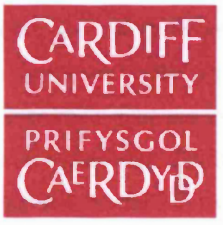
Sparse Modelling of Natural Images and Compressive Sensing

Nassir Mohammad

School of Mathematics, Cardiff University, UK
& Hewlett Packard Laboratories, Bristol, UK

A Thesis submitted for the degree of Doctor of Philosophy



June 9, 2011

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

This thesis concerns the study of the statistics of natural images and compressive sensing for two main objectives: 1) to extend our understanding of the regularities exhibited by natural images of the visual world we regularly view around us, and 2) to incorporate this knowledge into image processing applications.

Previous work on image statistics has uncovered remarkable behavior of the distributions obtained from filtering natural images. Typically we observe high kurtosis, non-Gaussian distributions with sharp central cusps, which are called *sparse* in the literature. These results have become an accepted fact through empirical findings using zero mean filters on many different databases of natural scenes. The observations have played an important role in computational and biological applications, where researchers have sought to understand visual processes through studying the statistical properties of the objects that are being observed. Interestingly, such results on sparse distributions also share elements with the emerging field of compressive sensing. This is a novel sampling protocol where one seeks to measure a signal in already compressed format through randomised projections, while the recovery algorithm consists of searching for a constrained solution with the sparsest transformed coefficients.

In view of prior art, we extend our knowledge of image statistics from the monochrome domain into the colour domain. We study sparse response distributions of filters constructed on colour channels and observe the regularity of the distributions across diverse datasets of natural images. Several solutions to image processing problems emerge from the incorporation of colour statistics as prior information. We give a Bayesian treatment to the problem of colorizing natural gray images, and formulate image compression schemes using elements of compressive sensing and sparsity. We also propose a denoising algorithm that utilises the sparse filter responses as a regularisation function for the effective attenuation of Gaussian and impulse noise in images. The results emanating from this body of work illustrate how the statistics of natural images, when incorporated with Bayesian inference and sparse recovery, can have deep implications for image processing applications.

Declarations

Statement 1

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in any candidature for any degree.

Signed 

Date 16.02.2011

Statement 2

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by explicit references and a bibliography is supplied.

Signed 

Date 16.02.2011

Statement 3

I hereby consent for my thesis, if accepted, to be available for photocopying and inter-library loan, and for the title and summary to be made available to outside organisations.

Signed 

Date 16.02.2011

Acknowledgments

It has been a privilege to study for a research degree in the mathematical field of signal processing. My gratitude goes to Professor Alexander Balinsky for being such a knowledgeable and enthusiastic supervisor. Your patience, expertise and most importantly, guidance, has been invaluable.

My gratitude also extends to Dr Stephen Pollard and Andrew Hunter from Hewlett Packard Labs in Bristol for their supervision, guidance and experience of the industrial technology sector. At home I would like to thank my family for supporting me and allowing me to pursue this Doctorate with ease, comfort and peace. My appreciation and thanks also extends to all the people at the School of Mathematics in Cardiff University for making it such a pleasant and interesting place of learning.

I would like to acknowledge the various sources of funding that have allowed me to fulfill my PhD without financial restriction and hindrance. Your assistance has made my years as a student a tremendously enjoyable experience: EPSRC, Hewlett Packard Labs, The Smith Institute Knowledge Transfer Network and Cardiff University.

In precisely built mathematical structures, mathematicians find the same sort of beauty others find in enchanting pieces of music, or in magnificent architecture. There is, however, one great difference between the beauty of mathematical structures and that of great art. Music by Mozart, for instance, impresses greatly even those who do not know musical theory; the cathedral in Cologne overwhelms spectators even if they know nothing about Christianity. The beauty in mathematical structures, however, cannot be appreciated without understanding of a group of numerical formulae that express laws of logic. Only mathematicians can read "musical scores" containing many numerical formulae, and play that "music" in their hearts...

Kiyosi Ito

To my 'Abba' and 'Amma'

Dissemination of Results

Publications

- A. Balinsky and N. Mohammad, Non-linear filter response distributions of natural colour images. Proceedings of the Computational Color Imaging Workshop, Saint Etienne, France, Lecture Notes in Computer Science, 5646, pp. 101-108 2009. Springer-Verlag Berlin Heidelberg 2009.

- A. Balinsky and N. Mohammad, Colorization of natural images via L1 optimization, proceedings of the Workshop on Applications of Computer Vision (WACV), 2009 IEEE Winter Vision Meetings, Snowbird, Utah, USA.

- A. Balinsky and N. Mohammad, Sparse natural image statistics and their applications to colorization and compression, proceedings of the IEEE International Conference on Image Processing (ICIP) 2010, 26th-29th Sep, Hong Kong.

- A. Balinsky and N. Mohammad, Chroma reconstruction from inaccurate measurements, proceedings of the 19th International Conference on Computer Graphics, Visualization and Computer Vision, 31st Jan-3rd Feb 2011, Plzen, Czech Republic.

- A. Balinsky and N. Mohammad, Non-Linear Filter Response Distributions of Natural Images and Applications in Image Processing, Appl. Math. Inf. Sci. 5, 3, (2011) pp. 367-389

Conference Presentations

- Statistical modelling of natural images, University of Wales Gregynog colloquium, UK, 25th May 2008.

- Statistical modelling of natural images, Alan Taylor day, Oxford, UK, 24th Nov 2008.

- Non-linear filter response distributions of natural colour images, 2009 Computational color imaging workshop, Saint Etienne, France, 26-27 March 2009.
- Non-linear filter response distributions of natural colour images and the colorization problem, Visual computing in Wales workshop 2009, Aberystwyth, UK, 1st April 2009.
- Natural image statistics and their applications to colorization and compression, University of Wales Gregynog colloquium, UK, 19th May 2009.
- Non-linear filter response distributions of natural colour images and applications, BMVA annual summer school 2009, Kingston University, Kingston, UK, 5th -10th July 2009.
- Natural image statistics and their applications to colorization and compression, (invited talk to) BMVA UK student vision workshop, Queen Mary University, London, UK, 11th Sep 09.
- Colorization of natural images via L1 optimization, Graduate School in Physical Sciences and Engineering Annual Lecture, Cardiff, UK, 11th Nov 09.
- Natural image statistics and their applications to colorization and compression, Alan Taylor day, Oxford, UK, 30th Nov 2009.
- Colorization of natural images via L1 optimization, Workshop on applications of computer vision, 2009 IEEE Winter Vision Meetings, Snowbird, Utah, USA, 7-10th Dec 09.
- Chroma reconstruction from inaccurate measurements, 19th International Conference on Computer Graphics, Visualization and Computer Vision, 31st Jan-3rd Feb 2011, Plzen, Czech Republic.

Seminars

- Statistical modelling of natural images, Cardiff School of Maths, Cardiff, UK, 29 Jan 2008.

- Statistical modelling of natural images and applications, HP Labs Bristol, UK, 29 April 2008.
- Statistical modelling of natural images and compressive sampling, Vision lunch, Cardiff School of Computer Science, Cardiff, UK, 26th Feb 2009.

Attendances

- British Machine Vision Association/EPSRC Annual Summer School, July 2009, Kingston University, UK.
- Mathematics in Medicine Study Group, Sep 2009, Imperial College London, UK.
- SIAM imaging conference, April 2010, Chicago, USA.
- London Mathematical Society (LMS) 2011 Invited Lectures on compressed sensing by Emmanuel Candes, 21-25 March 2011, Cambridge, UK.

Contents

Chapter 1

Introduction

1.1 Objective

Our aim in this thesis is to understand the statistics of natural images and to use this knowledge in image processing. For the former task we study statistics of natural colour images to learn about the inherent structures and symmetries found within them. We aim to extend fundamental results of non-Gaussian image statistics found in monochrome images, where the distributions are largely symmetric and exhibit high kurtosis, i.e. *sparse*. We also aim to incorporate ideas from the emerging field of compressive sensing to illustrate interesting connections between the sparse statistical properties of natural images and the new sensing modality. Subsequent work looks to incorporate image statistics into mathematical models in order to increase the performance of image processing applications. Problems we will consider include sensing, compression, denoising and colorization.

1.2 Motivation

No time in human history has ever witnessed such explosive influence and impact of image processing on society, the sciences and engineering [31]. Problems in this field occur in a host of diverse subjects due to the fact that vision plays

such an important role in human endeavours. As a few examples we see that in medicine, imaging techniques are utilised to non-invasively diagnose diseases and abnormalities in the human body. Security agencies utilise advances in imaging for gathering intelligence and identifying targets of interest. In astronomy image processing techniques are utilised to sense deep into space, while remote sensing is also utilised to monitor changes back on earth. The entertainment, publishing and digital communication industries utilise techniques to create realistic graphics, movies, automated documents, and image enhancement software. These examples illustrate that image processing has become an integral part of the current technological society.

As a branch of signal processing, image processing was initially built upon the machinery of Fourier and spectral analysis. Over recent years, there have emerged numerous novel competing methods and tools for successful image processing. They include, for example, stochastic approaches based upon Gibbs/Markov random fields and Bayesian inference theory, variational methods incorporating various geometric regularities, linear and nonlinear partial differential equations, as well as applied harmonic analysis centered around wavelets [31].

Recently there has been a great deal of interest in the properties of a particular class of images called 'natural images', which has arisen from both a biological and computational point of view. Natural images are set apart from man-made images due to the ease with which humans are able to distinguish between the two. This fact arises from the particular types of structures that appear in natural scenes and whose texture and colour are not completely realisable through artificial efforts. However, a full definition as to what constitutes a natural image is yet to be defined. Furthermore, natural images are very rare in the space of all possible images that could occur. Even a modestly sized 16×16 image composed of 256 gray levels gives rise to an astronomical 256^{256} number of images that could possibly occur. In this enormous space there exist natural images where if we tried to randomly construct an image on a computer we would practically never realise a natural scene, or even a tree.

In this body of work we loosely define natural images to be those images

which humans 'see' in their every day lives. Images of mountains, buildings, people, vegetation, animals, indoor rooms, etc. are all considered to be natural scenes. On the other hand images of random noise patterns, cartoons, video games, paintings, etc. are considered artificial. The classification is of course weak as a painting, once created, could be thought of as being a natural image since it has been visualised by the people viewing it. However, in our work we focus the meaning onto images that we 'regularly' see in our day to day lives, leaving aside an exact definition.

The fact that natural images appear to be non-random quantities motivates the drive to understand these images. This has led to the finding of some striking statistical symmetries and most notably the discovery that natural images display non-Gaussian image statistics, which is evident on virtually any histogram of the filter response images on virtually any database of natural images [64]. The single pixel intensity histograms constructed all essentially have kurtosis greater than that of the Gaussian distribution, typified by a sharp central cusp in the distribution. These distributions have taken the name of *sparse* due to connections with sparse coding.

Another interesting empirical finding has been the invariance of image statistics to scale. This means that any local statistic calculated on $n \times n$ images and on block averaged $2n \times 2n$ images should be the same. Although not exact, this result has been approximately confirmed on many large image databases. The reason this is exciting is that it implies that local image models describing the small-scale structures in images will work as global image models describing the large scale structures in images. This creates a surprising stability for image statistics [64].

The non-Gaussianity and scale invariance in natural images have been found to occur regardless of image content. The immediate benefits gained by the statistical properties have been appreciated in a host of scientific fields. From a visual perception or biological point of view the incorporation of the statistics of natural images has been viewed as a critical element in the optimisation goals of a visual system. The theory is based on construction of statistical models

of images combined with Bayesian inference. This approach originates from the ideas that the visual world has a great influence on the physical design of biological creatures' visual systems, with one of the most important aspects being the fluctuations of light intensity in the environments. It is in this signal that information about the surroundings are conveyed and Bayesian inference shows how we can use prior information on the structure of typical images to greatly improve image analysis, and statistical models are used for learning and storing that prior information [86]. Contributions in this area have shown how redundancy minimisation or decorrelation [38], maximisation of information transmission [11], sparseness of the neural encoding [80], and minimising reconstruction error [65] can predict visual processes.

From a computational perspective image priors have been, and still are, being successfully used in many image processing tasks. Investigators have utilised the knowledge to invent more effective denoising [90] and deblurring [66] algorithms, as well as improvements in realistic super-resolution [94], colorization [5] and inpainting [47] applications. The statistics of images have also been incorporated into various scene categorisation [97], object recognition, face detection and clutter classification tasks [92]. Additionally, models motivated by vision research have found applications in document processing in tasks for automatic keyword extraction, classification and registration [4].

The success of prior art in image processing motivates the main goals of the project: **to investigate mathematically and statistically some internal features of natural images and to incorporate some new mathematical tools and methods into real digital devices.**

1.3 Contributions

In the following paragraphs we concisely describe the novel findings of the thesis: Our first main result shows how non-linear filters operating on colour images display non-Gaussian and high kurtosis distributions. These statistics are then modelled using generalised Gaussian distributions whose shape parameters con-

firm the distributions to indeed be sparse. This empirical finding is justified through the use of a diverse dataset of natural images, captured or compressed by arbitrary forms. Essentially the result categorises a particular form of regularity exhibited by natural colour images with limited variability due to changing scenes. This is an extension and addition to the non-Gaussian behaviour of image statistics observed across monochrome images.

Subsequently, we develop a collection of applications which form connections with the emerging field of compressive sensing and sparse recovery: We begin by giving a novel Bayesian analysis for the problem of colorizing natural gray images. We show how this long standing problem in image processing can be solved using prior information provided by user interaction in the form of colour points, and utilisation of sparse priors. The derived non-convex optimisation problem is convexified using L^1 optimisation, and transformed into a constrained linear program in line with the ideas of recovering sparse signals. The importance of the method is that it illustrates how sparse priors can be incorporated into colorizing natural images.

The next application we develop turns around the idea of colorization to present a colour image compression scheme where standard (e.g. JPEG) compression is used for the monochrome component, while colour data is sampled and compressed using sparse binary random matrices. In the case of sampling 'seed' pixels the novel contribution shows that L^1 optimisation can be used as a decompression tool. Another main novelty is that compressive sensing measurements can be made of the colour channels and reconstruction accomplished by L^1 optimisation. The importance of this work is that it incorporates compressive sensing in a first presentation of a colour image compression scheme which utilises gray level information for reconstruction.

In a further contribution we present an effective image denoising algorithm whose novelty lies in utilising the non-linear filter response on the colour channels of images as a regularization function. This leads to an optimisation scheme that seeks a solution image with the sparsest filter response while also being encouraged to be close to the noisy measurement. We observe effective attenuation

(through the specified parameters) of noise in real images, as well as images artificially corrupted by white and impulse noise.

1.4 Outline of the thesis

The thesis is outlined in the following way: Chapter 2 reviews previous work on the statistics of natural images and presents the theories that have been proposed to model their behaviour: the generalised Gaussian distribution and the Bessel distributions. This chapter follows the works of [74, 60, 93, 46, 91]. Chapter 3 introduces the emerging field of compressive sensing from its basic underpinnings to some interesting new results that are utilised later in the thesis. This chapter follows the works of [21, 83, 10]. Chapter 4 presents the first of our contributions where we empirically show and model the non-linear filter response distributions of natural colour images. The following chapters then present various applications that exploit knowledge of image statistics. Chapter 5 gives a Bayesian analysis of the problem of colorizing natural gray images. Chapter 6 describes a compression scheme that utilises the sparse nature of image statistics and elements from compressive sensing and colorization. Chapter 7 presents a denoising scheme for colour images that removes unwanted noise in real images, and images affected by white and impulse noise. Chapter 8 discusses the results obtained in the thesis and summarises directions for future research. Chapter 9 give the tables, graphs and example images used throughout the thesis. Finally, the Bibliography provides all the references to this work.

Chapter 2

Statistics of Natural Images & Models

The study of natural scenes is an endeavour that seeks to find fundamental properties exhibited by natural images in order to aid us in understanding the visual world. This chapter presents some statistical properties of natural images which have aided technological advancements over the last few decades. These include the high redundancy found in image data [93], scale invariance [85] and non-Gaussianity [60] of image statistics, exhibition of symmetries, existence of transforms that can sparsely represent images for compression and models that have been proposed for fitting the observed distributions [92, 90].

The incorporation of image statistics arises through the formulation of vision problems as a problem in Bayesian inference [59, 73, 45]. This has become a widely established and accepted procedure in the computer vision community. However, the idea that the problem of recovering $3D$ or $2D$ information from a $2D$ image is ill-posed and requires inference, can be traced back several hundred years. The Arab scientist Ibn Al-Haytham (known to Europe as Alhazan) around the year 1000 [54] was the first to demonstrate that light rays are emitted by an external source and travel in straight lines, reflecting and refracting with physical objects until they reach the eye. This information is then required

to be interpreted in the human brain through a largely unconscious and very rapid inference process based on past visual experiences. This was a significant step in its time and did away with the conflicting theories that he had inherited from the Greeks. In modern times the inferences that underly visual perception have been intensely studied by several researchers, notable examples include H. Helmholtz, E. Brunswick and J.J. Gibson [74].

Following the work of [74], let us study this process through mathematics and Bayesian analysis, and take I to be static images of the visual world. This static assumption allows us to simplify the analysis, although it can be deemed a little unrealistic as living organisms are adapted to environments that require the handling of motion and changing scenes. The images I then record the light intensity falling at each point position on the $2D$ grid of pixels. Next we introduce variables w which stand for descriptors that describe objects, their textures, colour, shading, boundaries, orientation, etc. that generate images. This leads to two stochastic models which are learned from past experience: a prior model $\mathbb{P}(w)$ specifying what scenes are likely in the visual world we live in, and an imaging model $\mathbb{P}(I/w)$ which specifies what images should look like, given the scene.

Thus by Bayes's rule we have,

$$\mathbb{P}(w/I) = \frac{\mathbb{P}(I/w)p(w)}{\mathbb{P}(I)} \propto \mathbb{P}(I/w)\mathbb{P}(w). \tag{2.0.1}$$

The above concisely asks the following. Fixing the observed value I, find that w which maximises the measure $\mathbb{P}(w/I)$ or equivalently maximises $\mathbb{P}(I/w)p(w)$. Thus the vision problem we have is now interpreted as a problem in Bayesian inference, and requires us to establish a theory of stochastic models that can express all the variable patterns that w and I obey. This formulation also requires a model that can learn from past experience and efficient computational methods for finding the maximum of $\mathbb{P}(w/I)$. Bringing together these requirements into a fitting model enables accurate and efficient performance of vision orientated tasks. However, in this chapter we will largely concentrate on the building of stochastic models, with learning models and computing methods given via some of the image processing applications presented in later chapters.

We can proceed with the study of images in a number of ways. However, one approach is to look at ensembles of images of particular categories, or large variable databases of images. In such schemes we construct datasets that are deemed large enough to sample the possible images which that set represents. We can then analyse images through the use of zero mean filters, individual and joint histograms and can build stochastic models $\mathbb{P}(I)$ which replicate the properties of natural scenes.

Before proceeding further, let us give some definitions that will prove useful in later sections. Let X and Y be two random variables on $\mathbb{R}$ with μ and σ^2 representing the mean and variance of a random variable.

Definition 2.0.1. (Kurtosis) *We define the kurtosis of a random variable X to be*

$$k = \{\mathbb{E}(X-\mu)^4\}/(\sigma)^4. \tag{2.0.2}$$

Kurtosis is a measure of the 'peakedness' of the probability distribution of a real-valued random variable. A higher value implies that the peak of the distribution is more pronounced while a lower value gives wider shoulders to the distribution. In addition, a higher kurtosis value implies that the tails of the distribution are heavier, and that more of the variance is due to infrequent extreme deviations. Kurtosis can also be taken to be one measure of non-Gaussianity, however, due to the sensitivity to outliers, care must be taken when using this as a measure.

Definition 2.0.2. (Skewness) *We define skewness of a random variable X to be*

$$S = \{\mathbb{E}(X-\mu)^3\}/(\sigma)^3. \tag{2.0.3}$$

Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable. Positive skew implies the right tail is longer, i.e. the mass of the distribution is concentrated on the left of the figure. Negative

skew implies that the left tail is longer, the mass of the distribution is concentrated on the right of the figure. Skewness can also show the non-Gaussian nature of a distribution.

Definition 2.0.3. (Covariance) *We define the covariance between two random variables X and Y, with $\mathbb{E}(X) = \mu$ and $\mathbb{E}(Y) = \nu$ to be*

$$Cov(X, Y) = \mathbb{E}\{(X - \mu)(Y - \nu)\}, \tag{2.0.4}$$

equivalently

$$Cov(X, Y) = \mathbb{E}(XY) - \mu\nu, \tag{2.0.5}$$

where if X and Y are independent $Cov(X, Y) = 0$. Random variables whose covariance is zero are called uncorrelated.

In probability theory and statistics covariance is a measure of how two random variables vary together. From the definition we can see that if two variables tend to vary together, then the covariance between the two variables will be positive. However, if they vary in opposite ways with respect to their means then the covariance will pull towards negative.

2.1 Redundancy in Images

One of the earliest works on the study of natural images was performed by television engineers during the 1950's. Concerned with the real application and gains that could be made in increasing compression and transmission rates, they studied natural scenes in order to gauge their levels of predictability. Kretzmer [57], for example, studied simple image statistics such as histograms of image intensity and the correlation of nearby pixels to increasing distance, and showed that as pixel distances increased, the correlation between pixels decreased. Such initial studies showed that natural images showed signs of redundancy. Other pointers to redundancy were proposed by Attneave [2] where images were partitioned off so that the rare transition edges that occur between objects can be thought of as the unexpected events, and hence, in line with ideas of information theory can be thought of as the information bearing elements of an image.

The predictability of natural images was also studied experimentally by Kersten [58]. He was inspired by Shannon's illustration of redundancy exhibited by the English language [88], and wanted to apply similar techniques to derive similar results on the predictability of natural scenes. His method involved the removal of portions of intensity values from pixel's in an image, with humans subjects being asked to replace those values. The number of guesses until a correct response was tabulated to form a histogram and showed the single pixel redundancy to be within the range 67% to 81%.

In later work Ruderman [85] studied images of forest scenes in a first of several papers to study redundancy and scale invariance in natural images. He assumed that organisms have adapted to their environments and that they take advantage of the statistical regularities that the input signals display to enable efficient representation. Such quantities as local contrast in images were found to display scale invariance and exponential tails, while the power spectra of natural scenes also displayed non-trivial exponents, similar to the exponents found in phase transitions.

In further work Ruderman [86] computed the mutual information between pairs of pixels with increasing distance. The results confirmed earlier similar results that indicate information varying as a power law as separation distance between two pixels increases. His work calculated the redundancy in images to be around 10%, a much lower value than that obtained by Kersten [58] due to the wooded nature of his dataset from a forest environment.

Studies on the redundancy exhibited by images paved the way for future research into the statistical properties of natural scenes. Among several consequences and applications were that the predictability showed how high levels of compression were theoretically achievable. Additionally, the coupling with biological vision showed how systems optimised for sparse linear codes of natural scenes developed a complete family of localized, oriented, bandpass receptive fields, similar to those found in the primary visual cortex [79]. These results support Barlow's theory that the goal of natural vision is to represent the information in the natural environment with minimal redundancy [9]. They also

provide insight and connections to some surprising empirical observations which we present in the following sections.

2.2 Scale Invariance and Implications

This section closely follows the work of [74] where we discuss the remarkable property of images that they show approximate invariance to scale. This property can be described with the following example: Given 64×64 sized images from a database of images from the world we view around us, and a probability distribution model $p_{64}(I)$ in the Euclidean space $\mathbb{R}^{4096}$ of all such images, we can form a new database of the same set of images except that each image has been reduced to a size of 32×32 pixels. This can be accomplished by simply taking the central 32×32 pixel region of each image, or by filtering each image using an averaging 2×2 filter. Thus the statement that images are from a scale invariant process implies that the two resulting marginal distributions on both databases of images are approximately the same, and this experimental result should happen for images of any size.

