at it, you can attain it. He has instilled in me the belief that nothing is out of grasp if you work hard and persevere. I'm sad that this dissertation might be our last endeavor together.

There are two awesome individuals without whom this work would not have been possible. I am deeply grateful to James M. Hughes and Daniel J. Graham. I have bothered James and Dan with the most inane questions at all times of the day and night. This dissertation is as much a testament to their guidance as it is to their patience.

I would like to thank Professor Hany Farid who rescued this thesis from its ruins. I am extremely grateful to him for his patience and recommendations that allowed me to complete this work correctly and in time. I would also like to thank Nick Foti for his valuable insights into the analysis process and correcting me whenever I went wrong. I am also grateful to Eric Kee for providing me with the code to download entire image sets from Flickr. I can't imagine the state I would have been in if I had to download those images one at a time.

I would like to thank Professor Tom Cormen for his valuable suggestions and pointers on improving the proposal and defense manuscripts. CS 110 was one of my favorite classes at Dartmouth. Every time I write a paper or prepare a presentation, I make sure to remember at least a few of Professor Cormen's guidelines. I am also thankful to Professor Devin Balkcom. His enthusiasm for research is infectious and rubs off on all his students. I must also thank all the other faculty members who I've had the opportunity

to interact with.

Dartmouth College has provided me with many opportunities to grow and develop into the person I am today. I consider these two years as one of the best experiences of my life. The most important of life's lessons that I learnt at Dartmouth is managing my time efficiently. Whether it was staying up all night to complete a final project or finding time to volunteer at the senior center, Dartmouth's rigorous schedule kept me on my toes.

The CS department at Dartmouth has so much to offer; outstanding faculty, courses that are offered nowhere else, students who go out of their way to help you and an amazing administrative staff.

This acknowledgment would not be complete without mentioning a few special friends who helped me throughout my Dartmouth experience. I have to thank Aarathi Prasad, Amina Shabbeer, Umang Bhaskar, Priya Natarajan, Paritosh Kavathekar, Ranganath Kondapally, and Geethmala Sridaran for standing by me through thick and thin.

This research is dedicated to my wonderful parents; my *Dartmouth Experience* would not have been possible without their love and support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

It is well known that natural images share important statistical regularities. Among others, these statistical regularities include the characteristic shape and slope of the spatial-frequency power spectrum [3, 7, 23], sparseness in wavelet response [8], and local contrast statistics [9]. One of the most important characteristic properties of natural scenes is approximate scale invariance: regardless of the distance between the observer and the principal objects that form a scene, the distribution of energy across spatial frequency octaves in the 2D power spectrum of the image remains relatively constant [7]. The power spectrum measures the magnitude of the contribution of different spatial frequencies to the image as a whole. It is related to the autocorrelation function, which measures correlation in the intensities of neighboring pixels. For natural scenes, the energy content across spatial frequency f falls off with a form $1/f^p$, with p approximately 2 for the power spectrum (and approximately 1 for the amplitude spectrum) [3, 7, 8, 23]. In addition, natural images have significantly more power in the horizontal and vertical orientations (hereafter referred to as cardinal orientations) in both built and natural environments, due largely to the regularity of sky/ground relations and to the presence of trees and buildings [24].

Another reason that a statistical study of natural images has become important is the

great deal of research in the computer graphics community on the creation of realistic-looking artificial images. Previously, contributions have been made to realistic modeling of plant ecosystems [5], understanding the principles of global illumination [4], scattering models for large-scale simulations [6], and modeling of buildings [17]. Consequently, it is logical to ask whether there is some inherent property or set of properties of images that could be exploited to produce artificial images that are indistinguishable from real images. We do not provide a solution to this problem but list it as a secondary motivation for the work.

Prior work with images of natural scenes has made it possible to predict statistical regularities in natural images. The ability to predict statistical regularities suggests that the inputs to the human visual system constitute a very limited subset of the space of all possible images. Indeed, the human visual system exploits the redundant nature of this subset [1, 12, 16, 21], through methods of efficient coding (e.g., sparse coding [2, 15, 19]).

Until fairly recently, research into statistical regularities was limited to only natural scenes. Few studies investigated the presence of statistical regularities in other types of image categories. Recent work has begun to investigate images of visual media, and especially art. These images seem to share certain statistical properties with natural images [13, 14, 20]. In particular, these studies have confirmed the $1/f^p$ fall-off of the power spectrum for art images. The value of p for paintings is in the region of $2.5-2.8$, which implies that paintings have a steeper slope than natural images. The reason that paintings have a steeper slope is because of the presence of more low-frequency amplitude components as compared to natural images. Additionally, such images show anisotropy in their angular power distributions, with disproportionate amounts of power in the cardinal directions. These findings suggest that humans impose on visual media the same basic spatial regularities that exist in natural scenes. One view of the presence of basic natural scene regularities in visual media is that they are a prerequisite for mak-

ing perceptible images [11]. However, if we assume that the act of transforming natural objects and scenes into 2D visual media is specifically tailored to the processing of the human visual system, some regularities specific to such images may be tied to aesthetic judgment. That is, we can search for shared characteristics of images produced for human viewing.

In this study, we investigate the statistical properties of several previously unstudied image categories: astronomical images, underwater images, and microscale images (i.e., those that are taken using microscopes and depict structures at an extremely small spatial scale). The astronomical and microscale images occupy two different ends of a hypothetical spectrum of scale, and they give some idea of the inherent structure of the natural world at that scale. As such, they are interesting to consider in relation to natural scenes, which are known to show approximate scale invariance in spectral power over a wide range of scales (though just how many scales the invariance holds for, is not known). To get a better picture of the scale invariance of the natural world (in terms of basic spatial statistics), we need to understand scaling across a hypothetical spectrum of scale, from the extrema of the very small to the very large. It is obvious that images of very large and very small scales are highly modified: they require long exposures, powerful lenses, and sensitive electronics, and they might not even depict visual-wavelength data. Moreover, the focal plane of the astronomical images is approximately at infinity, whereas the focal plane of the microscale images is roughly at zero.

The underwater images we chose to examine exhibit the important characteristic that they are natural scenes occurring at approximately the same scale as scenes that humans routinely encounter. However, canonical scene relationships such as sky/ground relations and that objects cannot occur anywhere in 3D space, are absent or greatly altered underwater. The only bias that we used in selecting these images was the absence of human subjects in the image. As our criteria was that these images are of natural

scenes of the underwater world, we removed images featuring humans.

Although it is not unreasonable to expect that these factors might alter the basic statistical regularities of images of each of these realms, we show that there are some features, such as the $1/f^p$ fall-off of the power spectrum and angular power distribution, that appear to be regular for depictions of natural structure even at extremes of spatial scale and underwater. We find that all the image categories we studied possess a spread angular spectrum without significant power concentrated at particular orientations. This result follows from prior findings, which showed that even though individual images of various types contain significant power at oblique orientations, this significant power was lost when the spectra of several images were averaged [25].

Organization. The rest of this document comprises six chapters. Chapter 2 presents the image categories studied in this work, as well as introduces the analyses performed on these image categories. Chapter 3 is an illustrative section detailing the analyses for natural images. Chapter 4 is divided into multiple subsections; each subsection discusses a particular plot and the respective characteristics for each of the three image categories. Chapter 5 presents the results of the various classifiers in distinguishing between natural images and the other three image categories, as well as how well the classifier performs on all four image categories taken together. Chapter 6 details future research directions. Finally, Chapter 7 provides concluding remarks.

Chapter 2

Methods

In this section, we describe the image categories chosen for this study and the analysis performed on these categories.

