Noise-limited scene-change detection in images

A thesis
submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

at

Lincoln University

by

K. Irie

Lincoln University

2009

(empty)

Declaration

Parts of this thesis have been submitted and/or accepted for publication in advance of submission of the thesis:

(i) Conference presentation and proceedings

Irie, K., McKinnon, A. E., Unsworth, K. & Woodhead, I. M. (2006). A comparison of noise in CCD and CMOS image sensors, *Proc. Image and Vision Computing New Zealand* pp. 43–48. (Includes parts of Chapter 3).

Irie, K., McKinnon, A. E., Unsworth, K. & Woodhead, I. M. (2007). Measuring digital camera image noise, *Proc. 2nd International Conference on Sensing Technology.* (Includes parts of Chapter 3).

Irie, K., McKinnon, A. E., Unsworth, K. & Woodhead, I. M. (2007). Shadow removal for object tracking in complex outdoor scenes, *Proc. Image and Vision Computing New Zealand* pp. 25–30. (Includes parts of Chapter 4).

Irie, K., McKinnon, A. E., Unsworth, K. & Woodhead, I. M. (2008). An investigation into noise-bound shadow detection and removal, *Proc. Image and Vision Computing New Zealand* pp. 1–6. (Includes parts of Chapters 4 & 5).

(SUBMITTED to IVCNZ 2009) Irie, K., McKinnon, A. E., Unsworth, K. & Woodhead, I. M. (-). Measured effects of temperature on illumination-independent camera noise. (Includes parts of Chapter 3).

(ii) Journal Publications

Irie, K., McKinnon, A. E., Unsworth, K. & Woodhead, I. M. (2008). Measurement of digital camera image noise for imaging applications, *Sensors and Transducers* 90:185-194. (Includes parts of Chapter 3).

Irie, K., McKinnon, A. E., Unsworth, K. & Woodhead, I. M. (2008). A model for evaluation of noise in CCD digital-video cameras, *Measurement, Science, and Technology*, 19. (Includes parts of Chapter 3).

Irie, K., McKinnon, A. E., Unsworth, K. & Woodhead, I. M. (2008). A technique for evaluation of CCD video-camera noise, *IEEE Transactions on Circuits and Systems for Video Technology* 18(2):280-284. (Includes parts of Chapter 3).

(SUBMITTED to JOSA-A) Irie, K., McKinnon, A. E., Unsworth, K. & Woodhead, I. M. (-). A noise-bound method for detecting shadow-free scene changes in image sequences. (Includes parts of Chapters 4 & 5).

Abstract of a thesis submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy

Noise-limited scene-change detection in images

by
K. Irie

This thesis describes the theoretical, experimental, and practical aspects of a noise-limited method for scene-change detection in images. The research is divided into three sections: noise analysis and modelling, dual illumination scene-change modelling, and integration of noise into the scene-change model.

The sources of noise within commercially available digital cameras are described, with a new model for image noise derived for charge-coupled device (CCD) cameras. The model is validated experimentally through the development of techniques that allow the individual noise components to be measured from the analysis of output images alone. A generic model for complementary metal-oxide-semiconductor (CMOS) cameras is also derived. Methods for the analysis of spatial (inter-pixel) and temporal (intra-pixel) noise are developed. These are used subsequently to investigate the effects of environmental temperature on camera noise. Based on the cameras tested, the results show that the CCD camera noise response to variation in environmental temperature is complex whereas the CMOS camera response simply increases monotonically.

A new concept for scene-change detection is proposed based upon a dual illumination concept where both direct and ambient illumination sources are present in an environment, such as that which occurs in natural outdoor scenes with direct sunlight and ambient skylight. The transition of pixel colour from the combined direct and ambient illuminants to the ambient illuminant only is modelled. A method for shadow-free scene-change is then developed that predicts a pixel's colour when the area in the scene is subjected to ambient illumination only, allowing pixel change to be distinguished as either being due to a cast shadow or due to a genuine change in the scene. Experiments

on images captured in controlled lighting demonstrate 91% of scene-change and 83% of cast shadows are correctly determined from analysis of pixel colour change alone.

A statistical method for detecting shadow-free scene-change is developed. This is achieved by bounding the dual illumination model by the confidence interval associated with the pixel's noise. Three benefits arise from the integration of noise into the scene-change detection method:

1. The necessity for pre-filtering images for noise is removed;
2. All empirical thresholds are removed; and
3. Performance is improved.

The noise-limited scene-change detection algorithm correctly classifies 93% of scene-change and 87% of cast shadows from pixel colour change alone. When simple post-analysis size-filtering is applied both these figures increase to 95%.

Keywords: camera noise, image noise, noise models, scene-change detection, object segmentation, shadow detection, illumination modelling.

Preface

The ability of humans to process their surroundings using vision has been finely crafted and refined over millions of years. At a mere glance, and with virtually no conscious thought, we can extract details from our environment in a purely passive manner, recreating and forming objects inside our mind to represent virtually the physical objects around us. In addition, advancements by humans in the past 50 years have given us computers that can process millions of instructions per second, eclipsing the computing power of the human brain, and camera vision systems that routinely check and count thousands of items without fail every second, far beyond the capability of any living human.

Why then have we not developed robots, even slow ones, that can reproduce the most basic human visual tasks, such as analysing its surroundings, in a robust and reliable manner?

One answer may be due to the extreme complexity of the human brain, and its ability to continually learn. Even if the brain-speed of a new-born child was increased a million fold, it would not be able to perform a detailed analysis of its environment — not until it learned what its environment *was*. If we are to use the advancements in computing power provided to us today to replicate simple human abilities, then perhaps we too have to start with baby steps — we have to understand what it is that needs to be learned, and we need to implement them.

Scientists have been slowly unravelling the secrets of the human brain and in particular with regard to the study in this thesis, the understanding of the human vision system (HVS) and the interaction between the optical system and the brain. We know that the brain pre-processes the information captured by the retina prior to any high-level processing functions including form, movement, stereo depth, and colour (Sekular & Blacke 1988, Gregory 1988). This reduces the signals that are transmitted to the brain through the optic nerves from the eye's 120 million receptors to 1 million. Furthermore, Gregory states "No features *determine* depth or form; they can only increase the probabilities of seeing in particular ways". This suggests that an 'unlearned' image processing algorithm cannot determine an object from features alone.

We know that shadows provide important cues for interpretation of images in the HVS (Ramachandran 1998). This is demonstrated in Figure 1 which illustrates how shadows alone can allow us to extract an object's form. Aside from this function, shadows enable a single eye to have something close to binocular vision. Our brain continually measures the lengths of the shadows and deduces their heights and shapes, helping to deduce distance. The world around us looks quite flat when the light is directly behind us, for then there are no shadows, which makes for flat scenes and dull pictures. Further, David Brewster (1781-1868) in his "Letters on natural magic", describes the changes in the elevations and depressions on a model depending upon

whether the light is falling from above or below. In further experimentation of the effect Brewster found that this was more marked in adults than with children.

Figure 1 An example of extracting an object's form by shadow inference only.

In relation to robotics and their inability to accurately reproduce the simple human visual task of object tracking, I speculate that we start at the very beginning with baby steps, mimicking the pre-processing operations of the HVS, then allowing higher-level functions to operate on the extracted detail. One area of extreme importance to any image processing application is the way it treats image noise. For example, Figure 2 illustrates an image that we find easy to decouple from noise - we do not need to evaluate the image for long before we can extract the detail of the word 'Noise'. Yet with no other input it is difficult to logically program an algorithm that could extract the word 'Noise' alone.

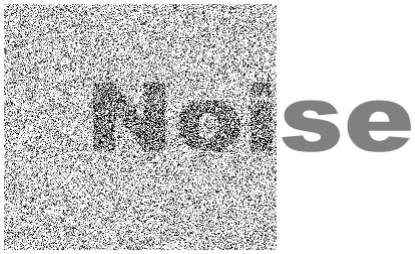


Figure 2 An example of the ability of the HVS to extract an object's form from within the presence of random noise.

This study is targeted firstly at our understanding of noise within digital cameras, and the ability to measure the noise content within captured images. Secondly, the study of noise is validated by applying the knowledge gained to a scene-change detection application that attempts to segment objects from their shadows. By doing this, features can be provided that could improve the probability that an algorithm will be able to detect the correct form of an object.

Aims and achievements

The overall aim of this research is to contribute to science a greater understanding of digital camera image noise and to demonstrate its application to increasing the robustness of image processing applications. The pathway to this aim includes the study and review of relevant background theory, the formulation of original ideas,

implementation and execution of these ideas, analysis of experimental results, and scrutiny of the work by the scientific community. Through this pathway the aim has been met.

Achievements of this research, and directly related to the work, include:

- Construction of new comprehensive models for CCD and CMOS camera noise.
- Development of a new method for measuring the sources of CCD camera noise from output images alone.
- Development of methods for measurement of spatial, temporal, and combined noise in any type of digital camera.
- Analysis of the effect of environmental temperature on camera noise, with results showing surprising noise variation both within and between cameras.
- Creation of a new model and method for scene-change detection in image sequences based on dual-illumination and colour chromaticity.
- Integration of the image noise model into the scene-change detection shadow model which creates a new method that is free from empirical thresholds and demonstrates improved performance over the original method.
- Presentation of this work at four international conferences: *Image and Vision Computing New Zealand 2006* (Irie, McKinnon, Unsworth & Woodhead 2006), *International Conference on Sensing Technology 2007* (Irie, McKinnon, Unsworth & Woodhead 2007a), *Image and Vision Computing New Zealand 2007* (Irie, McKinnon, Unsworth & Woodhead 2007b); *Image and Vision Computing New Zealand 2008* (Irie, McKinnon, Unsworth & Woodhead 2008a); with one under review, *Image and Vision Computing New Zealand 2009*; and presentation of two seminars at Lincoln University.
- Publication of papers in three refereed journals: *Measurement, Science, and Technology* (Irie, McKinnon, Unsworth & Woodhead 2008c), *IEEE Circuits and Systems for Video Technology* (Irie, McKinnon, Unsworth & Woodhead 2008d), *Sensors and Transducers* (Irie, McKinnon, Unsworth & Woodhead 2008b), (*JOSA-A* under review).

Thesis layout

The thesis is segmented into four parts. Part I presents an introduction and background pertinent to this research. Part II contains work related to the modelling and measurement of camera noise, Part III contains work related to scene-change detection based on dual-illumination and the integration of the noise model with the scene-change detection process. Part IV finishes the thesis with some conclusions and suggestions

for future study. Each of parts II & III have their respective literature reviews clearly delineated at the beginning of their chapters.

- Part I Introduction and background
  Chapter 1 introduces the importance of noise in image processing and the need for a deeper understanding of digital camera noise, as well as the necessity for robust shadow-free scene-change detection.

  Chapter 2 provides background on the the human vision system and its importance in digital imaging. Digital cameras are discussed, followed by the philosophy behind lighting and shadows.

- Part II Noise analysis
  Chapter 3 discusses camera noise in detail, providing a review of the current literature regarding sources of noise in digital cameras, particularly for the ubiquitous CCD- and CMOS-type devices. The formation of CCD and CMOS noise models are presented, followed by methods developed for measuring camera noise from output images alone. These methods are used to validate the CCD model experimentally. The effects of environmental temperature on camera noise is also investigated, which provides some surprising results.

- Part III Scene-change detection
  Chapter 4 discusses scene-change detection based upon colour change only. It presents a review of the current literature and methods for the shadow removal and scene-change detection problem in general, followed by a new method for detection of scene-changes in image sequences.

  Chapter 5 details an enhanced method of scene-change detection that integrates camera noise knowledge with the scene-change detection method.

- Part IV Conclusions and future work
  The thesis finishes with a concluding discussion of the work presented, as well as suggestions for future research.

Acknowledgements

I offer my sincerest thanks to Annette and Taylor, without whose support and understanding this thesis would not have been written. This work was as much a result of their support and encouragement as my effort.

My supervisors, Professor Alan McKinnon, Dr Keith Unsworth, and Dr Ian Woodhead, who have been outstanding in their guidance and support throughout the duration of this PhD study. I hesitate to think how much more difficult it would have been without their patience and dedication.

Thanks also to Lincoln Ventures Ltd, who supported me through my PhD studies whilst in full-time employment, and also to Lincoln University and the Department of Applied Computing.

Finally, I wish to acknowledge the support of the New Zealand Foundation for Research, Science, and Technology, who supported my work through their Research For Industry programme.

Glossary of terms

ADC	Analogue-to-Digital Conversion
APS	Active Pixel Sensor (CMOS)
Brightness	The visual perception of light output
CCD	Charge-Coupled Device
CFA	Colour Filter Array
Chrominance	A 2D-space for representing colour
CMOS	Complementary Metal-Oxide-Semiconductor
Daylight	The combination of sunlight and skylight
DSNU	Dark-Signal Non-Uniformity (also referred to as FPN)
DSP	Digital Signal Processor
FF	Full-Frame Transfer (CCD)
FT	Frame Transfer (CCD)
GMB	Gretag Macbeth colour chart
HID	High Intensity Discharge
HVS	Human Visual System
IC	Integrated Circuit
IL	Interline Transfer (CCD)
Intensity	The relative brightness of a portion of image or illumination
LBR	Log Band-Ratio
Lightness	The perception response to brightness
Luminance	Photometric measure of luminous intensity per unit area of light
Metamerism	The matching of apparent colour of objects with different SPDs
PPS	Passive Pixel Sensor (CMOS)
Quincunx	A geometric pattern consisting of five points, four of them forming a square or rectangle and a fifth at its center
Skylight	Solar radiation from the scattering of sunlight in the atmosphere
SNR	Signal-to-Noise Ratio
SPD	Spectral Power Distribution
Sunlight	Radiation given off by the sun
YC_RC_B	Digital luminance/chrominance colour space

Symbol and variable definitions

σ	Standard deviation
λ	Wavelength (of light)
μ	Mean
B	Blue component of an item
C	Sensor response
E	Spectial energy of an illuminant source
FPN	Fixed-pattern noise
G	Green component of an item
h	Planck's constant
k	Boltzmann's constant
L^*	Lightness
N_C	Column noise
N_D	Demosaicing effect on noise
N_{filt}	Filtering effects on image noise
N_Q	Quantization noise
N_{read}	Readout noise
P	RGB pixel measurement
$PRNU$	Photo-response non-uniformity
R	Red component of an item
S	Surface of an object
SN_{dark}	Dark-current shot noise
SN_{ph}	Photon shot noise
T	Temperature
Y	Luminance

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Part I

Introduction and background

Chapter 1

INTRODUCTION

Digital cameras have become ubiquitous sensors for automated technologies. Advancements in both vision and electronic technologies have increased the applicability of these sensors to almost every industry imaginable, as camera specification, sensitivity, and computing power is increasing, while effective cost is decreasing. Consequently, the expectations of the associated performance of image processing algorithms is also increasing. It would be fair to say that machine vision applications are no longer limited by optical hardware or processing power, but in the way that the image data is interpreted.

One confounding yet compelling aspect of digital cameras is that the photo-sensors themselves collect light in a passive manner. In uncontrolled environmental conditions, such as in an outdoor setting, a camera has no absolute reference point in terms of the data it is collecting. There is little use in analysing the absolute number of photons captured by each pixel as the total photon flux will vary greatly depending upon how much sunlight is available. Any subsequent image analysis must therefore be made on the relative differences in the number of photons captured by the pixels.

Further, the problem of image analysis is hampered by the addition of noise to the pixel data. Noise will always be present in any captured image, even in the perfect camera, as photons are discrete entities and are subject to Poisson variations. Noise becomes increasingly problematic as the contrast reduces between pixels being analysed and as the pixel values approach noise levels. An image processing algorithm will then be unable to distinguish true scene content from noise, and the algorithm may not know when it is failing. Indeed, this is one of the reasons why cameras are used less frequently in safety-critical applications — if the algorithm is reacting to noise then it will be too sensitive and falsely react, yet if it is insufficiently sensitive to noise it may not react to genuine scene changes when it should.

Somewhat surprisingly, most image processing algorithms do not adequately recognize or handle noise; rather they rely on heuristic or empirical thresholds which can directly affect algorithm robustness and applicability.

An unavoidable effect in any outdoor environment is shadowing, which can cause

large relative pixel differences depending upon the strength of the illumination source and whether a pixel is in shadow or not. Consequently, this can make it very difficult for an algorithm to distinguish whether changes in pixel values are caused by changes in the scene or changes in shadow.

Object tracking applications in outdoor environments suffer from noise and shadow effects. It has become more relevant in recent times due to the ubiquity of low-cost cameras being used for security purposes in urban areas where algorithms are expected to operate throughout the day and night. Object tracking, along with other outdoor image processing applications, can gain significant robustness from incorporating noise detail and shadow-free scene-change detection into their algorithms.

The human vision system (HVS) has an excellent ability to adapt to environmental conditions, filter noise, and segment shadows from objects. Applying this ability to computer vision has not been a trivial task for engineers and scientists, and development of completely autonomous vision systems has proven elusive.

The needs identified and addressed in this thesis are as follows:

- To understand image noise and its sources in standard off-the-shelf digital image sensors.
- To model and measure image noise types from digital camera images.
- To research a method for application-independent scene-change detection, whose performance is ideally limited only by image noise.

Chapter 2

BACKGROUND

This chapter provides a background to the principles of the human vision system, cameras, lighting, and shadows. In explaining these fundamentals, issues that affect computer vision applications will become apparent.

2.1 THE HUMAN VISION SYSTEM

2.1.1 The non-linear intensity response of the HVS

The human vision system (HVS) has a non-linear perception of intensity. Although light intensity can be easily measured, brightness is a subjective phenomenon. Studies show that there is a complex relationship between intensity and perceived brightness that depends on several factors, including the level of surrounding light (Stone 2003), but is often approximated as log intensity, or intensity powered to between 1/2 and 1/3 (Poynton 1996). The Commission Internationale de l'Éclairage (CIE) defines luminance, denoted Y, as radiant power weighted by a spectral sensitivity function that is characteristic of the HVS (CIE 2009). Furthermore, the CIE defines lightness, L^* (scale 0–100), the perceptual response to brightness, as predominately a modified cube root of luminance:

$$L^* = 116 \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} - 16, \qquad \frac{Y}{Y_n} > 0.008856 \tag{2.1}$$

where Y_n is the luminance of the white reference. This response is graphed in Figure 2.1 where Y_n is 100.

An example demonstrating the non-linearity of the HVS is demonstrated in Figure 2.2. The top bar is luminance, encoded linearly, where it appears that most of the values are light. The bottom bar is lightness which is perceptually encoded such that the photographic mid-grey appears at value 50, but is only 18% of the maximum luminance value 100.

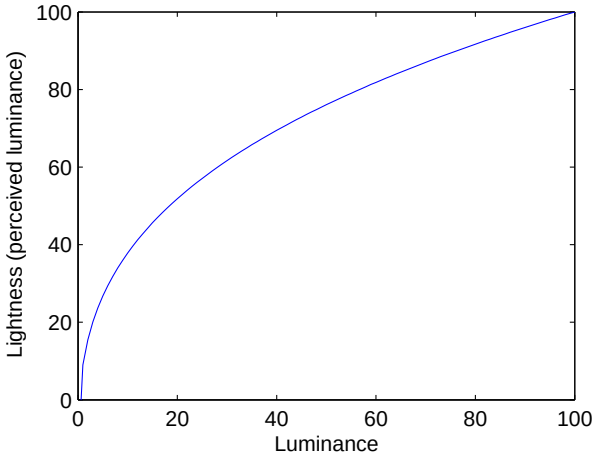


Figure 2.1 The human perceptual response to luminance as defined by the CIE.

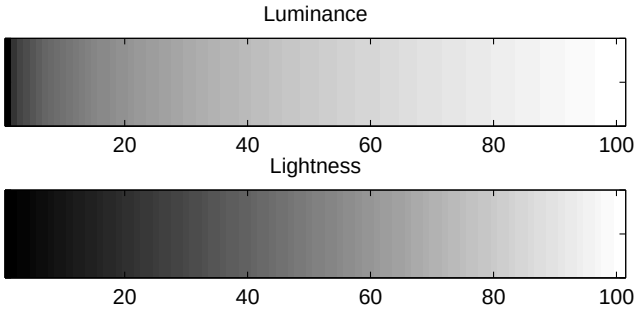


Figure 2.2 Linearly encoded intensity values (top) and linearly encoded lightness (perceptual) values (bottom).

2.1.2 Colour response of the HVS

Colour is what we see, and much more than just a physical attribute of the light our eyes detect. Sir Isaac Newton describes it as follows (Newton 1730):

> "For the Rays (of light) to speak properly are not coloured. In them there is nothing else than a certain Power and Disposition to stir up a Sensation of this or that Colour."

An example of this process is given by the colour purple. Purple is not a hue listed in the spectrum of colour (red, orange, yellow, green, cyan, blue, indigo (Gerritsen 1975)) but a combination of the red and blue/violet hues. This illustrates the difficulty in developing machine vision systems capable of mimicking the HVS (Huang & Wu 1998). The eye captures light in the rods and cones of the retina. The rods are highly light sensitive, and allow vision in dim lighting conditions. The reason we see monochromatically at night is because we only have one type of rod. There are three

varieties of cone: short, medium and long, which are sensitive to blue, green, and yellow-green light respectively (Trussell, Saber & Vrhel 2005) giving colour. These sensitivities form the basis for the trichromatic (R,G,B) colour primaries in digital imaging. The relative populations of the short, medium and long cones are approximately 1:20:40. The scarcity of short (blue) cones, which are absent entirely from the centre of the fovea, limits the sharpness of intensely blue colours (Stone 2003). The eye's colour sensitivity curve peaks at 555 nm, in the green hue of the spectrum (Gerritsen 1975).

2.1.3 Chromatic adaptation and image display

Chromatic adaptation describes the ability of the HVS to adapt to the colour of the light illuminating the scene (Stone 2003), helping to keep the perceived object colour constant when viewed in different lighting conditions. For instance, an apple may look green at midday when the sun is approximately white in colour, but also at sunset when the illumination has a red cast. This helps the HVS identify objects.

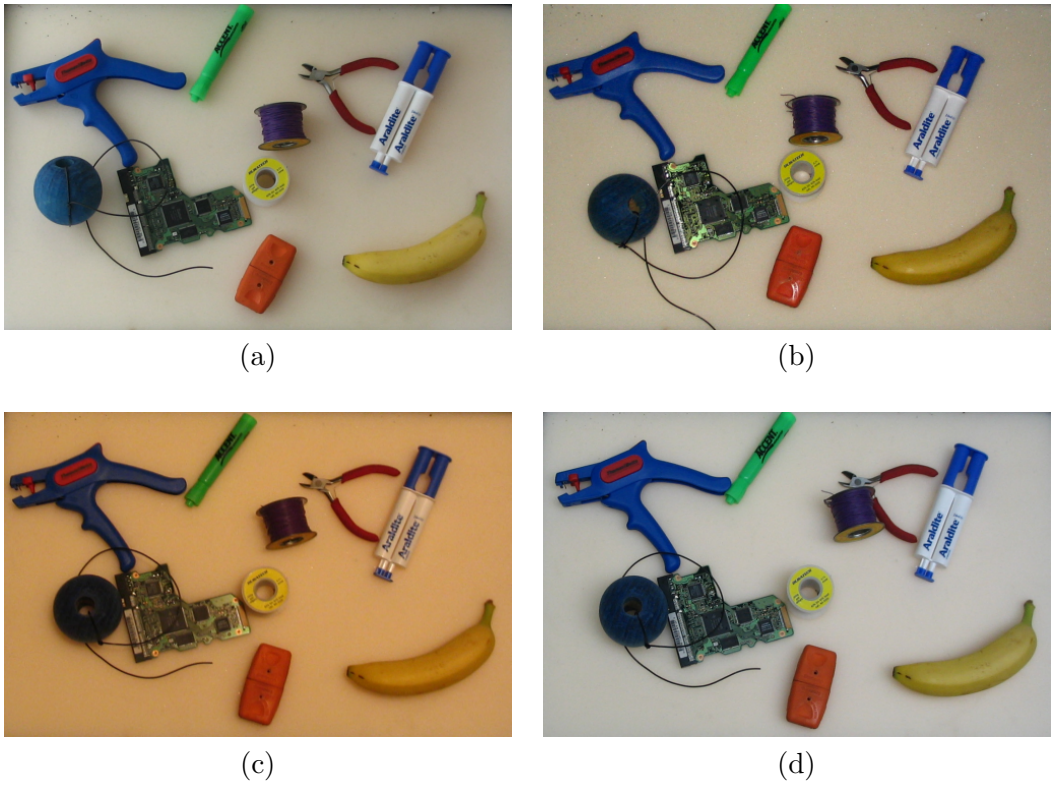
Most consumer-grade digital cameras have an automatic white-balance mode that is designed to adjust the captured images to appear as if illuminated by a pure white illuminant. Virtually all output devices are calibrated to a white reference. For example, the white used in standard monitors is defined as RGB(255,255,255), whereas the white for a standard printer is the colour of the paper. The process of adjusting a captured image so that it appears as if recorded under an ideal white illuminant is known as colour constancy.

The process of achieving colour constancy in digital images is a non-trivial task and is an active area of research (Funt, Barnard & Martin 1998, Barnard, Cardei & Funt 2002). Shown in Figure 2.3 is an example highlighting the difficulty of achieving colour constancy. Each image was taken in different illumination conditions with the camera set to full automatic mode. A perfect colour constancy process would adapt the colouring of each image to an ideal white illuminant yet there remain significant colour differences between the images. The differences would have been greater if the automatic white balance in the camera had been disabled. The eye would adapt to the colouring in each of the images if they totally encompassed the eye's field of view, but the images appear different when placed together on a background because the eye adapts to the paper as a white reference.

2.2 IMAGE CAPTURE

This thesis uses the term 'image' to define pictorial content typically produced by an optical or electronic device. The components required in capturing an image are:

- an object/scene



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

Figure 2.3 Images taken with a Canon Powershot A75 camera in different illumination conditions with automatic white balance enabled: (a) ambient sunlight, (b) flash, (c) incandescent, and (d) fluorescent lighting.

- an illuminant (light source)
- a sensing array

The sensing device, typically an eye or an electronic sensor, is limited in spectral bandwidth and sensitivity and can be affected by metamerism (colour aliasing, see Section 2.3) and saturation. The sensor of a camera system usually modifies the image non-linearly and digitizes the pixel data before export. The representation of an image of the scene will change as it passes through the camera and display system.

Figure 2.4 shows the pathway of scene irradiance from capture to reproduction for a digital imaging system. The scene to be imaged is subject to an illuminant such as the sun or sources of artificial lighting. The reflected scene illumination passes through the camera optics and the incident image falls upon the sensor. The sensor then converts the incident image photons into an electronic signal. This signal is amplified and is referred to as the captured image. Some form of processing typically occurs, such as gamma, gain and white balance adjustment, before the captured image is non-linearly encoded and the recorded image is available for export in a digital form (e.g., raw, JPEG, YC_RC_B). The recorded image is usually stored and/or transmitted to an output device such as a monitor or printer, which creates a rendered image for visual display.