The phenomenon of scale invariance has attracted many researchers and several explanations have been put forward. One interesting proposition is that the world we live in consists of objects organised in collections of differing sizes. These objects are viewed at arbitrary distances and angles and their arrangement leads to the scale invariance behaviour. For example, a picture of a room may contain a desk, which in turn may have books and papers, which in turn may have writing. Thus zooming into an image results in showing similar numbers of objects and which also occur with the same levels contrast.

The existence of scale invariant statistics has been found to be one of the most robust qualities of the visual world. This is a phenomenon that has been approximately found in different visual environments and experimental results indicate that it manifests itself the beyond second-order measure given by the power spectrum. A most surprising consequence of this observation is that researchers have found the property regardless of the types of image databases

considered. Furthermore, datasets constructed such that no two images were ever from same environment have shown this remarkable behaviour [86].

The reason this is exciting is that it implies that local image models describing the small-scale structures in images will work as global image models describing the large scale structures in images. Furthermore, as the histograms obtained are scale-invariant then so are all expected moments of the filter responses [64]. Another surprising feature in the data of natural image scaling is invariance to choice of *calibration*. This implies that when images vary from raw uncalibrated pixel values, to calibrated luminance and to the logarithm of luminance, the scale invariant feature of natural images still holds [86].

A consequence of scale invariance is the law for the decay of power at high frequencies in the Fourier transform of images. It states that the expected power as a function of frequency should fall like:

$$\mathbf{E}_f(|\hat{I}(|\lambda,\mu)|^2) \approx C/(\lambda^2+\mu^2) = C/f^2, \tag{2.2.1}$$

where $f = \sqrt{\lambda^2+\mu^2}$ is the spatial frequency [74].

One the earliest findings of power-law scaling in image power spectra was in 1957 by Deriugin [36] with regard to television signals. These findings were re-discovered in 1987 by Burton and Moorhead [20] and Field [42]. The experiments conducted on databases of natural images over the decades point to the following conclusions: Databases containing images from certain categories such as forestry, mountains, cities, sky, etc. are found to have different powers of best fit. These can range from $1/f^3$ to $1/f$ but with a high concentration near $1/f^2$, and with a large variance [74, 46]. However, the ensemble power spectrum across large databases containing varieties of images follows the general rule. In the following sections we will explore the power law phenomenon in greater detail in its equivalent form in the image domain under the context of non-Gaussian spatial statistics.

From a mathematical point of view the approximate invariance displayed by images have some interesting consequences with regard to the types of functions that can be used to model them. Indeed, images, due to their infinite resolution cannot be considered functions at all, but rather as 'generalised functions'. This

is due to the obvious fact that the values at any point in an image do not, in the limit approach any defined value, as we zoom further into an image [48].

However, several researchers have proposed various function spaces for random images. For example, Mumford and Shah [76] proposed that the observed images are a sum:

$$I(i,j) = u(i,j) + v(i,j), \tag{2.2.2}$$

where u is the piecewise smooth 'cartoon', representing the important content of the image, and v some L^2 noise. This formulation leads to the function spaces of images to be the space of functions with bounded variation after the removal of noise, i.e. $\int |\nabla I| didj < \infty$ [74]. However, one disadvantage of this model is that texture and noise are considered as one, whereas in real images textures form an important part of an image.

Thus more recent models introduced by DeVore and Lucier [37] have proposed that:

$$I(i,j) = u(i,j) + v(i,j) + w(i,j), \tag{2.2.3}$$

where u is the cartoon, v the true texture and w is the noise. Here they state that the true image $u+v$ belongs to a suitable Besov space, i.e. the space of functions $f(i,j)$ for which bounds are put on the L^p norm of $f(i+h,j+k) - f(i,j)$ for (h,k) small. This approach was also simplified by Carasso [28] who hypothesises that images I, after removal of 'noise' should satisfy:

$$\int |I(i+h,j+k) - I(i,j)| didj < C(h^2+k^2)^{\alpha/2}, \tag{2.2.4}$$

for some α as $(h,k) \to 0$.

The noise that we see in captured images can be argued as being nothing more than mere 'clutter'. This essentially describes certain masses of objects whose detail in a scene are too small to be fully resolved by the resolution of a given camera. Thus, what we describe as noise which does not usually follow any standard probability distribution are simply objects that have been blurred

beyond recognition. An ideal camera would be one that has infinite resolution, however, due to sensor limitations in the real world it is not a realistic option to capture. However, in modelling ideal images, because of the infinite level detail it would carry it cannot be considered a function. This is because a function such as this would have more higher and higher frequency content as the sensors were refined. Hence, the total energy would diverge and not be in L^2 in the limit. The labeling of objects, texture and noise in an image can also not be clearly made. In any given image, depending on the scale and context, a given object may be considered as being part of textural information, a little bit of noise or back to being an object [74]. (See [102] for an explicit study of this concept).

A final, perhaps simplest, model we can consider that includes both the properties of high kurtosis and scale invariance, is one proposed by Gidas and Mumford [48], called the random wavelet model:

$$I(i,j) = \sum_{\alpha} \psi_{\alpha}(e^{r_\alpha} i - i_\alpha, e^{r_\alpha} j - j_\alpha). \tag{2.2.5}$$

Here $(r_\alpha, i_\alpha, j_\alpha)$ is a uniform Poisson process in 3-space and ψ_α are samples from the auxiliary 'Levy process', a distribution on the space of scale and position normalised elementary image constituents, which one may call mother wavelets or textons. The components ψ_α represent elementary parts of an image, which can be considered to be Gabor patches, edgelets, curvelets or more complex shapes such as ribbons [74].

2.3 Non-Gaussianity

It has become a well documented and striking empirical fact that image statistics do not follow Gaussian distributions. Instead we often see distributions that have sharp central cusps and characterised by high values of kurtosis. This empirical result is observed when any linear filter F with zero mean is applied to images, and results in the values $x = (F \cdot I)(i,j)$ of the filtered image following a distribution with kurtosis larger than 3. These sparse histograms obtained from

the statistics of natural scenes have been observed by several early researchers, such as Daugman [35] and Field [43]. Other researchers Burr and Morrone [19] related the distributions to the sparsely occurring edges in natural images which are proposed to be the information bearing parts. Within the context of neural encoding of signals Field [42] expressed ideas that heavy tails correspond to the arrangements in the Fourier phases of natural images. This fundamental property thus has the effect of producing only a sparse set of neural activity within the brain, and verifies experimental results concerning the nature of the signal encoding. On the other hand, drawing images from a Gaussian distribution results in distributions taking random phases, and these do not correspond to any structural properties exhibited by natural scenes [43, 84].

The non-Gaussian behaviour of images has continued to hold over several decades of research, and has also been studied by Ruderman [85], Simoncelli and Adelson [90], Moulin and Liu [77] and Wainwright and Simoncelli [99]. These studies often utilised small datasets of images, but continually confirmed the empirical finding. However, in a move to robustly verify the observations of non-Gaussianity in large calibrated datasets of natural images, Huang and Mumford [60] utilised a large database of images provided by J.H. van Hateren [61]. The availability of such a database allowed the precise and intensive statistical studies of the local nature of images. Huang and Mumford reported results ranging from the simplest single pixel intensity to the joint distributions of 3 Haar wavelet responses. We will review some of the results of this work as it ties in with distributions that have been proposed for modelling the behaviour of image statistics.

2.3.1 Single Pixel and Marginal Distributions

The database of J. H. van Hateren [61] is a large set of over 4000 calibrated images of size 1024x1536 pixels where the pictures were taken by digital camera. This set consists of images of cities and country taken in Holland with the images containing a diverse range of natural objects. These include mixtures of trees, rivers, stones, buildings, sky, roads and many more objects, often within a single

image, with images taken from different angles. Use of such a large amount of data is aimed at stabilising and showing the robustness of the results that are obtained.

Calibrated datasets imply that images only measure light (intensity of light) in the world up to an unknown multiplicative constant in each image. Essentially implying that the amount of light present in a scene varies from image to image due to the images being taken outdoors, dependent on the present conditions. Note that the images are pure, in the sense that no further processes are applied to the measured intensity of light, e.g. gamma correction, which would render the images artificial. We only want to investigate systematically the exact statistics that underline natural images.

Figure 9.1 shows a few examples of the raw pictures taken from the van Hateren database. The images are very dark and interestingly show what a camera is really measuring before any further processes are applied. Indeed, figure 9.2 to shows the same set of images having been auto-corrected using Microsoft Picture Manager. Here we can observe some of the typical elements found within a scene.

To obtain results on these images or databases of images that are independent of the varying amount of light, we are required to work with statistics that do not contain this multiplicative constant, now an additive constant. i.e. we work with statistics of the *log-contrast* of images such as

$$I(i,j) = ln(\phi(i,j)) - < ln(\phi) >, \tag{2.3.1}$$

where $I(i,j)$ are the calibrated gray level values at position (i,j) in an image, $\phi(i,j)$ is the camera recorded intensity value at pixel position (i,j), and the expectation $< . >$ is taken over each image separately. We note that the statistics we use will be of a similar nature to the above though expectation of image intensity may be replaced with the log intensity of another pixel. This process prevents the overall illumination present in a scene from affecting the statistics.

Figure 9.3 shows a plot of ln(histogram) of random variable $ln(\phi(i,j))$ - $< (ln(\phi)) >$. The distributions are plotted with the logarithm of the probability,

or frequency of occurrence. This is important when studying the distributions as simply plotting raw probabilities leads to distributions that look alike. However, on a log scale sparse distributions identify themselves against Gaussian distributions since $ln(e^{-x^2}) = -x^2$.

Some constants associated with the statistic in figure 9.3 are:

$$\mu = 0, \sigma = 0.79, S = 0.22, k = 4.56. \tag{2.3.2}$$

From the log plot and the skewness S we see that this statistic is not symmetric. One important possible reason is the presence of sky that features in many images from the database, which is quite different from the rest of the parts of an image, always with a high intensity value. Hence it leads to a higher proportion of pixels having relatively larger intensity values. Another interesting feature is the linear tail in the left half. Obviously this statistic is non-Gaussian, although the centre part of the log plot does show a parabola shape. The kurtosis is bigger than 3 (the value for Gaussian's) but not very large.

The single pixel statistic is however not very informative. This is because we can strongly modify the histogram of an image without affecting much of its perception. Hence, we move on to look at statistics involving nearby pixels. An important one we can study is the marginal distribution of horizontal derivatives, which in the discrete case, is simply the difference between two adjacent pixels in a row, i.e. $D = ln(\phi(i,j)) - ln(\phi(i,j+1))$. Figure 9.4 shows the $ln(Histogram)$ of D, and here are some constants associated with it:

$$\mu = 0, \sigma = 0.260, S = 0.085, k = 17.43. \tag{2.3.3}$$

The statistic is observed to be symmetric but with very high kurtosis. We also observe the σ value to be of interest as it points to an analogy that a given image has approximately 75% stability across a line with 25% contributing to the variance [60]. It has also been noted in several articles that for large databases of natural images, the derivative statistic is surprisingly stable - consistent across different databases and categories of images [64].

2.3.2 One Dimensional Long Range Correlation

Aside from studying statistics which concern single or nearby pixels, we are also interested in studying the relations between pixels at increasing distances and orientations. One of the most important long range statistics is the covariance of two pixels, and in this section we extend on the results concerning the power spectrum of natural scenes to study correlations across spatial scales.

A key result states that the power spectrum of natural scenes takes the form of a power law in the spatial frequency. This scaling result can be expressed equally well in the spatial domain in terms of the function $C(r)$ as:

$$C(r) = -C_1 + C_2|r|^{-\eta}, \tag{2.3.4}$$

where r is the separation distance between pixels, η the same constant value as it did in the power spectrum, C_1 and C_2 constants [86].

The function is found experimentally as the expected product of the image at two pixels separated by a distance r:

$$C(r) = <<<< \phi(\bar{x}_0)\phi(\bar{x}_0 + \bar{x}) >_{\theta}>_{\bar{x}_0}>_{\varphi} . \tag{2.3.5}$$

This is the 'one dimensional case' of the covariance function. Here $\phi(\bar{x}_0)$ is the image value at position $\bar{x}_0$ (a two vector), and the triple expectation value is over (from the outside inwards) all images φ, all initial positions $\bar{x}_0$ and all displacement vectors $\bar{x}$ of length r (parameterised by angle θ). This large expression can be written schematically as

$$C(r) = < \phi(0)\phi(x) >, \tag{2.3.6}$$

with the ensemble average, shift over positions and average over angles implied [86].

Our images are only well defined up to an 'additive constant', so we replace the covariance statistic by the following function,

Definition 2.3.1. (One Dimensional Difference Function) *We define the one dimensional difference function schematically to be*

$$D(r) = < |\phi(0) - \phi(x)|^2 >, \tag{2.3.7}$$

with the ensemble average, shift over positions and average over angles implied, and with distance between two points of r.

The relation between $C(r)$, the covariance, and the difference function, $D(r)$, is simply

$$D(r) + 2C(r) = constant, \tag{2.3.8}$$

this is found by expanding the squared term in the expectation value of $D(r)$ [86]. From the above relation we have

$$\begin{aligned} D(r) + 2C(r) &= \alpha \text{ (constant)} \\ D(r) + 2[-C_1 + C_2|r|^{-\eta}] &= \alpha, \\ D(r) &= \alpha + 2C_1 - 2C_2|r|^{-\eta}. \end{aligned}$$

Hence we have

$$D(r) = D_1 - D_2|r|^{-\eta}, \tag{2.3.9}$$

where D_1, D_2 and η are constants. A power law spectrum thus yields a power law difference function, except for the presence of an added constant [86].

Calculation of the difference function is obtained by choosing pairs of n (e.g. $n = 5$ million) arbitrary pixels randomly in an image, of different distances and orientations and calculating the difference function by averaging $D(r)$ for each distance x over the pairs corresponding to that distance x. In the case of the database image difference function we include in the averaging, all the 'difference pairs calculated' at the same distance but from all images. So for each distance we would have the intensity difference summed over all the images, and then divided over the number of pairs at that particular distance. After getting this difference function, $D(r)$, the scaling model 2.3.9 is fitted.

2.3.3 Two Dimensional Long Range Correlation

In this section we generalise our one dimensional difference function to two dimensions.

Definition 2.3.2. (Two Dimensional Covariance Function) *The two dimensional covariance function is defined schematically to be*

$$C(x,y) =< \phi(x,y)\phi(0,0) > . \tag{2.3.10}$$

However, our images are samples of a distribution which is only well defined up to an additive constant. Hence, as in the previous section we replace this statistic by the 'difference function'.

Definition 2.3.3. (Two Dimensional Difference Function) *We define the two dimensional difference function schematically to be*

$$D(x,y) =< |\phi(x,y) - \phi(0,0)|^2 >, \tag{2.3.11}$$

which is related to the covariance by

$$D(x,y) + 2C(x,y) = constant, \tag{2.3.12}$$

when both are well defined [60].

The explicit calculation of the two dimensional difference function is a little different to that of the one-dimensional version. The expectation here is over all images and displacements only, with (x,y) representing a vector from an arbitrary anchoring point $\phi(0,0)$. Therefore, (x,y) represents a vectorial displacement: x units (positive or negative) from arbitrary anchoring point $(0,0)$ and y units (positive or negative) from arbitrary anchoring point $(0,0)$.

Calculation of the difference function on the whole van Hateren database by taking into account all possible pixel pairs within 500 pixels, results in a mesh plot and a contour plot (see figure 9.5). The statistics obtained are very stable and we can look more closely at the tail of the statistics, and even take delicate operations like derivatives on them. A relatively smooth 3-dimensional plane is achieved [60].

Taking cross sections through the origin (when $x = 0$, and when $y = 0$) of the mesh plot we have the horizontal and vertical cross sections. Graphs in figure 9.5 correspond to evaluating the difference functions below,

$$D(x,0) =< |\phi(x,0) - \phi(0,0)|^2 >, \tag{2.3.13}$$

and

$$D(0,y) =< |\phi(0,y) - \phi(0,0)|^2 > . \tag{2.3.14}$$

A comparison between the graphs shows a notable result in that the cross section along the vertical direction grows much faster than that along the horizontal direction. i.e. the average difference intensity along the vertical of the image changes more rapidly than the horizontal. One of the reasons put forward by Mumford [60] is that this is due to the fact that in many images from the van Hateren database there is a portion of sky at the top and ground at the bottom, and the large difference between the intensities contributes more to the difference function along the vertical direction.

Next we consider the positive part of the horizontal cross section of the mesh plot in figure 9.5, i.e. along $y = 0$, with only x displacements, and where $x > 0$. i.e. we only consider intensity displacements that are positive in relation to the arbitrary anchoring points.

Shown in figure 9.6 is a $log - log$ plot of the derivative of the positive part of the horizontal cross section. We have pixel distances in the range from 1 to 500. For pixel distances of 4 to 32, on the log scale we have 2 and 5. We observe that between 2 and 5 pixel distances the derivative is close to a straight line with a slope -1.19.

If we use the model

$$D(r) = D_1 - D_2|r|^{-\eta}, \tag{2.3.15}$$

we find that $\eta = 0.19$. This is obtained by the following method as done by Ruderman [86],

$$[D(r)]' = (-2C_2) \times (-\eta)|r|^{-\eta-1}, \tag{2.3.16}$$

taking logs we find the R.H.S to be

$$log(2C_2 \times (\eta)) + (-\eta - 1)log|r|, \tag{2.3.17}$$

recalling that the derivative between 4 and 32 pixel differences is close to a straight line we compare the above equation with a straight line,

$$log(y) = -1.19log(r) + C. \tag{2.3.18}$$

Equating coefficients we obtain $-\eta - 1 = -1.19 => \eta = 0.19$.

In the work of Ruderman, the anomalous exponent η is found to take the value 0.19, where the one dimensional difference function is calculated (with random orientations) for his database of 45 images of size 256×256 pixels. In the work of Huang and Mumford a similar result is obtained for a slightly different calculation of the one dimensional difference function, namely using horizontal displacements via the cross section.

An interesting observation is that in the works of Ruderman and Huang and Mumford, both results find the anomalous dimension to be the same. This is in contrast to the diverse and different datasets that are used. These pointers also suggest that natural images are universal in some sense i.e. they exhibit properties and structures that are universal to all images. Such results are surprising, while we have observed the robustness of the anomalous exponent to large databases of a variety of image types, these results further display the remarkable property that correlations of light intensity do not seem to be related to information content in an image [86].

However, for larger distances it appears that a linear term dominates due to the $log - log$ plot beginning to turn and becoming almost a horizontal line around log distance 8. Hence we can model this better as

$$D_l(r) = D_1 - D_2|r|^{-\eta} + D_3|r|. \tag{2.3.19}$$

Thus the power-law term dominates the short-range behaviour, while the linear term dominates at large pixel distances [60].

2.3.4 Distribution of the anomalous exponents

In a separate study [46], the authors confirm the results of Huang and Mumford [60] by calculating the one dimensional difference function using randomly

chosen 5 million pairs of points in the images that are separated by distances between 0 and 32 pixels, and oriented in different directions. Figure 9.7 shows the obtained average difference function values along with a power law fitting over 1400 images from the van Hateren database. In their measurements they obtain a value of $\eta = 0.19$ when fitting the difference function. This value varies a little due to averaging over different numbers of images. In their measurements it was of the order of $\Delta\eta = +\_0.01$. (For instance, for the first 1100 images of the archive the value of $\eta = 0.18$ is obtained.)

An interesting point is that the results correspond to those of Huang and Mumford, who obtained the anomalous dimension to be 0.19 through the horizontal cross section. A puzzling remark is that taking pairs of pixels in random orientations produces the same anomalous exponent as taking pairs of pixels oriented *only* in the horizontal direction. Another very interesting comparison is that Ruderman, with his relatively much smaller database of 45 images and size 256×256 pixels, obtains the anomalous exponent to be 0.19 whilst taking 5 million pairs at all orientations for the database of images.

The next result we can consider is whether the behaviour of the database difference function is the same for each individual image difference function. Taking into account the expectation that the image difference function will be noisier due to the fact that we are averaging on much smaller sizes, we observe in figure 9.8 fittings for some images with their corresponding η values. Clearly, we can see that the exponent differs from image to image, however, the same characteristic power law behaviour is observed. Hence, we can ask whether the database exponent can be considered as a representation of the individual image exponents. Such a conclusion requires that the distribution of the image exponents should be narrow. However, as can be see from the distribution values of η over 1500 images shown in figure 9.9, two obvious features are a maximum in the vicinity of $\eta = 0$ and a rather wide distribution giving significant weight to relatively high values of η. Hence, we can conclude that the database η is not a representative of the image η's, and additionally we observe that the distribution is also non-Gaussian.

2.4 Applications of Image Statistics

The statistics of natural images have been exploited in a number of diverse fields. In the subject of biological vision it has been used to explain some parts of the perceptual systems of humans through the assumption that such systems are designed to interpret scenes of the natural world. The sparse nature of the filter responses have also been exploited to increase image compression rates by capturing only a selection of the largest coefficients of a transformed image, while discarding the high number of lower coefficients without perceptual loss. Additionally, statistical regularities have been used to build models that seek to generate natural images, with applications in texture synthesis, image modelling and classification. In this section we will briefly detail some of these, and several more instances.

2.4.1 Biological Vision

The human visual system is capable of carrying out a host of image processing activities with relative ease. For example, filling in missing regions of images, extracting meaningful information from images and videos, matching shapes and objects from only a handful of views, are all tasks which humans are adept at doing. The way in which the human brain is able to carry out such tasks is of significant interest to researchers in the fields of artificial intelligence, biological and computer vision. For example, Shannon's classical work on the exploitation of redundancy in the English language led to Barlow [9] applying similar techniques to the study of images. He was one of the initiators to propose that the human visual system utilises fundamental redundancies exhibited by natural images from the earliest stages of visual interpretation right to the higher semantic levels.

The anatomical study of visual systems of mammals have illustrated that biological systems perform processes which take into account the statistical regularities of images. These studies point to the fact that the sensing of images in the brain is not merely conducted by storing point by point intensity values of

given images. Rather, it is hypothesisesd that cortical neurons encode the information through their inherent selectivity towards features such as orientation, motion, colour and light intensity [62]. Thus the visual system is rather exploiting redundancy rather than trying to compress the data which it receives. Other works such as Field [42] and van der Schaff [98] have also supported the ideas that cortical neurons are tailored towards redundancy minimisation schemes. Field, for example, related the distributions of spatial frequencies obtained from natural images to the way in which cortical neurons activate. Interestingly, the patterns of excitation point towards redundancy exploitation being a memory mechanism for the information it is tasked to deal with.

Others have also applied the fundamental ideas of redundancy minimisation to visual systems. Linsker [68], for example, took concepts from information theory to further understand properties of the visual system, while Atick [38] sought to uphold the idea that living organisms had visual systems designed to handle the inherent structure's found within natural images. Properties of images have also been further analysed to extract meaningful information, patterns and regularities by Tolhurst et al. [95], Ruderman [85], Field [43] and Baddeley [8].

Recent works on the studies of natural images have tried to utilise properties held by cortical neurons in trying to design processes that give the required outputs which are obtained by physiological studies. Interesting works by Bell and Sejnowski [11], Olshausen and Field [80], van Hateren and van der Schaff [61] took ideas obtained from separation of a mixture of signals into their individual constituent components, and applied them to studies of images. They performed Independent Component Analysis on databases of patches of natural images to derive the constituent basis patches whose linear combinations produce the actual image patches. Their predictive results showed how histograms of spatial frequency bandwidth, orientation tuning bandwidth, aspect ratio and length of the receptive fields matched experimental results well [61]. Hence, the works point to cortical neurons being selective to the particular types of structures and patterns of light found within natural images.