2.1 Image data

The astronomical images used in this study were collected from the National Aeronautics and Space Administration (NASA) Astronomy Picture of the Day Archive at `http://antwrp.gsfc.nasa.gov/apod/archivepix.html`, and from the Flickr groups *Deep Space Astrophotography* at `http://www.flickr.com/groups/deep_space_astrophotography/` and *Gazing into Outer Space* at `http://www.flickr.com/groups/stargazers/`. The astronomical images selected contained both dark black backgrounds with stars and other more colorful phenomena such as supernovae and gas clusters. The microscale images were collected from Nikon's Small World Galleries at `http://www.nikonsmallworld.com/gallery`, Olympus' Bioscapes Digital International Imaging Competition at `http://www.olympusbioscapes.com/`, and the Flickr group *Microscopy* at `http://www.flickr.com/groups/microscopy/`. These images were captured via optical microscopes using techniques such as phase contrast, polarized light, fluorescence, deconvolution, and interference contrast. Both

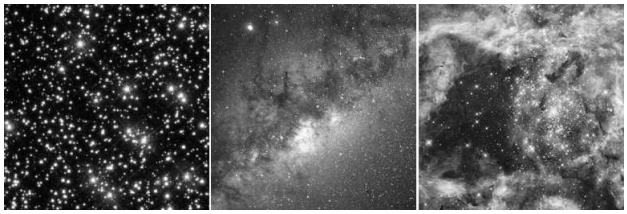
of these image categories reflect the choices of experts in the fields of astronomical and microscopic imaging and photography, and as such we believe that they are representative of each category, though neither set was obtained by purely random selection. The underwater images were collected from the Flickr groups *Underwater picture gallery* at `http://www.flickr.com/groups/underwater/` and *Underwater Indonesia* at `http://www.flickr.com/groups/underwater-indonesia/`. We chose these images because they most closely resembled the images in the two other categories in that they possess a fairly flat background, with little information beyond the primary subject of the image. The images of natural scenes were collected from the Flickr group *Natural Scenes* at `http://www.flickr.com/groups/naturalscenes/`. Unlike some of the previous work [24] we have not separated the images of natural scenes into natural and man-made objects. All the images are in Joint Photographic Experts Group (JPEG) format.

The images we analyzed were 256×256 pixels in size. Since most of the images were larger than 256×256 pixels, we first resized the smaller dimension of the image to 256 pixels while maintaining the aspect ratio. Then the image was cropped to 256×256. Our initial image data set consisted of 150 astronomical images, 162 underwater images and 203 microscale images.

Figures 2.1(a)–(c) show three exemplars from each image category. The initial data set was expanded to include 3500 images in each image category. The images were converted to gray scale before we analyzed them. We applied the following formula to every pixel in the image to convert the images from RGB format to gray scale

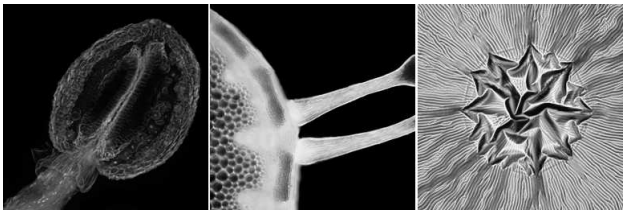
$$\text{Gray} = 0.2989R + 0.5870G + 0.1141B \ , \tag{2.1}$$

where R, G, and B are the red, green, and blue pixel channel values, respectively.



(a) Astronomical images.

(b) Underwater images.



(c) Microscale images.

Figure 2.1: Sample images

2.2 Image analysis

We first applied a pre-filtering to the input image $i(x, y)$ to reduce illumination effects and prevent local image regions from dominating the energy spectrum. The pre-filtering reduces the variance of the intensity of the pixels across the image, i.e., it reduces the regional variations in pixel intensities. This also reduced the overall variance of the image. The steps involved in this pre-filtering phase are taken from [10] and are as follows:

1. The input image $i(x, y)$ of size $N \times N$ was padded to size $P \times P$ where $P = 2N$. The original image $i(x, y)$ was placed in the top left quadrant of the new padded image.

2. The Fast Fourier Transform (FFT) $I(u, v)$ of the image was computed and centered.

3. The filter function of size $P \times P$ was generated.

4. The FFT $H(u, v)$ of the filter function was computed and centered.

5. The point-wise product $G(u, v)$ of the FFTs of the image and filter was formed, i.e., $G(u, v) = I(u, v) \times H(u, v)$.

6. The processed image was obtained as the inverse FFT $g_p(x, y)$ of $G(u, v)$.

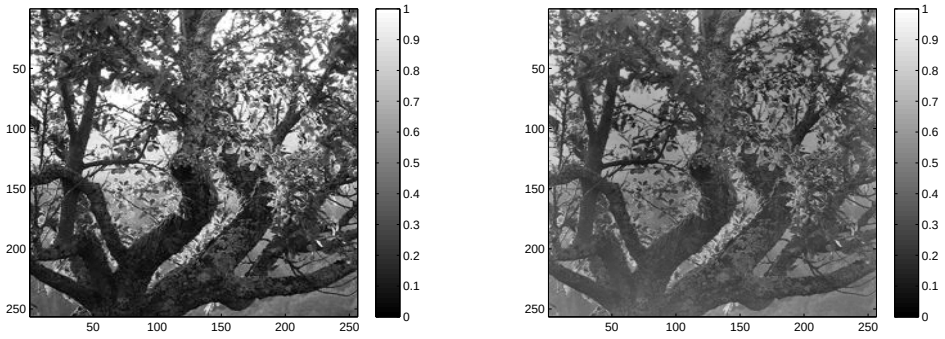
7. The final processed result $g(x, y)$ was found by extracting the $N \times N$ region from the top left quadrant of $g_p(x, y)$.

The pre-filtering function used is the following:

$$i'(x, y) = \frac{i(x, y) * h(x, y)}{\epsilon + \sqrt{(i(x, y) * h(x, y))^2 * g(x, y)}} , \tag{2.2}$$

where $*$ is the convolution operator, $i(x, y)$ is the intensity of the pixel, $g(x, y)$ is a low-pass Gaussian spatial filter with a radial cut-off frequency at 0.015 cycles/pixel, and $h(x, y) = 1 - g(x, y)$, i.e. $h(x, y)$ is a high-pass filter. ϵ is a constant that avoids noise enhancement in constant image regions and is taken as 20 for all images. The numerator is a high-pass filter that cancels the mean intensity value of the image and whitens the energy spectrum at the very low spatial frequencies. The denominator acts as a local estimator of the variance of the high-pass filter output. Figure 2.2 gives an example of the local normalization process. To verify the correctness of the pre-filtering process, we split each image into 16 regions of 64×64 and computed the variance of each region for both the original image and its filtered counterpart. We considered these 16 variances as a vector and computed the variance of this vector. We found a mean reduction in the variance from the original image to the filtered image of 0.5.

Next we applied a 256×256 Hann window to the locally normalized images to get rid of edge effects that could result in spectral leakage in the FFT spectrum. The FFT implements a Fourier transform at a discrete set of frequencies over a finite interval.



(a) Image before pre-filtering. (b) Image after pre-filtering.

Figure 2.2: Images before and after the pre-filtering process.

Thus, the FFT is not very frequency-selective. The FFT transform assumes that the finite data set is one period of a periodic signal. Windowing makes the endpoints of the waveform meet and therefore result in a continuous waveform without sharp transitions.

To generate a 2D Hann window, we take the outer product of the 1-D Hann window with its transpose. The Hann window used is displayed in Figure 2.3.

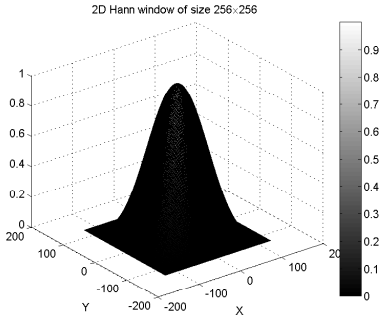


Figure 2.3: Hann window of dimension 256×256.

