Image Statistics and their Applications in Computer Graphics

Tania Pouli[1], Douglas W. Cunningham[2] and Erik Reinhard[1]

[1]Department of Computer Science, University of Bristol, UK
[2]Brandenburg Technical University Cottbus, Cottbus, Germany

Abstract
The statistics of natural images have attracted the attention of researchers in a variety of fields and have been used as a means to better understand the human visual system and its processes. A number of algorithms in computer graphics, vision and image processing take advantage of such statistical findings to create visually more plausible results. With this report we aim to review the state of the art in image statistics and discuss existing and potential applications within computer graphics and related areas.

1. Introduction

The field of natural image statistics studies images and their statistical regularities [Gei08, HHH09]. The human visual system has evolved in natural environments, i.e. those without man-made structures. Human vision, it is argued, has therefore evolved to be particularly good at observing natural scenes, as opposed to random images [van98a]. By studying the input to the human visual system, it is thought that the human visual system can be understood better. This has led to a rich literature on various aspects of natural image statistics [Rud94, HM99b, DF01].

The typical process by which natural image statistics are computed, involves collecting statistics of ensembles of images after passing each image through a particular transform. The transforms typically employed tend to correspond to a component of human vision. It is, for instance, possible to compute the power spectra of each image, and from these calculate the average power spectral slope for the whole set. This state-of-the-art report reviews this and many other transforms and statistics.

Aside from learning about human vision and the implications for perception, many of these statistical regularities have transferred to computer graphics and adjacent areas: several applications now successfully employ statistics of natural images to help solve engineering problems. This generally takes two forms. Statistical regularities of a particular ensemble can be used to ensure that the solution peoduced by an algorithm will approach that ensemble.

Alternatively, statistical properties of single images are computed and manipulated to achieve certain effects. As an example, deblurring algorithms exist that rely on optimisation involving priors that help push candidate images towards visually plausible solutions [SJA08]. Of course, visually plausible solutions are those that appear to humans in some sense natural. Image priors have therefore been designed with the aid of specific knowledge of natural image statistics, in this case regarding the average distribution of image gradients.

We think that the use of natural image statistics has so far shown great promise, with interesting applications making use of some of the available statistical regularities. We also see excellent scope for further future developments. In this report, we review the current state of natural image statistics, and give an overview of how statistics have been used in computer graphics, computational photography, visualisation and image processing so far.

Section 2 presents issues regarding image capture and calibration and discusses possible correction mechanisms. First order statistics, as well as their relevant findings and applications are discussed in Section 3 while second and higher order statistics are shown in Sections 4 and 5 respectively. Colour related statistics are presented separately in Section 6. Finally, we summarise the findings in Section 7.

2. Data Collection and Calibration

The statistical analysis of natural images typically involves the collection of large ensembles. Various image data sets are already available but in order to create new ensembles, sev-

eral points need to be considered. One might need to create new ensembles if the intended application involves a specific class of images, or if the study is to understand human vision under a specific viewing conditions.

Advances in camera technology have made the capture of higher resolution and higher dynamic range imagery possible. Each such type of image presents its own advantages in terms of information and its own specific issues regarding capturing and calibration. The type of image appropriate for each study largely depends on the type of information studied. In this report we will focus on the statistical analysis of 2-dimensional images and as such exclude from our discussion 3-dimensional data (such as range images or 3D scans) or video sequences. More specifically, we will consider LDR and HDR imagery. Issues relating to the collection of large image ensembles will be discussed in Section 2.1.

Before analysing the collected images, it is prudent to remove artifacts and irregularities that are due to the camera, lens or settings used to capture the scenes. These can include noise, geometric distortions due to imperfections of the lens or variations due to changes in illumination or even weather conditions. The calibration process and pre-processing steps necessary to account for such irregularities will be covered in Section 2.2.

2.1. Image Capture

A number of issues need to be considered before capturing images for an ensemble. The following largely depend on the aims of the study:

- **Type of imagery:** For most purposes, low dynamic range (LDR) imagery is sufficient, although HDR images can capture a wider range of luminance values, making them more appropriate for scenes of more extreme lighting (such as night scenes where artificial light sources are much brighter than the rest of the scene). Panoramas or wider angle images can capture a larger part of the field of view, or even the full scene in the case of spherical panoramas.
- **Number of images:** Although generally, larger data sets afford more accurate statistical findings, in the case of image statistics, the trade-off between number of images and data collection time often needs to be resolved. The number of images appropriate for a particular study largely depends on its aims. Some example dataset sizes are given in Table 1.
- **Camera:** Most digital cameras available on the market offer high resolutions. However, single lens reflex (SLR) cameras offer better image quality as well as a higher level of control.

HDR Capture High dynamic range scenes can be captured in a number of ways. The most commonly used approach involves capturing a series of differently exposed versions of the scene, known as exposure bracketing, that are then merged into the HDR image. The merging process was first proposed by Mann and Picard [MP95]. Several modern digital SLR cameras offer automated settings for bracketing, greatly simplifying the process. The number of brackets offered typically ranges from 2 or 3 to a maximum of 9 in higher range models of each brand. Table 2 shows the number of brackets and the maximum exposure step for a selection of cameras offering more than 3 brackets.

Recently, cameras capable of directly capturing HDR content have become available. Specialised cameras exist that are capable of capturing $360°$ spherical images. Some examples are the SpheroCamHDR by Spheron VR, which uses a single-line CCD and offers up to 26 f-stops of dynamic range, and the Civetta by Weiss AG, which uses a full frame sensor and a fish-eye lens to capture up to 30 f-stops of dynamic range. A more detailed account of available HDR cameras can be found elsewhere [RWPD05].

2.2. Post-processing and Calibration

When measuring the statistical regularities of an image ensemble, some care is necessary to ensure that irregularities and artifacts due to the camera, lens or shooting conditions do not affect the results. To avoid such an effect, the properties of the equipment used need to be accounted for. Moreover, to ensure the images are calibrated and therefore pixel values between them are directly comparable, properties of the scene need to be measured.

2.2.1. Lens Aberrations

The lens is responsible for focusing the light coming from the scene to an image plane (which may be a sensor in digital cameras or the film). For many applications, it is sufficient to model the camera as a simple pinhole whereby no lens is present and the aperture through which the light enters is a point. To model the various effects of an optical system to the image however, more complex lenses need to be considered.

Lenses can consist of a single element, in which case they are referred to as *simple lenses* or multiple element, called *compound lenses*. For a detailed description and model of lenses the reader is referred to [Wil94,SHL95,RKAJ08]. Assuming Gaussian optics (lenses are arranged along a single reference axis, and all light rays make only small angles ϕ with this axis, so that $\sin(\phi) \approx \phi$ and $\cos(\phi) \approx 1$), the relationship between the focal length of a lens f, the distance to an object s_0 and the distance behind the lens where the image is formed s_i is given by:

$$\frac{1}{f} = \frac{1}{s_0} + \frac{1}{s_i} \qquad (1)$$

Imperfections in the design or construction of such lenses can cause them to violate the assumption of Gaussian optics. These deviations are known as *aberrations* and can

Study	Image Type	Number of Images	Resolution
[ZL97]	8 bit CCD Video Frames	100	256 x 256
[RCC98]	Spectral Images	12 scenes x 43 images	192 x 165
[vv96]	8 bit CCD Video Frames	276	768 x 512
[HLM00]	Range Panoramic Images	205	444 x 1440
[HM99c]	Calibrated Grayscale Images (from [vv98])	4000	1024 x 1536
[DLAW]	HDR Spherical Panoramas	95 (Teller:2003) + 9 (Debevec)	various

Table 1: *Dataset sizes and resolutions from various image statistics studies.*

Camera Model	Brackets	EV Step	Camera Model	Brackets	EV Step
Canon 1D MKII / MKII N	3, 5 or 7	3	Olympus C5050 Zoom	3 or 5	1
Canon 1D MKIII	2, 3, 5 or 7	3	Olympus C5060 Wide Zoom	3 or 5	1
Canon 1Ds MKII	2, 3, 5 or 7	3	Olympus C730UZ	3 or 5	1
Canon 1Ds MKIII	2, 3, 5 or 7	3	Olympus C765UZ	5	1
			Olympus C-8080	3 or 5	1
Casio EX-F1	3 or 5	2 or 1	Olympus E-1	3 or 5	1
			Olympus E-3	3 or 5	1
Fuji S5 Pro	2 to 9	1	Olympus E-30	3 or 5	1
			Olympus SP-310	3 or 5	1
Kodak P712	3 or 5	1	Olympus SP-320	3 or 5	1
Kodak P880	3 or 5	1	Olympus SP-350	3 or 5	1
			Olympus SP-510UZ	3 or 5	1
Konica Minolta 7D	3 or 5	0.5	Olympus SP-550UZ/SP-560UZ	3 or 5	1
Leica M9	5 or 7	2	Panasonic DMC-L1	3 or 5	1
			Panasonic Lumix DMC-G1 / GH1	7	2/3
Nikon D200	2 to 9	1	Panasonic Lumix DMC-GF1	3, 5 or 7	2/3
Nikon D300	2 to 9	1	Panasonic Lumix DMC-G1K	3, 5 or 7	2/3
Nikon D300s	2 to 9	1			
Nikon D700	2 to 9	1	Pentax K7	3 or 5	2
Nikon D2H	2 to 9	1	Pentax K10D	3 or 5	2
Nikon D2X/D2Xs	2 to 9	1	Pentax K20D	3 or 5	2
Nikon D3	2 to 9	1			
Nikon Coolpix 5400	3 or 5	1	Samsung GX-10	3 or 5	2
Nikon 5700	3 or 5	1	Samsung GX-20	3 or 5	2
Nikon 8080	3 or 5	1			
Nikon 8400	3 or 5	1	Sony Alpha A-700	3 or 5	0.7
Nikon 8700	3 or 5	1	Sony DSC-R2	9	1
Nikon 8800	3 or 5	1			

Table 2: *The number of brackets and the maximum exposure step for cameras offering more than 3 brackets in Auto Exposure Bracketing mode (AEB). (Adapted from [AEB10].)*
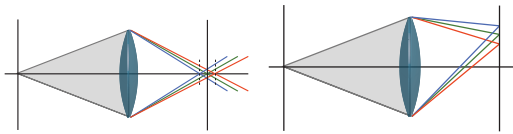
Figure 1: *Longitudinal (left) and lateral (right) chromatic aberrations are shown (adapted with permission from [RKAJ08]).*

be broadly categorised as chromatic, where the optical system behaves differently for different wavelengths, and monochromatic, where the distortions in the image are apparent in all channels.

Chromatic Aberrations Lens elements can have wavelength-dependent indices of refraction, resulting in different colours focusing in slightly different places. This causes *longitudinal* or *axial* chromatic aberration, which appears as chromatic image blur [BL09] (see Figure 1). Typically this form of aberration is managed optically, using a pair of thin lenses with different refractive indices, known as *achromatic doublets*. Such systems can reduce longitudinal aberration effects but not always completely remove them.

Another type of chromatic aberration is *lateral* chromatic aberration, which is the result of different wavelengths hitting the image plane at slightly different positions. This is more commonly known as *colour fringing* and can be reduced digitally. This can typically be achieved by realigning the colour channels of the image [MW07] which can compensate for the effect of the colour shifting to some extent, or for a more accurate reconstruction, a more complete characterisation of the optical system can be utilised [Kan07].

One particular case of chromatic aberration specific to digital cameras, especially ones using a charge coupled device sensor (CCD), is known as *purple fringing*. This effect is caused by a combination of factors that operate in addition to the lens aberrations described earlier. Highlights in the scene that are overly bright can cause some of the CCD quantum-wells to flood, creating the visual effect of blooming. Additionally, a false colour effect can be created by the demosaicing process [KP08].

A number of digital solutions are available for correcting the effects of chromatic aberrations in images with various levels of accuracy. DxO Optics handles chromatic aberrations by taking into account the characteristics of the lens where possible [DxO]. A similar approach is also found in PTLens [ePa]. Adobe Photoshop (version CS4 tested) provides simple sliders to account for red/cyan and blue/yellow shifting.

Monochromatic Aberrations A wider range of artifacts in images appears as the result of monochromatic aberra-

tions. These include blurring aberrations and geometric distortions. Blurring aberrations can be further divided into *spherical aberration*, *coma*, *astigmatism* and *Petzval field curvature*.

Spherical lenses are commonly used in optical systems as they are relatively easy to manufacture, but their shape is not ideal for the formation of a sharp image. Spherical aberrations are the consequence of light rays further from the lens axis (marginal) focusing at a different position than rays passing through the middle of the lens. When marginal rays focus nearer the lens than the principal rays, the effect is known as *positive* or *undercorrected* spherical aberration, while a marginal focus further away from the lens than the paraxial focus causes *negative* or *overcorrected* spherical aberration [van].

Comatic aberrations cause objects further from the optical axis to appear distorted. *Positive coma* occurs when marginal rays focus further away from the optical axis than principal rays, while *negative coma* is the oposite effect, namely the marginal rays focusing closer to the optical axis.

A third type of blurring aberration is known as *astigmatism* and is the result of off-axis rays of different orientations focusing at slightly different positions. The *meridional plane* for an off-axis object point is defined by that point and the optical axis. The plane orthogonal to the meridional plane is known as the *saggital plane*. Rays along these two planes focus at different positions on the optical axis. As such, if the lens is focused such that the meridional image is sharp, blur will occur in the saggital direction and vice versa.

A final aberration considered here is a consequence of the *field curvature*. The image and object planes are generally considered to be planar. However, in the case of spherical lenses they are in fact spherical, which is known as *Petzval field curvature*. Regions of the image further from the optical axis increasingly deviate from the planar assumption. This is more evident when a planar image plane is used (such as a sensor) in which case the image is less sharp towards the periphery [RKAJ08].

The effect of blurring aberrations can be reduced in postprocessing by sharpening the image or during the image capture by using smaller aperture settings and thus only allowing rays closer to the optical axis to enter. Optical solutions do exist and generally involve the use of lens doublets or carefully designed lenses. These solutions however, complicate the design of the optical system and are as a result found in more expensive lenses.

A different class of monochromatic aberrations consists of radial distortions. These cause straight lines in the scene to appear curved in the resulting image and can be corrected digitally either manually or using information about the characteristics of the lens that was used. Radial distortions are the result of deviations from a linear projection model and are more pronounced for wider angle lenses.
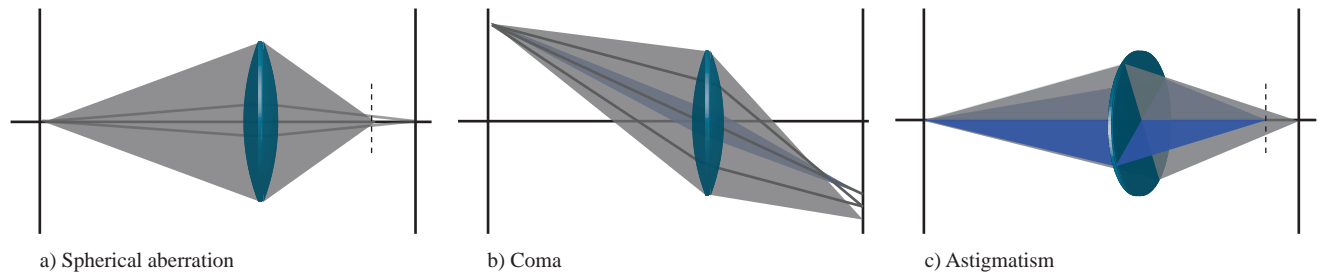
a) Spherical aberration   b) Coma   c) Astigmatism

Figure 2: *Three of the four blurring aberrations.*

Figure 3: *This figure demonstrates the effect of barrel (middle) and pincushion (right) distortions on an image. In this case, the distortions are exaggerated and overlaid with a grid for demonstration purposes. The undistorted image is shown left.*

Although interactions between the various lens elements of a complex optical system can produce some less well defined distortions, the two main types considered here are *barrel* and *pincushion* distortions, shown in Figure 3. Barrel distortion is the result of decreasing lateral magnification and it causes straight lines to curve away from the image center. Pincushion distortion occurs in lenses with increasing lateral magnification and causes straight lines to curve towards the image center [WCH92].

Radial distortions can be adequately estimated as a fourth degree polynomial:

$$x' = x(1 + \kappa_1 r^2 + \kappa_2 r^4) \qquad (2)$$

$$y' = y(1 + \kappa_1 r^2 + \kappa_2 r^4) \qquad (3)$$

where $r^2 = x^2 + y^2$, and $\kappa_{1,2}$ are the *radial distortion parameters* [Sze10]. Several methods have been proposed in the literature for estimating the distortion parameters of Equation (3). In the simplest case, by photographing a test scene containing straight lines, the distortion parameters can be adjusted until the lines in the resulting image are also straight [EMF03, BL09]. This process can of course be simplified even further if the lens characteristics are provided by the manufacturer.