Researchers have made significant progress in understanding biological visual processes through the understanding of the images on which such systems are trained upon. However, although much work has been done in this area, our understanding is limited. With all the power of scientific innovation we are still far from a full understanding of how the human brain is able to perform vision orientated tasks e.g. tracking, recognition and matching with such apparent ease.

2.4.2 Modelling Filter Response Distributions and Applications

In the usual case of filtering natural images using zero mean filters, the resulting distributions display high kurtosis and sharp central cusps. In the case of images where the clutter is less and the filter is matched to typical image features (like edges) the peak is much more pronounced, while if the the clutter is greater or the filter has no geometric significance the peak is less pronounced. In general the responses all form distributions that have high values of kurtosis.

Two explicit models have been proposed for modelling the marginal distributions of natural images e.g. figure 9.4. One being the generalised Gaussian distribution (GGD) proposed by Wainwright and Simoncelli [100] and the other being Bessel-K forms [51] developed by Grenander et al. In the former case the generic prior takes the form

$$C \cdot e^{-|x|^{\alpha}}. \tag{2.4.1}$$

Here α is in the range $(0.5, 1.0)$ and is called the shape parameter which can vary from image to image and between classes of images. C is a normalizing constant [74].

The alternative model, Bessel-K forms, models images through taking into account of the image formation process itself. Bessel representations explain this phenomena via a fundamental hypothesis: that images are made up of objects. This can be illustrated by using the following example from [91]: For an image I let us study the histogram of differences in the values of the (horizontally)

neighboring pixels. If two neighboring pixels have equal or similar intensities then their difference is small and it adds to the histogram bin containing zero (which also happens to be the median value). On the other hand, if the two neighboring values are quite different, as is the case for vertical edges, then the difference is high and it adds to the histogram at the tails. In general, regions with horizontal homogeneity add to the the central peak and the sharp, distinct edges map to the tails. It implies that for images with large objects with smooth, homogeneous foregrounds and sharp, distinct edges, the difference-histograms will have a sharp central peak and decaying tails. On the other hand, images with lots of blurred objects in the scene will have difference-histograms that are close to Gaussian (through central limit theorem). Object-based models for images allow for both the homogeneous regions and the edges in an image and hence can explain the observed non-Gaussianity.

The object based approach results in the following model given by Bessel functions:

$$C \cdot e^{-|x|^{\alpha}} \cdot K_s(|x|), \tag{2.4.2}$$

where K_s is the 'modified Bessel function' and $0 \leq s \leq 0.5$. Setting aside the theory of Bessel functions, this distribution is best understood as the Fourier transform of $1/(1+\xi^2)^{s+0.5}$. Although both these distributions seem alike, the GGD has larger tails when $\alpha < 1$ while the Bessel ones are all asymptotically like $e^{-|x|}$. In the case of deciding which model really fits the data well there is difficulty. This is because both models differ mostly in the tails where additionally the data is most noisy [74]. However, the two imaging models have been utilised in a host of image processing applications and analysis. Next we present a selection of successful applications that have made use of such knowledge and other related statistics of natural images.

Texture Synthesis

Texture synthesis is the process of accurately replicating a given patch of image that displays regular patterns. Following the review of [91] we observe that

newer statistical models have revolutionised this area by utilising the marginals of filtered images. For example, Faugeras et al. [41] used this approach for representing textures, while Bergen and Landy [13], Chubb et al. [33], and Heeger and Bergen [52] also advocated the use of histograms. In [106] Zhu et al. illustrated the sufficient characterisation of homogeneous textures by using a collection of filters and their respective marginal distributions. This approach of using the frequencies of occurrence suggests that information from histograms is sufficient while location information can be discarded. The representation of an image can thus be accomplished by extracting features using wavelet decompositions at several scales and orientations, and exploiting the periodicity of such patches. Other attempts that have utilised the statistics of natural images have involved the use of joint statistics of filter responses where the measured correlations of raw coefficients, as well as their magnitudes, have been used to develop an efficient algorithm for synthesizing random images [81].

Image Denoising and Deblurring

A traditional procedure in image denoising is to perform transformations, such as the wavelet transform, upon an image that allows the approximate sparse representation of an image. One can then perform soft or hard thresholding in order to shrink the desired coefficients towards zero with the resulting inverse transformation giving a more visually pleasing result [16, 40]. In the statistical approach Bayesian or (Maximum a-priori) MAP techniques are used. Here Simoncelli and Adelson [90] utilised prior models on the filter response of images within a Baysian framework to successfully advance the state of image denoising algorithms. Additionally Moulin and Liu [77] were able to incorporate the generalised Gaussian distribution model into their denoising schemes.

Many pyramid based approaches to image denoising have been developed due to the fact that the marginal statistics are easier to characterise than that of whole images. In such schemes images are decomposed into multi-scale representations and statistics of coefficients are used for denoising [81, 67] in individual frequency bands [91].

In other works of image recovery the sparse distributions have been successfully utilised as prior information within a Bayesian framework for deblurring [66]. Results here exceed classical approaches such as the Richardson-Lucy algorithm and the algorithms that assume Gaussian priors for deblurring gray images. This technique of using sparse priors has also been successfully used for the simultaneous deblurring and denoising of colour images [56].

Super-resolution

Filters exhibiting localised, oriented and bandpass characteristics, such as the Gabor filters, applied to natural images result in the classical sharply peaked distributions with high kurtosis. This phenomenon of regular sparse distributions across a wide category of imagery motivated their use within an image super-resolution scheme that incorporates the prior knowledge into graphical models to infer latent information. Specifically, factor graphs were used and derivative filters exploited to estimate a high resolution image from a single low resolution image. The resulting high resolution images gained good image quality with sharp edges and low reconstruction error. Techniques such as these have also been successfully exploited in the closely related problem of image demosaicing [94].

Image Classification

Problems in image understanding require us to develop models that are able to perform vision oriented tasks based on contextual and physical information. For example, in the problems of facial recognition the schemes are required to extract features of interest and match against representations within a database collection. These problems can be tackled from two perspectives, lower and higher level vision. The former case requires the building and analysis of individual constituents of an object, such as points, edges and shapes that can integrate together to create objects of interest. The latter approach see's one starting from know shapes and characteristics of objects and performing matching while taking into account the various nuisance variables such as occlusion, points of view

and clutter objects. This latter approach which relies upon physical principles has been successfully used for classical problems in image processing through the use of prior knowledge, such as facial recognition, pose estimation and object tracking. Naturally, such a scheme is inherently restricted by the number of known objects, or the often called targets of interest (TOI).

The issue of TOI appearing in an image together with sets of objects that add to the difficulty in performing tasks pose significant challenges for any image processing approach or algorithm. These variables, often termed nuisance variables or clutter objects, can appear in all parts and contexts of an image, and any model that wishes to gracefully handle such observances needs to be able model the characteristics of the TOI and the types of clutter they can often be associated with.

In the case of modelling clutter types Bessel K forms have been successfully utilised for establishing the types of clutter found in the problems of automated target recognition (ATR). In such situations, e.g. recognising a tank from its remotely sensed image, one requires automated recognition of the TOI amongst clutter objects. Knowledge of the clutter types, e.g. grass, sand, buildings, etc. can significantly aid algorithmic performance when searching. The procedure consists of utilising a host of filters, each tuned for selectivity and whose responses are then modelled using the Bessel-K forms. These distributions are parameterised, making pairwise comparisons and hence classifying the clutter types from their spectral images against known types of clutter efficient [51].

Pruning for Hypotheses Testing

This application deals with the problem of pruning a database of possible candidates that may match the features of a given object. Such pre-processing can significantly aid the speed at which matching can be performed. In the case of recognising people from their frontal infrared images of the faces, a natural way to proceed is to develop $3D$ polygonal meshes of the faces along with thermal values at each point. The objective is then to search over these physical representations for given test images. Recognition of objects from their observed

images corresponds to the selection hypothesis (mesh template) in the presence of nuisance variables. The hypothesis selection is performed using detailed models involving physical geometries, thermal variables, pose, and motion [92].

Pruning the possible candidates places significant probability only on a small subset of the possible targets and Bessel representations can be utilised in such situations. We follow the procedure of [92] outlined in the following: Let A be the set of objects in an image that we are interested in, and let S be the group of nuisance variables such as pose, translation and thermal variables. Define a probability mass function on A according to, for an $\alpha \in A$

$$\mathbb{P}(\alpha|I) = \frac{1}{Z}\exp(-min_{s\in S}\sum_{j=1}^{J} d(p_{abs}^{(j)}, c_{abs}^{(j)}, p_{\alpha,s}^{(j)}, c_{\alpha,s}^{(j)})^2/D), \tag{2.4.3}$$

where Z is a normaliser, d a suitably defined metric (or pseudo-metric) and D controls our confidence (analogous to the temperature in simulated annealing) in this probability. Here $(p_{abs}^{(j)}, c_{abs}^{(j)})$ are the estimated Bessel parameters for the filtered image $I^{(j)}$, and $(p_{\alpha,s}^{(j)}, c_{\alpha,s}^{(j)})$ are the estimated Bessel parameters for the filter $F^{(j)}$ and the target α rendered at the nuisance variable $s \in S$. Note that $(p_{abs}^{(j)}, c_{abs}^{(j)})$ can be precomputed offline for all $\alpha \in A$, $j \in \{1, 2, ..., J\}$, and a finite subset of S. With the goal of pruning for hypothesis selection in mind, on a database of images all objects with $P(\alpha|I)$ greater than some threshold can be short listed as candidates for classification.

Several other techniques have also been widely utilised for purposes of dimension reduction. For example, principle component analysis (PCA) and independent component analysis (ICA) have been widely used to obtain lower dimensional approximations of images. These approaches incorporate higher order statistical structures and decompose a given image into their respective constituents and discard ones that do not lead to significant loss of detail. Experiments in [92] indicate that Bessel K forms result in the best recognition rate among the three methods under general test conditions, and are typically less computationally expensive to implement.

Image Compression

The concepts of redundancy minimisation and human perception have important consequences with regards to image compression. Two popular image compression standards, JPEG and JPEG 2000, specifically utilise techniques that exploit these particularities. In the former case due to the densities of colour and brightness sensitive receptors in the human eye, humans can see considerably more fine detail in the brightness of an image than in the hue and colour saturation of an image. Using this knowledge, encoders can be designed to compress images more efficiently by for example converting an RGB image to a luminance-chrominance space such as YUV and down sampling the chroma channels. Redundancy exploitation is also highlighted by the usage and standardisation of JPEG 2000 as an image compression scheme. This format shows how once an image has been transformed into the wavelet domain, there are statistical properties that can be obtained from the corresponding coefficients. Most important is the finding that the distributions of these coefficients are highly non-Gaussian. This can be best exploited when using wavelets which produce few high amplitude coefficients and many small amplitude coefficients [72], thus allowing one to keep only the largest coefficients. Compression at high rates of orders of magnitude can be achieved with minimal perceptual loss.

Chapter 3

Compressive Sensing

Traditional methods of capturing signals or images follow the Shannon/Nyquist theorem, that to avoid losing information when capturing a signal, the sampling rate must be at least twice the maximum frequency present in the signal. This principle governs nearly all signal acquisition protocols, and indeed the Nyquist rate can be so high that too many samples result, making compression a necessary prior to storage or transmission [10]. There is already an extensive body of literature concerning data compression. In the context of imaging, the processes underlying the schemes rely upon transformations into an appropriate basis and then encoding only the important expansion coefficients. Two of the most notable representations of images have utilised sinusoids and wavelets, these transforms underly the classical JPEG and modern JPEG 2000 compression standards, respectively.

However, developments concerning the capture and representation of compressible signals at a rate significantly lower than the Nyquist rate have recently been emerging. This form of sampling, called *compressive sensing* [25], [83], [21], seeks to acquire a compressed signal representation without going through the intermediate stage of acquiring *all* samples. Here one exploits the idea that signals of interest exhibit redundancy (and are hence compressible) and can be expressed sparsely in some appropriate basis. This protocol aims to overcome the usual sampling method of acquiring large amounts of data and then per-

forming compression, by directly arriving at a compressed representation of the signal. In the following we give a concise treatment of this new sensing modality and show some surprising consequences of the theory.

3.1 Image Sampling and Sparse Representation

To settle the notation and generalise the notion of sampling an image, consider a real-valued finite dimensional discrete-time signal X. Here X can be thought of as an $N \times 1$ column vector in $\mathbb{R}^N$ with elements $X(n)$, $n = 1, 2, ..., N$. This signal structure equally applies to images where we would simply rasterise the image, row by row, to obtain one column vector. In the usual imaging setting one may collect points of the image at unique spatial locations or obtain averages over spatial areas such as pixels. However, we can generalise to consider a set of measurements y_k obtained by taking inner products between X and a different function ϕ_k:

$$y_1 = \langle X, \phi_1 \rangle, y_2 = \langle X, \phi_2 \rangle, ..., y_m = \langle X, \phi_m \rangle. \tag{3.1.1}$$

Here the notion of the test functions dictate the type of measurements we are acquiring. If we take ϕ_k to be sinusoids at different frequencies we are implying that we are measuring Fourier coefficients. If on the other hand they are taken to be delta ridges, then we are obtaining line integrals while in the case of a standard digital camera we would take the functions to be indicators on squares [83].

Next we consider the representation of signals of interest: Any signal in $\mathbb{R}^N$ can be represented in terms of a basis of $N \times 1$ vectors $\{\psi_i\}_{i=1}^N$. For simplicity we take the basis to be orthonormal and using $N \times N$ basis matrices $\Psi = [\psi_1 \psi_2 .... \psi_N]$ with the vectors $\{\psi_i\}$ as columns, a signal X can be expressed as follows:

$$X = \sum_{i=1}^{N} S_i \psi_i \quad or \quad X = \Psi S. \tag{3.1.2}$$

Here S is the $N \times 1$ column vector of weighting coefficients $S_i = \langle X, \psi_i \rangle = \psi_i^T X$, and ^{T} denotes transposition. In an appropriate basis Ψ we note that many signals of interest, such as natural images, can be expressed using few coefficients S_i where the S_i occur sparsely or the sorted magnitudes display quickening decay. Precisely, we can say that a signal is K sparse if it is a linear combination of only K non-zero S_i coefficients in 3.1.2 and $(N - K)$ are zero. In order for compressive sensing to be used as a practical tool, we are concerned with the scenarios where $K << N$.

A simple example concerning the wavelet basis can illustrate the ideas more clearly. Consider a digital image composed of pixel values which are almost all non-zero. However, expressing an image X in the wavelet basis leads to coefficients that are mostly close to zero with only a relatively few large coefficients that capture the essential information. This is a typical real life situation where the signals of interest, when expressed in an appropriate basis are not exactly sparse, but only *approximately* sparse. Hence, one can now see that when such a representation exists in an alternative basis one can compactly represent the original signal by discarding much of the smaller coefficients while keeping the fewer larger ones without significant perceptual loss [21].

This process of transform coding also displays some obvious drawbacks where one can see that the sampling has required taking a large number of N samples, even though the desired number of samples K is a much smaller set. Furthermore, such a process also requires that S_i coefficients be computed even though only K are to be kept, while the larger coefficients are also required to be adaptively encoded [10]. However, we will begin to see in the text how the notion of sparsity has deep implications for dealing with the acquisition process itself. Sparsity determines how efficiently we can acquire signals using a predetermined set of linear measurements *non-adaptively*, and which are orders of magnitude smaller than the total size of the signal.

3.2 Incoherent Sampling

Suppose we are given a pair of orthobases (Φ, Ψ) of $\mathbb{R}^N$ where the basis Φ is used for sensing the image X while Ψ is used to represent X. Then we have the following definition:

Definition 3.2.1. *The coherence between the sensing basis Φ and the representation Ψ is*

$$\mu(\Phi, \Psi) = \sqrt{N} \cdot max_{1 \leq k,j \leq N} \|\langle \phi_k, \psi_j \rangle\|. \tag{3.2.1}$$

In other words, the coherence measures the largest correlation between any two elements of Φ and Ψ. It follows from linear algebra that $\mu(\Phi, \Psi) \in [1, \sqrt{N}]$. If Φ and Ψ contain correlated elements, the coherence is large, otherwise it is small [21].

Compressive sensing is concerned with low coherent pairs of $\{\Phi, \Psi\}$. As an example consider the pairs of bases, Φ in which $\phi_k(y) = \delta(t-k)$ (spike basis) and Ψ in which $\psi_j(t) = n^{-1/2} e^{i2\pi jt/n}$ (Fourier basis). Since Φ is the sensing matrix this corresponds to the classical sampling scheme in time or space. The time-frequency pairs obey $\mu(\Phi, \Psi) = 1$ and therefore we have maximal incoherence. Furthermore, spikes and sinusoids are maximally incoherent in any dimensions, e.g. two, three, etc [21].

One part of the central problem in compressed sensing is the design of stable measurement matrices Φ that allow the reconstruction of the length-N signal X from $m < N$ measurements. This problem appears ill-conditioned. However, if X is K-sparse and the K locations of the nonzero coefficients in S are known, then the problem can be solved provided $m \geq K$. This brings us to the next important definition that has proved useful in the study of the general robustness of compressive sensing: *the restricted isometry property* (RIP) [23].

Definition 3.2.2. *For each integer $K = 1, 2, ...$ define the isometry constant δ_K of a matrix A as the smallest number such that*

$$(1 - \delta_K)||X||_{l_2}^2 \leq ||AX||_{l2}^2 \leq (1 + \delta_K)||X||_{l2}^2 \tag{3.2.2}$$

holds for all K-sparse vectors X.

We will loosely state that a matrix A obeys the RIP of order K if δ_K is not too close to one. When this property holds, A approximately preserves the Euclidean length of K sparse signals, which in turn implies that K sparse vectors cannot be in the null space of A. This is needed as otherwise there would be no hope of reconstructing these vectors. In order to view the connection between RIP and compressive sensing, consider acquiring K-sparse signals with A. Taking δ_{2K} to be sufficiently smaller than one implies that all pairwise distances between K-sparse signals must be well preserved in the measurement domain. Hence we have,

$$(1-\delta_{2K})||X_1-X_2||_{l_2}^2 \leq ||AX_1-AX_2||_{l2}^2 \leq (1+\delta_{2K})||X_1-X_2||_{l2}^2, \quad (3.2.3)$$

holds for all K-sparse vectors X_1 and X_2. This preservation of distances enables the existence of efficient and robust algorithms for discriminating K-sparse signals based on their compressive measurements [21].

Surprisingly, the construction of a measurement matrix Φ such that it satisfies the related properties of incoherency and RIP with high probability is to select *random matrices* [25]. As an example, let ϕ_i be independent and identically distributed (iid) random variables from a Gaussian distribution with mean zero and variance $1/N$. In this case our sampling procedure simply involves taking random linear combinations of the elements of X. This Gaussian matrix is observed to have two interesting properties [10]:

1. The matrix Φ is incoherent with the basis $\Psi = I$ of delta spikes with high probability. More specifically an $m \times N$ iid Gaussian matrix can be shown to have the RIP with high probability if $m \geq cKlog(N/K)$, with c a small constant.

2. The matrix Φ is universal in the sense that $\Theta = \Phi\Psi$ will be iid Gaussian and thus have the RIP with high probability regardless of the choice of orthonormal basis Ψ.

3.3 Sparse Signal Recovery

Having measured an image using a random matrix to form $y = \Phi X = \Phi\Psi S = \Theta S$, we now need to design a reconstruction algorithm that takes as input m linear measurements, the random matrix Φ and the basis Ψ. Since we are sampling a K-sparse vector S we can invert the measurement process and solve an optimisation problem:

$$min_{S'}\{i : S'(i) \neq 0\} \quad \text{subject to} \quad \Theta S' = y. \tag{3.3.1}$$

The objective term is a measure of the number of nonzero elements in the candidate solution S', and is often called the L^0 norm in the literature. This optimisation scheme can recover a K-sparse signal exactly with high probability from only $m = K + 1$ iid Gaussian measurements. However, while easily stated, this problem is both numerically unstable and a NP-hard problem, requiring an exhaustive enumeration of all (N, K) possible locations of the nonzero entries in K [23].

Alternatively, we could propose to minimise other norms of the reconstruction error. Defining the L^p norm of the vector S as $(||S||_p)^p = \sum_{i=1}^{N} |S_i|^p$ we could propose to find the vector in the translated null space with the smallest L^2 norm (energy) by solving

$$argmin_{S'}||S'||_2 \quad \text{subject to} \quad \Theta S' = y. \tag{3.3.2}$$

This optimisation has the convenient closed form solution $S' = \Theta^T(\Theta\Theta^T)^{-1}y$, but L^2 minimisation will unfortunately almost never find a K-sparse solution, returning instead a non-sparse S' with many non-zero elements [10].

However, surprisingly optimisation based on the L^1 norm can exactly recover K-sparse signals and closely approximate compressible signals with high probability using only $m \geq cKlog(N/K)$ iid Gaussian measurements:

$$argmin_{S'}||S'||_1 \quad \text{subject to} \quad \Theta S' = y. \tag{3.3.3}$$

Although we appear to have only substituted the sum of magnitudes in

place of size of support, this has yielded a convex optimisation problem that can be expressed as a linear program and solved using a variety of modern techniques. However, though they are not solvable as effectively as quadratic minimisation schemes there have been recent waves of activity trying to solve such optimisation programs as efficiently as possible. This still remains an area of active research that is continually propelling new algorithms.

One of the most important facets of this reconstruction scheme is that it enables us to apply theoretical ideas of sparse recovery to real world problems. Here we are able to sense in already compressed format, a signal, using only linear combinations of its elements. Furthermore, the measurement process is *universal* in the sense that we do not even have to know in what basis the signal is sparse in. This knowledge is postponed to the recovery phase when trying to reconstruct the original signal from its compressed representation. Hence in future if we are able to find a new and more useful basis representation of a signal, we can simply use this knowledge for reconstruction [83].

3.4 Geometry of L^1 optimisation

The geometry of compressive sensing in $\mathbb{R}^N$ helps *intuitively* to visualise why L^2 reconstruction fails to find the sparse solution identified by L^1 optimisation. This is illustrated by the sketches in figure 9.10 which show the case in $\mathbb{R}^2$. Figure 9.10(a) shows the equidistant measurement ball of L^1 with radius $\mathbf{r}$. Here any points on the edge of the ball are observed to have the same 'distance' from the origin. Similarly Figure 9.10(c) shows the metric ball distance created by the L^2 norm. Note that the L^1 ball is 'pointy' along the axis (anisotropic) compared to the L^2 ball which is spherical and hence isotropic.

To show how the L^1 algorithm finds the sparsest solution compared to L^2, observe the figure in 9.10(b). Here we pictorially show the recovery program in $\mathbb{R}^2$ where the point labeled α_0 is a sparse vector (only one of its components are nonzero) of which we make one measurement. The line H is the set of all α such that $\Phi\alpha = y$. Now, we wish to obtain the sparsest solution which by

definition lies on a point on the axes. Thus the point with minimal L^1 norm is found to be the intersecting point α_0 due to the anisotropy of the L^1 ball, and the flatness of the space H-precisely where sparse vectors are located.

On the other hand the point of minimal L^2 norm is found to be the intersecting point α_{l2}^* which is not generally a sparse solution. In higher dimensions this difference in solution becomes more pronounced and illustrates that two very seemingly similar measures of distance, sum of magnitudes and sum of magnitudes squared, are indeed quite different metrics.

Finally, we want to outline some reasoning (which is concreted in [23], [39] and [96]) to show that if α_0 is sparse, then for all α' with $||\alpha'||_1 \leq ||\alpha_0||_1$ we have $\Phi\alpha' \neq \Phi\alpha_0$. From figure 9.10 we can see that the program 3.3.3 will recover α_0 if the line H does not cut through the L^1 ball at α_0. Another way to say this is that for every h in the cone of descent from the facet of the L^1 ball on which α_0 lives (meaning $||\alpha_0 + h||_1 \leq ||\alpha_0||_1$ we will have $\Phi h \neq 0$. The key here is that all descent vectors h are concentrated on the same (relatively small) set as α_0, with the pointedness of the L^1 ball at the low-dimensional facet on which α_0 lies severely constraining how descent vectors can behave [83].