We computed the ensemble (i.e., average) power spectrum (or the power spectral density) for each image category via the formula

$$\mathrm{PwrSpec}(\omega, \theta) = \left(\frac{F(\omega, \theta) \cdot F^{*}(\omega, \theta)}{M} \right) , \tag{2.3}$$

where $F(\omega, \theta)$ is the value of the 2D discrete Fourier transform of an image at spatial frequency ω and orientation θ, z^* for $z \in \mathbb{C}$ denotes the complex conjugate, and M is the number of images in that particular image category. The typical parametrization of a 2D Fourier transform is given as $F(u, v)$, where u and v vary over the entire range of horizontal and vertical frequencies, respectively. Our formulation slightly differs from this form and is dependent on both spatial frequency and orientation, where $\omega = \sqrt{u^2 + v^2}$ and $\theta = \tan^{-1}(v/u)$.

We computed the spectral signature of the power spectrum to illustrate the structural aspects that the power spectrum captures. The spectral signatures were computed on the ensemble power spectrum of each image category as

$$E[PwrSpec(\omega, \theta)^2|S] \simeq \Gamma_s(\theta)/\omega^{\alpha_s(\theta)} \ , \tag{2.4}$$

where $E[I(f, \theta)^2|S]$ is the expected value of the power spectrum for a particular image category S [18]. Functions $\Gamma_s(\theta)$ and $\alpha_s(\theta)$ are obtained by a linear fitting of the ensemble power spectrum on logarithmic units for each orientation θ [25]. To calculate the values for $\Gamma_s(\theta)$ and $\alpha_s(\theta)$, we take the logarithm of Eq. (2.4):

$$\log\left(E[PwrSpec(\omega, \theta)^2]|S\right) \simeq \log\left(\Gamma_s(\theta)\right) - \left(\alpha_s(\theta)\right)^* \log\left(\omega\right) \ . \tag{2.5}$$

Eq. (2.5) is similar to the slope-intercept form of the equation of a line, $y = mx + c$. The slope of Eq. (2.5) is $-\alpha_s(\theta)$, and the y-intercept is $\log(\Gamma_s(\theta))$. To compute $\Gamma_s(\theta)$, we take the exponent of the y-intercept, i.e., $\Gamma_s(\theta) = \exp^{(\log(\Gamma_s(\theta)))}$.

The function $\Gamma_s(\theta)$ is a scaling factor revealing the dominant orientations of a scene category. The function $\alpha_s(\theta)$ represents the slope of the decreasing power spectrum values, from low to high spatial frequencies. Most of the steps to compute both the functions are the same. We first found the power contained in all the spatial frequencies along each orientation. This step is common to finding both $\Gamma_s(\theta)$ and $\alpha_s(\theta)$. To

compute $\Gamma_s(\theta)$, we found the intercept of the best-fit line to the plot of the orientations and the power, both plotted on logarithmic scales. The final $\Gamma_s(\theta)$ is obtained by taking the exponent of the intercepts calculated in the previous step and dividing the values thus obtained by the maximum value of the exponentiated intercept. To find $\alpha_s(\theta)$, we find the slope of the best-fit line to the plot of the orientations and the power, both plotted on logarithmic scales.

We define the rotational average of the power spectrum as

$$RA_F(\omega) = \frac{1}{n}\sum_{i=1}^{n} |PwrSpec(\omega, \theta_i)| \ , \tag{2.6}$$

which implies averaging the spectral power over orientation as a function of spatial frequency. We averaged over orientations θ from 0 to π for each spatial frequency ω between 0 and 127 (i.e., $M/2 - 1$) cycles per image. This particular frequency range was chosen to avoid artifacts in our analysis as a result of the pre-filtering stage [20]. We can also view the distribution of spectral energy as a function of orientation, which yields the angular average of the power spectrum, defined as

$$AA_F(\theta) = \frac{1}{n}\sum_{i=1}^{n} |PwrSpec(\omega_i, \theta)| \ , \tag{2.7}$$

where ω ranges over all spatial frequencies from 0 to $M/2 - 1$ for each orientation θ between 0 and π.

Principal component analysis (PCA) has been used in the field of vision (especially in face recognition) to produce a reduced set of orthogonal functions able to approximately represent the original function in a low-dimensional space. Image principal components (IPCs) decompose the image as

$$i(x, y) = \sum_{n=1}^{P} v_n \mathrm{IPC_n(x, y)} \ . \tag{2.8}$$

$\mathrm{IPC_n(x, y)}$ are the eigenvectors of the covariance matrix $T = \mathbb{E}\left[(i-m)(i-m)^T\right]$, where $\mathbb{E}$ is the expectation operator and m is the mean of the images. v_n are the coefficients for describing the image. P is the total number of IPCs. We used PCA to find the type of structures present in the particular image categories and verify if there is any pattern prevalent in the images.

We also computed a measure of the power distribution, which we refer to as the anisotropy index[1]. The anisotropy index gives us a percentage measure of the power concentrated in the cardinal orientations vs. the power concentrated in oblique orientations. To compute the anisotropy index, we take the power at 0 and $\pi/2$ (i.e. the cardinal power), and the power at $\pi/4$ and $3\pi/4$ (i.e. the oblique power). We sum the power at all 4 orientations and take the ratio of the cardinal power to the oblique power.

[1]The initial definition of the anisotropy index as given in the thesis proposal document, has been changed as per the suggestions by Prof. Farid

Chapter 3

Illustrative samples

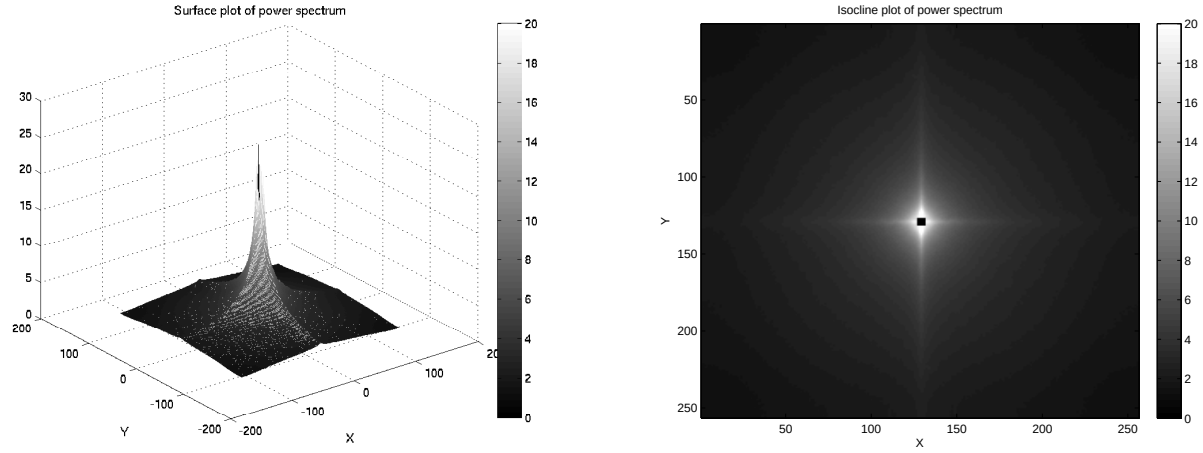
We illustrate a few of these metrics with the help of images of natural scenes (Figure 3.1). We collected 3500 images of natural scenes. The only bias while collecting these images was to have minimal human subjects in the images and only try to capture the natural or man-made structures in the scene.

Figure 3.1: Images of natural scenes.

3.1 Surface and isocline plots of natural scenes

We computed the metrics for the images of natural scenes to verify the correctness of our work. The first set of metrics we looked at were the surface and isocline plots (Figure 3.2). As can be observed from the surface and isocline plots, there is a bias towards the cardinal orientations, with little power concentrated along oblique orientations. In

particular, it has been shown that there are, on average, more cardinally oriented structures in natural images than obliquely oriented structures [22]. For the surface plot, the power spectrum has been raised to the power of 0.3. This was done to view the raised ridges along the cardinal orientations, signifying increased power concentration along these orientations. The isocline plot allows us to view the 3D structure of the surface plot in two dimensions from a top-side view. The surface plots are dominated by the central DC component which makes it difficult to view the other structures in the power spectrum. This is the reason we also view the isocline plot.



