Exploring the Visual Pathway and Its Applications to Image Reconstruction, Contrast Enhancement and Object Recognition

Asim Ali Khwaja

December 2010

A THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY OF
THE AUSTRALIAN NATIONAL UNIVERSITY

THE AUSTRALIAN NATIONAL UNIVERSITY

School of Engineering
College of Engineering & Computer Science

"Through dangers untold and hardships un-numbered, I have fought my way here to the castle beyond the goblin city ..."

(From the movie: Labyrinth)

This thesis is dedicated to my family who contributed in, various ways, not only to my PhD but to me! I don't know what I would have done without them.

To my parents,
Who made me what I am. They are the light of my life - here and hereafter.

To my aunt,
Who would always become restless on my sufferings and would see no limits for my comfort. She is no longer alive to see me finish. Oh, I miss you so much!

To my sister,
Who is like a mother to me. I still remember her little bed-time stories that she used to tell when I was a kid and console me when I would cry.

To my brother,
Who is like a father to me. He has always been a model and a source of inspiration in all respects.

To my wife,
Whose love I treasure very much. She has always been a very understanding wife. She is the best anyone can ever have!

To my kids,
Who are too young to understand what this is all about but pretend that they do when explained why dad can't be with them most of the time. They are the joy of my life as much as they drive me crazy!

Declaration

This thesis describes the results of research undertaken in the School of Engineering at the Australian National University, Canberra. This research was funded by scholarships primarily from National University of Science and Technology (NUST) and partly from Australian National University (ANU). The results and analyses presented in this thesis are my own original work, accomplished under the supervision of Dr. Roland Goecke except where otherwise stated. This thesis has not been submitted for any other degree.

Asim Ali Khwaja
School of Engineering
The Australian National University
21 December 2010

Acknowledgements

I am indebted to my supervisor and mentor, Dr. Roland Goecke for all his help, guidance and suggestions that resulted in the completion of this dissertation research. He was exceptionally approachable and utterly supportive and accommodating. I am particularly grateful to Dr. Ted Maddess of the Research School of Biological Sciences (RSBS), ANU for the invaluable and motivating, hour-long discussions I had with him on retina, visual cortex and contrast enhancement. I would like to thank Dr. Michael Ibbottson of RSBS for letting me sit in his neuroscience course which proved extremely beneficial. I am also thankful to Dr. Patrick Cavanagh of the Psychology Department of Harward University for putting the video lectures of his course 'Vision and Brain' online. They provided a turning point in my understanding of cognitive processes. Special thanks to my dear friend Manfred Doudar for his continuous encouragement, and help in reviewing a part of the thesis. Special thanks is due to James Ashton, IT Manager at Research School of Information Sciences and Engineering, who was very supportive and tolerant of my heavy downloads throughout the term and for the countless number of times that I have walked into his office for issues big and small, he never raised an eyebrow - always eager to help. I feel indebted to my wife for her excellent support throughout, and for bearing the burden of the household by herself while I was busy in study. I am also thankful to my brother for his moral and financial support. Finally, I could not have accomplished such a task without the love and prayers of my parents and much-missed aunt.

Abstract

The natural world is filled with perfectly working, functional systems that are robust, accurate, and adaptable; this work takes favour with the aforesaid and presents a biologically inspired approach to computer vision; in particular on the subjects of image reconstruction, contrast enhancement and object recognition.

The first half of this thesis takes an exploratory approach, on the example of image reconstruction, towards the understanding of the visual pathway from retina to the primary visual cortex (V1), investigating redundancy reduction, information preservation and contrast enhancement. The retina having approximately 130 million cells, is forced to discriminate with the incoming information. Programmed for concision, the primate eye encodes information with sparsity, yet remains information preserving by encoding only contrast.

By reconstructing an image from its contrast map pairs using gradient descent least squares error minimization, this work has shown that information is preserved across the optic nerve channel despite sparsification of the input image presented on the photoreceptors. By mimicking the irregularities of the eye's receptive fields, it has been shown that the neural architecture along the visual pathway is robust and fault tolerant against irregularities - a general characteristic of the entire nervous system.

Using non-linear and asymmetric gain control with the on- and off- centre contrast map pairs, it has been shown that the mean luminance of an image can be controlled and the aforesaid reconstruction can be used for straightforward

enhancement; thus reducing contrast enhancement to a scaling operation over the contrast domain. This has further been successfully applied to colour image contrast enhancement using a number of different models, including the neuro-physiologically proven representation of colour opponency, in the form of colour opponent contrast maps.

With the above work serving as a pre-processing stage, the second half of the thesis approaches the subject of object recognition; improving upon prior work in Sparse Representation Classification (SRC). Sparseness is a key feature of the brain's internal representation whereby it achieves its robustness and adaptability. This work replaces the mean square error measure for similarity comparison of images with a perceptually compatible structural error measure, as well as the conventional sparsifiers with a genetic algorithm of the original SRC algorithm. This has resulted in an improved recognition rate - owed in large part to a more effective similarity comparison and improved sparseness of the solution.

The approaches to the troika of reconstruction, contrast enhancement and object recognition strengthen both premise and belief that biologically inspired vision is dually meritorious and warrants greater appreciation and study by the Computer Vision community at large; not to be discounted as is often done. The hope is this work proves a seed for future endeavours.

Publications

- A. Khwaja, A. Asthana and R. Goecke. **Illumination and Expression Invariant Face Recognition Using SSIM Based Sparse Representation.** *Proceedings of the 2010 IEEE International Conference on Pattern Recognition ICPR2010*, Istanbul, Turkey, 23-26 Aug. 2010. IEEE.

- A. Khwaja and R. Goecke. **Biologically Inspired Contrast Enhancement Using Asymmetric Gain Control.** *Proceedings of the International Conference on Digital Image Computing: Techniques and Applications DICTA2009*, Melbourne, Australia, 1-3 Dec. 2009. APRS.

- A. Asthana, A. Khwaja and R. Goecke. **Automatic Frontal Face Annotation and AAM Building for Arbitrary Expressions from a Single Frontal Image Only.** *Proceedings of the 2009 IEEE International Conference on Image Processing ICIP2009*, Cairo, Egypt, 7-11 Nov. 2009. IEEE.

- A. Khwaja and R. Goecke. **Image Reconstruction from Contrast Information.** In *Digital Image Computing: Techniques and Applications (DICTA 2008)*, Canberra, Australia, 1-3 December 2008. APRS.

Abbreviations

AHE	Adaptive Histogram Equalization
CAT	Computer Aided Tomography
CLAHE	Contrast Limited Adaptive Histogram Equalization
DoG	Difference of Gaussians
fMRI	Functional Magnetic Resonance Imaging
GA	Genetic Algorithm
HDR	High Dynamic Range
HE	Histogram Equalization
JGACE	JND-Guided Adaptive Contrast Enhancement
JND	Just Noticeable Difference
L*a*b*	A colour space. L for lightness and a and b for colour-opponent dimensions
LDR	Low Dynamic Range
LGN	Lateral Geniculate Nucleus
MSE	Mean-Square Error
NP	Non Polynomial
PSF	Point Spread Function
PSNR	Peak Signal-to-Noise Ratio
RGB	Red Green Blue
RMS	Root Mean Square
SpaRSA	Sparse Reconstruction by Separable Approximation
SparsGA	Genetic Algorithm based Sparsifier

SRC	Sparse Representation Classification
SSIM	Structural Similarity
V1	Primary Visual Cortex
2D	Two Dimensional

Contents

List of Figures

1

Introduction

The serenity and beauty of a lake-side scene, the intricacies of design patterns on the wings of a butterfly, the delicious taste of a chocolate sundae, the vibrant fragrance of a perfume and the mesmerizing touch of a beloved. The vast richness of all such sensory experiences is possible only through a small piece of organic tissue - the brain. Among all the senses, the neural processing in the brain is strongly influenced by vision.

The biological visual system is responsible for two main tasks: first, to identify objects by forming invariant representations in the presence of convoluted changes in light intensities, distance, orientation and position; and second, to guide complex interactions with the environment using the aforesaid representations.

From time immemorial, humans have focussed on making tools that would facilitate life, machines that could work alongside humans. This requires a key attribute in them - that they are able to *interact* with their environment. Such an interaction is utterly complex, for a human environment is fraught with nature that is imperfect, noisy, and ambiguous. On the other hand, machines are designed to operate within a controlled environment underpinned by their own limitations and the constraints of the environment. In order for them to work in an unstructured environment, they need to be able to handle information through

the same modalities as humans do.

1.1 Research Motivation

The field of Computer Vision has developed rather rapidly in the last two decades. Two distinct research trends can be observed here: one that builds upon models of signal processing, also known as the conventional or traditional approach, and the other one inspired by nature, that tries to take licence from the animal nervous system; known as the neuro-physiological approach, or the biological approach in general. While the conventional Computer Vision methods have been largely successful in a variety of applications in a lab or controlled real world environment; they have failed miserably outside.

Consider face recognition - currently a topic of sustained focus in the Computer Vision community, owing largely to its immediate applicability to security and surveillance applications. Although many algorithms and techniques have been developed for recognizing human faces [Zhao et al., 2003] and while some of them have achieved good results in the laboratory (in some instances above 95% accuracy), these very same algorithms do not perform nearly as well on real-world datasets, falling in accuracy as low as 60% or less. Since animals, and humans in particular, excel in object and face recognition tasks, it was thought to model vision along patterns discovered in brain. There were tremendous modelling difficulties due to lack of understanding, and any systems designed on such claims as modelling the animal vision faced an even greater failure resulting in a heavy criticism and disbelief on the robustness of any such methodology, which prevails over a large mindsets to this day.

1.1.1 Failure of Face Recognition Systems on Real-World Data

As face recognition began gaining attention under exaggerated claims by industry, these algorithms failed to deliver expected results on real world data. The Tampa face-recognition system [Bowman, 2003] commissioned by the Police Department (Tampa Florida), and installed in 2001 to monitor criminal activity in one of its districts, was taken down in 2003 for the fear of false identifications, saying that the system never identified, alerted to, or caught any criminal. Two separate face-recognition systems installed in 2002 at Boston's Logan Airport [Willing, 2003, Leyden, 2003] gave disappointing performance of 61.4% in detecting volunteers who acted as potential terrorists as they passed security checkpoints during a three-month test period. As a result, the airport decided to explore other technologies for securing its checkpoints. Furthermore, the systems also proved to be absurdly ineffective for crowd surveillance [Greene, 2001, Feder, 2004]. It was only in very constrained and structured environments that the systems have shown some use. For instance in personal notebook logins, where the lighting is controlled, the subject is under instruction, and the image database is very small. These failures indicate that the technology is far behind what nature achieves. With recent advances in the understanding of the primate and human visual system, it seems to be a good point in time to investigate biological approaches again in more detail.

1.1.2 Arguments Against Biological Approaches

Successful engineering applications are often cited as arguments against biological approaches: aeroplanes do not flap their wings and cars do not run on legs. However, the basic principle of aeroplanes was indeed inspired by birds when

it was observed how birds take flight by initially flying close to earth as they gain speed and air pressure. In fact, humans would likely never have thought about flying if it were not for them. And although cars do not use legs, and are successful and efficient for most applications, their wheels give up on rough terrains, cluttered environments, and, most commonly, stairs. Thus, when it comes to unstructured environments, natural systems always have an edge over artificial ones and the deficiencies and failures in such systems are due to our ignorance and misunderstandings of nature while the designs of nature stand ever more elegant and triumphant.

► *1.1.3 The Rise and Fall of the Biologically-Inspired Approaches*

The nervous system has always been a source of inspiration with several attempts made to create systems, visual or otherwise, that were built on the working principles of the brain. While artificial systems have spurred much interests, and have been a focus of considerable research, none have been convincingly useful in real applications and seldom went beyond academia. *Artificial Neural Networks* were only one such endeavour that initially caused a lot of excitement in the 1980s and 1990s [Crick, 1989]. It was later realized that the highly simplified abstractions that these models were based on cannot take us any further into understanding and replicating the functionality of the brain.

A reason often cited for the collapse of the biological approach is that, historically, computing power has lagged behind the ability to simulate large neuronal processes to aid in unveiling the mysterious workings of the brain; yet despite the decrease in cost and increase in computing power, the world has yet to still witness progress in this regard.

Perhaps the biggest and possibly the only reason of such failures is a lag in

the neuroscience research itself. One of the primary challenges in any research that involves neurophysiologically-inspired techniques is a heavy dependence on the progress of the subject itself. Much about the brain remains unknown; which leads to the hypothesizing of the functionality of its different regions. Olshausen & Field [Olshausen and Field, 2004] write that:

> "The primary visual cortex of mammals has been the subject of intense study for at least four decades.... Given the magnitude of these efforts, one might reasonably expect that we would by now have a fairly concrete grasp of how V1 works and its role in visual system function. However, ..., there still remains so much unknown that, for all practical purposes, we stand today in nearly the same state of ignorance as did Hubel and Wiesel 40 years ago."

And as Sur of MIT [Sur, 2003] says in one of his lectures:

> "We have learned more about the brain in the last 20 years than in all of previous human history... (but) if we have a stroke, neurologists can do little more than hold our hands even today."

The reason neurophysiologically-based approaches have not been able to yield any successful Computer Vision applications is not that they are not practical for artificial systems, but because we are not properly equipped with the knowledge of how the visual cortex works. While we have gained much insight into the workings of the visual cortex, to paraphrase Olshausen, these are but just bits and pieces - drops from an ocean.

The extent of our progress can also be understood by observing that the visual cortex is the most explored area of the brain. From within the visual cortex, the primary visual cortex, also known as area V1, remains the most researched of all

visual regions of the brain. And yet we are not at all sure what the neurons in that area are actually doing. In fact, our knowledge about the retina remains insufficient.

1.1.4 Reasons for Lag in Neuroscience Research

The following reasons may be noted to primarily account for the current lag in neuroscience research:

Technological Difficulties

Firstly, experimental neuroscience is undoubtedly in itself very complex. Recording from brain cells is not an easy task. Only certain cortical areas of the brain can be reached safely, while keeping the animal alive on the surgical table. Even with today's technology, our access to the brain is all but extremely limited. As this is the organ that keeps the whole body going, a little carelessness is enough to kill the animal under experiment. CAT scans and fMRI scans have helped a lot to reveal the mysterious operation of the brain but the restricted physical access is proving to be a real obstacle. In particular, the mid and the hind brain are very obscured and since these control the life support functions, any damage to the neurons in there may seriously hamper with some major body function like respiration or the circulatory system immediately killing the animal and hence terminating the experiment.

Another technological complicatedness is in simultaneous recording from multiple neurons. It is very hard to make multiple simultaneous measurements. The trend pursued even today is single cell recordings where a neuron is picked out from the crowd, almost randomly, and its response observed. However, neurons do not work alone. Each is taking input from thousands of other neurons and sending output to a similar number of others with intricate feedback paths. Besides, the brain works as a distributed system, and to understand how the whole

system works, we need to study interactions between activity in different areas. With recent technology, however, multi-electrode experiments are gradually becoming popular.

Ethical Issues

Animals are and will continue to form an essential part of medical research and testing, however, ethical considerations place substantial limits on the ways, in which they can be used in experiments. Up until now only anesthetized animals were allowed to be probed. With the development of new technologies, awake animals are gradually being permitted to be used in certain experiments, but the matter is still controversial and many institutes are hesitant in granting such permissions. It is strongly believed that the neuronal activations recorded in an anesthetized animal will differ vastly from those in a conscious animal shedding new light and opening up fresh avenues of understanding.

Research Attitude

Olshausen and Field [Olshausen and Field, 2004] identify five problems with the current research attitude:

1. *Biased sampling of neurons:* The neurons selected for single-cell recordings are usually chosen with a certain bias. Typical biases include picking neurons that fire more strongly than the rest, whose activation pattern appears more 'rational' to the investigator, neurons with bigger cell bodies etc. This approach leave out other cells, for example those that do not fire as expected and, hence, one particular view picks up.

2. *Biased stimuli:* The general trend is to use highly simplified stimuli approximating the neurons to be linear systems. Stimuli like dots, bars, gratings are very popularly used in the hope that the neuronal response to these

could be generalized to more complex situations like natural scenes. However this view has been shown to be incorrect as neurons are not linear devices.

3. *Biased theories:* The theories formed about the visual system are primarily based on the data collected, when they should have taken into account the problems that the visual system was designed to solve. Categorization of cells like simple, complex and end-stopped were created out of the particular stimuli that were used. Hence, these theories stand shaken in front of the conflicting observations that are often not appropriately reported. Similarly based on such cell classifications, the view created of the V1 as a bar and edge detector is also not well-founded.

4. *Interdependence and contextual effects:* Most experiments conducted on the V1 are not designed to take into consideration the vast amount of intracortical input to the cortical V1 neurons, focussing only on the feed-forward pathway. The context dependence of V1 neurons' behaviour is now well established and the idea is that about 60-80% of the neuronal response is a result of internal feed-back pathways from other V1 neurons, while approximately only 35% is a function of the geniculate input [Sillito et al., 2006, Horn et al., 2000]. This kind of interdependence and contextual effects are largely ignored while designing experiments projecting a view that is too simplistic and unrealistic.

5. *Ecological deviance:* Often the data that are not in favour of a standard model are discarded as noise or experimental error and only those are published that strengthen it. Journals are often unprepared to publish such results that demonstrate the failure of a popular model, unless the study also suggests and experimentally validates an alternate one. On the other

hand, forming a model may take a significant period of time and how would the model be formed when the results are not surfaced. Hence, it is important to know not just those data that verify any accepted model, but also those that go against it. For example, studies that used natural scenes as opposed to simple stimuli have found the results to deviate from the predictions of standard models [Kayser et al., 2003].

► 1.2 Research Scope and Objectives

Intending to promote interests in the direction of biologically-inspired approaches towards Computer Vision, this work explores the visual pathway, builds models and applies them to image reconstruction, contrast enhancement and object recognition.

This thesis constrains itself to the study of a particular functionality of retina, namely the processing of foveal image modulo peripheral vision.

The component of motion throughout the visual pathway is disregarded, limiting the focus to static images; as are the plurality of pathways for electrical signals from retinal to various centres of the brain - studying only the pathway to the brain's lateral geniculate nucleus (LGN) leading to V1. This study also concerns itself with the mechanics by which the eye extracts contrast information. The constraints herein made are necessarily simple and fundamental to understanding any complex system.

The primary goal of this dissertation is to understand the information coding of the retinal ganglion cells over the optic nerve to the lateral geniculate nucleus (LGN) and the primary visual cortex (V1) from the perspective of Computer Vision investigating redundancy reduction, information preservation and contrast enhancement. The first two of these goals are achieved through the development of an image reconstruction algorithm. The image falling on the retina is filtered

out by simulated on- and off- centre receptive fields resulting in a contrast map that is sent over the optic nerve to various centres of the brain. The image reconstruction algorithm is used to reconstruct the original image back employing only its contrast map, thus showing that the coded signal is complete in its information content. Furthermore, the robustness of the neural circuitry is studied by simulating irregularities in the retinal receptive fields by the induction of artificial noise.

On- and off-centre maps are reconsidered in the context of contrast enhancement and a novel algorithm is developed whereby contrast enhancement is reduced to a scaling operation over the contrast domain. The gain control function is perceptually inspired with asymmetric gain factors that are applied to the two maps resulting in control over the mean luminance. Another interesting facet of the contrast enhancement algorithm developed here is that it is equally applicable to the colour-opponent representation of colour images known to exist within the retina and cortex.

Further, taking inspiration from the brains orientation to concision of representation, this research looks into sparse representation classification as an approach to object recognition using a perceptually compatible structural error measure and replacing the conventional sparsifiers with a genetic algorithm variant that allows literally any error function to be used resulting in improved recognition rates and improved sparseness of solution.

1.3 Thesis Overview

The remainder of this thesis is structured as follows:

Chapter 2 presents a reconstruction algorithm for images from its on- and off-centre contrast maps. A near-perfect reconstruction proves that there is no significant information loss. This chapter further discusses simulation results

employing the irregular receptive fields akin to those present in nature.

Chapter 3 talks about the application of the reconstruction algorithm developed in chapter 3 to image contrast enhancement as a model for gain control. Its application to both gray scale and colour images is explained. In colour images, the algorithm is applied using three different colour models - RGB, L*a*b* and the colour-opponent receptive field model present in primate brain. That this algorithm can directly deal in a representation known to exist in primates, makes it a study of interest.

Chapter 4 discusses face recognition as a sparse representation problem and extends it to use any error measure for l_1 minimization and the calculation of residuals. As an example, the structural similarity measure (SSIM) is taken to replace the mean-square error (MSE) measure because of its similarity to human perception. The robustness of SSIM over illumination and expression variations is demonstrated. Also demonstrated are the deficiencies of the MSE measure. Furthermore, a new sparsifier is developed that performs l_1 minimization using genetic algorithms. Experimental results are presented comparing the original algorithm with its extended version.

Chapter 5 summarizes the research work and proposes areas for future research.

2

Image Reconstruction from Contrast Information

The research here aims to investigate the information content of the coded image signal from the retinal ganglion cells to the cortex through the optic nerve. This is achieved through an image reconstruction approach. The coded image is used to reconstruct the original image and the fidelity of the reconstructed image is taken to be a measure of the presence or loss of information in the coded signal. The results shed light not only on the nature of retinal processing, they are an indication of a general behaviour of processing in nerve cells.

This chapter presents an iterative algorithm developed for the reconstruction of natural images given only their contrast map. The solution is neurophysiologically inspired, where the retinal cells, for the most part, transfer only the contrast information to the cortex. By being able to reconstruct the original image from these contrast maps, this research shows thus that the coded image sent to the brain through the optic nerve does indeed contain all the information from the original image. The image reconstruction algorithm is based on least squares error minimization using gradient descent. The chapter also builds its corresponding Bayesian framework for the underlying problem. Starting from an initial image, its contrast map is computed using the Difference of Gaussians

(DoG) operator at each iteration, which is then compared to the contrast map of the original image generating a contrast error map. This contrast map is put through a non-linearity to deal with saturation effects. Pixel values are then updated proportionally to the resulting contrast errors. Using a least squares error measure, a convex error surface with a single minimum is obtained, thus providing consistent convergence. The experiments show that the algorithm's convergence is robust to initial conditions but not the performance. A good initial estimate results in quick convergence. The algorithm is extended on to colour images using three different colour models. The research introduces the terminology of regular and irregular receptive fields as representing simulated and natural receptive fields. These are explored and reconstruction results are presented. The reconstruction algorithm provides a novel approach to manipulating an image in its contrast domain.

► 2.1 Salient Contributions of This Research

- Development of an iterative algorithm based on least squares error minimization to reconstruct a natural image from only its contrast map.
- Development of its Bayesian framework.
- Extension of this algorithm to colour image reconstruction using three different models, namely the RGB model, the L^*a^*b model and the neurophysiologically inspired colour-opponent receptive field model.
- Reconstruction using irregular receptive fields - a simulation of natural retinal processing.
- Part of the research presented here has been published [Khwaja and Goecke, 2008].

2.2 Overview and Background

Image reconstruction or restoration is an inverse problem which aims to recover the original image from its transformed version. It is an ill-conditioned problem at best and a singular problem at worst [Andrews and Hunt, 1977]. The transformation of the image could be the result of unwanted distortion like blurring, aliasing, or noise, or it could be a purposeful operation like compression, convolution, etc. Similarly, the transformation function may or may not be known. Furthermore, the transformation could be information-preserving or non-preserving. Inverse problems are generally ill-posed in that they might not have a unique solution or a solution at all.

Colloquially, it is often known that interesting information lies in the contrast. An image's contrast map contains within it everything that is needed for a faithful reconstruction. Here a technique is described which reconstructs an image given only its contrast map such that all the relative intensity values are restored. And not only that, even the ambient levels get restored to their original values. But how is that possible, one may ask, as the mere process of calculating contrast discards the ambient component of the luminance in the image? It turns out that even though the ambient component is thrown away, its information gets encoded in the relative intensity levels as the contrast map of any natural image is a highly constrained system when computed in an overlapping series of image patches. This enables a high fidelity reconstruction of the images from only their contrast information.