3.5 Compressive Sensing in Practice

In the practical case we are not interested in reconstructing signals that are themselves exactly sparse, but signals which are approximately sparse when transformed into some known basis. This transformation is accomplished by the following simple procedure. Given the random measurements $y = \Phi X$, instead of solving 3.3.3 we solve

$$argmin_{X'}||\Psi^* X'||_1 \quad \text{subject to} \quad \Phi X' = y. \tag{3.5.1}$$

In words we are searching for the length N signal with the sparsest transform that explains the measurements we have already observed. Here Ψ^* represents an orthonormal transform in which we expect our signals of interest to be compressible or sparse. In the case of images this may correspond to the DCT,

DFT, wavelet transform or replacement of the objective by a Total-Variation function.

In the practical case we also never observe completely pure signals. Instead, signals are always corrupted by some degree of noise and it is important that the compressive sensing scheme is able to be robustly handle such signals. For application in the real world we must, at the very least, be able to handle noise in the data such that small perturbations only lead to small perturbations in the reconstruction. Together with the imperfect sparseness of practical signals there is a natural concern that the recovery algorithm would be unstable due to its nonlinearity. However, the L^1 optimisation program is able to accurately recover signals that are only approximately sparse [23], [24]. Furthermore, when the measurements y are perturbed there are various ways to *relax* [23], [26], [96] the program 3.5.1 so that the recovery error is on the same order as the measurement error. These facts enable compressive sensing to be used as a practical and robust sensing scheme that is able to effectively handle noisy signals and those which are also not necessarily sparse in their original forms [83].

3.6 Extensions of Measurement Matrices

In the preceding work we have shown how measurement matrices constructed from e.g. Gaussian distributed entries can be used to compressively sense images. However, these matrices are extremely dense and even for modestly sized images of size 256×256, the resulting matrices can become far too large to solve or store explicitly. Indeed, choosing $m = 5481$ (approximately 8.5% of 65536) would make the measurement matrix itself requiring almost 3 gigabytes of memory if stored in double precision. Thus in practice using these dense matrices directly is not computationally feasible when reconstructing high dimensional signals such as images by convex programming [27].

Instead of creating the measurement matrix explicitly, we can provide *function handles* that take a vector X, and return AX. However, in view of work

in later chapters we can also consider the alternative use of matrices that are themselves sparse. As an example, we can create sparse binary $\{0, 1\}$ random (SBRM) matrices as measurement matrices. An $m \times N$ matrix A would be generated by assigning $d << m$ random ones to each column of A, and leaving all the other entries at zero. This allows storage of such matrices as *sparse* identities, hence the drastically reduced storage requirements. In [12] it is also shown that SBRM can be used effectively as measurement matrices and satisfy a weaker '$RIP - 1$' property, but nonetheless are essentially as 'good' as the dense ones. At the same time, they provide additional benefits, such as reduced encoding and decoding time. This is due to the fact that a typical LP is solved using the interior-point method, which repeatedly performs the matrix-vector multiplication.

Chapter 4

Non-Linear Filter Response Distributions of Natural Images

Statistical analysis of natural luminance images have revealed an interesting property: non-Gaussian behaviour of image statistics, i.e. high kurtosis, and sharp central cusps (see e.g. [60], [72], [105], [99], [79], [43]). This property has been extensively studied via the empirical distributions on large databases of natural images, establishing image statistics, under common representations such as wavelets or subspace bases (PCA, ICA, Fishers etc.), as non-Gaussian. For example, a popular mechanism for decomposing natural images locally, in space and frequency, using wavelet transforms leads to coefficients that are quite non-Gaussian with the histograms displaying rapidly decaying tails and sharp cusps at the median [91]. In this chapter we present results showing that this striking phenomenon readily follows across to natural colour images [6].

4.1 Filter Construction

Given an RGB image we convert it to the colour space YUV. (The chromacity images U and V are similar and so we only explain our workings for the U component, where analysis of the V component is obtained by substitution.) Our filter takes as input the chromacity channel U and the intensity image Y, the proposed filter is given below,

$$F(U)(\mathbf{r}) = U(\mathbf{r}) - \sum_{\mathbf{s}\in N(\mathbf{r})} w(Y)_{\mathbf{rs}} U(\mathbf{s}). \tag{4.1.1}$$

Here $\mathbf{r}$ represents a two dimensional point, $N(\mathbf{r})$ a neighborhood (e.g. 3x3 window) of points around $\mathbf{r}$, and $w(Y)_{\mathbf{rs}}$ a weighting function.

For our purpose we define two weights:

$$w(Y)_{\mathbf{rs}} \propto e^{-(Y(\mathbf{r})-Y(\mathbf{s}))^2/2{\sigma_{\mathbf{r}}}^2}, \tag{4.1.2}$$

and

$$w(Y)_{\mathbf{rs}} \propto 1 + \frac{1}{\sigma_{\mathbf{r}}^2}(Y(\mathbf{r}) - \mu_{\mathbf{r}})(Y(\mathbf{s}) - \mu_{\mathbf{r}}), \tag{4.1.3}$$

where $\mu_{\mathbf{r}}$ and $\sigma_{\mathbf{r}}^2$ are the weighted mean and variance of the intensities in a window around $\mathbf{r}$.

The proposed filter thus takes a point $\mathbf{r}$ in U and subtracts a weighted average of chromacity values in the neighborhood of $\mathbf{r}$. The $w(Y)_{\mathbf{rs}}$ is a weighting function that sums to one over s, large when $Y(\mathbf{r})$ is similar to $Y(\mathbf{s})$, and small when the two intensities are different. The filters are compatible with the hypothesis that the essential geometric contents of an image are contained in its level lines (see [29] for more details). These types of filters arose from the colorization problem by Levin et al. [63], where the authors wanted to automatically colorize a gray image through user interaction (see section 5). Their algorithm was developed under the assumption that areas of similar luminance should have similar colours and resulted in convincing colour extrapolation.

4.2 Filter response distributions

Figure 9.11 shows a sample of 8 natural colour images from our dataset of 25 images which are all bitmap uncompressed, captured using a Canon digital SLR camera, and were chosen to cover a wide spectrum of natural scenes in order to give some measure of robustness to our findings. We did not pay too much attention to the methods of capture, or any subsequent re-calibration as we wish to work with colour images captured via any mode, and believe that when images are considered to be natural this will have little effect on the general properties of the filter response.

Applying the non-linear filter on each of the colour channels, U and V, in the image outputs an intensity matrix on which we compute a histogram. We note that application of the filter is only possible within a boundary of the original image, dependent on the size of the neighborhoods used in the filter construction. In our case the filter was not computed on a one pixel boundary of the image. However, for applications the filter response of pixels on boundary points can be taken by using its neighboring pixels only.

Outputs of the filter using weighting function (4.1.2) on both the colour channels for two of our sample images, 'balloons' and 'objects', are shown in Figure 9.12 as gray-scale intensity images. These have additionally been normalised to a $[0, 1]$ range for viewing. The histograms of the single pixel intensity values of these response images is shown by the dots in figures 9.13 and 9.14. With the vertical axis on a log scale we clearly observe a distribution that is non-Gaussian and exhibiting high kurtosis.

4.3 Modeling sparse distributions

Considering filter responses of natural images, two models have been proposed for modelling their sparse behaviour. One is the Bessel distributions [51] which are derived analytically and have parametric forms that match the non-Gaussianity of image statistics. The derivation stems from the assumption of a fundamental hypothesis that images are made up of smooth objects that have intensity jumps

at edges. Another, more commonly utilised model is the 'Generalised Gaussian Distribution' (GGD). Experimental evidence elucidating which model best fits and describes the non-Gaussian distributions have not been satisfactorily made. The difficulty lies in the fact that both models differ most in the tails where the data is also most noisy.

However, for our purposes and especially due to its simplicity, we utilise the latter model:

$$J_\alpha(x) = \frac{1}{Z} e^{-|x/s|^\alpha}, \tag{4.3.1}$$

where Z is a normalising constant so that the integral of $J_\alpha(x)$ is 1, s the scale parameter and α the shape parameter. The GGD gives a Gaussian or Laplacian distribution when $\alpha = 2$ or 1, respectively. When $\alpha < 1$ we have a distribution which we call sparse.

The scale parameter and shape parameters are directly related to the variance and kurtosis by:

$$\sigma^2 \approx \frac{s^2\Gamma(\frac{3}{\alpha})}{\Gamma^2(\frac{1}{\alpha})} \quad \text{and} \quad k \approx \frac{\Gamma(\frac{1}{\alpha})\Gamma(\frac{5}{\alpha})}{\Gamma^2(\frac{3}{\alpha})}. \tag{4.3.2}$$

Figures 9.13 and 9.14 show the histograms of the filter response (using weighting function (4.1.2)) on each chromacity channel, U,V, for two of our sample images, with the GGD fitting overlaid. The responses are typically concentrated around zero and highly non-Gaussian, exhibiting large kurtosis and rapidly decaying tails, as compared with the normal distribution (see figure 9.15).

Table 9.1 shows the associated parameters for each filtered image using the first weighting function (4.1.2). We observe that kurtosis is greater than that of the normal distribution for all the images considered. α is seen to lie within the range $[0, 1]$, with the parameter varying from image to image, the only exception being the distribution of the U-filtered response for image: 'indoors'. This was the only component of an image to show $\alpha > 1$ in our diverse dataset of images. Generally, responses also exhibit some degree of skewness and have very low variance. Table 9.2 shows the statistics obtained by filtering the same set of 8

images, but using the second weighting function. Results are similar and show again that the filter response is highly non-Gaussian.

We next discuss the observed GGD distribution in relation to the filter response and structure. This is important for understanding the particularities of the distribution and for applications to image processing problems. An example illustrates the relationship clearly. In general, the filter response of a pixel chosen centrally in the 3×3 window of regions with colour homogeneity add to the central peak, and regions where colour differs from the central pixel map to the tails. The deviance from the median is large where colour contrasts together with luminance, as is usually assumed to be the case for natural images, and greatest where there is colour contrast but homogeneous intensity. The implication here is that images with smooth homogeneous objects of colour will have a sharp central peak and display high kurtosis. On the other hand, images with lots of objects in the scene with their variety of changing colours will tend to have histograms that are less peaked and give more probability in the tails. The GGD is able use its parameters obtained from histograms of the filter response to adaptively model these changes in natural images.

We also note here that the JPEG standard of image compression and storage is common place and hence we wanted to see how the filter holds under this form of compression. In order to do this we converted samples of the bitmap images from our dataset to the jpeg standard and filtered the images. Additionally, we used several standard test images in a variety of file forms from the image processing literature for further evaluation of the filter responses. Results were again the same: non-Gaussian, high kurtosis distributions of the filter response on both the chromacity channels of natural images.

Chapter 5

Colorization of Natural Images via L^1 Optimisation

Colorization of natural images has been a long standing problem in image processing. The initial process was invented by Wilson Markle and Brian Hunt and first used in 1970 to add colour to monochrome footage of the moon from the Apollo mission [18]. Early processes required an artist to begin by segmenting regions in an image and then choosing colours to fill these regions. This resulted in the time consuming tasks of image segmentation as the algorithms for this process were difficult to implement on natural images that typically have complex and fuzzy region boundaries. The techniques for colorization often gave images that were pale, flat and with washed out colour.

However, research in this area has been continuing over the past few decades and has resulted in significant advancements in the technology. Contemporary techniques used by the industrial sector are little known about in the public domain, but are still thought to rely on defining regions and tracking them between frames [89]. BlackMagic is an example of a commercial software for colorizing still images [78]. It provides the user with useful brushes and colour palettes, however the segmentation task is still left entirely to the user, and thus the desired automation of the process is still lacking. In response to this

problem Welsh et al. [101] described a semi-automatic technique for colorizing a grayscale image by transferring colour from a reference colour image. This resulted in the colour choosing aspect of the problem to instead be directed to finding a suitable image for colour transfer. Their approach examines the luminance values in the neighborhood of each pixel in the target image and transfers the color from pixels with matching neighborhoods in the reference image. This technique works well on images where differently colored regions give rise to distinct luminance clusters, or possess distinct textures. In other cases, the user must direct the search for matching pixels by specifying swatches indicating corresponding regions in the two images. Furthermore, the technique does not explicitly enforce spatial continuity of the colours, and in some images it may assign vastly different colours to neighboring pixels that have similar intensities [63].

In a bid to increase the accuracy and ease of colorization, a seminal approach that utilises 'scribbles' from a user was proposed by Levin et. al [63]. Here colorization is performed by optimization, and exploits a fundamental hypothesis that areas of similar luminance should have similar colours. This assumption, together with additional colour scribbles placed on the interior regions of objects in the gray image, is used to propagate colour to the rest of the image by minimisation of a quadratic cost function. The result is a visually pleasing image with a reduction in user interaction.

A number of recent advancements have since been made to improve the quality and efficiency of the colorization process. These works can roughly be divided into scribble based and example-based colorization. The former technique is used in [103] where a computationally simple, yet effective, approach is presented which works very fast and can be conveniently used 'on the fly', permitting the user to promptly get the desired results after providing a set of chrominance scribbles. [82] presents an interactive colorization system that makes it easy to colorize natural images of complex scenes. Their energy optimization propagates colour labels to intensity-continuous and texture-similar regions that may be far apart in image space and disconnected. This labeling

scheme drastically reduces the amount of interaction in scribbling the strokes. The paper [55] develops the method of transferring colour from a segmented example image, and uses the method in [63] to produce the finished colorized image. This method has the advantage of not relying upon the user's skill or experience in choosing suitable colours and strokes for a convincing colorization.

5.1 Bayesian Analysis of the Colorization Problem

Partially inspired by the work [63], in this chapter we give a Bayesian analysis of the colorization problem [5]: We begin with a gray level natural image in the RGB colour space where a user has placed their own points of colour. Converting to the YUV colour space we now have the gray image Y and points U_o on a subset of pixels S in the U channel which the user has marked (the procedure is similar for both U and V channels so we only explain for one).

Now the problem is to find an estimate U' on the whole image s.t.

(c1) $U'|_S = U_o$,

(c2) and the resulting colour image looks natural.

Formally we have the following: For any A let us denote by $P_Y(A)$ the conditional probability $P(A|Y)$. Then we wish to maximise $P_Y(U'|U_o)$. Applying Bayes' formula results in maximising $P_Y(U_o|U') \cdot P_Y(U')$, or equivalently to find

$$\arg\max_{U'} P_Y(U'), \tag{5.1.1}$$

under condition (c1).

To model the prior $P_Y(U')$ we utilise the sparse filter response of (4.1.1), which we modelled using a GGD. Hence we have the expression

$$P_Y(U') \propto e^{-\sum |F_i \cdot U'|^\alpha}, \tag{5.1.2}$$

where F_i is the filter operating on the i'th pixel in the image. Taking logs leads

to an equivalent minimisation objective,

$$\arg\min_{U'} \sum_i |(F_i \cdot U')|^\alpha \quad \text{s.t.} \quad U'|_S = U_o. \tag{5.1.3}$$

Here the parameter α now details the form of the prior assumed for the filter response. Taking $\alpha = 2$ gives the same optimisation problems solved in [63] which illustrates that their approach effectively assumed a Gaussian response of the filter F_i. However, the analysis and modeling of natural images in chapter 4 has shown that α is almost always less than one. Hence we arrive at the correct optimisation problem. Solving (5.1.3) for this case leads to a non-convex optimization problem that unlike least squares regression has no explicit formula for the solution. Instead we convexify the problem using L^1 optimization which often gives the same results for sparse signals [22].

Taking $\alpha = 1$ we can rewrite the objective term of (5.1.3) in the vectorial form

$$||AU'||_1, \tag{5.1.4}$$

where $||\cdot||_1$ represents the L^1 norm. A is an $N \times N$ matrix where the i'th row corresponds to the filter response of the i'th pixel in the image. The constraint term of (5.1.3) is incorporated into a matrix B of size $|S| \times N$ and with a column vector b holding the values of the marked pixels, U_o. This allows the problem to be written as a Linear Program (LP) through the addition of two slack variables ν_i and μ_i:

$$\begin{gathered} Min \sum_i \nu_i + \mu_i \\ \text{s.t. } AU' + \nu - \mu = 0 \\ BU' = b \\ \nu_i, \mu_i, U'(i), b_i \geq 0 \end{gathered} \tag{5.1.5}$$

The objective function and the first constraint allow us to find the smallest pairwise addition $\nu_i + \mu_i$, such that their difference is equal to $b(i) - A_{i\rightarrow}U'$. This occurs precisely when one of the ν_i or μ_i are zero and the other equal to $b(i) - A_{i\rightarrow}U'$, and allows us to handle both the positive and negative cases.

5.2 Colorization Examples

We solve the linear programs using the package LIPSOL [104] which is available through high-level programming environments Scilab and Matlab. Images in the region of 250×250 pixels each take a few minutes to colorize, hence our method is slower than the solvers used in [63]. However, our goal here is not to efficiently solve such problems, but only to state the correct optimization problem and to show that when such a problem is solved, the resulting colorized image is of a higher quality.

The results shown in the figures compare the quality of the colorization using L^1 optimisation against the approach of Levin et. al (L^2 optimization). Marking large regions of pixels gives similar results, however, using a much smaller set of marked pixels highlights the differences between the two methods. (We note here that since we are only concerned with the correct propagation of colour, and not the choosing of colour, we use the original colour channels of the images for marking colour points.)

Figure 9.16 shows an example where we colorize using a sparse set of marked pixels placed arbitrarily on the image regions. (a) shows the gray image with the marked colour pixels and (b) the original image for reference. (d) shows the improvement in colorization using L^1 optimization over the L^2 approach in (c). We observe more vibrancy in the colours in (d) against the general 'washed out' look of the colorization in (c). Colour blending is also apparent, especially in the green leaves (at the bottom and centre left) which have taken a red tinge from the pink petals and the red roses. Overall we have a sharper result and not an oversmoothed output as usually is the case for assuming a Gaussian prior.

Figure 9.17 shows another example where L^1 minimisation gives a sharper colorization compared to the L^2 approach. Here the latter method incorrectly colorizes the red balloon in the centre of the picture as purple. We also observe more vibrancy in the colours in (d) over (c).

Figure 9.18 shows similar results where colorization using L^2 optimization produces artifacts of 'washed out' colour against the sharper results of the L^1 approach. Here we see for example that the central red pepper has its colour

blended with the surrounding green peppers resulting in an incorrectly colorized image. L^1 has given an almost indistinguishable image from the original.

Figure 9.19 shows an example image taken from the paper [63] where we now colorize using a sparse set of marked pixels. Again we observe an overall 'washed out' result using the L^2 approach against the L^1 minimisation. As examples we see that the cushion in the background has had its blue colour blended with the brown from the boys hair and the yellow from the t-shirt. The child's left eye is also incorrectly colorized brown instead of blue, this has all resulted in an overall loss of colour vibrancy.

Finally Figure 9.20 illustrates the oversmoothed output obtained using a Gaussian prior against the L^1 optimization. We observe in particular that the blue feathers of the bird on the left have had their colours blended with the green and yellow, also the red feathers of the bird on the right exhibit much more colour vibrancy. The example illustrates the colour sharpness and vibrancy obtained when using L^1 optimization over L^2.

Interestingly our findings coincide with the results obtained in [15], where the authors explore the use of colorization as a means for compression. Here the gray image and colour seed pixels are stored as a compressed representation of the original image. Levin et. al's algorithm is then used for decompression by colorizing the gray image. Good levels of compression are achieved but at the cost of 'washed out' colour artifacts. Similarly in [32] and [53] the given examples illustrate these artifacts when using colorization as a tool for compression.

Our technique has some interesting consequences where few seed pixels are chosen for colorization. As a means for compression this is an interesting and natural avenue to pursue and develop. However, at present the effectiveness of the L^1 approach still needs to be improved. While we obtain sharper results in areas where colour information is sufficient, we also observe incorrect colour artifacts in regions where not enough colour information has been given. This is opposed to the L^2 approach which simply results in washed out colour artifacts. These areas then require additional colour markings in order to give convincing colorizations. In future it would also be useful to look into automatically select-

ing the required colour points in a given image, or automatically obtaining the information from reference images, and combining all this with the effectiveness of the L^1 approach. This would reduce the amount of labour and skill required for placing and choosing colours, and also lead to more natural looking images.

Chapter 6

Image Compression using Elements of Sparsity

The advent of digital imaging has led to an explosion in the amounts of data people are capturing, storing and transmitting across the world. A key element in these activities is compression. Compression algorithms are able to reduce data by many orders of magnitude and allow the efficient management of images. Of particular interest are lossy compression schemes, such as the popular JPEG standard, which aim for high data reduction with minimal perceptual loss. These schemes often take advantage of the sparse representation of images in a suitable basis, keeping the largest coefficients that capture the essential information whilst discarding the rest. In line with this philosophy of lossy compression, in this chapter we explore a method that uses elements of compressive sensing and colorization, as a tool for compression [7].

This atypical approach to colour compression was first explored, to the best of our knowledge, by the authors of [15]. Here they operate in the YUV colour space and use a variable grid of points to sample the colour information of the pixels at the intersections. Decompression is performed by using the sampled colour information as seeds for the colorization algorithm used in [63]. This optimisation scheme propagates colour by assuming pixels with similar intensity

should have similar colour, and consists of minimising a quadratic cost function constrained by the given colour information.

Colour compression was also proposed in [53] where chromatic information is sampled by generating a set of rectangular boxes in the CIELAB colour space. Pixel distribution is covered with the regular lattice points inside the min-max colour ranges of the distribution, and seeds chosen from each body center from equally divided unit boxes. Decompression is performed by utilizing two assumptions of natural images: Firstly, that if two arbitrary points in an image are close in the Euclidean distance, then the chrominance distance Δab is small. Secondly, that pixels with similar intensity should have similar colour. A weighted combination of distance functions incorporating these assumptions is then used for the propagation of colour.

The authors in [32] also experimented using colorization for compression where they sample colour information from a few representative pixels to learn a model which predicts colour on the rest of the pixels. A graph based inductive semi-supervised learning module is then used for the colorization, and a simple active learning strategy to choose the representative pixels.

With view of the previous work, we utilise the statistics of natural images and elements of compressive sensing to measure and reconstruct colour information. The following sections describe our measurement and subsequent reconstruction processes.

6.1 Sampling using Sparse Matrices

In the following schemes the U (and V) elements are sampled using random pixel selection or a random linear combination of the pixel values. Both processes can be approximately expressed as measurements in the compressive sensing framework using a sparse binary random matrix (SBRM) [12]. Beginning with the direct pixel selection in the spatial domain, we create a SBRM θ of size $m \times N$ ($m << N$) which only has one unique element $\{1\}$ in each row corresponding to sampling m pixels from each of the chroma channels. The parameter

m describes the rate of compression where smaller values imply less sampling and higher values more sampling. The rasterised chroma components are then multiplied by θ and the obtained measurements stored as our compressed data. This process can be considered within the compressive sensing framework with reconstruction accomplished in a similar fashion. However, for θ to truly be a measurement matrix it needs to satisfy the Restricted Isometry Property (RIP) [22] for accurate reconstruction using LP.

The second sampling scheme we consider is a SBRM matrix φ of size $m \times N$ formed in the following way: for each column, d random values between 1 and m are generated, and 1's are placed in that column, in rows corresponding to the d numbers. If the d numbers are not distinct, the generation for the columns is repeated until they are (this is not really an issue when $d << m$). We chose to use $d = 8$ and store the measurements $z = \varphi U$ as our coded colour data. By sampling random linear combinations of pixel elements this method increases the probability of our measurement matrices being suitable within the framework of compressive sensing and sparse recovery. Indeed, the matrix φ has been shown to satisfy a weaker form of the RIP [12].