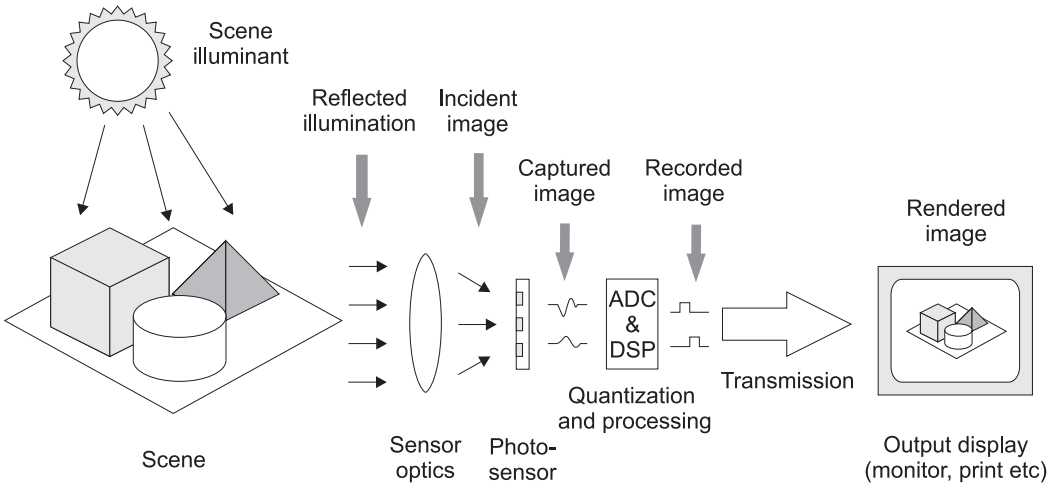


Figure 2.4 The image capture and reproduction pathway.

2.3 METAMERISM

The spectral energy P of an object recorded by a sensor is a function of the illumination E, the object surface response S, and the sensor response C, such that:

$$P(\lambda) = E(\lambda)S(\lambda)C(\lambda) \tag{2.2}$$

where λ is the wavelength of light (Funt & Finlayson 1995). The particular value of P for a specific surface S_1 and illuminant E_1 is not unique, as other surfaces and illuminants S_2 and E_2 will exist such that:

$$\frac{E_1(\lambda)}{E_2(\lambda)} = \frac{S_2(\lambda)}{S_1(\lambda)}. \tag{2.3}$$

This effect is known as metamerism, or colour aliasing. In practical terms it means that the colour image of two objects with different surface colourants may appear identical when placed under different illuminants. For standard colour imaging, the recorded spectral energy is a combination of the responses of the R, G, and B wavelengths:

$$P_{(R,G,B)} = \Big(P(\lambda_R), P(\lambda_G), P(\lambda_B)\Big). \tag{2.4}$$

Metamerism can occur on any one or more of the R, G, and B colour primaries. Multiple objects, having different spectral reflectance curves but the same measured spectral response, are defined as metameric objects (Billmeyer & Saltzman 1966).

2.4 DIGITAL CAMERAS

The purpose of digital imaging is to capture a scene in a digital format for accurate scene reconstruction at a later time. Most digital cameras have the ability to alter characteristics of the image such as white balance or contrast to visually improve the quality of the image when reproduced for the HVS. These adjustments can have significant impact on the machine processing of digital images as the captured image is no longer representative of the incident image.

Charge-coupled devices (CCDs) have traditionally been the dominant image-sensor technology. However, recent advances in the design of complementary metal-oxide-semiconductor (CMOS) technologies have led to its adoption in many consumer-grade devices, as well as being a viable alternative to professional CCD systems (El Gamal & Eltoukhy 2005). Both Nikon and Canon shifted to CMOS sensors in their professional SLR cameras from the D3 (Nikon) and EOS-1Ds Mk-III (Canon) cameras onwards.

2.4.1 A brief history of digital imaging

The following section is summarized from Flory (Flory 1985), Fossum (Fossum 1997), and El Gamal (El Gamal & Eltoukhy 2005).

The entire history of electronic image acquisition spans more than 80 years, starting with the iconoscope which was invented around 1925. The period 1945–1965 saw the era of the orthicon, which was a complex solution to the shortcomings of the iconoscope. Both of these devices used scanning electron beams to scan a photosensitive element. In 1965 the photoconductive sensor was developed, and it became the dominant image sensor technology until 1970, when the CCD was developed. It has since been the dominant sensor technology.

MOS photodiodes were developed in the late 1960s, but MOS image sensors started arriving only in the early 1980s. The arrival of CMOS passive-pixel sensors happened in the mid-1980s, where they were the technology of choice for only a very few scientific fields due to low dynamic range and low signal-to-noise ratios (SNR). In the early 1990s work began on the modern CMOS active-pixel sensor, originally conceived in 1968. It was quickly realized that adding an amplifier to each pixel significantly increases sensor speed and SNR, overcoming the limitations of passive-pixel CMOS. With the recent advent of deep submicron CMOS and integrated microlens technologies, active-pixel CMOS sensors have become viable alternatives to CCD image sensors. CMOS image sensors remain in an active state of research and development, and are likely to be the choice of imaging sensor in the future.

2.4.2 CCD/CMOS image capture

The CCD and CMOS sensor architectures share many similarities. Photons are captured and converted to charge in the photo detection process, the charge is conditioned/sampled, and the samples are digitally enhanced for output. This process is described comprehensively in the literature (El Gamal & Eltoukhy 2005, Flory 1985, Litwiller 2001, Blanc 2001, Kremens 2001, El Gamal, Fowler, Min & Liu 1998b, Catrysse, Wang & El Gamal 1999, Costantini & Süsstrunk 2004, Anaxagoras, Guerrini, Turchetta & Allinson 2005). Figure 2.5 details a typical digital image sensor for a CCD/CMOS digital camera system.

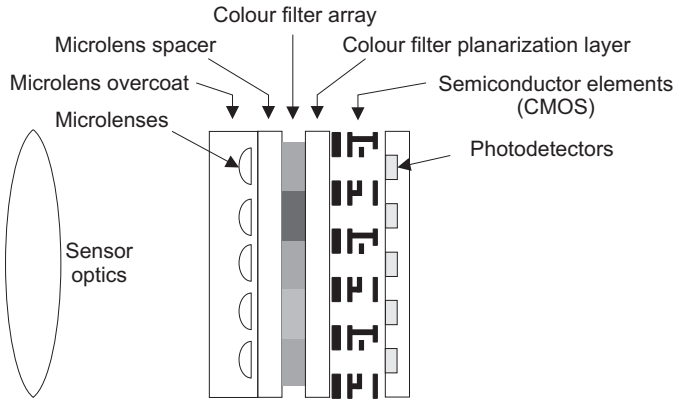


Figure 2.5 The layout of a typical digital image sensor.

Light passing through the optics falls onto the imaging sensor. Some image sensors use microlenses to increase the amount of light incident on the photodetectors (El Gamal & Eltoukhy 2005, Chen, Catrysse, El Gamal & Wandell 2000). This helps reduce the problem of vignetting, where light entering the sensor in a direction that is not parallel to the optical axis does not reach the photodetectors due to the optical tunnel formed by the sensor manufacturing process. For consumer grade colour cameras the light passes through a colour filter array (CFA) to generate trichromatic images. The filtered light then enters the photodetectors, where approximately half of the photons are converted to charge (Brouk & Nemirovsky 2004).

The CCD and CMOS fabrication technologies are significantly different. Specifically, CCDs are manufactured with technologies solely optimized for imaging and charge transfer, whereas CMOS devices are manufactured with mostly standard technologies and have the ability to incorporate logic and analogue processing on the same die as the photodetectors. Apart from the ability to integrate the sensor, control logic, and processing on the same chip, CMOS sensors enable each individual pixel to have its own transistors, amplifiers, charge-to-voltage conversion, or even digital signal processors (DSPs). The drawback is that the pixel fill rate, or the area that can be used to capture photons, is reduced when logic is added alongside each pixel.

Figure 2.6 (left) shows a typical interline CCD readout architecture. The charge from each pixel is read sequentially, with each charge moving along a column or row in a conveyer-type fashion. Analogue-to-digital conversion (ADC), storage, and enhancement are performed on supporting integrated circuits (ICs). There are three major readout architectures used for CCD sensors: full-frame (FF), frame-transfer (FT), and interline-transfer (IL). FF has a fill rate of up to 100% of the pixel area, but is prone to smearing during charge readout. FT shifts the image to a second image buffer at high speed prior to readout, thereby reducing image smear. IL CCD is the most popular image sensor for camcorders and digital still cameras, but suffers from a reduced fill rate due to charge storage buffers located beside each pixel (Nakamura 2006, *Kodak CCD Primer KCP-001* 1999). This reduction in fill rate varies, but fill rates between 20 and 50% are not uncommon. A microlens is often used to compensate for this reduction by increasing the effective area of incident light collection.

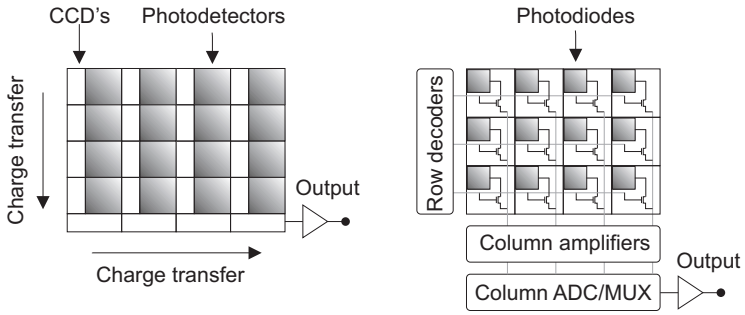


Figure 2.6 Typical architectures of interline CCD (left) and passive-pixel CMOS (right).

Current CMOS sensors can be divided into two main architectural types: passive-pixel sensors (PPS) and active-pixel sensors (APS) (Nakamura 2006, Yadid-Pecht & Fish 2004, El Gamal & Eltoukhy 2005). Figure 2.6 (right) shows a typical readout architecture for a CMOS PPS. Each pixel contains a photodiode and one MOS transistor. As in most CMOS image sensors, row and/or column decoders are used for addressing the pixels in the array. Although they have relatively high fill rates due to having only a single transistor, PPS devices suffer from high noise due to large capacitive bus loads during pixel readout. They are also prone to column fixed-pattern noise (FPN) from variations in the manufacturing process of the column amplifiers, which can result in objectionable vertical stripes in the recorded image.

CMOS APS can be divided into three main subtypes: photodiode, photogate, and pinned photodiode, where in each type three or more transistors are used in each pixel. APS devices typically have a fill rate of 50-70%, but the reduced capacitive bus load during readout due to the transistor amplifiers leads to lower readout noise, which increases the signal-to-noise ratio (SNR) and sensor dynamic range (DR). The pinned-photodiode APS has currently been reported as the most popular implementation for

CMOS image sensors (Nakamura 2006).

There are other types of APS available but they are currently not in widespread use. The logarithmic photodiode sensor (Loose, Meier & Schemmel 2001) operates continuously and provides increased dynamic range from logarithmic encoding of the photocurrent. However, low output sensitivity during low illumination and significant temperature dependence are serious drawbacks limiting the use of this method. Fowler et al. describe a 'digital-pixel sensor' that has 22 transistors and ADC functionality at each pixel (Fowler, El Gamal & Yang 1994).

2.5 LIGHT & SHADOWS

Light is a small component of the electromagnetic spectrum, consisting of two perpendicular oscillating magnetic and electric waves (Parisi, Sabburg & Kimlin 2004). However, light can exhibit properties of both waves and particles. This property is referred to as wave-particle duality, with references found in many textbooks, such as (Ditchburn 1976).

In this thesis light is referred to using both wave properties and as individual photons (light carriers). For example, in Section 2.5.1 light is discussed in spectral wavelengths, yet in Section 3.1.1.3 it is treated as an individual unit of energy that generates an electron, which itself exhibits wave-particle duality, in the photosensitive wells of a camera.

2.5.1 Colour temperature

An illuminant's colour can be described loosely by its colour temperature (Stone 2003), which is based on the colour output of a black-body radiator at a specific temperature. In general, low colour temperatures represent a reddish-yellow illuminant and high colour temperatures represent bluish-white illuminants. For example, candle flames have a relatively low colour temperature of 2,000 K. The tungsten in incandescent lighting has a colour temperature ranging from 2,500–2,800 K. Halogen lights range from 2,800 K to around 3,200 K, and fluorescent lamps typically operate between 2,500–6,500 K. Sunlight at midday light is approximated by a 6,500 K colour temperature.

The CIE-defined daylight spectrum for midday light is labelled D65. This term is sometimes used interchangeably with 6,500 K but the colours defined by the CIE are different from black-body radiators because they simulate not only the colour of the sun (a black body radiator of approximately 5,800 K) but also the scattering effect of the atmosphere.

The spectral emission of high-intensity discharge (HID) and fluorescent lamps are not appropriately described by colour temperature alone when used in photography, as their spectrum is composed of specific descrete wavelengths.

Figure 2.7 shows the CIE 1931 chromaticity space which is often used to approximately describe colours. The Planckian locus describes the colour emitted by a black-body radiator at increasing temperature. Colours falling in the vicinity of the lines crossing the locus can be described as a correlated colour temperature of the associated line's intersection with the locus.

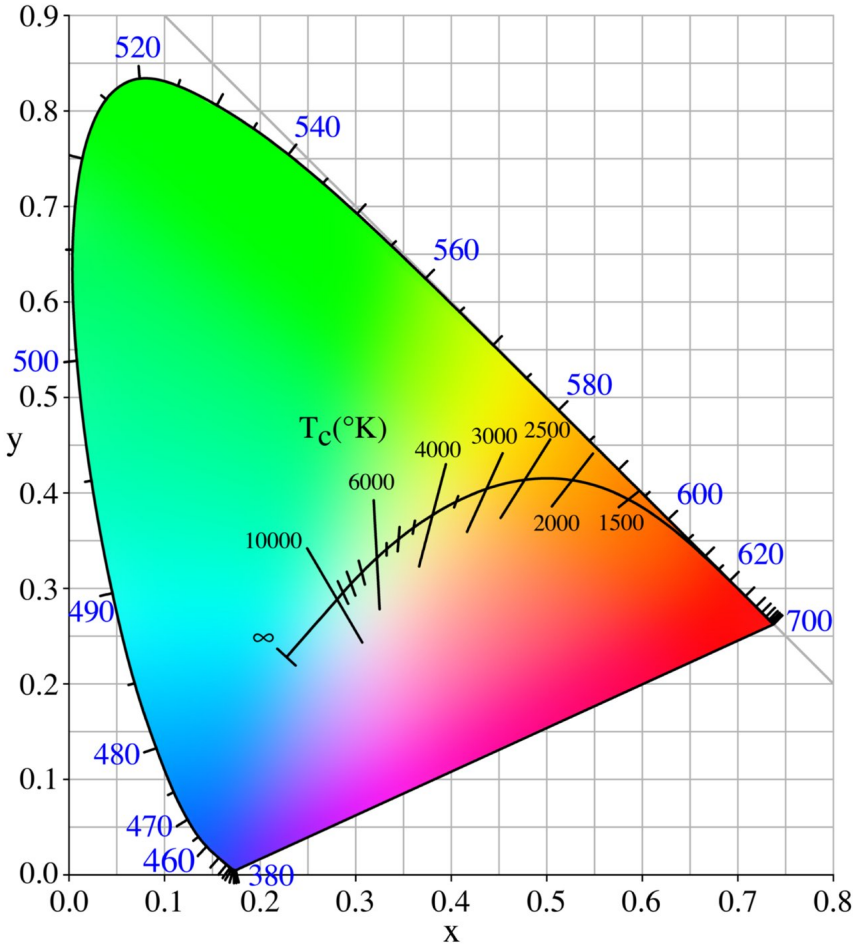


Figure 2.7 The CIE 1931 x,y chromaticity space, also showing the chromaticities of black-body light sources of various temperatures (Planckian locus), and lines of constant correlated colour temperature. (Image PlanckianLocus.png from http://www.wikipedia.org. Permission to freely distribute this image under the GNU Free Documentation license, Version 1.2 or later).

2.5.2 Sunlight, skylight, and daylight

The sun emits radiation that roughly approximates a black-body radiator from 5750–5800 K (Henderson 1977, Iqbal 1983). The intensity and spectrum of daylight changes primarily with atmospheric conditions (ignoring changes in the solar constant). The earth's albedo is roughly 40% (approximately 40% of the incoming solar radiation is reflected back into space). About 15% is absorbed in the atmosphere, and around 45%

reaches the earth's surface, where it is absorbed and re-radiated at infra-red wavelengths (Henderson 1977). The atmospheric filtering process is non-linear, hence the apparent colour of sunlight when viewed from Earth is typically yellow depending upon the number and size of particles in the atmosphere.

According to Rayleigh's theory of molecular scattering, the probability that a single photon of sunlight will be scattered from its original direction by an air molecule is inversely proportional to the fourth power of the wavelength (Lynch & Livingston 1995). That is, the shorter the wavelength of light (i.e., the bluer the light), the higher its chances of being scattered. This causes the sky to appear to irradiate a blue colour, called skylight, which is the Rayleigh-scattered component of the sun's light passing through the atmosphere. Figure 2.8 illustrates the illumination of earth from the sun, and the colouring of cast shadows from skylight.

On earth, daylight is a combination of sunlight and skylight. The colour of the daytime sky in clear conditions changes with the viewing angle. The sky viewed directly upwards is typically deep blue in appearance, with a 'fading' to a lighter shade of blue or even white when looking towards the horizon. Each molecule in the atmosphere can be considered a source of Rayleigh-scattered light. Looking towards the horizon increases the quantity of air mass that is viewed when compared to looking directly upwards, increasing the amount of scattered light seen by the observer. The maximum viewing distance before all light is scattered and the view becomes opaque is dependent upon wavelength. The shorter the wavelength, the shorter the distance until the transmitting medium becomes optically thick. As the view of the sky is shifted from directly above the observer towards the horizon, the blue end of the spectrum is the first to become opaque (i.e, maximum brightness from Rayleigh scattering), followed by green then red, until all wavelengths become opaque and the skylight illumination appears white. The red or yellow sky at dawn and dusk is primarily from the added scattering of light from dust, smoke, and water particles in the air. These particles attenuate the sunlight after the Rayleigh scattering has removed the shorter wavelengths of light (Lynch & Livingston 1995).

2.5.3 Artificial illumination

Incandescent and halogen light bulbs are essentially black-body radiators that use heated filaments to produce light, and compared to more modern methods of lighting are relatively crude and inefficient as a large amount of energy is lost as heat. They emit light over a broad spectrum extending beyond the 350–700 nm range of the HVS.

Fluorescent and high-energy discharge (HID) lamps emit light efficiently but only at specific wavelengths. Monochromatic LEDs are narrow band, and the increasingly popular white LEDs have spectral emission curves closer to that of the HVS.

The effect of the significant differences in spectral emission curves of the different

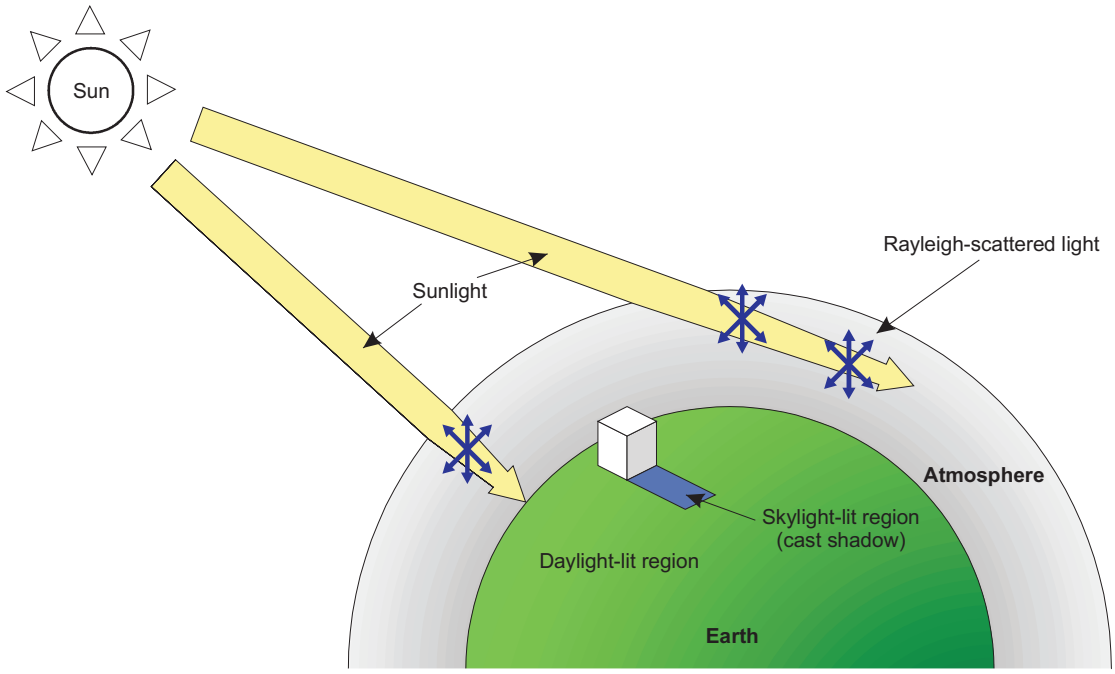


Figure 2.8 The generation of daylight illumination on earth.

forms of artificial illuminants is in part negated by the band-pass filtering of light into the three primary colours in the eye and in standard digital cameras.

2.5.4 Shadow classification

Shadows are a natural and common occurrence in almost every environment, and are caused by the occlusion of light. Shadows are typically categorised as being either 'cast' or 'self' shadows. Cast shadows are generated by objects partially or completely occluding the illuminating light sources, whereas self-shadows describe the shading that occurs on an object from the object itself, e.g., the side of a box on the opposite side from the illuminating source.

Cast shadows can contain areas of umbra, or 'hard' shadows, and penumbra or 'soft' shadows. The umbra is the area completely occluded from the light source. The penumbra is the area that is partially occluded from the light source because the light source (non-point) is only a finite distance away. In reality, all natural images will contain shadows with both umbra and penumbra shadows, although the size of the penumbra in an image may be very small depending upon the resolution, field-of-view of the imaging device, and the size and distance of the light source.

2.5.5 Shadows and imaging

The digital image capture of any non-trivial scene introduces several complexities that can confound the shadow identification and/or removal process:

- Imaging devices have limited dynamic range and may not be able to capture enough scene information to allow shadow algorithms to perform robustly.
- Many imaging devices automatically adjust settings and parameters so that the device's response to the given scene is maximized. Hence colour balance, focus, gain etc., may be used to alter the content and quality from image to image.
- The image noise may change from image to image, as it is highly dependent upon camera settings, parameters, temperature and scene content.

Part II

Noise analysis

Chapter 3

CAMERA NOISE

Digital cameras are commonly used as a data source in many image processing applications. The algorithms in these applications operate on the contrast differences between pixels, either across an image or between images. Fundamentally, the detectable difference in contrast between one pixel and another is limited by the signal-to-noise ratio (SNR) within the captured images. This chapter contains a detailed investigation into camera noise, particularly for the common CCD- and CMOS-based devices. On the path to a deeper understanding of camera noise, a comprehensive CCD noise model has been developed, along with methods to analyse and measure camera noise from the output images alone.

The chapter is organised as follows. Section 3.1 reviews the existing literature on camera noise and states the resulting objectives of the camera noise research. Section 3.2 presents a new CCD noise model, and Section 3.3 describes a process for measuring the CCD noise components from analysis of output images alone. Section 3.4 shows the resulting calibration of the noise model from the measurement of the noise components, and Section 3.5 presents a new CMOS noise model. Section 3.6 describes a process for measuring spatial, temporal, and combined noise for any set of images. The effect of temperature on camera noise is investigated in Section 3.7, and a discussion in Section 3.8 completes the chapter.

3.1 LITERATURE REVIEW

While most individual sources of camera noise are well understood the complete noise path from photon capture to digital image has not been extensively investigated. Section 3.1.1 describes the individual noise sources present in images captured with digital cameras, and Section 3.1.2 provides a review of the camera noise models presented to date. This is followed by a discussion and a set of objectives in Sections 3.1.3 and 3.1.4. The terminology used for sources of image sensor noise varies and is inconsistent. It depends upon the author and the approach used. Multiple terminology will be provided where appropriate.

3.1.1 Sources of noise in images captured with digital cameras

3.1.1.1 Offset fixed-pattern noise

Offset fixed-pattern noise (FPN) arises from changes in dark currents caused by device mismatches and variations in pixel geometry during sensor fabrication. Dark currents are leakage currents produced by surface generation and minority carriers thermally generated in the sensor well (El Gamal & Eltoukhy 2005). The expected value for a pixel's dark current is constant for a given operating condition although it increases with exposure time (Nakamura 2006). FPN is the pixel-to-pixel variations that occur across an image sensor, and is alternatively referred to as Dark Signal Non-Uniformity or DSNU (Costantini & Süsstrunk 2004, Chen et al. 2000, Kremens 2001). It increases exponentially with temperature and can be measured in dark conditions. Furthermore, CMOS image sensors have additional offset FPN caused by variations in pixel-level transistor characteristics.

3.1.1.2 Photo response non-uniformity

Photo response non-uniformity (PRNU) describes the difference in pixel responses (gain) to uniform light sources. In uniform illumination conditions, each photodetector cell of an image sensor should exhibit the same output voltage. However, this is not the case due to variations in pixel geometry, substrate material, and microlenses. Since PRNU is caused by the physical properties of a sensor, it is nearly impossible to eliminate and is therefore usually considered a normal characteristic of the sensor array used in any CCD or CMOS camera. The effect of PRNU is proportional to illumination, and is prominent under high illumination levels (El Gamal & Eltoukhy 2005). Variations in gain between transistors in CMOS image sensors generate additional illumination-dependent fixed-pattern noise. PRNU is sometimes referred to as gain fixed-pattern noise.

3.1.1.3 Shot noise

Shot noise is a Poisson process that arises from random fluctuations in sampling when discrete quanta are measured. Significant shot noise sources in an image sensor are in the capturing of photons in the photon-detection stage, in the temporal variation of dark currents, and in transistor semiconductors. Any electronic conductor exhibits shot noise when charges act independently when moving, e.g., when charges cross a barrier in a semiconductor. However, the current movement in a metallic conductor exhibits long-range correlations and therefore far less noise than within semiconductors (Horowitz & Hill 1989), so shot noise due to current flow through simple conductors is not considered a significant source of noise within an image sensor.

Dark current shot noise occurs due to leakage currents in each pixel of an image sensor. The number of dark-current shot noise electrons doubles with every 8°C rise in temperature (Nakamura 2006, Healey & Kondepudy 1994), and is proportional to the pixel integration time. Photon shot noise is dependent upon the mean number of captured photons, and therefore increases with sensor irradiance.

3.1.1.4 Readout noise

Readout or read noise is generally defined as the combination of the remaining circuitry noise sources between the photoreceptor and the analogue-to-digital conversion (ADC) circuitry. Readout noise varies only in the temporal domain, and includes pixel reset noise, thermal noise sources (Johnson–Nyquist), and other minor contributors like the frequency-dependent $1/f$ (flicker) noise sources and conductor shot noise.