Humans are visual beings. Our visual system is not only the most elegant example of vision in nature, it is the foremost example we have. If it were not for our ability to see, we would have no idea whatsoever of what actually is a seeing experience. Hence, the motivation of this research is derived from discoveries in

primate vision in the last fifty years.

Primate Retina

Faced with a load of information, the primate retina with an approximate 130 million cells has to decide about transmitting the most important information, disregarding those that are not needed. It turns out that what is transmitted tells the cortex about changes in the pattern of information reaching the eye [Kuffler, 1953, Hubel, 1995, Hubel and Wiesel, 1959]. These changes are either in space, such as a border between a bright region and a dark one, or in time, such as a sudden increase or decrease in light intensity. Thus, the retina is primarily a detector of change, both spatially and temporally. Hence, the main function of the vertebrate retina, in doing spatial analysis (as we are concerned), is to extract contrast from the luminance distribution, thus making the percept largely independent of the ambient illumination level. This is achieved through a layered architecture of bipolar and ganglion cells (Fig. 2.1) that perform spatial differentiation (broadly symmetric) on their inputs using a finite sized, roughly circular, receptive field with antagonistic centre and surround. Two such types of cells are found to exist in the retina which are termed as the on and the off-centre cells. The on-centre cells are activated when the centre of their receptive fields are brighter than their surround and deactivated otherwise. The off-centre cells work the opposite way by turning on when the surround is brighter than the centre and off otherwise. Together these two cell types capture all the spatial information that is available in an image.

Kuffler's Experiments

These cells in the retina and their peculiar behaviour were first discovered by Kuffler [Kuffler, 1953] in 1952 who realized that shining an all-invasive bright

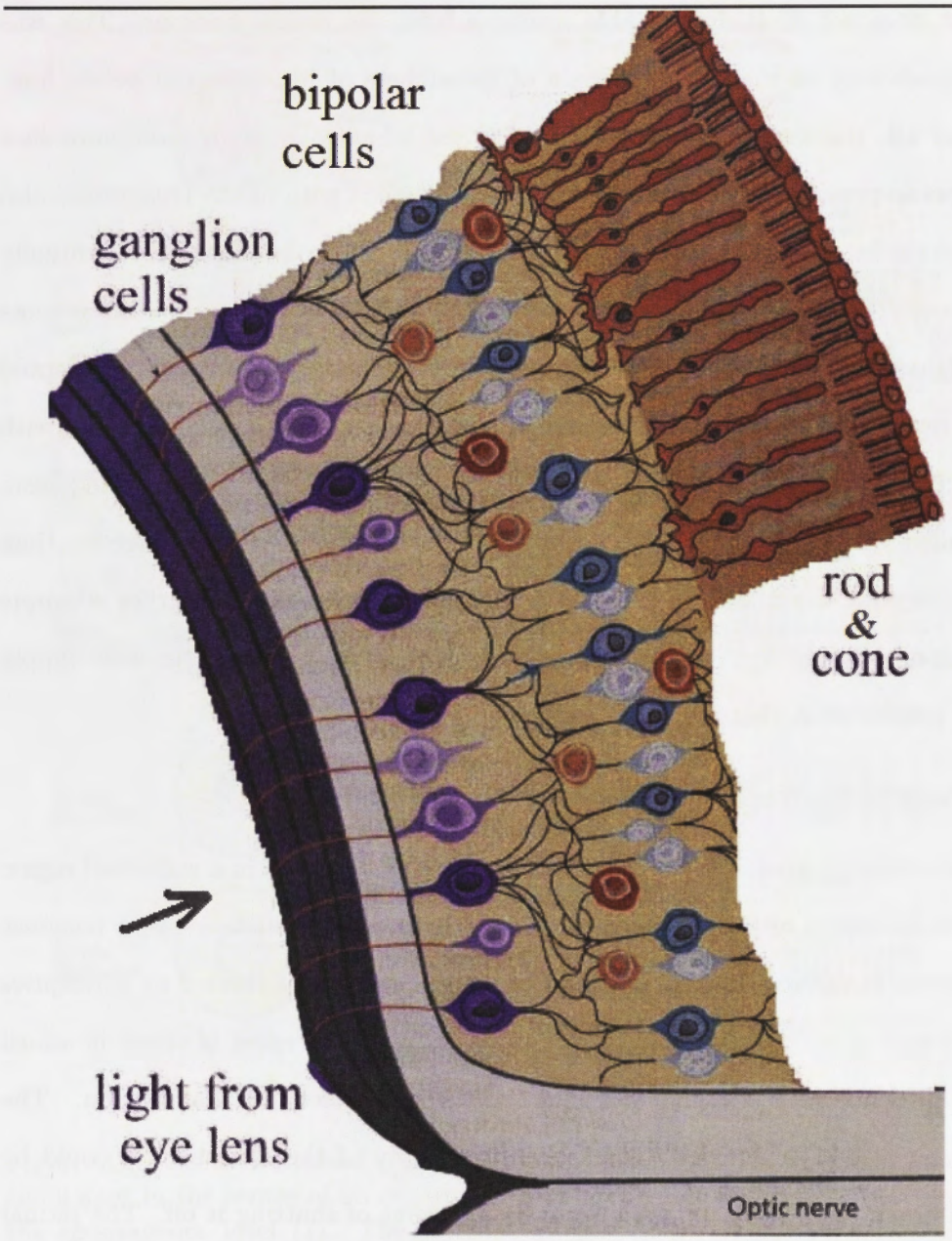


Figure 2.1: Retinal segment showing the translucent layers of neurons. Light has to pass through ganglion cells, amacrine cells, bipolar cells and horizontal cells before it finally reaches the photo receptors called rods and cones. The electrical signal pathway is in the reverse order. The rods and cones generate a signal that goes to the bipolars (modulated by horizontal cells) and then to the ganglions (modulated by the amacrine cells), finally leaving the retina through the optic nerve. (From *Eye, Brain, and Vision* [Hubel, 1995])

light does not elicit any notable response from the retinal neurons. This was contradictory to the common sense of researchers of his time and before him. After all, the retina is sensitive to light and what could be a more sure shot response-provoking stimulus then to flood the retina with light. That would also solve the issue of picking the correct axon for electrode placement. He eventually discovered the right stimulus. The retinal ganglion cells were interested in spots of light of precise diameters. By shinning small spots of light on the light-adapted cat retina, Kuffler showed that ganglion cells have concentric receptive fields with an 'on' centre and an 'off' periphery and vice versa. The mutual antagonism implied that a spot restricted to the centre of the field was more effective than one covering the whole receptive field. That was the reason why earlier attempts to shine a bright light resulted in no response. The retinal neurons were simply not interested in that kind of stimulus.

Receptive Fields

The retinal neurons were found to respond to stimulus only in a restricted region of the retina. Outside this region, the stimulus would fail to evoke a response however favourable it may be. This localized region was termed as a *receptive field* and so a receptive field of a sensory neuron is a region of space in which the presence of a stimulus will affect the firing pattern of that neuron. The affect could be excitatory, increasing the activity of the neuron or it could be inhibitory, decreasing its activity up to the point of shutting it off. The retinal ganglion cells have a receptive field structure that consists of a central region and a surrounding region. These are roughly symmetrical and round - as symmetrical and round as it can get owing to the development process it is subjected to (This is discussed in detail in a later chapter). The centre of the 'on' cells, as they are called, is excitatory, that is, light falling on the centre will increase the firing

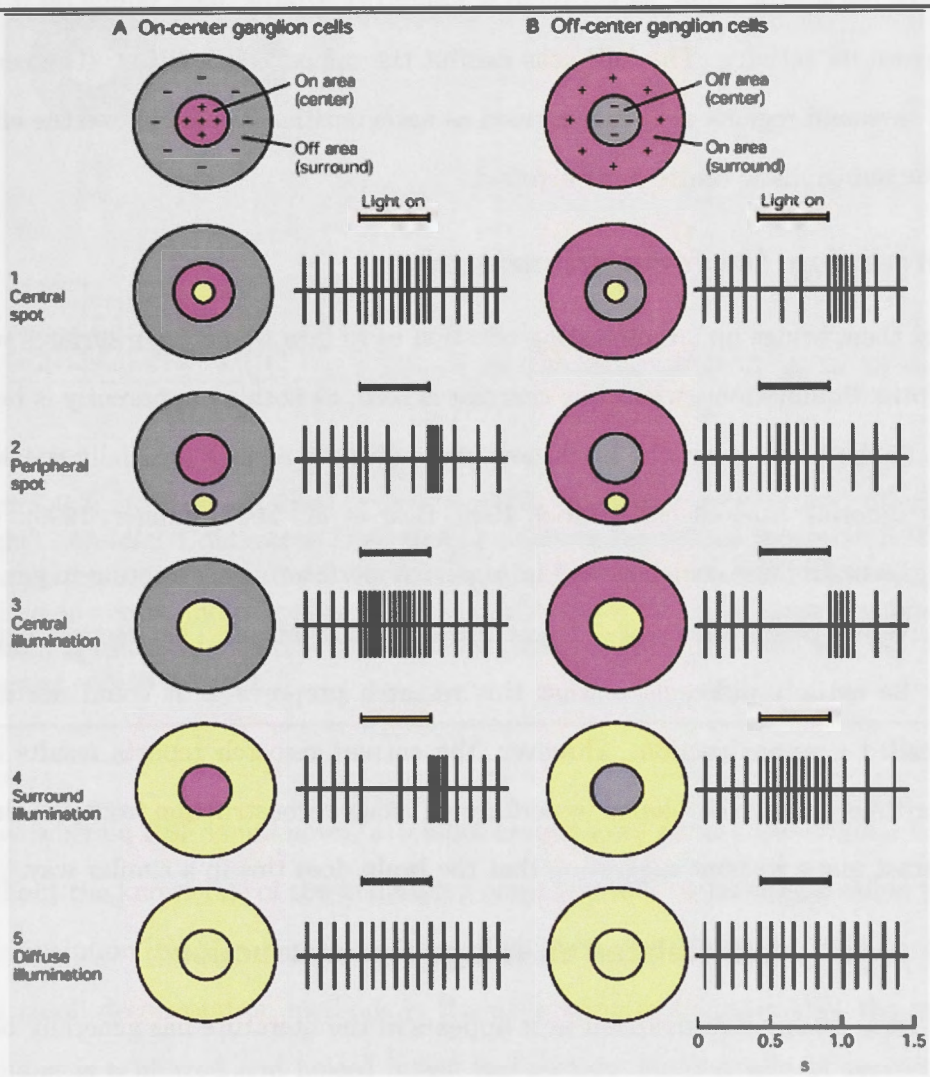


Figure 2.2: Centre-surround receptive fields depicting luminance response. A small spot in the centre of an on-centre cell increases its activity slightly beyond the spontaneous level (1). The same sized spot in the surround decreases its activity slightly below the spontaneous level (2). When the spot is enlarged to cover the entire centre but not the surround, the activity reaches its peak (3). When the entire surround is illuminated but not the centre, the activity gets suppressed to its limit (4). When both the centre and surround are illuminated entirely, the neuron maintains its spontaneous activity level - this kind of stimulus is least interesting to it (5). The off-centre cell exhibits the opposite behaviour. (From *Principles of Neuroscience* [Kandel et al., 2000])

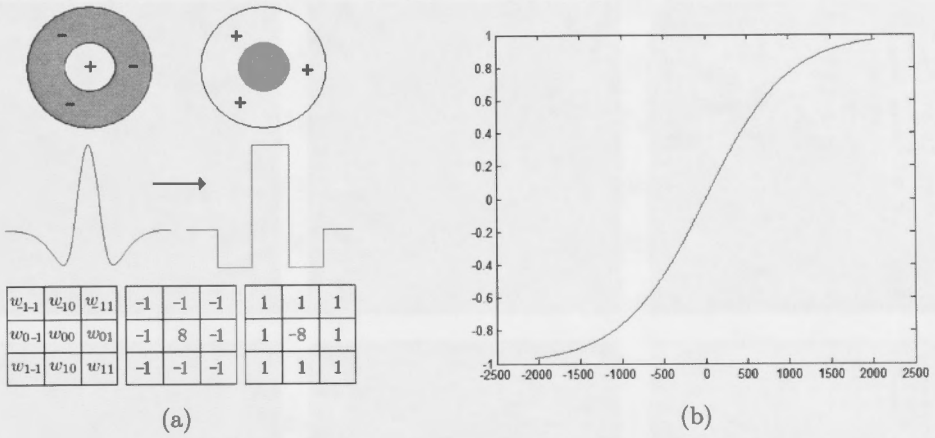
rate of the neuron while its surround is inhibitory, that is, light falling on it will decrease its activity. The 'off' cells exhibit the opposite behaviour. The centre and surround regions are hence termed as antagonistic. Fig. 2.2 shows the effect of the antagonistic centre and surround.

Perceiving Uniform Illumination

This, then, brings up an interesting question as to how we perceive surfaces with uniform illumination - where the contrast is zero, as nothing apparently is being sent to the brain about the brightness levels of those regions (spatially speaking only, ignoring time effects) [Hubel, 1995, Bear et al., 2007, Palmer, 1999]. For this the brain must somehow and in some way perform reconstruction to get the relative brightness values of different regions of the image. That reconstruction may be entirely different to what this research proposes or it could not even be called a reconstruction. However, the current research reports results and algorithms that are developed to perform an image reconstruction from just those contrast maps without suggesting that the brain does this in a similar way.

► 2.3 Prior Work Related to Image Reconstruction

The term *image reconstruction* as it appears in the literature has generally been used to imply some form of image filtration or reverse filtration to retrieve a denoised version of it. Deconvolution is usually performed for image restoration in many applications. Andrews and Hunt [Andrews and Hunt, 1977] describe in detail the linear algebraic restoration methods that include inverse filters, least-squares or Wiener filters, constrained least-squared filters and maximum entropy filters. Bates and McDonnell [Bates and McDonnell, 1986] describe these in terms of multiplicative and subtractive deconvolutions and other such techniques. In all these the point spread function (PSF) is assumed to be known *apriori*. In



(a) (b)

Figure 2.3: (a) *Top:* Retinal receptive fields. On-centre (left) and off-centre (right). *Middle:* Continuous DoG (left) modelling the retinal receptive field approximated by its discretized version (right). *Bottom:* a general 3x3 mask (left) and its on-centre and off-centre weights (centre and right). (b) Tangent sigmoidal squashing function with $\alpha = 1000$. With a mask of (a), the values will be in the range of -2040 to 2040.

cases where a PSF is unknown, attempts are made to restore the original image without the knowledge of the underlying noise function - a technique called blind deconvolution [Kundur and Hatzinakos, 1996, Ayers and Dainty, 1988]. But central to all deconvolution methods is the underlying assumption that the image at hand is a blurred and noised image and so they majorly aim at performing deblurring and denoising. Besides, the selection of an appropriate technique depends upon the form and extent of the PSF, the nature of the original image and the extent of distortions in it [Bates and McDonnell, 1986]. Furthermore, as [Bates and McDonnell, 1986] points out that in order to apply standard deconvolution methods effectively, it is almost always necessary to pre-process the given image in some way including enhancement, sectioning, windowing etc. We show a comparison of our algorithm in Fig. 2.4 with that of two deconvolution algorithms: the regularized filter deconvolution and the Wiener filter deconvolu-

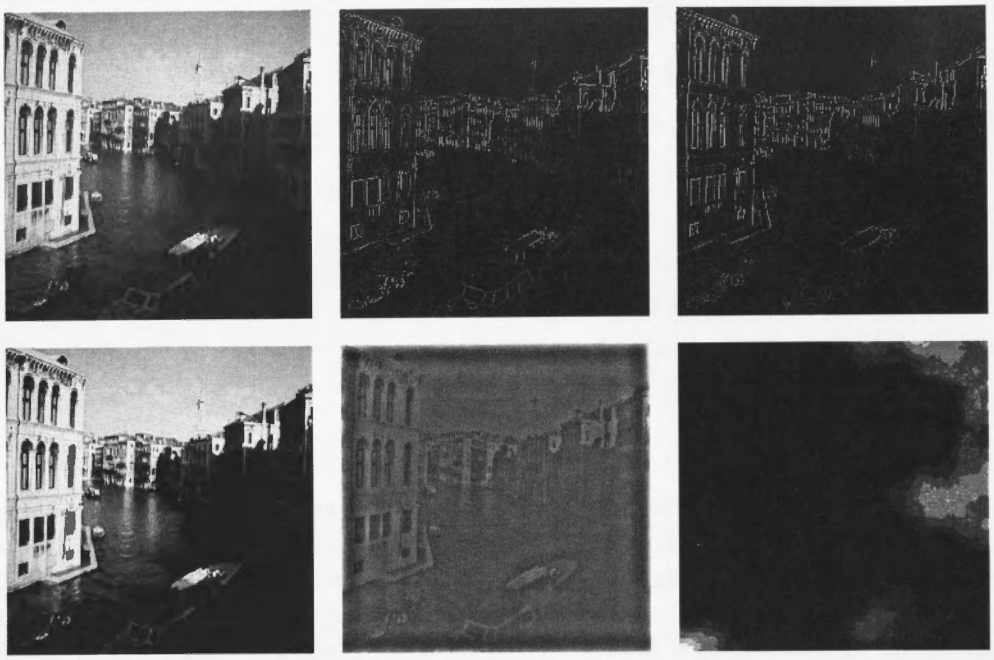


Figure 2.4: *(Top,left)*: Original image *(Top, middle & right)*: Its on- & off-centre map *(Bottom-left)*:Reconstruction using the proposed algorithm *(Bottom, mid)*: Reconstruction using Wiener filter deconvolution *(Bottom, right)*: A 50x scaled error image of our reconstruction giving a PSNR of 48.82db

tion[1]. As can be seen, the deconvolutional output does not restore the lightness levels and introduces additional noise. In addition, the non-blind and blind versions of Richardson-Lucy deconvolution failed to reconstruct the image at all. Some authors have mentioned image reconstruction from contrast information as an interesting issue in their work without resorting to any solution. Hubel, for one, raises this question in his book [Hubel, 1995] in which he wonders how we are able to see uniformly lit surfaces when all our cells respond to, is the local intensity differences. However, without proposing any answers to it, he quickly disposes off the topic by supposing that the centre-surround architecture results in it somehow.

[1]The Matlab implementations of these algorithms were used in the experiments

Other related work is perhaps the work of Land and McCann on *Retinex Theory* [Land and McCann, 1971] which was followed up by that of Horn [Horn, 1974] and Marr [Marr, 1974]. The same are also cited by Palmer [Palmer, 1999] when discussing the importance of intensity edges and luminance ratios. Retinex theory is a theory of colour constancy and brightness constancy in which luminance ratios are taken at edges and integrated across the visual field. Land and McCann [Land and McCann, 1971] mostly talks about colour constancy and about intensity calculations in one dimension. Horn [Horn, 1974] takes it to two dimensions but mentions that since the component related to ambient light is differentiated out, there would no way of calculating the brightness values of uniformly illuminated surfaces while Marr [Marr, 1974] focuses mostly on the primate retina. Although the Retinex theory and the work of Land and McCann deals with image reconstruction they do it within the context of colour constancy. Colour constancy is an important topic but quite irrelevant to the research presented here and so it will not be discussed any further.

Sadr et al. [Sadr et al., 2002] investigate local ordinal encoding and constraints. The authors show that such an approach can encode the most stable characteristics of an image and that it is neurally plausible while [Stanley et al., 1999] takes the neurophysiological data recorded from the LGN of cats and attempts to reconstruct the image to some extent with the motivation of unveiling the coded information in the action potentials of neurons.

The work presented here differs from the rest as we put forward a non-linear method of reconstruction given only the contrast maps of an image. This opens up avenues where by images can be transformed to contrast domain, manipulated and then transformed back. Calculating the contrast of an image significantly loses information and the contrast map does not lend itself to such enhancement techniques as needed to make it suitable to apply classical deconvolutional meth-

ods. Furthermore, this approach differs from [Sadr et al., 2002] and [Stanley et al., 1999] in that through the use of an iterative, non-linear technique it is able to reconstruct the image to such a higher fidelity which in a limited number of iterations produces an image which is visually indifferentiable to that of the original. Further on a Bayesian framework is developed for this method. Bayesian Inferencing has become a popular tool lately in image processing and computer vision, so much so that vision [Hanson, 1993, Tipping, 2004] now is being believed by some to be a Bayesian inferencing process [Mamassian et al., 2002].

▶ 2.4 Natural Image Reconstruction with Symmetrical Regular Receptive Fields

▶ *2.4.1 The Algorithm*

We compute on and off-centre contrast maps from the original image. If $\mathbb{M}$ represents the mask and $\mathbb{I}$ the image, then a contrast map is given by

$$\mathbb{C} = \mathbb{M} * \mathbb{I} \tag{2.1}$$

where $*$ is the convolution operator. Algorithm 1 depicts the step-by-step reconstruction procedure. The following sections work out the math in detail.

▶ *2.4.2 Least Squares Error Minimization*

Using two 2D Gaussians with zero means,

$$\left.\begin{aligned} G_{\sigma_1}(u,v) &= \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(u^2+v^2)}{2\sigma_1^2}} \\ G_{\sigma_2}(u,v) &= \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(u^2+v^2)}{2\sigma_2^2}} \end{aligned}\right] \tag{2.2}$$

Algorithm 1 Image Reconstruction From Contrast Information

```
1: img_in ← input
2: rf ← receptive_field_mask
3: contr_d ← compute_image_contr(img_in, rf)
4: eta ← 0.8
5: img_out ← initial_value
6: while stopping_condition ≠ true do
7:   for all [x, y] in img_out do
8:     contr_a ← compute_pixel_contr(x, y, rf)
9:     contr_e ← contr_d[x, y] − contr_a
10:    if contr_e ≠ 0 then
11:      img_out[x, y] ← img_out[x, y] + eta ∗ contr_e
12:      if img_out[x, y] < 0 then
13:        img_out[x, y] ← 0
14:      end if
15:      if img_out[x, y] > 255 then
16:        img_out[x, y] ← 255
17:      end if
18:    end if
19:  end for
20: end while
```

The centre-surround receptive field R is defined as a Difference of Gaussians (DoG) at any point (u, v) as

$$R(u, v) = G_{\sigma_1}(u, v) - G_{\sigma_2}(u, v) \tag{2.3}$$

The contrast at any pixel i at location (x, y) is defined as a convolution of the receptive field R with the image intensities I in the neighbourhood of the pixel (x, y) (see Sec. 3.3 for contrast definition)

$$C(x, y) = I(x, y) * R(u, v) \tag{2.4}$$

$$C(x, y) = \int_v \int_u I(x - u, y - v) R(u, v)\, du dv \tag{2.5}$$

For the discrete case, the receptive field $R(u,v)$ is approximated by the mask $M(u,v)$ and the convolution then becomes

$$C(x,y) = \sum_v \sum_u I(x-u, y-v) M(u,v) \tag{2.6}$$

where u, v range for the size of the mask. However, because the combined input representing the entire neighbourhood influence can exceed limits, one must also specify a squashing function to keep resultants within reasonable range. A tangent sigmoidal non-linearity (Fig. 2.3(b)) is used as a squashing function

$$y(x) = \frac{1 - e^{-2x/\alpha}}{1 + e^{-2x/\alpha}} \tag{2.7}$$

where α is the constant for range adjustment and is dependent on a mask of particular size and weights. In certain applications, α could be made to vary as a function of local contrast serving as a dynamic range adjustment factor. The cost function is defined as

$$C_{SE}(x,y) = \frac{1}{2}[C_D(x,y) - C_A(x,y)]^2 \tag{2.8}$$

The original image's contrast map is taken as the desired contrast C_D and the contrast of the reconstructed image as the actual contrast C_A which is evaluated at every iteration and their squared difference is the squared contrast error measure C_{SE} at any pixel (x,y). Based on this error measure the actual pixel intensities are updated until C_A converges to C_D.