6.2 Decompression by Sparse Recovery

The reconstruction process involves solving a convex optimisation problem where we seek the solution to the program,

$$\arg\min_{U'} \sum_{i} ||(F_i \cdot U')||_1 \quad \text{s.t. } \phi U' = z = \phi U, \tag{6.2.1}$$

where F_i is the filter (4.1.1) operating on the i'th pixel and ϕ is the measurement matrix which is either θ or φ. In words (6.2.1) is searching for the N-pixel image with the sparsest filter response that explains the measurements we have observed. This problem is similar to the one (indeed, identical when $\phi = \theta$) solved in the Bayesian analysis of the colorization problem outlined in chapter 5, where the formulation leads to solving (5.1.3) for the convex case $\alpha = 1$. Hence, (6.2.1) can be written in a vectorial form and solved using LP as in chapter 5. The reconstruction process exploits the fact that the filter responses of natural

images observed in Chapter 4 have a sparse distribution. Hence the U (and V) component is compressible using the random matrices and reconstructible using L^1 optimisation.

6.3 Compression Examples

Figure 9.21 shows examples where an uncompressed bitmap image is compressed using randomised seed selection and compressive sensing. Here the monochrome image is stored in uncompressed format and colour information sampled at a rate of around 5% of the original image. Further compression can be achieved with visually indistinguishable results by storing the gray component using JPEG. Sampling at lower rates resulted in increased artifacts in the decompressed images. The results show that convincing reconstructions can be made from a small amount of compressed data using L^1 optimisation. The PSNR values quantify the results and show acceptable values for a lossy compression scheme.

We note here that the compression scheme sampling seed pixels at a rate of 5% gives similar results when decompressing using L^2 or L^1 optimisation. Reducing the rate further leads to washed out colour artifacts with the former method and incorrect colours using the latter. However, with results from section 5, in future it would be useful to incorporate the selection of as few seed pixels as possible together with L^1 optimisation in order to increase the rate of compression. In the case of compressively sensing the chroma components, L^2 reconstruction fails as it almost never returns a sparse solution.

Chapter 7

Chroma Reconstruction from Inaccurate Measurements

Denoising is a fundamental problem in image processing due to the fact that images, no matter their content, usually contain some degree of noise. This is often regarded as a form of image degradation and the goal of denoising algorithms are to form an estimate x' of the the original image x given the observed noisy version x^*, modeled as

$$x^* = x + n, \tag{7.0.1}$$

where n is the matrix of the random noise pattern.

The principal causes of noise in digital images arise during image acquisition (digitization) and/or transmission. This can be caused by several factors such as low light levels, sensor temperature, electrical interference, malfunctioning pixels and interference in the channels used for transmission. The distribution of noise can be several, such as white, impulse or multiplicative, each giving its own characteristic form of degradation [49].

Various algorithms have been introduced with success over the past few

decades for denoising images. The proposals, in their original form, have sparked an abundant literature resulting in many improvements in quality and speed. These algorithms can be categorized into several groups including Wavelets, Bilateral filtering, Anisotropic diffusion, Total variation and Non-local methods. Readers are advised to see [17] and [71] for comprehensive reviews and comparisons of the best available versions together with powerful novel approaches.

Some recent algorithms to mention include [70] where the authors propose a unified framework for two tasks: automatic estimation and removal of colour noise from a single image using piecewise smooth image models. Their segmentation-based denoising algorithm is claimed to outperform current methods. This paper also contains an interesting introduction that discusses the current state of the art methods for image denoising. Another recent algorithm which claims to lead to excellent results is C-BM3D [34]. In this scheme the authors propose an effective colour image denoising method that exploits filtering in a highly sparse local 3D transform domain in each channel of a luminance-chrominance colour space. For each image block in each channel, a 3D array is formed by stacking together blocks similar to it. The high similarity between grouped blocks in each 3D array enables a highly sparse representation of the true signal in a 3D transform domain, thus a subsequent shrinkage of the transform spectra results in effective noise attenuation.

The importance of denosing in image processing has also led to many commercial and freely available software. These include Neat Image, Noise Ninja, DenoiseMyImage, Photoshop, Topaz Denoise, Gimp and many more. The programs often incorporate a host of image enhancement tools to collectively remove typical forms of image degradation. A full evaluation of so many programs is difficult, especially since each has parameters which a user can change for subjective suitability. However, from general usage and reading it has been found that Noise Ninja and Neat Image are among the best used noise reduction programs. DenoiseMyImage is also a current alternative that uses a modified form of the state of the art non-local means method. Readers may view [3] for a comprehensive *user* comparison of current software.

Denoising algorithms are usually fed a noisy RGB image corrupted in each channel. Most methods have been formulated as a channel by channel or vectorial model. In the former case the RGB values are mapped to a colour space such as YUV or Lab or any other suitable space to separate the luminance and chroma, with the denoising algorithm *usually* applied to each band. Since the luminance channel contains the main structural information and chroma noise is more objectionable to human vision (as opposed to the film grain appearance of luminance noise), separation allows more intensive denoising of the chroma channels without too much loss of detail. These models take into account the human perception of colour and allow us to handle the particular characteristics of the noise affecting each component. Methods based on their luminance-chromatic decomposition are well known for their excellent results with [34] being a recent example. Furthermore, in the process of transmission, the reduction of bandwidth for the chroma allows errors and artifacts to be more easily compensated for than using a typical RGB model.

In this chapter we propose a novel algorithm for removing noise from real images and also white and impulse noise from the chroma channels of an image in the colour space YUV, where a *good* version of the Y component is obtainable. (Due to the similarity of the colour components, from here on we interchangeably mention either the U or V channel, where analysis of the other is obtained by substitution). Algorithms such as those in [34], [44] and [14] have successfully exploited the information in the luminance channel for effectively filtering the chroma components. In line with this philosophy our approach utilises the non-linear filter response distributions observed in chapter 4 as a regularization term (a prior, in Bayesian analysis) to penalize solutions that don't give a *desired* sparse solution when filtered.

7.1 Reconstruction Procedure

We consider real noisy RGB images that have been corrupted by unknown noise which are then transformed to the YUV colour space. Due to the properties

of the underlying natural colour images, such as high correlation between R, G, and B channels, we note that Y has higher SNR than U and V and that it contains most of the valuable information such as edges, shades, objects, texture patterns, etc. The U and V contain mostly low-frequency information with iso-luminant regions, i.e. variation in only U and V, being unlikely. Thus removing chroma noise through knowledge of gray information is plausible. We chose to use Neat Image or DenoiseMyImage when appropriate to denoise the Y channel when needed. We additionally used them as a benchmark for testing our algorithm. Furthermore, our algorithm is also tested against images in the YUV space suffering from impulse noise only in the chroma channels.

Thus, given the noisy chroma component U^* and a denoised gray image Y, our task is to recover a good approximation U' of the original element U. This model results in the following optimisation scheme,

$$argmin_{U'} \ ||F \cdot U'||_1 + \lambda ||U' - U^*||_d. \tag{7.1.1}$$

Given an $n \times m$ image, (we abuse the notation a little and have) F here is an $nm \times nm$ matrix whose rows correspond to filtering a single pixel where U' and U^* are $nm \times 1$ column first rasterized vectors. U' is the estimate we seek of U, while U^* is the noisy observation of U.

The first term is our penalizing function which takes small values for desirable solutions and the second is the *fidelity* term. The parameter d is taken to be either 2 or 1 reflecting the norms proposed in the measurement of the distance between the two vectors. In words, this optimisation scheme searches for the estimate image U' with the sparsest filter response and with the second term encouraging the solution to be close to a noisy chroma measurement U^*.

For an image assumed to be corrupted by Gaussian noise our reconstruction process involves solving (7.1.1) with $d = 2$, where the fidelity term encourages solutions to be close to the noisy version in the L^2 sense. When the noise is taken to be impulsive and affecting the image at random points by taking extrema values, we solve (7.1.1) with $d = 1$. Modifying the fidelity term to $d = 1$ (i.e. L^1 norm) has been studied with success within the Total Variation framework, as reviewed in [30].

An important parameter in our algorithm is the value of λ which controls the relative weight of the difference between the noisy channel and the solution. Too small a value and the optimisation results in an overly smoothed output, while too high a value results in a solution that is too close to its noisy version. We found experimentally that $\lambda \in (0, 5]$ gave the best results, with half-integer increments for optimality.

7.2 Denoising Examples

Our optimisation problem was solved using CVX [50] which is a convex programming package implemented in Matlab. The images that we used are of sizes in the region of 200×200 pixels, which took up to a couple of minutes to denoise. However, our aim here is not to pose a fast algorithm but only to show the applicability of such a scheme for denoising chroma channels. The algorithm is parameterised by the value of λ whose value is given in the text accompanying the figures.

Fig. 9.22(a) shows an example RGB image which is made severely noisy by adding Gaussian noise of mean zero and variance 0.01 to all the channels as shown in (b). (c) shows the denoised image obtained using Neat Image and (d) the result obtained using DenoiseMyImage. Neat Image was used at maximum setting while DenoiseMyImage was used at an adjusted medium level to obtain the best results. Neat image still left considerable noise like artifacts in the image, while DenoiseMyImage gave a less noisy but much smoother output. The result using our algorithm is shown in (e) where we used DenoiseMyImage to denoise the gray component. Visually comparing the results shows that our algorithm gives an intermediate result which is better than using NeatImage, while the colours are much more vibrant and appear sharper than when using DenoiseMyImage. This is also further justified by the peak signal to noise ratios (PSNR) which quantify the results, and shows our algorithm having a higher but similar value.

The next examples focus on real world images where the type of noise af-

fecting the image is unknown. We begin with Fig. 9.23(a) which shows an image that is severely affected by colour noise. This is typical of an image taken in low light conditions with high ISO settings. (b) shows the image having been denoised using Neat Image. This program requires a suitable region to be selected for noise estimation, after which luminance and chrominance noise reduction can be individually adjusted. We required 100% noise reduction on all components due to the high amount of noise present in the image. (c) shows our algorithm where the luminance channel was denoised using Neat Image and the filter matrix F constructed from it for reconstructing the chroma channels. (d) shows the result of using DenoiseMyImage. We observe that our algorithm gives similar noise reduction compared to the existing methods.

Fig. 9.24(a) has been taken from some examples given on the Neat Image website. This is a crop of a television frame captured with a computer TV card. The image has strong colour banding visible across all the image caused by the electric interference in the computer circuitry. Similar banding is sometimes observed in digital camera images (caused by interference too). The banding degradation does not affect the luminance, however all channels still show grain like noise. (b) shows the best Neat Image result obtainable by denoising the chroma and luminance at 100%. However, the banding is still evident in the result. (c) is the result of our algorithm which clearly removes the noise. (d) is the best result obtainable using DenoiseMyImage which is still unable to remove the banding noise.

Our algorithm is able to remove this type of noise by filtering only the chroma channels and using Neat Image for clearing the fine grain luminance noise. The result is free of the colour banding and (f) shows that the V channel does not display any of this degradation against the V channel when using Neat Image (e). We are able to attain this result as we are filtering the chroma channels through taking account of the underlying gray level structure. Since the colour banding is not appearing in the luminance, minimisation of the filter response favours areas of homogeneous colours while the fidelity term bounds the colours to being close to the original.

The final two examples illustrate the flexibility of the model in handling chroma noise taking a different distribution. Fig. 9.25 shows an example of a clean image (a) which is transformed to the YUV colour space and impulse noise of density 0.05 added to the U and V channels only. Our algorithm with the fidelity term measuring L^1 norm is able to denoise such that the recombined RGB image shown in (b) is visually identical to the original. The detailed look of the chroma components reveals no sign of the impulse noise, while the PSNR is of a good value.

Fig. 9.25 shows another example of an image that has been corrupted by impulse noise and reconstructed. (a) shows the original image, (b) the RGB image with noise having been added to only the chroma channels and (c) shows our reconstructed image. The results illustrate again that noise has been successfully removed to a very high standard with good PSNR values, and this is further justified by looking at the chroma channels which have had their impulse noise removed. Neat Image and DenoiseMyImage are unable to effectively denoise the images affected by impulse noise. Instead we obtain a 'washed out' look with the impulse points still remaining. An example is shown by (d).

We have illustrated how knowledge of the statistics of natural images can be incorporated into an effective denoising scheme. Our objective was to propose a novel algorithm for removing chroma noise from digital images by operating in a luminance-chrominance colour space. The flexibility of the model was also shown by its ability to handle different types of noise very effectively. This was accomplished by altering the fidelity term to measure L^1 or L^2 norm as required, and shows concentration on gray level denoising gives sufficient information for colour channel reconstruction. In future it would be most useful to robustly test this approach across diverse datasets of images and also in other colour spaces where we may observe increased performance. We are also looking at algorithms for solving the optimisation scheme much more quickly and looking at applying the approach to denoising hyperspectral images.

Chapter 8

Discussion

This thesis has acquainted us with the fact that images of the natural world around us are far from random entities. They are predictable, as shown by the high redundancy which they exhibit. They also display non-Gaussian, high kurtosis distributions and the statistics of ensembles of natural images have been observed to be approximately scale invariant. These properties of natural images give us a certain level of regularity and provide stable platforms for image modelling. In this chapter we will summarise the main contributions of the thesis and present an outlook for future work, as well as possible applications in other related areas of signal processing.

The sparse distributions observed across databases of grayscale and colour images have been shown to increase the performance of image processing applications. This has been accomplished by utilising Bayesian analysis and incorporating the image statistics as prior information. In our work we showed how non-linear filtering of colour images produces highly kurtotic responses which build upon previous decades of work showing the sparse nature of image statistics. This was empirically found to hold on our diverse dataset of images and also on arbitrarily chosen images obtained from standard test datasets and the Internet. The finding that the distributions are sparse was an important step, however, we also observe large variability in the shape value α across different types of images. Here we can note that second order statistics involving the

power spectra of natural scenes has already been utilised for scene classification objectives [97]. Additionally, it has been observed that the marginal distributions across different categories of databases have shown to exhibit similar, yet particular types of distributions [64]. Hence it would be interesting and useful to explore the average distributions, and their respective α values, of the filter responses for categories of images. This information could be utilised for improving the image processing applications presented in this thesis. In future we will also study the filter under the context of the human visual system. This is due to the success researchers have had with the Laplacian of Gaussian and Gabor filters in modelling early vision.

Our approach to colorization has been the first to couple sparse image statistics and L^1 optimisation for this long standing problem in image processing. The L^1 optimisation scheme enables users to place fewer colour scribbles on a grayscale image, while also producing colorization results that display sharpness and colour vibrancy. In future we will be looking to overcome some of the drawbacks of the method. Namely, we would like to increase the speed of the solvers and automatically choose and reduce the number of points in which a user is required to place colour scribbles. The former problem is already being tackled by many researchers due to its relation to the L^1 sparse recovery problem in compressive sensing. Already, the tool of Bregman iterations is being used to give fast, accurate iterations for constrained L^1-like minimization problems, and to improve the regularization quality of nonsmooth regularizers such as L^1, total variation, and their variants.

The presentation of a compression scheme for colour images has clearly shown the high redundancy found in colour data. Our algorithm utilised earlier elements of the sparse nature of the non-linear filter response distributions, together with elements of colorization and compressive sensing. We were able to present two similar, but quite different, compression schemes that utilised sparse random binary matrices (SBRM). The first simply turned around the idea of colorizing a natural gray image, by storing few colour points as our compressed data and effectively using colorization as a mode to decompress

the images. The second approach took compressive sensing measurements using SBRM, and utilised L^1 constrained sparse recovery for the decompression stage. This atypical approach to compression will need to be studied further and it shows most promise in compressing hyperspectral images where currently the data transfer streams are too high for efficient processing.

The final application we presented deals with the classical problem of removing unwanted noise from natural images. Here, working in a luminance-chrominance colour space we were able to show how information from the gray level image can be utilised to effectively denoise the chroma channels of an image. The model utilised the sparse nature of the statistics of natural images as a regularisation function, and enabled us to form an L^1 constrained optimisation scheme. This problem was again interpretable as a linear program or a convex optimisation scheme and the solution showed how real images corrupted by noise could be effectively denoised. We also experimented with images artificially corrupted by Gaussian and impulse noise and the flexibility of the model to handle different noise distributions was illustrated. In future we will look to increase the performance of the application and explore its use in alternative colour spaces such as *Lab*.

Although substantial progress has been made over recent decades in understanding the complex statistical properties of natural images, we are still quite far from a full probability model. As an example, the existing models do not allow us to accurately capture the variety and complexity of natural images by drawing from the sample distributions. Beyond the univariate and bivariate densities of image statistics the computational burden of the models increases exponentially. Many of the models concerning natural images have described statistical behaviour on ensembles of images, however, their applications for analysis of individual images need to be clarified. Furthermore, aside from the applications presented in the thesis there remain many outstanding problems where developments in statistical understanding will be important [91].

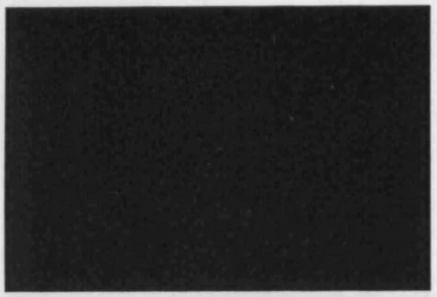
The work presented in the thesis is a starting point for exploring filters on colour channels and their applications to image processing. In future the explo-

ration of image statistics in general, and in particular, colour image statistics will be made. We will look to further understand natural images and propose models that can be utilised in a host of applications.

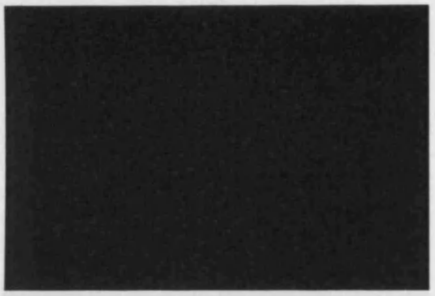
Chapter 9

List of Tables and Figures

This chapter gives examples and tables of results that have been used throughout the thesis. See the corresponding text in the chapters for additional details on each of the tables, sketches, graphs and images.



(a)



(b)

(c)

Figure 9.1: A sample of raw images taken from the van Hateren database. These are the images actually recorded by the camera before any artificial processes are applied such as gamma correction. The images are very dark, and may not be of sufficient contrast when printed.

(a)

(b)

(c)

Figure 9.2: Here are the same pictures in figure 9.1, but after performance of auto-correction using Microsoft Picture Manager, where we are able to see much more detail in the images.

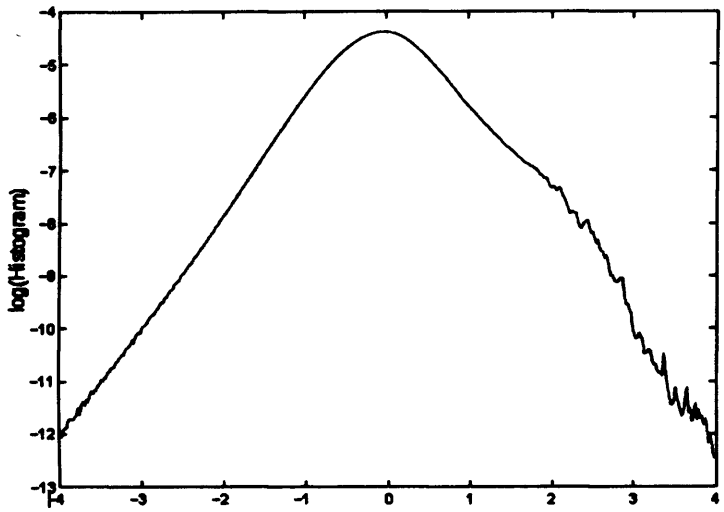


Figure 9.3: We show here the single pixel distribution of intensity values over the whole van Hateren database. The plot is of the ln(histogram) of random variable $ln(\phi(i,j))$ - $< (ln(\phi)) >$. We observe this database statistic to be non-Gaussian-explained by the linear tail on the left and the jagged tail on the right. The kurtosis value is also observed to be higher than that of the Gaussian (see text). (Figure obtained from [60]).

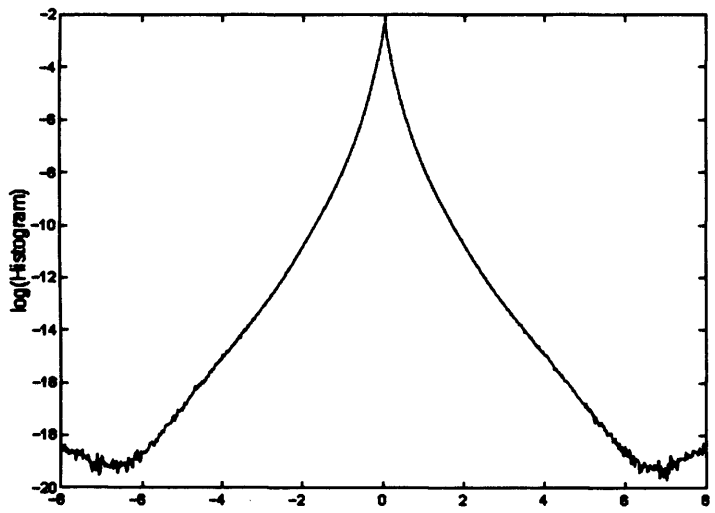


Figure 9.4: We plot here the random variable $D = ln(\phi(i,j)) - ln(\phi(i,j+1))$. This is the log difference of the intensity between two adjacent pixels in a row, and is computed across the whole image and for all images from the van Hateren database. We observe the often seen high kurtosis non-Gaussian distribution which has been commonly found to hold over many different datasets of natural images. (Figure obtained from [60]).

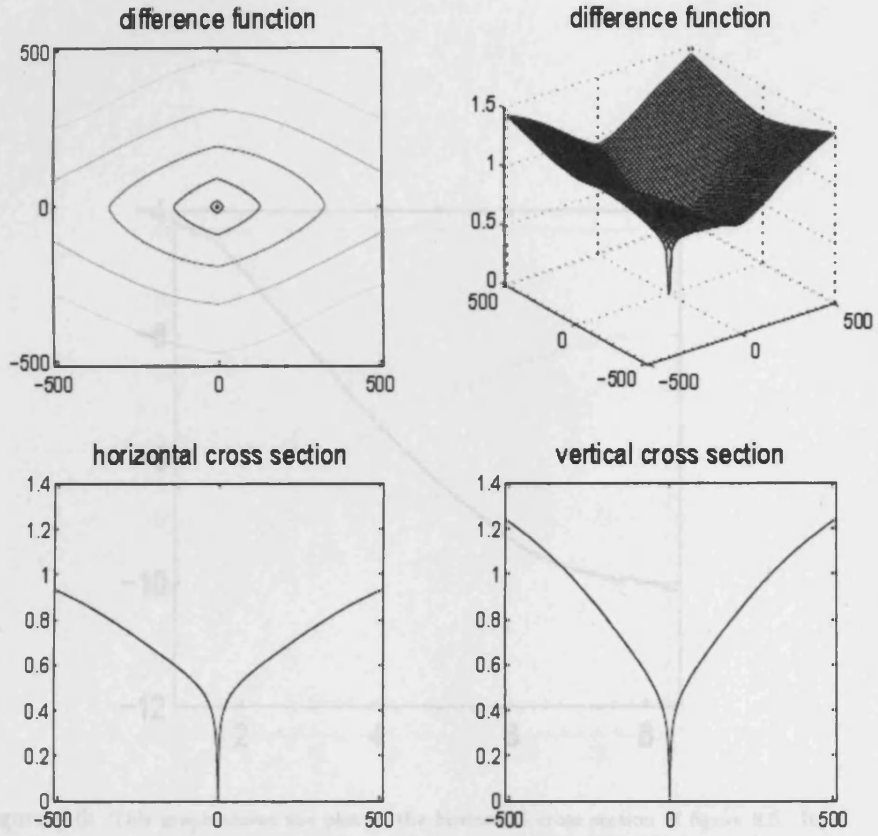


Figure 9.5: The above graphs plot the difference function $D(x,y) =< |\phi(x,y) - \phi(0,0)|^2 >$ and also illustrate a pair of cross sections. The top left shows the contour map of the differences, while the top right image shows us the three dimensional mesh plot formed. Taking the horizontal and vertical cross sections we obtain the bottom two graphs which show that the cross section along the vertical falls much more steeply than that along the horizontal. These plots correspond to evaluating the following difference functions, $D(x,0) =< |\phi(x,0) - \phi(0,0)|^2 >$ and $D(0,y) =< |\phi(0,y) - \phi(0,0)|^2 >$. (Figure obtained from [60]).