More complex solutions have also been proposed, combining image alignment with distortion parameter estimation [SK99, Ste97] or by recovering the radial distortion characteristics together with intrinsic camera parameters [CF05, Stu05, HK07].

Barrel and pincushion distortions are not the only possible geometric distortions. The interactions of the various lens elements and their resulting distortions can result in effects that require more complex solutions. In these cases, approaches using calibrated targets can still be of use.

Various software packages offer solutions that reduce or remove the effect of radial distortions. Options such as DxO Optics or PTLens use lens-specific information while other solutions exist that require more manual input, such as for instance Hugin or the Lens Correction module in Adobe Photoshop.

Vignetting Vignetting is a common artifact appearing in photographs, which causes areas of the image further from the center to appear darker. Although this effect often goes unnoticed and sometimes can even be added to images for artistic purposes, the magnitude differences it causes can be large. It has been found that for certain optical systems, the amount of light reaching the corners of the image can be several times less than the light going through the center [ZLK10]. Such differences may affect the results in statistical or computer vision applications.

Several factors contribute to vignetting [BL09]. *Mechanical vignetting* can be caused by lens extensions or accessories that physically block parts of the lens. *Optical vignetting* arises when rays arrive from an oblique angle. Parts of the lens may be obstructed by the rim of the barrel, reducing the effective aperture and thus allowing less light to enter the lens for these angles. Finally, a third cause of vignetting is known as *natural vignetting* and is the effect of natural light falloff. The severity of this effect depends on the

angle θ that light rays exit the rear element and hit the image sensor and for simple lenses can be modeled as $\cos^4(\theta)$ falloff [GC05].

Both optical and natural vignetting can be reduced digitally after the image is captured. The simplest approaches for estimating the vignetting for a given optical system involve capturing a photograph of a uniformly coloured and illuminated surface [BL09, Sze10]. Using such an image, the vignetting function can be computed and its effect can then be removed from subsequent images taken using the same system.

A limitation of this approach is that it requires such a calibration image to be captured with the same camera and lens as the image(s) to be corrected [AAB96, KW00]. This may not always be possible. To that end, information from a single image can be used to reduce the effects of vignetting. This can be achieved by locating slow intensity variations in the radial direction [ZLK10]. Alternatively, in cases where multiple overlapping images are available (for instance when capturing a panorama), corresponding pixels that appear in more than one images can help recover the vignetting function [LS05, GC05].

2.2.2. Noise

Noise can be a major cause of image degradation and can be defined as any random variation in brightness or colour in the image caused by the sensor, circuitry or other parts of the camera. Noise can have various sources, each requiring different treatment. Some common types of image noise will now be discussed.

Fixed Pattern Noise Fabrication errors can cause different pixels of a sensor array to have different responses to constant illumination. This is known as *fixed pattern noise* and appears in two forms. *Dark signal non-uniformity* (DSNU) is the fixed offset from the average across each pixel of the image sensor, occurring even when no illumination is present. *Photo response non-uniformity* (PRNU) is the variation in the responsivity of pixels under constant illumination.

Fixed pattern noise can depend on camera settings and conditions such as exposure time and temperature but can be removed, either in the camera hardware or in post-processing. In the latter case, DSNU can be removed simply by capturing and subtracting an image with no illumination (for instance by keeping the lens cap on) in otherwise the same conditions as the images to be calibrated. To additionally remove noise due to PRNU under various illumination levels, an image of a flat, uniformly lit surface can be captured with the illumination set such that it nearly saturates the sensor pixels. Then using the DSNU calibration image described previously and the new image, the response of each pixel can be linearly interpolated for a particular set of conditions (temperature, camera settings, etc.) [Vis07].

Dark Current Even if no light reaches the sensor, electrons can still reach the quantum-wells, generating *dark current shot noise*. The severity of this type of noise is affected by temperature and consequently can be reduced by cooling the sensor [RKAJ08].

Photon Shot Noise The number of photons collected by each individual quantum-well of an image sensor is not constant for a given intensity. Due to this randomness, a given pixel can receive more or fewer photons than the average. The distribution of photons arriving at the sensor can be modeled by Poisson statistics (mean μ equals variance σ^2). As such, if the intensity of the source increases, so will the variance in the signal. Since the signal-to-noise ratio can be defined as $SNR = \mu/\sigma = N/\sqrt{N} = \sqrt{N}$ where N is the mean number of photons hitting the image sensor, large values of N cause the *SNR* to be large too, thus reducing the effect of this type of noise.

Most of the noise sources discussed do not depend on illumination levels in the scenes. As such, in darker scenes, the signal-to-noise ratio becomes smaller, effectively amplifying the effect of noise in the image. Although dark current and fixed pattern noise can be minimised though the use of appropriate calibration, other noise sources (including photon shot noise) cause less predictable patterns. The reader is referred to the work by Healey and Kondepudy [HK94] for a detailed account on noise estimation and camera calibration.

2.2.3. Radiometric Calibration

In order to map pixel values to real-world radiometric quantities, the response function of the camera, which is the mapping between sensor illuminances to pixel values, needs to be known. Although this information is not generally available, the non-linearities for a particular camera can be recovered through the use of calibrated input.

The relationship between scene radiance and pixel values can be recovered in two ways and the process is known as *camera characterisation*. Specialised equipment can be used, such as for instance a monochromator and a radiance meter, to measure the response of the camera to particular wavelengths [RKAJ08]. More commonly, a specific colour target (such as the GretagMacbeth ColorChecker Chart) can be used in conjunction with a spectrophotometer to match XYZ values to the RGB values captured by the camera [Fai07].

Several approaches exist for finding the transformation between the measured XYZ values and the captured RGB values. Look-up tables store the data of such measured/captured pairs. If a large enough number of samples is collected so that the camera gamut is densely populated, then XYZ values corresponding to RGB samples not in the look-up table can be estimated using three-dimensional interpolation [RKAJ08]. Alternatively, if the number of samples available is not sufficient for interpolation, a function can

be fitted to the available data. This is however much more computationally expensive.

2.2.4. HDR Merging

If the accurate representation of the scene illumination is necessary, a single exposure may not be sufficient to cover the full dynamic range in a scene. A series of exposures can be captured as described in Section 2.1, which can then be merged into a single high dynamic range image. Each individual exposure follows the non-linearities of the camera. As such, before they can be merged into an HDR image, the different exposures need to be linearised using the response curve of the camera. The techniques described above could be used to recover this response curve.

However, a sequence of differently exposed images can be used directly to recover the response curve. A commonly used method is proposed by Debevec and Malik [DM97]. By exploiting a property of both physical film and digital sensors known as *reciprocity* (meaning that halving the irradiance hitting the sensor E and simultaneously doubling the exposure time Δt will result in the same pixel values p), a linear system is derived. The solution to that is the inverse function g^{-1} to the camera response curve. Other methods for recovering the response curve of a camera are known [MP95, MN99, RBS03].

After the response curve is obtained, the image irradiance E_n for a given pixel of an exposure n can be computed given the exposure time Δt_n and the value of that pixel p_n as follows:

$$E_n = \frac{g^{-1}(p_n)}{\Delta t_n} \qquad (4)$$

If a sufficient number of exposures is collected, each part of the scene should be correctly exposed in at least one of the images. The irradiance value for a given pixel p in the HDR image can then be computed as a weighted average of the corresponding pixels of each exposure.

Several weighting functions have been proposed in the literature. Mann and Picard [MP95] merge the exposures using *certainty functions* computed from the derivatives of the response curves for each differently exposed image. A simpler approach is used in [DM97] where a hat function suppressing values that are likely to be under- or over-exposed is applied. Finally, inspired by signal theory, Mitsunaga and Nayar [MN99] propose a weighting function that takes into account the signal-to-noise ratio. Despite their differences, any of these approaches will yield satisfactory results.

Many software packages are capable of merging exposures into an HDR image. Photoshop offers a 'Merge to HDR' function since its CS2 version that allows some manual control over the image alignment. Other packages include HDRSoft's Photomatix [HDR] and HDRShop [Deb], both of which offer batch processing functionality. Finally, a free option that is only available for Apple computers is Greg Ward's Photosphere [Any], which is also an excellent viewer for HDR image collections.

3. First Order Statistics

The simplest regularities in images that can be explored through statistical analysis are properties of single pixels. The distribution of pixel values in images is typically represented with histograms. For a given image I, we define its histogram H with B bins of width V as follows:

$$H = \{(h(1), v(1)), ..., (h(B), v(B))\} \qquad (5)$$

$$B = \left\lceil \frac{\max(I) - \min(I)}{V} \right\rceil \qquad (6)$$

$$h(i) = \sum_{p=1}^{N} P(I(p), i), \quad i \in [1, B] \qquad (7)$$

$$v(i) = \min(I) + (i-1)V \qquad (8)$$

$$P(I(p), i) = \begin{cases} 1 & i = \left\lfloor \frac{I(p) - \min(I)}{V} + 1 \right\rfloor \\ 0 & \text{otherwise} \end{cases} \qquad (9)$$

where H is the set of all pairs $(h(i), v(i))$ for all $i \in [1, B]$ corresponding to the number of elements and value of the ith bin of the histogram. $I(p)$ is the value of the pth pixel of image I which contains a total of N pixels and $P(I(p), i)$ represents the probability of a pixel $I(p)$ belonging to a bin i.

First order statistics do not capture relationships between pixels and as such are not suitable for studying spatial regularities within images. Figure 4 shows an example of two very different images that would result in identical first order statistics. The right image is constructed by randomly permuting the pixels of the left image, yet it appears unnatural. Statistical analysis of the distribution of pixel intensities can however lead to interesting insights.

Statistical moments are commonly employed to quantitatively describe the shape of a distribution. The kth moment of a distribution can be computed as follows:

$$m_k = \sum_{p=1}^{N} \frac{I(p) - c)^k}{N} \qquad (10)$$

where c can be any constant. Generally, if $c = 0$ then the above equation computes the raw moments of the distribution, while setting $c = \mu$ gives us the central moments (i.e. centered at the mean). The first raw moment is the mean μ of the distribution and the second is the variance σ^2.

The meaning of further moments is less straight forward but the skewness S and kurtosis κ of a distribution relate to the third and fourth moments respectively. More specifically, the skewness and the kurtosis are defined as:

$$S = \frac{m_3}{\sigma^3} \qquad (11)$$

$$\kappa = \frac{m_4}{\sigma^4} \qquad (12)$$

Figure 4: *The right image was created by randomly permuting the pixels of the image on the left, resulting in identical first order statistics.*

respectively, where m_3 and m_4 are the third and fourth central moments respectively.

Huang and Mumford analysed more than 4000 greyscale images of natural scenes (taken from the database created by J. H. van Hateren [vv98]) by computing central moments of the log histograms [HM99c]. The values found for the mean, standard deviation, skewness and kurtosis were respectively $\mu = 0$, $\sigma = 0.79$, $S = 0.22$ and $\kappa = 4.56$. The value of the skewness shows that the distribution is not symmetrical, which can be attributed, at least partly, to the presence of sky in many of the images, resulting in a bias towards higher intensities. In addition to that, the values of both the skewness and the kurtosis show that the distribution is non-Gaussian.

A less skewed distribution was found by Brady and Field [BF00] in their analysis of 46 logarithmically transformed natural images, while the linear images resulted in a distribution skewed towards darker values. Although no exact values were provided for the moments of the distributions, Figure 5 shows the resulting histograms for both cases.

As can be seen from the above examples, although in both cases natural images were used and analysed, the results vary. Generally, the distribution of log intensities for natural images does not deviate far from symmetric. Results do depend however on the choice of images.

One particular case of interest was explored by Dror et al. [DLAW] where high dynamic range spherical panoramas were analysed. Two separate datasets were used consisting of 95 images taken from the Teller database (http://city.lcs.mit.edu/data) [TAB*03] and 9 from Debevec's Light Probe Image Gallery (http://www.debevec.org/Probes/) [DHT*00]. Figure 6 shows the resulting log intensity histograms for the two datasets. The values found for the standard deviation,
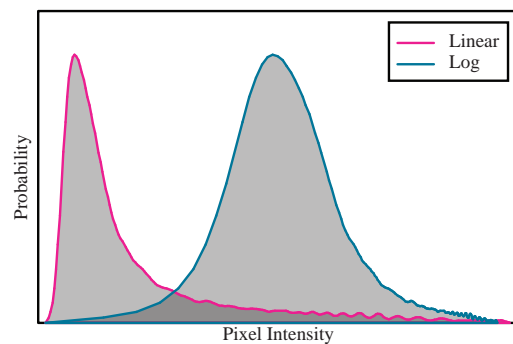

Figure 5: *The linear and log intensity histograms found by Brady et al. from their analysis of 46 natural images. (Adapted from [BF00].)*

skewness and kurtosis of the distributions respectively for the Teller dataset were $\sigma = 1.04$, $S = 0.02$, $\kappa = 4.04$ and for the Debevec dataset $\sigma = 1.32$, $S = 0.36$, $\kappa = 12.49$.

The shape of a histogram can also be associated with particular qualities of the depicted surface. Motoyoshi et al. [MNSA07] studied the perceived surface qualities of various materials and analysed their relation to the associated histograms. They found that materials that are perceived as glossier tend to have a positively skewed distribution while matte surfaces result in a negatively skewed distribution. The correlation found between the surface properties and the skewness of the histograms is shown in Figure 7.

3.1. Histogram Adjustments

Despite their simplicity, first order statistics have now found several applications in image processing. Studies have shown correlations between first order statistical regularities
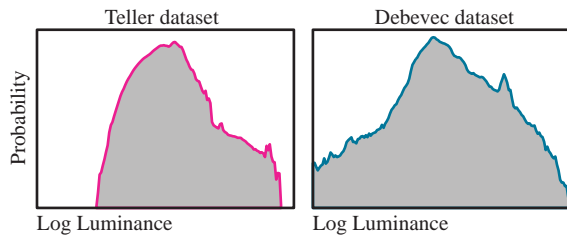
Figure 6: *The log luminance histograms found by Dror et al. for the Teller (left) and the Debevec (right) datasets. (Adapted from [DLAW].)*
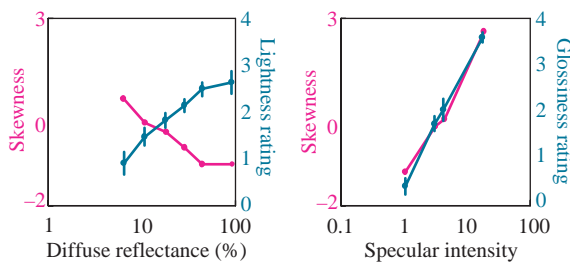

Figure 7: *As the diffuse reflectance of the surface increases, the lightness rating as perceived by the observers also increases while the corresponding histogram skewness decreases(left). An increase in specular intensity also results in an increased rating of glossiness as well as higher skewness value (right). (Adapted from [MNSA07].)*

in images and properties of the illuminant [RKAJ08,APS98] which has proved useful in areas such as white balancing. Moreover, transferring statistical moments between images in appropriate colour spaces has been demonstrated in what is now known as colour transfer [RAGS01]. These colour-related application are discussed in detail in Section 6.

First order statistics can also be computed on a single image basis. By manipulating the distribution of values of a single image, a variety of effects can be achieved. For instance, the contrast of an image that only covers a small portion of the available range of intensities can be increased by adjusting the pixel values such that the full range of intensities is more equally represented. This is known as *histogram equalisation* and an example can be seen in Figure 8.

A more general version of histogram equalisation is known as *histogram matching* and it allows the histogram of a source image (I_s) to be matched to that of a given target (I_t). First, the cumulative histograms of the source and target
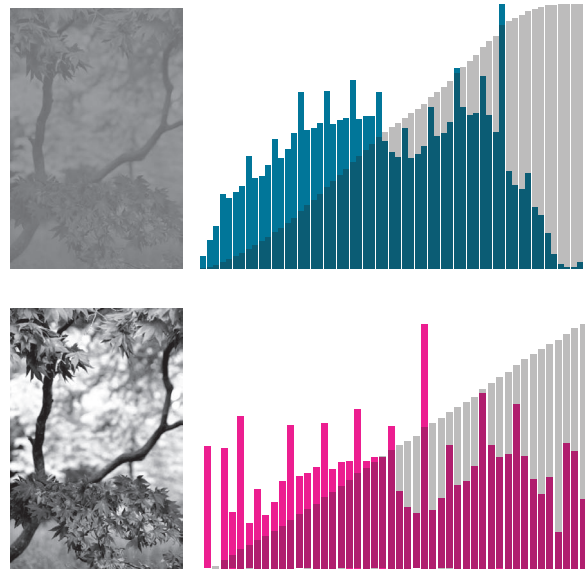

Figure 8: *Histogram equalisation can increase the contrast of a low contrast image (top left) by reshaping the intensity distribution more equally. The equalised resulting image is shown at the bottom left.*

are computed:

$$C_s(i) = \sum_{i=1}^{B} h_s(i) \qquad (13)$$

$$C_t(i) = \sum_{i=1}^{B} h_t(i) \qquad (14)$$

$$I_0(p) = v_t \left(C_t^{-1} \left(C_s \left(\frac{I(p) - \min(I) + 1}{V} \right) \right) \right) \qquad (15)$$

where a cumulative histogram C is defined as a function mapping a bin index to a cumulative count. The inverse function C^{-1} acts as a reverse lookup on the histogram, returning the bin index corresponding to a given count. An example of this technique applied on a source-target pair of colour images is shown in Figure 9.