The gradient of this error w.r.t. the actual pixel intensity I_A at (x,y) is

$$\frac{\partial C_{SE}(x,y)}{\partial I_A(x,y)} = -[C_D(x,y) - C_A(x,y)] \frac{\partial C_A(x,y)}{\partial I_A(x,y)} \tag{2.9}$$

so that,

$$\frac{\partial C_A(x,y)}{\partial I_A(x,y)} = R(0,0) \tag{2.10}$$

But $R(0,0) \approx M(0,0) = w_{00}$ for a mask of finite size, leading to the following error expression

$$\frac{\partial C_{SE}(x,y)}{\partial I_A(x,y)} = -[C_D(x,y) - C_A(x,y)]w_{00} \tag{2.11}$$

Using gradient descent, the update rule then becomes

$$\Delta I_A(x,y) = \eta w_{00} C_E(x,y) \tag{2.12}$$

where η is the update constant and $C_E(x,y) = [C_D(x,y) - C_A(x,y)]$ is the contrast error at pixel (x,y). Starting with an initial image local changes are performed using this pixel update rule until some convergence criterion is met.

▶ 2.4.3 Bayesian Framework

This section develops a Bayesian framework for the least-squares error minimization of Section 2.4.2. Given the desired contrast map (of the original image) $\mathbb{C}_D$, the posterior distribution of the reconstructed image at each iteration, given Bayes' rule, is

$$p(\mathbb{I}_A|\mathbb{C}_D, \mathbb{M}) = \frac{p(\mathbb{C}_D|\mathbb{I}_A, \mathbb{M})p(\mathbb{I}_A, \mathbb{M})}{p(\mathbb{C}_D|\mathbb{M})} \tag{2.13}$$

The term $p(\mathbb{C}_D|\mathbb{I}_A, \mathbb{M})$ is the probability distribution of obtaining the desired contrast map from the reconstructed image at any iteration. Since the contrast at any pixel i depends only the image intensities at i and its defined neighbourhood and is independent of contrasts calculated at any other pixel, the likelihood, as

well, at every pixel is independent of the others, hence

$$p(\mathbb{C}_D|\mathbb{I}_A, \mathbb{M}) = \prod_i p(C_{D_i}|I_{A_{i\epsilon N}}, \mathbb{M}) \tag{2.14}$$

where, $\mathbb{I}_A$ is the image pixel intensity, $\mathbb{M}$ is the mask approximating the DoG (Eq. 2.3), C_{D_i} is the desired contrast at any pixel i, I_{A_i} is the actual intensity at pixel i and $i\epsilon N$ represents the pixels in the neighbourhood N of i. The contrast error at the i_{th} pixel, defined in terms of the desired contrast (C_{D_i}) and the actual contrast (C_{A_i}), is

$$C_{E_i} = C_{D_i} - C_{A_i} \tag{2.15}$$

having a probability distribution function given by a normal distribution with zero mean $p(C_{E_i}|\sigma^2) = N(0, \sigma^2)$, so that

$$p(C_{D_i}|I_{A_{i\epsilon N}}, \mathbb{M}) = N(C_{A_i}, \sigma^2) \tag{2.16}$$

$$p(\mathbb{C}_D|\mathbb{I}_A, \mathbb{M}) = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(C_{D_i} - C_{A_i})^2}{2\sigma^2}} \tag{2.17}$$

Using a normal distribution here could, theoretically, cause pixel values to cross the valid range of $0 - 255$ because it is unbounded while an image is a strictly bounded system. Although other distributions exist, like the beta distribution that may be more appropriate in such circumstances, they are generally harder to manipulate. However, the desired boundedness is imposed algorithmically by limiting all pixels to 0 on the low side and to 255 on the high side.

Since no prior information exists about what the reconstructed image will look like, the best prior, in this case, would be a uniform random distribution in

the range of $0 - 255$

$$p(\mathbb{I}) = \prod_i p(I_{A_i}) \tag{2.18}$$

where,

$$p(I_{A_i}) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{255} & \text{for } 0 \leqslant x \leqslant 255 \\ 0 & \text{for } x > 255 \end{cases} \tag{2.19}$$

Now, maximizing $p(\mathbb{C}_D|\mathbb{I}_A, \mathbb{M})$ is the same as minimizing the negative log likelihood

$$-log\ [p(\mathbb{C}_D|\mathbb{I}_A, \mathbb{M})] = \frac{N}{2} log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_{i=1}^{N} (C_{D_i} - C_{A_i})^2 \tag{2.20}$$

The first term in Eq. 2.20 is independent of the pixel intensity and so the second term is the sum of squared error for the entire image and for a single pixel, minimizing the squared error would maximize the log likelihood

$$-log\ [p(C_{D_i}|I_{A_i}, \mathbb{M})] = \frac{1}{2\sigma^2} (C_{D_i} - C_{A_i})^2 \tag{2.21}$$

This is essentially a least-squares expression. For the case when the probability distribution function is normal and there is no prior information available, the Bayesian paradigm reduces to the maximum likelihood which gives the same result as the least-squares method.

► 2.4.4 Implementation

Parameters

The difference of Gaussian function is approximated by a 3x3 mask of values with a balanced centre and surround as in Fig. 2.3, the centre consisting of a single pixel only. The parameter α of the tangential squashing function was chosen to

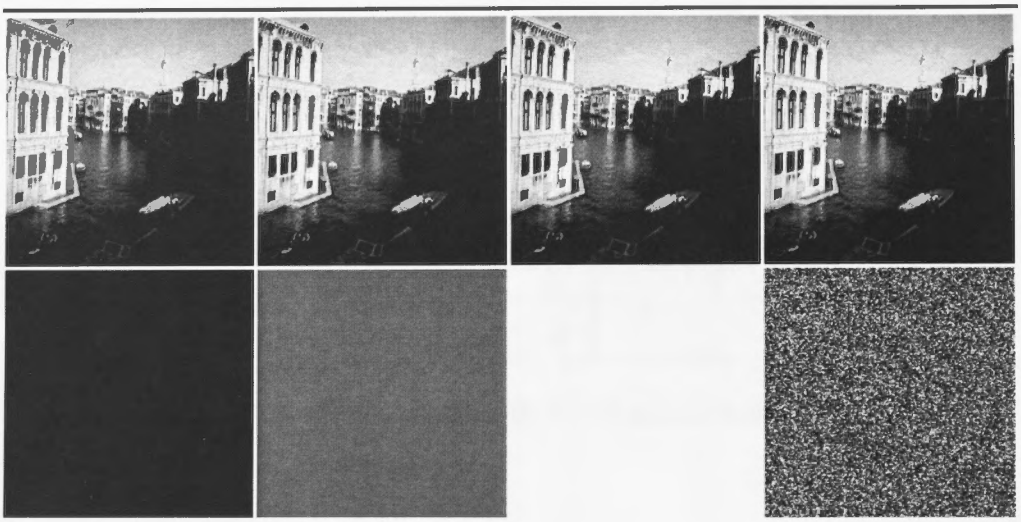


Figure 2.5: Reconstructed images in the top row with their corresponding initial images in the bottom.

be 1000. This choice was purely empirical to limit most of the output within the linear region of the function.

Initialization Values

The algorithm works by starting from some initial values for the pixels and progressively modifying them until converging to an image with the same contrast map as the given one. While the choice of initial values does not influence the final convergence result, a good initial guess can help in reducing the number of iterations needed for convergence. Experiments were conducted with a few different initializations. As shown in Fig. 2.5, these were all black pixels, all white pixels, all mid-level gray pixels and pixel values from a uniform random distribution in the range 0 to 255. It turned out that the algorithm converges relatively faster when an initial image comprising of pixels from all mid-level gray or uniform random distribution is chosen.

Stopping Condition

Let the following be the images at different iterations:

$$\mathbb{I}_0, \mathbb{I}_1, \mathbb{I}_2, \ldots, \mathbb{I}_{n-1}, \mathbb{I}_n \tag{2.22}$$

and let the absolute differences in these images be defined by the sequence

$$|\Delta\mathbb{I}_0|, |\Delta\mathbb{I}_1|, |\Delta\mathbb{I}_2|, \ldots, |\Delta\mathbb{I}_{n-1}| \tag{2.23}$$

where $|\Delta\mathbb{I}_{n-1}| = |\Delta\mathbb{I}_n| - |\Delta\mathbb{I}_{n-1}|$. Then for some threshold θ, we define the stopping condition as

$$\frac{|\Delta\mathbb{I}_{n-1}|}{|\Delta\mathbb{I}_0|} \leqslant \theta \tag{2.24}$$

Convergence

Although the least squares error cost function used here is non-linear, it is still a convex uni-modal function. In addition, the choice of α for the squashing function defined in Eq. 2.7 constrains most of the allowable values to the quasi-linear range around the origin, with few values curving off on the extremes as in Fig. 2.3(b). Together, these factors constrain the system of equations to solve in such a way that a unique solution is provided and convergence to the solution is guaranteed.

Normalized Masks

The pixel update rule in Eq. 2.12 depends on the centre weight of the convolution mask. This has the disadvantage that, if the mask is changed, the update will get severely affected. To make the update rule independent of such factors, normalized masks can be used with a centre weight of unity. This gives the following update rule:

$$\Delta I_A(x, y) = \eta C_E(x, y) \tag{2.25}$$

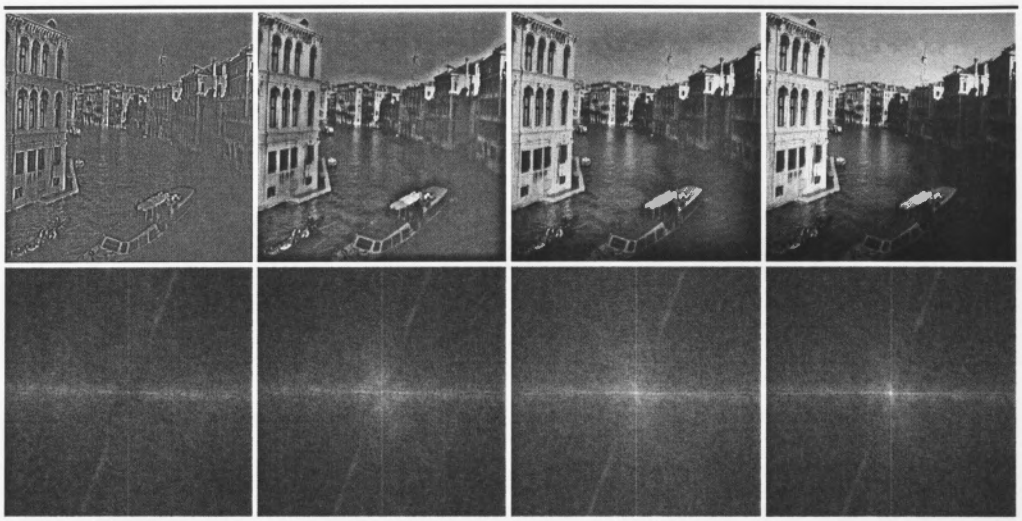


Figure 2.6: Reconstructed images and their FFT at iteration 1, 10, 100 and 500 showing the gradual reconstruction of low frequency components.

Doing so will require changes in the range adjustment constant α as well that is used in the squashing function (Eq. 2.7).

Spectral Analysis

The centre-surround filter acts as a high-pass filter keeping the high frequency components in the contrast information while discarding a band of low frequencies. These low frequency components get reconstructed through a diffusion process. The high frequencies determine the detail of an image. As the details are already preserved, information from the contrast diffuses to nearby areas, filling in, thereby reconstructing the image. Fig. 2.6 depicts the spectrum of the first few iterations showing the iterative reconstruction of low frequencies.

Peak Signal-to-Noise Ratio (PSNR)

PSNR is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. Because many

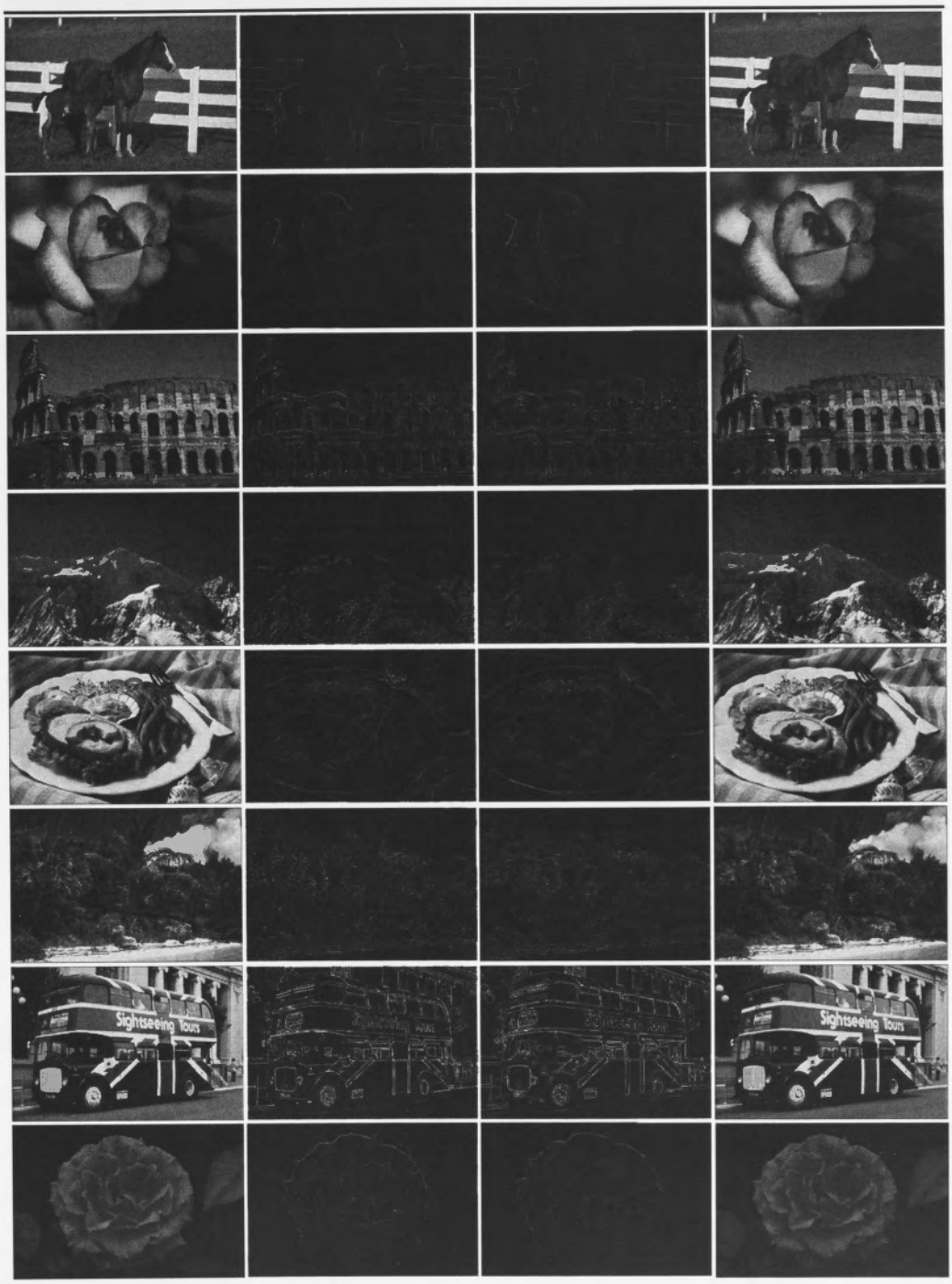


Figure 2.7: COREL images: *(Left):* Original, *(Centre):* Their on- & off-centre contrast maps and *(Right):* Their reconstructions.

signlas have a very wide dynamic range,PSNR is usually expressed in terms of the logarithmic decibel scale. Here, it is used as a measure of the quality of reconstruction.

Results

Tests were performed on all images from the COREL 1000 image database (Fig. 2.7), generally used for content-based image retrieval. Each image was 256x384 pixels in size and was converted to gray scale before processing. Stopping criterion used (Sec. 2.4.4) during experiments was $\theta < 0.5\%$ which generated in a typical mean square error of less than 100. This resulted in reconstructions that are visually indifferentiable to the original image. However the algorithm requires a huge number of iterations to reconstruct such high fidelity images and typical values range anywhere from few hundred iterations to a thousand or more. This is due to the fact that there is a kind of diffusion process in effect where changes diffuse out from areas of high contrast to those of low or zero contrast. Similarly, images that are composed of large uniformly illuminated areas require an unusually high iteration count to produce reconstructions that are anywhere near the original.

► 2.5 Colour Extensions

The above methodology can be easily extended to colour images. Three different approaches were experimented with namely the RGB Space, the L*a*b* Space and the colour-opponent receptive fields.

► *2.5.1 RGB Space*

This is a natural extension of the gray-scale algorithm to colour images by applying it to each of the three colour planes individually. Each colour plane of the RGB system can be treated as a gray scale image in itself [Foley et al., 1995]. An

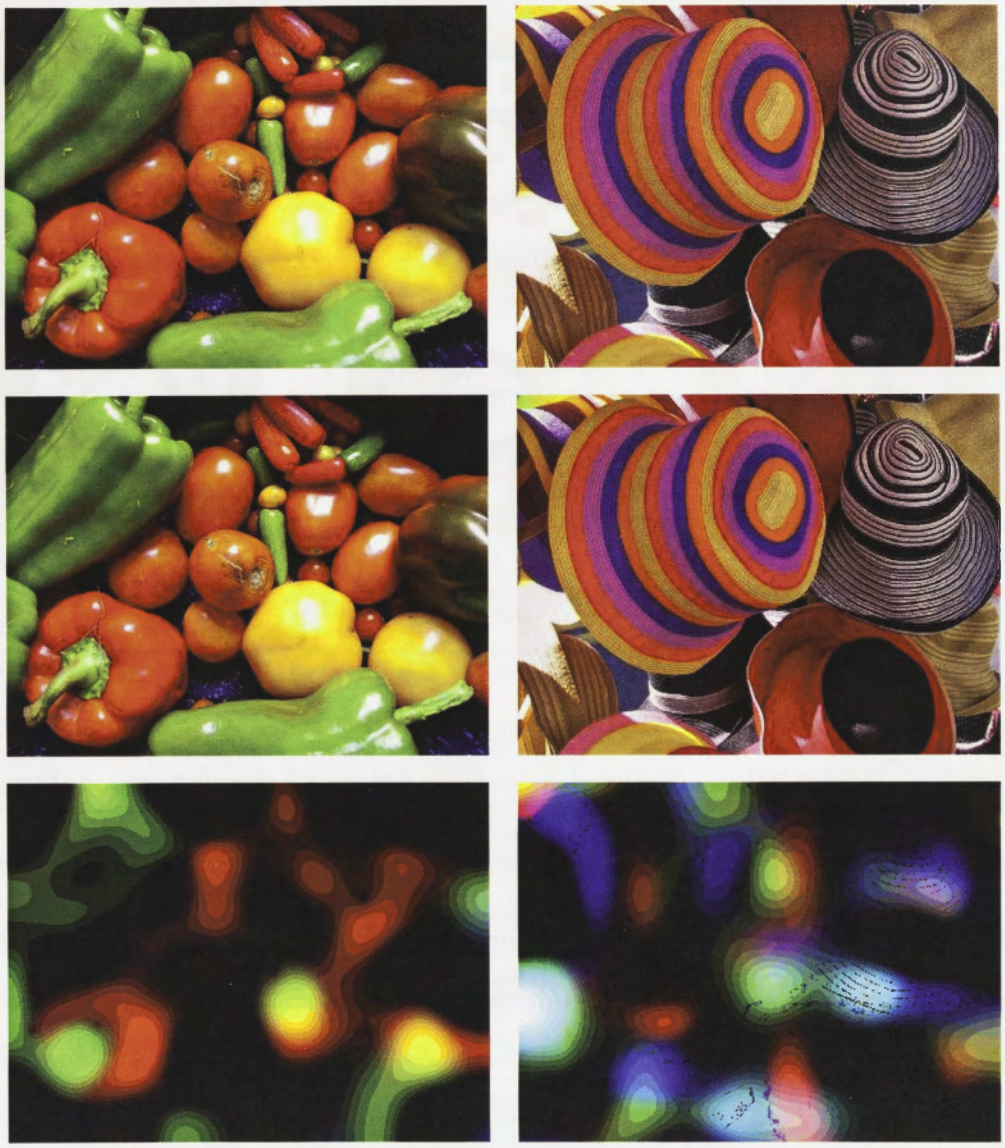


Figure 2.8: Original (top), reconstruction (middle) and 30x scaled error image (bottom) with a PSNR of 34.60db and 38.43db, respectively.

on-centre and off-centre map for each plane is computed resulting in six contrast maps. This approach involves three times more computations than for gray scale images. Some initial results are shown in Fig. 2.8. It should be noted, however, that this is not the way a primate brain is known to handle colour information

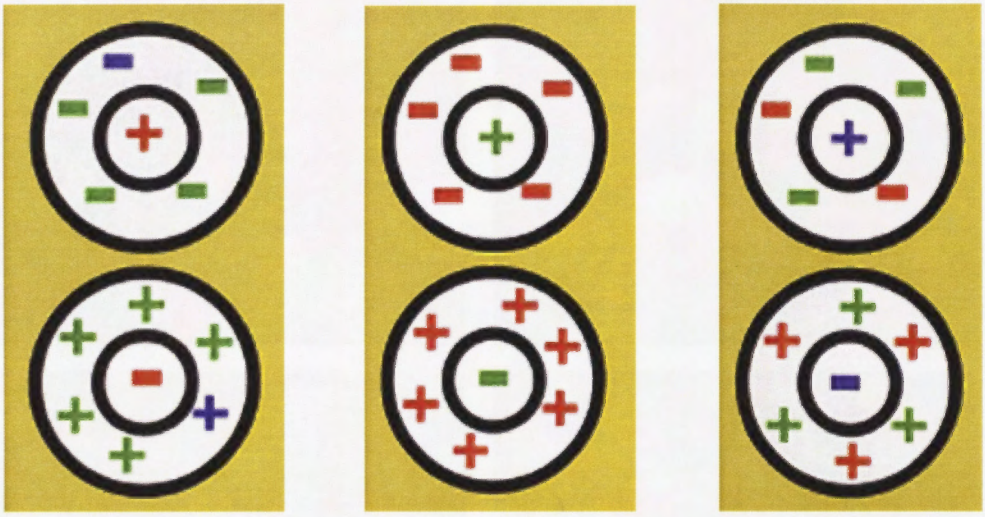


Figure 2.9: Colour-opponent Receptive Fields

[Livingstone, 2002]. In the RGB space, the contrast maps are computed using the three colours independently with no dependencies whatsoever among them.

▶ 2.5.2 $L^*a^*b^*$ Space

In the L*a*b* Space, the luminance and colour components are separated out [Anagnostopoulos et al., 2005]. The luminance component then acts like a gray-scale image. Contrast map is calculated of the luminance component and the reconstruction algorithm is then applied. The reconstructed gray-scale image is then merged with the colour component to generate the final colour image.

▶ 2.5.3 Colour-Opponent Receptive Fields

These receptive fields exist in the primate retina and have centre and surround that respond to different frequencies [Livingstone, 2002, Hubel, 1995]. Three different types of receptive fields can be found there. One has a red-green opponency the other has green-red opponency and the third one has blue-yellow opponency.

-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	**48**	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1
-1	-1	-1	-1	-1	-1	-1

-0.838	-0.999	0.000	-1.065	-0.886	-0.951	0.000
-0.828	-1.019	-0.957	-1.068	-0.915	-0.943	0.000
-1.022	-1.151	-0.930	-0.990	-0.945	-1.067	-0.907
0.000	-1.084	-1.144	**48.231**	-1.131	-0.955	-1.129
-0.954	-0.984	-1.066	-0.975	-1.103	-0.989	-0.874
-0.990	-1.084	-1.043	-0.984	-1.140	-0.932	-0.903
-0.989	-0.970	-0.929	-1.055	-0.894	0.000	-0.950

Figure 2.10: An example of a 7x7 regular receptive field (left) and its corresponding irregular version (right) where the weights are modified by random noise under a Gaussian distribution with a standard deviation of 0.1 and mean of -1 for the surround and 48 for the centre. The centre for both is highlighted in bold. Along the periphery of the surround, elements are zeroed out at random simulating asymmetricity.