Reset noise in CCDs is a specific type of thermal noise arising from the kT/C fluctuations when resetting the charge sense capacitor to a reference voltage (Hynecek 1990). A commonly used method called correlated double sampling (CDS) is used to reduce the effect of reset noise (Nakamura 2006, Dalsa 2006). CDS samples the noise value on the sensing capacitor after reset, and subtracts it from the sample of pixel data after charge transfer.

Thermal, or Johnson–Nyquist noise, arises from fluctuations of an electric current inside an electrical conductor due to the random thermal motion of the charge carriers. It is independent of illumination and occurs regardless of any applied voltage (Horowitz & Hill 1989, Ott 1988).

$1/f$ or flicker noise is generated in the photo-diodes and in the low-bandwidth analogue operation of MOS transistors due to imperfect contacts between two materials (Ott 1988, Nakamura 2006). It is pink-coloured (the power spectral density is inversely proportional to the frequency), and the level of noise is dependent upon the frequency of the pixel sampling rate. Flicker noise is not completely understood, and is still a subject of analysis and debate (Tian & El Gamal 2000).

3.1.1.5 Column noise

CCD sensors usually have two column amplifiers, one for odd columns and one for even columns (Diwa 2007). Imperfect manufacturing can result in different gains for the two amplifiers, which may result in column-to-column variations.

In CMOS image sensors, the column parallel readout architecture is reported to be used most commonly (Nakamura 2006). Pixels are read out in a row simultaneously and processed in parallel allowing for faster processing (or reduction in power consumption by using lower clock frequencies) and to reduce the number of bus lines going to each

pixel. Variation in the performance of the parallel segments introduces offset and gain fixed-pattern noise that cause vertical stripes to appear in the image.

3.1.1.6 Demosaicing

Many single-sensor colour cameras use colour filter arrays (CFAs) to restrict the pixel bandwidths to a particular range in the optical spectrum. A commonly used CFA is the Bayer matrix shown in Figure 3.1 that reduces each pixel's bandwidth to approximately 1/3 of the visible wavelengths of light. A method of colour interpolation called demosaicing is then employed to generate full-colour values at each pixel in an image (Gunturk, Glotzbach, Altunbasak, Schafer & Mersereau 2005). The demosaicing process used to interpolate the colour data for each pixel can be application dependent, with the methods used in consumer-grade digital cameras (for example general purpose personal cameras, web cameras and cell phone cameras) to create visually pleasing images often being different from those in industrial-grade cameras which try to recreate accurate colour data. A common method used in industrial-grade cameras and in lower cost cameras in general is the relatively simple bilinear interpolant, which has low computational requirements. Figure 3.2a illustrates bilinear-interpolation of green pixels of the Bayer matrix. Pixel G_a is calculated as the mean of the four surrounding green pixel values:

$$G_a = \overline{G_1 + G_2 + G_3 + G_4}. \tag{3.1}$$

The bilinear interpolation of the red and blue channels is slightly more complex, due to the relatively sparse number of red and blue samples on the quincunx grid shown in Figure 3.2b. Pixels R_a, R_b, R_d, and R_e are interpolated from their two nearest neighbour red pixels:

$$\begin{aligned} R_a &= \overline{R_1 + R_2} \\ R_b &= \overline{R_1 + R_3} \\ R_d &= \overline{R_2 + R_4} \\ R_e &= \overline{R_3 + R_4} \end{aligned} \tag{3.2}$$

whereas R_c is interpolated as the mean of the 4 closest red pixels:

$$R_c = \overline{R_1 + R_2 + R_3 + R_4}. \tag{3.3}$$

Bilinear interpolation of the blue channel is performed using the same process as the red channel.

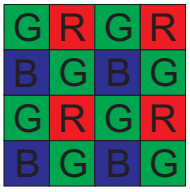
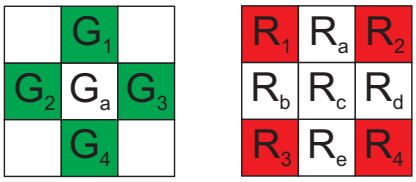


Figure 3.1 The Bayer colour filter matrix.



(a) (b)

Figure 3.2 The demosaicing of the green (a) and red (b) channel of the Bayer matrix, where numbered pixel-subscripts represent captured pixel data, and lettered pixel-subscripts represent pixels that require estimation.

There are many other methods for demosaicing, e.g., (Gunturk et al. 2005), but they are often computationally expensive and designed to optimise the resulting interpolated image for the human vision system.

3.1.1.7 Quantization

Digital images are often quantized to 8–16 bits per colour channel for export from the camera. Where the quantization step is very small compared to variations within the image, the quantization process adds noise (σ) to the image (Baher 1990):

$$\sigma^2 = \frac{q^2}{12} \tag{3.4}$$

where q is the quantizing step (dimensionless).

3.1.1.8 Summary

The noise sources described above can be segmented into illumination independent, illumination dependent, and digital processing categories. A summary is given in Tables 3.1–3.3.

Table 3.1 The illumination-independent noise types.

Noise type	Origin	Manifestation	Dependencies
N_{read}, readout noise (thermal noise N_{therm}, reset noise N_{reset}, and other minor contributors N_{other})	CCD Sensor and CCD support ICs	Additive temporal and spatial variance	Temperature, CCD readout rate
SN_{dark}, dark-current shot noise	CCD sensor	Additive temporal and spatial variance	Temperature, exposure time
FPN, offset fixed-pattern noise	CCD sensor	Additive spatial variance only	Temperature, exposure time

Table 3.2 The illumination-dependent noise types.

Noise type	Origin	Manifestation	Dependencies
$PRNU$, photo-response non-uniformity	CCD support ICs	Multiplicative spatial variance only	Incident pixel illumination
SN_{ph}, photon shot noise	CCD sensor	Additive temporal and spatial variance	Incident pixel illumination

3.1.2 Existing camera noise models

The literature reveals several noise models for camera sensors, both partial and full, that attempt to explain the sources and their relevance to the output from the camera. Very few models were experimentally validated, and no model was found that included all sources and effects. Many of the earlier noise models are based upon analogue-output television cameras, and were either incomplete or were unable to be validated due to limitations in digital image capture technology. A list of each of these camera noise models follows in chronological order. Noise types denoted with an asterisk are described in the existing literature using different terminology than that used in this thesis.

Flory (Flory 1985) describes the primary electronic sensor noise as a combination of shot and amplifier noise. This relatively early paper does not describe camera noise in detail.

Boie and Cox (Boie & Cox 1992) analyze camera noise for CCD and vidicon cameras. Their model includes photon shot and electronic shot noise, column noise, dark-current, and readout noise, and results suggest noise is localized and stationary.

Healey and Kondepudy (Healey & Kondepudy 1994) model camera noise for measuring scene variation, and include FPN*, photon and dark-current shot noises, PRNU*, a read-noise equivalent, and quantization noise. They produce a general estimate of spatial and temporal noise which were measured, though the effect of quantization

Table 3.3 The digital processing noise types.

Noise type	Origin	Manifestation	Dependences
N_D, demosaicing noise	CCD support IC	Multiplicative noise amplification or attenuation	Demosaicing implementation, combined sensor noise
N_{filt}, digital filter noise	CCD support IC	Multiplicative noise effect	Gain parameters for enhancement, combined sensor noise
N_Q, quantization noise	CCD support IC	Additive noise. Image content dependent	Variance of image data. Sets lower noise limit for non-trivial image content

noise was not actually included in the process. Individual noise components were not evaluated. Their analysis was in part restricted by the use of analogue frame-grabbers in the 1990s, which reduced the integrity of the signals after export from the camera.

Mendoza (Mendoza 1997) developed a model to analyze intrinsic CCD noise for application in pressure sensitive paint instrumentation systems, and includes thermal noise, reset noise, photon shot noise, and PRNU.

El Gamal et al. (El Gamal, Fowler, Min & Liu 1998a) model FPN noise components in CMOS image sensors. Though they do not model the whole noise path from photon integration to image output, their work is included for completeness.

Catrysse et al. (Catrysse et al. 1999) developed a CMOS image sensor noise model to measure colour reproduction errors and analyze the contributions of different sources to the error. Their model includes photon and electronic shot noise, readout noise, offset FPN, and quantization.

Tsin et al. (Tsin, Ramesh & Kanade 2001) reference and use Healey and Kondepudy's noise model and add a colour-balance term to it. They group several noise sources together for experimental measurement, much like Healey and Kondepudy. They do not validate their model, but provide the estimated results of noise from a series of images. This is the only method that uses an iterative technique to estimate parameters of their model.

Tian et al. (Tian, Fowler & Gamal 2001) analyze temporal noise only in CMOS photodiode APS sensors, and model reset noise as dark current, photon and electronic shot noise in the reset transistor.

Farrell et al. (Farrell, Xiao, Catrysse & Wandell 2004) describe a digital camera simulation tool which includes optics, sensor, and processor modules. The sensor module models sources of noise including readout noise, dark-current, offset FPN*, PRNU, and photon shot noise.

Costantini and Süsstrunk (Costantini & Süsstrunk 2004) developed an image noise

model for simulations that include readout noise, photon and electronic shot noise, FPN*, PRNU, and dark-current shot noise.

Finally, El Gamal and Eltoukhy (El Gamal & Eltoukhy 2005) describe sensor noise at the end of charge integration as the sum of electronic shot noise, reset noise, readout noise, offset FPN, and PRNU*. They do not provide a model, but are included for completeness.

The noise models described above are summarised in Table 3.4.

Table 3.4 A list of camera noise-models with their noise sources and effects.

Model	**FPN**	**PRNU**	**Photon shot noise**	**Dark current shot noise**	**Readout noise**	**Column noise**	**Demosaic effects**	**Quantiz. noise**	**Model Validated?**
Flory 1985			Y		Y				
Boie & Cox 1992			Y	Y	Y	Y			
Healey & Kondepudy 1994	Y	Y	Y		Y			Y	(2)
Mendoza 1997		Y	Y		Y				
El Gamal, Fowler, Min & Liu 1998 (1)	Y					Y			
Catrysse et al. 1999	Y		Y	Y	Y			Y	
Tsin, Ramesh & Kanade 2001	Y	Y	Y		Y		Y	Y	(2)
Tian, Fowler & Gamal 2001 (1)					Y				
Costantini & Süsstrunk 2004	Y	Y	Y	Y					(2)
Farrell, Xiao, Catrysse & Wandell 2004	Y	Y	Y	Y	Y				
El Gamal & Eltoukhy 2005 (1)	Y	Y		Y	Y	Y			N/A

(1) Partial noise model only

(2) The combined model was tested experimentally as a whole, but not validated.

3.1.3 Discussion

None of the models described in Table 3.4 incorporates all of the noise sources described in Section 3.1.1. Further, no models were validated. Several of the existing models take some measurement data and then apply their parameters as a 'best fit', though usually with grouped noise types (e.g., all temporal noise, or all spatial noise). In these cases either high-end professional cameras (Costantini & Süsstrunk 2004, Tsin et al. 2001) or custom-fabricated sensors (Tian et al. 2001) were used.

The models referenced in Table 3.4 cover both CCD and CMOS technologies that have been developed over the past 20 years, and the more comprehensive papers discuss the prominent noise sources at various illumination levels and arrive at similar conclusions (El Gamal & Eltoukhy 2005, Flory 1985, Mendoza 1997): low-illumination level noise is dominated by read noise and dark current; medium-illumination level noise is dominated by photon and electronic shot noise; and high-illumination level noise is dominated by PRNU.

3.1.4 Objectives

The goal of this section of work is to provide a greater understanding of image noise in standard commercially available cameras. There are no existing models that include all of the noise sources present in digital cameras. Consequently, analysis of individual noise components is not possible without measurement of noise sources in the electronic circuitry, and methods for measuring individual or combined noise components from the output images alone cannot be developed until complete noise models, from photon-to digital image, have been defined.

The objectives for noise-related camera research are:

- To develop comprehensive noise models for the commonly available CCD and CMOS cameras.
- To develop a non-invasive method to validate the model(s), if possible, using only the output images from the cameras.
- To develop methods for measuring combined total sources of noise for any camera, in a form suitable for use in image processing applications.

Further, many sources of camera noise are dependent upon temperature, and given that successful methods are developed to measure combined sources of noise, a further objective is:

- To investigate the effect of temperature on camera noise.

3.2 CCD NOISE MODEL

A new model based on the physical characteristics in the acquisition and sampling of photons through the image capture process in a digital CCD camera is developed by integrating the effect of each noise source as photons pass through the CCD camera, from the CFA through to digital output. Although blooming and smearing in CCDs are destructive to images, they are artefacts of charge overflow and are avoidable. Hence they are not considered as part of this noise model.

After passing through the sensor optics, the photons are usually filtered through the CFA to allow colour information to be captured by the photodetectors. The photon to electron conversion is linear (Nakamura 2006), so after passing through the CFA, the captured image I becomes:

$$I_{CFA,k,t} = I \times CFA_k, \quad k = 1, \ldots, n \tag{3.5}$$

where CFA_k is the kth colour filter on the image sensor, and t is time.

The photons captured in the image $I_{CFA,k}$ are subject to variations in gain as a result of PRNU. The noisy image, I_{PRNU}, after incorporating the effect of PRNU (and dropping the subscript k from hereon for clarity) for each colour channel becomes:

$$I_{PRNU} = I_{CFA} + I_{CFA} \times PRNU. \tag{3.6}$$

The capture of photons is a Poisson process, with noise variance equalling the mean sample photon count. The image captured after photon conversion, I_{ph}, in the wells of the photodetectors includes photon shot noise SN_{ph}:

$$I_{ph} = I_{PRNU} + SN_{ph}(I_{CFA}). \tag{3.7}$$

In addition to the signal I_{ph}, dark currents in the well sensor generate additive electrons that vary from pixel-to-pixel. The noisy image captured, I_{FPN}, after inclusion of FPN is:

$$I_{FPN} = I_{ph} + FPN. \tag{3.8}$$

The dark currents that generate FPN themselves also contain shot-noise variations in the number of electrons generated. The noisy image capture model, I_{dark}, after the addition of dark current shot noise, SN_{dark}, is:

$$I_{dark} = I_{FPN} + SN_{dark}. \tag{3.9}$$

The electrons stored in the well of each pixel are then read out and amplified. The

noisy image model, I_{read} , after the addition of read noise, N_{read} (reset noise, thermal noises, and other minor contributors), is:

$$I_{read} = I_{dark} + N_{read}. \tag{3.10}$$

Variations in column amplifiers can result in different gains, adding column noise N_C:

$$I_N = I_{read} \times N_C, \tag{3.11}$$

where I_N is the noisy image after incorporation of column noise. Following read out of pixel data the camera may demosaic the image and apply filtering. The demosaicing effect, N_D, of the the commonly-used bilinear method is proportional to image content and multiplies the noise as follows:

$$I_D = I_N \times N_D, \tag{3.12}$$

where I_D is the noisy image captured after demosaicing.

The effects of digital modification such as image gain and colour balance multiply image content and therefore image noise[1]. The equation for noisy image capture, I_{filt}, after digital filtering is:

$$I_{filt} = I_D \times N_{filt}. \tag{3.13}$$

Finally, the image is quantized for export out of the system. For a quantization step of $q = 1$, Equation 3.4 becomes

$$\sigma_{quantization} = 0.29. \tag{3.14}$$

The result is an additive noise source dependent upon image content, N_Q, that completes the noisy image capture model, I_{cap}:

$$I_{cap} = I_{filt} + N_Q(I_{filt}). \tag{3.15}$$

Unwrapping each of the noise source stages gives the full equation for a noisy image:

$$\begin{aligned} I_{cap} = &(I_{CFA} + I_{CFA} \times PRNU + SN_{ph}\left(I_{CFA}\right) + FPN + SN_{dark} + N_{read}) \\ &\times N_C \times N_D \times N_{filt} + N_Q\left(I_{filt}\right), \end{aligned} \tag{3.16}$$

[1]The common brightness or black level adjustment simply alters the DC offset of the image and does not affect its noise detail.

for each colour channel defined by the CFA. A diagram of the model is given in Figure 3.3.

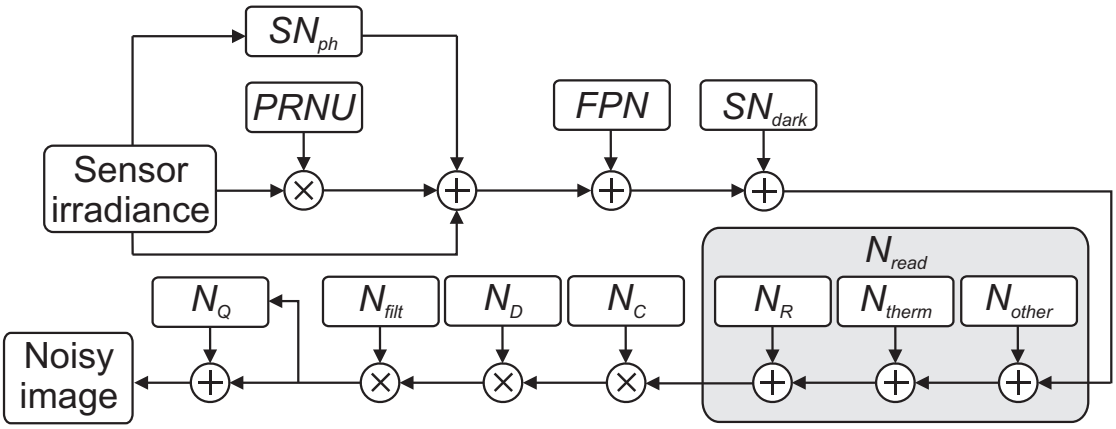


Figure 3.3 The noise model for image capture in a standard CCD digital camera.

3.3 MEASUREMENT OF CCD CAMERA NOISE

The second objective for Part II is to develop methods of validation for the defined noise models. This section attempts to validate the CCD noise model defined in Section 3.2 by use of measurements made on output images from a camera. The measurement of image noise requires a series of static images containing areas of constant detail. The Gretag Macbeth Color Checker (GMB chart), shown in Figure 3.4, provides a useful series of colour patches suitable for noise measurement. The lower portion of the chart consists of grey-scale panels that provide a range of reflectances (3.1 to 90.0 CIE Y values) that can be used for analysis of both colour and monochromatic cameras. Sample sets of images were captured for noise measurement analysis.

3.3.1 Sample requirements

Standard statistical methods can determine an appropriate number of samples required to achieve a desired confidence interval and error for the observed sample mean (Bhattacharyya & Johnson 1977).

For a population with an unknown mean μ and unknown standard deviation σ, the confidence interval for the population mean, based on a random sample of size n, is:

$$\overline{X} \pm t \times \frac{s}{\sqrt{n}} \tag{3.17}$$

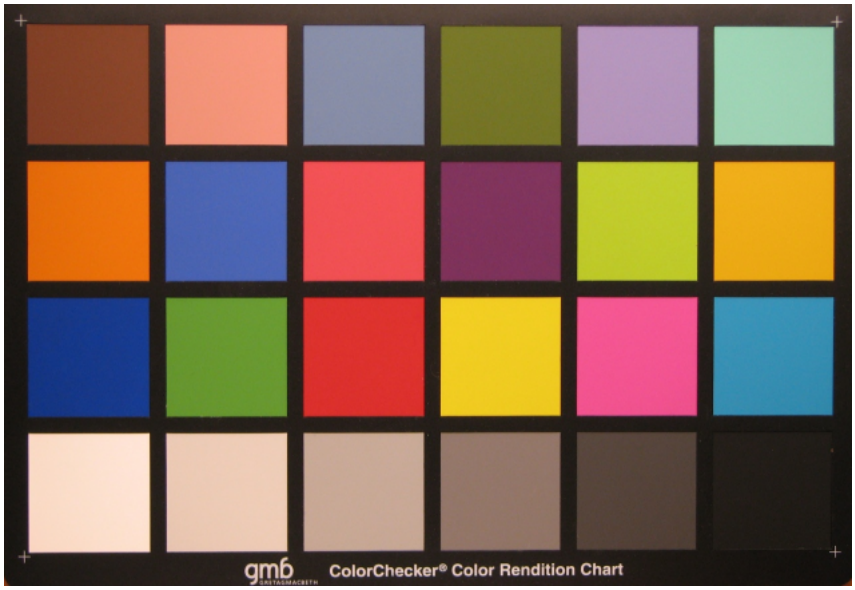


Figure 3.4 The Gretag Macbeth colour chart.

where $\overline{X}$ is the mean of the sample population, s is the estimated standard deviation derived from the sample population, and t is the $(1 - C)/2$ critical value for the t-distribution with $n - 1$ degrees of freedom with C as the desired confidence interval.

A standard confidence interval of 95% is used for this research. It was observed that s is rarely greater than five in most digital video images (with cameras in 'auto' mode, data values 0-255) in canonical lighting conditions. A suitable measurement or error is the quantization step in the digital data, which is a value of one. From expression (3.17) above, this yields the following upper bound for the error in the mean:

$$1 > t \times \frac{5}{\sqrt{n}}. \tag{3.18}$$

The t-distribution is a function of n and C. Using standard t-distribution tables and choosing $n = 100$ gives a value of 1.962. Substituting these values into the right hand side of (3.18) gives a value of 0.981 and inequality (3.18) is satisfied. Note that this holds only for estimated standard deviations less than or equal to 5. Therefore an appropriate sample size to achieve a 95% confidence interval for noise measurement is 100 samples.

3.3.2 Measurement process

Experiments were conducted to validate the CCD noise model using a Unibrain Fire-i400 colour camera (Table 3.5), defocused to reduce the effect of high-frequency scene content in the observed image that could affect the noise analysis. All digital camera

effects such as colour balance and gamma were disabled in the experiments and images were transferred to a computer in the RGB format, so that $N_{filt} = 1$.

Video images of a GMB chart were taken with controlled fluorescent and incandescent lighting and the chart was positioned to fill the image frame. The experimental setup is shown in Figure 3.5 where the illumination sources were positioned above the camera and directed towards the chart such that the image was free from direct specular reflection. Simulated images were generated for the analysis of quantization and demosaicing effects (discussed ahead in Section 3.3.3).

Table 3.5 Unibrain Fire i400 camera details.

Parameter	**Value**
Sensor type	Sony Wfine ICX098BQ 1/4" colour CCD (Bayer colour filter)
Native resolution	CCD 640 × 480
Video mode	24-bit RGB (8–bits per colour channel)
Interface	IEEE-1394a (Firewire)
Operating temperature	−10°C to 50°C

Stuck or 'hot' pixels were identified by taking long exposure images, increasing gain, and detecting saturated pixel values. The pixel data at hot-pixel locations were removed from any subsequent processing.

An area of 50 × 50 pixels was extracted from the centre of the image of each GMB grey-scale panel for analysis. An example panel extraction image is shown in Figure 3.6, and the resulting extracted panels are shown in Figure 3.7. Experiments were performed at an ambient temperature of approximately 22°C. The CCD temperature was measured using a thermistor placed on the CCD sensor surface, and was found to be consistently 7°C above the ambient temperature.

Methods for analysis of temporal noise, spatial noise, and combined noise were formulated for each row and column of image data. The standard deviation (σ) was chosen as the measure of image noise as it can be easily related to the magnitude of pixel variations. The mean value of each panel, μ_p, taken as the mean of the extracted 50 × 50 panel, was used as a measure of I. The following describes the method used to calculate each noise value:

(i) Temporal noise, $\mu(\sigma_t)$:

1. σ of temporal data, σ_t, for each pixel was calculated over the set of 100 images.
2. σ_t for all pixels were then averaged, giving a value $\mu(\sigma_t)$ for the mean temporal variation for the panel.

(ii) Spatial (fixed-pattern) noise, $\sigma(\mu_t)$:

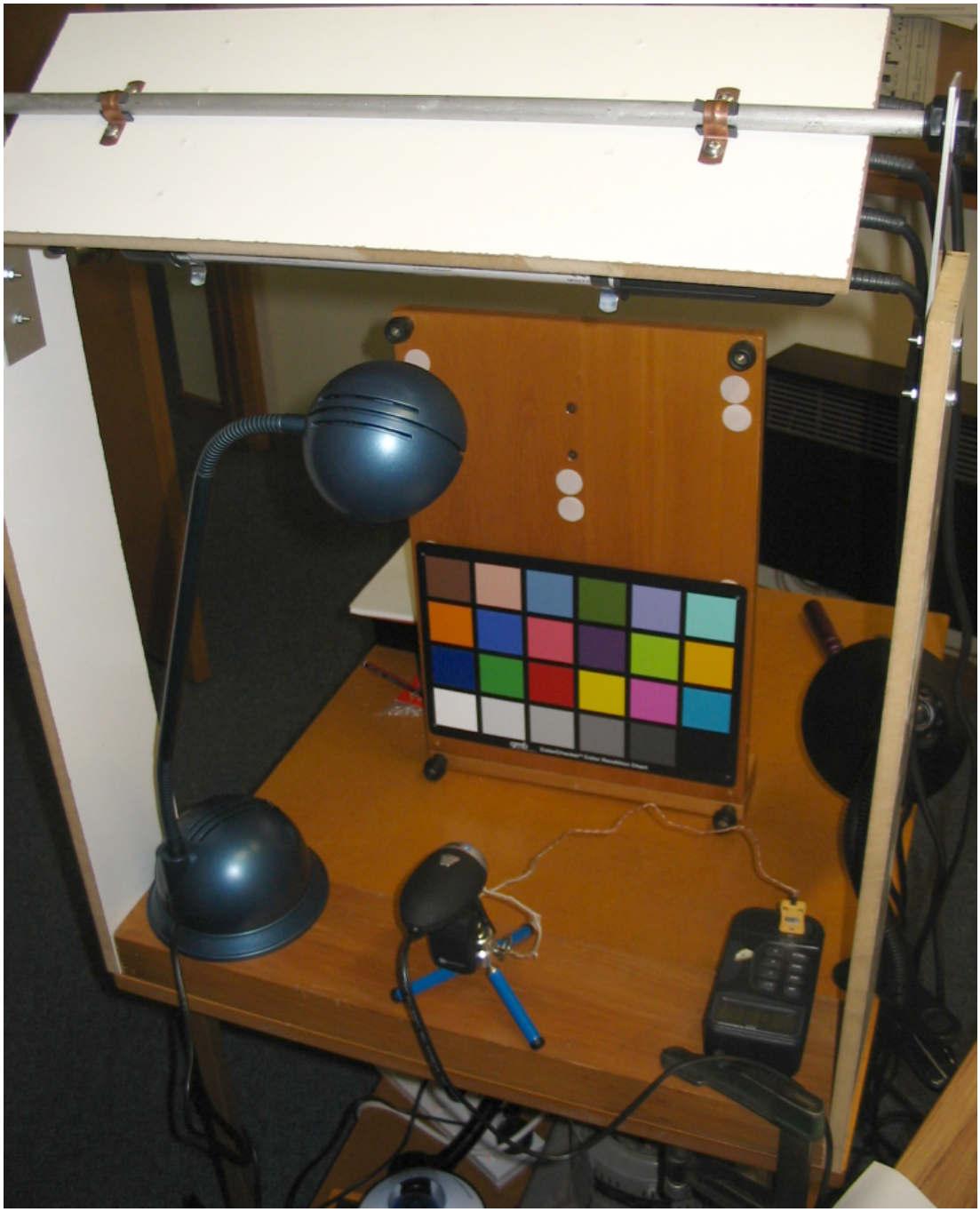


Figure 3.5 The experimental setup for capture of GMB chart images. The fluorescent lights are behind the white panel at the top of the image.