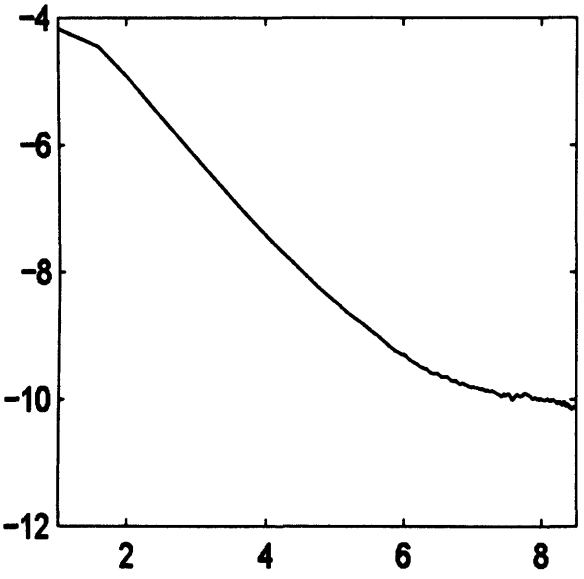


Figure 9.6: This graph shows the plot of the horizontal cross section of figure 9.5. It is plotted on a log-log scale. For pixel distances of 4 to 32, on the *log* scale we have 2 and 5, we observe that the derivative is close to a straight line with a slope -1.19. For larger distances the $log - log$ plot begins to turn and becomes almost a horizontal line around *log* distance 8. (Figure obtained from [60]).

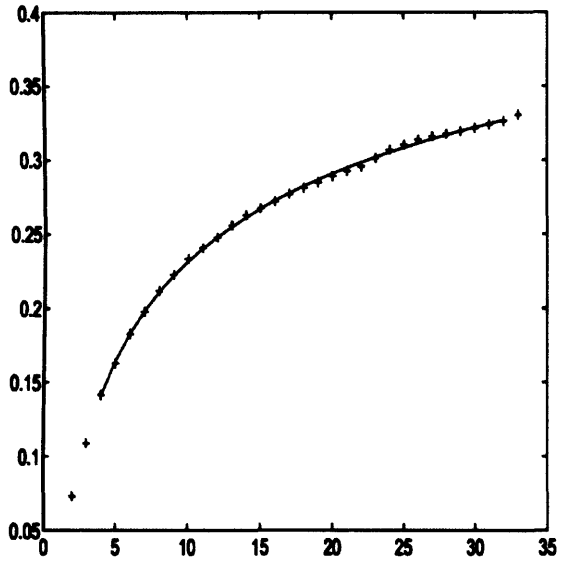


Figure 9.7: We plot the average value of the one dimensional difference function over the pixel range $0-32$ and over 5 million pixels from the van Hateren database using 1400 images. The line of best fit is made using model 2.3.9 and the value of η is found to be 0.19. (Figure obtained from [46]).

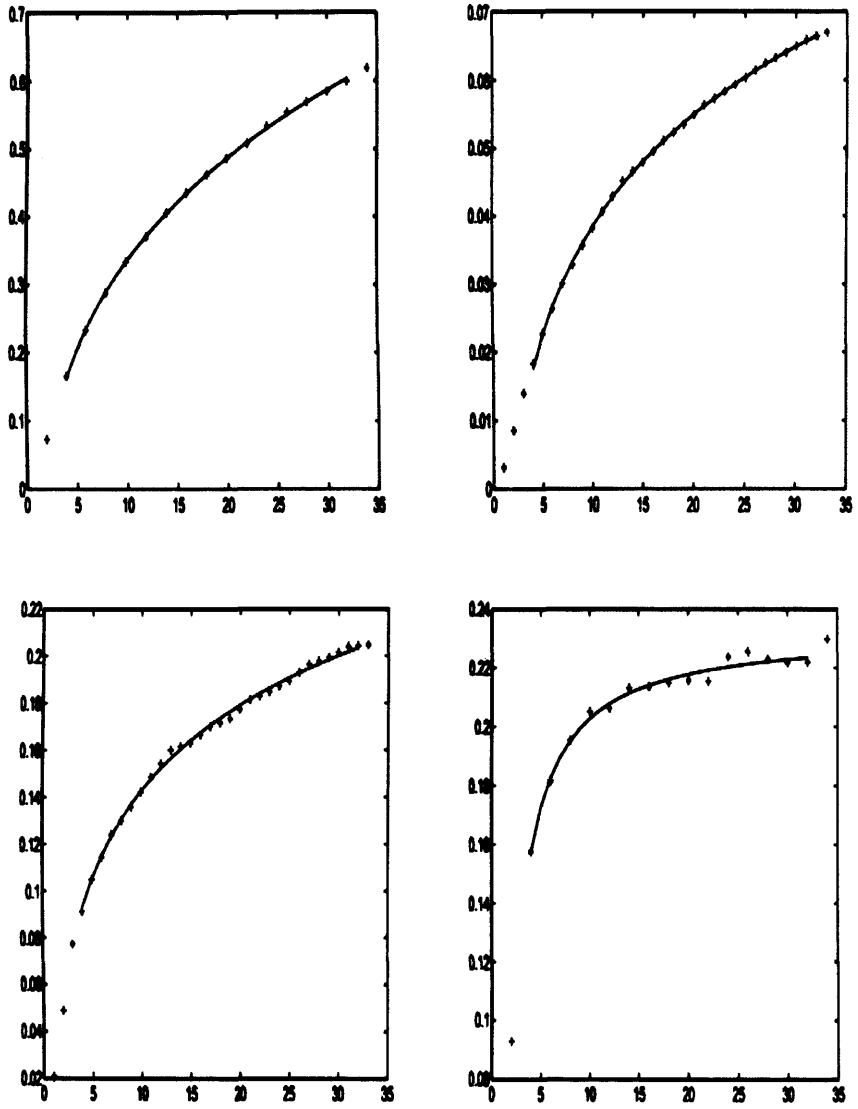


Figure 9.8: We plot here the average one dimensional difference function values obtained for four single images over the pixel range $0 - 32$. We obtain a power law fitting for each of the images although the individual eta values for the images vary. From top left and clockwise: Image 1401, $\eta = -0.2$, Image 262, $eta = -0.05$, Image 712, $eta = 1.01$, Image 59, $eta = 0.04$. (Figure obtained from [46]).

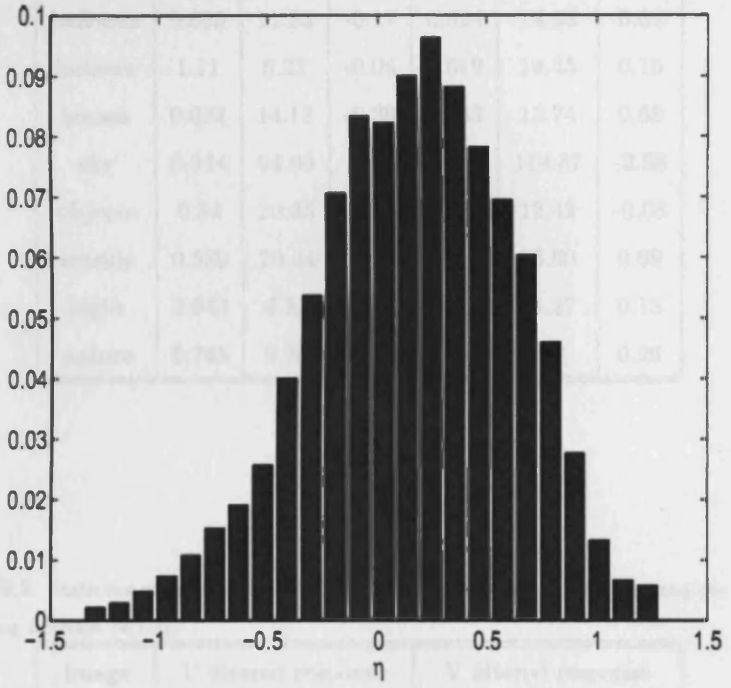


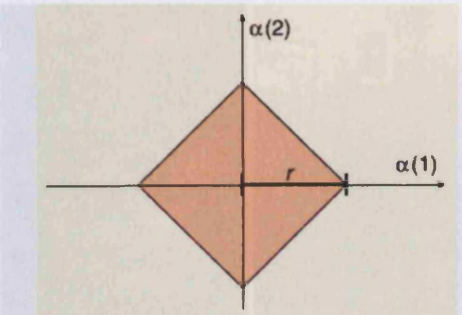
Figure 9.9: The histogram of the individual image η values is shown. We observe a relatively wide distribution giving more weight to relatively higher values of η, and a maximum in the vicinity of $\eta = 0$. We can also conclude that the database η value is not representative at all of the image η's with the distribution of the anomalous dimension being non-Gaussian. The following statistics are associated: Mean = 0.155, Variance = 0.21, Skewness = -0.28, Kurtosis = 2.9, Kurtosis Excess = -0.1. (Figure obtained from [46]).

Table 9.1: Statistics of the non-linear filter response for our sample images using the first weighting function (4.1.2).

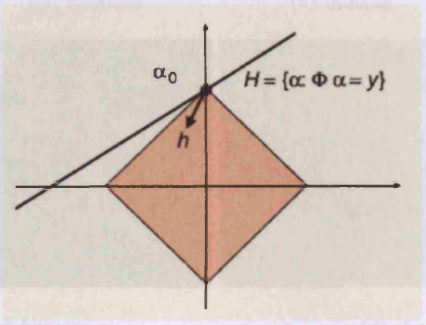
Image	U filtered response			V filtered response		
	α_U	k_U	S_U	α_V	k_V	S_V
balloons	0.695	11.23	-0.17	0.624	14.23	0.03
indoors	1.11	5.22	-0.05	0.619	14.45	0.16
houses	0.624	14.18	-0.39	0.633	13.74	0.68
sky	0.344	94.00	0.64	0.328	114.87	-2.58
objects	0.54	20.35	0.68	0.662	12.43	-0.08
seaside	0.539	20.44	0.63	0.491	26.60	0.09
night	0.944	6.52	0.03	0.561	18.37	0.13
nature	0.745	9.76	-0.11	0.826	8.11	0.26

Table 9.2: Statistics of the non-linear filter response for our sample images using the second weighting function (4.1.3).

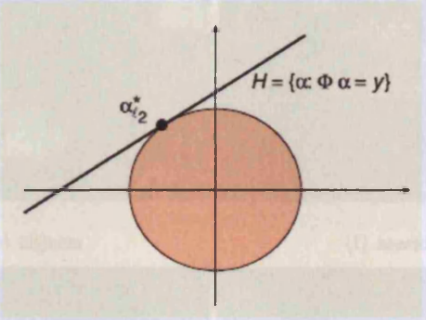
Image	U filtered response			V filtered response		
	α_U	k_U	S_U	α_V	k_V	S_V
balloons	0.685	11.57	-0.19	0.624	14.19	-0.01
indoors	1.094	5.31	-0.07	0.599	15.62	0.16
houses	0.61	14.98	-0.41	0.607	15.14	0.76
sky	0.339	99.38	0.78	0.321	126.27	-2.48
objects	0.534	21.02	0.62	0.654	12.79	-0.07
seaside	0.54	20.39	0.60	0.489	26.84	0.14
night	0.931	6.66	0.07	0.556	18.85	0.18
nature	0.736	10.00	-0.07	0.811	8.37	0.26



(a)

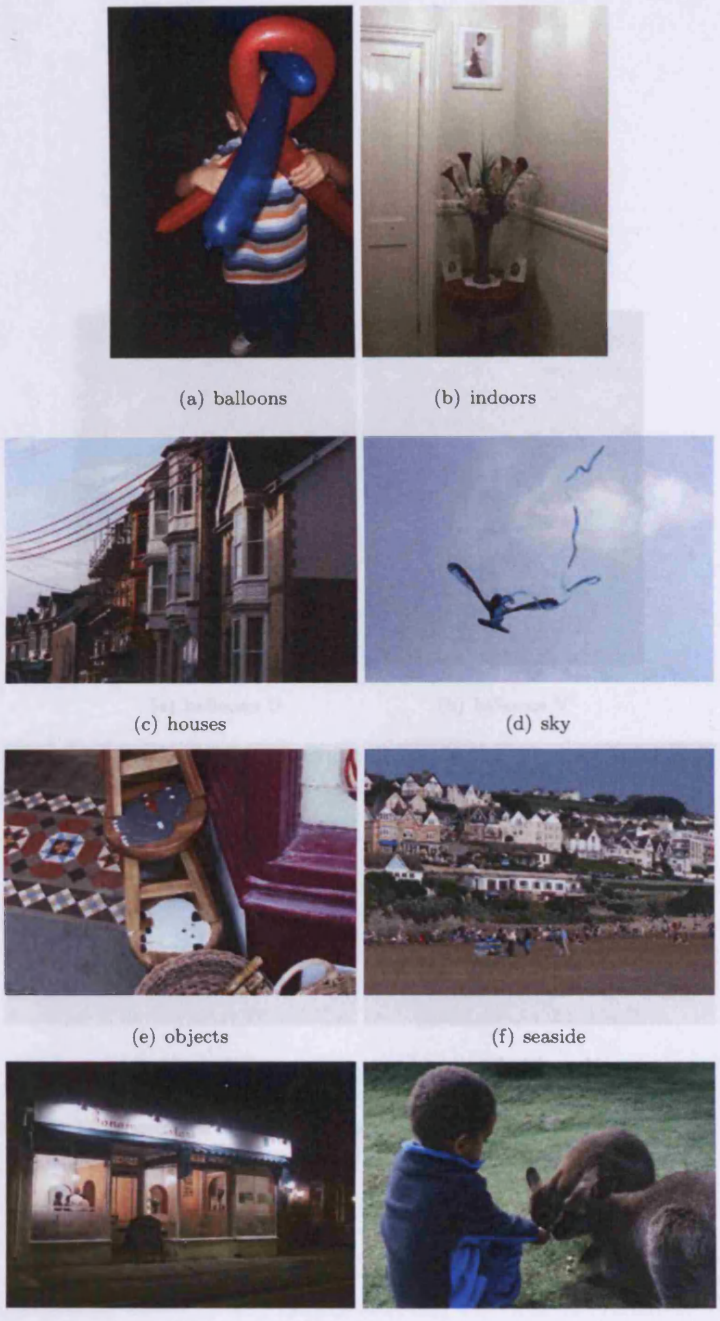


(b)



(c)

Figure 9.10: Geometry of L^1 and L^2 recovery algorithms. (a) shows the L^1 ball of radius $\mathbf{r}$ with the orange region defining the area where $\alpha \in \mathbb{R}^2$ such that $|\alpha(1) + \alpha(2)| \leq \mathbf{r}$. (b) illustrates the solution of the L^1 minimisation scheme 3.3.3 which recovers sparse vectors such that $\Phi\alpha_0 = y$. We clearly observe that the anisotropy of the L^1 ball intersects the space H (of points $\alpha : \Phi\alpha = y$) at α_0. Note that the descent vectors h pointing into the L^1 ball from α_0 will be concentrated on the support of α_0. In (c) we see that finding the intersecting point with the space H that has minimum L^2 norm results in the solution α^*_{l2} which is not generally sparse. (Figures obtained from [83]).



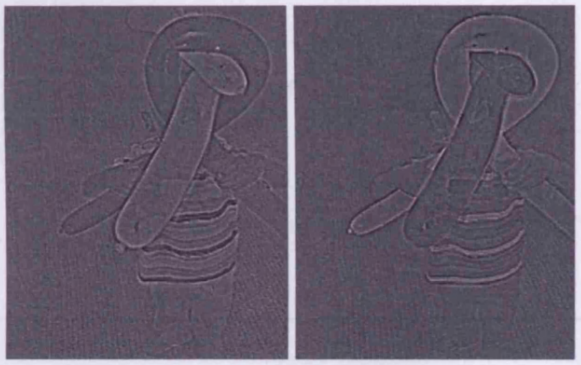
(a) balloons (b) indoors

(c) houses (d) sky

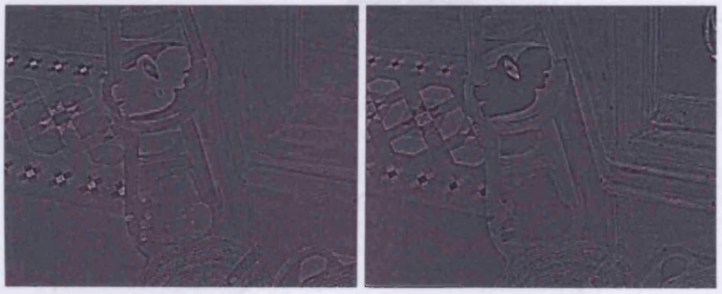
(e) objects (f) seaside

(g) night (h) nature

Figure 9.11: Here we display a sample of 8 pictures taken from our dataset of 25 images. In order to give a measure of robustness to our findings we chose pictures covering a wide spectrum of natural scenes, ranging from natural landscapes to urban environments. Images shown here are all truecolour RGB obtained by a Canon digital SLR camera of varying resolutions in uncompressed bitmap format, and reduced to sizes in the region of 200x200 pixels using Adobe photoshop.



(a) balloons U (b) balloons V



(c) objects U (d) objects V

Figure 9.12: Filter response of each of the colour channels, U and V, of two of our sample images, 'balloons' and 'objects', using the first weighting function (4.1.2).

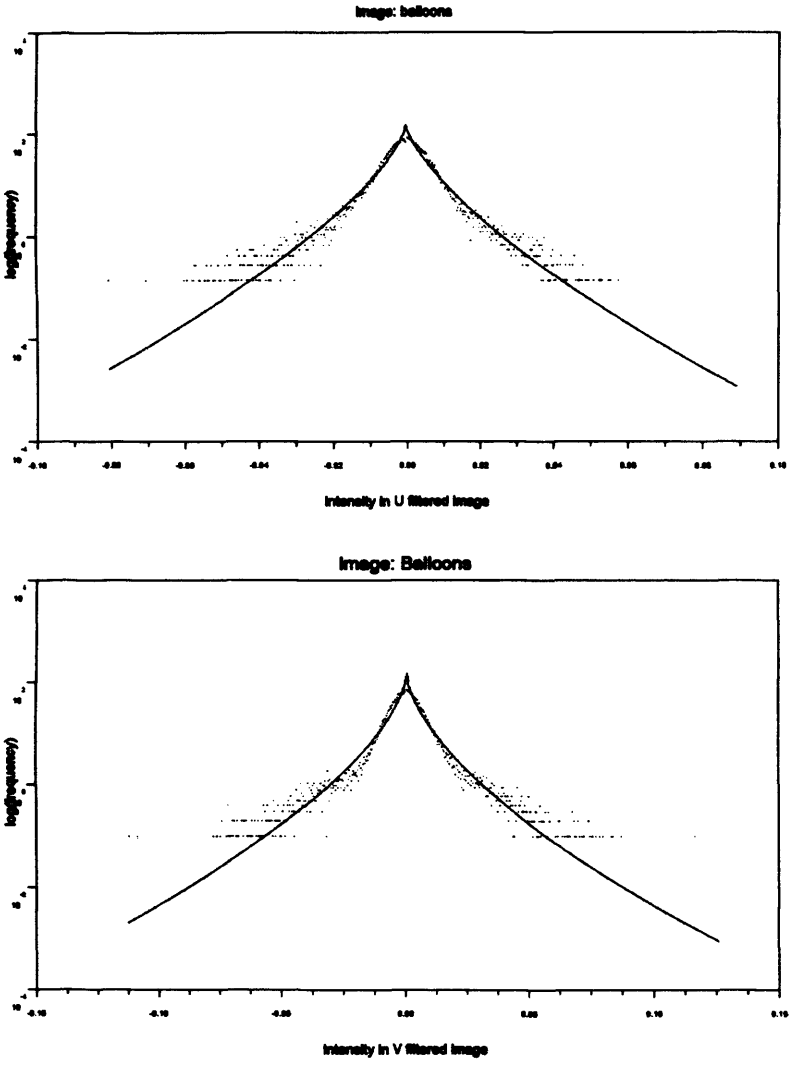


Figure 9.13: Distribution of the filter response for both chromacity channels U and V for the image 'balloons' from figure 9.11 using the first weighting function (4.1.2).

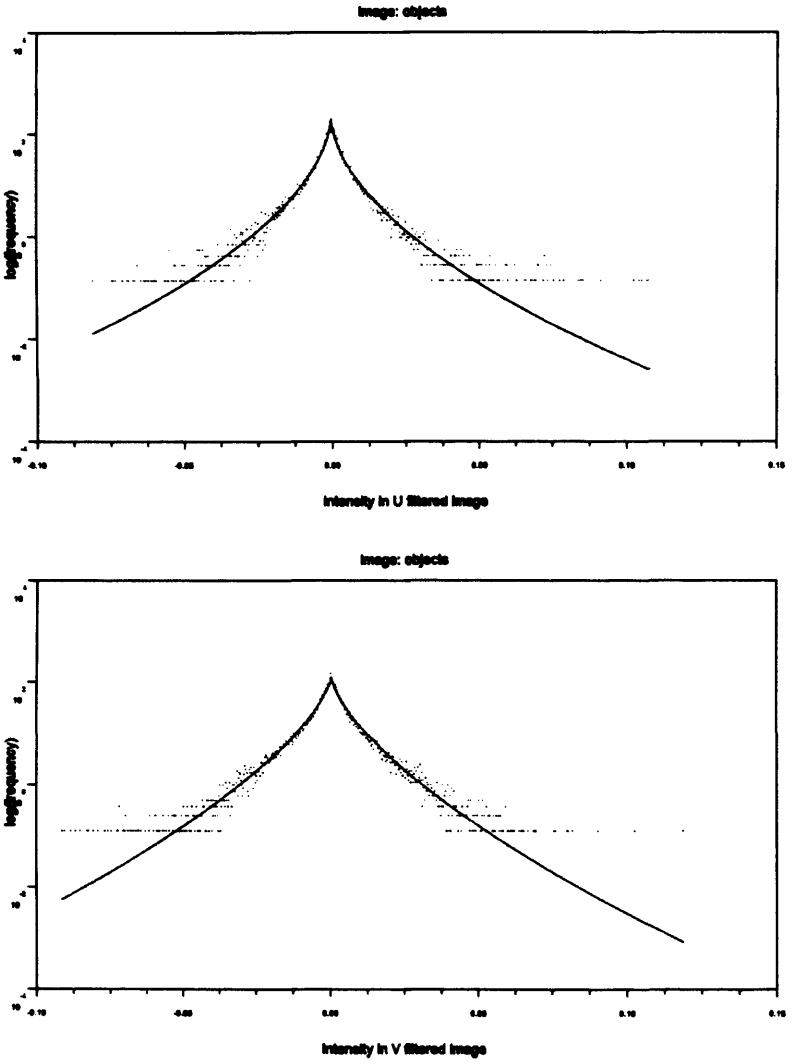


Figure 9.14: Distribution of the filter response for both chromacity channels U and V for the image 'objects' from figure 9.11 using the first weighting function (4.1.2). The non-Gaussian, high kurtosis distribution is clearly observed and is typical of the images used in our dataset.