Opponency means that, for example, for a red-green cell, if a red spot shines on the centre it activates the cell. If this red spot is increased to cover the surround as well, it shows no effect and the cell continues to be active. However, if the surround is shown a green light, it inhibits the cell. The centre and surround respond only to a fixed wavelength. In a green-red cell, green light on the centre activates the cell while red on the surround inhibits it. Same goes for the blue-yellow where a blue light on the centre activates the cell while a yellow (red+green) light on the surround inhibits it. Each cell has an on-centre and off-centre version making a total of six receptive fields. Applying the reconstruction algorithm using these, requires calculating six contrast maps each of which is dependent on two or more colours. Note how this is different than the six contrast maps in the RGB space where each colour is treated as independent to the other in contrast calculation. The reconstruction is then performed in the usual way using these contrast maps.

▶ 2.6 Natural Image Reconstruction with Irregular Receptive Fields

The reconstruction algorithm developed and discussed in the previous sections used receptive fields that were an idealized approximation to the real receptive fields that exist in a mammalian retina. The approximations made, had taken all the surround weights to be the same, the receptive field to be perfectly symmetrical and all the receptive fields to be the exact copy of each other. In reality none of these assumptions hold. The following three imperfections can be identified in a mammalian retina [Liu et al., 2009]:

1. The weights of the surround can hardly be same so that no two connections influence the neurons output alike.

2. The receptive fields are never perfectly symmetrical. This is an outcome of the developmental process at work.

3. No two receptive fields can be found that are exactly the same. This too, is directed by the developmental process.

In short the real biological systems are full of imperfections which can be imagined by taking note of how a nervous system actually develops within the foetus all the way to birth and beyond.

Naturally, this aroused curiosity for understanding the nature of the distortions caused by such imperfections and whether the image could still be reconstructed after going through such noisy processing. In other words, how much of the information is retained in the image and if it is at all reconstructable then under what conditions and to what fidelity.

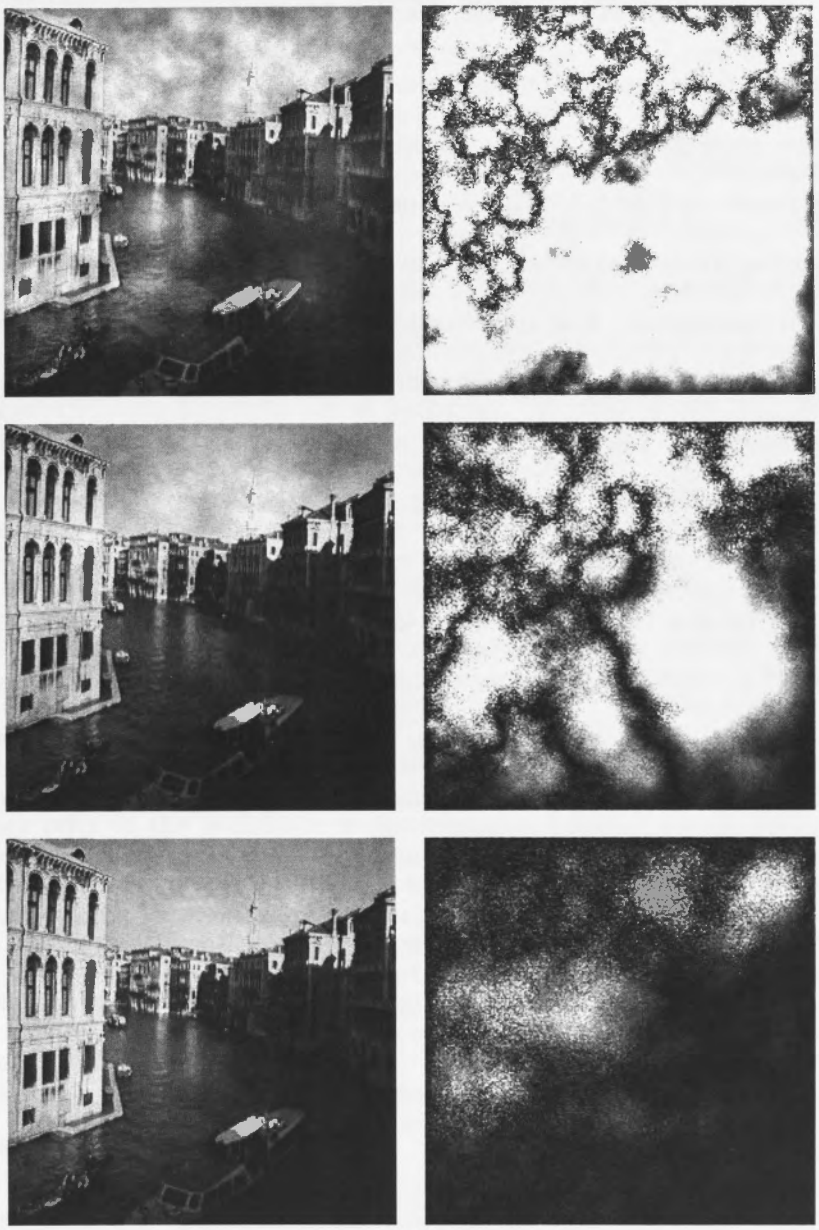


Figure 2.11: Reconstructions using irregular receptive fields generated with a standard deviation of 0.1 using a size 5x5 receptive field (top), size 7x7 receptive field (middle) and a size 11x11 receptive field (bottom) with their 20x scaled error images on their corresponding right giving a PSNR of 25.64db, 28.35db and 36.01db, respectively. A higher PSNR indicates a smaller error. Note that as the receptive field size increases, it becomes more inert to the irregularities.

2.6.1 Irregular Receptive Field Model

An irregular receptive field model was created by modifying the ideal one. Ideally the receptive field was modelled by a *nxn* convolution kernel. A 3x3 kernel is shown in Fig. 2.3. This has one element representing the centre and the remaining for the surround. All the surround elements are weighed equally to -1. To simulate the first imperfection, noise was introduced to each of the surround element under normal distribution with a fixed mean of -1 for the surround and 8 for the centre, under a fixed standard deviation. These two parameters are then kept the same for the whole image. The second one was handled in a like manner. The shape of a receptive field can be described by weights having a very small value or the value of zero. Thus weights at the periphery of a receptive field were randomly zeroed out to create un-symmetrical receptive field. The third imperfection was simulated by generating an entire array of such kernels, rather than using just one for the entire image as was done early on. The size of this array equalled the resolution of the image such that no two kernels were ever alike. In addition the array of convolution masks (simulating the receptive fields) used to reconstruct the image from the contrast maps were different than those used to calculate the contrast maps in the first place introducing an added level of uncertainty. An example 7x7 receptive field is shown in Fig. 2.10.

2.6.2 Experiments and Discussion

Experiments were performed on the 256x256 Venice image of Fig. 2.4 using three different receptive field sizes of 5x5, 7x7, and 11x11. A normal random distri-

bution was used to introduce noise into the generated receptive fields to create the effect of irregularities under a standard deviation of 0.1 and a mean of -1 for the surround and a corresponding mean for its centre. Fig. 2.11 shows results of the reconstructions. These reconstructions are compared to the original using peak signal-to-noise ratio (PSNR). The corresponding error images, enhanced 20 times for better perception, are shown on the right. It was observed that for a given standard deviation of errors introduced, the fidelity of the reconstruction increases as the receptive field size is increased. In other words, the bigger the receptive field, the larger the error it could handle. This can be explained by the errors actually cancelling each other out with some being positive and others negative when summed over a huge array of inputs. Mammalian brain cells are known to have an astronomical number of inputs ranging in thousands. This can, in part, explain their remarkable tolerances to noise. It can also possibly explain the variety of cell types mentioned by Hubel and Wiesel [Hubel and Wiesel, 1961] when they refer to the variations in the receptive fields of simple cells of which one particular form exhibits itself as asymmetricity in the two antagonistic regions. This, in the light of the experiments mentioned above, can be viewed as a manifestation of imperfections and irregularities in the brain rather than a new cell type.

► 2.7 Conclusions

An algorithm has been presented by which images (gray-scale and coloured) can be reconstructed given only their contrast information. Starting with an initial guess and using an iterative gradient descent algorithm, an image is progressively updated by comparing its contrast map with that of the original image with the difference being used as a cost function. A spectral analysis is presented, which shows the gradual reconstruction of the low frequency components which

were discarded by the DoG filter while calculating the contrast map. Tests were performed on all images from the COREL 1000 image database and results were presented.

The algorithm was extended on to the reconstruction of colour images. These images are complicated by the presence of more than one colour planes. A number of models are in use in computers to represent and manipulate colour images out of which two were picked (RGB and L*a*b*) due to their significantly different nature and a third model was inspired from neurophysiology known as the colour-opponent receptive field model. The colour-opponent receptive field model is known to exist within the cortex for colour image representation. The algorithm was applied successfully to colour images using all three models.

Furthermore, irregular receptive fields as reported to exist in the mammalian retina were simulated by injecting Gaussian noise with a fixed mean and standard deviation and reconstruction results were presented. It was discovered that the tolerance to a particular noise level (selected by the standard deviation) increases with the increase in the receptive field size.

Even though the image reconstruction method developed in this research can be neurophysiologically plausible as it uses only local computations, no such claims are made. This method can be viewed as providing a transformation and its inverse by which images can be transformed into their respective contrast domain, manipulated in interesting ways, and transformed back into image domain.

3

Biologically-Inspired Contrast Enhancement

This chapter presents a neuro-physiologically inspired model for the contrast enhancement of gray-scale and colour images. The contrast of an image is calculated using simulated on- and off-centre receptive fields whereby obtaining the corresponding two contrast maps. An adaptive asymmetric gain control function is proposed that is applied to the two contrast maps which are then used to reconstruct the image resulting in its contrast enhancement. The image's mean luminance can be adjusted as desired by adjusting the asymmetricity between the gain control factors of the two maps. The model performs local contrast enhancement in the contrast domain of an image where it lends itself very naturally to such adjustments. Being neuro-physiologically plausible, this model can be beneficial in theorising and understanding the gain control mechanisms in the primate visual system. The results are compared with the standard CLAHE algorithm.

▶ 3.1 Salient Contributions of This Research

- Development of a biologically-inspired contrast enhancement algorithm that uses the receptive field contrast operator and is based on the image reconstruction algorithm both of which were developed in chapter 2 suitable for

images having either a uni or multimodal histograms.

- Incorporation of both contrast and mean luminance control in the algorithm.
- Extension of the algorithm to colour images using three different colour models.
- The research presented here has been published [Khwaja and Goecke, 2009].

▶ 3.2 Overview and Background

Contrast is the dissimilarity or difference between things. In vision, it is the difference in visual properties that makes an object or its representation in an image distinguishable from other objects and the background. The visual properties that make a prominent contribution to the determination of an object's contrast with its surroundings are its colour and luminance. Contrast is an important attribute whose subjective and quantitative impact on the image quality is enormous, determining the visibility of its details as well as its aesthetic appeal.

The dynamic range of the visual stimuli in the real world is extremely large. On the contrary, image capturing and display devices are very restricted in their range, compressing the luminance levels to a very small range, thereby reducing the contrast between them significantly. Even for high dynamic range (HDR) images, displaying such an image on customary displays results in either its under- or over-exposure. This requires enhancement both to the contrast of an image as well as to its mean luminance level [Debevec and Malik, 1997]. Contrast in high spatial frequencies determines the visibility of details while that in the low spatial frequencies affects the appeal of it.

The primate visual system excels in tasks to a level so far unachievable by the best man-made systems. Among those is its amazing adaptability to a wide

range of luminance levels despite having a limited dynamic range. Here, a neuro-physiologically inspired model for local contrast enhancement of under- or over-exposed images is presented. The model gives excellent performance in enhancing the contrast of images having both uni- and multi-modal histograms. The algorithm works by computing two contrast maps of an image using the on and off-centre receptive fields of neurons, as found in the primate visual system. Contrast enhancement then becomes a straightforward operation where we rescale the contrast values in the map using a gain control function that is modelled after psychometric adaptability curves. These transformed maps are then used to reconstruct the image using our image reconstruction algorithm from [Khwaja and Goecke, 2008] which results in a contrast enhanced image. Feedback is used at every iteration to modify the gain control function. A unique feature of the model is that the image's mean luminance can be controlled by asymmetrically boosting the gain control factors that are applied to the two contrast maps. While it is not known that such asymmetrical boosting occurs at the neuro-physiological level, this research shows that superior results are achieved in this way as compared to symmetrical boosting on both maps.

► 3.3 Contrast Definition

There are several definitions of contrast and all of them are considered correct. The definition of contrast is often selected on the basis of the stimulus used [Kukkonen et al., 1993]. For spots it is based on the Weber fraction or light energy, for periodic stimuli, like gratings, in terms of Michelson contrast and for aperiodic and complex stimuli in terms of RMS contrast or contrast energy. However, in principle Michelson contrast, RMS contrast and contrast energy could be applied to all of these stimuli. Contrast can be defined in terms of luminance alone or it could also involve colour. Luminance contrast is a measure of the

perceived lightness or brightness difference between two colours. Common to most definitions of contrast is a measure of luminance difference relative to average or background luminance representing some kind of ratio [Peli, 1997]:

$$\frac{LuminanceDifference}{AverageLuminance} \tag{3.1}$$

The rationale behind this is that a small difference is negligible if the average luminance is high, while the same small difference matters if the average luminance is low.

The following are some of the varieties in which luminance contrast is defined:

Weber Contrast

The Weber contrast is defined as,

$$\frac{I - I_b}{I_b} \tag{3.2}$$

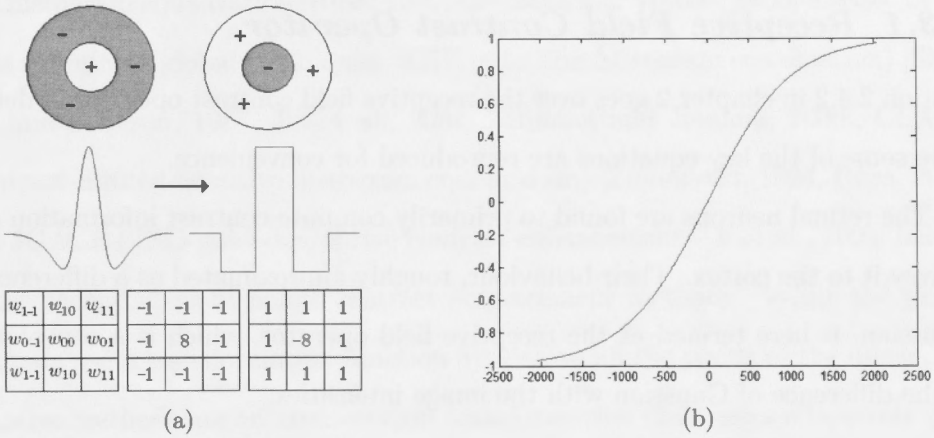
where I is the luminance of the features and I_b the background luminance. It is commonly preferred where small features are present on a large uniform background - in other words, the average luminance of an image is roughly the same as its background.

Michelson Contrast

The Michelson contrast defines contrast using the peak luminance values of the stimulus. It is defined as,

$$\frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{3.3}$$

where I_{max} and I_{min} represent the highest and lowest luminance. The denominator represent twice the average of the luminance. This particular metric is commonly used for images where both bright and dark features are present in



(a) (b)

Figure 3.1: (Repeat of Fig. 2.3 but added for convenience) The Receptive Field Operator that is used as the local contrast operator. The output of the convolution with the discretized DoG mask is sent through the sigmoidal non-linearity (a) *Top:* Retinal receptive fields. On-centre (left) and off-centre (right). *Middle:* Continuous DoG (left) modelling the retinal receptive field approximated by its discretized version (right). *Bottom:* a general 3x3 mask (left) and its on-centre and off-centre weights (centre and right) (b) Tangent sigmoidal squashing function.

equivalent measures.

RMS Contrast

Root mean square (RMS) contrast is defined as the standard deviation of the pixel intensities:

$$\sqrt{\frac{1}{MN}\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}(I_{ij}-\bar{I})^2} \tag{3.4}$$

where I_{ij} are the ith and jth element of the two dimensional image of size M b N. $\bar{I}$ is the average intensity of all pixel values in the image. The image I is assumed to have its pixel intensities normalized in the range $[0, 1]$.

3.3.1 Receptive Field Contrast Operator

Section 2.4.2 in chapter 2 goes over the receptive field contrast operator in detail. Here some of the key equations are reproduced for convenience.

The retinal neurons are found to primarily compute contrast information and convey it to the cortex. Their behaviour, roughly approximated as a difference of Gaussian, is here termed as the receptive field operator, which is a convolution of the difference of Gaussian with the image intensities:

$$C(x,y) = I(x,y) * R(u,v) \tag{3.5}$$

$$C(x,y) = \int_v \int_u I(x-u, y-v) R(u,v)\, du dv \tag{3.6}$$

This equation when discretized, replaces the receptive field $R(u,v)$ by a mask $M(u,v)$:

$$C(x,y) = \sum_v \sum_u I(x-u, y-v) M(u,v) \tag{3.7}$$

Because a neuron is a highly non-linear device, the result of this convolution is made to pass through a non-linearity. Historically, the sigmoidal function has been found to relate best to the neuronal output in the light of research data. The output, thus, of the convolution through the sigmoidal non-linearity is taken to be the contrast of the underlying image and the overall operator is termed as the receptive field contrast operator as depicted in Fig. 3.1.

3.4 Related Work

Processing techniques for image-contrast enhancement can be classified, according to one classification, into *global* and *adaptive* methods. Linear contrast stretch

and histogram equalization [Russ, 1995, Gonzalez and Woods, 2008] are two of the most prominent global techniques. AHE (adaptive histogram equalization) [Sherrier and Johnson, 1987, Jin et al., 2001, ZHiming and Jianhua, 2006], CLAHE (contrast-limited adaptive histogram equalization) [Zuiderveld, 1994, Reza, 2004] and JGACE (JND-guided adaptive contrast enhancement) [Ji et al., 1994] belong to the second class of image contrast enhancement methods. While the global methods use a transformation function applied to all the pixels of the image, the adaptive methods use an input-output transformation that varies adaptively with the local characteristics of the image.

Linear contrast stretch is one of the simplest piecewise linear functions which increases the dynamic range of the gray levels in the image being processed. Then there are histogram methods. The histogram of an image represents the relative frequency of occurrence of gray levels within an image. Histogram modelling techniques modify an image so that its histogram has a desired shape. This is useful in stretching the low-contrast levels of an image with a narrow histogram, thereby achieving contrast enhancement. In histogram equalization (HE), the goal is to obtain a uniform histogram for the output image, so that an *optimal*, so to say, overall contrast is perceived. However, the feature of interest in an image might need enhancement locally.

Adaptive histogram equalization (AHE) provides a local contrast enhancement by computing the histogram of a local window centred at a given pixel to determine the mapping for that pixel. A generalization of AHE, contrast limiting AHE (CLAHE) has more flexibility in choosing the local histogram mapping function. By selecting the clipping level of the histogram, undesired noise amplification can be reduced [Pizer and Amburn, 1987]. In addition, by method of background subtraction, the boundary artefacts can also be reduced [Rehm and Dallas, 1989]. The CLAHE algorithm partitions the images into contextual

regions and applies the histogram equalization to each one. This evens out the distribution of used grey values and thus makes hidden features of the image more visible.

Another contrast enhancement technique is tone mapping [Devlin, 2002]. Tone mapping methods more commonly work with HDR images, mapping them onto the low dynamic range (LDR) domain or using images captured at multiple exposures generate an output image that best captures all the detail. Among the unconventional approaches are those that use wavelets [Laine et al., 1995] and fuzzy logic [Gonzalez and Woods, 2008].

Relatively less work exists using the biological approaches but with increase in computing power and better understanding of the brain, there is a recent upsurge of interest in this domain. [Vonikakis et al., 2008] uses a centre-surround receptive field model for contrast enhancement. [Ling et al., 2007] uses the retinex algorithm to visualise HDR images. Some have developed methods that are based on human perception and contrast sensitivity functions [Huang et al., 2004]. Perhaps the work that comes closest to our approach is [Pattanaik et al., 1998], which used Difference of Gaussians (DoG) to calculate contrast maps and applied some sort of gain control to it. Their approach, however, is more towards computing spatial frequency components at different scales and applying selective gain controls to those.

3.5 Contrast Enhancement of Gray-Scale Images

3.5.1 Proposed Method

Summary

The proposed method of contrast enhancement is summarized in Fig. 3.2. By applying the receptive field contrast operator, the on and off-centre contrast maps

are obtained. These contrast maps are enhanced by applying the adaptive asymmetric contrast enhancement algorithm described shortly. The enhanced contrast maps are used to then reconstruct the contrast-enhanced image.

Details

Convolving the mask with the discretised approximation of DoG in Fig. 2.3(a) with the image and passing the result through the hyperbolic tangent nonlinearity in Fig. 2.3(b):

$$y(x) = \frac{1 - e^{-2x/\alpha}}{1 + e^{-2x/\alpha}} \tag{3.8}$$

the on and off-centre contrast maps of the image are obtained, to which is then applied the asymmetric gain control function.

Psychophysical studies of contrast adaptation have shown that prolonged exposure to high-contrast patterns leads to a decrease in the perceived contrast [Crowder et al., 2006]. This contrast adaptation is often indicated by a right shift of psychometric curves (Fig. 3.3(a)) indicating a decrease of sensitivity to low contrast. The gain control function is modelled by an adaptable sigmoidal function that is a reverse of the contrast adaptation function (Fig. 3.3(b)). The adaptation is based on the contrast values of the image at any pixel (x, y), as the image contrast is modified through iterations. Increase in image contrast at any pixel (x, y) will shift its contrast gain function to the left decreasing its gain, while those pixels whose contrast levels are still low will continue to enjoy a high gain, thus, providing local gain control for every pixel. Gain control G as a function of contrast c at any pixel (x, y) is modelled as:

$$G(c) = (g + \beta) - \frac{(g + \beta - 1)}{1 + e^{-(|c| - \gamma)/\alpha}} \tag{3.9}$$

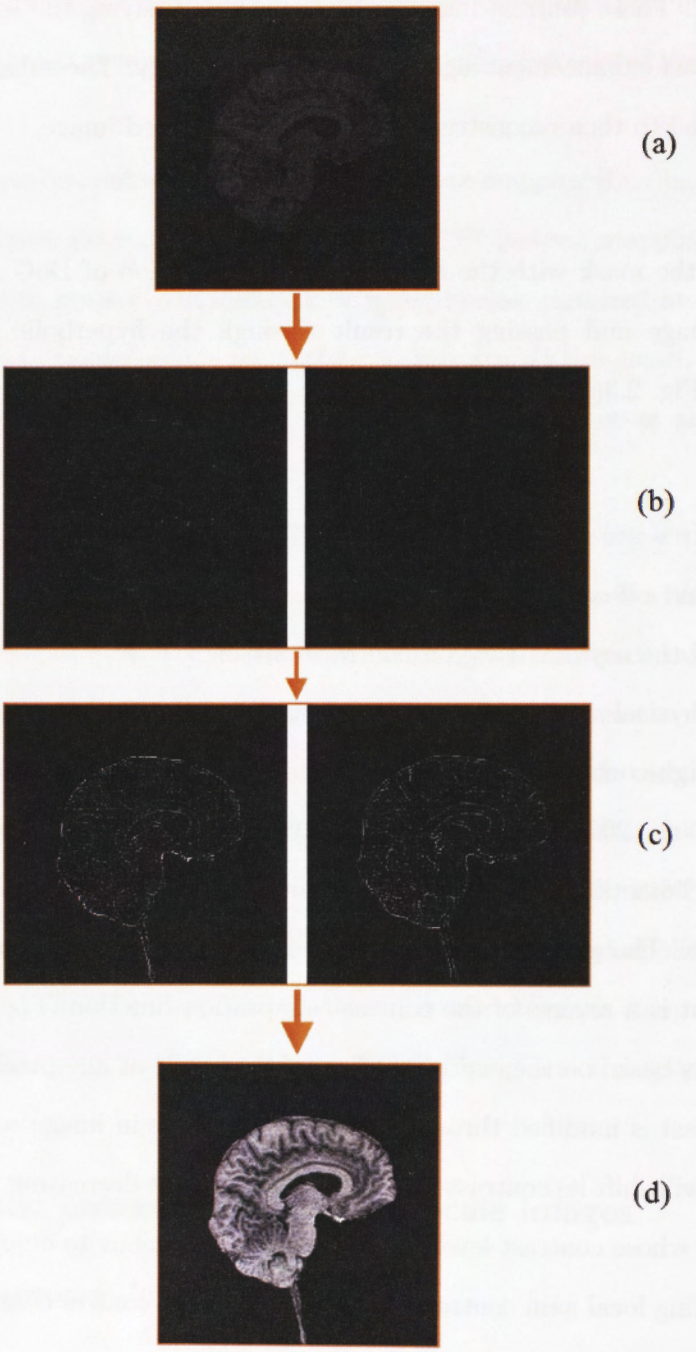


Figure 3.2: The contrast enhancement process. *(a)* Original low contrast image *(b)* On- & off-centre contrast maps obtained from the original image by applying the receptive field contrast operator *(c)* Enhanced contrast maps after performing contrast enhancement *(d)* Reconstructed image from the enhanced contrast maps.

where γ and α are constants that set the initial position of the curve on the contrast axis and its slope, respectively, while $(g + \beta)$ is the maximum value of the gain factor.