1. The mean of the temporal data, μ_t, for each pixel was calculated over the set of 100 images.

2. A second-order polynomial fit for each column of μ_t was calculated and subtracted from the data to remove optical effects such as vignetting and $1/R^2$ illumination fall-off expected from the use of a discrete illumination source.

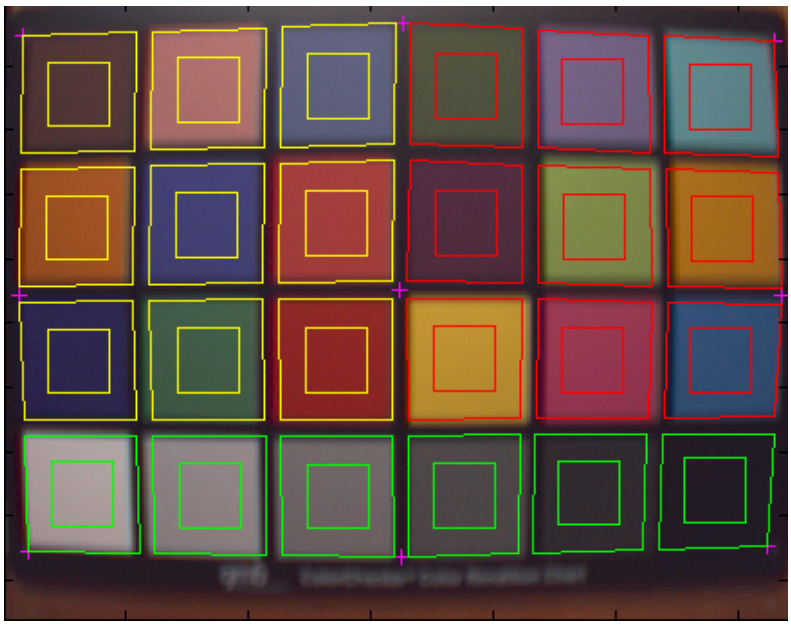


Figure 3.6 An image of a GMB chart with calibration lines overlaid showing the approximate boundaries of each panel (outer squares) and the area used for panel extraction (inner squares). The magenta pluses on the chart are manually selected centre-points for geometric calibration.

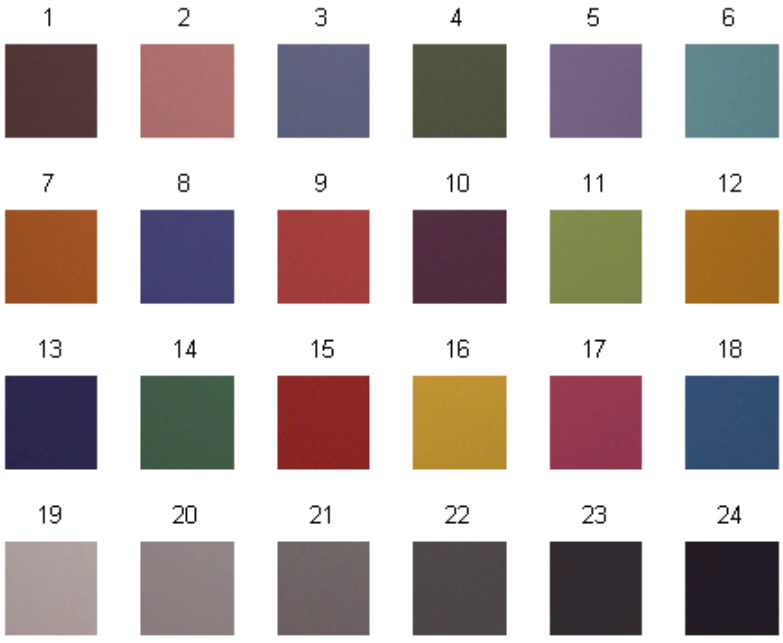


Figure 3.7 The extracted centre panels of the GMB chart shown in Figure 3.6.

3. The residuals after subtraction of the polynomial-fitted data were concatenated and σ calculated to determine a value of $\sigma(\mu_t)$ for the mean spatial variation for the panel.

(iii) Total image noise for each row and column, $\mu(\sigma_r)$, $\mu(\sigma_c)$:

1. A second-order polynomial fit was calculated for each data row and column of an image of a GMB panel, and the fitted line was subtracted from the data to remove optical effects.

2. The residuals after subtraction of each fitted line were concatenated and σ of the concatenated data calculated to give row and column noise values σ_r and σ_c respectively.

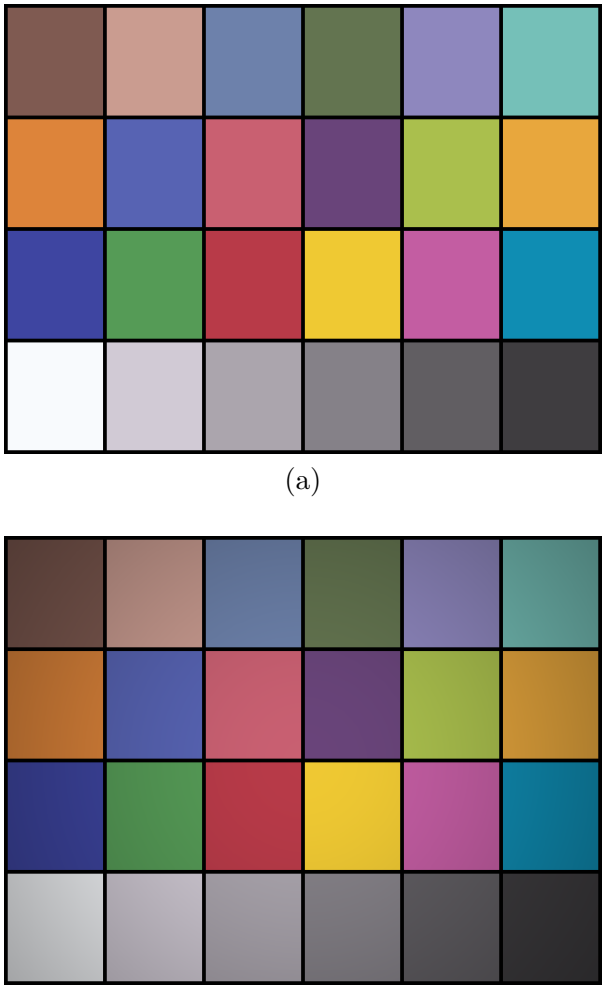
3. σ_r and σ_c for the panel were averaged over 100 test images to derive the final noise figures $\mu(\sigma_r)$ and $\mu(\sigma_c)$ for the panel for each image set.

3.3.3 Quantization noise

From Equation 3.14, any data with significant variation will exhibit quantization noise up to $\sigma = 0.29$. A tool was created using MATLAB that generated images of the GMB chart that included point-source illumination. Images generated using the tool allowed experiments to be conducted to measure the effect of quantization using simulated images both with and without significant variations in the form of lighting gradients across the image. Example images are shown in Figure 3.8.

Figure 3.9 shows the results of quantization error from analysis of simulated images with increasing additive Gaussian noise in dark conditions, and Figure 3.10 shows the resultant error due to quantization from the actual noise value. The methods for measuring spatial and temporal image noise described in Section 3.3.2 were used (100 images generated for each batch in steps of $\sigma = 0.05$). As there is no significant variation in the image data (e.g., no lighting gradient, or visible scene data) there is an initial dip in the error curves caused by the rounding of noise to zero. Both spatial and temporal curve errors then peak at $\sigma_{added} = 0.5$. For $\sigma > 0.5$ the temporal curve approaches zero and the spatial curve continues a downward trend below zero due to the polynomial fitting to noise.

Figures 3.11 & 3.12 shows the same analysis as for Figures 3.9 & 3.10 but with the addition of simulated point-source lighting. This increases the dynamic range of the data by creating a gradient across the measured panel, reducing the significance of the quantization step for low additive noise values. This lighting gradient clearly affects measured noise values as the non-trivial image data now exhibit a minimum spatial noise level of $\sigma = 0.29$, as calculated in Equation 3.14. Temporal noise measurement is



(a)

(b)

Figure 3.8 Synthetic image of the GMB chart for use in analysis of quantization noise with (a) no lighting gradient; (b) point-source lighting.

also affected with continual over-estimation of the real noise level. Both curves exhibit similar responses to those in Figures 3.9 & 3.10 for added noise values $\sigma > 0.5$.

For each experiment utilizing quantization noise, N_Q, the appropriate quantization noise curve illustrated in Figures 3.10 and 3.12 was used to estimate the true noise by subtracting the error from the measured value.

3.3.4 Demosaicing

The effect of demosaicing on image noise was analyzed by measuring noise levels on simulated sets of GMB color chart images. The Bayer array measures the green image on a quincunx grid and the red and blue images on a rectangular grid. Half of the green pixels and three quarters of the red and blue pixels are interpolated. There are many methods for demosaicing the Bayer array (Ramanath, Snyder & Bilbro 2002) with the

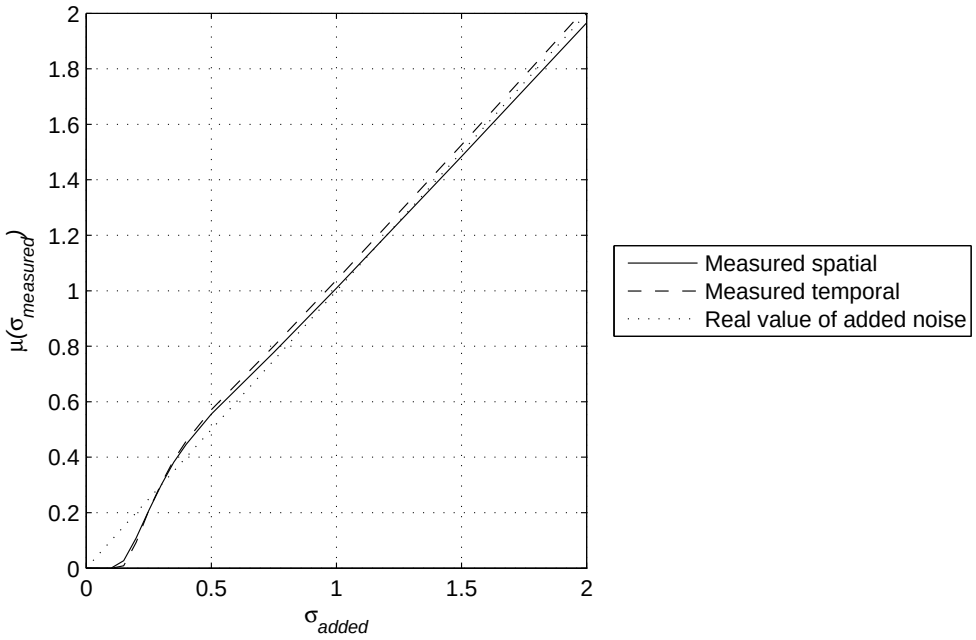


Figure 3.9 Measured spatial and temporal quantization noise for noisy, simulated images without a lighting source.

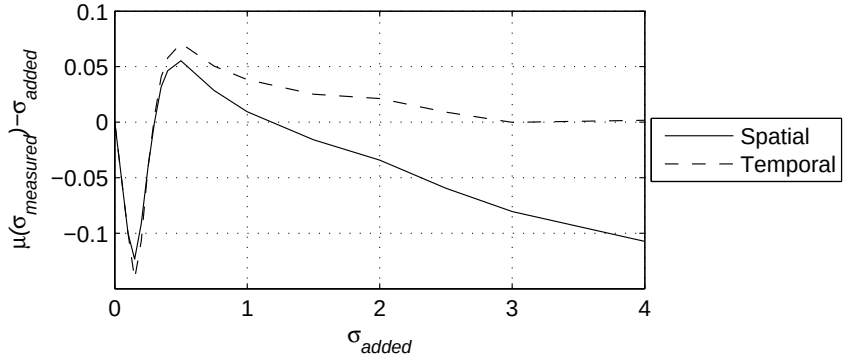


Figure 3.10 Measured error due to quantization for noisy, simulated images without a lighting source.

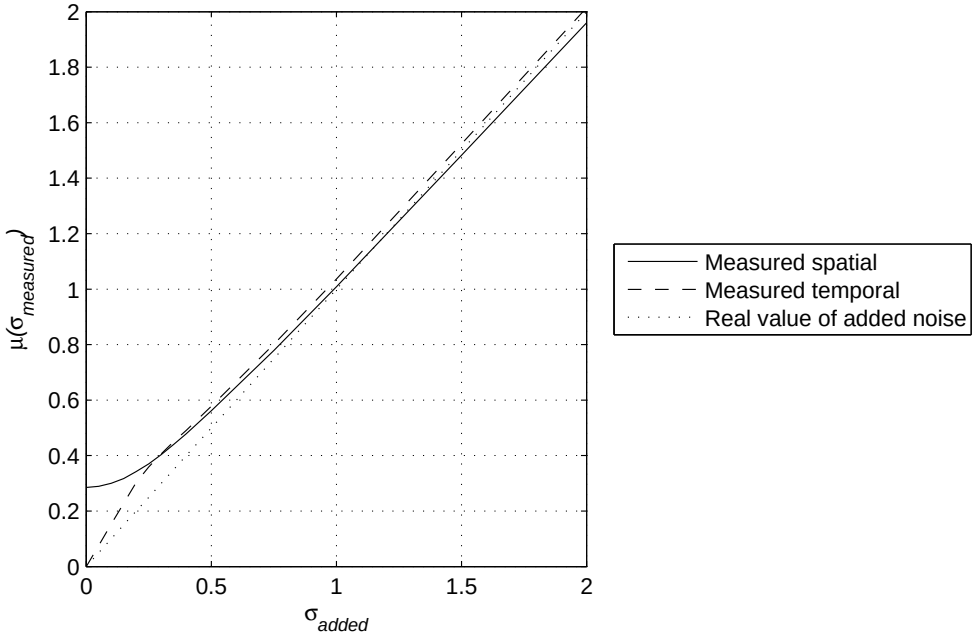


Figure 3.11 Measured spatial and temporal quantization noise for noisy, simulated images with point-source lighting.

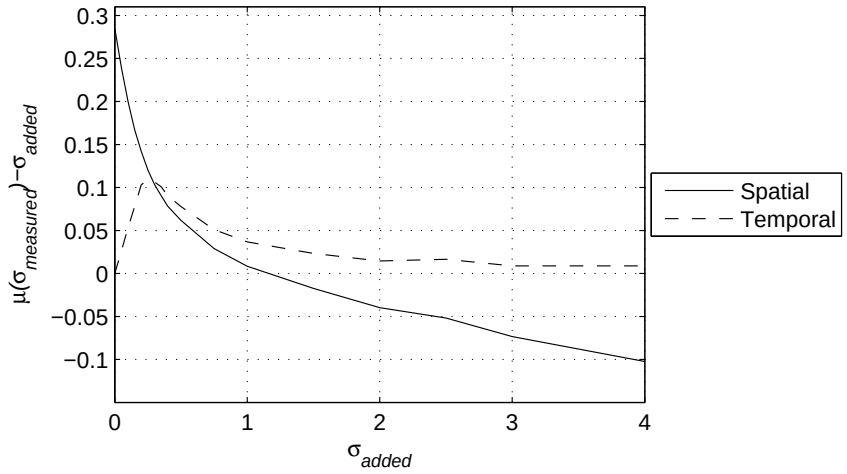


Figure 3.12 Measured error due to quantization for noisy, simulated images with point-source lighting.

bilinear method (BL) being commonly used due to its simplicity. For BL demosaicing the interpolated green pixels are derived from four surrounding measured green pixels. Assuming that individual pixel noise values are independent, the resultant noise in the interpolated green pixel is given by:

$$\sigma_{interp,G} = \sqrt{\left(\frac{\sigma_{G_1}}{4}\right)^2 + \left(\frac{\sigma_{G_2}}{4}\right)^2 + \left(\frac{\sigma_{G_3}}{4}\right)^2 + \left(\frac{\sigma_{G_4}}{4}\right)^2} \tag{3.19}$$

and since σ_G for each of pixels G_1–G_4 is identical:

$$\sigma_{interp,G} = \frac{\sigma_G}{2}. \tag{3.20}$$

The effect of BL demosaicing on the entire green channel noise (50% interpolated) is given by:

$$\sigma_{demosaic,G} = \frac{1}{2}\left(\sigma_G + \frac{\sigma_G}{2}\right) = 0.75\sigma_G. \tag{3.21}$$

In each BL demosaiced image one quarter of red and blue pixels are interpolated from four surrounding measured pixels, and half of the red and blue pixels are interpolated from two surrounding pixels, depending on their location within the Bayer matrix (see Section 3.1.1.6). For the entire red channel:

$$\sigma_{demosaic,R} = \frac{1}{4}\left(\sigma_R + \sqrt{2}\sigma_R + \frac{\sigma_R}{2}\right) = 0.73\sigma_R. \tag{3.22}$$

Similarly, for the entire blue channel:

$$\sigma_{demosaic,B} = 0.73\sigma_B. \tag{3.23}$$

To verify the above, a series of noisy simulated images was generated with added Gaussian noise levels from $\sigma = 0$ to 5, a range empirically determined by observation of noise levels across several cameras. Figure 3.13 shows measured blue-channel spatial-noise characteristics on simulated, noisy, GMB images. The removal of quantization effects result in approximately linear noise response, and measurement of the gradients of the demosaiced noise curves provides the following BL demosaicing values:

$$\begin{aligned} \sigma_{demosaic,R}(spatial) &= 0.71\sigma_R \\ \sigma_{demosaic,G}(spatial) &= 0.77\sigma_G \\ \sigma_{demosaic,B}(spatial) &= 0.71\sigma_B. \end{aligned} \tag{3.24}$$

Figure 3.14 shows the analysis of temporal noise on the blue channel of simulated,

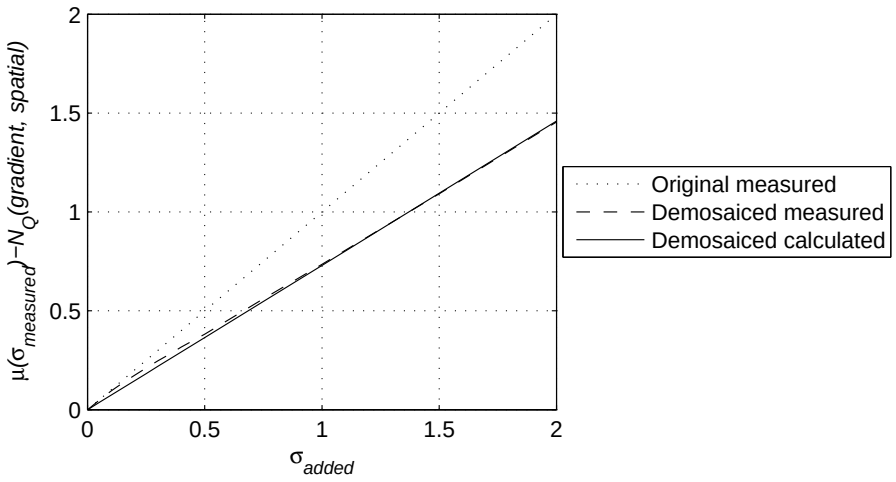


Figure 3.13 Measurement of the effects of bilinear demosaicing on spatial noise analysis on simulated noisy images (blue channel).

noisy, GMB images, and as with the spatial analysis the removal of quantization effects provides an approximately linear noise response. The results for temporal noise analysis give the following:

$$\begin{aligned} \sigma_{demosaic,R}(temporal) &= 0.73\sigma_R \\ \sigma_{demosaic,G}(temporal) &= 0.75\sigma_G \\ \sigma_{demosaic,B}(temporal) &= 0.73\sigma_B. \end{aligned} \tag{3.25}$$

The temporal results match the values given in Equations 3.21–3.23, with measured spatial results varying by no more than 0.02 from the theoretical values.

3.3.5 Column noise

Noise for the i400 was measured by column and by row to observe any variances in column amplifier gain. Figure 3.15 shows virtually identical responses of noise whether measured by column or row, giving $N_C = 1$, in the i400 camera.

3.3.6 Offset FPN

FPN was measured by analyzing a set of images taken in dark conditions with no illumination (n.i.). Equation 3.16 becomes:

$$N_{n.i.} = (FPN + SN_{dark} + N_{read}) \times N_C \times N_D + N_Q(I_{filt}). \tag{3.26}$$

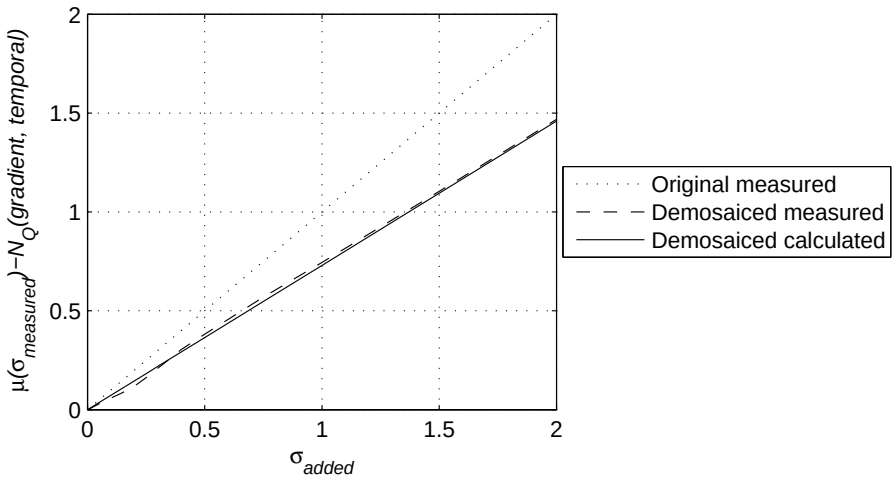


Figure 3.14 Measurement of the effects of bilinear demosaicing on temporal noise analysis on simulated noisy images (blue channel).

The temporal averaging of images effectively removes the terms SN_{dark} and N_{read} giving:

$$N_{n.i.,avg} = FPN \times N_C \times N_D + N_Q(I_{filt}). \tag{3.27}$$

Subtracting $N_Q(I_{filt})$, the quantization noise from rounding of data, derived in Section 3.3.3, from Equation 3.27 and dividing by N_D and N_C (1 for the i400) leaves FPN, shown in Figure 3.16. The measured FPN noise for the i400 camera was taken as being the mean measured value of $\left(\sigma\left(\mu_t\right) - N_Q\left(I_{filt}\right)\right)/N_D$ for each channel:

$$\begin{aligned} FPN_R &= 0.300 \\ FPN_G &= 0.123 \\ FPN_B &= 0.247. \end{aligned} \tag{3.28}$$

3.3.7 Dark current shot noise and readout noise

SN_{dark} is generated within the photodetector and is dependent upon the leakage photocurrent. The value of dark-current shot noise voltage can be sampled at the end of photocurrent integration, where the mean square value of noise voltage is proportional to the integration (exposure) time (Tian et al. 2001).

Equation 3.26 gives the measurable noise component given no illumination. Taking the temporal variations over the set of images removes FPN, giving combined dark-shot

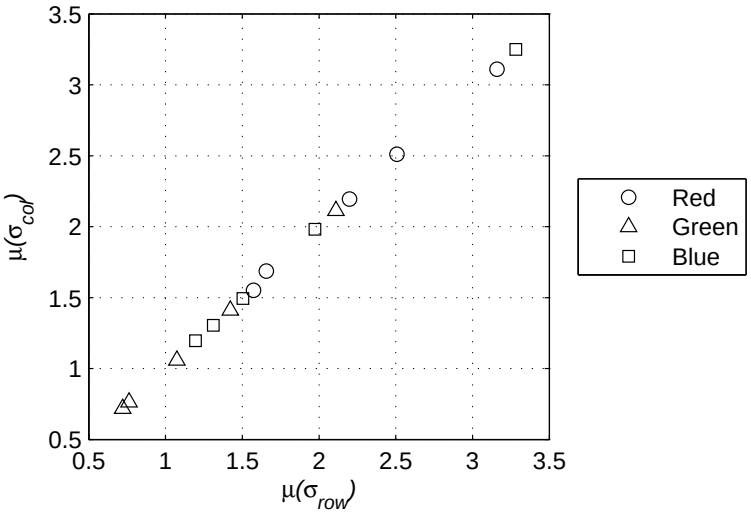


Figure 3.15 Graph of row noise vs column noise for the i400 camera.

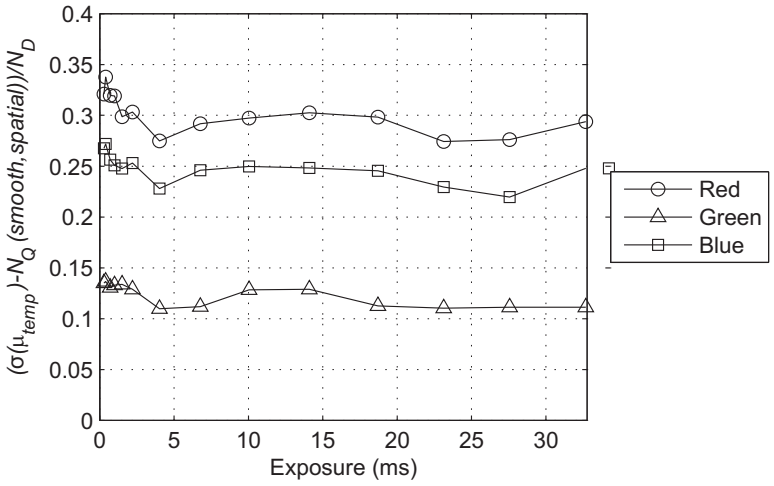


Figure 3.16 Measured fixed-pattern noise for the i400 camera.

and read noise:

$$N_{n.i.,t} = (SN_{dark} + N_{read}) \times N_C \times N_D + N_Q(I_{filt}). \tag{3.29}$$

where $N_{n.i.,t}$ is the temporal noise without illumination. Subtracting $N_Q(I_{filt})$ and dividing by N_D and N_C gives $SN_{dark} + N_{read}$, as shown in Figure 3.17. As SN_{dark} is a function of exposure, a value of N_{read} can be determined by extrapolation to zero of the temporal noise curve. Like FPN, the shape of each $SN_{dark} + N_{read}$ noise curve is approximately flat showing no apparent trend of increasing noise with exposure, suggesting that SN_{dark} is swamped by readout noise, effectively removing SN_{dark} from the noise model for the i400 camera. With $N_C = 1$ for the camera, the measured N_{read} was taken as the mean measured value of $(\mu(\sigma_t) - N_Q(I_{filt}))/N_D$ for each channel:

$$\begin{aligned} N_{read,R} &= 1.61 \\ N_{read,G} &= 0.613 \\ N_{read,B} &= 1.24. \end{aligned} \tag{3.30}$$

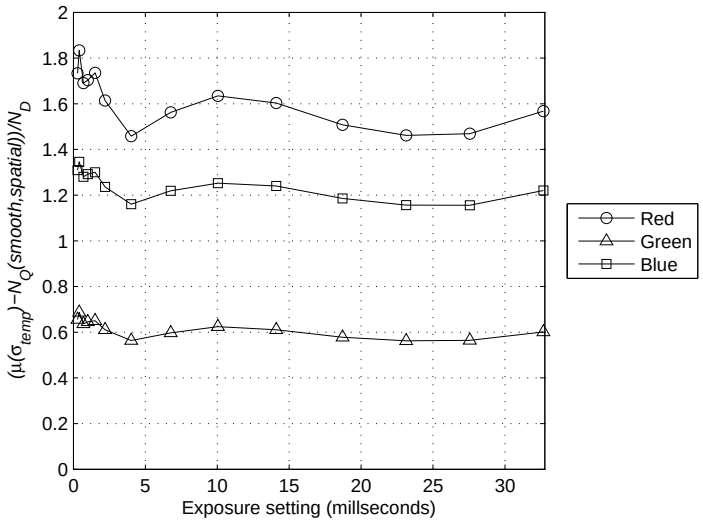


Figure 3.17 Measured dark-current shot-noise and read noise for the i400 camera.