(a) Surface plot of natural scenes.

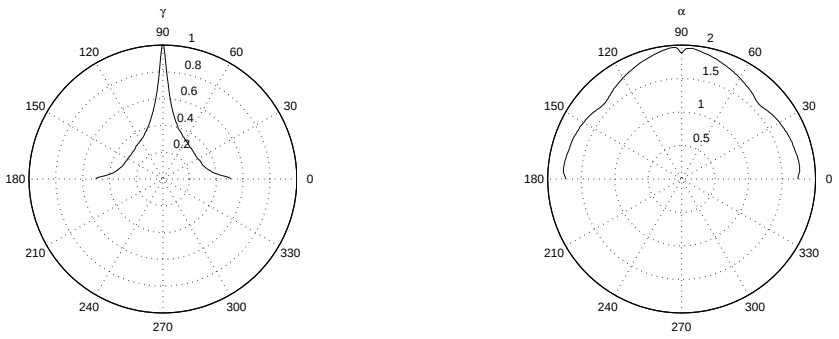
(b) Isocline plot of natural scenes.

Figure 3.2: Surface and isocline plots for natural scenes. Both the plots display significant power in cardinal orientations, which is expected given the abundance of horizontal and vertical structures in nature (e.g. trees, mountains, coastlines etc).

3.2 Spectral signature of natural scenes

The next metric we computed was the spectral signature, i.e., the $\Gamma_s(\theta)$ and $\alpha_s(\theta)$ plots. We computed the spectral signatures to illustrate structural aspects that are captured by the power spectrum [18]. $E[PwrSpec(\omega, \theta)^2]$ in Eq. (2.4) is used to approximate the value of the spectral signature. The plots are displayed in Figure 3.3. From the $\Gamma_s(\theta)$ plot we can infer that the vertical orientation is dominant and there is some bias towards

the horizontal orientation. Apart from these two orientations, the rest of the spectrum is quite flat. The second component of the spectral signature is $\alpha_s(\theta)$, which represents the slope of the decreasing power spectrum. The $\alpha_s(\theta)$ plot displays a uniform value at all orientations, with a slight drop at the cardinal orientations.

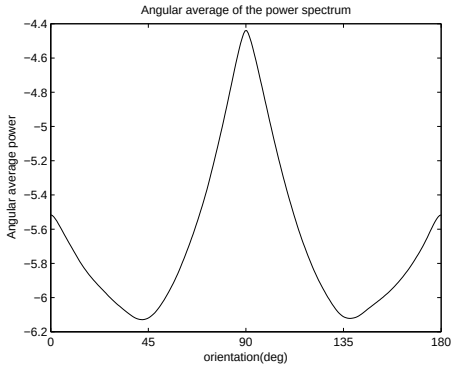


(a) $\Gamma_s(\theta)$ plot for natural scenes. (b) $\alpha_s(\theta)$ plot for natural scenes.

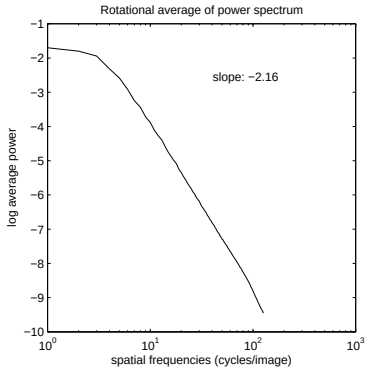
Figure 3.3: $\Gamma_s(\theta)$ and $\alpha_s(\theta)$ plots for natural scenes. The $\Gamma_s(\theta)$ plot displays a marked bias for the vertical orientation. The $\alpha_s(\theta)$ plot displays a uniform distribution with a slight drop at the cardinal orientations.

3.3 Angular and rotational average of power spectrum

The next set of metrics we investigated were the angular and rotational averages of the power spectrum (Figure 3.4). The angular average of the power spectrum, similar to the $\Gamma_s(\theta)$ plot, shows a significant peak at $\pi/2$. The slope of the rotational average of the power spectrum was computed to be 2.16, which is close to the value of 2 reported in past work. The slope is computed for the spatial frequency range 10 to 127 cycles/image.



(a) Angular average for natural scenes.



(b) Rotational average for natural scenes.

Figure 3.4: Angular and rotational averages of power spectrum of natural scenes. The angular average of the power spectrum shows a significant spike at $\pi/2$.The rotational average has a slope of -2.16.

3.4 Anisotropy index of natural scenes

The next metric we computed was the anisotropy index. The anisotropy index gives a measure of the power concentrated in cardinal orientations vs. the power concentrated in oblique orientations. The anisotropy index for our natural image set was 15.51 and the percentage power concentration in cardinal orientations was 1.63%.

3.5 Image principal components of natural scenes

We looked at the first eight IPCs of natural scenes. The IPCs in Figure 3.5 primarily show horizontal and vertical components, which is expected since natural scenes consist of more cardinal structures than oblique structures. As discussed by [8], stationarity of natural images is responsible for IPC shape which corresponds to Fourier basis.

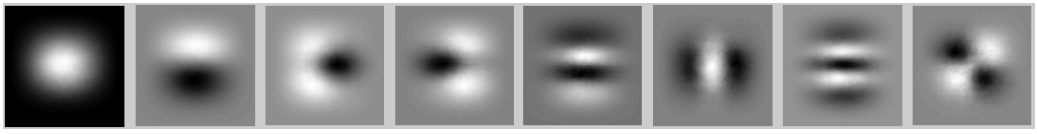


Figure 3.5: The first eight IPCs.

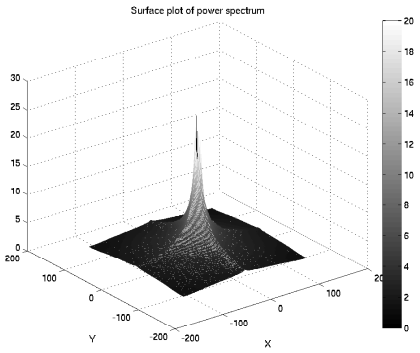
Chapter 4

Results

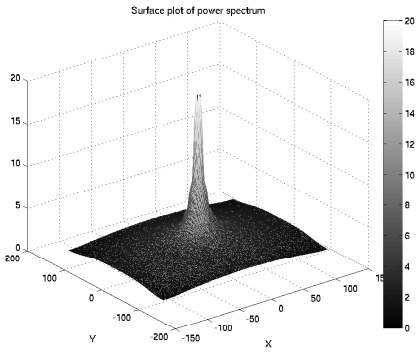
In this chapter, we show the eight types of plots described in chapters 2 and 3 for the three image categories. Section 4.1 shows the surface and contour plots. Section 4.2 presents the spectral signatures (in particular, the functions $\Gamma_s(\theta)$ and $\alpha_s(\theta)$). Section 4.3 shows the rotational and angular averages of the power spectrum. Section 4.4 introduces the anisotropy indices for the various image categories studied. Finally, Section 4.5 presents the IPCs for the three image categories.