β is the *asymmetricity factor*. For zero β, the same gain factor G will be applied to both the on- and off-centre contrast maps. This would give a resultant image with approximately the same mean luminance L_m as the original one. Generally, this is what would be preferred if the L_m of the input image does not fall too much to the extremes. In case the image is too dark or too bright, its L_m would also need adjustment along with its contrast. To shift L_m towards brightness, a higher gain factor is applied to the on-centre map, as compared to the off-centre one. Similarly shifting L_m towards darkness, a higher gain factor is applied to the off-centre map. This is congruent with an on-centre receptive field performing figure-ground separation of bright objects, while the off-centre one does the same for dark objects. Thus, the gain applied to the two maps differs by the value of β, if non-zero:

$$\begin{aligned} \beta_{\text{on}} &= \kappa e^{-(l_m - 255)^2/\sigma} \\ \beta_{\text{off}} &= \kappa e^{-(l_m)^2/\sigma} \end{aligned} \tag{3.10}$$

The curves corresponding to the on- and off-centre β are given in Fig. 3.3(c) and 3.3(d), respectively. κ sets the maximum asymmetricity applied to the maps while σ determines the region extent for which β is non-zero. l_m is the mean local luminance around any given pixel at any particular iteration. As the mean luminance of the image shifts away from the extremes, β eventually becomes zero.

These boosted contrast maps are then used to reconstruct the image using our reconstruction algorithm [Khwaja and Goecke, 2008] that, given a composite contrast map of an image, is able to iteratively reconstruct the image from it. The algorithm stops when the change in image through successive iterations is

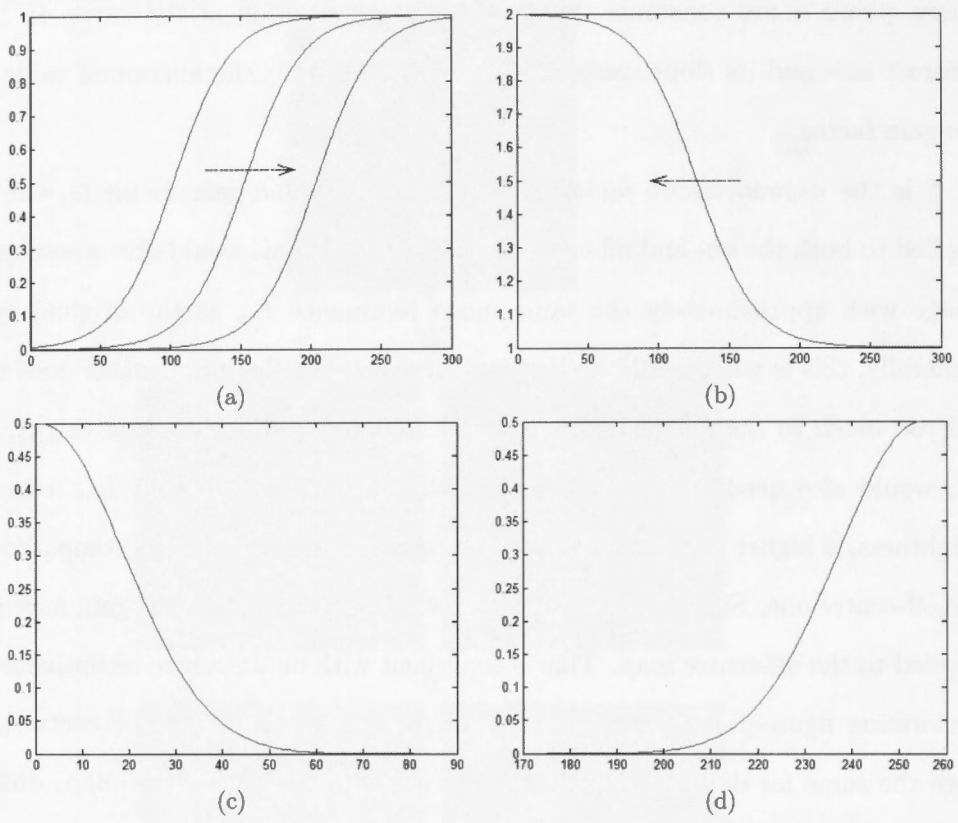


Figure 3.3: (a) Psychometric contrast adaptation curves from [Crowder et al., 2006]. (b) Sigmoidal gain curve used by our algorithm. (c) Left-asymmetricity exponential curve for on-centre map. (d) Right-asymmetricity exponential curve for off-centre map.

below a certain threshold. Reconstruction is accomplished by a gradient descent algorithm using the difference between the original contrast map and the contrast map calculated at each iteration as the error function. The pixel update rule, as in [Khwaja and Goecke, 2008], is:

$$\Delta I_A(x,y) = \eta w_{00} C_E(x,y) \tag{3.11}$$

where ΔI_A is the change in the actual pixel intensity at (x,y), η is the update

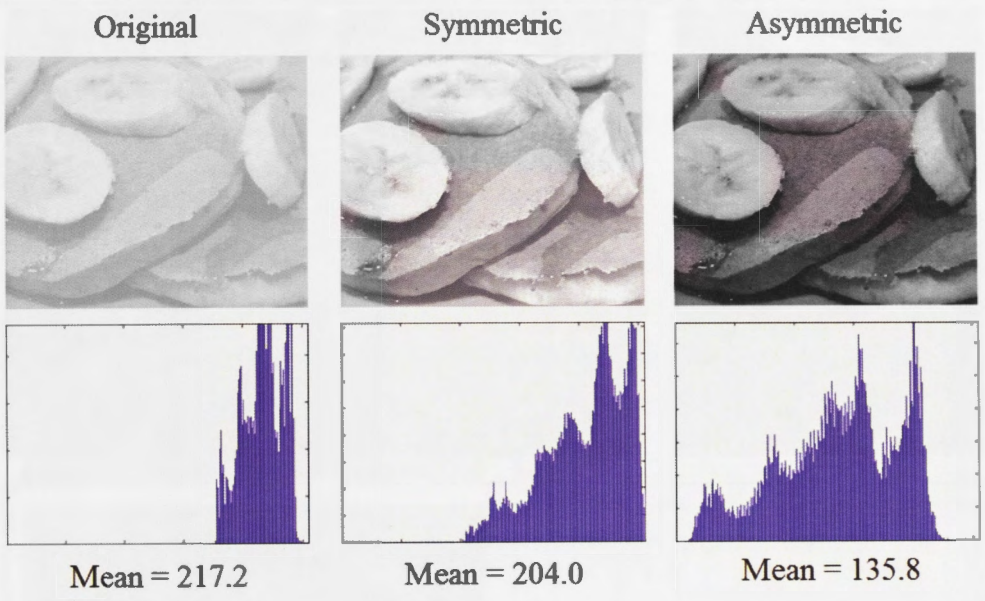


Figure 3.4: The effects of asymmetric gain control on mean luminance control and contrast enhancement. Symmetrically increasing the gain keeps the mean luminance roughly in the vicinity of the original image, which is undesirable in this case. The maximum intensity level being 255, there is no room for contrast expansion on the right hand side. Applying asymmetric gain control pulls the mean luminance to the centre of the spectrum resulting in a better image.

constant and $C_E(x,y) = [C_D(x,y) - C_A(x,y)]$ is the contrast error at pixel (x,y) while w_{00} is the centre weight of the DoG mask. Details of the reconstruction algorithm are in Chapter 2.

▶ *3.5.2 Experiments*

To validate the proposed approach experimentally, the values of the parameters were empirically set to $g = 2.0$, $\gamma = 0.5C_m$ and $\alpha = 0.08C_m$ in Eq. 3.9 and $\kappa = 0.5$ and $\sigma = 2.5C_m$ in Eq. 3.10, where C_m is the maximum possible contrast whose value depends upon the mapping performed by the tangent hyperbolic non-linearity. In this case, C_m is mapped to a value of 255. Some results are shown in Fig. 3.6, 3.5 and 3.8. Although image quality is a subjective and context

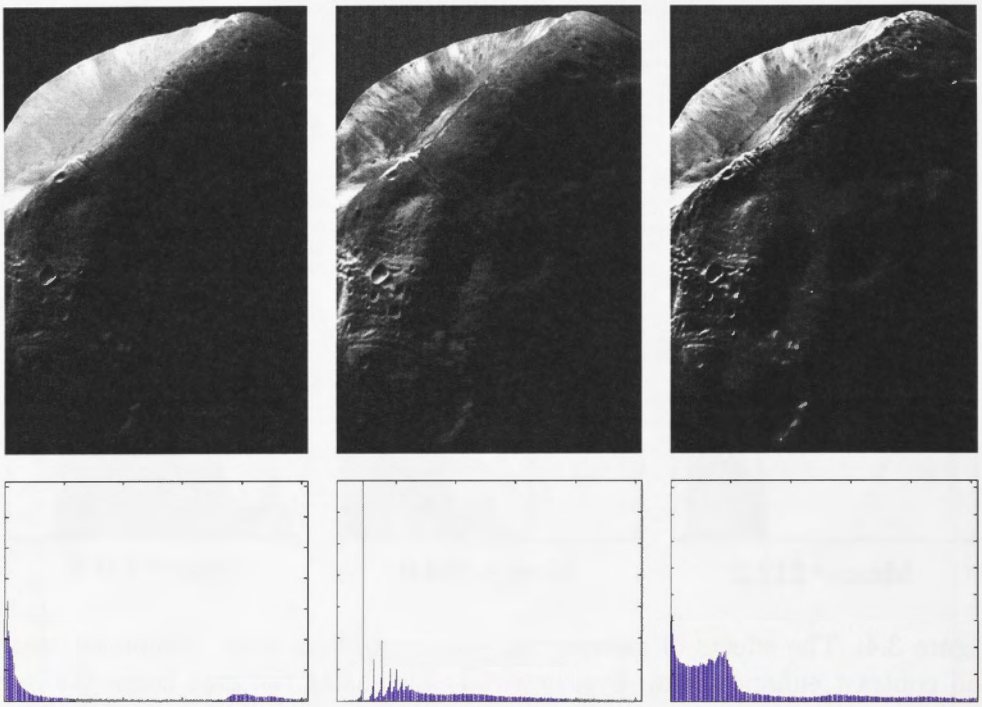


Figure 3.5: *(Top)*: Example moon image (left), CLAHE output (middle) and output of our algorithm (right) *(Bottom)*: Corresponding histograms.

dependent measure which is very hard to quantify, a few necessary, though not sufficient, parameters can be identified that are indicative of good image contrast, namely, the spread of the image histogram and its density. It can be seen from the histograms of the examples that their contrast spreading is wider than that of CLAHE and that they are quite dense with no visible gaps. Also, the developed algorithm is not known to produce any perceivable halo effect that some of the other contrast enhancement algorithms generate.

3.6 Contrast Enhancement of Colour Images

Next, the same gain control model as developed in Section 3.5.1 was applied to colour images. Three different colour spaces were experimented with and their results are compared:

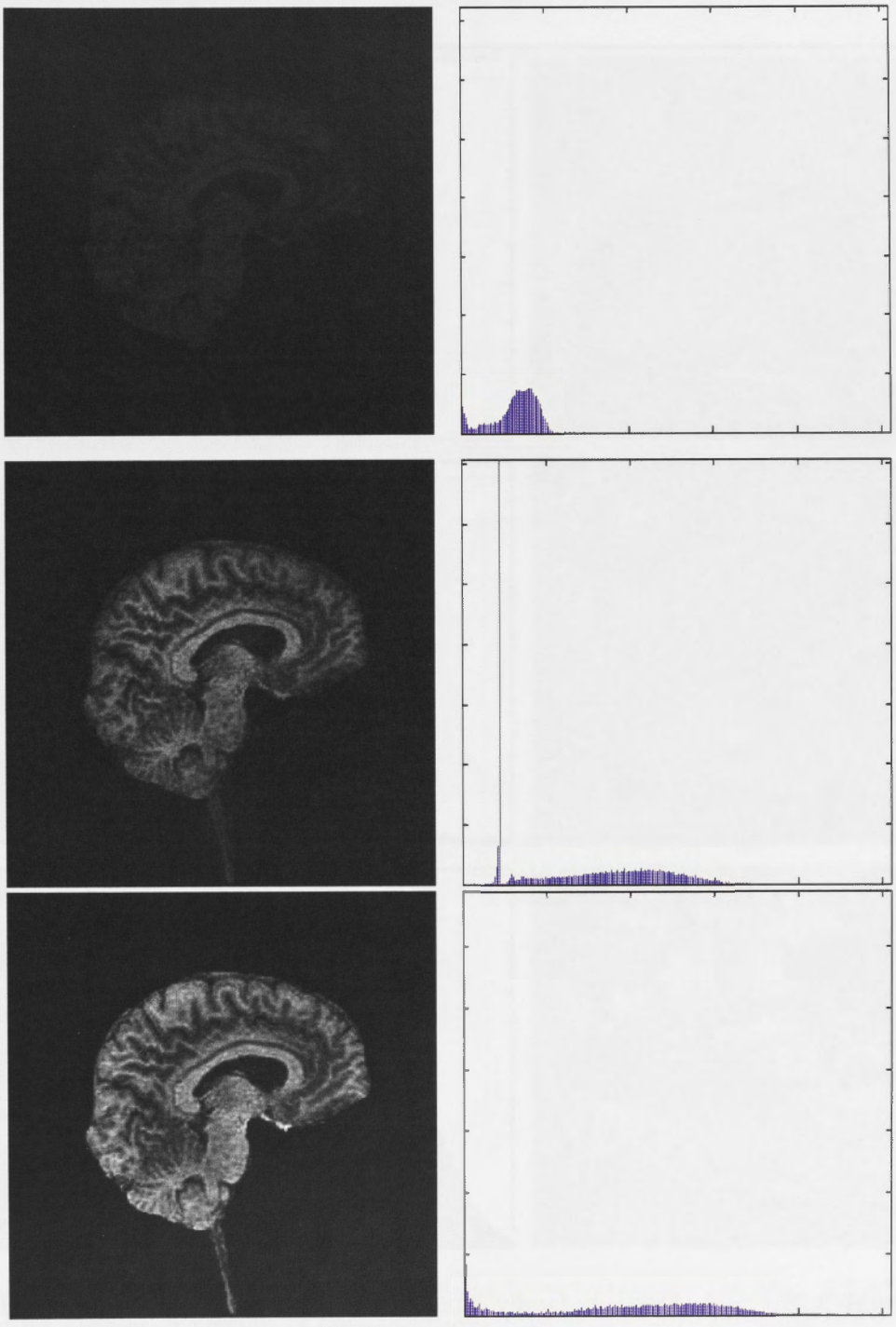


Figure 3.6: *(Top)*: Example brain image & its histogram. *(Middle)*: CLAHE output. *(Bottom)*: Output of our algorithm.

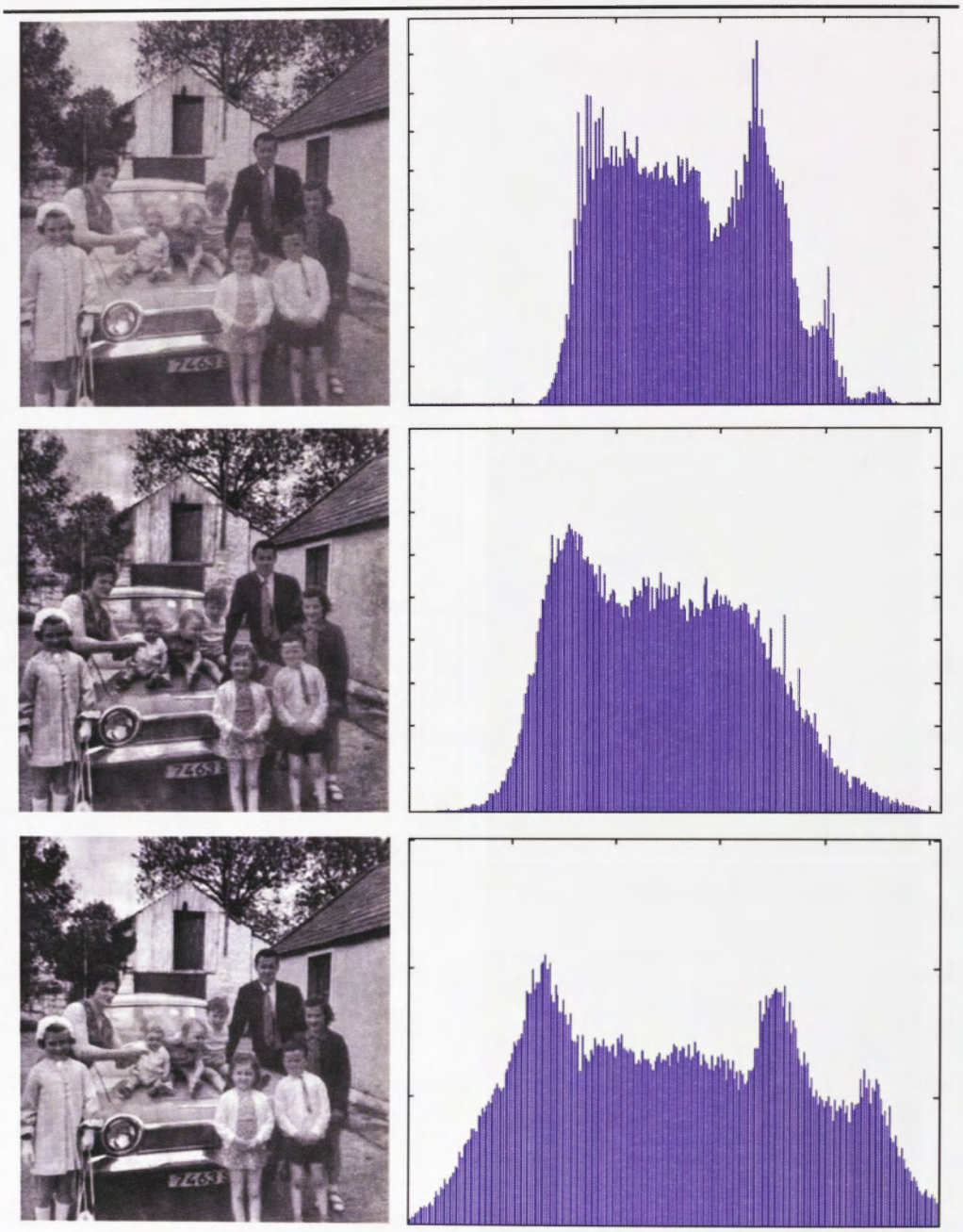


Figure 3.7: *(Top)*: Example family image & its histogram. *(Middle)*: CLAHE output. *(Bottom)*: Output of our algorithm.

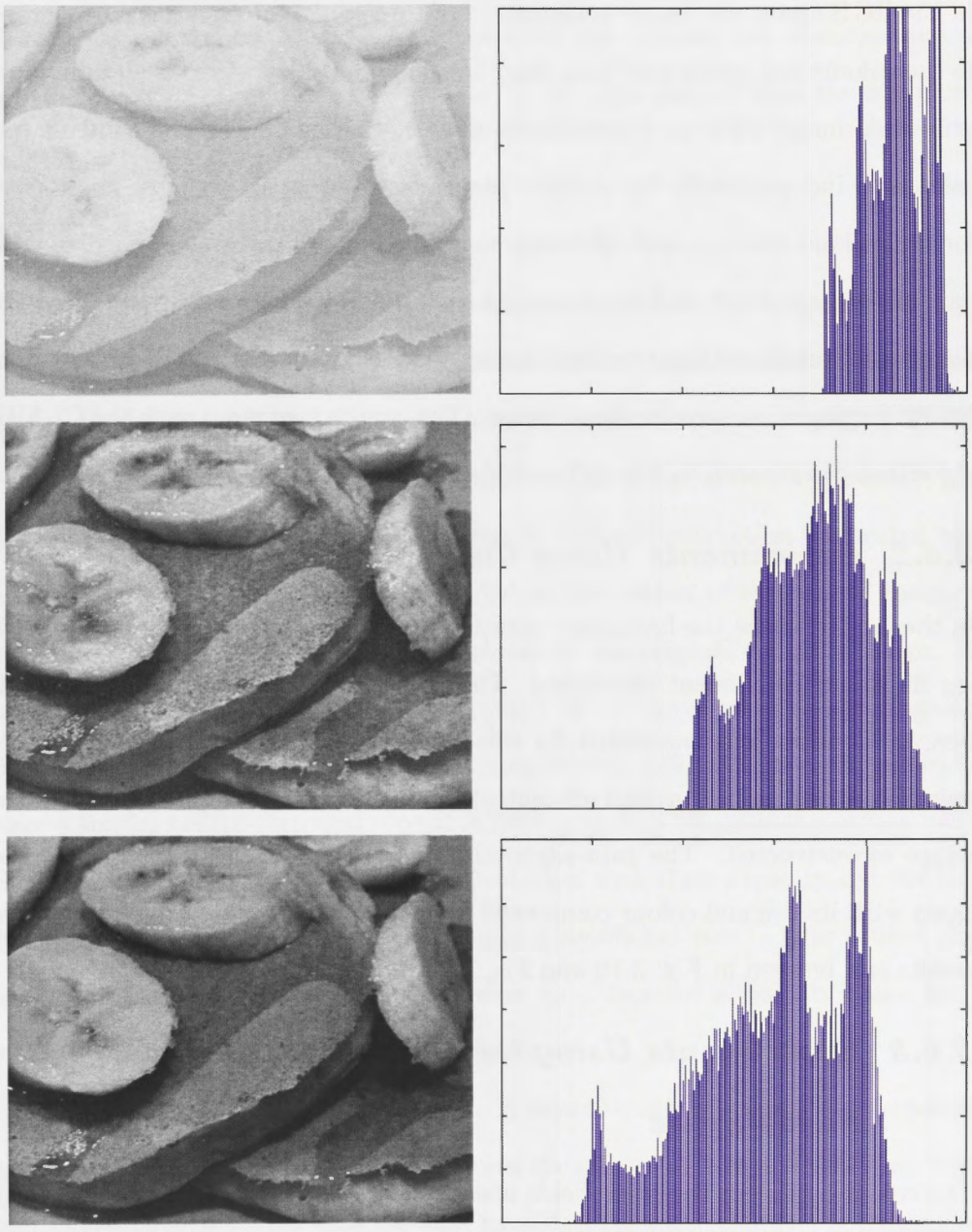


Figure 3.8: *(Top)*: Example pancake image & its histogram *(Middle)*: CLAHE output *(Bottom)*: Output of our algorithm.