3.3.8 Photon shot noise

For a given illumination, measurement of temporal values only in Equation 3.16 removes I along with spatial noise terms FPN and PRNU giving:

$$N_t = (SN_{ph}(I) + SN_{dark} + N_{read}) \times N_C \times N_D + N_Q(I_{filt}). \tag{3.31}$$

Photon shot noise, SN_{ph}, can be measured by subtracting $N_Q(I_{filt})$, dividing by N_D and N_C, then subtracting the previously measured SN_{dark} (zero for the measured camera) and N_{read}. Figure 3.18 shows photon shot-noise for images captured at maximum exposure (32.68 ms) with different illumination intensities for the i400 camera. A trend can be seen in each of the colour channels which has noise consistent with $\sigma \propto \sqrt{\mu}$, the relationship between noise and sample mean for the Poisson sampling of discrete quanta. A square-root curve was fitted for each colour giving the equations for measured SN_{ph}:

$$\begin{aligned} SN_{ph,R} &= 0.212\sqrt{\mu_{p,R}} + 0.468 \\ SN_{ph,G} &= 0.515\sqrt{\mu_{p,G}} + 0.009 \\ SN_{ph,B} &= 0.215\sqrt{\mu_{p,B}} + 0.040 \end{aligned} \tag{3.32}$$

.

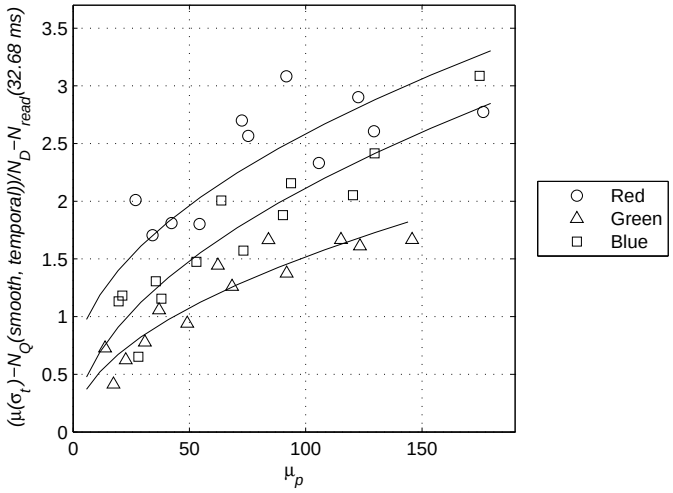


Figure 3.18 Measured photon shot-noise for the i400 camera.

3.3.9 PRNU

Measurement of spatial variations removes I and any temporal terms from Equation 3.16 giving the value for spatial noise without filtering (s,n.f.):

$$N_{s,n.f.} = (I \times PRNU + FPN) \times N_C \times N_D + N_Q(I_{filt}). \tag{3.33}$$

The amplitude of PRNU for a particular irradiance can be calculated by measuring along the columns of pixel data to remove N_C, subtracting $N_Q(I_{filt})$, dividing by N_D,

and subtracting FPN. Figure 3.19 shows the measured PRNU for the i400 camera, which increases approximately linearly with pixel value. PRNU noise for the sensor is defined as the best-fit line for PRNU noise values:

$$\begin{aligned} PRNU_R &= 0.0104\mu_{P,R} + 0.0746 \\ PRNU_G &= 0.0063\mu_{P,G} + 0.122 \\ PRNU_B &= 0.0128\mu_{P,B} + 0.0235. \end{aligned} \tag{3.34}$$

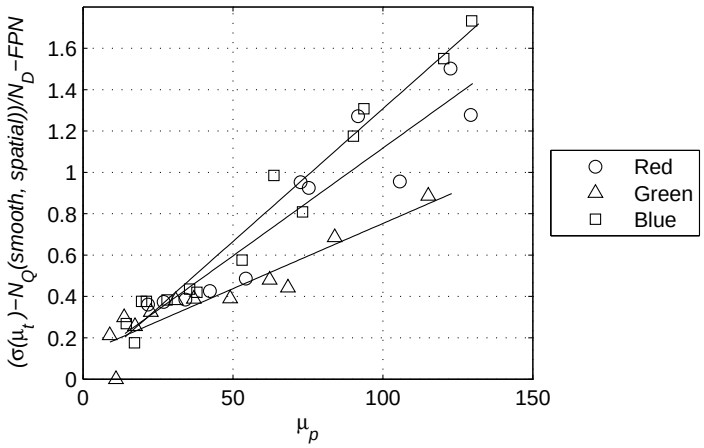


Figure 3.19 Measured photo-response non-uniformity for the i400 camera.

3.4 CCD NOISE MODEL CALIBRATION

The camera noise model in Equation 3.16 was calibrated from the measured data described in Section 3.3, where sensor irradiance becomes a function of measured pixel value. Figure 3.20 shows measured total image noise, and noise curves generated from the calibrated noise model. The values of total image noise $\mu(\sigma_r)$ and $\mu(\sigma_c)$ are highly correlated, hence only the row analysis is shown. The measured noise data supports the validity of the model in Equation 3.16. Maximum values for μ_p are restricted to below 150 as this was the maximum value the i400 camera would output on the green channel when all digital effects were disabled or set to neutral.

Figure 3.21 shows the relative magnitudes of the individually measured noise components on the blue channel for the i400 camera. N_Q represents the maximum potential contribution of quantization noise to overall noise and is dependent only upon final image content. Bilinear-interpolative demosaicing results in a scaling factor and is not included in the graph, and SN_{dark} is insignificant and is not shown as a contributor

to noise in the i400 camera. Image noise for this camera is clearly dominated by N_{read}, $PRNU$, and SN_{ph}.

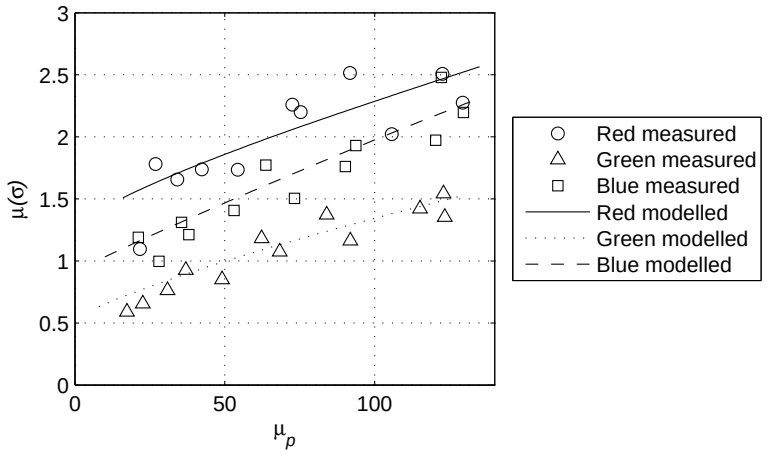


Figure 3.20 Measured i400 camera noise and the predicted noise curve from the calibrated noise model.

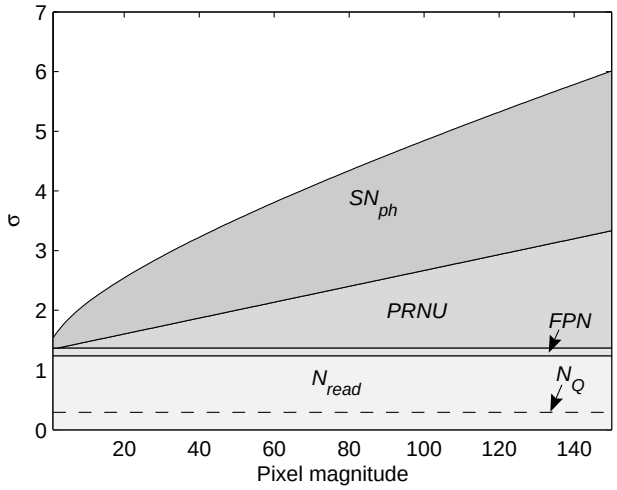


Figure 3.21 The relative magnitudes of noise components for the blue channel of the i400 CCD camera at 22°C environmental temperature. N_Q denotes the maximum potential quantization noise value, and SN_{dark} is zero.

3.5 CMOS NOISE MODEL

The design of a CMOS image sensor, and in particular the layout and number of transistor elements surrounding each pixel, are manufacturer dependent and generally unknown. However, a generalised noise model can be developed that extends the CCD model by incorporating the expected additional CMOS noise sources. In effect, each

transistor around a CMOS pixel will have its own characteristics in terms of an offset and gain. The extra noise sources of active gain FPN ($AFPN_{gain}$), active offset FPN ($AFPN_{off}$), column gain FPN ($CFPN_{gain}$), and column offset FPN ($CFPN_{off}$) can therefore be defined. These sources can be incorporated into the CCD noise model after N_{read}, as shown in Figure 3.22.

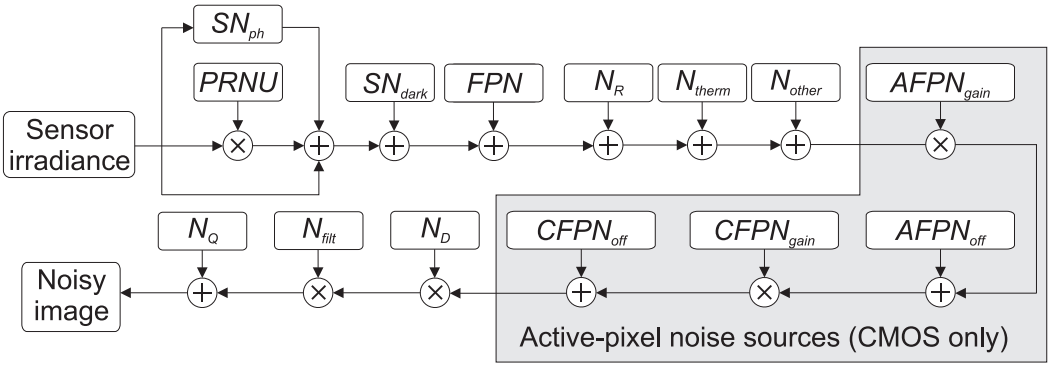


Figure 3.22 The noise model for image capture in a standard CMOS digital-video camera.

From Figure 3.22, the full equation for a noisy CMOS image is therefore:

$$\begin{aligned} I_{cap,CMOS} = (((&I_{CFA} + I_{FCA} \times PRNU + SN_{ph}(I_{CFA}) + FPN + SN_{dark} + N_{read}) \\ &\times AFPN_{gain} + AFPN_{off}) \times CFPN_{gain} + CFPN_{off}) \\ &\times N_D \times N_{filt} + N_Q(I_{filt}). \end{aligned} \tag{3.35}$$

Due to the added complexity of the CMOS model when compared to the CCD model, it is difficult to differentiate between the individual noise components from analysis of output images alone. Hence the general CMOS noise model was not experimentally validated.

3.6 MEASUREMENT OF COMBINED NOISE

The analysis of modelled image noise becomes more difficult with increasing camera complexity, so a method of noise analysis was developed that does not require a specific model and is dependent only upon output images from a camera.

3.6.1 Measurement of image noise

The sources of noise can be segmented into three general categories for the purpose of noise measurement: spatial noise, temporal noise, and total image noise (see Tables 3.1–

3.3 for the list of noise types). Measurement of spatial noise is calculated along the rows of an image to ensure that any column noise is included.

3.6.1.1 Spatial noise

Spatial noise is exhibited as variations between the pixels in an image given a constant illumination across the sensor. Knowledge of spatial noise is useful in applications where images are averaged prior to analysis (e.g., low-light applications where multiple images may be captured to increase effective exposure time). Images of spatial noise can be achieved by averaging a series of images containing smooth areas of constant reflectance (e.g., the panels of a colour chart) and removing the temporal variations. The generation of an image of spatial noise, $\overline{image(i,j)}$, is given by:

$$\overline{image(i,j)} = \frac{\sum_{k=1}^{n} P_k(i,j)}{n}, \tag{3.36}$$

where n is the number of images and $P_k(i,j)$ is the pixel value for row i, column j in the k^{th} image. A second-order polynomial can be fitted to each row of $\overline{image(i,j)}$ to describe any optical effects such as shadowing or vignetting. The residuals after subtraction of the polynomial-fitted data can be concatenated for each row and the standard deviation (σ) calculated, giving the value of the spatial noise $\sigma_{spatial}$.

3.6.1.2 Temporal noise

Temporal noise varies between images and is dependent on illumination. Knowledge of temporal noise is useful in applications where images are compared or subtracted before analysis (e.g., in a subtraction-based motion detection algorithm). Temporal noise can be measured by taking the average value of the variations exhibited by a pixel over a series of images of the same scene. The equation for temporal noise σ_{temp} for a given sensor irradiance is therefore:

$$\sigma_{temp} = \frac{\sum_{i=1}^{x}\sum_{j=1}^{y}\sigma(i,j)}{x \times y} \tag{3.37}$$

where x and y are the number of rows and columns respectively and $\sigma(i,j)$ is the standard deviation of the pixel value at (i,j) over n images.

3.6.1.3 Combined noise

The combined noise in an image is the combined effect of all spatial and temporal noise present in an image. A measure of the combined noise is useful in applications where a single image is used for processing (e.g., in an edge detection algorithm). To measure

total image noise, the process of fitting a second-order polynomial to each of the n images is applied in the same fashion as the calculation of spatial image noise $\sigma_{spatial}$ in $\overline{image(i,j)}$. A series of standard deviations for each row of the k^{th} image of the n image set is calculated, giving σ_k. The average of σ_k over all n images gives σ_{total}.

3.6.2 Experimental results for a CCD camera

Figure 3.23 illustrates the spatial, temporal, and combined noise measurements for the i400 CCD camera (specifications in Table 3.5). Experiments found that the green channel of the CCD camera was limited to values below approximately 150 when all digital effects were disabled, hence the illumination was set to ensure the green channel measurements were below this value.

3.6.3 Experimental results for a CMOS camera

Figure 3.24 illustrates the spatial, temporal, and combined noise measurement for the uEye CMOS camera (Table 3.6), with each of the three colour channels showing similar noise responses.

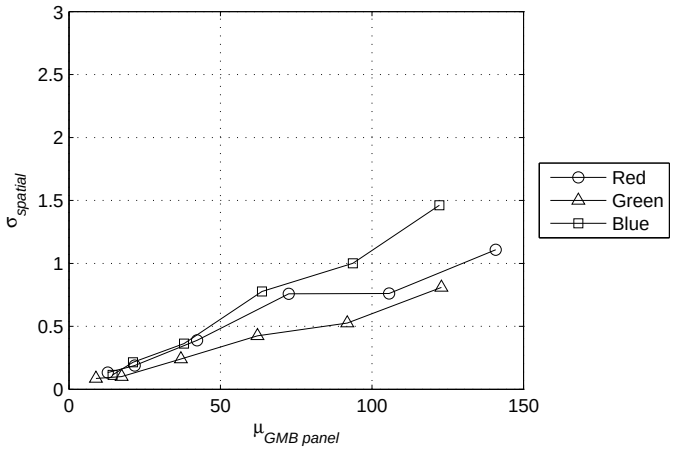
Table 3.6 uEYE UI1210-C camera details.

Parameter	**Value**
Sensor type	1/2" Zoll colour CMOS (Bayer colour filter)
Native resolution	CCD 640 × 480
Video mode	24-bit RGB (8–bits per colour channel)
Interface	USB 2.0
Operating temperature	0°C to 50°C

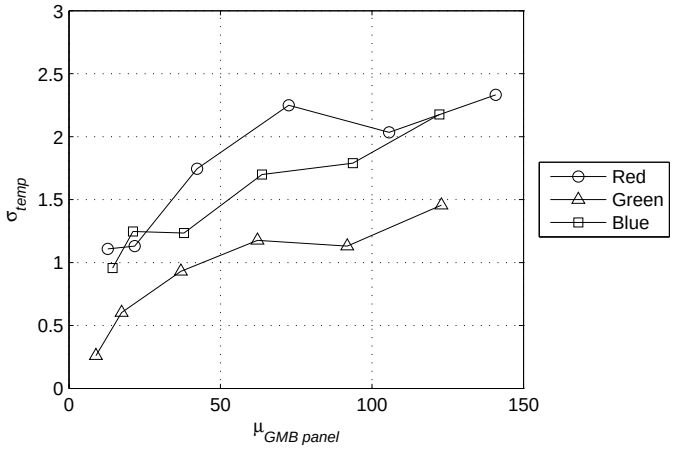
3.7 THE EFFECT OF TEMPERATURE

One of the major contributors to image noise is heat within the sensor, where variations in thermally generated currents results in a fluctuation of measured pixel values. To help better understand the effect that temperature has on image noise, the noise characteristics of 4 different industrial cameras were measured from 21°C to 55°C.

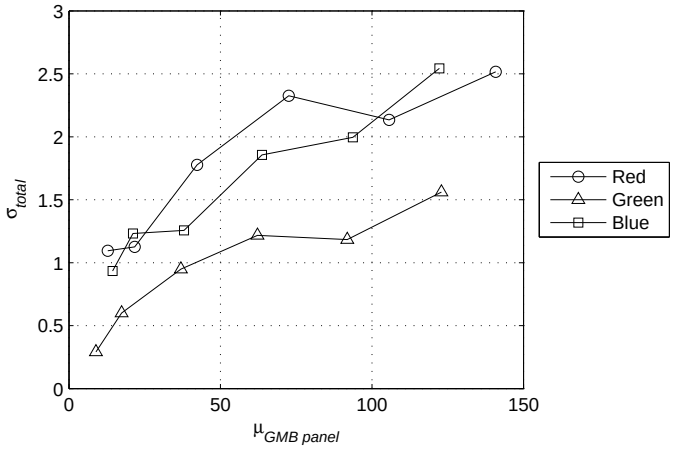
The effect of temperature can be significant, with some noise sources (such as dark current shot noise) doubling with every 8°C rise in temperature (Nakamura 2006). If image noise is to be appropriately managed within an image-processing application, then the effect of temperature on spatial and/or temporal camera noise must be clearly understood.



(a)



(b)



(c)

Figure 3.23 i400 CCD camera combined noise: (a) spatial, (b) temporal, (c) combined.

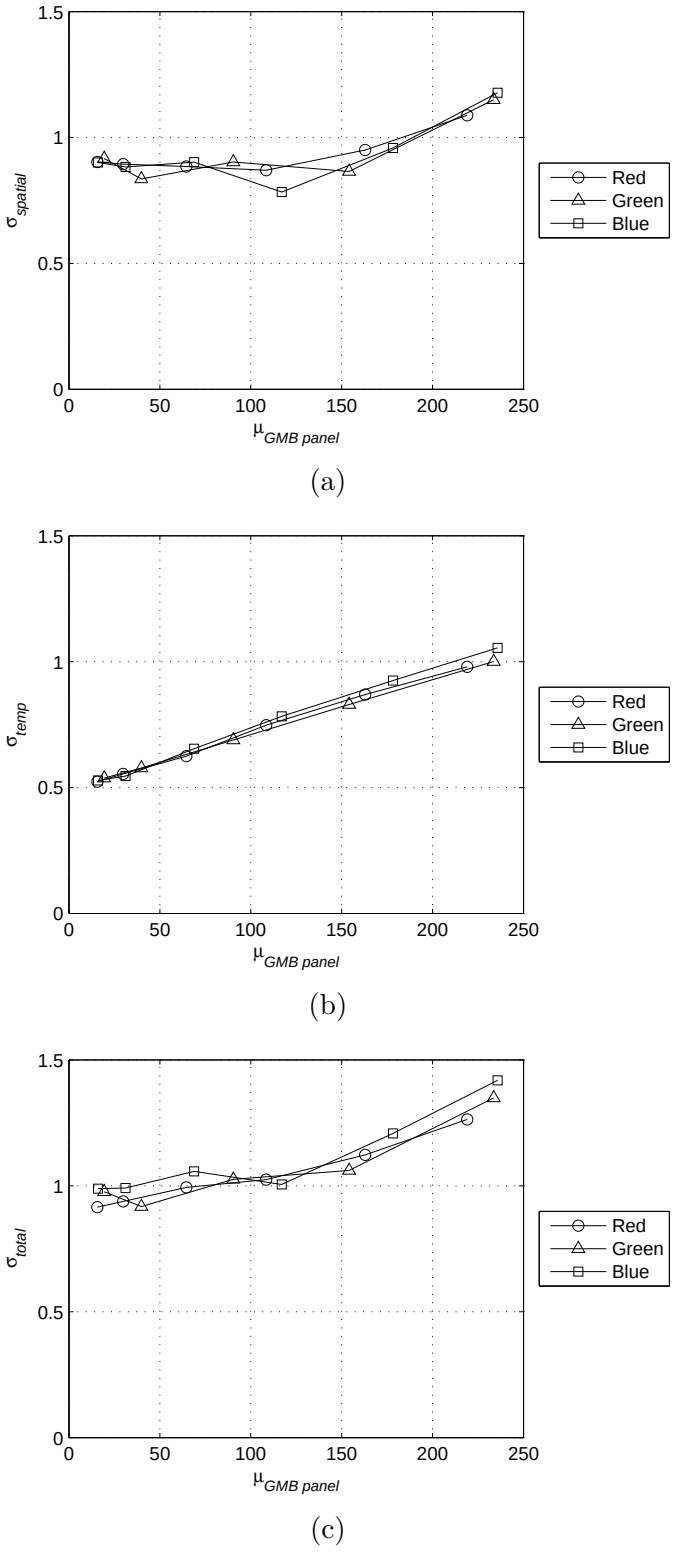


Figure 3.24 uEye CMOS camera combined noise: (a) spatial, (b) temporal, (c) combined.

3.7.1 Experimental setup

The total spatial and temporal noise characteristics for 4 different industrial cameras (specifications given in Tables 3.5, 3.6, 3.7 & 3.8) were measured using the method described in Section 3.6.1, where 100 images were captured across a range of environmental temperatures for each camera. The resulting analysis returns the standard deviation of temporal noise σ_{temp} and spatial noise $\sigma_{spatial}$. The 4 cameras were placed in a sealed, temperature-controlled thermal chamber, as shown in Figure 3.25. Each camera lens aperture was covered to prevent any light entering the sensor. The cameras were powered on and the temperature of the chamber was allowed 10 minutes to settle between experiments. Temperature measurements using thermistors placed inside the cameras demonstrated that this was ample time for the image sensors to reach thermal equilibrium. Multiple computers then captured a series of images at approximately the same time for each of the four cameras from 21.1°C to 55.2°C in steps of approximately $1-2$°C.

Table 3.7 AVT Guppy F044C camera details.

Parameter	**Value**
Sensor type	Sony ICX419AKL 1/2" colour CCD, (CMYG colour filter)
Native resolution	752 × 580
Video mode	RAW 8-bits (binned pairs)
Interface	IEEE-1394a (Firewire)
Operating temperature	5°C to 45°C

Table 3.8 AVT Guppy F080C camera details.

Parameter	**Value**
Sensor type	Sony ICX204AK 1/3" colour CCD (Bayer colour filter)
Native resolution	1032 × 778
Video mode	RAW 8-bits
Interface	IEEE-1394a (Firewire)
Operating temperature	5°C to 45°C

All cameras had filtering disabled (e.g., gamma turned off, neutral colour balance) with all options set to manual mode. Exposure time was set to 33 ± 1 ms for all cameras, and all images were captured in an uncompressed format.

3.7.2 Results

Patches of 500 × 150 pixels that were identified free from 'hot pixels' were extracted from the captured images for subsequent noise analysis. Values of σ_{temp} and $\sigma_{spatial}$

Figure 3.25 The temperature-controlled thermal chamber with the 4 cameras under test.

were calculated, and the results are shown in Figures 3.26–3.29. Each camera was pushed beyond its specified operational temperature range, which is indicated by the dashed lines in the figures. The spatial noise results demonstrate reasonable consistency with an expected increase in noise with temperature for the F044C (CCD), F080C (CCD) and 1210-C (CMOS) cameras, at least up to their maximum rated operating temperatures. The CMOS camera results exhibit a classical exponential increase, up to and beyond the maximum operating temperature, although its overall spatial noise is up to an order of magnitude higher than that of the CCD cameras. The i400's spatial noise response demonstrates an unexplained bump between 37°C and 50°C.

Temporal noise results demonstrate a higher consistency of values across the cameras, though more unexplained trends are present. The F044C's temporal noise decreases with temperature, while the F080C's noise increases semi-exponentially up to the maximum rated temperature before exhibiting substantial amounts of oscillation. The 1210-C has a slight exponential trend with increasing temperature while the i400 demonstrates very low temporal noise for the red channel, with very high relative temporal noise for the blue channel. Again, from 37°C and upwards the i400

demonstrates an unexpected bump, followed by a sudden jump when operating above its recommended operating temperature.

3.7.3 Discussion

Each camera measured has a unique noise response to temperature, with only the uEye 1210-C CMOS camera (Figure 3.29) demonstrating what could be described as a 'classical' increase in noise with temperature. It is likely that the electronic circuitry required for charge transfer and readout of CCD image sensors contributes significantly to the noise characteristics displayed by them. The F080C and i400 CCD cameras demonstrated unexpected behaviour when operated in an environment beyond their rated temperature. Further, all CCD cameras exhibited lower temporal noise responses at 51°C than at 41°C. The i400 shows a significant jump in both thermal and spatial noise when in a 55°C environment – just 5°C higher than its rated maximum operating temperature.

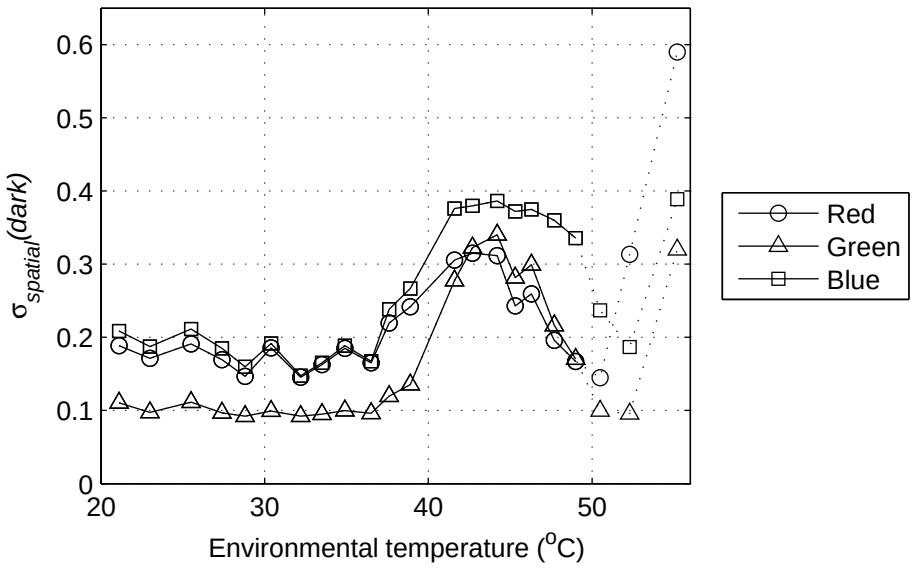
These experiments highlight that noise can vary significantly between different models of camera, with some cameras demonstrating unpredictable noise characteristics when running beyond their maximum rated temperature.