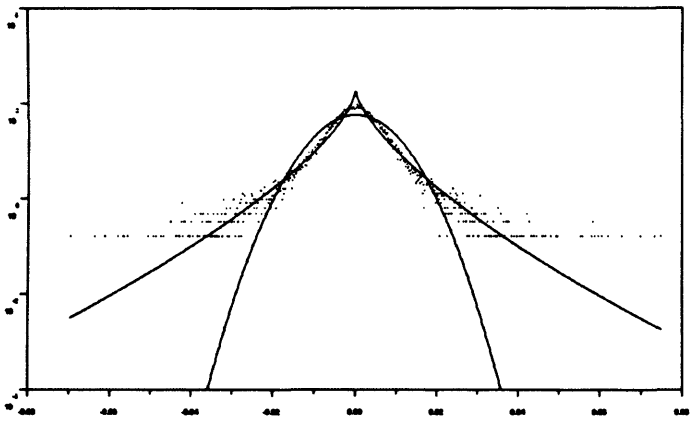
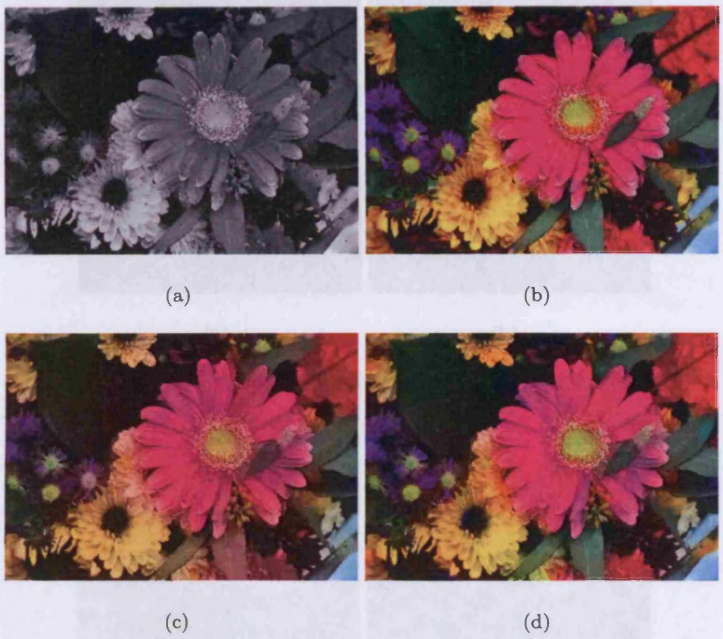


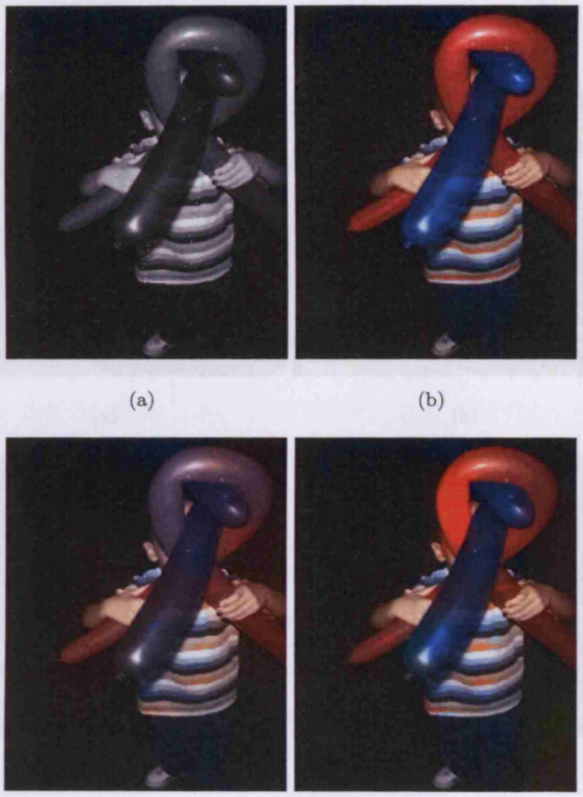
Figure 9.15: Here we show in blue the histogram of the filter response of the U component of image 'balloons'. This distribution is typical of natural images and is shown with the vertical axis on a log scale to better show the nature of the tails. Fitted to the data is the GGD distribution that takes the form of a sparse distribution function. For comparison we have also overlaid the parabola shaped Gaussian distribution which illustrates the difference in the tails between the two models.



(a) (b)

(c) (d)

Figure 9.16: Colorization example. (a) The gray image marked by a sparse set of colour pixels; (b) the original image for reference; (c) colorization using L^2 optimization; (d) L^1 optimization. Here we have colorized using a sparse set of arbitrarily placed marked pixels. We observe more vibrancy in the colours in (d) against the general 'washed out' look of the colorization in (c). Colour blending is also apparent, especially in the green leaves (at the bottom and centre left) which have taken a red tinge from the pink petals and the red roses. Overall we have a sharper result and not an oversmoothed output as usually is the case for assuming a Gaussian prior.



(a) (b)

(c) (d)

Figure 9.17: Colorization example. Here we have a comparison of the visual quality produced by L^1 and L^2 optimization. (a) is an example gray image marked by a sparse set of coloured pixels arbitrarily placed; (b) the original colour image for reference; (c) shows colorization using L^2 optimization; (d) L^1 optimization. We observe a more accurate colorization in (d), e.g. the red balloon in the centre of the image is correctly colorized against the purple colorization in (c). We also observe more vibrant and sharper colours in (d) over (c).

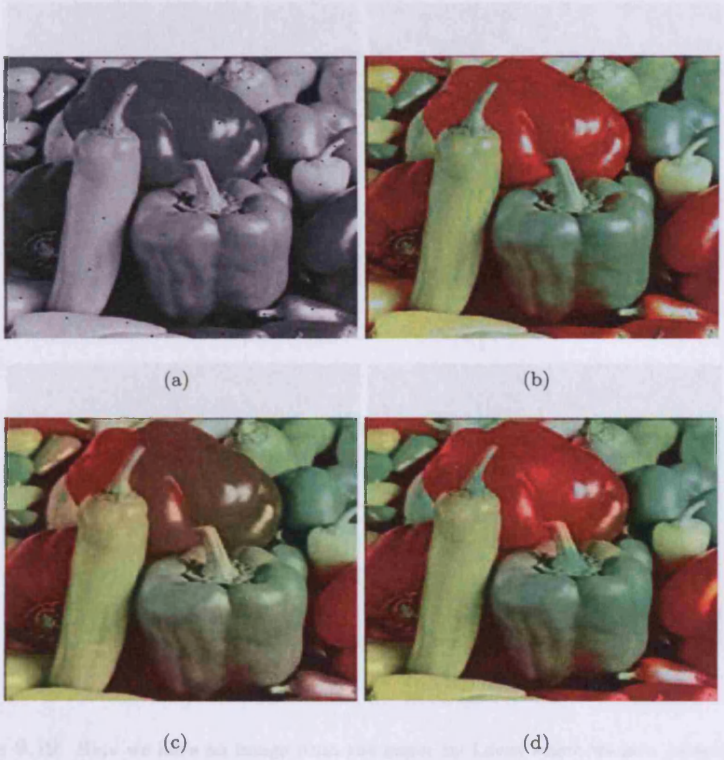


Figure 9.18: Here we compare the visual quality produced by L^1 and L^2 optimization. (a) is a gray image marked by a sparse set of coloured pixels arbitrarily placed; (b) the original colour image; (c) shows colorization using L^2 optimization; (d) L^1 optimization. We observe colour blending between the red and green in the L^2 approach whilst L^1 colorization creates an almost indistinguishable image from the original.

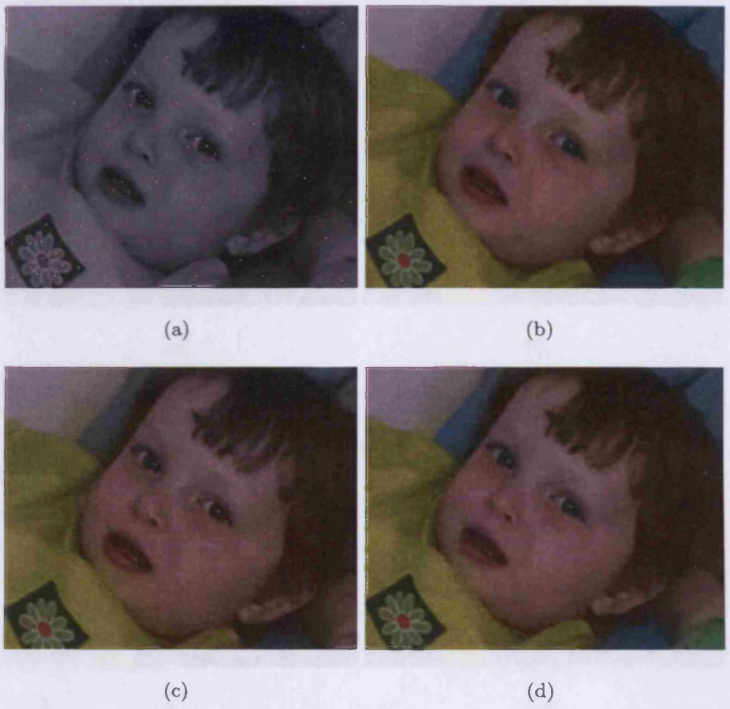


Figure 9.19: Here we have an image from the paper by Levin where we now colorize using a sparse set of marked pixels. (a) is the marked gray image; (b) the original colour image for reference; (c) shows colorization using L^2 optimization; (d) L^1 optimization. Again we observe an overall 'washed out' result using the L^2 approach against the sharper and more accurate result using L^1 minimisation. As examples we see that the cushion in the background has had its blue colour blended with the brown from the boys hair and the yellow from the t-shirt. The child's left eye is also incorrectly colorized brown instead of blue. There is an overall loss of colour vibrancy in the image.

(a) (b)

(c) (d)

Figure 9.20: Here we compare the visual quality produced by L^1 and L^2 optimization. (a) is a gray image marked by a sparse set of coloured pixels arbitrarily placed; (b) the original colour image for reference; (c) shows colorization using L^2 optimization; (d) L^1 optimization. We observe in particular that the blue feathers of the bird on the left have had their colours blended with the green and yellow, also the red feathers of the bird on the right exhibit much more colour vibrancy. This example illustrates the colour sharpness and vibrancy obtained using L^1 optimization over L^2.

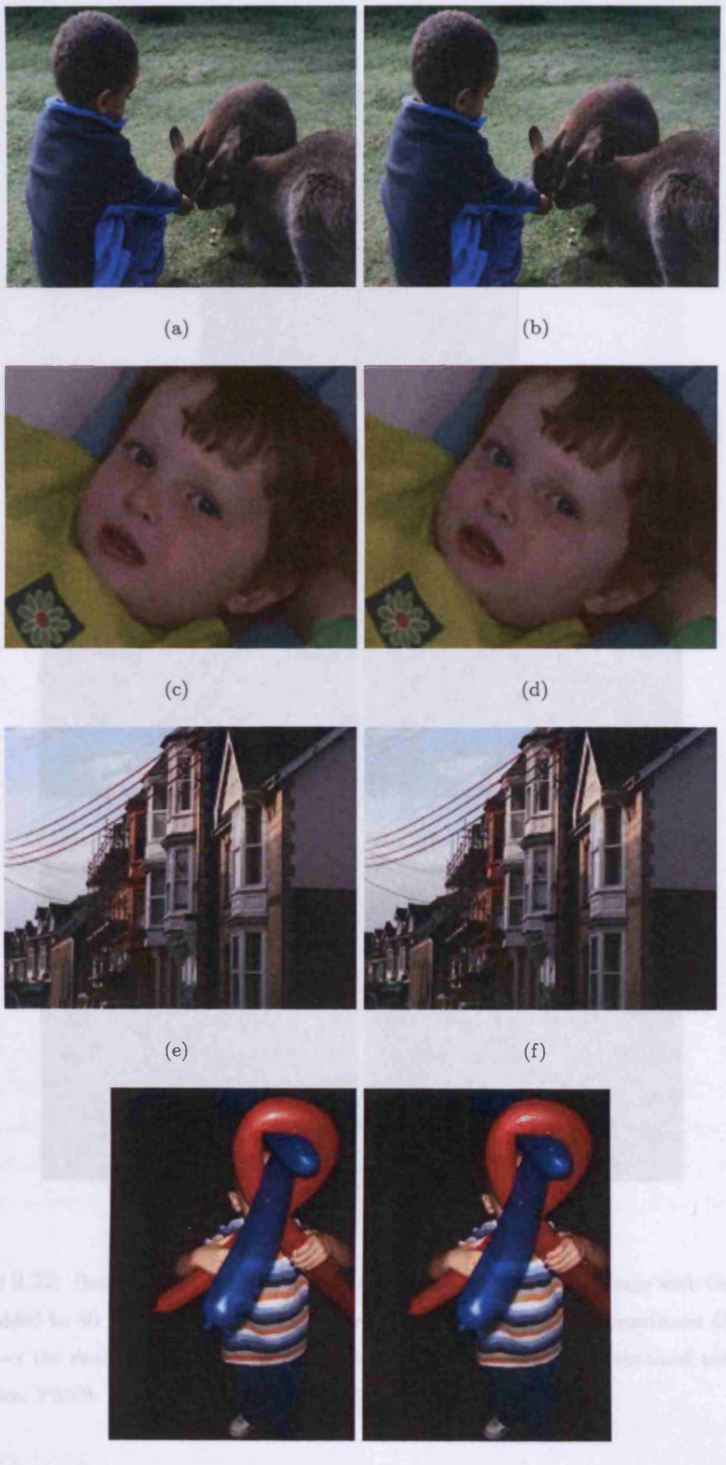


Figure 9.21: Chromacity channel compression. (a) the original image; (b) reconstruction from 5% of random seed measurements. (c), (e) and (g) original images; (d),(f) and (h) reconstruction from 5% of compressive sensing measurements. PSNR: (b) 31.81, (d) 38.91, (f) 31.87, (h) 32.76.

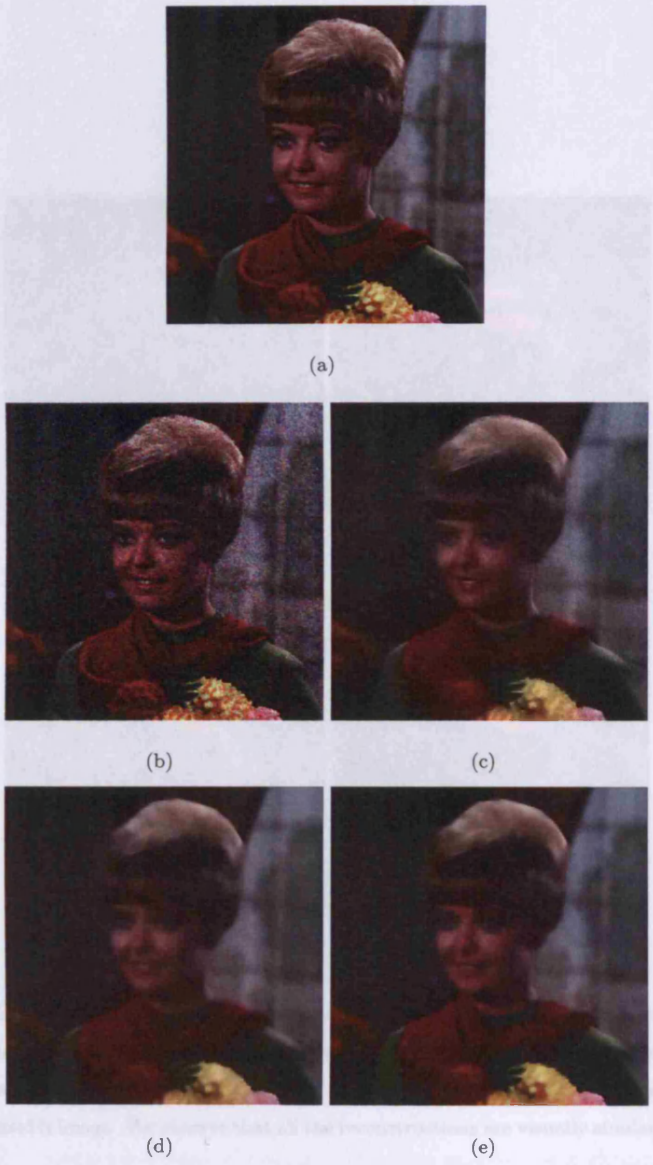
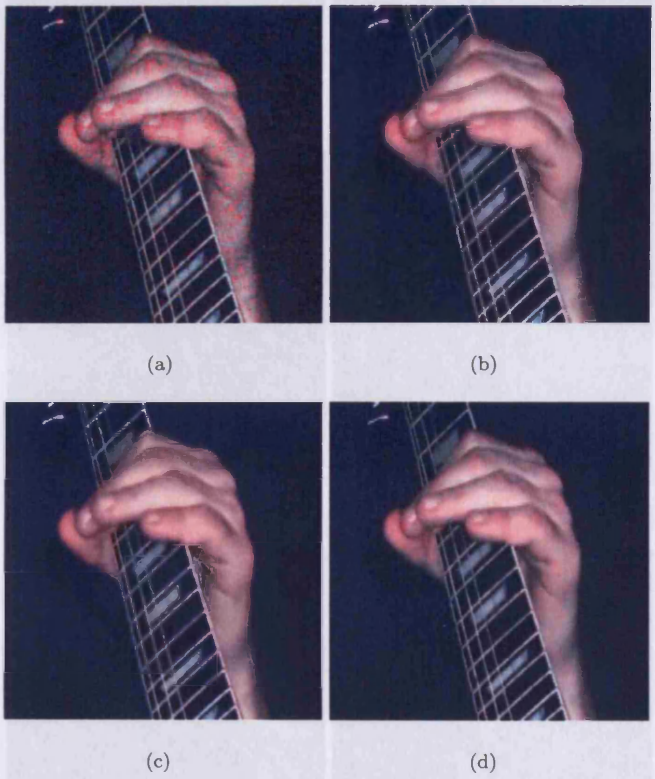


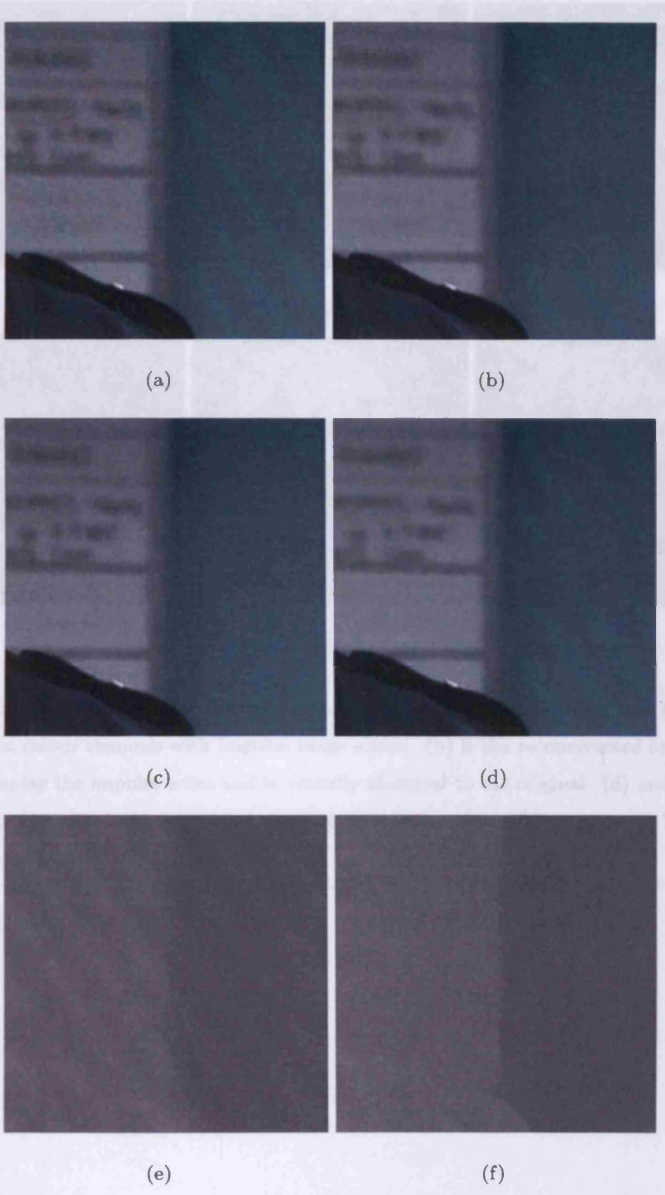
Figure 9.22: Denoising example. (a) shows the original image, (b) the image with Gaussian noise added to all *RGB* channels. (c) is the result using Neat Image at maximum filtering. (d) shows the denoising result using DenoiseMyImage. (e) is the result obtained using our algorithm. PSNR: (c) 26.69, (d) 26.35, (e) 27.20 ($\lambda = 5$)



(a) (b)

(c) (d)

Figure 9.23: Real image denoising example. (a) is an image that has been affected by severe chroma noise resulting in the appearance of 'blotches' of colour. (b) shows the denoised image obtained using Neat Image and (c) is obtained using our algorithm. (d) is the result obtained using DenoiseMyImage. We observe that all the reconstructions are visually similar. ($\lambda = 0.5$)



(a) (b)

(c) (d)

(e) (f)

Figure 9.24: Real image denoising example. (a) shows an example image affected by chroma noise that appears as bands in the colour channels. (b) is the result obtained using Neat Image which still leaves evident colour banding. (c) is our result which is able to remove the noise leaving a clean image as the colour banding does not correlate with the luminance structure. (d) is the best result obtained using DenoiseMyImage. (e) shows the banding still remaining in the V channel of the image when using Neat Image, while (f) clearly shows that the banding structure has been removed in our reconstructed V channel. ($\lambda = 0.1$)

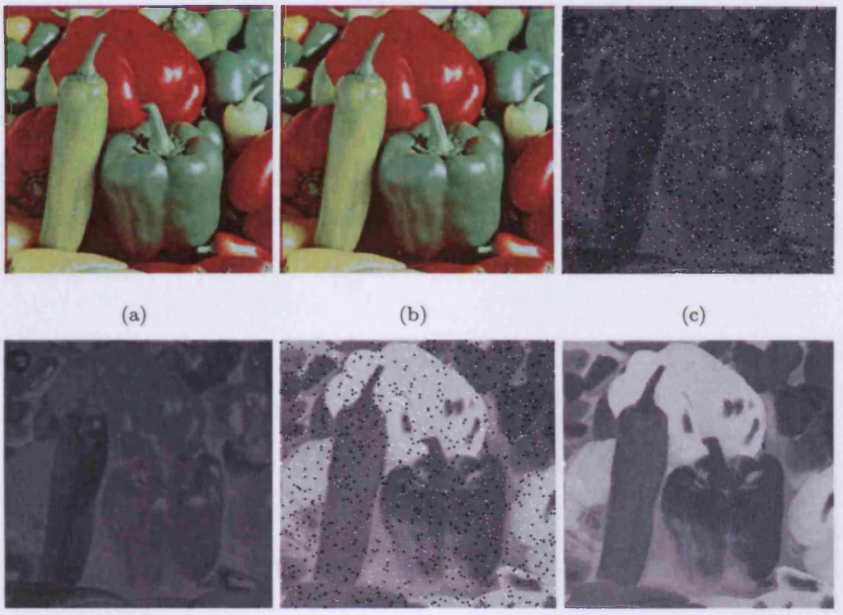


Figure 9.25: Impulse noise removal example. (a) shows the original image and (c) and (e) illustrate the colour channels with impulse noise added. (b) is the reconstructed image which does not display the impulse noise and is visually identical to the original. (d) and (f) shows the denoised chroma channels which have had their noise successfully removed. PSNR: (b) 37.68. ($\lambda = 0.5$)

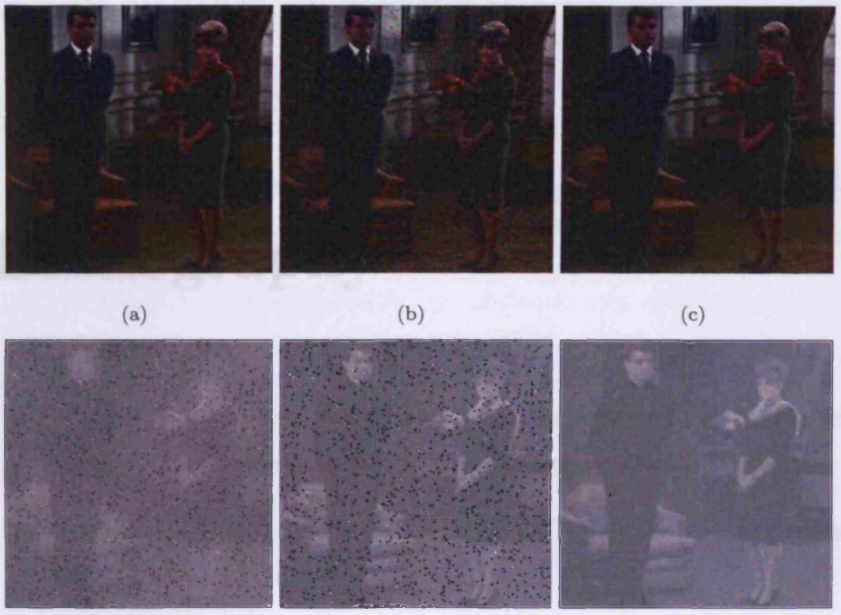


Figure 9.26: Impulse noise removal example. (a) shows an original colour image and (b) a noisy version that has had impulse noise added to the chroma channels in the YUV space. (c) is our reconstructed image which is virtually identical to the original. (d) is a typical result obtained using Neat Image or DenoiseMyImage. The impulse noise affecting the chroma is illustrated by (e) while the success of our algorithm for impulse removal is shown by (f). PSNR: (c) 42.20. ($\lambda = 0.5$)

Bibliography

[1] Atick, J. and Reedlich, A. (1992). Convergent algorithm for sensory receptive Field development. Neural Computation, 5, 45-60.