▶ 3.6.1 Experiments Using the RGB Space

In the RGB space, the image is composed of three different planes corresponding to the colours red, green and blue. Each image plane was treated as an individual gray-scale image with no dependencies with any other plane. The contrast was calculated independently for all three planes resulting in six contrast maps - two for each plane (the on- and off-centre maps). Gain control was applied to each pair of contrast maps and the image of each plane was reconstructed using the reconstruction algorithm described earlier. The three reconstructed images were finally displayed as a single colour image. The results compared with the CLAHE algorithm can be seen in Fig. 3.9 and Fig. 3.12.

▶ 3.6.2 Experiments Using the $L^*a^*b^*$ Space

In the L*a*b* space the luminance component is extracted from the image leaving its colour component untouched. This luminance component is a gray-scale image. Contrast was calculated for this single gray-scale image giving a single pair of contrast maps (on- and off-centre). Gain control was applied to it and the image reconstructed. The gain-adjusted reconstructed image was then merged again with its original colour component to produce the final colour image. The results can be seen in Fig. 3.10 and Fig. 3.13.

▶ 3.6.3 Experiments Using the Colour-Opponent Receptive Fields Space

The colour-opponent receptive fields are inspired from the primate brain. They are found in the fovea of the retina where there is a majority of colour-sensitive cells called cones. Six contrast maps are obtained in here as well like that of the RGB space with the difference that while in the RGB space each of the pair of contrast maps were computed treating each colour independently from

the other two, in here the contrast maps are computed such that each contrast map is dependent on two or more colours acting in opposition to the other. Because of this opponent dependency between the colours, the result using the colour-opponent receptive fields come out to be more natural than the other two approaches (see Fig. 3.11 and Fig. 3.14).

3.7 Discussion

Although image reconstruction is not what goes on in the brain to the best of our knowledge as the brain deconstructs rather than constructs the input image coming through the retina which is an important step in analysis, it (the reconstruction) is used as a central vehicle to comprehend and hypothesize the functional units of the brain. In Chapter 2, this reconstruction apparatus was used to demonstrate that the image coded at the output of the retinal ganglion cells does indeed retain the full information of the original receptor image. In here, the gain control model that is applied on to the contrast maps proposes one of the possible roles of the LGN among others. LGN cells were reported to have a similar centre-surround receptive field as that of the retinal ganglion cells. Since they are modelled as doing a convolution with their input image, for this to be true, it is speculated that LGN plays a significant role in gain control and selective attention. This also makes sense as it receives about six times more connections from the cortex as it sends to it.

Gain control is also a key operation in selective attention tasks in which focussing on a desired object or scene reduces the attention to everything else. This can be restated by saying that the gain to all but the desired object is reduced to a level that is needed for its proper perception and then action. This gain control is an ever-happening phenomenon and in fact many of the visual illusions can be explained merely by the action of global or selective gain control. As the cortex

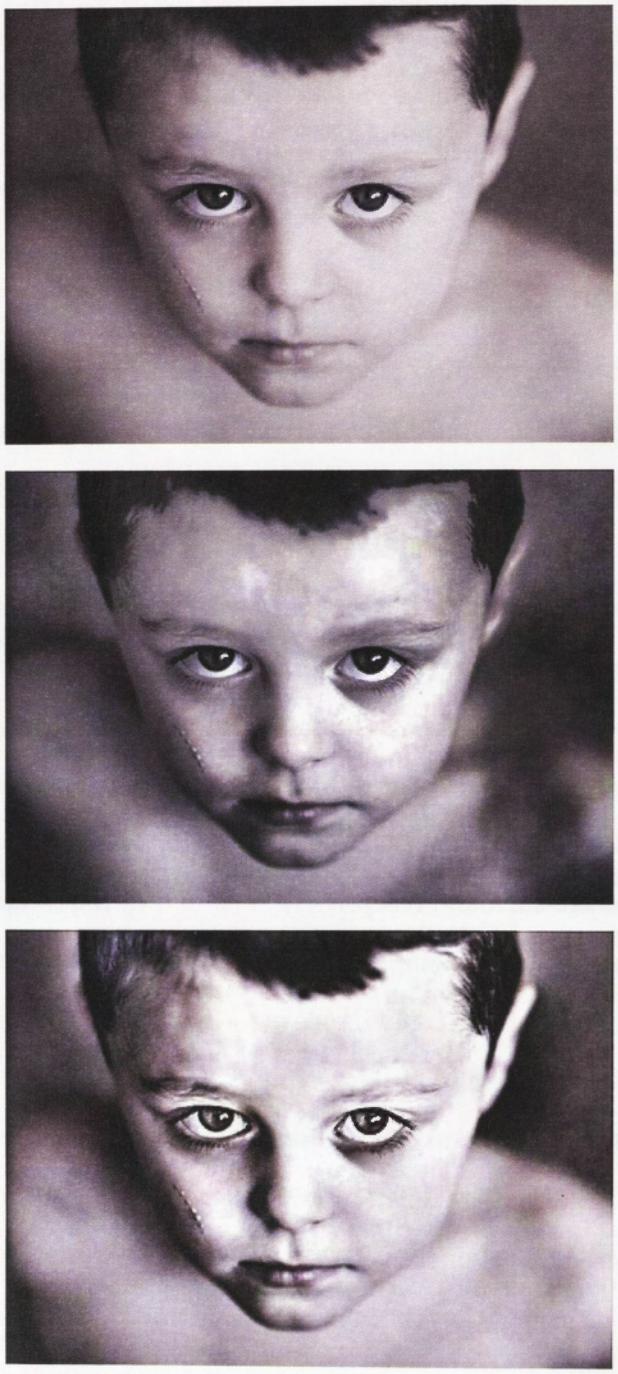


Figure 3.9: *(Top)*: Example boy image. *(Middle)*: CLAHE output in the RGB space. *(Bottom)*: Output of the proposed algorithm in the RGB space.

Figure 3.10: *(Top)*: Example boy image. *(Middle)*: CLAHE output in the L*a*b* space. *(Bottom)*: Output of the proposed algorithm in the L*a*b* space.

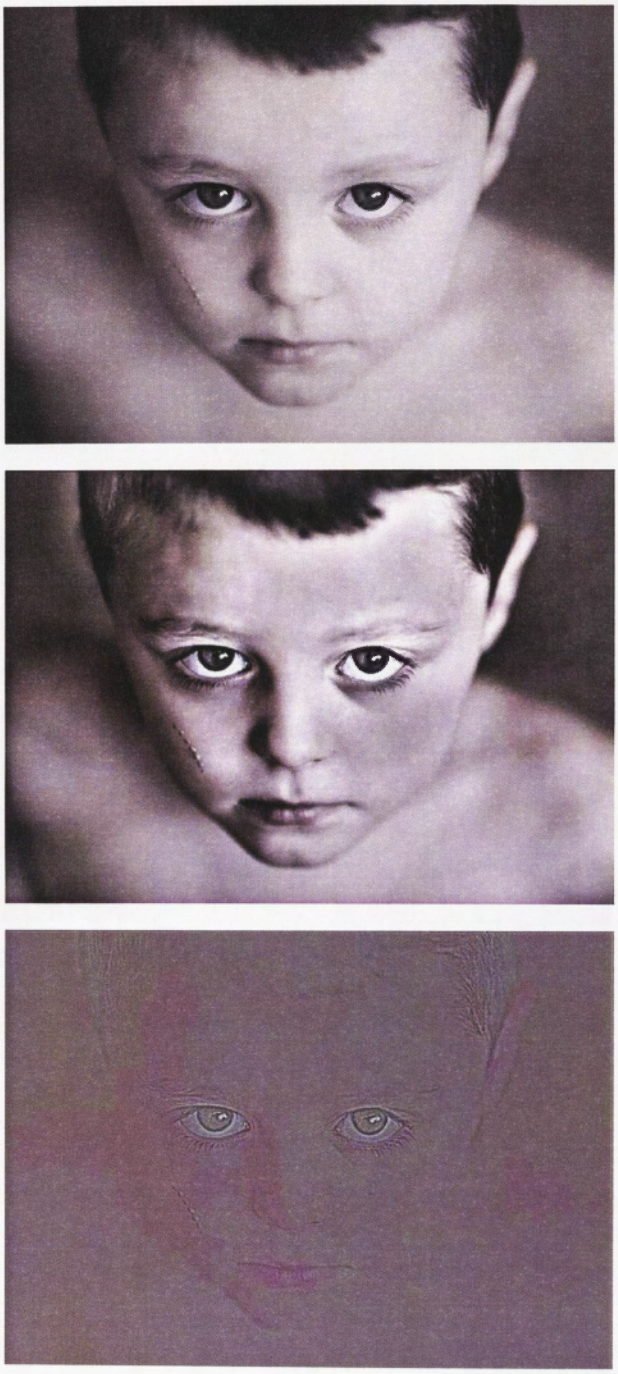


Figure 3.11: *(Top)*: Example boy image. *(Middle)*: Output of the proposed algorithm using the colour-opponent receptive fields. *(Bottom)*: Corresponding contrast image

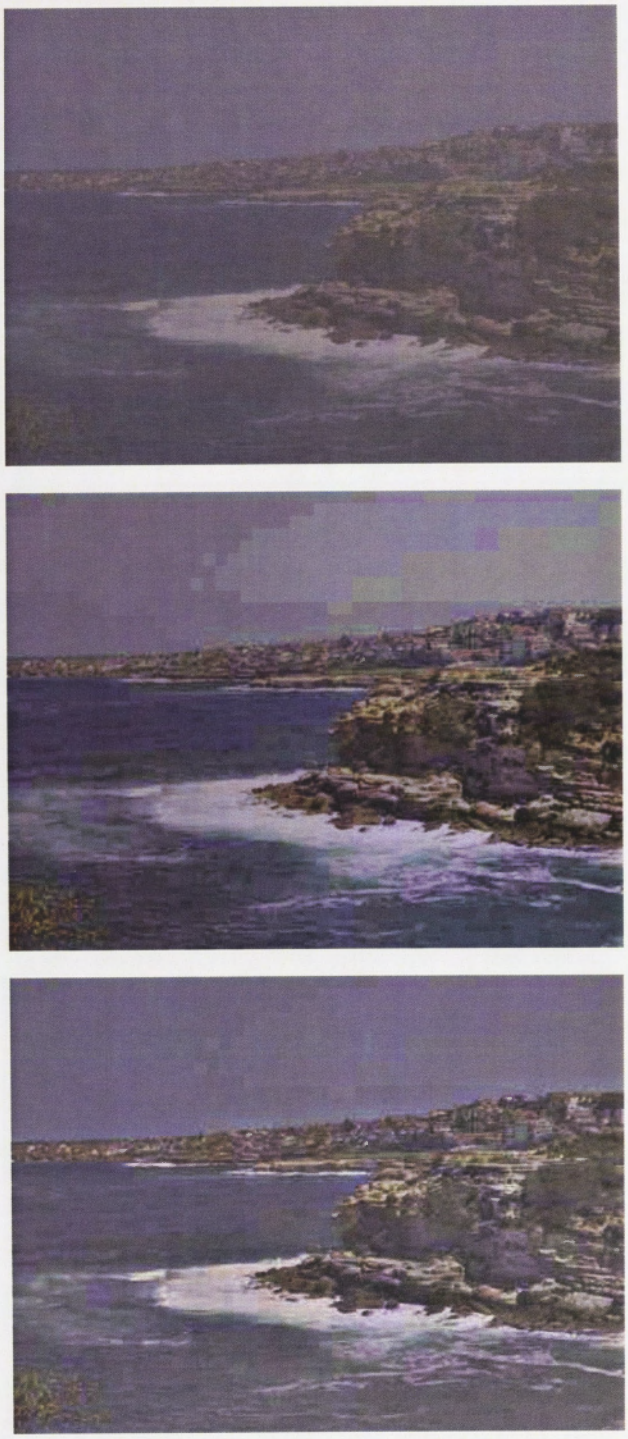


Figure 3.12: *(Top)*: Example seaside image. *(Middle)*: CLAHE output in the RGB space. *(Bottom)*: Output of the proposed algorithm in the RGB space.

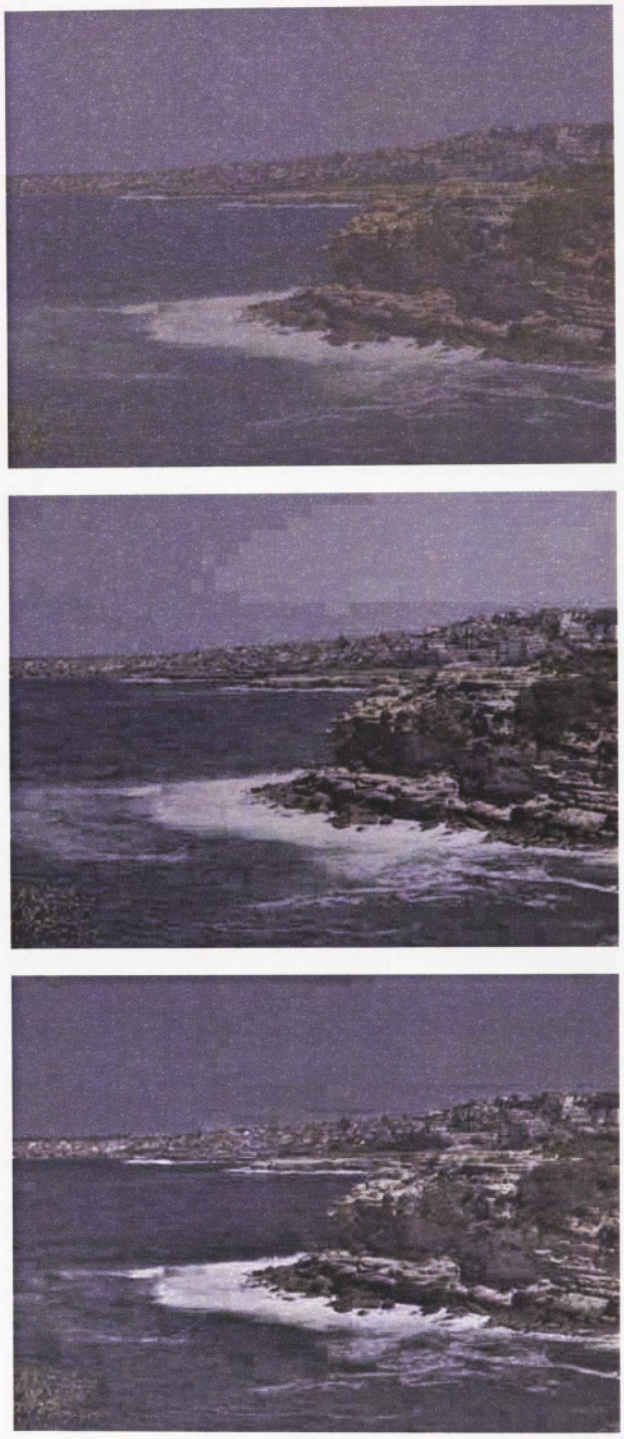


Figure 3.13: *(Top)*: Example seaside image. *(Middle)*: CLAHE output in the $L^*a^*b^*$ space. *(Bottom)*: Output of the proposed algorithm in the $L^*a^*b^*$ space.

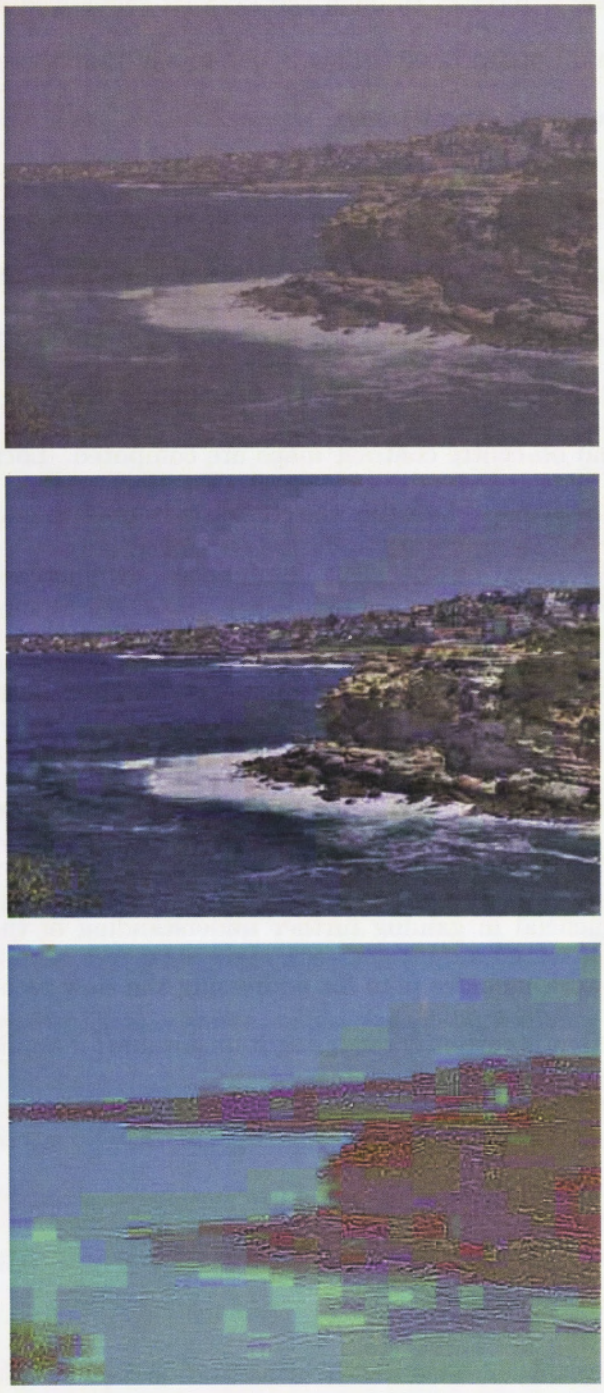


Figure 3.14: *(Top)*: Example seaside image. *(Middle)*: Output of the proposed algorithm using the colour-opponent receptive fields. *(Bottom)*: Corresponding contrast image

analyses the scene it receives through the LGN, it sends feedback to it thereby adjusting the gain appropriately. This is not to say that there cannot be local gain control occurring in the V1 area.

▶ 3.8 Conclusion

A model of contrast enhancement is developed using the image reconstruction algorithm of Chapter 2 that is biologically inspired. A key feature of this model is the control of the mean luminance value of the image. Starting with the original image, its on- and off-centre contrast maps are computed. Local gain control is then applied to these maps, boosting the contrasts in the two maps symmetrically or asymmetrically based on the mean luminance. Symmetrical boosting keeps the mean luminance roughly the same, while enhancing the contrast, whereas asymmetrical boosting shifts the mean luminance towards the stronger weighted side. A higher gain on the off-centre map shifts the mean luminance towards the darker values, while that on the on-centre moves it towards brighter ones. The model is also biologically plausible since it uses only local computations and, thus, can be beneficial in gaining further understanding of the primate visual system. Future work includes plan for addressing the slow reconstruction phase of the algorithm that is currently rendering it unsuitable for applications requiring quick enhancements.

4

Face Recognition as a Sparse Representation Problem

The sparse representation technique has provided a new way of looking at object recognition [Wright et al., 2009]. As is demonstrated here, however, the mean-squared error (MSE) measure, which is at the heart of this technique, is not a very robust measure when it comes to comparing images, in particular images which differ significantly in luminance values, as it only performs pixel-by-pixel comparisons. This requires a significantly large training set with enough variations in it to offset the drawback of the MSE measure. A large training set, however, is often not available. A replacement of the MSE measure by the structural similarity (SSIM) measure in the sparse representation algorithm is proposed, which performs a more robust comparison using only one training sample per subject. In addition, since the off-the-shelf sparsifiers are also written using the MSE measure, a new sparsifier is developed using genetic algorithms (SparsGA) that use the SSIM measure. One of the main issues in any optimization algorithm is the satisfaction of multiple objectives, as generally all but the very trivial problems are multi-objective in nature, leading to suboptimal convergence. This concern is addressed by employing a novel non-linear coding of the problem parameters thus transforming a multi-objective problem to a uni-objective one. The modi-

fied algorithm is applied to the Extended Yale Face B database as well as to the Multi-PIE database with expression and illumination variations. The improved performance demonstrates the effectiveness of the proposed modifications. Further the sparse representation classification technique is analysed objectively for pros and cons. The research in this chapter is based on the work of Wright et al. [Wright et al., 2009].

▶ 4.1 Salient Contributions of This Research

- Replacement of the mean-squared error measure in the sparse representation classification algorithm by a more robust structural similarity measure.
- Development of a genetic algorithm based sparsifier (SparsGA) which can work with literaly any image comparison measure.
- Development of a novel non-linear coding of genetic algorithm parameters for the transformation of multi-objective optimization problem to a uni-objective one.
- The inclusion of a pre-processing stage to offset image degradation before the recognition stage.
- An analysis and critique of the sparse representation classification technique - its pros and cons.
- Part of the research presented here has been published [Khwaja and Goecke, 2009].

▶ 4.2 Introduction

A novel and comprehensive approach for face recognition was recently proposed based on the sparse representation theory [Wright et al., 2009]. Here, the face

recognition problem is treated as searching for a sparse representation of a given test image in terms of the training images. Each face is represented by a set of features sufficiently characterizing each individual. With the prior knowledge that faces of the same individual are similar to each other, a test face can be considered as being approximated by a linear combination of the training images of the same individual. A test face ideally, thus, has an inherent sparse representation in the training set. Given enough training images per individual and enough features extracted per image, it promises significant data corruption and occlusion tolerances. The technique is also independent of the kind of features used. Almost any feature, including random transformations, can be used to obtain recognition rates of similar accuracy, provided the number of features is sufficient. In addition, the technique is claimed not to require any pre-processing beforehand.

In spite of the promising advantages, the sparse representation technique does not deal effectively with all the practical issues that arise in face recognition. Firstly, it is highly sensitive to image alignment in both translation and rotation [Huang et al., 2008]. Secondly, it cannot handle pose variations. Thirdly, it is sensitive to illumination and facial expression variations [Li et al., 2009]. Some of these can be offset by a large training set covering all the expected variations in terms of sample images. However, such comprehensive training sets are not generally feasible to obtain for realistic tasks.

In this research, modifications are proposed to the original sparse representation algorithm to improve its performance under illumination and expression variations given only one training sample per individual. The main contributions of this research are the following: 1) Potential problems are identified in the error measure used in Eq. 4.7 & 4.8 and propose to replace it by another error measure called the *Structural Similarity* Index. 2) Issues are identified with the off-the-shelf sparsifiers or l_1 minimization algorithms and a genetic algorithm

based sparsifier is developed. 3) A pre-processing stage is added and it is asserted that when the images are expected to be degraded, it is vital to pre-process images before attempting any recognition. The alignment issues, however, are not addressed and it is assumed that the images provided are already sufficiently aligned.

▶ 4.3 Sparse Representation - Background and Related Work

Sparse representations are representations or coding schemes that account for most or all the information of a signal with a linear combination of a small number of elementary signals called *basis functions* or *atoms*. Often, the basis functions are chosen from a so called over-complete dictionary. Formally, an over-complete dictionary is a collection of basis functions such that the number of these functions exceeds the dimension of the signal space. Thus, this is an over-representation approach where more modelling parameters exist than are needed. Hence, any signal can be represented by more than one combination of different basis functions. Recent results in brain research show that the brain makes heavy use of sparseness within itself for information representation. This could well be one of the reasons of its robust and fault-tolerant behaviour.

Sparseness is one of the reasons for the extensive use of popular transforms such as the Discrete Fourier Transform, the wavelet transform and the Singular Value Decomposition. The aim of these transforms is often to reveal certain structures of a signal and to represent these structures in a compact and sparse representation. Sparse representations have therefore increasingly become recognized as providing extremely high performance for applications as diverse as: noise reduction, compression, feature extraction, pattern classification and blind source separation. Sparse representation ideas also build the foundations of wavelet denoising and methods in pattern classification, such as in the Support

Vector Machine [Boser et al., 1992] and the Relevance Vector Machine [Tipping, 2001], where sparsity can be directly related to learnability of an estimator.