3.8 SUMMARY

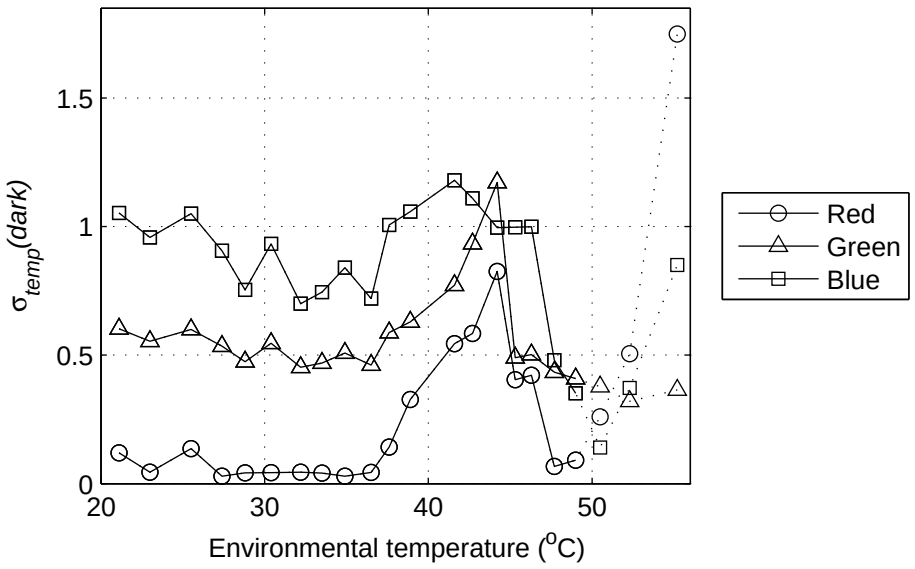
This chapter has provided a comprehensive study of camera noise in several standard industrial digital cameras. Noise sources and existing noise models have been reviewed, and new noise models for CCD and CMOS cameras derived. Methods to analyse the individual components of CCD noise from output images alone were developed, along with general methods of measuring spatial and temporal noise in images. The effects of temperature on image noise were measured using four different cameras.

A technique for measuring noise in a CCD camera has been developed that includes noise sources from the CCD and supporting ICs, colour processing, and quantization. The noise components were grouped into measurable quantities and measured on a commercially available CCD camera. The derived CCD noise model was then calibrated with the measured data, generating a noise response which was dependent only upon pixel value. The modelled response compares favorably with the measured total image noise. The CMOS model was not validated experimentally due to the complexity of the model, which makes it difficult to 'unwrap' each noise layer by analysis of output images alone.

The camera temperature has a significant effect on the quantity and quality of base-line noise in a camera. Further, significant variances in the noise characteristics were found when four different cameras were analysed.



(a)



(b)

Figure 3.26 The effect of temperature on i400 CCD camera noise: (a) spatial, (b) temporal.

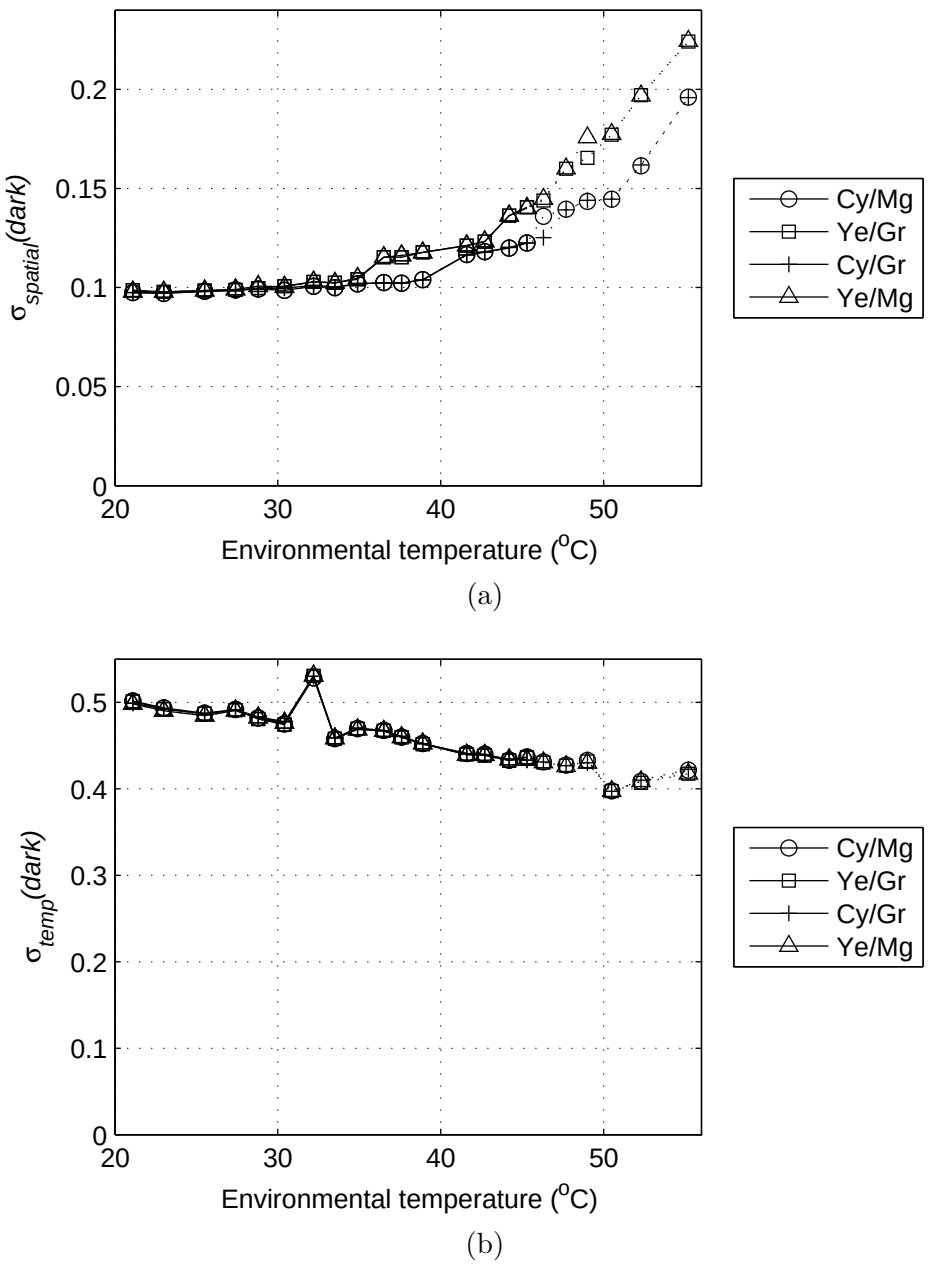
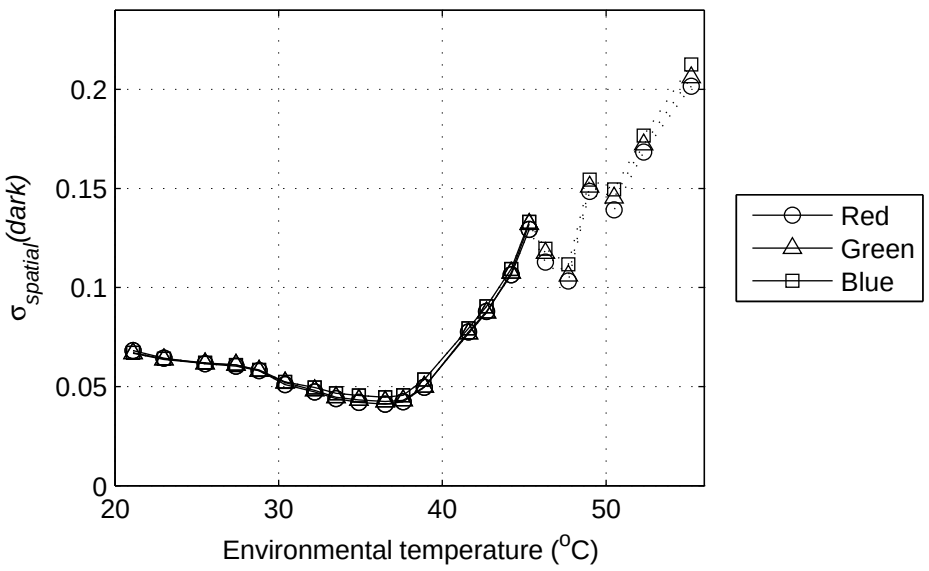
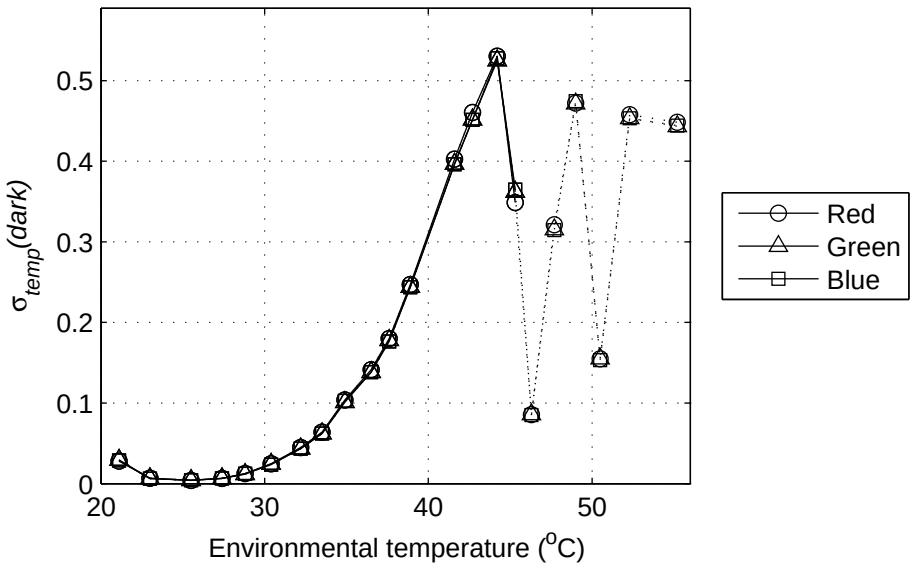


Figure 3.27 The effect of temperature on F044C CCD camera noise: (a) spatial, (b) temporal.

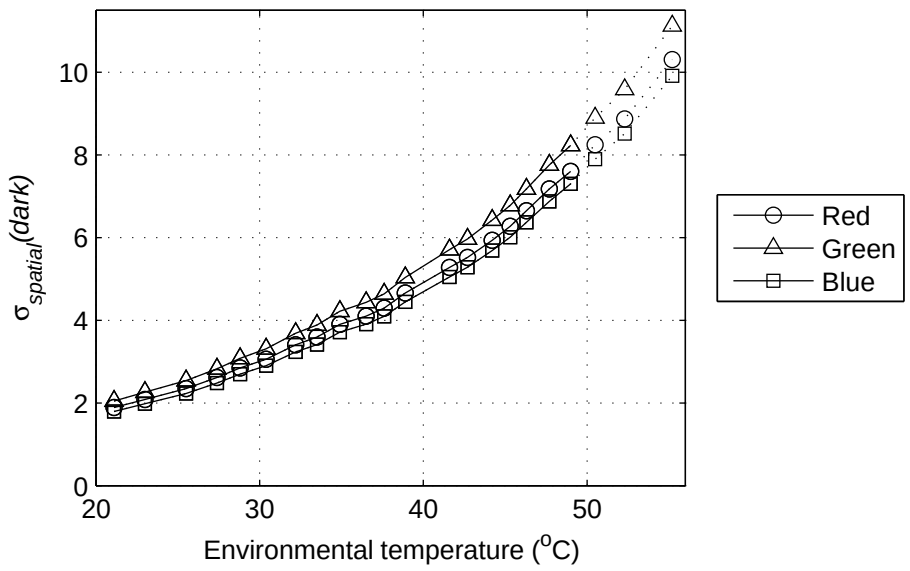


(a)

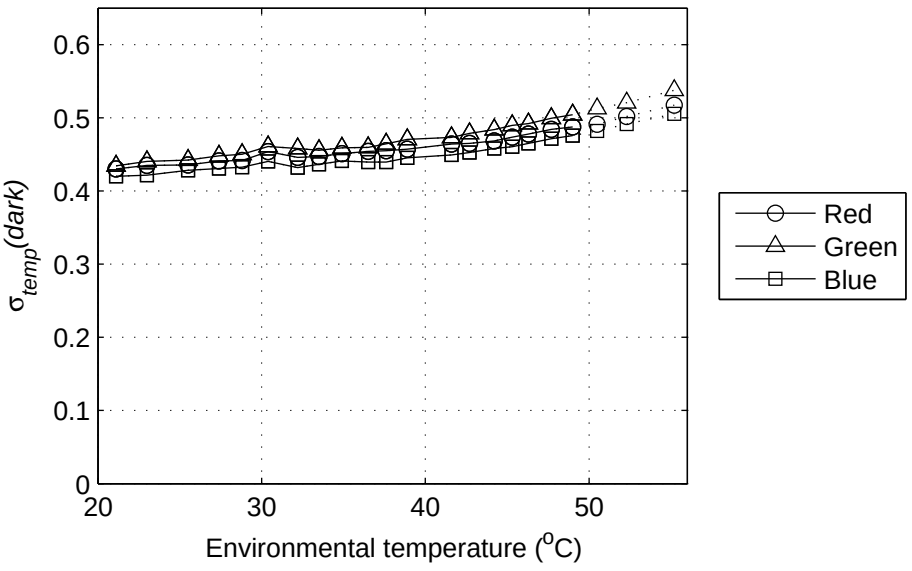


(b)

Figure 3.28 The effect of temperature on F080C CCD camera noise: (a) spatial, (b) temporal.



(a)



(b)

Figure 3.29 The effect of temperature on 1210C CMOS camera noise: (a) spatial, (b) temporal.

The noise information gathered by the methods developed in this chapter can be used in image processing applications to tune an algorithm to a particular image or images from a digital video camera. For example an edge-detection algorithm could benefit from knowledge of the spatial noise content, or an object tracking function could utilize knowledge of the temporal noise to improve its reliability. A natural application for this technique is in the design of noise removal filters for images and video, where *a priori* knowledge of noise content could increase accuracy and robustness. It is important to ensure that any noise analysis that is used for the purpose of an image processing application occurs at the camera temperature that the camera will be operating in. The results of this work have shown considerable differences in noise characteristics between the CCD and CMOS cameras analysed.

Part III

Shadow-free scene-change detection

Scene-change detection is commonly used to detect moving objects in image sequences. However, the process of detecting scene change suffers from the effect of cast shadows on the background. In all natural environments and almost every artificially illuminated environment there will be shadows cast by any non-trivial object, changing the quality and quantity of light being integrated and measured by a camera. The shadows will darken the apparent surface of other objects within the scene to various levels, depending upon the strength and direction of the illumination sources. These reductions in surface intensities can appear as objects to a scene-change detection algorithm, causing errors in the object detection and identification process.

The use of user-defined thresholds in the scene-change detection process allows the segmentation between shadows and objects to be empirically set to tune an algorithm's performance. However, the values of these thresholds usually relate to an esoteric internal parameter that have no direct relation to anything meaningful, and are dependent upon what objects are present in the scene and how important the false-positive or true-negative detection of objects is.

The goal for Part III of the thesis is to research a robust method for shadow-free scene-change detection that is ideally free from any arbitrary or empirical thresholds, and can potentially lend itself to automated machine-vision applications. Chapter 4 defines a new method of scene-change detection that is based on colour-change only, whilst Chapter 5 integrates the fundamental concept of image noise into the new method, allowing for the development of a novel algorithm for scene-change detection.

Chapter 4

COLOUR-BASED SCENE-CHANGE DETECTION

Automated scene surveillance and monitoring is an important tool for both data analysis and security applications. Camera-based solutions are attractive due to their relative low cost and maintenance, and potential for minimal operator interaction during operation. However, robust image analysis techniques are difficult to develop as not only is the image-capture process imperfect but object segmentation of real-world images is a challenging problem. Many camera-based applications attempt to interpret pixel information so that decisions can be made from the images. However, many important tasks such as object identification and tracking suffer from the detrimental effects of shadowing. Shadow and scene-change detection can play a vital role in helping these

tasks perform robustly and accurately, by identifying or removing the shadowed areas of an image.

In images, shadows appear to extend the shape of objects or merge multiple objects together, and many shadow detection or removal algorithms use edges to classify scene changes as either objects or shadows. The draw-back with these methods is that they rely on distinct boundary conditions, which limits the type of shadow they are trying to detect: methods that look for shadow gradients on soft shadows will be unable to locate the sharp edges of hard shadows, and vice versa. Further, arbitrary or empirical thresholds are almost invariably required to ensure a good level of segmentation performance. This can severely limit the applicability of an algorithm, as a threshold set for one scenario may not be suitable for another without experienced user-intervention.

This chapter describes the development of a method for scene-change detection that can operate in both indoor and outdoor environments by classifying pixel change as either due to genuine scene change or as a new shadow cast by an object onto the scene.

4.1 OBJECTIVES

The objectives for Part III are to investigate techniques for scene-change detection with the intention of deriving a more robust method that:

- is independent of object shape;
- is independent of additional shadows cast upon a scene;
- operates in both indoor and outdoor environments;
- does not require manual interaction;
- is free from arbitrary internal thresholds;
- does not require significant training.

4.2 LITERATURE REVIEW

Object tracking and identification in outdoor scenes can be difficult when cast shadows from an object fall on to the surrounding terrain. The shadows can grossly extend the apparent shape of the object due to the reduction in illumination intensity, and objects in close proximity can appear joined. A large proportion of the literature on scene-change detection is focused upon the detection or removal of shadows, highlighting the significance of the problem. Hence this literature review covers both the areas of scene-change and shadow detection. It applies to both Chapters 4 & 5 of this thesis.

Scene-change, by definition, requires a reference image which a changed image is compared to. Some of the algorithms discussed in the literature review are not specific to scene-change, and may only use a single image for shadow detection or removal. The outputs of these algorithms could conceivably be used as input to a scene-change detection method, and in the review no distinction is made between the methods that use reference images or not.

Recent state-of-the-art advancements in shadow detection have been made by a core group of researchers that includes Graham Finlayson, Steven Hordley, and Mark Drew. The scene-change detection algorithm developed in Part III of the thesis uses an extension of their work as a basis for detection of shadows in image sequences. This literature review starts with a discussion of prior work in shadow detection for scene-change analysis, followed by a more in-depth review of the works by the Finlayson, Hordley and Drew group.

4.2.1 Beginnings

Early research on shadows in images was centred around adding shadows to computer-generated scenes to add realism. One of the earliest literature reviews of these algorithms is by Crow in 1977 (Crow 1977), with another by Woo et al. in 1990 (Woo, Poulin & Fournier 1990). Most algorithms reviewed dealt with generation of cast shadows, i.e., shadows with sharp boundaries that are caused by objects partially or completely occluding the illuminating light sources. Woo et al. also describe algorithms for creating shadows from other objects such as glass. Elementary image analysis for shadows started appearing around the 1980s, where the target was removing shadows from aerial photographs.

4.2.2 Intrinsic methods of shadow removal

Intrinsic methods of shadow removal attempt to develop an underlying 'intrinsic scene' (usually by using a long series of images for training), which by its nature is shadow-free. The performance of these methods typically improves with increasing exposure to different illumination conditions.

Weiss' paper, entitled 'Deriving intrinsic images from image sequences' (Weiss 2001) attempts to recover intrinsic images from a sequence of images by using edge filters to locate detail in the image, which are averaged over a period of time where there is substantial shadow change (e.g., over the course of a day). The edge-image details from temporally changing objects such as shadows and moving objects are reduced in intensity by the averaging function, whereas constant edges, such as at the boundaries of fixed objects in the scene, are retained. Weiss' method for scene-change detection is limited in its application as a long series of images is required to derive the

intrinsic scene (in outdoor conditions), and it does not take into account objects which may be stationary for intermediate periods of time. Also, by definition, intrinsic scenes do not differentiate possible scene-change as being due to either objects or shadows.

Matsushita et al. (Matsushita, Nishino, Ikeuchi & Sakauchi 2004) uses Weiss' method as an input to an algorithm for a video surveillance application for tracking vehicles in a city. Their method is specifically looking to remove the effects of strong cast shadows from the likes of large buildings, which they calibrate prior to processing for vehicle detection. Their method inherits the restrictions of Weiss' method, and also requires *a-priori* information on cast-shadows that are to be removed. Hence cast-shadows from temporal objects (such as long shadows thrown by vehicles themselves) are not handled.

Porikli and Thornton (Porikli & Thornton 2005) use background subtraction and a video stream of images to recursively update a shadow model and intrinsic scene which, for the most part is parameter-free, but requires a substantial number of images over the course of a day. Their system does not address dynamically changing illumination within the video stream, and assumes that shadows are simply a reduction in intensity, and that the hue of a shadow is unchanged from the underlying scene.

Methods of intrinsic scene recovery by Eiseman and Durand (Eisemann & Durand 2004) and Lu et al. (Lu, Drew & Finlayson 2006) use a pair of flash/no-flash images to provide distinct and controlled illumination conditions from which to extract the underlying scene content. By subtracting the images a 'flash only' or 'ambient shadow' image can be obtained. In the majority of both outdoor and indoor environments it is not pragmatic to use flash-lighting for continual object tracking and scene-change detection.

4.2.3 Methods based on illumination colour-change

Many recent methods have used illumination colour change to assist in the detection of shadowed areas, particularly for outdoor environments where shadows are blue-coloured due to skylight illumination.

Withagen, Groen and Shutte (Withagen, Groen & Schutte 2007) provide an interesting method for shadow detection that uses band-ratios (e.g., values of red/green and blue/green) to move pixels into a plane for segmentation of shadowed regions of an image. They make a significant assumption in that the colour response of the camera is flat across each colour band. In practice, colour cameras have non-linear spectral responses that cause the band-ratio plane to warp into a curved 3D surface. This in turn causes the accuracy of the system to depend upon the camera being used. Their method also uses various empirical thresholds to segment shadow regions, potentially limiting a set of hand-tuned thresholds to a particular set of images.

Marchant and Onyango (Marchant & Onyango 2000) use a CIE black-body approximation of daylight to create a grey-scale image from a colour camera that is effectively independent of the illumination intensity. This is done by determining which black-body curve a pixel falls on in chromaticity space, and assigning a value to that pixel based upon the number of the black-body curve. As the reflectance of each object in the scene is constant whether it is illuminated by daylight or skylight, the theory is that the pixel will always lie on that one specific curve. Their method assumes that the camera's spectral response is narrow, and that skylight can be modelled by a black-body radiator. In reality, the majority of cameras are wide-band, and skylight is not always well approximated by a black-body radiator (Hernandez-Andres, Lee & Romero 1999). Their research is applied to distinguishing the difference between plants and soil, and their results show that a distinction can be made between a particular colour of plant and soil when shadows were present.

Nadimi and Bhanu (Nadimi & Bhanu 2004) propose a physical model for moving-shadow and object detection in video. They essentially define all moving objects in an image, and test whether the change has a blue tint to it. It requires an initial collection of frames to calculate mean pixel values, and then uses manually defined areas of shadow for which to determine the 'blue tint'. The problem with this method is the requirement for manual interaction, the requirement for an initial sequence of image frames, and the reliance upon motion detection rather than scene-change detection.

4.2.4 Manually-guided methods

Many algorithms utilise human input to help guide the shadow-removal process. These methods are mainly utilised in removing visually distracting shadows in images, and are not typically suitable for machine-vision applications. For example, Finlayson et al. (Finlayson, Hordley & Drew 2002b) manually specify where the boundary of a shadow is within in image so that a process of 're-lighting' can be implemented in an attempt to remove the shadow. Chuang et al. (Chuang, Goldman, Curless, Salesin & Szeliski 2003) use blue-screen images to easily identify and extract shadowed regions from an object, and overlay them onto another natural scene (for special effects in films). Wu and Tang (Wu & Tang 2005) use a Bayesian technique where the user selects four different areas of the image: shadow regions; shadow-free regions; items to be excluded; or unknown items, for a specific image. The algorithm then uses the shadow smoothness to determine the extent of the shadow, and provides a bound region for it.

4.2.5 Edge-based methods

The majority of scene-change methods use some form of edge detection, including many that are described in the previous sections of the literature review. Some use

shadow direction to try to isolate shadows from objects, such as in the work by Wang et al. (Wang, Chung, Chang & Chen 2004) which attempts to determine the location of the sun from edges of shadows cast by cars. Most, however, use edges to determine objects within scenes and then subsequently process the objects for classification as either genuine objects or cast shadows. The disadvantage with these types of edge-based methods is that they rely on distinct contrasts between shadowed and non-shadowed areas. This creates a problem with soft shadows, as the performance of the algorithm is dependent upon sensitivity thresholds set in the edge-detection algorithm. Another example is given by Xiao, Han, and Zhang (Xiao, Han & Zhang 2007), who detect outlines of vehicles and shadows on highways and subsequently attempt to determine whether the components causing the outlines are vehicles or shadows. Only images with hard shadows are present in the paper. Similarly, Javed and Shah (Javed & Shah 2002) use a threshold-based gradient method to determine regions of detail in a surveillance application, which are then analysed for shadow/object segmentation.

Conversely, the method presented by Stauder et al. (Stauder, Mech & Ostermann 1999) requires a significant penumbra for detection of shadow edges. They state that the light sources must be of 'nonnegligible' size. All of the examples in their paper are from indoor environments where artificial lighting creates a large visible penumbra, and their method would not be suitable for outdoor images where the penumbra is often negligible in size. Similarly, Xu et al. (Xu, Liu, Li & Tang 2005) use gradient filters to detect a smooth reduction in intensity across the penumbra for images captured in indoor environments. Their method also presents results from images captured in environments with artificial lighting and shadows with large penumbra. Further, they require the setting of edge-detection parameters for their scene-change detection process.

Funka-Lea and Bajcsy (Funka-Lea & Bajcsy 1995) use a known probe in the image to recover shadow direction and detail within an image. Their method attempts to recover the umbra and penumbra shadow components separately, and requires simple scenes with plain-coloured backgrounds only. Their method would likely perform poorly in natural or cluttered environments.

Complex scenes containing shadows cast on detailed surfaces can cause a performance reduction with edge-based methods, as multifaceted shadows will increase the number and diversity of shadow edges detected in the image.

4.2.6 Texture-based methods

Texture-based methods utilise complex and predictable areas of the image as a cue for identification of areas with similar properties. While edge-based methods are best served by plain, uncomplicated image content for robust operation, texture-based methods perform best on structured, detailed scenes.