4.1 Surface and isocline plots

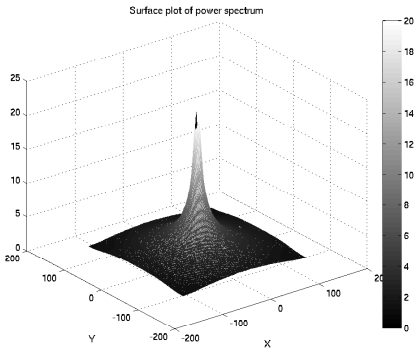
Figures 4.1(a)–(d) display the surface plots of the ensemble power spectra for all three image categories. The surface plots for all three types of images show uniform distribution of power at all orientations, a property very different from that of natural images. The surface plots though are hampered by the central DC component and hence we are unable to view the structures of the power spectrum clearly. The contours or isoclines of the power spectrum are plotted to clearly view these structural aspects of the power spectrum. The contour plots of all the image categories we studied were primarily isotropic (Figures 4.2(a)–(d)). But the microscale images display a slight bias for the cardinal orientations. The astronomical and underwater images do not display bias for



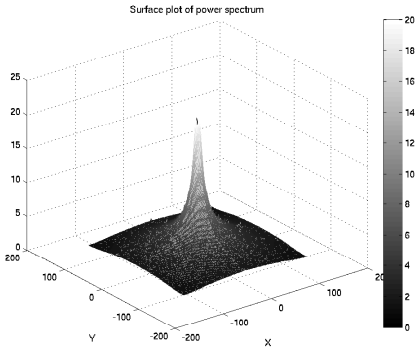
(a) Natural images.



(b) Astronomical images.

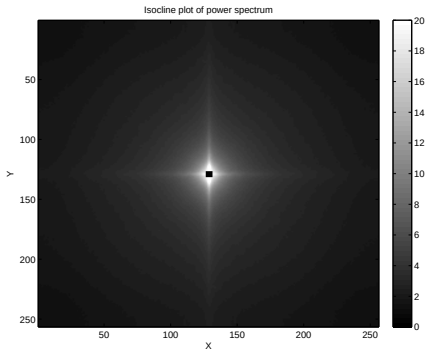


(c) Underwater images.

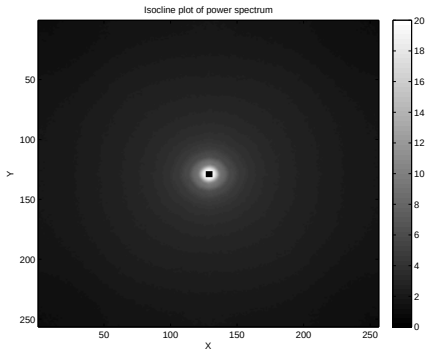


(d) Microscale images.

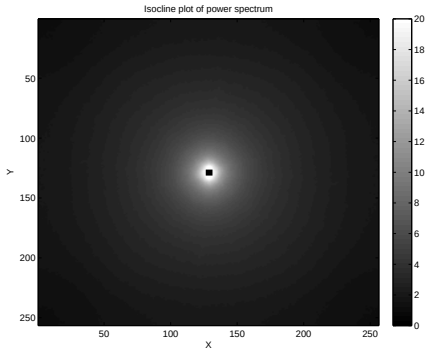
Figure 4.1: Surface plots. The surface plots for the ensemble power spectrum for the three image categories depict a 3D visualization of the power spectrum. The power spectrum is plotted along the z-axis and the spatial frequencies are plotted along the x and y axes.



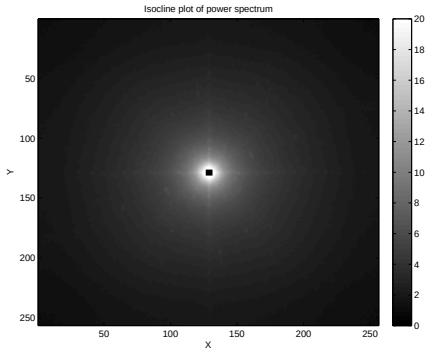
(a) Isocline plot of natural scenes.



(b) Astronomical images.



(c) Underwater images.



(d) Microscale images.

Figure 4.2: Isocline plots. The isocline plots for the three image categories depict a 2D visualization of the surface plots where the power is depicted in rings of increasing radii (or varying spatial frequency).

any of the orientations. This was an interesting observation since the expected result was that the microscale images would not have any orientation bias. The microscale images, while arguably natural, depict scenes not typically encountered by humans, and as such, humans lack an *a priori* model to understand how the microscale phenomena *should* be viewed. Therefore, it is not unreasonable to assume that the creators of these images rendered them in such a way as to match the statistics of the natural environment. Indeed, these are the very statistics that the human visual system is tuned to perceive. In microscale images, we observe more power at higher spatial frequencies as compared to the astronomical images 4.2(d)).

The underwater images also have a significant amount of power concentrated at higher spatial frequencies as compared with the astronomical images (see Figure 4.2(c)). The isocline plot of the ensemble power spectrum possesses the expected isotropic distribution of power, i.e., power is uniformly distributed over all orientations.

4.2 Spectral signatures

Figures 4.3(a)–(d) display the values for the function $\Gamma_s(\theta)$ for all three image categories. The shapes of the plots vary a bit among the image categories. For the astronomical images, we see a broad structure with value close to 1 for all orientations. For the underwater images, the peak values are concentrated closer to the vertical orientation, and the horizontal orientation values are relatively smaller. For the microscale images, we see a slight bias towards the cardinal orientations with the maximum value of about 1. This is in line with the earlier orientation bias in the isocline plot.

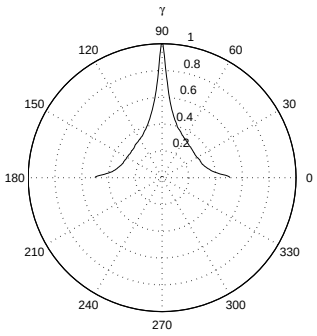
Figures 4.4(a)–(d) display the values for the function $\alpha_s(\theta)$ for all three image categories. For all three image categories, we can see that the plots are fairly uniform with a similar drop in the power spectrum as the frequency increases along each orientation. We can classify the image categories based on their slope characteristics into gradual

slope ($\alpha \sim 1.5$) and relatively steeper slope ($\alpha \sim 2$). The astronomical images belong to the first category, and the underwater and microscale images belong to the latter category.

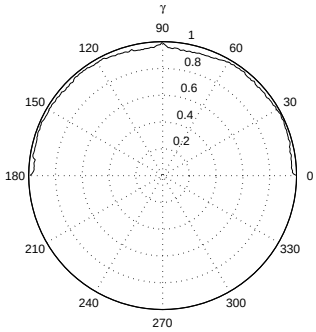
4.3 Angular and rotational average of power spectrum

The ensemble power spectrum averaged over spatial frequency ω, expressed as a function of orientation θ from 0 to π for all image categories is given in Figure 4.5. This is the angular average of the power spectrum. We are using the formulation given in prior research [25]. For all three image categories, we can observe that the angular power is well distributed over all orientations. For all image categories, there is a slight peak at $\pi/2$. This peak is not significant when considering the variance of the power spectrum values for the entire range of orientations. The underwater images have the highest variance and the astronomical images have the lowest variance in power values. For the purpose of reference, we also plot the angular average of natural scenes.

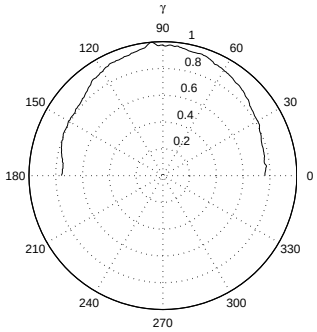
The ensemble power spectrum averaged over orientation θ, expressed as a function of spatial frequencies ω from 0 to 127, i.e., $M/2$ cycles/image for each image category is given in Figures 4.6(a)–(d). This is the rotational average of the power spectrum. Again, we are using the formulation given in prior research [25]. The log average power is plotted vs. spatial frequency (which is plotted on a log scale). The slope of the best-fit line varies from 1.72 for astronomical images to 2.37 for microscale images. This range is in accordance with the slope of the power spectrum of natural scenes. The data points lie approximately on a straight line, which means that spectral power and spatial frequency are related by a power law. This plot clearly exhibits the $1/f^p$ fall-off of the power spectrum with increase in spatial frequency. Also, it can be seen that the log average power for all the image categories have nearly similar peaks. The astronomical images have a gradual fall-off as compared to the underwater and microscale images.