The technique of finding a representation with a small number of significant coefficients is often referred to as *Sparse Coding*. Decoding merely requires the summation of the relevant basis, appropriately weighted, however, unlike a transform coder with its invertible transform, the generation of the sparse representation with an over-complete dictionary is non-trivial. Indeed, the general problem of finding a representation with the smallest number of basis from an arbitrary dictionary has been shown to be NP-hard [Amaldi and Kann, 1999]. This has led to considerable effort being put into the development of many sub-optimal schemes. These include algorithms that iteratively build up the signal approximation one coefficient at a time [Alder et al., 1997], e.g. Matching Pursuit [Berinde et al., 2008], Orthogonal Matching Pursuit [Blumensath and Davies, 2007], and those that process all the coefficients simultaneously, e.g. Basis Pursuit [Chen et al., 2008], Basis Pursuit De-Noising [Rakotomamonjy] and the Focal Underdetermined System Solver family of algorithms [Gorodnitsky and Rao, 1997].

The sparse representation classification (SRC) presented in [Wright et al., 2009], on which this research is based on, consider face recognition with frontal face images from k individuals that have been properly cropped and normalized. Each image is of size $w \times h$ and can be viewed as a point in the space $\mathbb{R}^m$ with $m = w \times h$. The n_i images associated with subject i are stacked as vectors $\boldsymbol{v}_{i,1}, \boldsymbol{v}_{i,2}, \ldots, \boldsymbol{v}_{i,n_i} \in \mathbb{R}^m$. Any new test image of the same subject, stacked as a vector $\boldsymbol{y} \in \mathbb{R}^m$, can be represented as a linear superposition of the gallery images associated with subject i:

$$\boldsymbol{y} = \alpha_{i,1}\boldsymbol{v}_{i,1} + \alpha_{i,2}\boldsymbol{v}_{i,2} + \ldots + \alpha_{i,n_i}\boldsymbol{v}_{i,n_i} \tag{4.1}$$

for some scalars $\alpha_{i,j} \in \mathbb{R}, j = 1, \ldots, n_i$. Collecting all the n_k gallery images as column vectors of one matrix:

$$A \doteq [\boldsymbol{v}_{i,1}, \boldsymbol{v}_{i,2}, \ldots, \boldsymbol{v}_{i,n_1}, \boldsymbol{v}_{2,1}, \ldots, \boldsymbol{v}_{k,n_k}] \in \mathbb{R}^{m \times n} \tag{4.2}$$

Then ideally the test image $\boldsymbol{y}$ of subject i can be represented in terms of all the images in the training set as:

$$\boldsymbol{y} = A\boldsymbol{x}_0 \in \mathbb{R}^m \tag{4.3}$$

where $\boldsymbol{x}_0 = [0, \ldots, 0, \alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,n_i}, 0, \ldots, 0]^T \in \mathbb{R}^n$ is a coefficient vector whose entries are mostly zero except those associated with the i^{th} subject. The following simple observation is therefore exploited: A valid test image can be sufficiently represented using only the gallery images of the same subject. This representation is naturally sparse if the number of subjects k is reasonably large.

Now, since the data dimensions are *very high*, dimension reduction techniques need to be used to reduce the data dimension m to $d \ll m$. There are myriad such techniques such as Eigenfaces [Turk and Pentland, 1991], Fisherfaces [Belhumeur et al., 1997], Laplacianfaces [He et al., 2005] or even random projections [Baraniuk and Wakin, 2007] projecting high-dimensional face images to low-dimensional feature spaces. The authors compare results for these dimension reduction or feature extraction techniques and report that for the SRC algorithm the choice of a particular feature extraction technique is irrelevant and selected to employ random projections for its ease of use and low computational cost. For any face feature extraction technique, the projection from the image space to the feature space can be represented as a matrix $R \in \mathbb{R}^{d \times m}$ with $d \ll m$:

$$\tilde{\boldsymbol{y}} \doteq R\boldsymbol{y} = RA\boldsymbol{x} = \tilde{A}\boldsymbol{x}_0 \in \mathbb{R}^d \tag{4.4}$$

This desired sparse coefficient vector $\boldsymbol{x}_0$ is obtained as a solution to the following optimization problem:

$$min\|\boldsymbol{x}\|_0 \quad subject\ to \quad \tilde{\boldsymbol{y}} = \tilde{A}\boldsymbol{x} \tag{4.5}$$

where $\|.\|_0$ denotes the l^0 norm, which simply counts the number of nonzero entries in a vector. Solving Eq. 4.5 is NP-hard and even difficult to approximate by polynomial-time algorithms [Amaldi and Kann, 1999]. In the general case, no known procedure for finding the sparsest solution is significantly more efficient than exhausting all subsets of the entries for $\boldsymbol{x}$.

Recent development in the emerging compressed sensing theory [Donoho, 2006, Candes et al., 2006, Candes and Tao, 2006] reveals that if the solution $\boldsymbol{x}_0$ sought is sparse enough, the combinatorial problem of Eq. 4.5 is equivalent to the following l^1 minimization problem:

$$min\|\boldsymbol{x}\|_1 \quad subject\ to \quad \tilde{\boldsymbol{y}} = \tilde{A}\boldsymbol{x} \tag{4.6}$$

This problem can be solved in polynomial time by standard linear programming or quadratic programming methods [Chen et al., 2008]. Several algorithms exist that allow recovery of sparse signals via l^1 minimization. SpaRSA [Wright et al., 2008], l^1-Magic [Candes et al., 2006], and Sparsify [Blumensath and Davies, 2007] are some of the solvers that are available free of charge and can be used along with the SRC algorithm under discussion. Since real data are noisy, it may not be possible to express the test sample exactly as a sparse superposition of the gallery samples. The Eq. 4.6 can thus be modified to explicitly account for this by employing an l_2 norm:

$$min\|\boldsymbol{x}\|_1 \quad subject\ to \quad \|\boldsymbol{y} - A\boldsymbol{x}\|_2 \leqslant \epsilon \tag{4.7}$$

Algorithm 2 Sparse Representation-based Classification (SRC) - Original

1: **Input:** a matrix of gallery images $A = [A_1, A_2, \ldots, A_k] \in \mathbb{R}^{m \times n}$ for k classes, a test sample $\boldsymbol{y} \in \mathbb{R}^m$, (and an optional error tolerance $\epsilon > 0$).
2: Normalize the columns of A to have unit l^2 norm.
3: Solve the l^1 minimization problem:

$$min \|\boldsymbol{x}\|_1 \quad subject\ to \quad \|\boldsymbol{y} - A\boldsymbol{x}\|_2 \leqslant \epsilon$$

using available solvers (like SpaRSA, l^1-Magic, Sparsify etc).
4: Compute the residuals $r_i(\boldsymbol{y}) = \|\boldsymbol{y} - A\delta_i(\boldsymbol{x})\|_2 \quad for \quad i = 1, \ldots, k.$
5: **Output:** $identity(\boldsymbol{y}) = min_i[r_i(\boldsymbol{y})]$.

where y is the probe image, A is the matrix of gallery images and x is the sparse vector sought for. The sparse coefficients returned in x are compared again for minimum residuals using the same l_2 norm:

$$r_i(\boldsymbol{y}) = \|\boldsymbol{y} - A\delta_i(\boldsymbol{x})\|_2 \quad for \quad i = 1, \ldots, k \tag{4.8}$$

The image class corresponding to the smallest residual is taken to be the class of the probe image. The above discussion can be summarized as in Algorithm 2.

Modifications in this algorithm have been proposed by a few in different directions. The alignment sensitivity of this algorithm was explored by [Huang et al., 2008] and a modified algorithm was proposed that attempted to align the images simultaneously while retrieving their class. This required a threefold increase in gallery matrix size for only translation invariance (for up to 7 pixels) and sixfold for affine invariance (for 2°, 4°, and 6° rotations).

The expression sensitivity of the sparse representation approach was investigated by [Li et al., 2009], which used 3D surface meshes and a feature pooling and ranking scheme. By applying the sparse representation framework to the collected low-level features, they report satisfactory results.

Here, results are provided that reveal the sensitivity of the algorithm to illu-

mination variations in particular and expressions in general. Experiments conducted with two different databases show the poor performance of the sparse representation algorithm [Wright et al., 2009] under lighting changes requiring significant gallery set size with different variations to offset it. Modifications are proposed in the algorithm for improved handling of such illumination variations and expressions under a single gallery image per individual.

▶ 4.4 The Error Measure

Both Eq. 4.7 & 4.8 in the original SRC algorithm use the mean squared error (MSE) or l_2 norm as the preferred error measure. MSE has traditionally been the dominant quantitative performance metric in the field of signal processing and remains the method of choice for comparing competing signal processing methods and for the assessment of signal quality and fidelity. This is true despite the fact that in many of these applications, the MSE exhibits weak performance and has been widely criticized for serious shortcomings, especially when dealing with perceptually important signals such as speech and images. Yet it continues to receive significant attention and usage due to many attractive features it possesses:

1. It is simple, parameter free and inexpensive to compute.

2. It is also memory less — squared error can be evaluated at each sample, independent of other samples.

3. All l_p norms are valid distance metrics in $\mathbb{R}^N$, which allow for consistent interpretation of similarity by satisfying the following convenient conditions:

   - nonnegativity: $l_p(x, y) \geqslant 0$
   - identity: $l_p(x, y) = 0$ if and only if $x = y$

- symmetry: $l_p(x, y) = l_p(y, x)$
- triangular inequality: $l_p(x, z) \leqslant l_p(x, y) + l_p(y, z)$

4. It has a clear physical meaning — it is the natural way to define the energy of an error signal.

5. The MSE is also an excellent metric in the context of optimization due to possessing the very satisfying properties of convexity, symmetry, and differentiability.

6. Finally, the MSE is widely used simply because it is a convention, providing a convenient standard against which the results of new algorithms may be compared to the old ones.

Despite these favourable properties which are attractive predominantly for computational conveniences, the MSE has long been criticized for its poor correlation with perceived image quality [Pappas et al., 2005, Wang and Bovik, 2006, Wu and Rao, 2005, Girod, 1993]. The wherefores for this quandary lies in the following assumptions implicit in the use of MSE:

1. Signal fidelity is independent of any spatial relationship between image signal samples. As a result, changing the spatial ordering of the image signal samples does not affect the distortion measurement

2. Signal fidelity is independent of any relationships between the image signal and the error signal. As a result, for the same error signal, no matter what the underlying image signal is, the distortion measure remains the same.

3. Signal fidelity is determined by the magnitude of the error signal only. As a result, changing the signs of the error signal samples has no effect on the distortion measurement.

4. All signal samples are of equal importance to signal fidelity.

Unfortunately, not one of these assumptions holds when signal fidelity is taken to be the perceptual image quality which is a highly desirable metric in image processing and computer vision applications. Failure of each of the above assumptions have convincingly been demonstrated by illustrative examples [Wang and Bovik, 2009, 2006]. From a perceptual outlook as well, we can see that the MSE is completely blind to any underlying structure in the image since it determines the best fit by very simple, isolated pixel comparisons. This lack of any cognition for structure makes it highly sensitive to illumination variations and noise.

Lately, there has been considerable progress in our understanding of the functions of our perceptual systems and in incorporating psychophysical and neurophysiological findings into mathematical models of these functions and while we are still at a stage of infancy in this domain, it is becoming evident that herein lies a promising future. One recently proposed approach to image fidelity measurement that is compatible to human perception is the Structural SIMilarity (SSIM) index [Wang et al., 2005, 2004]. It is, hence, proposed that the MSE measure in both Equations 4.7 and 4.8 be replaced by the SSIM measure. Hence, the modified equations can now be seen as:

$$min\|\boldsymbol{x}\|_1 \quad subject\,to \quad SSIM(\boldsymbol{y}, A\boldsymbol{x}) \leqslant \epsilon \tag{4.9}$$

$$r_i(\boldsymbol{y}) = SSIM(\boldsymbol{y} - A\delta_i(\boldsymbol{x})) \quad for \quad i = 1, \ldots, k \tag{4.10}$$

The SSIM performs three different local similarity measurements of luminance, contrast and structure on image patches, and thereafter combines them to obtain

a single local SSIM index:

$$SSIM(x,y) = [l(x,y)]^{\alpha}.[c(x,y)]^{\beta}.[s(x,y)]^{\gamma} \tag{4.11}$$

where $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ are parameters used to adjust the relative importance of the three components and x and y are the compared images. These three components are defined in terms of the mean luminance, standard deviation and cross correlation:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{{\mu_x}^2 + {\mu_y}^2 + C_1}, \tag{4.12}$$

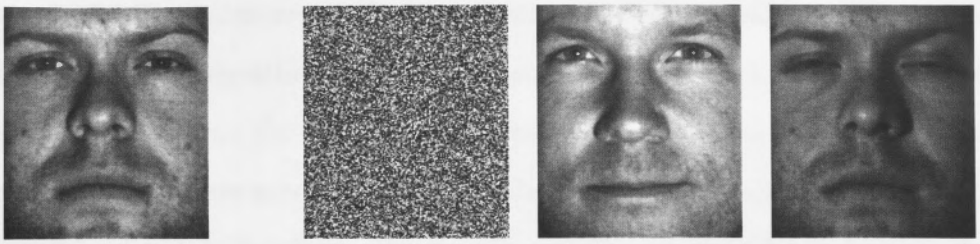
$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{{\sigma_x}^2 + {\sigma_y}^2 + C_2}, \tag{4.13}$$

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{4.14}$$

where σ is the standard deviation and μ is the mean.

The SSIM is symmetrical ($S(x,y) = S(y,x)$), bounded ($S(x,y) \leqslant 1$) and has a unique maximum ($S(x,y) = 1$ iff $x = y$) of 1 when both the images are same. The SSIM index is computed locally within a sliding window that moves pixel-by-pixel across the image, resulting in an SSIM map. The SSIM score of the entire image is then computed by pooling the SSIM map. The pooling strategy employed in this case is a simple average of all the SSIM values across the image. This value decreases as the images start to differ up to a minimum value of -1. This range of $-1 \leq SSIM \leq 1$ is mapped for conveniences of comparison with the MSE to $0 \leq SSIM_{inv} \leq 2$ and the new function is called $SSIM_{inv}$:

$$SSIM_{inv} = 1 - SSIM \tag{4.15}$$



(a) Template image (b) MSE = 4.2341e3 $SSIM_{inv}$ = 0.9936 (c) MSE = 4.8684e3 $SSIM_{inv}$ = 0.6682 (d) MSE = 8.9227e3 $SSIM_{inv}$ = 0.5174

Figure 4.1: Image similarity ranking using two different error measures MSE and $SSIM_{inv}$. (a) 192x168 template image of a subject from Extended Yale Face B database. (b) Randomly generated image of the same size. (c) Image of another subject. (d) A different image of the same subject.

Despite its simplicity, the $SSIM_{inv}$ index performs remarkably well across a wide variety of image and distortion types. Luminance shifting and contrast-stretching, which generally do not degrade image structure, lead to a very low $SSIM_{inv}$ value. An illustrative example is given in Fig. 4.1 comparing the MSE and $SSIM_{inv}$ functions. The face images are taken from the Extended Yale Face B database. Three images - a random image (Fig. 4.1(b)) over a uniform distribution and two face images (Fig. 4.1(c) & 4.1(d)) are compared and ranked against a template image (Fig. 4.1(a)). It can clearly be seen that the $SSIM_{inv}$ scores are more consistent and perceptually attuned than MSE scores which ranks the three images in an entirely opposite order with regard to similarity with the template image.

▶ 4.5 SparsGA — The Genetic Algorithm Based Sparsifier

The sparsifier for l_1-minimzation algorithm is an arch component of the SRC algorithm. Given a probe image and an error measure, the sparsifier, using the l_1 minimization algorithm, is used to identify a small number of gallery images

whose linear combination is optimally close to the given image. It is expected that at least one of the identified images will belong to the subject of the probe image.

A plethora of sparsifiers are available for this purpose as mentioned in Section 4.3 some of which are very efficient as well. However, all share the following common traits which make them unsuitable candidates for image processing and computer vision tasks, particularly, object recognition:

1. They all internally employ the l_2 norm which was already demonstrated above to be deficient in providing robust comparisons of images.

2. They are generally based on some kind of gradient-descent algorithms. These types of algorithms besides being rather restrictive in their choice of error functions, also are prone to getting stuck in local minima.

3. A stern limitation that they impose on the error functions is that of first-order continuity i.e. the error functions must be differentiable. This debars all those functions that have any kind of singularities in them or even zero order continuity.

The last item above is a consequential constraint. Many error functions are quite complex and computing their gradient can become a formidable task. Furthermore, guaranteeing them to be free of singularities is a non-trivial and elusive endeavour.

Owing to the above rationales, a new sparsifier was developed based on genetic algorithms (GAs). Genetic algorithms (GAs) are search algorithms falling into the category of optimization methods. Conventional AI search methods proceed by building a search tree along their way. The traversal is usually done by a fixed traversal scheme. These methods, in general, do not perform a directed search.

Solutions that are obtained during the traversal are judged and discarded and new solutions are searched for until the optimal solution is found. Gradient-based calculus methods on the other hands start with a initial solution and traverse the parameter space proceeding in the direction that reduces the error, hence obtaining increasingly better solutions. These methods, while being very directed, are highly localized and solution quality commonly depends on the initial solution. Because of this, they are highly susceptible to getting stuck in a local minimum in problems having a multi-modal error surface.

GAs use a directed search like gradient-based calculus method but do not rely on derivatives. Thus, they do not require the parameter space to be continuous. Along with that, they are global search methods. While gradient-based methods pick one initial solution, GAs pick a whole population of randomly generated initial solutions so that the whole space is searched in parallel. This prevents GAs to get stuck in a local minimum with a certain probability depending on the size of the population. There is, however, the problem of so-called premature convergence which is closely reminiscent of the local minima problem. In a premature convergence situation, the GA loses diversity in its population of solutions resulting in no improvement for a couple of generations until random mutation reintroduces some of the missing elements. While this impedes the performance of GA and the quality of solution significantly, it is not difficult to handle and various methods are proposed by researchers to get around it effectively.

The key concept in genetic algorithms is that problems are solved at a level different than that at which they are created. There is always some kind of coding scheme involved with which the parameters of one particular solution are encoded into strings called *chromosomes.* Genetic algorithms usually work with fixed sized populations of chromosomes. An initial population is generated entirely randomly. The quality of a chromosome is measured by a fitness function

which evaluates the nearness of the corresponding solution to the desired solution with a proper distance measure. Then, three genetic operators are applied to produce the next generation of solutions (chromosomes). These are *selection/reproduction*, *crossover* and *mutation*. Selection or reproduction is used to build the next generation using the fitness value of the chromosomes as their probability of survival. The more fit an individual is the more chance it has to be a part of the next generation. The crossover operator picks two parent chromosomes randomly from the new generation, chooses a random crossover point and swaps the chromosome strings after the crossover point. This way the operator produces two new child chromosomes. Then, a mutation operator is applied with a certain probability which changes one or more elements of the chromosomes randomly. Finally, the fitness values of the next generation's chromosomes are re-evaluated. This process is repeated for a fixed number of generations after which the chromosome with the best fitness value is taken to be the solution. For an excellent introduction to GAs, refer to [Goldberg, 1989, Mitchell, 1996, Michalewicz, 1996].

Since GAs do not rely on derivatives, almost any error function can be used with them. Furthermore, GAs tend to find the global minimum rather than a local one which gives a high likelihood of finding an optimal solution under most error measures [Michalewicz, 1996]. The main issue with GAs is encoding the problem parameters. It is in the encoding lies the magic. Here encoding implies techniques beyond simple binary-float representation quandary, which affects the performance of GAs in both accuracy and execution time. The problem at hand is what is known as a multiobjective optimization problem (Eq. 4.7). Perhaps the most intuitive approach to solving this type of multiobjective problem is constructing a single aggregate objective function by combining all of the objective functions into a single functional form. A well-known combination is the

weighted linear sum of the objectives which specifies scalar weights to each of the objectives. This aggregate function can then be solved by any single-objective optimizer. Clearly, the solution obtained will depend upon the relative values of the weights specified. Thus, it may be noticed that the weighted sum method is essentially subjective, in that someone needs to supply the weights.

This work uses a novel approach to transform the multiobjective optimization problem to a uniobjective one by incorporating one of the two constraints as part of the parameter coding leaving only one constraint to optimize. This transforms the problem parameter radically. With only one constraint left for the optimization algorithm to deal with, the problem complexity reduces drastically. The two constraints that need to be optimized for the SRC algorithm are the (1) sparsity constraint ($min\|\boldsymbol{x}\|_1$) and (2) the error constraint ($SSIM(\boldsymbol{y}, A\boldsymbol{x}) \leqslant \epsilon$), $\boldsymbol{x}$ being the parameter solved for. Conventionally, $\boldsymbol{x}$ is taken to be a random variable over a uniform distribution. The gradient-descent sparsifiers initialize $\boldsymbol{x}$ randomly over a uniform distribution and then proceed to modify its value under both constraints hoping to converge on an optimal value. This often leads to suboptimal results as the convergence could easily drift towards one constraint or the other or away from both. The proposed approach views $\boldsymbol{x}$ over a normal distribution and models it as a mixture of Gaussians. Because a Gaussian is an unbounded function, to keep $\boldsymbol{x}$ sparse, each Gaussian is truncated when its value drops below half its maximum value and scaled by a set factor. Thus for a certain mean and standard deviation, this process keeps only a limited set of values around the mean thereby ensuring sparsity. Such a truncated and scaled Gaussian is termed here as a *sparsity coefficient*. The vector $\boldsymbol{x}$ consists of a predetermined number of such sparsity coefficients combined together using a max function. The means and standard deviations of all the sparsity coefficients that comprise the vector $\boldsymbol{x}$ are then encoded into the form of a GA chromosome which

Algorithm 3 Sparse Representation-based Classification (SRC) - Proposed

1: **Input:** a matrix of gallery images $A = [A_1, A_2, \ldots, A_k] \in \mathbb{R}^{m \times n}$ for k classes, a test sample $\boldsymbol{y} \in \mathbb{R}^m$, (and an optional error tolerance $\epsilon > 0$).
2: Pre-process the test image $\boldsymbol{y}$ by a contrast enhancer such as one developed earlier in Chapter 3 or CLAHE.
3: Solve the sparsification problem:

$$\boldsymbol{x} \;\; subject\,to \;\; SSIM(\boldsymbol{y}, A\boldsymbol{x}) \leqslant \epsilon \;\; and \;\; \|\boldsymbol{x}\|_0 = \kappa$$

using SparsGA — the newly developed genetic algorithm based solver, given the values of ϵ and κ. κ is the number of sparse coefficients searched for. This second constraint is incorporated in the coding of $\boldsymbol{x}$.
4: Compute the residuals $r_i(\boldsymbol{y}) = SSIM(\boldsymbol{y} - A\delta_i(\boldsymbol{x})) \;\; for \;\; i = 1, \ldots, k.$
5: **Output:** $identity(\boldsymbol{y}) = min_i[r_i(\boldsymbol{y})]$.

denotes one possible solution. Thus, such an encoding is inherently sparse and the genetic operators are applied on it to vary the means and standard deviations of the various sparsity coefficients to satisfy the error constraint alone. This algorithm is named as *SparsGA* and its code was developed in Matlab from scratch for this research. It should be noted that SparsGA generates a sparse solution without performing an l_1 minimization at all.

▶ 4.6 Experimental Validation & Discussion

The original SRC algorithm 2 was modified using the above developments resulting in algorithm 3. As can be seen, a pre-processing stage was added to improve the contrast of the sample image which is one of the commonly occurring non-structural distortions. This could be either replaced or supplemented by other pre-processing stages demanded by the particularities of the application domain. The MSE measure has been replaced in both the l_1 minimization algorithm and the residual calculations by the SSIM measure. In addition the traditional sparsifiers were replaced by the newly developed genetic algorithm based sparisifier called *SparsGA* which can literally handle any type of error measure function. The genetic algorithm in SparsGA was operated on a population size of 50 and

Recognition Results			
Algorithm	**Extended Yale B**	**MPIE Non-Canonical**	**MPIE Canonical**
Original	50.6%	32.4%	22.5%
Proposed	80.0%	58.0%	48.1%

Figure 4.2: Recognition results on the Extended Yale Face B dataset and the Multi-PIE database.

upto 200 generations. These figures were picked up empirically to give a moderate accuracy to execution time ratio. A single sparsity coefficient was used with a standard deviation of 2.0 and a scaling factor of 1.3. The mean of the sparsity coefficient was determined by the GA runs.