Due to their limitations to single pixel properties, first order statistics are the simplest to compute and interpret. To further explore the relations between pixels, their spatial configurations and their associations with aspects of human vision more complex transforms are necessary. Second and higher order statistics as well as their applications in computer graphics, image processing and other relevant areas are presented in Sections 4 and 5 respectively.

4. Second Order Statistics

While the first order statistics of an image provide considerable information, their focus on individual pixels restricts
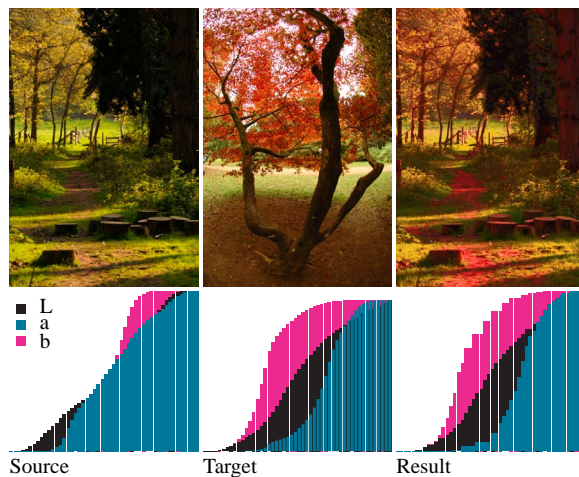
Figure 9: *The source image is matched to the target using histogram matching in the CIELab color space. The resulting image as well as the histograms for all three are shown.*

Figure 10: *The average power spectra for 133 natural images. Reused by permission from [RSAT04]*

them to the "unstructured" aspects of the image – equivalent to an investigation of individual rays of light. As James J. Gibson [Gib79] famously pointed out, since the interaction of light rays with an object systematic alters light according to the surface properties of the object, the pattern of change across a set of light rays – which he referred to as the structured optic array – provides direct, reliable information about the object. As is demonstrated in Figure 4, this higher-order structure is of central interest to the human visual system. It should not be surprising, then, that considerable attention has been directed towards investigating second and even higher order statistics. It has been shown, for example, that people can discriminate almost without effort between images that differ in the second order statistics [JC79,Cae81]. This is the basis for the well-known "pop-out" phenomenon.

The first and most obvious way to look at image structure is to examine the relationship between pairs of pixels. This is the core of second order statistics. There are two main second order statistics: The power spectrum (4.1) and gradients (4.2). Note that image contrast (e.g., the standard deviation of all pixel intensities divided by the mean intensity) is also often calculated. Image contrast can be derived from the power spectrum [van98b], and will not be discussed here.

4.1. Power Spectra

Images contain information at different spatial scales, from the large (e.g., the presence of a tree trunk on the right in Figure 4) to very small (e.g., the pattern of color change on the bird's feathers). As is well known from Fourier analysis, sharp edges, s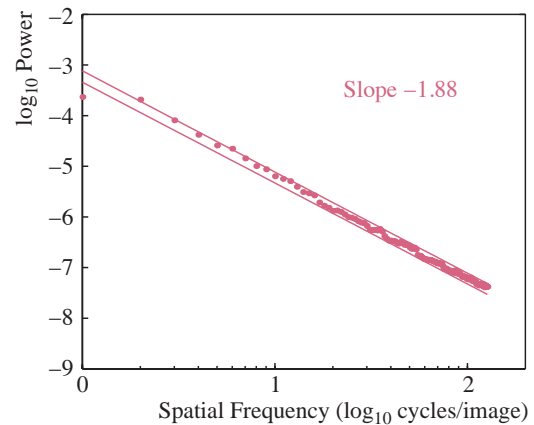uch as at the edges of the tree trunk, can be described by a weighted sum of sinusoids, with higher frequencies being weighted less. In other words, natural images contains information at many spatial frequencies. An examination of the relative power of the different spatial frequencies reveals several interesting trends, which are so prominent, that most work on image statistics provides an analysis of the power spectrum.

The power spectrum $S(u,v)$ of an N x N image is given by:

$$S(u,v) = \frac{|F(u,v)|^2}{N^2} \qquad (16)$$

where F is the Fourier transform of the image, and (u,v) are pixel indices in Fourier space. Two-dimensional frequencies can also be represented with polar coordinates (f,θ), using $u = f\cos(\theta)$ and $v = f\sin(\theta)$. While the sprectra of individual images varies considerably (which may play a role in the rapid detection of certain scenes or objects, [BCT08]), when the averaging over a sufficiently large number of images (and across orientation), a clear pattern arises: The lower frequencies contain the most power, with power decreasing as a function of frequency. In fact, on a log-log scale, amplitude as function of frequency lies approximately on a straight line. That is, the averaged spectrum tends to follow a power law, which can be well modeled with $P = 1/f^\beta$, where P is the power as function of frequency f, and β is spectral slope (see, e.g., Figure 10). Increasingly, the spectral slope is being used as a low dimensional descriptor of texture [BCHT01, CG87, KP88, RV90, TSW*05]. Perceptually, the primary effect of increasing the spectral slope is to increase the coarseness of the texture [BCHT01](see Figure 11). A number of studies have reported values for the spectral slope for different image ensembles. While the particulars of the image ensembles vary considerably, they tend to focus on natural scenes (which usually means simply
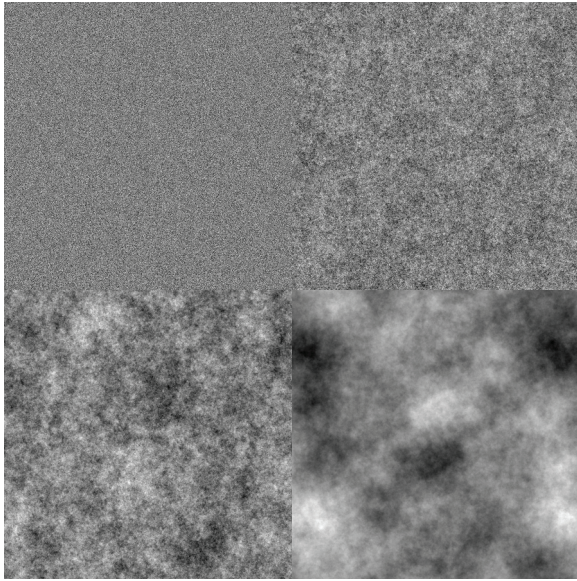
Figure 11: *Static, random phase patches produced by $1/f^\beta$ spatial-frequency filtering of random white noise. From top to bottom, the values of the spectral slope are 0.8, 1.6, 2,4, and 3.2 for the top left, top right, bottom left, and bottom right respectively.*

Study	# of images	$\beta \pm 1\text{sd}$
[BM87]	19	$2.1\pm.24$
[DA95]	320	2.30
* [DWA]+	95	2.29
[Fie87]	6	2.0
[Fie93]	85	2.20
[FB97]	20	2.20 ± 0.28
[van92]	117	2.13 ± 0.36
[HM99a]	216	1.96
[PBT98]	29	2.22 ± 0.26
[RST01]	133	1.88 ± 0.42
[RB94, Rud94]	45	1.81
[vv96]	276	1.88 ± 0.42
[TF97]	82	2.38
[TTC92]	135	$2.4.\pm0.26$
[TO03]	12,000	2.08
[WM97]	48	2.26

Table 3: *Spectral slopes for natural image ensembles. *Note that [DWA]'s work was on high dynamic range illumination maps.*

eliminating carpentered environments). The average spectral slope varies from 1.8 to 2.4, with most values clustering around 2.0 (see Table 3) [BM87, DA95, DWA, Fie87, Fie93, FB97, van92, HLM00, PBT98, RST01, RSAT04, RB94, Rud94, vv96, TF97, TTC92, TO03, WM97] One of the most prominent facets of such a power law is that these images tend to be scale invariant. That is, they share features at every scale, which is also a hallmark of fractal images (note that a simple linear transform of the spectral slope yields the image's fractal dimension [CG87]). In practice, this means that we should be able to zoom in and out of an image, or travel through a natural scene, and expect that the statistics will remain roughly constant. It is also worth noting that, according to the the Wiener-Kintchine theorem, the power spectrum and the auto-correlation function form a Fourier transform pair. This means that the spectral slope of image ensembles can be interpreted as describing relations between pairs of pixels. Intuitively, the means that since since the surface of a given object tends to be rather homogenous, it is expected that neighboring pixels will be similar and that the farther apart two pixels are, the less similar they will be.

The 1/f property of images seems to arise from several sources. Edges, for example, show 1/f spectra [SMS78]. Likewise, foliage and landscapes tend to exhibit fractal properties [Man83]. The clustering of independent objects is also such that the distribution of sizes in many scenes also tends to follow a power law [FB97, Rud97].

There is also some emerging evidence that the slope of the power spectrum is distinct for different scenes or objects [HM99a, TO03, WM97]. For example, Huang and Mumford [HM99a] examined 216 images which had been painstakingly segmented into pixels representing 11 different categories and found that although the image ensemble had an average slope of 1.96, there were systematic differences in the slopes across categories. Specifically, the slopes were 2.3, 1.8, 1.4, and 1.0 for man-made, vegetation, road, and sky pixels, respectively. Likewise, Webster and Miyahara [WM97] analyzed the power spectra and rms-contrast of 48 natural scenes. They found significant differences in both spectral slope and contrast across the three scene types (2.15, 2.23, and 2.4 for the forest, close-up, and distant meadow scenes, respectively). Furthermore, when the power spectra are not averaged over all orientations, it is clear that there is some variation as a function of angle: Natural images tend to concentrate their power in horizontal and vertical angles [BAM*97, OT01, Rud97, SMS78, vv96]. For example, in examining over 6000 man-made and 6000 natural scenes, [TO03] found that the slope varied as a function of both scene type and orientation (with slopes of 1.98, 2.02, and 2.22 for horizontal, oblique, and vertical angles in natural scenes and 1.83, 2.37, and 2.07 for man-made scenes). Thus, the spectral slope may be be useful in object or scene discrimination [BCT08]. It is critical to mention that if two images have similar spectral slopes, then swapping their power spectra will not affect recognition as long as phase information is preserved [TTC92, TT93].

It has also been shown that the pattern of change over time also follows a power law. That is, if the contrast mod-

ulation over time for a given pixel is examined, the power spectra can also be modeled with $P = 1/f^{\alpha}$, where P is the power as function of frequency f, and α is **temporal** spectral slope [BGK01,DA95,EBW92]. Temporal spectral slopes between 1.2 and 2.0 have been reported for natural image sequences. The temporal spectral slope relates, perceptually, to the apparent speed and jitter of the scene and to some degree with the apparent "purposefulness" of the motion (the higher the slope is, the more persistent the motion will be).

4.1.1. Human perception

There is an extremely large body of work examining the relationship between natural image statistics and human perception. At the simplest level, our ability to discriminate two random phase textures based solely on changes in the spectral slope has been examined [BCHT01, BCT08, KFK90, TT94]. Humans are most sensitive to slopes around 2.8 to 3.2, which would represent an image with much less high spatial frequency content than natural images. There is some evidence (albeit controversial) for a second minima near 1.2 Rainville and Kingdome examined the ability to detect symmetry for white noise images with different spectral slopes and found that one participant was bets for slopes near 2.8, consistent with the image discrimination data. The other participant was best for slopes between 1 and 2, consistent with the potential second minima. Regardless, it is clear that humans are not maximally sensitive to changes in spectral slopes representing natural images. The reasons for this are still unclear, although several hypotheses have been forwarded including that the shift from 2 to 2.8 reflects blur perception [Bil00].

Instead of attempting to determine the tuning of the visual system by measuring discrimination, one can approach the problem more indirectly: one can estimate the sensitivity of the spatial perception mechanisms from an autocorrelation analysis of contrast sensitivity function (which has been referred to as a modulation transfer function of the human visual system). It is generally accepted that human spatial perception is mediated by several, partially overlapping spatial frequency channels (at least 7 at each orientation: see, e.g., [Bil00]). Since similar frequencies are likely to be processed by the same channel, the sensitivity to similar frequencies should be similar. The less similar the frequencies are, the less correlated their sensitivity thresholds should be [BH96,OSS83,PWK95,SWO84]. Billock [Bil00], examined the correlation between contrast sensitivity thresholds as a function of the spatial frequency separation, and found that (for up to 5 octaves) the correlation functions were power laws with slopes ranging from 2.1 to 2.4. This held not only for static stimuli, but also for slowly flickering stimuli (up to 1 Hz). These slopes are much more in line with the slopes found in natural images, suggesting that human spatial frequency mechanisms may be optomized for natural images. Interestingly, more rapid flicker yielded higher slopes (around 2.6). As mentioned above, higher slopes reflects an attenuation of the higher spatial frequencies. Billock suggested that the higher slopes for rapidly flickering images may represent motion deblurring.

In contrast, the discrimination of temporal spectral slopes appears to be more straight forward. Humans are most sensitive to differences in temporal spectral slope for slopes between 1.8 and 2.0, which is very similar to the range of slopes in natural image sequences.

4.1.2. Fractal Forgeries

Thanks in large part to the seminal work of Mandelbrot [Man83], many areas of computer graphics use fractals to synthesize textures, surfaces, objects, or even whole scenes (see, e.g., [DCCH05,DL05,PS88]). A subset of this work focuses specifically on fractal brownian motion in general and and fractal brownian textures in specific which bear striking resemblance to real surfaces and textures. Since the fractal dimension of such a fractal texture is a linear transform of the spectral slope, these works are essentially relying on the regularities in power spectra. Many of these techniques either explicitly or implicitly choose parameters so that the spectral slope will be similar to natural images. Perhaps the most famous of these synthetic scenes are the eerily familiar landscapes produced by in Voss's "fractal forgeries" of [Vos85].

In the class of landscape synthesis, a white noise field is filtered in Fourier space with the desired spectral slope to produced the terrain texture of field [PS88]. In such methods, the role of natural image spectra is obvious. In a second class of approaches, the midpoint displacement method [FFC82], the role of power spectra is perhaps less obvious. In this approach, terrain is essentially generated by iteratively decreasing the spatial scale and (random) displacement. In general, the procedure is capable of producing a wide range of results, some of which look realistic and others which do not. A closer examination of the method shows that changing the size of the displacement decrement effectively allows one to control the coarseness of the texture. Since the coarseness of a texture is directly related to its spectral slope, it should be possible to automatically determine which displacement decrements produce natural terrain. The results of a simple perceptual study support this, showing that people felt that synthetic terrains that had a spectral slope of near 1.8 appeared to be the most natural [RSAT04].

Perhaps the second most common application of power spectra, again through the use of fractal dimensions, is the automatic generation of plants. Considerable literature has been written on plant synthesis (see, e.g., [DL05] for a review). Overall, it is well known that plants tend to exhibit a strong degree of self-similarity, and many descriptive systems have been developed to taker advantage of this characteristic. The most prominent of these are L-systems. Just as with fractal terrains, the system allows one to generate a

wide variety or results, some of which are perceptually plausible and some of which are not.

4.1.3. Image Processing and Categorization

While a power law description clearly captures the regularities in large collections of image, individual images tend not to be 1/f. It has been suggested that differences in the spectral slope between parts of an image allow people to rapidly make some simple discriminations (e.g., "pop-out" effect, see [BCT08, Cae81, JC79]). Others have speculated on the evolutionary advantage of being able to detect spectral slope [Bil00,BCT08,CHJ78,HB96,RV90]. Just as knowing about the statistics of natural images in generasl can inform us about how the human visual system works, and how we might build more efficient computer vision and computer graphics algorithms, so too will an understanding of the cause of variations in the statics provide insights.

A number of potential sources for 1/f patterns and their variations have been identified [Man83, FB97, HM99a, Rud97, SMS78, TO03, WM97]. For example, [HM99a] and [WM97] found different average spectral slopes for different scene categories (both within as well as between images): at the very least, there seem to be differences between man-made, general vegetation, forest, meadow, road, and sky elements. It has also been shown that underwater scenes have different spectral slopes [BG03]. To help further distinguish between object or scene categories, one can look at the interaction between power spectra and other characteristics [BAM*97, OT01, Rud97, SMS78, vv96]. For example, [OT01] suggested that a "spatial envelope" of a scene can be constructed from the interaction between power spectra and orientation combined with some information from PCA. This envelope yields perceptual dimensions such as naturalness, roughness, expansion, and similar categories tend to cluster together in this scene space. In [TO03], the approach was extended to estimate absolute depth using the relationship between power spectra and orientation and some information from wavelets.