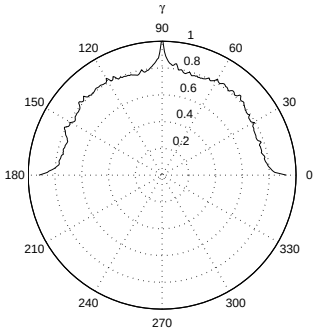
(a) Natural images.



(b) Astronomical images.

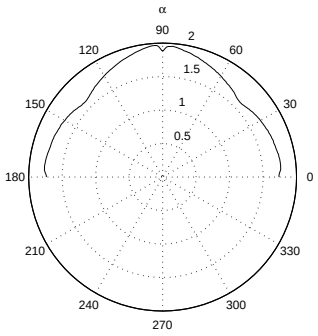


(c) Underwater images.

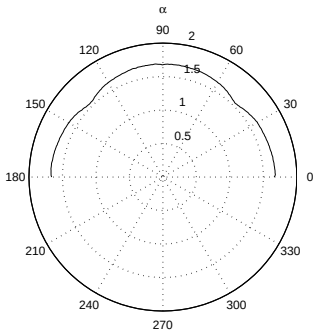


(d) Microscale images.

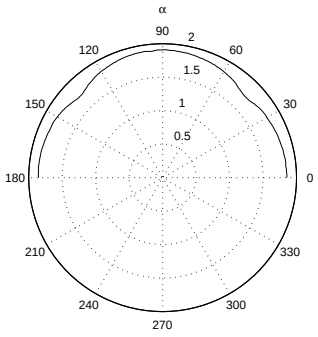
Figure 4.3: $\Gamma_s(\theta)$ plots. The $\Gamma_s(\theta)$ plot reveals the dominant orientations of an image category. The astronomical images are quite broad. The underwater images are slightly biased toward the vertical orientation. This could be because of the presence of water in the background. The microscale images display a slight bias towards the cardinal orientations.



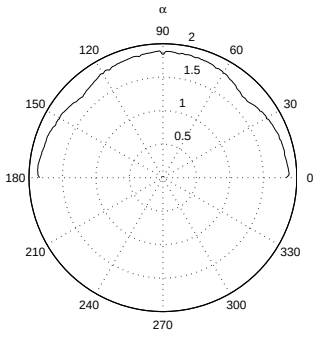
(a) Natural images.



(b) Astronomical images.



(c) Underwater images.



(d) Microscale images.

Figure 4.4: $\alpha_s(\theta)$ plots. The $\alpha_s(\theta)$ plots depict the slope of the decreasing power spectrum, with increasing spatial frequencies. It is evident from these plots that all the image categories display similar fall-off of the power spectrum for all orientations in the range 0 to π. The astronomical images have a gradual slope ~ 1.5. The underwater and microscale images have a relatively steeper slope ~ 2.

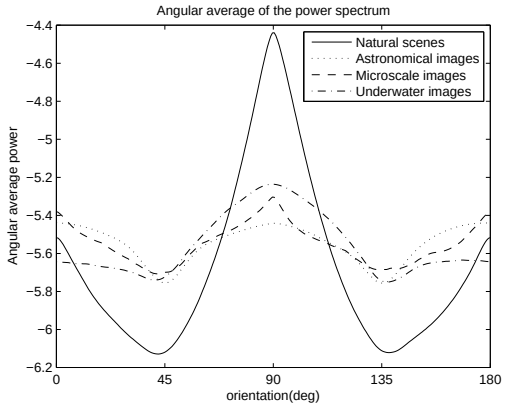


Figure 4.5: Angular average of the power spectrum. The power spectrum is averaged over all spatial frequencies, expressed as a function of orientation. The angular power is well distributed across all orientations.

	Anisotropy Index	Percentage Power in Cardinals
Microscale	3.72	0.94
Underwater	1.95	0.87
Astronomical	1.67	0.84

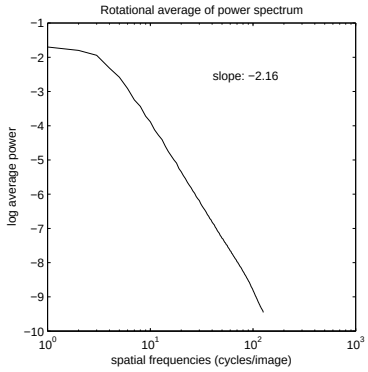
Table 4.1: Anisotropy indices and percentage power in cardinal orientations for all image categories.

4.4 Anisotropy index

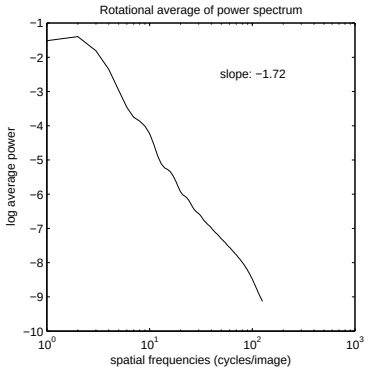
Table 4.1 displays the anisotropy index and the percentage angular power in the cardinal orientations for all image classes. From the table it can be inferred that all the image classes have little power in the cardinal orientations. The anisotropy index of the microscale images is higher than the other two image categories. This observation supports the evidence seen before regarding microscale images having a slight bias towards cardinal orientations. The percentage power for all image categories is very low.

4.5 Image principal components

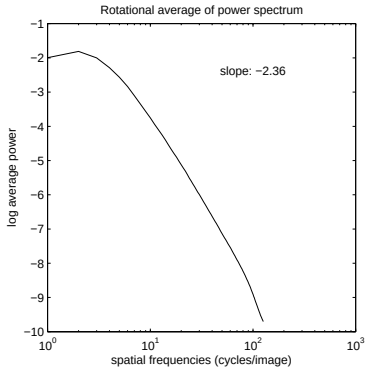
To plot the IPCs for the image categories, we consider the basis images constructed using eigen vectors for the covariance matrix for all the images from all four categories taken together. The first eight IPCs for all four image categories taken together are



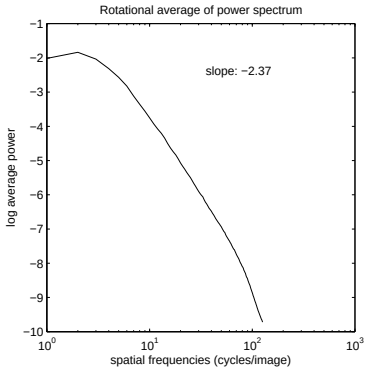
(a) Natural images.



(b) Astronomical images.



(c) Underwater images.



(d) Microscale images.

Figure 4.6: Rotational average of the power spectrum. The power spectrum is averaged over all orientations expressed as a function of spatial frequencies. The data points lie approximately on a straight line, which means that spectral power and spatial frequency are related by a power law. The underwater images have a slightly more gradual drop-off of the power spectrum as compared to the other two image categories.

displayed in Figure 4.7. The fist component shows the opposite trend to that of the first IPC of natural images. The first IPC of natural images has a light region in the middle surrounded by darker region, whereas the IPC for all the images taken together has a dark region in the center and lighter region surrounding it. The second components also display this behavior. We can also notice that the IPCs for natural images have horizontal and vertical components, whereas the IPCs for all the image categories taken together have more oblique shaped structures. From this we can infer that the image categories investigated tend to have few cardinal components and have primarily obliquely inclined structures.

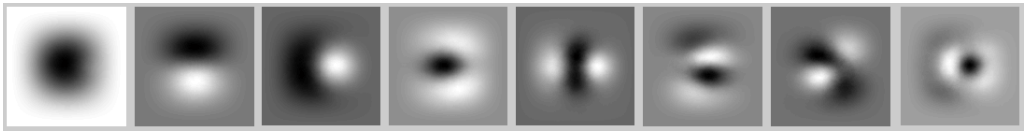


Figure 4.7: Image principal components for all image categories taken together.