[2] Attneave, F. (1954). Informational aspects of visual perception. Psychological Review, 61, 183193.

[3] Almond, M. http://www.michaelalmond.com/Articles/noise.htm

[4] Balinsky, A., Balinsky, H., and Simske, S. On Helmholtz's principle for Document Processing, 10th ACM Symposium on Document Engineering (DocEng2010), Manchester, UK, 21-24 September 2010.

[5] Balinsky, A., and Mohammad, N. Colorization of natural images via L^1 optimization, Proc. WACV, 2009 IEEE Winter Vision Meetings, Snowbird, Utah, USA.

[6] Balinsky, A., and Mohammad, N. Non-linear filter response distributions of natural colour images. *LNCS* 5646, pp. 101-108. *Springer-Verlag Berlin Heidelberg* 2009.

[7] Balinsky, A., and Mohammad, N. Sparse natural image statistics and their applications to colorization and compression, Proc. International Conference on Image Processsing (ICIP), 2010, Hong Kong.

[8] Baddeley, R. (1996). Searching for filters with interesting output distributions: an uninteresting direction to explore? Network, 7, 2, 409-421.

[9] Barlow, H. (1961). Possible principles underlying the transformation of sensory messages. In Rosenblith, W., editor, Sensory Communication. MIT Press, Cambridge, MA.

[10] Baraniuk, R. Compressive sensing, IEEE Signal Processing Magazine, 25(2), pp. 21-30, March 2008.

[11] Bell, A. J., and Sejnowski, T. J. The independent components of natural scenes are edge filters, Vision Research, Volume 37, Issue 23, December 1997, Pages 3327-3338.

[12] Berinde, R., and Indyk, P. Sparse recovery using sparse random matrices, {preprint} (2008).

[13] Bergen, J.R. and Landy, M.S. Computational Modeling of Visual Texture Segregation, MIT Press: Cambridge, MA, 1991, pp. 253271.

[14] Borkowski, D. Chromaticity Denoising using Solution to the Skorokhod Problem, Image Processing Based on Partial Differential Equations, Proceedings of the International Conference on PDE-Based Image Processing and Related Inverse Problems, CMA, Oslo, August, 2005.

[15] Brooks, S., Saunders, I., and Dodgson, N. A. Image compression using sparse colour sampling combined with non-linear image processing, *Proc. of SPIE*, Vol.6492,pp. 64920F (2007).

[16] Bayer, B.E. and Powell, P.G. A method for the digital enhancement of unsharp, grainy photographic images, Advances in Computer Vision and Image Processing, Vol. 2, pp. 3188, 1986.

[17] Buades, A., Coll, B., and Morel, J. M. A review of image denoising algorithms, with a new one, Multiscale Modeling & Simulation, vol. 4, no. 2, pp. 490530, 2005.

[18] Burns, G. Colorization. Museum of Broadcast Communications: Encyclopedia of Television, http://www.museum.tv/archives/etv/index.html.

[19] Burr, D. C., and Morrone, M. C. Feature detection in biological and artificial visual systems, Vision: Coding and Efficiency ed. C Blakemore (Cambridge: Cambridge University Press).

[20] Burton, G.J. and Moorhead, I.R. Color and spatial structures in natural scenes, Applied Optics, Vol. 26, No. 1, pp. 157170, 1987.

[21] Candes, E.J., and Wakin, M.B. An Introduction To Compressive Sampling, IEEE Signal Processing Magazine, Volume: 25, Issue: 2, page(s): 21 - 30, ISSN: 1053-5888, March 2008.

[22] Candes, E. "Compressive Sampling," *Int. Congress of Mathematics,* 3, pp. 1433-1452, Madrid, Spain, 2006.

[23] Candes, E.J., Romberg, J., and Tao, T. Stable signal recovery from incomplete and inaccurate measurements, Comm. Pure Appl. Math., vol. 59, no. 8, pp. 12071223, 2006.

[24] Candes, E.J., and Tao, T. Near-optimal signal recovery from random projections: universal encoding strategies?, IEEE Trans. Inform. Theory, vol. 52, no. 12, pp. 54065245, Dec. 2006.

[25] Candes, E.J., Romberg, J., and Tao, T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inform. Theory, vol. 52, no. 2, pp. 489509, Feb. 2006.

[26] Candes, E.J., and Tao, T. (2005). The Dantzig selector: Statistical estimation when p is much larger than n. Ann. Statist. 35 23132351.

[27] Candes, E., Romberg, J. L1-magic: Recovery of Sparse Signals via Convex Programming, accompanying notes, October 2005.

[28] Carasso, A. Singluar integrals, image smoothness amd the recovery of texture in image deblurring. SIAM Journal Applied Mathematics, 64:1749-1774, 2004.

[29] Caselles, V., Coll, B., and Morel, J.M.: Geometry and color in natural images, J. of Math. Imaging and Vision, Volume 16, Issue 2 (March 2002), Pages: 89-105.

[30] Chan, T. F. and Esedoglu, S. Aspects of total variation regularized L^1 function approximation, SIAM J. Appl. Math., 65 (2005), pp. 18171837.

[31] Chan, F. T, Shen, J. Image Processing and Analysis Variational, PDE, Wavelet, and Stochastic Methods, 2005.

[32] Cheng, L., and Vishwanathan, S.V.N. Learning to compress images and videos, Proc. of the 24th Int. conference on Machine learning (2007), pp. 161-168.

[33] Chubb, C., Econopouly, J. and Landy, M.S. Histogram contrast analysis and the visual segregation of IID textures, J. Opt. Soc. Am. A, Vol. 11, pp. 23502374, 1994.

[34] Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space, ICIP 2007. (Matlab code available at www.cs.tut.fi/foi/GCF-BM3D).

[35] Daugman, J. G. Complete discrete 2D Gabor transforms by neural networks for image analysis and compression, IEEE Trans. Acoust., Speech, Signal Processing. 36 1169-79, 1998.

[36] Deriugin, N.G. The power spectrum and the correlation function of the television signal, Telecommunications, Vol. 1, pp. 112, 1956.

[37] DeVore, R. A., and Lucier, D. J. Classifying the smoothness of images. In Proceedings of IEEE International Conference of Image Processing (ICIP), pages 6-10. IEEE Press 1994.

[38] Dong, D. W., and Atick, J.J. Temporal decorrelation-a theory of lagged and nonlagged responses in the lateral geniculate nucleus, Network, 6, 159-178, 1995.

[39] Donoho, D. L., and Huo, X. Uncertainty principles and ideal atomic decomposition, IEEE Trans. Inform. Theory, vol. 47, no. 7, pp. 28452862, 2001.

[40] Donoho, D. Denoising by soft-thresholding, IEEE Trans. Info. Theory, Vol. 43, pp. 613627, 1995.

[41] Faugeras, O.D. and Pratt, W.K. Decorrelation methods of texture feature extraction, IEEE Pat. Anal. Mach. Intell. Vol. 2, No. 4, pp. 323332, 1980.

[42] Field, D. J. Relations between the Statistics of Natural Images and the Response Properties of Cortical Cells, J. Opt. Amer., Vol. 4, no. 12, pp. 2379-2394, 1987.

[43] Field, D. J. (1994). What is the goal of sensory coding? Neural Computation, 6, 559601.

[44] Foi, A., Katkovnik, V., and Egiazarian, V. Pointwise Shape-Adaptive DCT for High-Quality Denoising and Deblocking of Grayscale and Color Images, IEEE Trans. Image Process., vol. 16, no. 5, May 2007.

[45] Forsyth, D., and Ponce, J. Computer vision. Prentice Hall, Upper Saddle River, NJ, 2002.

[46] Frenkel, G., Katzav, E., Schwartz, M., and Sochen, N. Distribution of anomalous exponents of natural images. Physical Review Letters, 2006 Sep 8;97(10):103902.

[47] Geman, S., and Geman, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Analysis and Machine Intelligence 6: 721-741. 1984.

[48] Gidas, B., Mumford, D. Stochastic models for generic images. Qaurterly of Applied Mathematics, 59:85-111, 2001.

[49] Gonzalez, R.C., Woods, R.E. Digital Image Processing (2nd Edition), Chapter 5, Image restoration.

[50] Grant, M. and Boyd, S. http://cvxr.com/cvx/

[51] Grenander, U., and Srivastava, A. Probability models for clutter in natural images. IEEE transactions PAMI, 23:424-429, 2001.

[52] Heeger, D.J. and Bergen, J.R. Pyramid-based texture analysis/synthesis, in Proceedings of SIGGRAPHS, 1995, pp. 229238.

[53] Horiuchi, T., and Tominaga, S. Color image coding by colorization approach, J. on Image and Video Proceess., Vol. 2008 , Issue 2 (Feb. 2008), Article No. 18.

[54] Ibn Al-Haytham, Kitab al-Manzir, c.1000. In Arabic translated by A.Sabra, The Optics of Ibn Al-Haytham, London: The Warburg Institute, 1989.

[55] Irony, R., Cohen-Or, D., and Lischinski, D. 2005 Colorization by example. In Rendering Techniques 2005, IEEE Computer Society Press, 201-210.

[56] Joshi, N., Zitnick, C. L., Szeliski, R., and Kriegman, D.J. Image deblurring and denoising using color priors. Computer Vision and Pattern Recognition, 2009. CVPR 2009. pages 1550 - 1557.

[57] Kretzmer, E. (1952). Statistics of television signals. Bell Systems Technical Journal, 31, 751763.

[58] Kersten, D. (1987). Predictability and redundancy of natural images. J. Optical Soc. of America A., 4, 23952400.

[59] Pattern theory, a unifying perspective. In D. Knill and W. Richards, editors, Perception as Bayesian Inference. Cambridge Iniversity Press, Cambridge, UK, 1996.

[60] Huang, J., and Mumford, D. Statistics of Natural Images and Models, *Proc. IEEE CVPR*, vol. 1, Fort Collins, CO, pp. 541-547 (1999).

[61] van Hateren, J.H. and Schaaf, A. van der. Independent component filters of natural images compared with simple cells in primary visual cortex, Proc. Royal Society of London, B 265:359-366, 1998.

[62] Hubbel, D. and Wiesel, T. (1959). Receptive fields of single neurons in the cat's striate cortex. Journal of Physiology, 148, 575-591.

[63] Levin, A., Lischinski, D., Weiss, Y. Colorization using optimization. *ACM Transactions on Graphics,* Volume 23, Issue 3, pp. 689694, 2004.

[64] Lee, A. B., Mumford, D., and Huang, J. Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model, International Journal of Computer Vision archive, Volume 41 , Issue 1-2, Pages: 35 - 59, 2001.

[65] Linsker, R. Sensory processing and information theory. In Grassberger, P. and Naadal, J.P. (Eds), From statistical physics to statistical inference and back, pp 237-247. Dordrecht: Kluwer.

[66] Levin, A., Fergus, R., Durand, F., and Freeman, W.T. Image and Depth from a Conventional Camera with a Coded Aperture, ACM Transactions on Graphics, SIGGRAPH 2007 Conference Proceedings, San Diego, 2007.

[67] Leporini, D. Pesquet, J.-C. and Krim, H. Best basis representations with prior statistical models, in Lecture Notes in Statistics: Bayesian Inferences in Wavelet Based Models, P. Muller and B. Vidakovic (Eds.), Springer Verlag: Berlin, 1999.

[68] Linsker, R. 1986, Proceedings of the National Academy of Sciences, USA, 83:7508-7512.

[69] Li, Y., Lizhuang, M., and Di, W. Fast Colorization Using Edge and Gradient Constrains, *Proc. of WSCG'07,* pp. 309-315.

[70] Liu, C., Szeliski, R., Kang, S.B., Zitnick, C.L., Freeman, W.T. Automatic Estimation and Removal of Noise from a Single Image, IEEE Transactions on pattern analysis and machine intelligence, vol. 30, NO. 2, Feb 2008

[71] Mairal, J., Elad, M., and Sapiro, G. Sparse representation for color image restoration, IEEE Transactions of image processing, vol. 17, No. 1, January 2008.

[72] Mallat, S.G. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation, IEEE Trans. Pattern Anal. Machine Intell., vol 11, pp. 674-693, July 1989.

[73] Mumford, D. Pattern theory: the mathematics of perception. In proc. Int. Congress of Math., Beijing, China, 2002. Higher Education Press, Beijing.

[74] Mumford, D. Empirical Statistics and Stochastic Models for Visual Signals, New Directions in Statistical Signal Processing From Systems to Brains, MIT Press, 2006.

[75] Mumford, D. Emperical Investigations into the Statistics of Clutter and the Mathematical Models it Leads To, Lecture for the Review of ARO Metric Pattern Theory Collaborative, Brown Univ., Providence, RI, 2000.

[76] Mumford, D., Shah, J. Optimal approximations of piecewise smooth functions and associated variational problems. Communications in Pure and Applied Mathematics, 42:577-685, 1989.

[77] Moulin, P. and Liu, J. Analysis of multiresolution image denoising schemes using a generalized Gaussian and complexity priors, IEEE Trans. Info. Theory, Vol. 45, pp. 909919, 1999.

[78] NeuralTek, 2003. BlackMagic photo colorization software, version 2.8. http://www.timebrush.com/blackmagic.

[79] Olshausen, B.A., and Field, D.J.: Natural Image Statistics and Efficient Coding, New.: Computation Neural syst., vol. 7, 1996.

[80] Olshausen, B. A., and Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images, Nature 381, 607609, 1996.

[81] Portilla, J. and Simoncelli, E.P. A parametric texture model based on join statistics of complex wavelet coefficients, International Journal of Computer Vision, Vol. 40, No. 1, pp. 4970, 2000.

[82] Qing, L., Wen, F., Cohen-Or, D., Liang, L., Xu, Y., and Shum, H. Natural image colorization. In Proc. Eurographics Symposium on Rendering (2007).

[83] Romberg, J. Imaging via Compressive Sampling, IEEE Signal processing magazine, Volume: 25, Issue: 2, page(s): 14 - 20 ISSN: 1053-5888, March 2008.

[84] Ruderman, D. L. Natural ensembles and sensory signal processing, PhD Thesis, University of California, Berkeley.

[85] Ruderman, D.L. The statistics of natural images, Network, Vol. 5, pp. 517548, 1994.

[86] Ruderman, D.L. Origins of scaling in natural images, Vision Research, Vol. 37, No. 23, pp. 33853398, 1997.

[87] Ruderman, D.L., and Bialek, W. Scaling of natural images: Scaling in the woods, Physical Review Letters, Vol. 73, No. 6, pp. 814-817, 1994.

[88] Shannon, C. (1949). Communicating in the presence of noise. Proc. IRE, 37, 1021.

[89] Silberg, J. 1998. The Pleasantville post production team that focussed on the absence of color. Cinesite Press Article, http://www.cinesite.com/core/press/articles/1998/10 00 98-team.html.

[90] Simoncelli, E.P., and Adelson, E.H. Noise removal via Bayesian wavelet coring, in Third Intl. Conf on Image Proc., Lausanne, IEEE Sig Proc Society, 1996, Vol. I, pp. 379382.

[91] Srivastava, A., Lee, A. B. and Simoncelli, E. P., and Zhu, S.C. On Advances in Statistical Modeling of Natural Images, Journal of Mathematical Imaging and Vision archive, Volume 18, Pages: 17 - 33, 2003.

[92] Srivastava, A. Stochastic Models for Capturing Image Variability, IEEE Signal Procesing Magazine, Volume: 19, Issue: 5 On page(s): 63-76, 2002.

[93] Sullivan, J. A Bayesian Framework for Object Localisation in Visual Images, PhD Thesis, University of Oxford, 2000.

[94] Tappen, M. F., Russel, B. C., and Freeman, W. T. Exploiting the sparse derivative prior for super-resolution and image demosaicing, Proc. IEEE Workshop on Statistical and Computational Theories of Vision, 2003.

[95] Tolhurst, D., Tadmoor, Y., and Chao, T. (1992). Amplitude spectra of natural images. Opthalm. Physiol. Opt., 12, 229-232.

[96] Tropp, J. Just relax: Convex programming methods for identifying sparse signals in noise, IEEE Trans. Inform. Theory, vol. 52, no. 3, pp. 10301051, 2006.

[97] Torralba, A., and Oliva, A. Statistics of natural image categories, Network. 2003 Aug;14(3):391-412.

[98] van der Schaaf, A. (1998). Natural Image Statistics and Visual Processing. PhD thesis, Groningen University.

[99] Wainwright, M.J., Simoncelli, E.P., Willsky, A.S.: Random Cascades on Wavelet Trees and their use in Analysing and Modelling Natural Images, Appl. Comput. Harman, Anal., Vol. 11, pp, 89-123, 2001.

[100] Wainwright, M.J., and Simoncelli, E.P. Scale mixtures of Gaussians and the statistics of natural images, Advances in Neural Information Processing Systems, S.A. Solla, T.K. Leen, and K.-R. Muller (Eds.), 2000, pp. 855861.

[101] Welsh, T., Ashikhmin, M., and Mueller, K. 2002. Transferring color to greyscale images. ACM Transactions on Graphics 21, 3 (July), 277.280.

[102] Wu, Y., Zhu, S.C., Guo, C. E. From information scaling laws of natural images to regimes of statistical models, Quarterly of Applied Mathematics, 2007.

[103] Yatziv, L., and Sapiro, G. Fast image and video colorization using chrominance blending. IEEE Trans. on Image Processing, Vol. 15, No. 5, pp. 1120-1129 (May 2006).

[104] Zhang. Y. Solving large-scale linear programs by interior-point methods under the MATLAB environment. Optim. Methods Software 10, 1-31 (1998).

[105] Zhu, S.C. and Mumford, D. Prior learning and Gibbs reaction diffusion, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 19, No. 11, pp. 12361250, 1997.

[106] Zhu, S.C., Wu, Y.N. and Mumford, D. Minimax entropy principles and its application to texture modeling, Neural Computation, Vol. 9, No. 8, pp. 16271660, 1997.

Short details

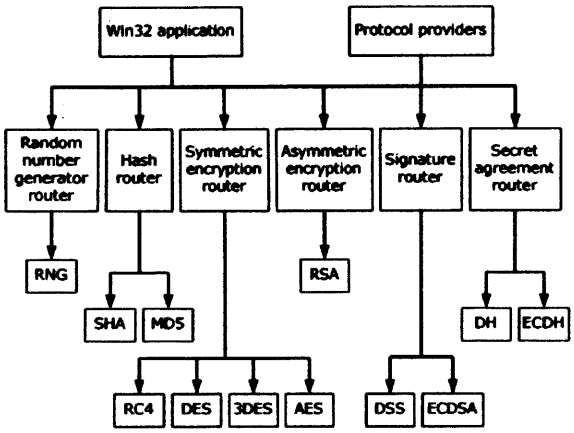
- AES operates on a 4◆4 array of bytes.
- For encryption, each round of AES (except the last round) consists of four stages: AddRoundKey, Subbytes, Shift rows and Mix columns.
- At each stage, the bytes are manipulated and processed for the next level.

Crytography API: Next Generation(CNG)

CNG provides a set of APIs that are used for performing basic cryptographic operations, such as creating hashes, encrypting, and decrypting data.

Each algorithm class in **CNG** is represented by a primitive router. Applications making use of the primitive APIs will link to the router binary (*Bcrypt.dll* in user mode, or *Ksecdd.sys* in kernel mode), and make calls to the various **CNG** primitive functions. All of the algorithm primitives are managed by various router components. These routers keep track of each algorithm implementation that has been installed on the system. The router will route each function call to the appropriate primitive provider module.

The following illustration shows the design and function of the **CNG** cryptographic primitives.



CNG provides primitives for the following classes of algorithms:

- Random Number Generator: This class is used to represent pluggable random number generation (RNG).
- Hashing: This class represents algorithms used for hashing, such as SHA1 and SHA2.
- Symmetric encryption: This class represents algorithms used for symmetric encryption. Some examples are AES, 3DES, and RC4.
- Asymmetric encryption: This class represents asymmetric (public key) algorithms that support encryption, like RSA.
- Signature: This class represents signature algorithms such as DSA and ECDSA. This class can also be used with RSA.
- Secret Agreement: This class represents secret agreement algorithms such as Diffie-Hellman (DH) and elliptical curve Diffie-Hellman (ECDH).

Using the code

Using a RSA CryptoService Provider (CAPI)

In CAPI, all cryptographic algorithms are predefined in *wincrypt.h* which makes it very difficult to extend cryptographic functionality to suit your application's need. Adding a custom symmetric algorithm is not easy. Secondly the CAPI requires **Microsoft** to sign the implementation, so that it can be a part of a security namespace.

`Encrypt` and `Decrypt` in a traditional way, with `RSACryptoServiceProvider`.

⊟ Collapse | Copy Code

```
RSACryptoServiceProvider MyAsymmetricAlgorithm = new RSACryptoServiceProvider();
byte[] PlainTextBytes;
byte[] CipherTextBytes;

private void Encrypt()
{
    PlainTextBytes = System.Text.Encoding.UTF8.GetBytes(TextBoxOriginal.Text);
    CipherTextBytes = MyAsymmetricAlgorithm.Encrypt(PlainTextBytes, true);
    TextBoxEncrypted.Text = TextBoxEncrypted.Text +
    + Convert.ToBase64String(CipherTextBytes);
              ShowPublicPrivate();
    // rest of the code removed for brevity
}

private void Decrypt()
{
    PlainTextBytes = MyAsymmetricAlgorithm.Decrypt(CipherTextBytes, true);
    TextBoxOriginal.Text = System.Text.Encoding.UTF8.GetString
    (PlainTextBytes);
}

private void ShowPublicPrivate()
{
   RSAParameters MyParameters = new RSAParameters();
   MyParameters = MyAsymmetricAlgorithm.ExportParameters(true);
   TextBoxPrivateKey.Text = Convert.ToBase64String(MyParameters.D);
   TextBoxPublicKey.Text = Convert.ToBase64String(MyParameters.Modulus);
```