In a related line of work, Dror and colleagues used a variety of natural image statiscs to estimate the reflectance of an object (e.g., metal versus plastic) under conditions where the illumination is unknown [DAW01]. They employ a wide range of image statistics from simple intensity distributions through oriented power spectra to wavelet coefficient distributions.

It has been suggested that the differences between the average spectral slope of 2.0 and the peak of human sensitivity to changes in slope (at 2.8) is due to deblurring. Furthermore, it has been suggested that the higher slopes for an autocorrelation analysis of human contrast sensitivity is due to motion deblurring. In an indirect examination of this claim, Dror and Colleagues examined the properties of Reichardt correlator motion detections [DOL00, DOL01]. While there is considerable evidence that the human visual system uses such correlators for the low-level detection of motion, it has also been shown using typical moving gratings that they signal temporal frequency and not velocity. Dror demonstrated that when stimuli that have natural image statics are used, the response properties of Reichardt detectors is better correlated with velocity and suggest that they make a much better basis for the synthetic processing of motion than previously assumed.

4.2. Gradients

Perhaps the simplest way to examine information contained in more than a single pixel, however, is to look at the relationship between pairs of pixels. This is a discrete approximation of the first derivative or gradient of the intensity function of an image. There are three different discrete appoximations: forward/backward difference, central difference, a soebel operator. The most straightforward is the forward difference method. Here, the gradient at a pixel (i,j) is calculated from the difference between it and the next pixel forwards:

$$Dx(i,j) = ln(I(i+1,j) - ln(I(i,j))) \quad (17)$$
$$Dy(i,j) = ln(I(i,j+1)) - ln(I(i,j)) \quad (18)$$

where I(i,j) is the luminance for pixel i, j, Dx is the horizontal gradient and Dy the vertical gradient. An obvious variant on this, referred to as the the backward difference method, is to use the previous pixel (e.g., I(i-1,j) rather than the next one. Ib both cases, it is common to calculate the mean gradient magnitude at a given location from the horizontal and vertical components:

$$D(i,j) = \begin{vmatrix} Dx(i,j) \\ Dy(i,j) \end{vmatrix} \quad (19)$$

$$D(i,j) = \sqrt{(Dx(i,j)^2 + Dy(i,j)^2)} \quad (20)$$

although one may choose to keep the vertical and horizontal gradients separate (see, for example, [Lev07]). An example of the gradient distribution for an image ensemble can be seen in Figure 12.

A slightly more robust method is to calculate the central differences:

$$Dx(i,j) = 1/2(ln(I(i-1,j)) - ln(I(i+1,j))) \quad (21)$$
$$Dy(i,j) = 1/2(ln(I(i,j-1)) - ln(I(i,j+1))) \quad (22)$$

Finally, one can use a Sobel operator, which also looks at the diagonal pixels:

$$Sx = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \quad (23)$$

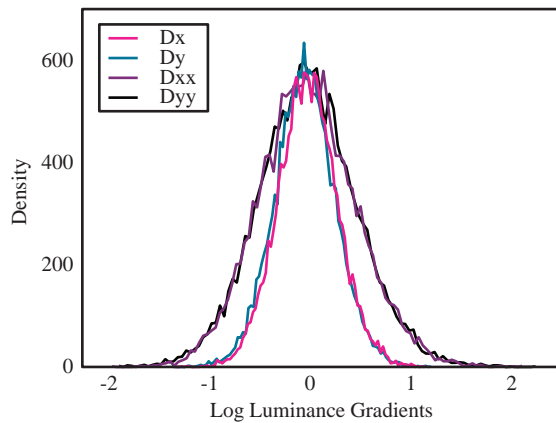$$Sy = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (24)$$

Figure 12: *Gradient distributions for a collection of natural images. Dx and Dy are the horizontal and vertical gradients (first derivative of the intensity function), respectively. Dxx and Dyy are the horizontal and vertical second derivatives.*

Just as was the case for power spectra, image gradients have been extensively investigated (see, e.g., [RB94, Rud94, HM99a, HLM00, DLAW, DWA]). It has been repeatedly found that the gradient histogram has a very sharp peak at zero (see Figure 12) and then falls off very rapidly (with long tails at the higher gradients). The distribution can be modeled as $e^{-x^{\alpha}}$ with $\alpha < 1$ [LZW02, Sim05]. The reason for the specific shape of the gradient distribution seems to be that images contain many, large surfaces which tend to be smooth and somewhat homogenous (so that the gradients tend to be near zero) along with a few high contrast edges (which will yield very high gradients) [BG00]. An object often finds itself over a consistent background, which means that the transition from the object's surface to the background will be similar all along the object's silhouette. This is reflected in the symmetry of the gradient distribution around the central peak.

4.2.1. Deblurring

When taking a photograph of a scene, it is not uncommon that either the camera or an object in the scene moves. The longer the aperture is open, the more likely this is to be the case. As a result, all or part of the image will be blurred. A number of approaches for sharpening an image have been proposed. One type of approach, blind motion deconvolution, essentially treats the photograph as the result of a convolution between an unknown sharp image and an equally unknown blurring kernel. The goal is to estimate the blur kernel so that it can be deconvolved with the blurred image to yield the sharp image. Naturally this is an underspecified problem, so additional constraints are needed and recently, a number of researchers have employed natural image statistics to provide them. For example, [CNR02] assume that

the sharp image's power spectra follows a power law distribution, while [JBFZ02, NCB04] use an interaction between power spectra and wavelets. Fergus et al. [FSH*06] estimate the blur kernel by optimizing for the gradient distributions using priors derived from natural images.

All of these approaches assume camera motion. That is, that there is a single blur kernel for the entire image. In an alternate approach, Levin [Lev07] use the gradient structure to find those pixels that are blurred and segment them from the rest of the image. Additionally, rather than use statistics of sharp images, of the blurring process. Specifically, they model how the the gradient distribution changes as a function of blurring to discovering the blurring kernel for a given image. One primary features of motion blurring is the attenuation of higher frequencies. This shows in the slope of the power spectra (by increasing β) as well as in gradient histogram (e.g., in particular by removing the longer tails at the higher gradients). Levin attempts to recover the blur kernel by applying different blurs to the image to find the one that produces a gradient histogram that matches the blurred one. This requires a non-blurred image or image region. Using a completely different image tends not to produce reliable results (due to differences in the underlying gradient histogram). Since much terrestrial motion tends to be horizontal, the majority of the effects of motion blurring are also horizontal. Thus, the vertical gradients can under some circumstances be used to calculate the blur of the horizontal components.

In a further refinement, [SJA08] employ an iterative alternation between optimizing the blur kernel and the sharp image using priors from natural image statistics. super resolution

In related work, [MFTF03] use the gradient distribution to fill in the holes produced when a high-resolution image is produced from a low resolution image (a process called super resolution). They also apply the same procedure to yield a full color image from a achromatic image.

Likewise, there are many approaches for removing the noise in an image. Recently, several researchers have had considerable success using gradient distributions (and occasionally other natural image statictics) to denoise images [RB05, RB09, Sim05].

4.2.2. Image Inpainting

One image processing application area where image statistics have been used successfully is known as *image inpainting*. In order to remove an unwanted object from an image, the texture that would otherwise be behind that object need to be re-created, a task that can be particularly difficult if the holes are large.

Among the various approaches, several have used image statistics to find optimal replacements for the missing texture and structure, usually by assuming that a patch similar to the missing portion can be found somewhere in the

remaining image. For example, [HT96], use a combination of spectral and spatial information to find the replacement patch. [LZW03] select the missing section based on the gradients at the boundary region as well as to maximize the match the global gradient histogram. [SJZW] fill in the missing region by completing the gradient maps and then reconstructing the image by solving a Poisson equation.

5. Higher Order Statistics

Whereas second order statistics take into account pairs of pixels through various transformations (in particular gradients and the amplitude spectrum), it is possible to find statistical regularities in larger groups of pixels using suitable image manipulations. In particular, gradient statistics can be extended in a hierarchical fashion to consider blocks of pixels by means of wavelet transforms. Further, it can be shown that much of the structure of images is located in the phase spectrum, as opposed to the power spectrum.

Principal Components Analysis (PCA) can be used to capture structure in images. It effectively allows the most important sources of commonality between images or within a single image to be assessed. Each principal component represents a dimension along which variability can be measured. They are typically ordered so that the first principal component accounts for most of the data, whereas each subsequent component accounts for progressively less of the structure in the data. Each of the components are orthogonal with respect to each other.

The orthogonality of the axes can be seen as a limitation. A better, albeit computationally much more involved technique is Independent Components Analysis (ICA). Rather than producing decorrelated axes that are orthogonal as achieved with PCA, Independent Components Analysis finds axes that are more or less independent, but which are not necessarily orthogonal.

In the following sections, each of these four approaches are outlined in more detail.

5.1. Phase Structure

It can be argued that although statistical regularities are present in power spectra of natural images, much of the perceptually relevant information is encoded in phase spectra [Tho99]. As an example, we have swapped the phase spectrum of two images, shown in Figure 13. As can be seen, much of the image structure has swapped. Thus, the image structure is largely encoded into the phase spectrum. Second order statistics such as the auto-correlation function (and therefore power spectra in Fourier space) and variance are insensitive to signal phase. They therefore do not adequately measure image structure.

We could try to measure higher moments such as skew and kurtosis, but these would be valid only if the image

Figure 13: *In this demonstration, the phases of the top two images are swapped to produce the bottom to images. The amplitude spectra are retained. Note that much of the image structure is located in the phase spectrum, and has therefore been swapped.*

exhibits stationary statistics, i.e. each image region has the same statistical properties. Natural images, however, are almost never stationary. As an example, this means that if we have an image with two regions characterised by a Gaussian distribution with means and standard deviations of (μ_1, σ_1) and (μ_2, σ_2) respectively, then each region will have zero skewness and kurtosis [NP96]. However, the image itself may still have non-zero skewness and kurtosis. In particular, when the means μ_1 and μ_2 are different, the skewness of the entire image may be non-zero. If the variances σ_1 and σ_2 are different, then the kurtosis will be non-zero. Here, the kurtosis is effectively a measure of the variance of the variances [Bad96b].

As a consequence, it is not possible to analyse the phase structure by computing higher-order measures. However, to gain access to phase information without polluting the results with first and second order information, we can whiten the images first [Tho99, Tho01]. This amounts to adjusting the spectral slope to become flat. The auto-correlation func-

tion will therefore be zero everywhere, except at the origin. By removing the second moment from consideration, it is now possible to compute skewness and kurtosis on the whitened signal. The whitened skew S_w and whitened kurtosis κ_w are thus a measure of variations in the phase spectra.

The results of applying this procedure to a set of natural images, leads to the conclusion that the whitened images are almost always positively skewed and are always positively kurtosed. In contrast, if the phase spectrum is randomised on the same images, the whitened skewness and kurtosis are close to zero.

While positive skewness and kurtosis of phase spectra points in the direction of the presence of statistical regularities in the phase spectrum, these results are relatively weak and do not easily explain aspects of human vision. Furthermore, they do not appear to have found employ in any graphics related applications that we are aware of (although it may lead to higher-order constraints in texture synthesis algorithms. In the following section, however, we will discuss how extending the present analysis to be localised in space, leads to further and stronger insights.

5.2. Wavelets

In the preceding section, we have seen that second order statistics, and in particular the amplitude spectrum of image ensembles, show striking statistical regularities. The amplitude spectrum can be analysed for different orientations, but as it is computed in Fourier space, it provides information of whole images rather than regions of images.

In this section we study transforms that allow us to analyse images in local regions of space, as well as for different orientations and spatial frequencies. Such models and transforms started with Gabor who introduced the notion of time-frequency representations [Gab46]. It is noted that such systems are in some sense similar to how images are analysed by portions of the human visual system. Many wavelet transforms are self-similar, consisting of spatially localised band-pass basis functions which vary only by a dilation, translation or rotation [Fie93]. They have found many direct applications in image analysis and image processing. In particular, they have been successfully used for the transmission of the chromatic component of HDTV signals [Wat90, Wat91], texture seggregation [BCG90], compact coding [Wat87,Dau88,ASF91] and in identification systems [Dau91].

Wavelet transforms using a variety of basis functions have also been used to model the spatial response properties of cells in the mammalian visual system [Fie93]. In particular, Gabor functions — sinusoids weighted by a Gaussians — are one of the more popular transforms. Nonetheless, these particular bases are non-orthogonal, making them difficult to invert and use in modelling [Fie93]. As a result, orthogonal bases were developed [ASH87, WA89, ASF91].

Cells in the visual system respond to certain patterns of light and dark, with each cell responding optimally to a specific pattern. It has been demonstrated that groups of cells in the visual cortex of several different species of mammal respond to wavelet-like structures, in particular the Gabor function [Mar80]:

$$f(x,y) = \sin\left(2\pi kx + \Theta\right)\exp\left(\frac{-x^2}{2\sigma_x^2} + \frac{-y^2}{2\sigma_y^2}\right) \qquad (25)$$

A Gabor filter and an image processed with it are shown in Figure 14. As can be seen, the local oriented structure in the image is analysed at the specific frequency determined by the sinusoidal component of the filter.

It is possible to extend this description by making the bandwidth proportional to the frequency, so that the variances σ_x, σ_y in the above equation are related to frequency band k:

$$\sigma_{x,y}(k) = \beta k^{-\alpha} \qquad (26)$$

This allows us to build a stack of filters, each level in the stack tuned to a different set of frequencies, in proportion to the position in the stack. Each level in the stack is also known as the scale.

Field has shown that both phase and amplitude can be measured and correlated across scales and location [Fie93]. Although the amplitude spectrum has shown scale invariance, the phase structure revealed by Fourier space analysis remains somewhat of a mystery. However, by analysing wavelets, which are simultaneously localised both in space and frequency, it can be shown that natural images are scale in variant in *both* phase and amplitude.

This means that in local image regions, the location of edges are correlated across scales. At higher frequencies neighbours in frequency become more highly correlated, whereas neighbours in space become less correlated. In other words, phases in natural images contain alignment that extend over greater ranges at higher frequencies [Fie93]. Phase structures tend to change somewhat in orientation as well as position when moving across scales. Such behaviour is well modeled with wavelets. It is therefore possible that this may help explain why mammalian visual system use wavelet-like structures in the visual cortex. It may well be that this is why wavelets are useful in a range of applications, as outlined above.

It was found that the distributions of histograms of wavelet coefficients are highly non-Gaussian [Fie87,Mal89], showing in particular long tails, i.e. having high kurtosis. In particular, these histogram distributions can be modeled by a generalised Laplacian [Mal89,SA96,Sim99b]:

$$P(x) = \frac{\exp\left(-|c/s|^p\right)}{\frac{2s}{p}\,\Gamma\left(\frac{1}{p}\right)} \qquad (27)$$

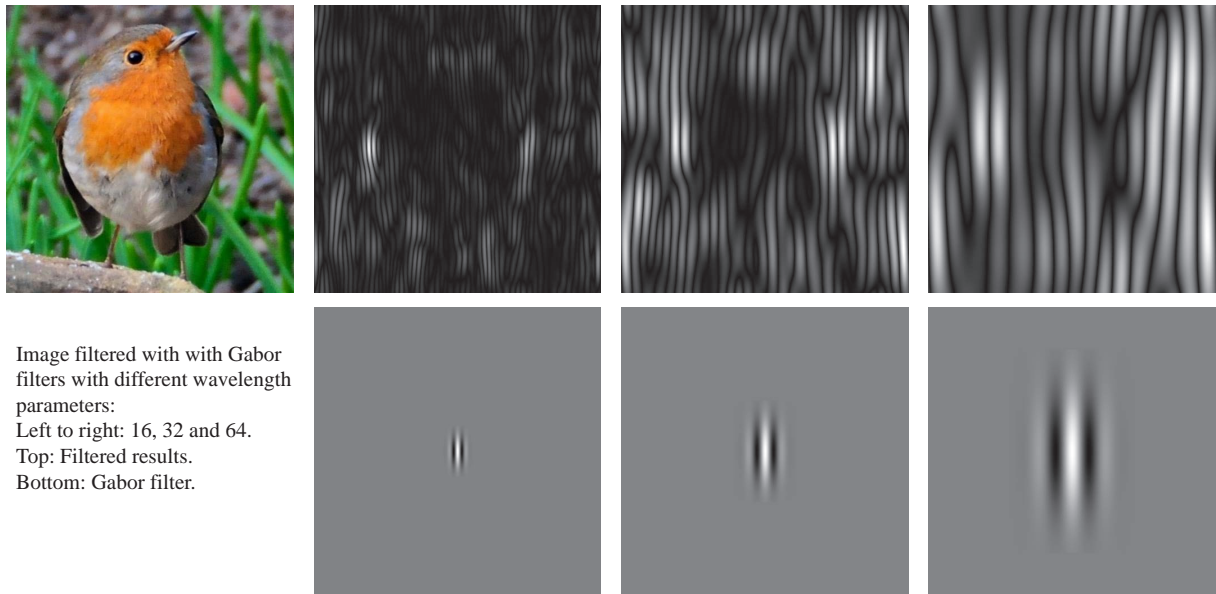where Γ is the gamma function and the parameters s and p

Image filtered with with Gabor filters with different wavelength parameters:
Left to right: 16, 32 and 64.
Top: Filtered results.
Bottom: Gabor filter.