Chapter 5

Classifiers

All the statistics we have computed thus far were based on the ensemble power spectrum. We have seen that the ensemble statistics for natural scenes are different from the statistics of the other three image categories. We wanted to test whether these same statistics could be used to distinguish between natural scenes and the three other image categories. To achieve this goal, we built four classifiers, 2 linear and 2 non-linear. We used the following types of classifiers

1. linear discriminant analysis (LDA),
2. k-nearest neighbor with Mahalanobis distance (knn-M),
3. Naive Bayes (NB),
4. decision tree (DT).

The features used for classification include

1. anisotropy index,
2. maximum value of $\alpha_S(\theta)$,
3. $\Gamma_S(\theta)$ values at 0 and $\pi/2$,

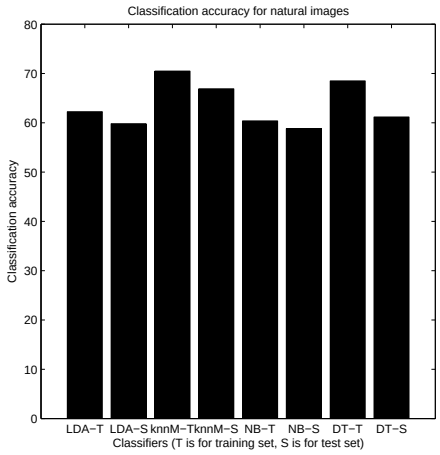
4. angular average of power spectrum at 0 and $\pi/2$,

5. slope of the rotational average

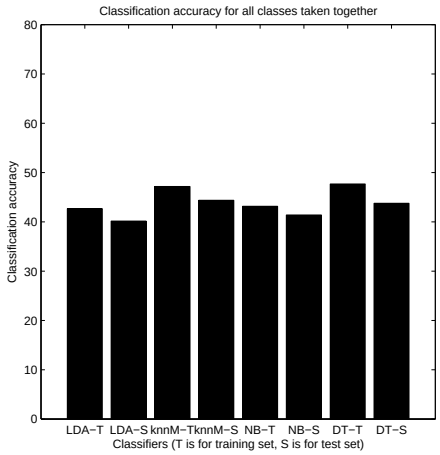
We used a training set of 2000 samples per image category (i.e., natural, astronomical, underwater, and microscale) and testing set of 1500 samples per image category. We tested the accuracy of the classifier in correctly recognizing natural images, and also the accuracy of recognizing the other three image categories. Thus we created a 4×4 confusion matrix with the accuracy rates for all four classes. We also performed 10-fold stratified cross-validation testing on the training set. The confusion matrices for the four classifiers on the training set, test set and the 10-fold cross-validated training set are presented in the Appendix A.

As was expected the non-linear classifiers, namely the k-nearest neighbor with Mahalanobis distance and decision tree, performed better than the linear classifiers, namely linear discriminant analysis and Naive Bayes. knnM had the best accuracy of 70.5% and 66.91% on the training and test sets respectively, of natural images. DT had an accuracy of 68.5% and 61.18% on the training and test sets respectively, of natural images. Although NB did not have as high classification accuracy as knnM and DT, it had the smallest fall in classification accuracy from training to test set for natural images of 4.08%. For all the classifiers, the 10-fold stratified cross-validation testing on the training set had a higher accuracy rate than the classifier where a separate testing set was used. Figure 5.1 presents the classification rates for natural scenes as well as all image classes taken together. Table 5.1 displays the fall-off rates from the training to the test set for all the classifiers for natural images, and all categories taken together.

Apart from natural images, microscale images were the only other category which had moderately higher than chance classification accuracy. This fact ties in well with the statistics we have presented throughout this document and also verifies the fact that individual images do tend to follow similar statistics to the ensemble image. The underwater and astronomical images got confused primarily with each other and this is



(a) Classification accuracy for natural images.



(b) Classification accuracy for all classes.

Figure 5.1: Classification accuracy for natural images only and for all classes taken together.

what we expected from the observations. These image categories have a spread power spectrum with no orientation bias and thus there were no features particular to either image category which could distinguish them from other categories.

A few samples of the natural images that were incorrectly classified are presented in Figure 5.2. The image misclassified as an underwater image has a butterfly in a river. The presence of water might have been the reason the image's computed statistics were similar to statistics of underwater images. The image misclassified as an astronomical image was a close-up of a flower. The statistics of this image would not have been

	Natural images fall-off	All images fall-off
LDA	3.90%	7.48%
knnM	5.09%	5.93%
NB	2.62%	4.08%
DT	10.69%	8.22%

Table 5.1

very similar to natural images since there are few cardinal oriented structures. The flower's petals are obliquely oriented and this would have been similar to statistics of astronomical images. The image misclassified as a microscale image is an image of sand with oblique lines. Again, since this image has no cardinal structures, it would not have matched the statistics of natural images. The lines towards the top left are less oblique and this could be the reason that this image got classified as a microscale image. As we have shown microscale images have a slight bias towards cardinal orientations. Perhaps this slight bias in the image allowed it to be classified as a microscale image.

Figure 5.2: Natural images misclassified as underwater, astronomical or microscale images.

Chapter 6

Future work

A future research direction for this work would be the following task. [1] The task would be to perform a psychophysical experiment to determine whether human subjects do indeed prefer images with power concentrated in the cardinal orientations as compared to images with power concentrated in oblique orientations. The experimental setup would consist of 40 images, 20 having power primarily concentrated in the cardinal orientations, and 20 having power concentrated in oblique orientations. This set of images will give 400 image pairs. The images will be from both the astronomical and microscale image classes. Human subjects will be shown two images together, one from each class. One of the images will have power primarily concentrated in the cardinal orientations, whereas the other image will have power concentrated primarily in oblique orientations. The human subjects will then be asked to choose one of the images based on preference. A few correlation tests will be run to verify which of the two types of images the human subjects preferred. It is expected that the human subjects will prefer the images with cardinal power.

The classifier results we present are encouraging but not entirely satisfying, especially in the case of astronomical and underwater images. The confusion between these

[1] This experiment was part of my thesis proposal. The proposal committee recommended dropping this experiment in favor of the classifier.

two classes merits further analysis of features that might lead to better disambiguation between them. It is possible that using certain oblique angles might lead to a better classification between astronomical and underwater images.

We have investigated three image categories and expanded this field a bit more. There are quite a few image categories that have not been investigated yet and it is possible some of them might share frequency statistics with natural images.

Chapter 7

Conclusions

In this work, we showed that the power spectra of highly disparate categories of images such as astronomical, underwater, and microscale images, display statistical regularities such as the $1/f^p$ fall-off of the power spectrum. These regularities are similar to those present in natural scenes. These three image classes have isotropic spectral signatures since they have few cardinal structures; cardinal structures are a prerequisite for an anisotropic and cardinally-oriented spectral signature. The microscale images have a slight bias towards the cardinal orientations and we show this with the help of multiple statistics. We compute various metrics such as isocline plots, spectral signatures, image principal components etc. We also present a new metric, coined anisotropy index, which measures the ratio of power concentrated in cardinal orientations vs. power concentrated in oblique orientations. Finally, we show that using linear and non-linear classifiers, we are able to distinguish between natural images and the three image categories. The best classification accuracy of 70.5% and 66.91% on the training and test sets respectively was achieved using k-nearest neighbor with Mahalanobis distance.

Appendix A

Classification matrices

	Natural	Astronomical	Microscale	Underwater
Natural	62.25	8.55	16.35	12.85
Astronomical	13.60	30.65	24.75	31.00
Underwater	10.75	21.40	44.30	23.55
Microscale	8.80	30.85	26.85	33.50

Table A.1: Classification matrix for LDA on training set.

	Natural	Astronomical	Microscale	Underwater
Natural	59.82	12.64	15.81	11.73
Astronomical	14.46	28.18	28.90	28.46
Underwater	8.18	24.91	41.18	25.73
Microscale	12.00	31.09	25.46	31.45

Table A.2: Classification matrix for LDA on test set.