By means of the thus proposed algorithm, experiments were performed using the Extended Yale Face B database [Lee et al., 2005] and the Multi-PIE database [Gross et al., 2008]. From the Extended Yale Face B database, cropped and pre-aligned images (size 192x168) of 38 subjects were used under 51 illumination conditions with neutral faces, eliminating those that were visually hard to recognize by human judgement. Out of these 51, the image under 0° illumination were preselected as the training image and the remaining 50 images as the test set.

From the Multi-PIE database, 100 subjects were randomly chosen, cropped and pre-aligned, with 21 images per subject in varying illumination conditions under two different expressions of neutral and smiling. Out of these, the neutral face image under 0° illumination was pre-selected as the training image and the remaining 20 images as the test set. Using this Multi-PIE subset, two further datasets were prepared. In one set, the *canonical dataset* (image size 140x140), all the images were mapped and warped to a canonical face, hence, minimizing the effects of expressions. In the *non-canonical dataset* (image size 130x120), the images were used without any mapping or warping. Experiments were run on all three datasets using the original sparse representation algorithm and the

Figure 4.3: Some recognition results from the Extended Yale Face B database. *(Top)*: Test images. *(Middle)*: Results from the original algorithm. *(Bottom)*: Results from the proposed algorithm.

proposed variations comparing algorithm performances. The experiments, thus, also outline performance differences under the canonical / non-canonical mapping.

The original sparse representation algorithm [Wang and Bovik, 2009] was tested under the following conditions: Five random features, each a vector of size 200, were extracted from each image. l_1 minimization was performed using the SpaRSA [Wright et al., 2008] algorithm. The proposed version of the algorithm was tested under the following conditions: Since the SSIM metric was used for image comparisons, which does not work under random transformations, the images were downsized to 64x64 and used it as a single extracted feature from the original images. Each test image was pre-processed for contrast enhancement using the CLAHE algorithm [Reza, 2004] implemented in Matlab. In the proposed genetic algorithm based sparsifier, the population size was kept to 50 and the generation count to 200. Three coefficients of sparsity were used.

The results in Fig. 4.2 show that the use of SSIM along with GA-based sparsifier have boosted the overall performance by a good margin. As can be seen in the sample results in Fig. 4.3 the original algorithm is not very robust to significant illumination variations owing particularly to the nature of the MSE error measure. Some inconsistencies can be seen in the results; for example, the classification of

the images in column 5 and 10, which are under similar illumination conditions. This can be explained by the local minima that gradient-based sparsifiers often get stuck in. Similar irregularities, however, are also exhibited by the proposed algorithm, for example, in the images in column 1 and 8. Although GA does not get into local minima, it does, however, need a lot of generations to converge optimally, limiting which sometimes lead to a suboptimal solution.

4.7 Conclusion

In this research, the sparse representation algorithm was extended to use any error measure for similarity comparison and the calculation of the residuals. The robustness of the structural similarity (SSIM) measure over illumination and expression variations was demonstrated. A new sparsifier was also developed that performed sparsification using genetic algorithms. An innovative method was presented to simplify the multiobjective optimization into uniobjective optimization by encoding one of the constraints as part of the problem parameter structure thereby making sparsity an inherent attribute of the generated solution. Further, the use of GAs allowed most error functions to be used without reservation. This gives tremendous flexibility in experimenting with different similarity measures, which act as a key classifier in this approach. Also demonstrated successfully were the deficiencies of the MSE measure adopted by the original algorithm and the gradient-based sparsifiers that are commonly used for this purpose. Experiments, that were conducted on two different databases, provide superior results showing marked improvement over the original approach.

5

Conclusions

The aim of this research was to explore the visual pathway in a mammalian brain, model certain aspects of the different modules that exist in this pathway and apply these models to selected problems in Computer Vision. Approaches to image reconstruction, contrast enhancement and object recognition were the subjects of primary investigation in this work. The research was motivated by the yet unsurpassed performance of the mammalian brain in a wide spectrum of scenarios. It discussed possible reasons for the failure of neuro-physiologically inspired artificial systems up till now. Neuroscience research, which holds strong bindings to the success of such artificial systems, was itself identified to be lacking concrete results owing in large part to a particular research attitude in trend. This dissertation contributed by analysing retinal processing through an image reconstruction approach, successfully applied the same reconstruction algorithm for contrast enhancement of gray-scale and colour images using colour-opponent receptive field model and improved upon the sparse representation classification algorithm by using a structural error measure along with a genetic algorithm based sparsifier.

▶ 5.1 Summary of Contributions

A summary of the main contributions of this thesis is outlined below:

▶ 5.1.1 Natural and Synthetic Image Reconstruction from Contrast Information

Retinal processing, in this work, was perceived as a process of information reduction that throws out uninteresting portions of the input image, passing only its contrast to the brain. This information reduction was also understood to be a general pattern in neural processing across different layers of the cortex. This, if true, would result in a gradual decrease of information that is passed along neural pathways. By being able to reconstruct the images from these contrast maps through the development of an image reconstruction algorithm, it was shown here that the processing prevalent in the retina and in V1 is information preserving. The algorithm (Alg. 1 in Chapter 2) was based on an iterative least squares error minimization using gradient descent, which was also analysed by providing its corresponding Bayesian framework. The Difference of Gaussians (DoG), when processed through a sigmoidal non-linearity, was taken as a contrast operator operating on input images. This contrast operator produced a composite contrast map consisting of an on- and off-centre contrast map. The computation and usage of the composite contrast maps were a key element to the reconstruction algorithm. In addition, a lot of emphasis has been placed on plotting out exact receptive field shapes, particularly in V1. By employing simulated irregular receptive fields during the reconstruction of images from their contrast maps and, thereby, obtaining successful reconstructions out of them, it was shown that such irregularities persist within the brain, and that the brain is robust to errors caused by such irregularities.

5.1.2 Biologically-Inspired Contrast Enhancement of Gray-Scale and Colour Images

The contrast enhancement algorithm developed 3.2 works on the principle of image modification in the contrast domain and uses the reconstruction algorithm described in Chapter 2 to perform an inverse transformation. The gain control function used was inspired by the psychometric contrast sensitivity curves. This gain control function was applied asymmetrically to both the on- and off-centre contrast maps resulting in a shift in luminance along with contrast improvement. Experimental results were reported on gray-scale and colour images. For the colour images, three different colour spaces were experimented with: RGB, L*a*b*, and the colour-opponent receptive fields.

Contrast enhancement algorithms that deal with colour images generally work using a particular colour model. In some models, enhancement is done on all planes of the image - e.g. RGB, while in others the enhancement is carried out on the luminance portion alone, which is later combined with the colour portion - e.g. L*a*b*. The issue with such an approach to enhancement is that either colours are enhanced independently of all other colours or they are left out initially, only later to be combined to generate the final image. In general, such algorithms are not able to work with the neurophysiological colour-opponent model that represent colour images in the form of colour-opponent contrast maps due to the fact that they generally operate on the image in its luminance domain. Because the contrast enhancement algorithm developed in this research is based on the aforementioned approach to image reconstruction (Alg. 1 in Chapter 2), and internally operates using such contrast maps, it naturally supports colour-opponent contrast maps by design. Operating on images in their contrast domain facilitates a straightforward approach to contrast enhancement transforming it to a mere

scaling operation. Furthermore, this approach also allows a shift in the mean luminance by applying different gain control factors to the two maps, which in itself is a novel approach.

5.1.3 Improvement of the Face Recognition Algorithm using Sparse Representation

The means to object recognition by sparse representation classification was first taken by Wright et al. [Wright et al., 2009], where they viewed face recognition, and in general object recognition, as a sparse representation problem. Although the original work disposed of any pre-processing and feature extraction steps labelling them to be unessential to sparse representation classification, the findings reported in this research speak otherwise. In addition, the original algorithm was found to be very sensitive to noise and ambiguities in real data sets. This research proposed modifications using early stages of pre-processing, employing image enhancements and feature extractions. The research discussed in detail the shortcomings of the mean-square error measure when it comes to comparing images and replaced it with one that is more aware of the underlying image structure - namely the Structural SIMilarity (SSIM) measure, which boosted the recognition rate significantly. Also a genetic algorithm based sparsifier (SparsGA) (Section 4.5) was developed and shown to outperform the gradient based algorithms in a final end result. SparsGA could work on literally any error measure function unlike its gradient-based predecessors, which operate using a derivative of the error measure requiring it to be continuous all the way, not to mention the extreme difficulty in calculating the derivatives of some of the more complicated error measures. Experimental results were reported on real data sets.

5.2 Future Work

This section identifies and briefly expounds on some of the potential future work related to the research presented in this dissertation. These future directions address expanding the scope of current research, as well as sowing seeds for new offshoots from this work.

Image Reconstruction

The image reconstruction approach from contrast maps provides a novel scenario, where images can be manipulated in their contrast domain and transformed back. This opens up innumerable possibilities for interesting image manipulations that could be valuable in Computer Vision. One of the applications that could be explored in object recognition is analysing the image at different contrast levels. It is not unusual for an image to contain objects that are at different contrast levels. Humans are very adept at attending to patterns at a particular contrast in the image. This includes selectively ignoring patterns of dominant contrast in order to discover subordinate ones. The proposed algorithm can be used to take an image to its contrast domain, filter out the unneeded levels, and reconstruct the filtered image. Due to the iterative nature of the algorithm, it is not as fast as some applications may demand. For this purpose, further work is needed to accelerate the convergence process leading to the reconstructed image.

Contrast Enhancement and Dynamic Range Adjustment

This same reconstruction algorithm is shown to be useful for contrast enhancement. A closely related problem is the dynamic range adjustment of high dynamic range (HDR) images for their display onto low dynamic range (LDR) displays. Several methods are used to achieve that: *contrast reduction*, where the HDR images are displayed on LDR devices by simply reducing the contrast; *clipping*

and compressing dynamic range, where the dynamic range is cropped, cutting off the darkest and brightest details, or alternatively with an S conversion curve that compresses contrast progressively and more aggressively in the highlights and shadows, while leaving the middle portions of the contrast range relatively unaffected; *tone mapping*, which attempts to reduce the dynamic range while retaining localized contrast, tapping into research on how the human eye and visual cortex perceive a scene. All of these methods have their pros and cons. Further insights into the proposed contrast enhancement algorithm could provide yet another promising alternative for HDR-to-LDR mapping that was not pursued during this research due to scope limitations.

Automated Human Face Recognition

A vital part in any object recognition algorithm is a similarity comparison metric. The accuracy of an algorithm is usually as good as the metric it uses. In Chapter 4, two different metrics were analysed and compared - the *MSE* and the *SSIM*. Although, the SSIM metric outperformed the MSE in comparing a particular class of images - namely the frontal facial images, it still falls short of an ideal measure by a large margin, particularly in confirmation to human perception. Thus, there is a need for sustained improvement in image similarity metrics currently at hand. The sparse recognition model currently does not handle recognition with pose variations. Similarly, translations and rotations are also not catered for. While work exists that has addressed these issues, it is far from an optimal solution. Another area that need to be extended on is the sparsifier. SparsGA or the genetic algorithm based sparsifier, holds lots of promises and work is required for its generalization.

Bibliography

J. Alder, B. D. Rao, and K. Kreutz-Delgado. Comparison of basis selection methods. *IEEE Signals, Systems and Computers*, 1997.

E. Amaldi and V. Kann. On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems. *Theoretical Computer Science*, 209: 237–260, 1999.

C. Anagnostopoulos, A. Koutsonas, I. Anagnostopoulos, V. Loumos, and E. Kayafas. Tile classification using the cielab color model. 3514:83–112, 2005.

H. C. Andrews and B. R. Hunt. *Digital Image Restoration.* Prentice-Hall Inc., 1977.

G. R. Ayers and J. C. Dainty. Iterative blind deconvolution method and its applications. *Optics Letters*, 13(7):547–549, 1988.

R. G. Baraniuk and M. B. Wakin. Random projections of smooth manifolds. *Foundations of Computational Mathematics*, 9(1):51–77, 2007.

R. H. T. Bates and M. J. McDonnell. *Image Restoration and Reconstruction.* Oxford Press, 1986.

M. F. Bear, B. W. Connors, and M. A. Paradiso. *Neuroscience: Exploring the Brain.* Lippincott Williams & Wilkins, 3 edition, 2007.

P. Belhumeur, J. Hespanda, and D. Kriegman. Eigenfaces vs fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

R. Berinde, P. Indyk, and M. Ruzic. Practical near-optimal sparse recovery in the l1 norm. *Allerton*, 2008.

T. Blumensath and M. E. Davies. In greedy pursuit of new direction: (nearly) orthogonal matching pursuit by directional optimisation. *EUSIPCO*, 2007.

B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal classifiers. *5th Annual ACM Workshop on COLT (ed. Haussler, D.)*, pages 144–152, 1992.

L. M. Bowman. Tampa drops face-recognition system. *CNET News, http : // news.cnet.com/Tampa − drops − face − recognition − system/2100 − 1029_3 − 5066795.html?tag = nl*, 8 2003.

E. Candes and T. Tao. Near-optimal signal recovery from random projections. *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006.

E. Candes, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Math*, 59(8): 1207–1223, 2006.

S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM Review*, 43:129–159, 2008.

F. Crick. The recent excitement about neural networks. *Nature*, 337:129–132, 1989.

N. A. Crowder, N. S. C. Price, M. A. Hietanen, B. Dreher, C. W. G. Clifford, and M. R. Ibbotson. Relationship between contrast adaptation and orientation tuning in v1 and v2 of cat visual cortex. *Journal of Neurophysiology*, 95:271–283, 2006.

N. W. Daw, R. J. Jensen, and W. J. Brunken. Rod pathways in mammalian retinae. *Trends in Neuroscience*, 13(3):110–115, 3 1990.

P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. *SIGGRAPH*, 1997.

K. Devlin. A review of tone reproduction techniques. *Dept. Computer Sci. Univ. of Bristol, Bristol, U.K.*, 2002.

D. Donoho. For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution. *Communication on Pure and Applied Math*, 59(6):797–829, 2006.

B. Feder. Technology strains to find menace in the crowd. *The New York Times, http : //www.nytimes.com/2004/05/31/technology/31face.html?ex = 1401336000&en= 100298736d182ca4&ei = 5007&partner = USERLAND*, 05 2004.

J. D. Foley, A. V. Dam, S. K. Feiner, and J. F. Hughes. *Computer Graphics: Principles and Practice in C.* Addison-Wesley, 2 edition, 1995.

M. S. Gazzaniga, R. B. Ivry, and G. R. Mangun. *Cognitive Neuroscience: The Biology of the Mind.* W. W. Norton & Company, 2 edition, 2002.

B. Girod. What's wrong with mean-squared error? *Visual Factors of Electronic Image Communications*, 1993.

D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison-Wesley, 1989.

R. C. Gonzalez and R. E. Woods. *Digital Image Processing.* Pearson Prentice Hall, 3rd edition, 2008.

I. F. Gorodnitsky and B. D. Rao. Energy localization in reconstructions using focuss: A recursive weighted norm minimization algorithm. *IEEE Trans. on Signal Processing*, 1997.

T. Greene. Face recognition useless for crowd surveillance. *The Registrar, http : // www.theregister.co.uk/2001/09/27/face_recognition_useless_for_crowd/*, 09 2001.

R. Gross, I. Matthews, J. F. Cohn, T. Kanade, and S. Baker. Multi-pie. *Proceedings of the Eighth IEEE International Conference on Automatic Face and Gesture Recognition FG'08*, sep 2008.

K. M. Hanson. Introduction to bayesian image analysis. *Medical Imaging: Image Processing*, 1898:716–731, 1993. Proc SPIE.

X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face recognition using laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):328–340, 2005.

B. K. P. Horn. Determining lightness from an image. *Computer Graphics and Image Processing*, 3(1):277–299, 1974.

Susan Van Horn, Alev Erisir, and S. Murray Sherman. Relative distribution of synapses in the a-laminae of the lateral geniculate nucleus of the cat. *The Journal of Comparative Neurology*, 416:509–520, 2000.

J. Huang, X. Huang, and D. Metaxas. Simultaneous image transformation and sparse representation recovery. *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008.*, pages 1–8, 2008.

K. Huang, Q. Wang, and Z. Wu. Color image enhancement and evaluation algorithm based on human visual system. *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004 (ICASSP '04)*, 3: iii–721–4, 2004.

D. H. Hubel. *Eye, Brain, and Vision.* Scientific American Library, 1995.

D. H. Hubel and T. N. Wiesel. Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology*, 148(3):574–591, 1959.

D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154, 1961.

R. Jain, R. Kasturi, and B. G. Schunck. *Machine Vision.* McGraw-Hill, 1995.

T. L. Ji, M. K. Sundareshan, and H. Roehrig. Adaptive image contrast enhancement based on human visual properties. *IEEE Transactions on Medical Images*, 13(4):573–586, 1994.

Y. Jin, L. Fayad, and A. Laine. Contrast enhancement by multi-scale adaptive histogram equalization. *Proceedings of SPIE - The International Society for Optical Engineering*, 4478:206–213, 2001.

E. R. Kandel, J. H. Schwartz, and T. M. Jessell. *Principles of Neural Science.* McGraw-Hill, Inc, 4 edition, 2000.

Christoph Kayser, Rodrigo Salazar, and Peter Konig. Responses to natural scenes in cat v1. *Journal of Neurophysiology*, pages 1910–1920, 2003.

A. A. Khwaja and R. Goecke. Image reconstruction from contrast information. *Proceedings of the 2008 Digital Image Computing: Techniques and Applications, DICTA 2008*, pages 226–233, 2008.

A. A. Khwaja and R. Goecke. Biologically inspired contrast enhancement using asymmetric gain control. *Proceedings of the 2009 Digital Image Computing: Techniques and Applications, DICTA 2009*, pages 424–430, 2009.

S. W. Kuffler. Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology*, 16(1):37–68, Jan 1953.

H. Kukkonen, J. Rovamo, K. Tiippana, and R. Nasanen. Michelson contrast, rms contrast and energy of various spatial stimuli at threshold. *Vision Research*, 33(10):1431–1436, July 1993.

J. J. Kulikowski and K. Kranda. Image analysis performed by the visual system: Feature versus fourier analysis and adaptable filtering. *Visual Neuroscience*, pages 381–404, 1986.

D. Kundur and D. Hatzinakos. Blind image deconvolution. *IEEE Signal Processing Magazine*, 13(3):43–64, 1996.

A. Laine, J. Fan, and W. Yang. Wavelets for contrast enhancement of digital mammography. *IEEE Engineering in Medicine and Biology Magazine*, 14(5): 536–550, 1995.

E. H. Land and J. J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, 1971.

K. C. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 27(5): 684–698, 2005.

J. Leyden. Anti-terror face recognition system flunks tests. *The Register, http : //www.theregister.co.uk/2003/09/03/antiterror_face_recognition_system_flunks/*, 09 2003.

X. Li, T. Jia, and H. Zhang. Expression-insensitive 3d face recognition using sparse representation. *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009*, 2009.

L. Ling, Z. Yinqing, and L. Jingwen. Visualization of high dynamic range image with retinex algorithm. *Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications*, pages 1215–1218, 2007.

Y. S. Liu, C. F. Stevens, and T. O. Sharpee. Predictable irregularities in retinal receptive fields. *PNAS*, 106(38):16499–16504, 2009.

M. Livingstone. *Vision and Art: The Biology of Seeing.* Harry Abrams Inc., 2002.

P. Mamassian, M. Landy, and L. T. Maloney. Bayesian modelling of visual perception. *Probabilistic Models of the Brain: Perception and Neural Function*, pages 13–36, 2002.

D. Marr. The computation of lightness by the primate retina. *Vision Research*, 14(12):1377–1388, 1974.

D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London*, 207(1167):187–217, 2 1980.

Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs.* Springer Verlag, 3 edition, 1996.

M. Mitchell. *An Introduction to Genetic Algorithms.* MIT Press, 1996.

B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

B. A. Olshausen and D. J. Field. What is the other 85% of v1 doing? *Problems in Systems Neuroscience*, 23, 2004.

S. E. Palmer. *Vision Science - Photons to Phenomenology.* MIT Press, 1999.

T. N. Pappas, R. J. Safranek, and J. Chen. Perceptual criteria for image quality evaluation. *Handbook of Image and Video Processing 2nd ed.*, 2005.

S.N. Pattanaik, J.A. Ferwerda, M.D. Fairchild, and D.P. Greenberg. A multiscale model of adaptation and spatial vision for realistic image display. *SIGGRAPH '98: Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 287–298, 1998.

E. Peli. In search of a contrast metric: Matching the perceived contrast of gabor patches at different phases and bandwidths. *Vision Research*, 37(23):3217–3224, 1997.

S. M. Pizer and E. P. Amburn. Adaptive histogram equilization and its variations. *Computer Vision, Graphics, and Image Processing*, 39:355–368, 1987.

A. Rakotomamonjy. Algorithms for multiple basis pursuit denoising. *SPARS*.

K. Rehm and W. J. Dallas. Artifact suppression in digital chest radiographs enhanced with adaptive histogram equilization. *SPIE, Medical Imaging III*, 1989.

A. M. Reza. Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *Journal of VLSI Signal Processing*, 38:35–44, 2004.

J. C. Russ. *The Image Processing Handbook*. CRC Press, 2 edition, 1995.

J. Sadr, S. Mukerjee, K. Thoresz, and P. Sinha. The fidelity of local ordinal encoding. *Advances in Neural Information Processing Systems*, 14, 2002.

R. H. Sherrier and G. A. Johnson. Regionally adaptive histogram equalization of the chest. *IEEE Transactions on Medical Images*, 6(1):1–7, 1987.

Adam Sillito, Javier Cudeiro, and Helen Jones. Always returning - feedback and sensory processing in visual cortex and thalamus. *TRENDS in Neurosciences*, 29(6):307–316, 6 2006.

G. B. Stanley, F. F. Li, and Y. Dan. Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *The Journal of Neuroscience*, 19(18):8036–8042, 1999.

M. Sur. *MIT Video Lecture*. URL of video, 2003.

M. Tipping. Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, pages 211–244, 2001.

M. E. Tipping. Bayesian inference: An introduction to principles and practice in machine learning. *Advanced Lectures on Machine Learning*, pages 41–62, 2004.

M. Turk and A. Pentland. Eigenfaces for recognition. *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 1991.

V. Vonikakis, I. Andreadis, and A. Gasterator. Fast centre-surround contrast modification. *Image Processing, IET*, 2(1):19–34, 2008.

Z. Wang and A. C. Bovik. *Modern Image Quality Assessment*. Morgan and Claypool, 2006.

Z. Wang and A. C. Bovik. Mean squared error: love it or leave it? - a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, 2009.

Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

Z. Wang, A. C. Bovik, and H. R. Sheikh. Structural similarity based image quality assessment. *Digital Video Image Quality and Perceptual Coding, Chap 7*, 2005.

D. K. Warland, P. Reinagel, and M. Meister. Decoding visual information from a population of retinal ganglion cells. *Journal of Neurophysiology*, 78(5):2336–2350, 1997.

R. Willing. Airport anti-terror systems flub tests. *USA Today, http : // www.usatoday.com/travel/news/2003/09/02 – air – secur.htm*, 09 2003.

J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.

S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo. Sparse reconstruction by separable approximation. *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008*, pages 3373–3376, 2008.

H. R. Wu and K. R. Rao. *Digital Image Video Quality and Perceptual Coding.* CRC, 2005.

W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.

W. ZHiming and T. Jianhua. A fast implementation of adaptive histogram equalization. *8th International Conference on Signal Processing, 2006*, 2:16–20, 2006.

Karel Zuiderveld. *Contrast limited adaptive histogram equalization*, pages 474–485. 1994.