Figure 14: *Examples of Gabor filters and an image processed with them.*

can be used to fit the model to a specific image or image ensemble.

If the bandwidths of the wavelet filters are chosen to be around 1-2 octaves, that the kurtosis of the histograms of wavelet transforms is maximal. This means that under these conditions the resulting transform is maximally sparse. In clusters of neurons, sparisity is an important feature. It means that much of the variability in the input signal is explained by a small number of neurons [Fie94]. Such neuronal activity is metabolically efficient [Bad96a], minimises wiring length [Fol95], and increases capacity in associative memory [BMW88].

Similarly, in a wavelet transform, sparsity means that most coefficients are small, with only a few coefficients having larger values. This is for instance the basis for compression algorithms: by replacing small coefficients with zero coefficients a lossy image compression algorithm can be constructed. This approach is used for instance in the JPEG2000 image compression standard [TM01]. Furthermore, wavelet-like receptive fields are discovered when training a neural network for sparseness [OF96].

Of course, we are still free to choose the basis functions. In many applications, for reasons of computability a much simpler set of basis functions are used. In particular, the Haar bases are used. In a Haar decomposition, an image is split into averages and differences with those averages. The averages are computed by a set of scaling func-

tions [SDS95, SDS96]:

$$\Phi_i^k = \phi\left(2^k x - i\right) i = 0, \dots 2^k - 1 \qquad (28)$$

$$\Phi(x) = \begin{cases} 1 & \text{for } 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \qquad (29)$$

The differences are computed with the mother wavelet function:

$$\psi_i^k = \phi\left(2^k x - i\right) i = 0, \dots 2^k - 1 \qquad (30)$$

$$\psi(x) = \begin{cases} 1 & \text{for } 0 \leq x < 1/2 \\ -1 & \text{for } 1/2 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \qquad (31)$$

An example of a Haar wavelet decomposition with three scales is shown in Figure 15.

5.2.1. Image Denoising

An example application of wavelet statistics is image denoising [Sim99a, PS00], where a Bayesian approach is employed to remove noise. Assuming that the wavelet coefficients of natural images follow a generalised Laplacian distribution, as outlined above, Gaussian image noise translates into Gaussian polution of wavelet coefficients. Each coefficient can therefore be written as $y = x + n$, with x drawn from the density given by Equation (27) [Sim99a]. The noise n is Gaussian. A standard estimator for x given the corrupted im-

Figure 15: *Top: Original image. Bottom: An example of a Haar wavelet decomposition with three scales. The top left image in the decomposition represents averages, whereas the remaining panels indicate differences in horizontal, vertical and diagonal directions.*

age y is the maximum a posteriori (MAP) estimator:

$$\hat{x}(y) = \arg\max_c P(x|y) \tag{32}$$

$$= \arg\max_c P(y|x)\, P(x) \tag{33}$$

$$= \arg\max_c P_n(y - x)\, P(x) \tag{34}$$

Thus, by applying Bayes rule, the estimator is rewritten as the product of the Gaussian noise density $P_n(y - x)$ and the prior density of the signal coefficient $P(x)$. If we assume that the parameter p in Equation (27) is set to 2 the estimator becomes linear:

$$\hat{x}(y) = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_n^2}\, y \tag{35}$$

For other values of p the estimator is non-linear. For instance $p = 0.5$ and $p = 1$ lead to hard thresholding and soft-thresholding respectively. In television and video engineering such thresholding is known as coring [BP86].

This approach assumes that wavelet coefficients are inde-

pendent. Although wavelet coefficients are on average zero, showing that their values are decorrelated, they are not independent. Although beyond the scope of this report, it can be shown that correlations between x and p can be accounted for using a joint statistical model for the purpose of denoising [Sim99a], image compression [Sim97], as well as texture synthesis [SP98].

5.2.2. Object Detection

An application area where wavelets have found extensive use is object detection. An important problem in that area is finding an effective and compact way of characterising objects in images. Some object categories, such as faces, show surprisingly little variation in terms of overall structure and colour, other types of objects however can be more complex.

To overcome this issue, wavelets, and in particular Haar wavelets, can be used to provide a standardised template for characterising objects in images. Using a training scheme, a small subset of Haar wavelets can be derived that can appropriately describe the structure of a particular class of images, such as for instance faces or pedestrians. This set of basis functions then forms the classifier that can be used to detect the object in question in new image [OPS*97, POP98].

More recently, a real-time face detection algorithm was proposed by Viola and Jones [VJ01] where features are applied on the image in a cascade fashion, ensuring that more complex classifiers are only used on instances that have been already accepted by simpler ones. Although the features used in this work are similar to Haar wavelets, some are more complex, as shown in Figure 16.

Image-based retrieval using the wavelet decomposition of each image has also been demonstrated by Jacobs et al. [JFS95]. An *image querying metric* is developed using truncated, quantised versions of the wavelet decompositions of the images. This forms a compact signature for each image that is fast to compute and captures sufficient amount of information to allow searching.

5.3. Principal Components Analysis (PCA)

One of the main properties of natural images is that pixel values are correlated. Nearby pixels tend to have similar values while these correlations tend to be less strong as the distance between pixels increases [HBS92]. One way to analyse these correlations is by decomposing the image data to a set of components that are as decorrelated as possible. Each of these components will then capture some particular mode of variation within the data.

Principal component analysis (PCA – also known as the *Karhunen-Loève* transform) is a solution commonly employed to that end. By computing the eigenvectors and corresponding eigenvalues of the covariance matrix of the image collection, a set of orthogonal, decorrelated components can
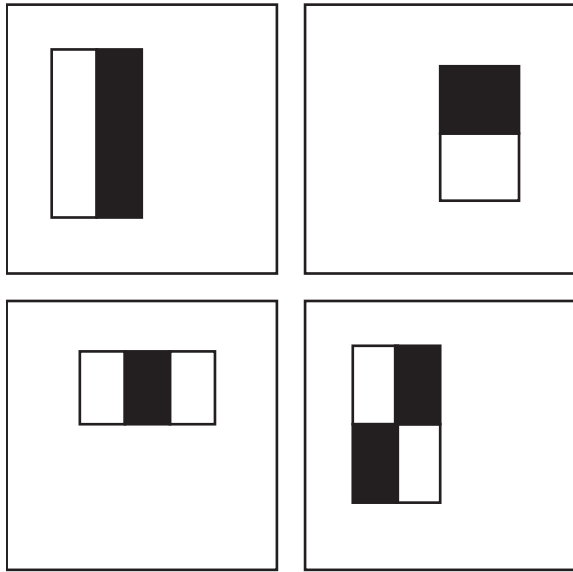
Figure 16: *The features used by Viola and Jones to detect faces in images.*

Figure 17: *A 2D representation of matching in the "face space". The axes u_1 and u_2 correspond to the two eigenvectors (eigenfaces).*

be derived. The covariance between two variables X and Y, each consisting of n samples, is defined as:

$$\text{cov}(X,Y) = \frac{\sum n_{i=1}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1} \qquad (36)$$

where n is the number of elements in X, Y. Note that $\text{cov}(X,X)$ gives the variance of X. To capture the covariance between more than two variables, a covariance matrix C can be computed. For m variables $\{I_1, I_2, \ldots, I_m\}$, each consisting of n samples, the covariance matrix $\mathbf{C}$ is given by:

$$\mathbf{C} = \begin{bmatrix} \text{cov}(I_1,I_1) & \text{cov}(I_1,I_2) & \cdots & \text{cov}(I_1,I_m) \\ \text{cov}(I_2,I_1) & \text{cov}(I_2,I_2) & \cdots & \text{cov}(I_2,I_m) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(I_m,I_1) & \text{cov}(I_m,I_2) & \cdots & \text{cov}(I_m,I_m) \end{bmatrix} \qquad (37)$$

The matrix of eigenvectors $\mathbf{V}$ of $\mathbf{C}$ and the diagonal matrix of corresponding eigenvalues $\mathbf{D}$ can then be computed such that $\mathbf{C} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$.

Components with a larger eigenvalue correspond to more important features while lower eigenvalues signify less important components. This representation offers the advantage that a relatively small number of components is sufficient to encode most of the information in the images.

Nevertheless, although the computed eigenvectors are orthogonal, statistical independence cannot be guaranteed. Natural images are non-Gaussian and thus the decomposition offered by PCA can only decorrelate the data [HHH09]. The main consequence of this is that meaningful information
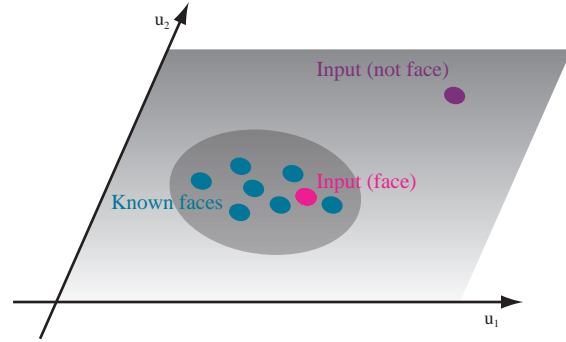
is only captured by the first few components, while further ones mostly correspond to high frequency features.

5.3.1. Face Recognition

As discussed, PCA can be used to reduce the dimensionality of a given set. This observation has been used in object and in particular face recognition applications successfully. The face recognition method known as *Eigenfaces* [TP91b, TP91a] is one example. The database of face images is decomposed into a set of eigenvectors and eigenvalues, each of which corresponds to some dimension along which faces vary. Faces tend to have a lot of common characteristics in terms of structure, making them an ideal subject for dimensionality reduction.

In order to match input faces to samples in the database, the new face can be projected to the computed "face space". It then becomes a simple matter of computing the distance between this new point and the database to decide whether the input is in fact a face. Recognition can then take place by finding the points closest to the input. Figure 17 shows a simplified 2-dimensional version of this process.

5.4. Independent Components Analysis (ICA)

While Principal Components Analysis measures covariances of pixels separated by varying distances, the components that are found constitute a linear transform. This means that although the data may be decorrelated, there is no guarantee that the resulting components are in any way independent. In particular, only if the input data happens to have a Gaussian distribution, then the resulting principal components are both decorrelated and independent.

As we have seen, many of the statistics of natural images that are currently known point in the direction of high kurtosis, i.e. they are highly non-Gaussian. This has

given rise to the use of Independent Components Analysis (ICA) [Com94, HKO01], which is a technique to separate a multivariate signal into a set of components that will sum to the original signal under the assumption that the input is non-Gaussian. Like PCA, ICA is a special case of blind source separation, which is a general collection of techniques that try to separate a set of signals from a mixed signal without prior knowledge of the source signals or the mixing process [Ach08].

ICA algorithms tend to require a set of preprocessing steps to make the problem more tractable. In particular, the data need to be centred and whitened. Also, data reduction is often applied. The whitening can be achieved by running the PCA algorithm first. By keeping only the first n components, data reduction is achieved. It ensures that only those components are used that are meaningful to the problem being solved. Moreover, it speeds-up the computations required to determine independent components.

Several algorithms are known, including infomax [BS95], extended infomax [LGS99], fastICA [Hyv99], Jade [Car99], kernel ICA [BJ02] and RADICAL [LMF03]. Although the implementations vary, their goals are the same: represent multivariate data as a sum of components in such a way that the components reveal the underlying structure that gave rise to the observed signal.

The number of ICA algorithms available, and their computational and algorithmic complexity make an in-depth discussion of this class of algorithms beyond the scope of this report. However, we will discuss some of their implications for natural image statistics as well as human vision.

One could ask what are the independent components in the context of natural images. It could be argued that these are the objects that comprise a scene [OF97]. It is exactly these that provide the structure that gives rise to the pixel array that is analysed by computer algorithms as well as by the human visual system. Both PCA and ICA, however, are generative models which are based on linear superposition. Thus, they cannot take into account translation and rotation of objects [OF97].

However, it has been found that small image patches of natural image ensembles can be analysed using ICA, revealing structures that resemble those found in the human visual system [BS97]. In particular, this technique yields Gabor-like patches — elongated structures that are localised in space, orientation and frequency. Their size, aspect ratio, spatial frequency bandwidth, receptive field length and orientation tuning bandwidth are similar to those measured in cortical cells [vv98]. These results lend credence to the argument that the human visual system appears to represent natural images with independent variables, each having a highly kurtotic distribution that leads to a metabolically efficient sparse coding.

6. Colour Statistics

We have so far considered statistical regularities in images either without specifically including the notion of colour, or deliberately ignoring colour and assuming a single luminance channel. The reason is that colour properties of images have their own peculiarities, warranting a separate treatment [BM87, PBTM98, CCO00, COVC00].

In general, images are formed as light is transduced to electric signals. This happens both in the retina as well as in electronic image sensors. Before that, light is emitted and usually reflected several times off different surfaces before it reaches a sensor. This process is modeled by the rendering equation, given by:

$$L_o(\lambda) = L_e(\lambda) + \int_\Omega L_i(\lambda)\, f_r(\lambda)\, \cos(\Theta) d\omega \qquad (38)$$

This equation models the behaviour of light at a surface: light L_i impinging on a surface point from all directions Ω is reflected into an outgoing direction of interest, weighted by the bi-directional reflectance distribution function (BRDF) f_r. The amount of light L_o going into that direction is then the sum of the reflected light and the light emitted by the surface point (which is zero unless the surface is a light source). Here, we have specifically made all relevant terms dependent on wavelength λ to indicate that this behaviour happens at all wavelengths, including all visible wavelengths. This means that $L_o(\lambda)$ can be seen as a spectral distribution.

Under photopic lighting conditions, the human visual system contains three cone types, each with a different peak sensitivity. Thus, each cone type integrates the incident light according to a different weight function:

$$L = \int_\lambda L_o(\lambda)\, \bar{l}(\lambda)\, d\lambda \qquad (39a)$$

$$M = \int_\lambda L_o(\lambda)\, \bar{m}(\lambda)\, d\lambda \qquad (39b)$$

$$S = \int_\lambda L_o(\lambda)\, \bar{s}(\lambda)\, d\lambda \qquad (39c)$$

where $\bar{l}(\lambda)$, $\bar{m}(\lambda)$ and $\bar{s}(\lambda)$ are the weight functions (responsivities) which peak at wavelengths that are perceived roughly as red, green, and blue. The letters L, M and S stand for "long", "medium" and "short" wavelengths.

A spectral distribution $L_o(\lambda)$ therefore gives rise to a triple of numbers (L, M, S) which represent the signal that is passed on from the photoreceptors to the remainder of the human visual system. Such triples are called tristimulus values. An important implication of this behaviour is that different spectral distributions can integrate to the same tristimulus values. Such spectral distributions are then necessarily perceived to be identical, which is known as metamerism. It is the reason that we are able to build display devices with three primaries that are not the same as the responsivities of the cones: tristimulus values in one colour space can be converted to tristimulus values in another colour space, and

although they do not represent the same spectral distributions, through metamerism they lead to the same percept. Further, rather than study spectral distributions, we can limit ourselves to the study of tristimulus values.

There are two interesting and relevant findings in natural image statistics that pertain to colour. Both of these relate to the relationships of values between the components of tristimulus values. In other words, there exist statistical relationships between the three axes of colour spaces. Essentially, correlations between color channels can be exploited in colour constancy / white balancing algorithms. By decorrelating images or image ensembles, new colour spaces can be constructed, which have applications in for instance colour difference metrics as well as colour transfer algorithms.

6.1. Colour Constancy and White Balancing

Humans tend to effortlessly discount the illumination of their environments. In other words, the colour of the light sources in a particular viewing environment are to a large extent ignored, so that human vision is fairly good at detecting surface colours. This phenomenon is known as colour constancy. Computational models that aim to achieve the same are known as white balancing algorithm. They intend to remove colour casts from images due to illumination.

As light is reflected off surfaces according to the rendering equation, reconstructing either surface reflectance and/or illumination is an underconstrained problem. This means that white balancing can only be achieved by means of statistical assumptions. These are discussed in the following sections.