	Natural	Astronomical	Microscale	Underwater
Natural	61.15	10.65	15.90	12.30
Astronomical	14.60	29.35	27.00	29.05
Underwater	9.75	22.50	43.35	24.40
Microscale	11.85	27.90	28.10	32.15

Table A.3: Classification matrix for LDA on 10-fold stratified cross-validated training set.

	Natural	Astronomical	Microscale	Underwater
Natural	70.50	6.45	12.70	10.35
Astronomical	14.50	33.55	20.05	31.90
Underwater	14.85	17.65	48.20	19.30
Microscale	15.25	25.35	22.90	36.50

Table A.4: Classification matrix for knn-M on training set.

	Natural	Astronomical	Microscale	Underwater
Natural	66.91	5.46	16.46	11.17
Astronomical	16.09	31.36	22.37	30.18
Underwater	17.91	16.09	45.64	20.36
Microscale	15.36	28.27	22.73	33.64

Table A.5: Classification matrix for knn-M on test set.

	Natural	Astronomical	Microscale	Underwater
Natural	68.95	5.50	14.10	11.45
Astronomical	14.60	32.75	22.55	30.10
Underwater	16.25	17.60	46.90	19.25
Microscale	15.90	26.40	22.30	35.40

Table A.6: Classification matrix for knn-M on 10-fold stratified cross-validated training set.

	Natural	Astronomical	Microscale	Underwater
Natural	60.40	10.75	16.50	12.35
Astronomical	14.95	36.70	19.05	29.30
Underwater	16.85	18.75	42.35	22.05
Microscale	16.05	28.35	22.35	33.25

Table A.7: Classification matrix for NB on training set.

	Natural	Astronomical	Microscale	Underwater
Natural	58.82	8.54	18.64	14.00
Astronomical	14.73	34.36	21.00	29.91
Underwater	15.55	21.90	40.09	22.46
Microscale	15.82	29.64	22.18	32.36

Table A.8: Classification matrix for NB on test set.

	Natural	Astronomical	Microscale	Underwater
Natural	60.10	7.70	18.50	13.70
Astronomical	13.05	35.65	21.80	29.50
Underwater	16.40	20.90	41.15	21.55
Microscale	16.10	27.45	23.85	32.60

Table A.9: Classification matrix for NB on 10-fold stratified cross-validated training set.

	Natural	Astronomical	Microscale	Underwater
Natural	68.50	6.30	15.20	10.00
Astronomical	13.60	38.10	21.00	27.30
Underwater	14.40	21.10	47.40	17.10
Microscale	15.50	26.75	21.05	36.75

Table A.10: Classification matrix for DT on training set.

	Natural	Astronomical	Microscale	Underwater
Natural	61.18	12.46	15.27	11.09
Astronomical	15.73	36.55	20.00	27.72
Underwater	15.46	22.09	44.27	18.18
Microscale	16.73	27.55	22.63	33.09

Table A.11: Classification matrix for DT on test set.

	Natural	Astronomical	Microscale	Underwater
Natural	65.18	8.46	14.27	12.09
Astronomical	13.54	37.15	22.00	27.31
Underwater	15.26	20.09	46.27	18.38
Microscale	15.89	26.64	22.38	35.09

Table A.12: Classification matrix for DT on 10-fold stratified cross-validated training set.

Bibliography

[1] J. J. Atick and A. N. Redlich. What Does the Retina Know About Natural Scenes? *Neural Computation*, 4(2):196–210, 1992.

[2] A. J. Bell and T. J. Sejnowski. The “Independent Component” of Natural Scenes are Edge Filters. *Vision Research*, 37(23):3327 – 3338, 1997.

[3] G. J. Burton and I. R. Moorhead. Color and Spatial Structure in Natural Scenes. *Applied Optics*, 26(1):157–170, 1987.

[4] P. Debevec. Rendering Synthetic Objects into Real Scenes: Bridging Traditional and Image-based Graphics with Global Illumination and High Dynamic Range Photography. In *Proc. SIGGRAPH ’98*, pages 189–198, 1998.

[5] O. Deussen, P. Hanrahan, B. Lintermann, R. Měch, M. Pharr, and P. Prusinkiewicz. Realistic Modeling and Rendering of Plant Ecosystems. In *SIGGRAPH ’98*, pages 275–286, 1998.

[6] C. Donner, J. Lawrence, R. Ramamoorthi, T. Hachisuka, H. W. Jensen, and S. Nayar. An Empirical BSSRDF Model. *ACM Transactions on Graphics*, 28(3):1–10, 2009.

[7] D. J. Field. Relations Between the Statistics of Natural Images and the Response Properties of Cortical Cells. *Journal of the Optic Society of America A*, 4:2379–2394, 1987.

[8] D. J. Field. What is the Goal of Sensory Coding? *Neural Computation*, 6(4):559–601, 1994.

[9] R. A. Frazor and W. S. Geisler. Local Luminance and Contrast in Natural Images. *Vision Research*, 46(10):1585–1598, 2006.

[10] R. C. Gonzalez and R. E Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.

[11] D. J. Graham. *The Relationship Between Efficient Coding of Natural Scenes in the Human Visual System and Statistical Regularities in Art*. PhD thesis, Cornell University Graduate School, 2008.

[12] D. J. Graham, D. M. Chandler, and D. J. Field. Can the Theory of "Whitening" Explain the Center-surround Properties of Retinal Ganglion Cell Receptive Fields? *Vision Research*, 46(18):2901 – 2913, 2006.

[13] D. J. Graham and D. J. Field. Variations in Intensity Statistics for Representational and Abstract Art, and for Art from the Eastern and Western Hemispheres. *Perception*, 37:1341–1352, 2008.

[14] D. J. Graham and D. J.Field. Statistical Regularities of Art Images and Natural Scenes: Spectra, Sparseness and Nonlinearities. *Spatial Vision*, 21(1-2):149–164, 2007.

[15] P. O. Hoyer and A. Hyvärinen. Sparse Coding of Natural Contours. *Neurocomputing*, 44-46:459–466, 2002.

[16] S. B. Laughlin. A Simple Coding Procedure Enhances a Neuron's Information Capacity. *Zeitschrift fr Naturforschung C - A Journal of Biosciences*, 36(9-10):910–912, 1981.

[17] P. Muller, P. Wonka, S. Haegler, A. Ulmer, and L. V. Gool. Procedural Modeling of Buildings. In *SIGGRAPH '06*, pages 614–623, 2006.

[18] A. Oliva and A. Torralba. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.

[19] B. A. Olshausen and D. J. Field. Natural Image Statistics and Efficient Coding. In *Network: Computation in Neural Systems*, volume 7, pages 333–339, 1996.

[20] C. Redies, J. Hasenstein, and J. Denzler. Fractal-like Image Statistics in Visual Art: Similarity to Natural Scenes. *Spatial Vision*, 21(1-2):97 – 117, 2007.

[21] M. V. Srinivasan, S. B. Laughlin, and A. Dubs. Predictive Coding: A Fresh View of Inhibition in the Retina. *Proceedings of the Royal Society B: Biological Sciences*, 216(1205):427–459, November 1982.

[22] E. Switkes, M. J. Mayer, and J. A. Sloan. Spatial Frequency Analysis of the Visual Environment: Anisotropy and the Carpentered Environment Hypothesis. *Vision Research*, 18(10):1393 – 1399, 1978.

[23] D. J. Tolhurst, Y. Tadmor, and C. Tang. Amplitude Spectra of Natural Images. *Ophthalmic Physiological Optics*, 12(2), April 1992.

[24] A. Torralba and A. Oliva. Statistics of Natural Image Categories. In *Network: Computation in Neural Systems*, pages 391–412, August 2003.

[25] A. van der Schaaf and J. H. van Hateren. Modelling the Power Spectra of Natural Images: Statistics and Information. *Vision Research*, 36(17):2759 – 2770, 1996.