Leone, Distante and Buccolieri (Leone, Distante & Buccolieri 2006) use a multi-layered approach to remove shadows from objects in a surveillance application. They initially create a difference image and then threshold it to remove noise (a fixed-threshold filter for the whole image), then define scene-change and shadow regions which are segmented using a 2D-Gabor texture analysis. Their algorithm requires the setting of four empirical parameters. The method was trialled only on a single test from an indoor environment with very soft shadows on a wooden floor, with the parameters tuned to optimise performance for their image set.

Liu and Gleicher (Liu & Gleicher 2008) use a human-guided method to select areas of shadow before re-lighting the shadow areas using a texture-based method. Though the results presented are visually acceptable, the necessity for complete human segmentation of shadow regions makes it impractical for use in an automated scene-change detection system.

4.2.7 Other methods

Methods exist that use only grey-scale images for shadow and scene-change detection, such as the method by Ibrahim and Rajagopal (Ibrahim & Rajagopal 2007). Although these methods could be applied to colour images by treating each colour channel independently, they are not discussed in detail in this literature review as their performance would be severely limited by metamerism when colour images are used.

Similarly, many papers describe shadow-detection research based upon the intensity-reduction principle that shadows are simply darker pixels of their underlying scene. These elementary methods typically require several user-set thresholds for shadow-detection and scene-change detection, such as in that by Horprasert, Harwood and Davis (Horprasert, Harwood & Davis 1999). These pure intensity-based methods of scene-change detection are not discussed in detail in this review.

Salvador, Cavallaro and Ebrahimi (Salvador, Cavallaro & Ebrahimi 2001) use a colour-invariant method to compare the edge maps of original and colour-invariant images to detect shadow regions. They use plain, non-textured backgrounds in their research, and the results cannot be applied to real-world scenarios without a significant reduction in performance due to these limitations.

Chang et al. (Chang, Hu, Hsieh & Chen 2002) describe a method based on Gaussian models to segment shadows from people in an image that has already been successfully subtracted from a reference image of the background. This method uses the angle between an object and its cast shadow to determine which part of the detected blob is a shadow. This algorithm requires a flat surface and large, strongly-cast shadows.

Arbel and Hel-Or (Arbel & Hel-Or 2007) describe a method for removing shadows on curved surfaces. Their method requires the shadow regions to be detected prior to analysis, after which they detect the penumbra region and re-light the shadowed region

accordingly. They are specifically looking to relight soft shadows on curved surfaces only.

The methods discussed in this section are not considered to be appropriate for automated scene-change detection, and are therefore not discussed in further detail.

4.2.8 The work of Finlayson, Hordley, and Drew

Finlayson, Hordley, and Drew have published many articles in the area of shadow-detection since 2002 (Finlayson et al. 2002b, Finlayson, Hordley & Drew 2002a, Drew, Chen, Hordley & Finlayson 2002, Jiang & Drew 2003b, Jiang & Drew 2003a, Finlayson, Drew & Lu 2004), with others extending their work (Erbou, Sorensen & Stage 2005). In 2006, they presented a landmark paper entitled 'On the removal of shadows from Images' (Finlayson, Hordley, Lu & Drew 2006) that was a significant advance in the area of shadow removal in images. They used a method of illumination invariance that maps an RGB colour space onto a 2D chromaticity plane, which can then be reduced to a 1D line for Planckian illumination changes (i.e., for illumination changes that run along the Planckian locus in Figure 2.7). They showed that a particular surface colour in the scene will reduce to the same point on the 1D line, as long as the illuminant is Planckian in colour. A more detailed explanation follows.

The equation for a camera RGB pixel measurement, P, is described by the product of illumination, surface reflectance, and camera sensitivities (Gershon, Jepson & Tsotsos 1987) giving:

$$P_l = \int E(\lambda)S(\lambda)C_l(\lambda)d\lambda, \quad l \in \{R, G, B\} \tag{4.1}$$

where E is the illumination spectral power distribution (SPD), S is the surface spectral reflectance function, C_l is the camera response, l is an RGB colour band of the camera, and λ the wavelength of light.

If the camera sensitivities are Dirac delta functions, $C_l(\lambda) = c_l\delta(\lambda - \lambda_l)$, and Equation 4.1 simply becomes

$$P_l = E(\lambda_l)S(\lambda_l)c_l. \tag{4.2}$$

A Planckian illuminant is described by Plank's law:

$$E(\lambda) = Ic_1\lambda^{-5}\frac{1}{\exp(\frac{hc}{kT\lambda}) - 1}, \tag{4.3}$$

where c_1 is a constant, h is Planck's constant, v the frequency, k Boltzmann's constant, T the temperature of the body, c is the speed of light, and I is a variable controlling

the overall intensity of light. Finlayson et al. use an approximation of Plank's law known as Wien's approximation, which holds when $hc \gg kT\lambda$ so that

$$E(\lambda) \approx Ic_1\lambda^{-5}\frac{1}{\exp(\frac{hc}{kT\lambda})}. \tag{4.4}$$

With this approximation, the pixel measurement becomes

$$P_l = Ic_1\lambda_l^{-5}\exp(-\frac{hc}{kT\lambda_l})S(\lambda_l)c_l. \tag{4.5}$$

Forming the band-ratios defines the two-vector X chromaticities for an RGB image:

$$X_j = \frac{P_j}{P_G}, \quad j = R, B, \tag{4.6}$$

i.e., $X_R = P_R/P_G$ and $X_B = P_B/P_G$. Substituting the expressions for P_l from 4.5 into 4.6 forms the chromaticity coordinates:

$$X_j = \frac{\lambda_j^{-5}\exp(-\frac{hc}{kT\lambda_j})S(\lambda_j)c_j}{\lambda_G^{-5}\exp(-\frac{hc}{kT\lambda_G})S(\lambda_G)c_G} \quad j = R, B. \tag{4.7}$$

Taking the logarithm X_j' of X_j gives

$$X_j' = \log\frac{s_j}{s_G} + \frac{1}{T}(e_j - e_G), \quad j = R, B, \tag{4.8}$$

where $s_j \equiv \lambda_j^{-5}S(\lambda_j)c_j$ and $e_j \equiv -\frac{hc}{k\lambda_j}$. Putting 4.8 into vector form gives:

$$\underline{X}' = \underline{s} + \frac{1}{T}\underline{e}, \tag{4.9}$$

where $\underline{s}$ is a two-vector which depends on the surface and camera, and $\underline{e}$ is a two-vector which depends upon the camera. As illumination temperature changes, the log-chromaticity vector for a particular surface moves along a straight line.

Finlayson et al. then discuss the process of creating a grey-scale illumination invariant image, which is used to 're-colour' areas of an image detected to be in shadow, by using edges found in the colour image to define shadow boundaries. The results in their papers demonstrate good shadow removal for strong shadows, but recreating their experiments raised some differences between their results and the results reported in this thesis. This can be demonstrated by referring to the measured points ahead in Figure 4.6. The figure shows the log-chromaticity points on a GMB chart when undergoing illumination change from daylight to skylight. If Finlayson et al.'s assumptions are valid, then the lines connecting each pair of start and end points should be parallel. Even though their data is noisy, it is clear that the lines would not be parallel, even if

the noise was removed. This is likely due to three assumptions made in their paper, each of which are described below:

1) The camera sensitivities are Dirac-delta functions. Finlayson et al. acknowledge that cameras do not have Dirac-delta colour sensitivities, and that this can reduce the accuracy of their model.

2) Skylight is well-represented as a Planckian black-body radiator. In outdoor scenes the colour temperature of sunlight is approximately Planckian (Henderson 1977) whereas the blue skylight that colours outdoor shadows is primarily caused by Rayleigh scattering of sunlight in the atmosphere (Lynch & Livingston 1995), a process which is not modelled by black-body radiation. However, the colour temperatures of skylight can (but not always) closely match the colour temperatures of a black-body radiator (Hernandez-Andres et al. 1999).

3) Wien's approximation is a good fit to the Planckian representation of visible light. Wien's approximation is valid when $h\nu \gg kT$, and Finlayson et al. state that the approximation is valid for the range of canonical lights between 2,500K and 10,000K. Using the wavelength for green light (550 nm), the wavelength that human vision is most sensitive to, and a standard daylight colour temperature of 6500K:

$$h\nu = 6.6 \times 10^{-34} \times \frac{3 \times 10^{8}}{550 \times 10^{-9}} = 3.6 \times 10^{-19} \text{ J}. \tag{4.10}$$

and

$$kT = 1.38 \times 10^{-23} \times 6500 = 0.9 \times 10^{-19} \text{ J}. \tag{4.11}$$

From Equations 4.10 & 4.11 $h\nu$ is only 4 times larger than kT, i.e., $h\nu \gg kT$ is not valid. This effectively makes the use of Wien's assumption a relatively poor approximation for the black-body radiation for visible light.

The combination of wideband colour filters, the associated errors arising from Wien's approximation and variation of skylight from the Planckian curve can result in varying degrees of curvature and angle in the log-chromaticity plot of the camera band ratios for outdoor illumination (as can be seen in Figure 4.6). Despite these approximations, Finlayson et al. demonstrate reasonable shadow removal results when this process is integrated with their edge-based method of shadow detection.

4.2.9 Summary

The methods described above for scene-change and shadow-detection are summarised in Table 4.1 along with their respective restrictions, based on the objectives given in Section 4.1. No method has been found in the literature that meets all of these objectives.

Aside from the requirements outlined in the objectives, there are specific challenges that an algorithm must handle to obtain robust performance:

- Variable ground planes that complicate the shape, size, and intensity of shadows.
- Interacting shadows from different objects.
- Changing shadow colour, depending upon the strength of the direct and ambient light sources and the colour of the surfaces under the shadow area. Ambient light-source colour is affected by atmospheric conditions (including cloud cover) and surface reflectance and colour of neighbouring objects and the ground.
- Colour gradients across the penumbra/umbra areas of a shadow, from colour differences between direct and ambient light sources.
- Anti-aliasing of shadow edges by the integration effect of a pixel area, merging true scene detail with shadow.
- Changing of shadow intensity with direct lighting intensity.
- Multiple shadows cast by multiple light sources (e.g., artificial lighting at a sports stadium, or street lighting in traffic applications).

4.3 SCENE-CHANGE DETECTION BASED ON DUAL ILLUMINATION

A generalized method of shadow detection is now developed by extending the Planckian-restricted method of Finlayson et al. to incorporate pairs of differently coloured illuminants that do not necessarily fall upon the Planckian locus, and by using cameras with broadband colour filter responses. This increases the applicability of illumination invariance to incorporate artificial illumination and standard commercially available cameras.

Two illuminants are defined: an ambient source E_a that illuminates all surfaces in a scene (including cast shadows) and a directional source E_d that only illuminates surfaces in its direct line of sight. Let n represent the proportion of added illuminant E_d, where $0 \leq n \leq 1$. $E(n, \lambda)$, the illumination function, may now be defined:

$$E(n, \lambda) = E_a(\lambda) + nE_d(\lambda), \quad n \in [0, 1]. \tag{4.12}$$

Substituting 4.12 into 4.1 defines the generalized model for dual illumination, where the camera's pixel response becomes:

$$P_l = \int (E_a(\lambda) + nE_d(\lambda))S(\lambda)C_l(\lambda)d\lambda, \quad l \in \{R, G, B\}, \quad n \in [0, 1]. \tag{4.13}$$

Table 4.1 A list of relevant scene-change and related shadow-detection algorithms and their performance requirements with regards to the objectives of Part III. 'User input' indicates whether any manual interaction is required for the operation of the algorithm (excluding calibration). 'Arb. thresh' indicates that setting of arbitrary or empirical thresholds is required for operation of the algorithm. 'Signif. training reqd.' indicates that the method requires substantial amounts of training (e.g., images taken over the course of a day) for operation of the algorithm.

Authors	**Uses object shape**	**Uses edges**	**User input**	**Arb. thresh.**	**Indoor & Outdoor operation**	**Signif. training reqd.**
Bhattacharyya & Johnson	Y	Y		Y	Y	
Chuang et al.		Y	Y	Y	Y	
Eisemann & Durand				Y		
Erbou et al.	Y	Y		Y	Y	
Finlayson et al. 2002a		Y	Y	Y		
Finlayson et al. 2002b	Y	Y		Y	Y(2)	
Finlayson et al. 2004	Y	Y		Y	Y(2)	
Finlayson et al. 2006	Y	Y		Y	Y(2)	
Funka-Lea & Bajcsy	Y	Y	Y	Y	Y	
Javed & Shah	Y	Y		Y	Y	
Jiang & Drew	Y	Y		Y	Y	
Leone et al.		Y		Y	Y	
Liu & Gleicher		Y		Y	Y	
Levine & Bhattacharyya	Y	Y		Y	Y	
Lu, Drew & Finlayson		Y		Y		
Lu, Xin, Kong, Li & Wang	Y	Y		Y	Y	
Marchant & Onyango	Y	Y		Y	Y (2)	
Matsushita, Nishino, Ikeuchi & Sakauchi	Y	Y	Y	Y	Y	Y
Nadimi & Bhanu	Y	Y		Y	Y	
Porikli & Thornton					(1)	Y
Stauder et al.	Y	Y		Y	Y	
Wang et al.	Y	Y		Y	Y	
Weiss		Y		Y	Y	Y
Withagen et al.				Y	Y	
Wu & Tang			Y	Y	Y	
Xiao et al.	Y	Y		Y	Y	
Xu et al.	Y	Y		Y	Y	

(1) Assumes shadows are a reduction in intensity only.

(2) Yes, though restricted to Planckian lighting.

Section 3.6.1.3 has shown that the green channel in standard colour cameras with the commonly used Bayer colour filter array (CFA) can exhibit less noise than the red or blue channels. As the effect of pixel noise is compounded with small denominators when forming band ratios, the green channel has been chosen as the denominator, resulting in $x = P_R/P_G$ and $y = P_B/P_G$, so that:

$$x = \frac{\int (E_a(\lambda) + nE_d(\lambda))S(\lambda)C_R(\lambda)d\lambda}{\int (E_a(\lambda) + nE_d(\lambda))S(\lambda)C_G(\lambda)d\lambda}, \quad n \in [0, 1] \tag{4.14}$$

and

$$y = \frac{\int (E_a(\lambda) + nE_d(\lambda))S(\lambda)C_B(\lambda)d\lambda}{\int (E_a(\lambda) + nE_d(\lambda))S(\lambda)C_G(\lambda)d\lambda}, \quad n \in [0, 1] \tag{4.15}$$

so that

$$y = x\frac{\int (E_a(\lambda) + nE_d(\lambda))S(\lambda)C_B(\lambda)d\lambda}{\int (E_a(\lambda) + nE_d(\lambda))S(\lambda)C_R(\lambda)d\lambda}, \quad n \in [0, 1] \tag{4.16}$$

or

$$\begin{aligned} \log(y) = &\log(x) + \log\left(\int \left(E_a(\lambda) + nE_d(\lambda)\right) S(\lambda)C_B(\lambda)d\lambda\right) \\ &- \log\left(\int \left(E_a(\lambda) + nE_d(\lambda)\right) S(\lambda)C_R(\lambda)d\lambda\right), \quad n \in [0, 1] \end{aligned} \tag{4.17}$$

which defines a space that distributes the colour data relatively evenly, enabling easier colour segmentation. Each different surface will have a characteristic shadow curve in the band-ratio or log band-ratio (LBR) spaces defined in Equation 4.17, dependent upon the illuminants and the colour response of the camera. One restriction on this model is that the green pixel response, P_G, must be greater than zero to avoid singularities in x and y. Singularities are unlikely to occur given canonical illumination, as most RGB cameras have relatively broadband colour filters (Nakamura 2006) and will therefore exhibit some response in the green channel if the blue or red channels are illuminated.

An experiment was conducted using artificial illumination in controlled conditions to test the validity of the generalized dual-illuminant model. A fluorescent tube (8W ZhaoGuang F8T5/D 6500K) was used to approximate the ambient light source E_a and a halogen bulb (12V 50W Bi-globe[1]) was used to approximate the point source E_d. The halogen source was rotated around a GMB chart and a series of images of the chart recorded, using the setup shown in Figure 4.1, to simulate the effect of changing n from one (halogen perpendicular to the chart) to zero (halogen behind the chart). 27

[1]No specifications were available for colour temperature, but it is expected to be near the typical colour temperature of 3,000K for halogen lighting.

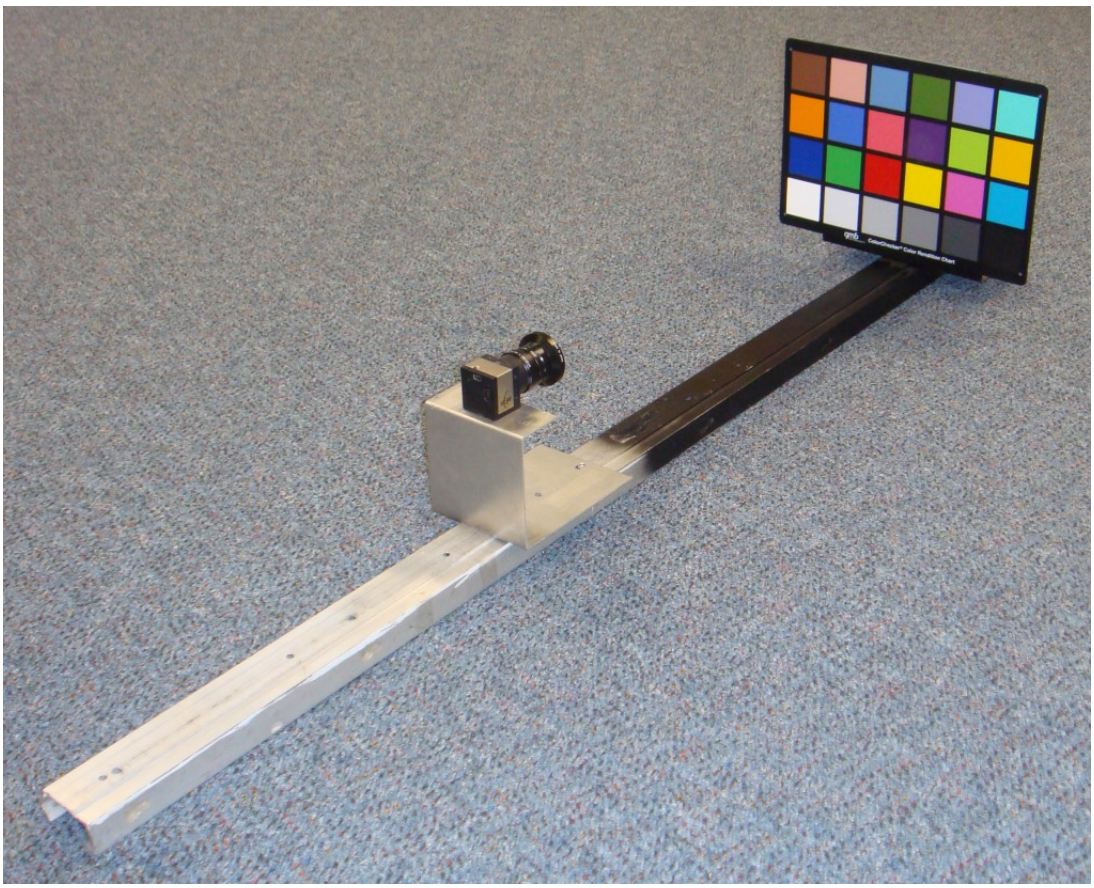


Figure 4.1 The camera and GMB chart setup for measuring the characteristic shadow curves of dual-illuminant environments. The camera records images of the GMB chart as the direct illumination that falls upon the chart varies.

Images were captured as the light was moved at a constant angular speed around the chart. The absolute angle of the light is not important to the analysis due to the use of band-ratios to calculate the chromatic position in x/y space, providing there is an adequate number of images captured. The measured $n = 0$ and $n = 1$ points are used as an input into the model in Equation 4.17, which allows the x and y points for any value of n to be calculated.

Figure 4.2 demonstrates the close agreement between modelled curves using Equation 4.17 and the measured data points. Note that Finlayson et al.'s method for shadow detection would be expected to fail in this environment as the curves are neither straight or parallel. This would be due to one or a combination of the narrow-band camera assumption, Wien's black-body approximation error, or the non-Planckian colour of the fluorescent lighting.

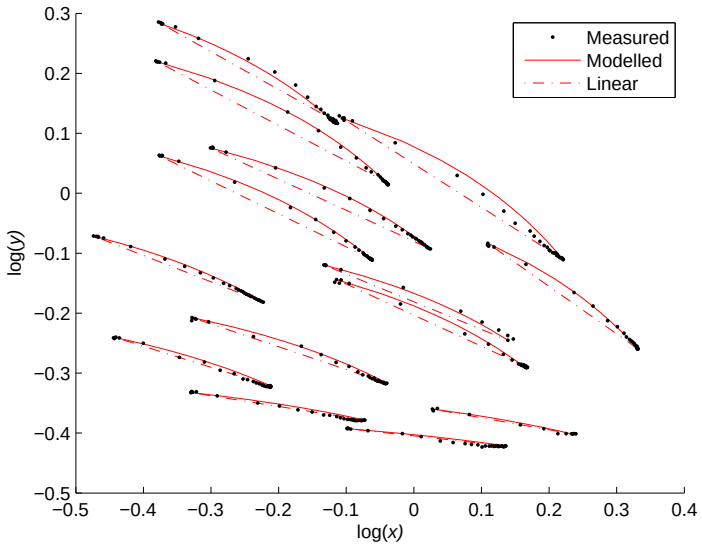


Figure 4.2 The measured and modelled characteristic shadow curves (changes in band ratios) for changing intensities of halogen lighting on a GMB chart in an environment with ambient fluorescent lighting (only panels 1-14 shown for clarity). The linear band-ratio changes are also shown for comparison.

4.3.1 Scene-change detection

The dual-illumination model can be used to detect scene-changes from a sequence of images containing variations in cast shadows. A reference image is used as a basis for comparing subsequent images' details to identify whether a pixel's change is due to a change in the scene or to a cast shadow. A shadow-free image is constructed by selecting pixels from either the reference or a comparison image, as determined by the scene-change detection algorithm.

4.3.1.1 Assumptions

It is assumed that changes between images of a scene are due to either cast shadows or a physical object that has entered the scene. It is also assumed that all cast shadows are illuminated in part by at least the ambient illumination source, and that the illuminants are constant between the reference image and the image being processed.

4.3.1.2 Model calibration

The illumination-invariant camera model requires calibration prior to shadow removal for a given pair of illuminants. This can be achieved with the use of a colour chart such as a GMB chart that contains multiple coloured patches: two images can then be recorded, one with ambient illumination only and one with combined ambient and direct illuminants, for each of the panels in the chart. The average value of the centre 50×50 pixels in each panel (as per Figure 3.6) is used to calibrate the start and end points only of the characteristic shadow curves (shown in Figure 4.2). The curve between the points arises from the model.

4.3.1.3 Pixel invariance and shadow removal

Pixel P_{ref} in the reference image has a curve in LBR space $\left(L_{P_{ref}}(n)\right)$ that is associated with it. This curve is given by Equation 4.17, with the endpoint $L_{P_{ref}}(1)$ obtained directly from the reference image. The curve represents the pixel colour change that would occur for the pixel P_{ref} as the illuminant changes from $E_a + E_d$ to E_a.

Suppose P_{cal} is a pixel in the calibration image, and $L_{P_{cal}}(n)$ is its closest characteristic shadow curve in LBR space. Given the limited number of unique colour panels in the illumination model calibration, we must approximate $L_{P_{ref}}(n)$ by $\hat{L}_{P_{ref}}(n)$ where

$$\hat{L}_{P_{ref}}(n) = L_{P_{cal}}(n) + d \tag{4.18}$$

and

$$d = L_{P_{ref}}(1) - L_{P_{cal}}(1) \tag{4.19}$$

and P_{cal} is chosen so that $|d|$ is minimized amongst all the possible choices of P_{cal}. In effect the closest calibration curve end-point to $L_{P_{ref}}(1)$ is selected by means of mimimal Euclidean distance in LBR space. The value of $n = 1$ is used as the reference image as it is the base from which shadows are detected in subsequent images, and is likely to be fully illuminated.

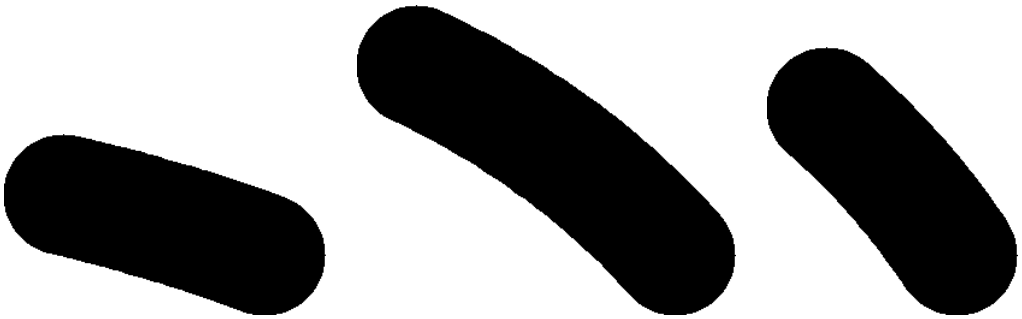
A threshold distance t defines an area around each characteristic shadow curve to enable segmenting of shadowed pixels. Consider a pixel P_{comp} in the comparison

image. Let ρ be its point in LBR space. Thus $\rho = (\log(x), \log(y))$ where x and y are calculated using Equations 4.14 & 4.15 respectively. Then P_{comp} is defined as being in a cast shadow if it has intensity less than that of P_{ref} and if ρ lies inside P_{ref}'s threshold area, defined as any point within a fixed distance of $\hat{L}_{P_{ref}}$ (set by value t). A pixel P_{free} can be defined that is free of changes in cast shadows between P_{ref} and P_{comp} by:

$$P_{free} = \begin{cases} P_{ref} & \text{if } \rho \in \hat{L}_{P_{ref}}(n) \otimes circ(t) \text{ AND } V(P_{comp}) < V(P_{ref}) \\ P_{comp} & otherwise \end{cases} \tag{4.20}$$

where $circ(t)$ is a disc of radius t, $\otimes$ is the convolution operator, and V is pixel intensity. Therefore a full colour image free of changes in cast shadows between the reference and comparison images can be created from independently analyzed pixels.

Examples of three threshold-bound regions in LBR space are given in Figure 4.3.



(a) (b) (c)

Figure 4.3 Examples of fixed-value shadow regions in LBR space of panels from the GMB chart: (a) red, (b) orange, (c) purplish-blue.

4.3.1.4 Noise

Section 3.6 demonstrated that noise within images is exhibited as temporal and spatial variations in pixel values, which can cascade into the resultant shadow-removed images if left unfiltered. Image noise is particularly detrimental to analysis of shadows due to the division by small green pixel colours in the formation of the colour band ratios in Equations 4.14 & 4.15. A series of filters was applied to the images during shadow removal to improve the performance of the algorithm in the presence of noise:

- A 2-D Gaussian filter was used to reduce temporal and spatial variations due to noise in the reference and comparison images.
- Small isolated areas set to P_{comp} from Equation 4.20 were considered erroneous and removed.