6.1.1. The Grey World Assumption

In colour spaces such as the aforementioned *LMS* space, equal values of the three components denote achromatic colours. One way to achieve such neutral colours is to start with an equal energy spectrum, i.e. a spectral distribution which has the same value L_o for each wavelength λ. This could happen if a scene is illuminated by an equal-energy light source, a source that emits the same energy at all wavelengths. If the BRDF then also reflects all wavelengths equally, a neutral colour would be achieved.

In practice, a scene is illuminated by only one or at most a few light sources, with an emission spectrum that is off-white. The BRDFs in a scene are much more variable. However, a surprising finding is that when a collection of BRDFs are averaged, the result ends up being a distribution function that is close to grey. This is known as the grey-world assumption [RKAJ08].

The implication of this finding is that if we were to average the colours of all pixels in an image, and the grey-world assumption holds, the average tristimulus value $(\bar{L}, \bar{M}, \bar{S})$ is a good indicator of the colour of the light source. Of course, when averaged over natural image ensembles, the grey world assumption does tend to hold. In single images this assumption may or may not hold. In particular, if the surface colours in the scene are biased towards some specific saturated colour, the average reflectance will not tend toward grey.

The grey world assumption is often used to aid white balancing. After all, if we know the colour of the illuminant, then we can correct all pixels by simply dividing all pixels by the average image value. Moreover, if we know that the display has a different white point, say $(L_{d,w}, M_{d,w}, S_{d,w})$, white balancing can be implemented as follows:

$$L_{\text{wb}} = L \frac{L_{d,w}}{\bar{L}} \tag{40a}$$

$$M_{\text{wb}} = M \frac{M_{d,w}}{\bar{M}} \tag{40b}$$

$$S_{\text{wb}} = S \frac{S_{d,w}}{\bar{S}} \tag{40c}$$

The grey-world assumption is a statistical argument that is necessary to perform white balancing on images in the absence of further information about the illuminant, given that white balancing is by itself an underconstrained problem [FBM98].

Note that we have glossed over the fact that this procedure is best applied in a perceptual colour space, such as *LMS*, thereby mimicking chromatic adaptation processes that occur in the human visual system. If an image is given in a different colour space, most likely the sRGB space, the image should first be converted to *LMS*.

The approximation of the illuminant can be improved by excluding the most saturated pixels from the estimation [APS98]. Alternatively, the image can be subjected to further statistical analysis to determine if the color distribution is due coloured surfaces or due to a coloured light source [GS04]. Here the image is first converted to the CIELAB colour opponent space. Ignoring the lightest and darkest pixels as they do not contribute to a reliable estimate, the remaining pixels are used to computed a two-dimensional histogram $F(a,b)$ on the two chromatic channels a and b. In each channel, the chromatic distributions are modeled with:

$$\mu_k = \int_k k F(a,b) \, dk \tag{41}$$

$$\sigma_k^2 = \int_k (\mu_k - k) \, F(a,b) \, dk \tag{42}$$

where $k = a, b$. These are the mean and variances of the histogram projections onto the a and b axes. In CIELAB neutral colours lie around the $(a,b) = (0,0)$ point. To assess how far the histogram lies from this point, the distance D can be computed:

$$D = \mu - \sigma \tag{43}$$

where $\mu = \sqrt{\mu_a^2 + \mu_b^2}$ and $\sigma = \sqrt{\sigma_a^2 + \sigma_b^2}$.

Figure 18: *Examples images used to demonstrate the correlation between channels. The first two images are reasonable examples of natural images, whereas the third image is an example of an image taken in a built-up area. Built environments tend to have somewhat different natural image statistics compared with natural scenes [ZL97, ZL98]. Figure taken from [RKAJ08], courtesy AK Peters, Ltd.*
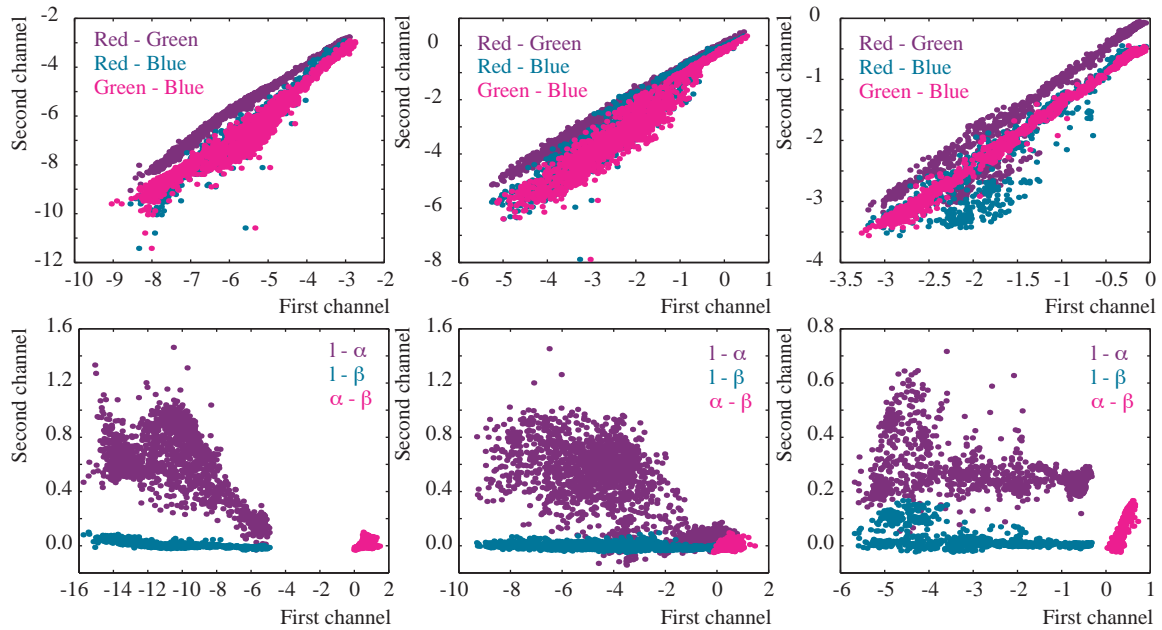

Figure 19: *Random samples plotted in RGB color space (top) and $L\alpha\beta$ color space (bottom). The top to bottom order of the plots is the same as the order of the images in Figure 18. Figure taken from [RKAJ08], courtesy AK Peters, Ltd.*

The measure $D_\sigma = D/\sigma$ can be used to assess the strength of the cast. If the spread of the histogram is small, and lies far away from the origin, the image is likely to be dominated by strong reflectances rather than illumination.

6.1.2. Generalised Grey-World and White Patch Assumptions

It was found that while grey-world algorithms work well on some images, alternate solutions such as the white-patch algorithm [Lan77] perform better on texture-rich images. The white-patch algorithm assumes that the lightest pixels in an image depict a surface with neutral reflectance, so that its colour represents the illuminant.

Both the grey-world and white patch algorithms are spe-

cial instances of the Minkowski norm [FT04]:

$$L_p = \left(\frac{\int f^p(x)dx}{\int dx} \right)^{1/p} = ke \qquad (44)$$

where $f(x)$ denotes the image at pixel x. The average of the image is computed for $p = 1$, thereby implementing the grey-world assumption. The maximum value of the image is computed by substituting $p = \infty$, which represents the white-patch algorithm.

A further generalised assumption can be made about images, which is that the average difference between two pixels evaluates to grey. This is known as the grey-edge assump-

tion [vG05], and can be formulated as follows [GG07]:

$$\left(\int \left| \frac{\partial^n f^\sigma(x)}{\partial x^n} \right|^p \, dx \right)^{1/p} = k e^{n,p,\sigma} \qquad (45)$$

Here, n is the order of the derivative, p is the Minkowski-norm and $f^\sigma(x) = f \otimes G^\sigma$ is the Gaussian blurred image where the size of the filter kernel is given by σ. With this formulation several colour constancy algorithms can be constructed:

$e^{0,1,0}$ This implements the grey-world algorithm.

$e^{0,\infty,0}$ This is the white-patch algorithm.

$e^{n,p,\sigma}$ This is known as the general grey-world algorithm. The value of n is typically 0, 1 or 2 for zero, first and second order statistics.

In each case, the value of n gives the order of the statistics.

6.1.3. White Balance Algorithm Selection

Having noted that many color constancy and white balancing algorithms exist, with none of them universally applicable, Gijsenij and Gevers use natural image statistics to classify an image, and then use this classification to select the most appropriate white balancing algorithm for the image at hand [GG07]. In particular, they use the finding that the distribution of edge responses in an image can be modeled by means of a Weibull distribution [GS05]:

$$f(x) = \frac{\gamma}{\beta} \left(\frac{x}{\beta} \right)^{\gamma-1} \exp \left(\frac{x}{\beta} \right)^\gamma \qquad (46)$$

The parameters β and γ have meaning in this context. The contrast of an image (or image ensemble) is given by β whereas γ is an indicator of grain size (i.e. the peakedness of the distribution). This means that higher values for β represent images with more contrast. Higher values for γ indicate finer textures.

Fitting a Weibull distribution involves computing a Gaussian derivative filter in both x and y directions. This results in the $(\beta_x, \beta_y, \gamma_x, \gamma_y)$ set of parameters for each colour channel. The Gaussian derivate filter can be first, second or third order. However, it was found that the order of the chosen filter is relatively unimportant: high correlations exist between the fitted parameters [GS05].

It is now possible to fit Weibull parameters to the derivaties of a large number of images, and correlate their values with the white balancing algorithm that performs best for each image. The parameter space tends to form clusters where a specific algorithm tends to produce the most accurate result. This means that the Weibull distribution can be used to select the most appropriate white balancing algorithm for each image individually [GG07].

6.2. Statistical Decorrelation

The success of the grey-world assumption has a second implication. If the average surface reflectance is grey, and light sources tend towards white, then in the *LMS* colour space and similar RGB-like spaces, values in one colour channel tend to be good predictors of values in the other two colour channels. In other words, if we find a high value for a red pixel, then chances are that the green and blue values are high as well. This means that the colour channels in such colour spaces are highly correlated.

An interesting experiment was conducted whereby a set of spectral images was converted to log *LMS* space, before being subjected to Principal Components Analysis (PCA) [RCC98]. The logarithm was taken as a data-conditioning measure. Applied in this manner, PCA rotates the axes to the point where they are maximally decorrelated. The resulting colour space is termed $L\alpha\beta$ where L is a lightness channel, and α and β are colour opponent channels. By starting in *LMS* cone space, the rotation yields a new colour space which surprisingly corresponds closely to the color opponent space found in the ganglion cells (these are the cells that transport the visual signal from the retina through the optic nerve to the lateral geniculate nucleus, thought to be a relay station in the brain).

Colour opponent spaces are characterised by a single achromatic channel, typically encoding luminance or lightness, and two chromatic axes which can be thought of as spanning red-green and yellow-blue (although the axes do not precisely correspond to these perceptual hues). The chromatic axes can have positive and negative values; a positive value can for instance denote the degree of redness, whereas a negative value in the same channel would denote the degree of greenness. A consequence of colour opponency is that we are not able to simultaneously perceive an object to have green and red hues. Although we may describe objects as reddish-blue or yellowish-green, we never describe them as reddish-green. The same holds for yellow and blue.

The conversion from XYZ to LMS, and then from LMS to the aforementioned decorrelated opponent colour space $L\alpha\beta$ is given by:

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.3897 & 0.6890 & -0.0787 \\ -0.2298 & 1.1834 & 0.0464 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \qquad (47)$$

and the inverse conversion from LMS to XYZ is given by:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 1.9102 & -1.1121 & 0.2019 \\ 0.3710 & 0.6291 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix} \begin{bmatrix} L \\ M \\ S \end{bmatrix} \qquad (48)$$

Converting from LMS to $L\alpha\beta$ involves a further matrix multiplication:

Figure 20: *The top-left image is decomposed into the L channel of the Lαβ color space, as well as L + α and L + β channels in the bottom-left and bottom-right images. Figure from [RKAJ08], courtesy AK Peters, Ltd.*

$$\begin{bmatrix} L \\ \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{6}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -2 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} \log L \\ \log M \\ \log S \end{bmatrix}$$

$$(49)$$

while the inverse transform is given by:

$$\begin{bmatrix} \log L \\ \log M \\ \log S \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -2 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{3} & 0 & 0 \\ 0 & \frac{\sqrt{6}}{6} & 0 \\ 0 & 0 & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} L \\ \alpha \\ \beta \end{bmatrix}$$

$$(50)$$

For 2000 random samples drawn from each of the images shown in Figure 18, their distribution is plotted in Figure 19. The point clouds form more or less diagonal lines in RGB space when pairs of channels are plotted against each other, showing that the three color channels in RGB space are almost completely correlated for these images. This is not the case for the same pixels plotted in $L\alpha\beta$ space.

The colour opponency of the $L\alpha\beta$ space is demonstrated in Figure 20, where the image is decomposed into its separate channels. The image representing the α channel has the β channel reset to 0 and vice versa. We have retained the luminance variation here for the purpose of visualization. The image showing the luminance channel only was created by setting both the α and β channels to zero.

The fact that natural images can be transformed to a decorrelated colour space, which coincides with a colour decomposition occurring in the human visual system, points to a beautiful adaptation of human vision to its natural input.

It also points to several applications that are possible due to the fact that opponent spaces are decorrelated. In particular, we highlight colour transfer, an algorithm that attempts to transfer the look and feel of one image to another on the basis of its colour content. Due to the three-dimensional nature of colour spaces, this is in the general sense a complicated three-dimensional problem. However, by converting to a colour opponent space, the data in each channel will be decorrelated from the other channels. As a first example, the histogram projection onto the *a* and *b* axes as outlined in Section 6.1.1 yields meaningful results as the CIELAB space used here is an example of a colour opponent space.

6.2.1. Colour Transfer

Perhaps one of the most direct applications of the use of decorrelated colour spaces is colour transfer between images. The aim is to transfer some statistical properties of one image to another. To make this process straightforward, images can be converted to a decorrelated colour space, such as the $L\alpha\beta$ space proposed by Ruderman [RCC98]. In this space, the means and standard deviations of all the pixels can be computed separately in each of the three axis. Doing this for both a source and a target image yields a set of factors and terms that can be used to shift and scale the pixel data in one image to match the means and standard deviations of the other image [RAGS01]. An example of this procedure is shown in Figure 21. Applications of colour transfer, aside from producing aesthetically pleasing images, include creating night-time images from day-time images, as well as augmented reality applications, where colour transfer provides an inexpensive way to make rendered objects fit into a captured environment.

Although the $L\alpha\beta$ space is on average decorrelated, individual images may deviate arbitrarily from this ideal. As such, many images may still show significant correlations
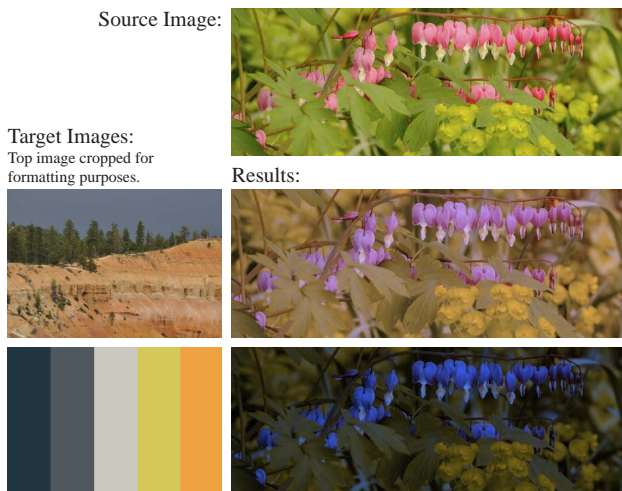
Source Image:

Target Images:
Top image cropped for
formatting purposes.

Results:

Figure 21: *Example of colour transfer.*

between the axis, even in a colour opponent space. It would be possible to individually subject images to Principal Components Analysis to find the most decorrelated axes for each specific image. Such a technique may provide plausible results for a wider range of inputs [AK04, AK07].

7. Conclusions

Natural images are not random. They contain striking statistical regularities, which are surveyed in this report. The human visual system has evolved to observe natural images and therefore contains neural circuitry that is specifically adapted to analyse natural images. Thus, the field that studies natural image statistics has emerged to understand the input to the human visual system first, for the purpose of gaining a better understanding of the inner workings of the human visual system itself.

The discovery of natural image statistics follows several steps. First an ensemble of natural images is collected. We have outlined several issues that need to be resolved in this step. Second, the images are transformed into a specific domain. Examples include the computation of power spectra, phase spectra or wavelets, although many more are known and listed in this report. Finally, statistical regularities are computed in this domain.

There are a couple of important observations to make about human visual processing of natural images. For metabolic reasons, the HVS appears to have evolved to use a sparse coding. In essence, this means that most cells in the visual cortex lie dormant most of the time, while for each input only a few cells are active. This helps minimise energy consumption.

Thus, the HVS also applies several transforms to its input so as to induce a sparse coding. In the statistical sense, sparse

coding is often associated with probability density distributions that have a high kurtosis. Indeed, we have shown several transforms that result in highly kurtotic signals, and which approximate human visual processes. In particular, it can be shown that wavelets that analyse regions locally in space, orientation as well as frequency induce high kurtosis. Moreover, such analysers are thought to operate in the visual cortex.

Further, the human visual system appears to reduce redundancy in its input. The photoreceptors transduce a signal which is recombined several times in the retina, producing a spatial center-surround mechanism as well as colour opponency. It has been shown in particular that opponent colour spaces tend to be decorrelated. For natural images, the opponent signals tend to be close to independent. As a result, the information carried in its three channels is stripped of much of its redundancy.

Thus, the importance of discovering natural image statistics for understanding the human visual system is undeniable. However, the use of natural image statistics is not limited to the study of human vision. Particularly, in application areas such as computer graphics, machine vision as well as image processing, images are synthesized or analysed. By observing the aforementioned image statistics, we are able to produce more plausible images, enable new applications, or detect objects in images with higher success rates. Examples of these are given throughout this report.

Here, we highlight some of the more salient applications that employ some form of natural image statistics. Of these, wavelets are particularly useful for many tasks. They are used for instance in denoising algorithms as well as for object detection. However, perhaps its greatest success story lies in its use in lossy compression algorithms. As wavelets induce a sparse coding: histograms of wavelet coefficients are highly kurtotic so that most coefficients are small. Clamping these coeffients to zero means that an image can be reconstructed from a very small set of non-zero coefficients, leading to good opportunities of encoding only a small set of coefficients and thereby obtaining high compression factors.

A second example is found in the use of the sparseness of the distribution of gradients. Modeling the distribution of gradients of image ensembles has resulted in a simple functional form that can be used in optimisation problems as priors. This has been demonstrated in blind deconvolution algorithms which aim to remove blur from images due to camera shake. As an infinite number of solutions exist, these algorithms require additional constraints. In particular, they can drive the solution towards some measure of naturalness, such as the aforementioned distribution of gradients.

Finally, the decorrelation afforded by opponent colour spaces has been used directly in colour transfer algorithms. Here, the plausible transfer of colour palettes is made simpler by treating each channel independently.

We hope that this overview of the field of natural image statistics, and its link with graphics-related applications may help in the development of further applications in graphics.

References

[AAB96] ASADA N., AMANO A., BABA M.: Photometric calibration of zoom lens systems. In *Proceedings of the 13th International Conference on Pattern Recognition* (1996), vol. 1, pp. 186–190.

[Ach08] ACHARYYA R.: *A New Approach for Blind Source Separation of Convolutive Sources: Wavelet Based Separation Using Shrinkage Function*. VDM Verlag Dr. Müller Aktiengesellschaft & Co. KG, Saarbrücken, Germany, 2008.

[AEB10] Auto exposure bracketing settings by camera model. http://hdr-photography.com/aeb.html, 2010.

[AK04] ABADPOUR A., KASAEI S.: A fast and efficient fuzzy color transfer method. In *Proceedings of the 4th IEEE International Symposium on Signal Processing and Information Technology* (2004), pp. 491–494.

[AK07] ABADPOUR A., KASAEI S.: An efficient PCA-based color transfer method. *Journal of Visual Communication and Image Representation 18*, 1 (2007), 15–34.

[Any] ANYHERE SOFTWARE: Photosphere. http://www.anyhere.com/.

[APS98] ADAMS J., PARULSKI K., SPAULDING K.: Color processing in digital cameras. *IEEE Micro 18*, 6 (1998), 20–30.

[ASF91] ADELSON E., SIMONCELLI E., FREEMAN W. T.: Pyramids and multiscale representations. In *Representations and Vision*, Gorea A., (Ed.). Cambridge University Press, Cambridge, 1991, pp. 3–16.

[ASH87] ADELSON E., SIMONCELLI E., HINGORANI R.: Orthogonal pyramid transforms for image coding. In *SPIE Visual Communications and Image Processing* (1987), vol. II, p. 845.

[Bad96a] BADDELEY R. J.: An efficient code in V1? *Nature 381* (1996), 560–561.

[Bad96b] BADDELEY R. J.: Searching for filters with "interesting" output distributions: An uninteresting direction to explore? *Network: Computation in Neural Systems 7*, 2 (1996), 409–421.

[BAM*97] BADDELEY R. J., ABBOTT L. F., MICHAEL C. A., BOOTH C. A., SENGPIEL F., FREEMAN T., WAKEMAN E. A., ROLLS E. T.: Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proceedings of the Royal Society B: Biological Sciences 264*, 1389 (1997), 1775–1783.

[BCG90] BOVIK A. C., CLARK M., GEISLER W.: Multichannel texture analysis using localized spatial filters. *IEEE Pattern Analysis and Machine Intelligence 12*, 1 (1990), 55–73.

[BCHT01] BILLOCK V. A., CUNNINGHAM D. W., HAVIG P. R., TSOU B. H.: Perception of spatiotemporal random fractals: an extension of colorimetric methods to the study of dynamic texture. *Journal of the Optical Society of America A 18(10)* (2001), 2404–2413.

[BCT08] BILLOCK V. A., CUNNINGHAM D. W., TSOU B. H.: What visual discrimination of fractal textures can tell us about discrimination of camouflaged targets. In *Proceedings of Human Factors Issues: In Combat Identification and in Image Analysis Using Unmanned Aerial Vehicles* (2008).

[BF00] BRADY M., FIELD D. J.: Local contrast in natural images: normalisation and coding efficiency. *Perception 29* (2000), 1041–1055.

[BG00] BALBOA R. M., GRZYWACZ N. M.: Occlusions and their relationship with the distribution of contrasts in natural images. *Vision Research* (2000).

[BG03] BALBOA R. M., GRZYACZ N. M.: Power spectra and distribution of contrasts of natural images from different habitats. *Vision Research 43*, 24 (2003), 2527–2537.

[BGK01] BILLOCK V. A., GUZMAN G. C. D., KELSO J. A. S.: Fractal time and 1/f spectra in dynamic images and human vision. *Physica D 148* (2001), 136–146.

[BH96] BILLOCK V. A., HARDING T. H.: Evidence of the spatial and temporal channels in the correlational structure of human spatiotemporal contrast sensitivity. *Journal of Physiology 490*, 2 (1996), 509–517.

[Bil00] BILLOCK V. A.: Neural acclimation to 1/f spatial frequency spectra in natural images transduced by the human visual system. *Physica D 137* (2000), 379–391.

[BJ02] BACH F. R., JORDAN M. I.: Kernel independent component analysis. *Journal of Machine Learning Research 3* (2002), 1–48.

[BL09] BRADY M., LEGGE G.: Camera calibration for natural image studies and vision research. *Journal of the Optical Society of America A 26*, 1 (2009), 30–42.

[BM87] BURTON G. J., MOORHEAD I. R.: Color and spatial structure in natural scenes. *Applied Optics 26*, 1 (1987), 157–170.

[BMW88] BAUM E. B., MOODY J., WILCZEK F.: Internal representations for associative memory. *Biological Cybernetics 59* (1988), 217–228.

[BP86] BAYER B. E., POWELL P. G.: A method for the digital enhancement of unsharp grainy photographs. *Advances in Computer Vision and Image Processing 2* (1986), 31–88.

[BS95] BELL A. J., SEJNOWSKI T. J.: An information-maximization approach to blind separation and blind deconvolution. *Neural Computation 7*, 6 (1995), 217–234.

[BS97] BELL A. J., SEJNOWSKI T. J.: The 'independent components' of natural scenes are edge filters. *Vision Research 37* (1997), 3327–3338.

[Cae81] CAELLI T.: *Visual Perception: Theory and Practice*. Pergamon Press, Oxford, 1981.

[Car99] CARDOSO J. F.: High-order contrasts for independent component analysis. *Neural Computation 11*, 1 (1999), 157–192.

[CCO00] CHIAO C.-C., CRONIN T. W., OSORIO D.: Color signals in natural scenes: characteristics of reflectance spectra and the effects of natural illuminants. *Journal of the Optical Society of America A 17*, 2 (2000), 218–224.

[CF05] CLAUS D., FITZGIBBON A.: A rational function lens distortion model for general cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2005), pp. 213–219.

[CG87] CUTTING J., GARVIN J.: Fractal curves and complexity. *Perception and Psychophysics 42* (1987), 365 – 370.

[CHJ78] CAMPBELL F. W., HOWELL E. R., JOHNSON J. R.: A comparison of threshold and suprathreshold appearance of gratings with components in the low and high spatial frequency range. *Journal of Physiology 284* (1978), 193 – 201.

[CNR02] CARON J. N., NAMAZI N. M., ROLLINS C. J.: Noniterative blind data restoration by use of an extracted filter function. *Applied Optics 41*, 32 (2002), 6884–6889.

[Com94] COMON P.: Independent component analysis: A new concept. *Signal Processing 36*, 3 (1994), 287–314.

[COVC00] CHIAO C.-C., OSORIO D., VOROBYEV M., CRONIN T. W.: Characterization of natural illuminants in forests and the use of digital video data to reconstruct illuminant spectra. *Journal of the Optical Society of America A 17*, 10 (2000), 1713–1721.

[DA95] DONG D., ATICK J.: Statistics of time-varying images. *Network: Computation in Neural Systems 6* (1995), 345 – 358.

[Dau88] DAUGMAN J. G.: Complete discrete 2D gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech and Signal Processing 36*, 7 (1988), 1169–1179.

[Dau91] DAUGMAN J. G.: Self-similar oriented wavelet pyramids: Conjectures about neural non-orthogonality. In *Representations and Vision*, Gorea A., (Ed.). Cambridge University Press, Cambridge, 1991.

[DAW01] DROR R. O., ADELSON E. H., WILLSKY A. S.: Surface reflectance estimation and natural illumination statistics. In *In Proc. of IEEE Workshop on Statistical and Computational Theories of Vision* (2001).

[DCCH05] DEUSSEN O., COLDITZ C., COCONU L., HEGE H.: *Efficient Modeling and Rendering of Landscapes Visualization in Landscape and Environmental Planning*. Tylor & Francis, 2005.

[Deb] DEBEVEC P.: Hdrshop. http://www.hdrshop.com/.

[DF01] DONOHO D., FLESIA A.: Can recent innovations in harmonic analysis 'explain' key findings in natural image statistics? *Network: Computation in Neural Systems 12* (2001), 371–393.

[DHT*00] DEBEVEC P., HAWKINS T., TCHOU C., DUIKER H.-P., SAROKIN W., SAGAR M.: Acquiring the reflectance field of a human face. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (2000), pp. 145–156.

[DL05] DEUSSEN O., LINTERMANN B.: *Digital Design of Nature: Computer Generated Plants and Organics*. Springer-Verlag, New York, 2005.

[DLAW] DROR R., LEUNG T., ADELSON E., WILLSKY A.: Statistics of real-world illumination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 164–171.

[DM97] DEBEVEC P., MALIK J.: Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques* (1997), pp. 369–378.

[DOL00] DROR R. O., O'CARROLL D. C., LAUGHLIN S. B.: The role of natural image statistics in biological motion estimation. In *Biologically Motivated Computer Vision* (2000), pp. 492–501.

[DOL01] DROR R. O., O'CARROLL D. C., LAUGHLIN S. B.: Accuracy of velocity estimation by reichardt correlators. *J. Opt. Soc. Am. A 18*, 2 (2001), 241–252.

[DWA] DROR R. O., WILLSKY A. S., ADELSON E. H.:.

[DxO] DxO IMAGE SCIENCE: Automated optical corrections dedicated to your lenses. http://www.dxo.com/us/photo/dxo_optics_pro/optics_geometry_corrections.

[EBW92] ECKERT M. P., BUCHSBAUM G., WATSON A. B.: Separability of spatiotemporal spectra of image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence 14*, 12 (1992), 1210–1213.

[EMF03] EL-MELEGY M., FARAG A.: Nonmetric lens distortion calibration: Closed-form solutions, robust estimation and model selection. *Ninth International Conference on Computer Vision (ICCV 2003)* (2003), 554–559.

[ePa] EPAPERPRESS: Ptlens. http://epaperpress.com/ptlens/.

[Fai07] FAIRCHILD M.: The HDR Photographic Survey. In *Proceedings of the 15 th Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications* (2007), pp. 233–238.

[FB97] FIELD D. J., BRADY N.: Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural images. *Vision Research 37* (1997), 3367 – 3383.

[FBM98] FUNT B., BARNARD K., MARTIN L.: Is machine colour constancy good enough? In *Proceedings of the 5th European Conference on Computer Vision* (1998), pp. 445–459.

[FFC82] FOURNIER A., FUSSELL D., CARPENTER L.: Computer rendering of stochastic models. *Communications of the ACM 25*, 6 (1982), 371–384.

[Fie87] FIELD D. J.: Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A 4*, 12 (1987), 2379–2394.

[Fie93] FIELD D. J.: Scale-invariance and self-similar 'wavelet' transforms: An analysis of natural scenes and mammalian visual systems. In *Wavelets, fractals and Fourier transforms*, Farge M., Hunt J. C. R., Vassilicos J. C., (Eds.). Clarendon Press, Oxford, 1993, pp. 151–193.

[Fie94] FIELD D. J.: What is the role of sensory coding? *Neural Computation 6* (1994), 559–601.

[Fol95] FOLDIAK P.: Sparse coding in the primate cortex. In *The Handbook of Brain Theory and Neural Networks*, Arbib M. A., (Ed.). MIT Press, Cambridge, MA, 1995, pp. 895–989.

[FSH*06] FERGUS R., SINGH B., HERTZMANN A., ROWEIS S., FREEMAN W.: Removing camera shake from a single photograph. *ACM Transactions on Graphics (TOG) 25* (2006), 787–794.

[FT04] FINLAYSON G., TREZZI E.: Shades of gray and colour constancy. In *Proceedings of the Twelfth Color Imaging Conference* (2004), pp. 37–41.

[Gab46] GABOR D.: Theory of communication. *JIEE London 93*, III (1946), 429–457.

[GC05] GOLDMAN D., CHEN J.-H.: Vignette and exposure calibration and compensation. In *Tenth IEEE International Conference on Computer Vision (ICCV 2005)* (2005), vol. 1, pp. 899–906.

[Gei08] GEISLER W. S.: Visual perception and the statistical properties of natural scenes. *Annual Reviews in Psychology 59* (2008), 167–192.

[GG07] GIJSENIJ A., GEVERS T.: Color constancy using natural image statistics. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2007), pp. 1–8.

[Gib79] GIBSON J. J.: *The ecological approach to visual perception*. Lawrence Erlbaum, Hillsdale, NJ, 1979.

[GS04] GASPARINI F., SCHETTINI R.: Color balancing of digital photos using simple image statistics. *Pattern Recognition 37* (2004), 1201–1217.

[GS05] GEUSEBROEK J., SMEULDERS A.: A six-stimulus theory for stochastic texture. *International Journal of Computer Vision 62*, 1-2 (2005), 7–16.

[HB96] HAMMETT S. T., BEX P. J.: Motion sharpening: Evidence for the addition of high spatial frequencies to the effective neural image. *Vision Research 36*, 17 (1996), 2729 – 2733.

[HBS92] HANCOCK P., BADDELEY R. J., SMITH L.: The principal components of natural images. *Network: Computation in Neural Systems 3*, 1 (1992), 61–70.

[HDR] HDR SOFT: Photomatix. http://www.hdrsoft.com/.

[HHH09] HYVÄRINEN A., HURRI J., HOYER P. O.: *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision*. Springer, 2009.

[HK94] HEALEY G., KONDEPUDY R.: Radiometric ccd camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 16*, 3 (1994), 267–276.

[HK07] HARTLEY R., KANG S. B.: Parameter-free radial distortion correction with center of distortion estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 29*, 8 (2007), 1309–1321.

[HKO01] HYVÄRINEN A., KARHUNEN J., OJA E.: *Independent Component Analysis*. John Wiley & Sons, 2001.

[HLM00] HUANG J., LEE A., MUMFORD D.: Statistics of range images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2000), vol. 1, pp. 324–331.

[HM99a] HUANG J., MUMFORD D.: *Image Statistics for the British Aerospace Segmented Database*. Tech. rep., Division of Applied Math, Brown Univeristy, 1999.

[HM99b] HUANG J., MUMFORD D.: Statistics of natural images and models. In *Proceedings CVPR 99* (1999), pp. 541–547.

[HM99c] HUANG J., MUMFORD D.: Statistics of natural images and models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (1999), vol. 1, pp. 541–547.

[HT96] HIRANI A. N., TOTSUKA T.: Combining frequency and spatial domain information for fast interactive image noise removal. In *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (1996), pp. 269–276.

[Hyv99] HYVÄRINEN A.: Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks 10*, 3 (1999), 626–634.

[JBFZ02] JALOBEANU A., BLANC-FÉRAUD L., ZERUBIA J.: Estimation of blur and noise parameters in remote sensing. In *In Proc. of Int. Conf. on Acoustics, Speech and Signal Processing* (2002), pp. 249–256.

[JC79] JULESZ B., CAELLI T.: Perception. *On the limits of Fourier decompositions in visual texture perception 8* (1979), 69 – 73.

[JFS95] JACOBS C. E., FINKELSTEIN A., SALESIN D. H.: Fast multiresolution image querying. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1995), ACM, pp. 277–286.

[Kan07] KANG S.: Automatic removal of chromatic aberration from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2007), pp. 1–8.

[KFK90] KNILL D. C., FIELD D., KERSTEN? D.: Human discrimination of fractal images. *J. Opt. Soc. Am. A 7*, 6 (1990), 1113–1123.

[KP88] KUBE P., PENTLAND A.: On the imaging of fractal surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence 10* (1988), 704–707.

[KP08] KIM B., PARK R.: Automatic detection and correction of purple fringing using the gradient information and desaturation. In *Proceedings of the 16th European Signal Processing Conference* (2008).

[KW00] KANG S. B., WEISS R.: Can we calibrate a camera using an image of a flat, textureless lambertian surface? In *Sixth European Conference on Computer Vision (ECCV 2000)* (2000), pp. 640–653.

[Lan77] LAND E.: The retinex theory of color vision. *Scientific American 237*, 6 (1977), 108–128.

[Lev07] LEVIN A.: Blind motion deblurring using image statistics. *Advances in Neural Information Processing Systems* (2007).

[LGS99] LEE T.-W., GIROLAMI M., SEJNOWSKI T. J.: Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources. *Neural Computation 11*, 2 (1999), 417–441.

[LMF03] LEARNED-MILLER E. G., FISHER III J. W.: ICA using spacings estimates of entropy. *Journal of Machine Learning Research 4* (2003), 1271–1295.

[LS05] LITVINOV A., SCHECHNER Y. Y.: Radiometric framework for image mosaicking. *Journal of the Optical Society of America A 22* (2005), 839–848.

[LZW02] LEVIN A., ZOMET A., WEISS Y.: Learning to perceive transparency from the statistics of natural scenes. In *In NIPS-15; The 2002 Conference on Advances in Neural Information Processing Systems* (2002), MIT Press.

[LZW03] LEVIN A., ZOMET A., WEISS Y.: Learning how to inpaint from global image statistics. In *Proceedings of the Ninth IEEE International Conference on Computer Vision* (2003), pp. 305–312.

[Mal89] MALLAT S. G.: A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Pattern Analysis and Machine Intelligence 11* (1989), 674–693.

[Man83] MANDELBROT B. B.: *The Fractal Geometry of Nature*. Freeman, New York, 1983.

[Mar80] MARCELJA S.: Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America 70* (1980), 1297–1300.

[MFTF03] M. F. TAPPEN B. C. R., FREEMAN W. T.: Exploiting the sparse derivative prior for super-resolution and image demosaicing. In *Third International Workshop on Statistical and Computational Theories of Vision at ICCV* (2003).

[MN99] MITSUNAGA T., NAYAR S.: Radiometric self calibration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (1999), vol. 1, pp. 374–380.

[MNSA07] MOTOYOSHI I., NISHIDA S., SHARAN L., ADELSON E. H.: Image statistics and the perception of surface qualities. *Nature 447*, 7141 (2007), 206–209.

[MP95] MANN S., PICARD R.: On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proceedings of IS&T 48th annual conference* (1995), pp. 422–428.

[MW07] MALLON J., WHELAN P.: Calibration and removal of lateral chromatic aberration in images. *Pattern Recognition Letters 28*, 1 (2007), 125–135.

[NCB04] NEELAMANI R., CHOI H., BARANIUK R.: ForWaRD: Fourier-wavelet regularized deconvolution for ill-conditioned systems. *Signal Processing, IEEE Transactions on 52*, 2 (2004), 418–433.

[NP96] NIKIAS C., PETROPOLU A.: *Higher-Order Spectra Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1996.

[OF96]   OLSHAUSEN B. A., FIELD D. J.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature 381* (1996), 607–609.

[OF97]   OLSHAUSEN B. A., FIELD D. J.: Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research 37* (1997), 3311–3325.

[OPS*97]   OREN M., PAPAGEORGIOU C., SINHA P., OSUNA E., POGGIO T.: Pedestrian detection using wavelet templates. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on* (1997), pp. 193–199.

[OSS83]   OWSLEY C., SEKULAR R., SIEMSEN D.: Contrast sensitivity throughout adulthood. *Vision Research 23* (1983), 689 – 699.

[OT01]   OLIVA A., TORRALBA A.: Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision 42*, 3 (2001), 145–175.

[PBT98]   PÁRRAGA C. A., BRELSTAFF G., TROSCIANKO T.: Color and luminance information in natural scenes. *Journal of the Optical Society of America A 15*, 3 (1998), 563–569.

[PBTM98]   PÁRRAGA C. A., BRELSTAFF G., TROSCIANKO T., MOOREHEAD I. R.: Color and luminance information in natural scenes. *Journal of the Optical Society of America A 15*, 3 (1998), 563–569.

[POP98]   PAPAGEORGIOU C. P., OREN M., POGGIO T.: A general framework for object detection. In *Computer Vision, 1998. Sixth International Conference on* (1998), pp. 555–562.

[PS88]   PEITGEN H.-O., SAUPE D.: *The Science of Fractal Images*. Springer-Verlag, New York, 1988.

[PS00]   PORTILLA J., SIMONCELLI E. P.: Image denoising via adjustment of wavelet coefficient magnitude correlation. In *Proceedings of the 7th International Conference on Image Processing* (2000), pp. 277–280.

[PWK95]   PETERZELL D. H., WERNER J. S., KAPLAN P. S.: Individual differences in contrast sensitivity functions: Longitudinal study of 4-, 6- and 8-month-old human infants. *Vision Research 35*, 7 (1995), 9651–979.

[RAGS01]   REINHARD E., ASHIKHMIN M., GOOCH B., SHIRLEY P.: Color transfer between images. *IEEE Computer Graphics and Applications 21* (2001), 34–41.

[RB94]   RUDERMAN D., BIALEK W.: Statistics of natural images: Scaling in the woods. *Physical Review Letters* (1994).

[RB05]   ROTH S., BLACK M.: Fields of experts: A framework for learning image priors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2005), pp. 860–867.

[RB09]   ROTH S., BLACK M.: Fields of experts. *International Journal of Computer Vision 82*, 2 (2009).

[RBS03]   ROBERTSON M., BORMAN S., STEVENSON R.: Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *Journal of Electronic Imaging 12* (2003), 219–228.

[RCC98]   RUDERMAN D. L., CRONIN T. W., CHIAO C.: Statistics of cone responses to natural images: Implications for visual coding. *Journal of the Optical Society of America A 15*, 8 (1998), 2036–2045.

[RKAJ08]   REINHARD E., KHAN E. A., AKYÜZ A. O., JOHNSON G. M.: *Color Imaging: Fundamentals and Applications*. A K Peters, Wellesley, 2008.

[RSAT04]   REINHARD E., SHIRLEY P., ASHIKHMIN M., TROSCIANKO T.: Second order image statistics in computer graphics.

In *Proceedings of the 1st Symposium on Applied perception in graphics and visualization* (2004), pp. 99–106.

[RST01]   REINHARD E., SHIRLEY P., TROSCIANKO T.: *Natural image statistics for computer graphics*. Tech. Rep. UUCS-01-002, 2001.

[Rud94]   RUDERMAN D. L.: The statistics of natural images. *Network: Computation in Neural Systems*, 5 (1994), 517–548.

[Rud97]   RUDERMAN D. L.: Origins of scaling in natural images. *Vision Research 37* (1997), 3385–3398.

[RV90]   ROGOWITZ B., VOSS R.: Shape perception and low-dimension fractal boundaries. *Proceedings of the SPIE 1249* (1990), 387 – 394.

[RWPD05]   REINHARD E., WARD G., PATTANAIK S., DEBEVEC P.: *High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting*. Morgan Kaufmann Publishers, San Francisco, 2005.

[SA96]   SIMONCELLI E. P., ADELSON E. H.: Noise removal via bayesian wavelet coring. In *Proceedings of the Third International Conference on Image Processing* (1996), vol. I, pp. 379–392.

[SDS95]   STOLLNITZ E. J., DEROSE T. D., SALESIN D. H.: Wavelets for computer graphics: A primer. *IEEE Computer Graphics and Applications 15*, 3 (1995), 76–84.

[SDS96]   STOLLNITZ E. J., DEROSE T. D., SALESIN D. H.: *Wavelets for Computer Graphics: Theory and Applications*. Morgan Kaufmann, San Francisco, 1996.

[SHL95]   SHIH S., HUNG Y., LIN W.: When should we consider lens distortion in camera calibration. *Pattern Recognition 28*, 3 (1995), 447–461.

[Sim97]   SIMONCELLI E. P.: Statistical models for images: Compression, restoration and synthesis. In *31st Asilomar Conference on Signals, Systems and Computers* (1997), pp. 673–678.

[Sim99a]   SIMONCELLI E. P.: Bayesian denoising of visual images in the wavelet domain. In *Bayesian Inference in Wavelet Based Models*, Müller P., Vidakovic B., (Eds.). Springer-Verlag, New York, 1999.

[Sim99b]   SIMONCELLI E. P.: Modeling the joint statistics of images in the wavelet domain. In *Proceedings of the SPIE 44th Annual Meeting* (1999), vol. 3813, pp. 188–195.

[Sim05]   SIMONCELLI E.: Statistical modeling of photographic images. In *Handbook of Image and Video Processing*, Bovik A., (Ed.). Elsevier Academic Press, 2005, ch. 4.7, pp. 431–441.

[SJA08]   SHAN Q., JIA J., AGARWALA A.: High-quality motion deblurring from a single image. *ACM Trans. Graph. 27*, 3 (2008), 1–10.

[SJZW]   SHEN J., JIN X., ZHOU C., WANG C. C. L.: Gradient based image completion by solving poisson equation. *Computers and Graphics*, 119–126.

[SK99]   SAWHNEY H., KUMAR R.: True multi-image alignment and its application to mosaicing and lens distortion correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence 21*, 3 (1999), 235–243.

[SMS78]   SWITKES E., MAYER M. J., SLOAN J. A.: Spatial frequency analysis of the visual environment: Anisotropy and the carpentered environment hypothesis. *Vision Research 18*, 10 (1978), 1393–1399.

[SP98]   SIMONCELLI E. P., PORTILLA J.: Texture characterization via joint statistics of wavelet coefficient magnitudes. In *Proceedings of the 5th International Conference of Image Processing* (1998), vol. I, pp. 62–66.

[Ste97] STEIN G.: Lens distortion calibration using point correspondences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (1997), pp. 602–608.

[Stu05] STURM P.: Multi-view geometry for general camera models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (June 2005), pp. 206–212.

[SWO84] SEKULER R., WILSON H. R., OWSLEY C.: Structural modeling of spatial vision. *Vision Research 24*, 7 (1984), 689–700.

[Sze10] SZELISKI R.: *Computer Vision: Algorithms and Applications.* Springer, 2010.

[TAB*03] TELLER S., ANTONE M., BODNAR Z., BOSSE M., COORG S., JETHWA M., MASTER N.: Calibrated, registered images of an extended urban area. *International journal of computer vision* (2003).

[TF97] THOMSON M., FOSTER D.: Role of second- and third-order statistics in the discriminability of natural images. *JOSA A 14* (1997), 2081 – 2090.

[Tho99] THOMSON M. G. A.: Visual coding and the phase structure of natural scenes. *Network: Computation in Neural Systems 10* (1999), 123–132.

[Tho01] THOMSON M. G. A.: Beats, kurtosis and visual coding. *Network: Computation in Neural Systems 12* (2001), 271–287.

[TM01] TAUBMAN D., MARCELLIN M.: *JPEG2000: Image Compression Fundamentals, Standards and Practice.* Springer, 2001.

[TO03] TORRALBA A., OLIVA A.: Statistics of natural image categories. *Network: Computation in Neural Systems 14* (2003), 391–412.

[TP91a] TURK M., PENTLAND A.: Eigenfaces for recognition. *Journal of cognitive neuroscience 3*, 1 (1991), 71–86.

[TP91b] TURK M., PENTLAND A.: Face recognition using eigenfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (1991), pp. 586–591.

[TSW*05] TAYLOR R., SPEHAR B., WISE J., CLIFFORD C., NEWELL B., HAGERHALL C., PURCELL T., MARTIN T.: Perceptual and physiological responses to the visual complexity of fractal patterns. *Nonlinear Dynamics Psychol Life Sci 9* (2005), 89 –114.

[TT93] TADMOR Y., TOLHURST D. J.: Both the phase and the amplitude spectrum may determine the appearance of natural images. *Vision Research 33*, 1 (1993), 141–145.

[TT94] TADMOR Y., TOLHURST D.: Discrimination of changes in the second order statistics of natural and synthetic images. *Vision Research 34* (1994), 541 –554.

[TTC92] TOLHURST D. J., TADMOR Y., CHIAO T.: Amplitude spectra of natural images. *Ophthalmic and Physiological Optics 12* (1992), 229–232.

[van] VAN WALREE P.: Spherical aberration. http://toothwalker.org/optics/spherical.html.

[van92] VAN HATEREN J.: Theoretical predictions of spatiotemporal receptive fields. *Journal of Comparative Physiology A 171* (1992), 151 – 170.

[van98a] VAN DER SCHAAF A.: *Natural image statistics and visual processing.* PhD thesis, Rijksuniversiteit Groningen, The Netherlands, 1998.

[van98b] VAN DER SCHAAF A.: Natural image statistics and visual processing.

[vG05] VAN DE WEIJER J., GEVERS T.: Color constancy based on the grey-edge hypothesis. In *Proceedings of the International Conference on Image Processing* (2005).

[Vis07] VISION CLUB OF FINLAND: Electronic shuttering for high speed cmos machine vision applications. http://www.automaatioseura.fi/jaostot/mvn/mvn2007/parameter.html, 2007.

[VJ01] VIOLA P., JONES M.: Robust real-time object detection. *International Journal of Computer Vision* (2001).

[Vos85] VOSS R.: Random fractal forgeries. In *Fundamental Algorithms for Computer Graphics*, Earnshaw R. A., (Ed.). Springer, Berlin, 1985, pp. 805–835.

[vv96] VAN DER SCHAAF A., VAN HATEREN J.: Modelling the power spectra of natural images: statistics and information. *Vision Research 36*, 17 (1996), 2759–2770.

[vv98] VAN HATEREN J., VAN DER SCHAAF A.: Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London B 265* (1998), 359–366.

[WA89] WATSON A. B., AHUMADA JR. A. J.: A hexagonal orthogonal oriented pyramid as a model of image representation in visual cortext. *IEEE Transactions on Biomedical Engineering 36*, 1 (1989), 97–106.

[Wat87] WATSON A. B.: Efficiency of an image code based on human vision. *Journal of the Optical Society of America A 4*, 12 (1987), 2401–2417.

[Wat90] WATSON A. B.: Perceptual components architecture for digital video. *Journal of the Optical Society of America 7* (1990), 1943–1954.

[Wat91] WATSON A. B.: Multidimensional pyramids in vision and video. In *Representations of Vision*, Gorea A., (Ed.). Cambridge University Press, Cambridge, 1991, pp. 17–26.

[WCH92] WENG J., COHEN P., HERNIOU M.: Camera calibration with distortion models and accuracy evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 14*, 10 (1992), 965–980.

[Wil94] WILLSON R.: *Modeling and calibration of automated zoom lenses.* PhD thesis, Pittsburgh, PA, USA, 1994.

[WM97] WEBSTER M., MIYAHARA E.: Contrast adaptation and the spatial structure of natural images. *Journal of the Optical Society of America A 14* (1997), 2355–2366.

[ZL97] ZIEGAUS C., LANG E. W.: Statistics of natural and urban images. In *Proceedings 7th International Conference on Artificial Neural Networks* (Berlin, 1997), vol. 1327 of *Lecture Notes in Computer Science*, Springer-Verlag, pp. 219–224.

[ZL98] ZIEGAUS C., LANG E. W.: Statistical invariances in artificial, natural and urban images. *Z. Naturforsch 53a* (1998), 1009–1021.

[ZLK10] ZHENG Y., LIN S., KANG S.: Single-image vignetting correction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2010), vol. 1, pp. 461–468.