- Morphological opening (using a 3x3 unity structuring element) was applied to the resultant areas set to P_{comp} in Equation 4.20 to fill small gaps left by the shadow detection algorithm.

Values for parameters were empirically determined by trial and error for the size of the kernel used for Gaussian filtering (9x9, 2 standard deviations) and the minimum size for isolated areas of the size of P_{comp} in Equation 4.20 (0.25% of the total image area).

4.3.2 Results

The method of shadow removal described in Section 4.3.1.3 was applied to images captured in controlled artificial lighting conditions and also to daylight-illuminated outdoor images. A standard commercially available colour camera (specifications in Table 3.6) was used. All digital effects such as gamma and colour balance were disabled, and the black-level output of the camera was calibrated to ensure an RGB output of (0,0,0) for dark conditions. Saturated pixels in the reference image were removed from processing as they would generate false colour band ratios. The value t was empirically determined by varying t, processing a batch of images captured in both indoor and outdoor conditions for each t, and choosing the value that provided approximately equal responses to false-positive shadow and true-negative object detection. t was fixed to 0.05 for all subsequent processing (results given ahead in Tables 5.2 & 5.3 justify the selection of $t = 0.05$).

4.3.2.1 Artificial illumination

An environment using fluorescent ambient and directional halogen illuminants was used to generate coloured shadows for objects on textured backgrounds. The calibration data for this environment is shown in Figure 4.2. The processes described in Section 4.3.1 was applied to the images of fruit added to a highly-textured background shown in Figures 4.4a & 4.4b. To allow the application of the noise filters, a mask was created using Equation 4.20 where black represents areas in shadow or areas not affected by the insertion of objects into the image. The raw mask is shown in Figure 4.4c, created from the reference image in Figure 4.4a and comparison image in Figure 4.4b. Post-processing filters were applied to the raw mask and the resultant filtered mask is shown in Figure 4.4d. By using the black areas of the mask to indicate pixels from the reference image and the white parts to indicate pixels from the comparison image, the final cast-shadow free image in Figure 4.4e was created.

All cast shadows from the halogen light illuminating the fruit have been removed in Figure 4.4e, along with a significant proportion of 'dark' shadows underneath the fruit. The remaining shadowed areas in the final image are areas in the umbra of both the

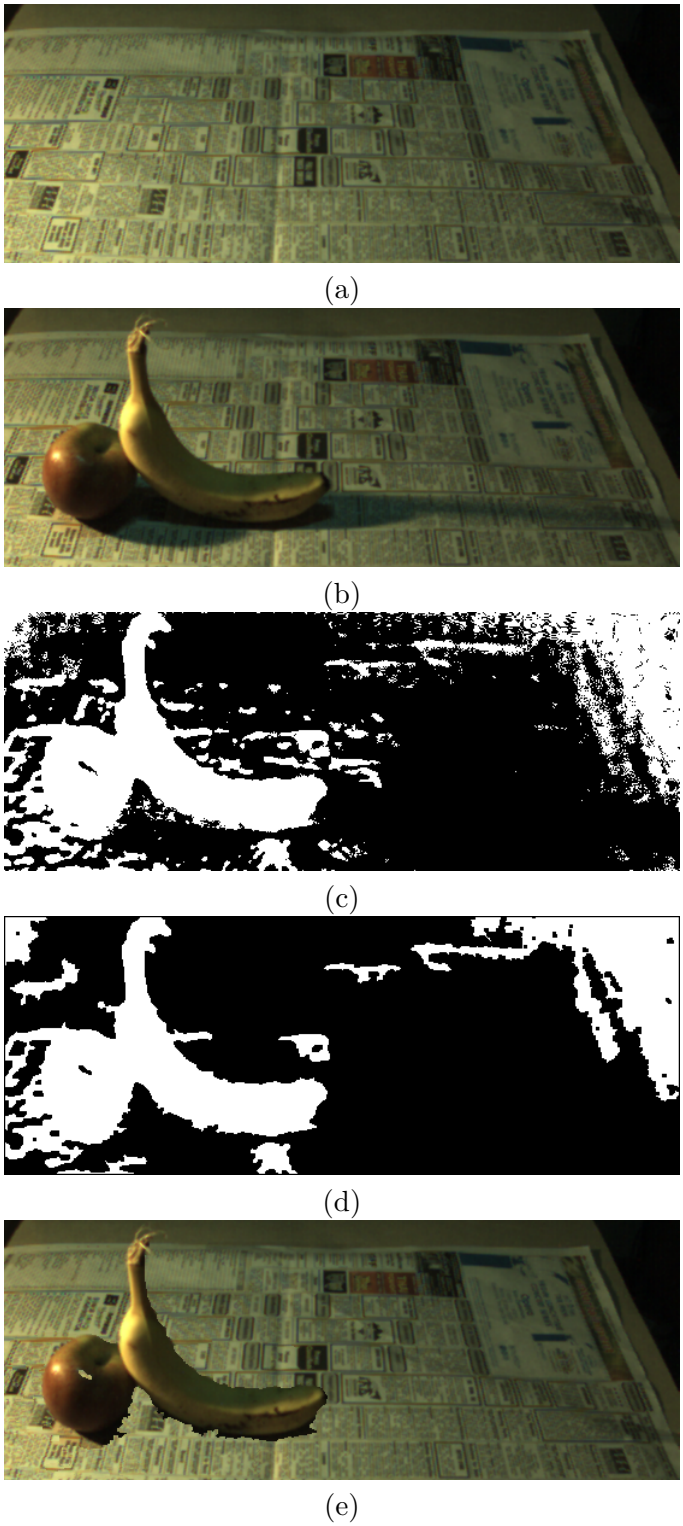


Figure 4.4 The shadow removal from objects on a textured surface with artificial illumination. (a) reference image; (b) image containing objects and cast shadows; (c) raw mask created using Equation 4.20 showing pixels from the shadow-containing image (white) to be fused with pixels from the reference image (black); (d) filtered mask used to create the final image in (e) with shadows removed.

(a)

(b)

(c)

Figure 4.5 Shadow removal from a cup on a textured surface. (a) Reference image; (b) image with objects and cast shadows; (c) image with cast-shadows removed.

halogen and ambient fluorescent tube, effectively creating colourless voids whose pixel colour does not fall on the approximated shadow curve of P_{ref}, $\hat{L}_{P_{ref}}(n)$, for a given pixel. The fruit objects have almost completely been detected as valid scene objects, with only a specular glint near the top of the apple wrongly detected as a cast shadow.

Figure 4.5 illustrates another shadow-removal example using a coloured textured scene. The comparison image contains cast shadows as well as a purple cup and a hand. Almost all of the cast-shadow region has been successfully removed except the area underneath the fingers and handle where the ambient fluorescent lighting was unable to illuminate the background. Several small patches on the cup and hand, on areas in the textured background that were of a similar hue, were wrongly detected as cast shadows due to metamerism.

4.3.2.2 Natural illumination

Nature presents its own form of dual-illumination in outdoor environments in the form of sunlight and ambient skylight. Figure 4.6 shows the shadow curves for the 18 colours and a neutral panel of the GMB chart measured in an open field (grass) using the method described in Section 4.3.1.2. Panels 4 and 13–15 in particular exhibit curves that are not modelled accurately. This is due to a third illumination source of reflected light from the grass that strikes the GMB chart when it is directed toward the sun. Panels 13-15 are the vivid red, green, and blue panels of the chart that will have the spectral bandwidths most likely to correlate with the Bayer-array colour filters, increasing their sensitivity in Equations 4.14 & 4.15 to changes in the illumination SPD. Similarly, panel 4 is foliage coloured and will therefore have a spectral reflectance close to that of the grass-coloured reflected light, increasing its sensitivity in Equations 4.14 & 4.15 to changes in the illumination SPD to foliage-coloured spectra. Ideally, calibration would occur without any reflective objects causing changes in the illuminants' SPD. The impact of the modelled errors in Figure 4.6 is likely to be small for cast shadows, as the model is calibrated with the start and end points of the shadow curves (n=0 and n=1) where this third light source is either absent or swamped by sunlight.

Figures 4.7 and 4.8 show the results of shadow removal on outdoor images using the calibration data shown in Figure 4.6. The shadows have been removed on the red bricks and grey-coloured decking in Figure 4.7c, with occasional incorrect detection of cast-shadows on the person. The black shoe and dark grey-brown pants on grey-decking, and the brown shirt and skin tone on yellow leaves are occasionally registered as cast shadows due to metamerism. The outdoor scene shown in Figures 4.8a & 4.8b incorporates objects and cast shadows on highly variable surfaces, including autumnal leaves, grass, brick, and concrete. The resultant image shown in Figure 4.8c demonstrates almost complete shadow removal and object integrity, with only occasional missed shadow removal around the fringe of the light-blue coloured cushion, and

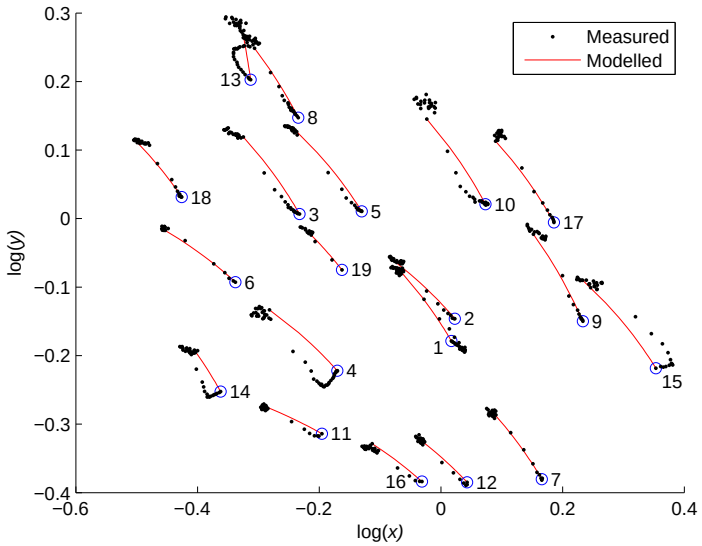


Figure 4.6 Measured and modelled changes in characteristic shadow curves for changing intensities of outdoor lighting on panels of a GMB colour chart (14 panels shown for clarity) using the colour space described in Equation 4.16.

an almost imperceptible border of misclassified cast-shadows between the cushion and the box.

4.3.2.3 Discussion

Using regions around the shadow-curves to threshold shadows affects the metamerism of genuine objects appearing as shadows. The threshold t was empirically set to a fixed-value ($t = 0.05$) which provided an even balance between shadow and object detection. The pre-filtering of images using a Gaussian kernel creates a thin blurred boundary around objects that subsequently causes an outline of wrongly-detected areas of cast shadow, effectively increasing their size. Further, the post-filtering of the scene-change detection masks reduces the sensitivity detection of small shadows. An improved noise filter and thresholding technique that has parameters tuned to the noise characteristics of a pixel would likely further improve the performance of the shadow algorithm by targeting the required noise filtering.

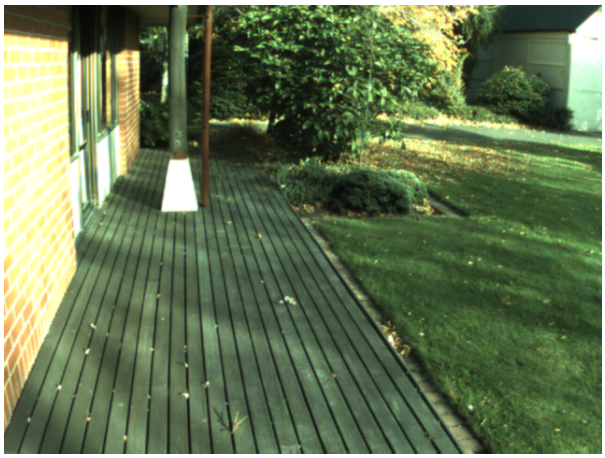
The analyses of shadows within a scene are specific to the camera calibration for a particular pair of direct and ambient illuminants. This affects the application of the technique to automated shadow removal in outdoor images where the atmospheric conditions can change. For example, the ambient SPD of an overcast day will be different from a clear day as the amount of Rayleigh scattered light is reduced due to the occlusion of light by the cloud. In this instance a method of identifying the SPD of the ambient illuminant will be required so that an appropriate illumination-invariant

calibration can be selected for shadow removal.

Effects such as image saturation, specular reflections, or errors in image registration between the reference and comparison images are likely to cause misclassified areas when Equation 4.20 is applied.

The colour-based scene-change detection algorithm that has been developed addresses several of the challenges identified in Section 4.2.9:

- Shadows on non-flat ground planes can be detected.
- Interacting shadows from different objects do not affect performance, as the method is not object-based.
- Changing shadow colour due to variations in the relative strengths of the direct and ambient light sources and the colour of the surfaces under the shadow area are inherently addressed in the design of the algorithm.
- Colour gradients across the penumbra/umbra areas of a shadow, from colour differences between direct and ambient light sources, are also inherently addressed in the algorithm design.
- The use of the dual-illumination concept for shadow detection allows the algorithm to detect anti-aliasing of shadow edges due to integration of pixel area in the camera.
- Any shadow intensity (below the saturation level of the image sensor) can be detected as the algorithm is chromaticity based, not intensity based.
- The direction and number of cast-shadows from an object is not limited, as no assumption is made about the position of the direct illumination source.



(a)

(b)

(c)

Figure 4.7 Shadow-free scene-change detection from 'Outdoor scene 1'. (a) Reference image; (b) image with objects and cast shadows; (c) image with cast-shadows removed.

(a)

(b)

(c)

Figure 4.8 Shadow-free scene-change detection from 'Outdoor scene 2'. (a) Reference image; (b) image with objects and cast shadows; (c) image with cast-shadows removed.

4.3.2.4 Conclusion

A colour-based method for scene-change detection has been developed based on two arbitrarily coloured illuminants. It does not require edge detection or manual guidance, and removes cast shadows from objects within a scene.

The new method has been demonstrated on images illuminated with artificial lighting and on naturally illuminated outdoor scenes. It uses the characteristics of a direct and an ambient illuminant to classify scene changes as shadow on a pixel-by-pixel basis. Experiments conducted with a standard commercially-available colour camera have shown that a straightforward calibration procedure is required to characterize the camera's response to the illumination spectra. The robustness of the method is primarily dependent on levels of image noise that become increasingly prominent for small-valued pixels. Revisiting the objectives laid out in Section 4.1 shows that 5 out of 6 of the objectives for scene-change detection have been met. The method:

- is independent of object shape;
- identifies and removes cast-shadows;
- is able to operate in both indoor and outdoor environments;
- does not require manual interaction;
- does not require a significant number of images for training.

The algorithm requires the setting of thresholds for the pre-processing filters (Gaussian convolution, morphological operators, particle filtering) and for the value of t, and therefore fails to meet the last objective of being free from arbitrary internal thresholds. This is addressed in the following chapter.

Chapter 5

NOISE-LIMITED SCENE-CHANGE DETECTION

A major drawback of the scene-change algorithm described in Chapter 4 is the need for arbitrary or empirically determined thresholds, the values of which can have a significant impact on the performance of the algorithm. It is very common for thresholds to be used in image processing applications; sometimes their relation to performance is obvious, such as when a simple threshold is used to segment high-valued pixels from low-valued pixels, but they are usually embedded deeply into the algorithm and have no obvious connection to performance. The threshold t in the scene-change algorithm described in Chapter 4 is a good example of this, as it defines an area of relative hue-change around a curve on a 2-D chromaticity plane. This makes it difficult to determine an optimal value of t other than by experimentation. Furthermore, changes in cameras or environment may adversely affect the performance of the algorithm for a given value of t.

One reason why thresholds are commonly used for attenuating the effects of noise in images could be due to the unavailability of suitable methods and processes for easily measuring image noise, hence the application of fixed-value, global thresholds to dichotomize data into 'probably genuine' and 'probably noise' categories (as demonstrated by the use of t in Chapter 4). From the experimental results in Chapter 3, this is a non-optimal approach as noise is complex and is highly dependent upon pixel value and the context of how the data is being used.

This chapter of work is targeted at integrating the knowledge developed in Chapters 3 & 4 to complete the objectives defined in Section 4.1. Sections 5.3 & 5.4 describe the noise-limited method, with results comparing the original fixed-threshold and noise-limited methods given in Section 5.5. The interpolation of calibration data is discussed in Section 5.6, with experimental results using the interpolated data given in Section 5.7. The chapter is completed with a discussion and conclusion in Sections 5.8 & 5.9.

5.1 INTRODUCTION

Many image processing applications are confounded by sensor noise and shadows cast upon the scene. All image sensors add noise to a captured image that can reduce algorithm sensitivity and performance, and global filters or fixed thresholds are often applied to limit their effects. Cast shadows can appear as scene changes and are difficult to adequately detect and remove from images and image sequences. In this chapter image noise statistics are coupled with the dual-illumination scene-change algorithm developed in Chapter 4 to provide a novel colour-based method for shadow-free scene-change detection, whose performance is bound by metamerism and image noise.

5.2 OBJECTIVE

The objective of this chapter is to extend the method of scene-change detection developed in Chapter 4 so that it is free from arbitrary internal thresholds. If this is accomplished then all of the objectives in Section 4.1 will have been met.

5.3 MODIFIED BAND-RATIO SPACE

The model presented in Section 4.3 uses a fixed-threshold region (bound by t) for scene-change detection. The value of this threshold (0.05) was determined empirically, with the intention of balancing the performance of genuine scene-change detection versus shadow detection. A logarithmic band-ratio space was used in previous work by Finlayson et al. (Finlayson et al. 2006), that demonstrated that the use of logarithms assisted in linearising black-body illumination change in this space, as do the results from the scene-change method developed in Chapter 4.

By integrating camera noise into the dual-illumination method for scene-change detection, the logarithmic space used in Equation 4.17 is no longer necessary as t will be calculated using the statistical noise model.

The calibration data used in Figure 4.2 can be remapped to the non-logarithmic x/y band ratio space shown in Equation 4.16. The resulting measured and modelled characteristic shadow curves in the non-logarithmic band-ratio space are shown in Figure 5.1.

5.4 NOISE-LIMITED SCENE-CHANGE DETECTION REGIONS

A change in scene content can be modelled by any variation in a pixel's colour from its band-ratio line (or from the $n = 1$ point for single-illuminant scenes). Image noise adds regions of uncertainly around the lines, that vary with pixel intensity.

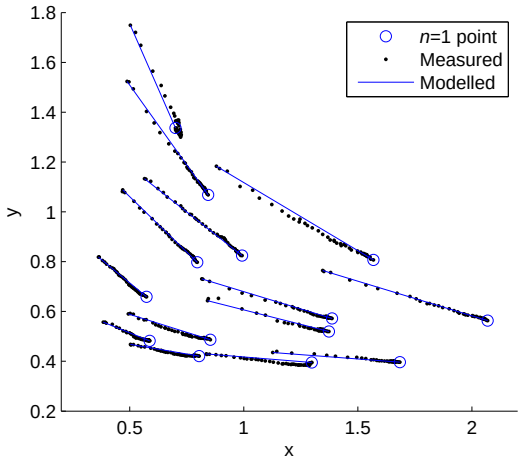


Figure 5.1 Measured and modelled changes in characteristic shadow curves for changing intensities of halogen lighting on panels of a colour chart in an environment with ambient fluorescent lighting (14 panels shown for clarity) using the colour space described in Equation 4.16.

The fixed threshold value t can be replaced by a value determined by pixel noise, which was found to vary with pixel value (see Section 3.6). Full knowledge of the temporal noise characteristics present at each pixel for a particular camera allows the use of standard propagation of errors to determine the noise characteristics, σ_x and σ_y, for each point in x/y space:

$$\sigma_x = \sqrt{\left(\frac{\sigma_R}{P_R}\right)^2 + \left(\frac{\sigma_G}{P_G}\right)^2}\, x \tag{5.1}$$

and

$$\sigma_y = \sqrt{\left(\frac{\sigma_B}{P_B}\right)^2 + \left(\frac{\sigma_G}{P_G}\right)^2}\, y. \tag{5.2}$$

Confidence intervals can then be used to determine noise thresholds for the desired application. A standard value of 2σ was used to provide a 95.45% confidence interval for categorization of scene change. Values of P_R and P_B can be calculated for any value of P_G on any given calibration line in x/y space (from the band-ratios $x = P_R/P_G$ and $y = P_B/P_G$). This allows σ_x and σ_y to be calculated for any position on the calibration line, resulting in a region in x/y space that can be used to statistically determine if a pixel change is likely to be caused by noise or a change in the scene content. Using the measured temporal noise data for the *uEye* CMOS camera (Figure 3.24b) a series of scene-change regions for an example calibration line from Figure 5.1 is shown in Figure 5.2.

Note that the scene-change detection areas vary with the green value of the pixel

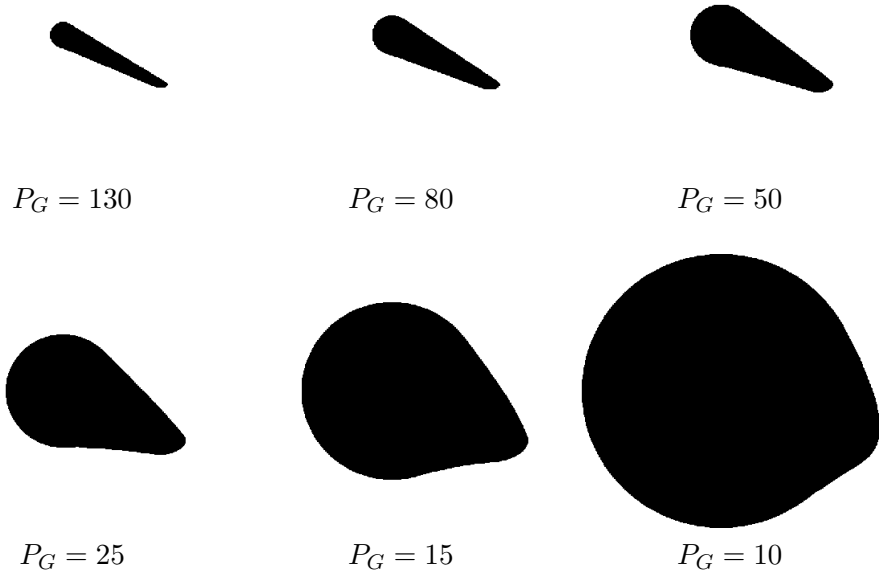
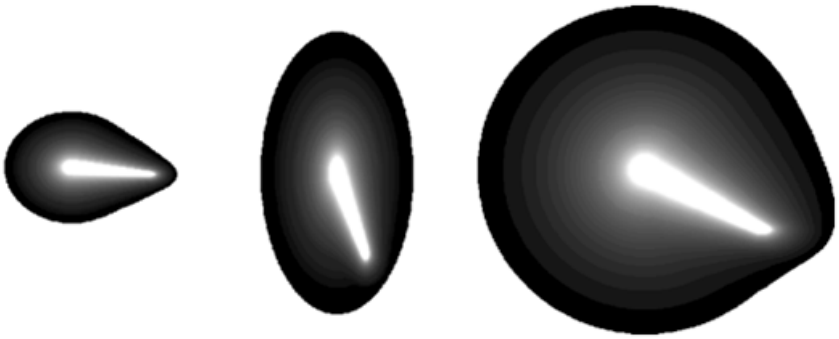


Figure 5.2 Scene-change detection regions for a line in x/y space (direct halogen and fluorescent ambient lighting) for given values of P_G, on a calibration line from Figure 5.1. Regions for larger values of P_G (brighter intensity) are smaller, showing that temporal image noise has a smaller effect on bright pixels and a greater effect on darker pixels.

(P_G). As shown in Figure 3.24b, pixel SNR is greater at high pixel values, which is consistent with the shapes shown in Figure 5.2. The scene-change regions for all values of P_G can be combined to create a composite mask, where each shade of grey defines the mask for its P_G-value (0=black, 255=white). Every RGB combination has its own unique scene-change detection mask for a given illumination, and several example masks are given in Figure 5.3. A pixel from the comparison image is classified as a change in scene if it falls outside the shadow-region determined by the initial P_G value of the corresponding pixel from the original reference image. The masks provide a convenient implementation for scene-change detection as they can be easily pre-calculated for each calibration curve, significantly reducing the number of processing operations required for the operation of the scene-change algorithm.

One significant benefit from using a noise-limited method for detecting the scene-change boundary is that the need for image pre-processing is removed. The Gaussian filtering and morphological operations performed by the fixed-threshold method of scene-change detection (see Section 4.3.1.4) are no longer required.



(a) (b) (c)

Figure 5.3 Examples of overlaid noise-limited scene-change detection regions from Figure 5.1. Each shade of grey represents the P_G-value of the original pixel (gamma of 1.5 applied to the images to improve visual clarity). Figure (c) is the single overlaid region for the regions given in Figure 5.2. Images are to scale with each other.

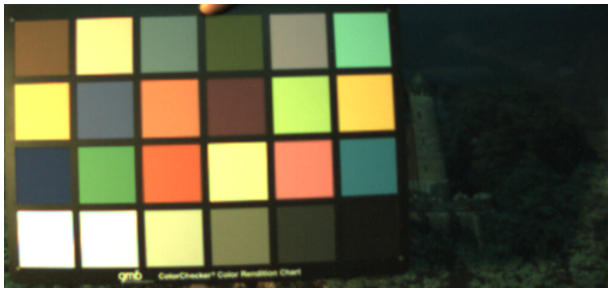
5.5 SCENE-CHANGE METHOD COMPARISON USING INITIAL CALIBRATION DATA

Eight image sets were collected that contained a variety of different objects placed in front of one of four backgrounds. Two of the backgrounds were large printed photographs of natural scenes (a castle on a hill, and a beach scene), the third was a coloured pattern on material, and the fourth was the same as the third background with a GMB chart included. The objects included a pencil sharpener, cup, a colourful stuffed toy, a GMB colour chart, and an arm with a wristwatch or ring.

The new noise-limited method for scene-change detection was tested on the 8 image sets and the results are compared to the output from the fixed-threshold algorithm. Both algorithms used the same calibration data from Section 4.3. Shown in Figure 5.4a is the original background image for the image set 'Castle 2', which is used to provide a coloured and textured background against which scene changes and cast shadows are to be identified.

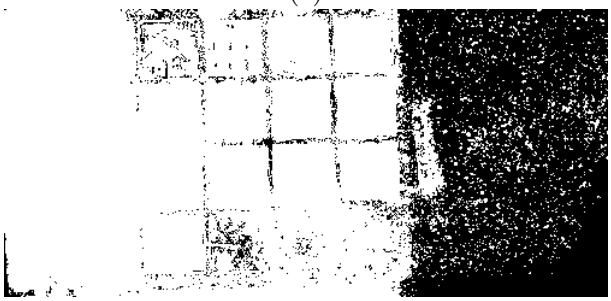
Figure 5.4b shows the scene of 5.4a with the addition of the GMB chart that is casting a strong shadow onto the right side of the image. A perfect scene-change detection algorithm would identify only the GMB chart as genuine scene change. The method of scene-change detection taken by the fixed-threshold algorithm is to threshold all pixel colour changes an equal distance away from the characteristic shadow curve in the log band-ratio space. The threshold was empirically chosen so that the majority of shadowed pixels are removed, balanced against the undesirable metameric effect of genuine scene change being incorrectly detected as shadow. The resultant scene-change mask from processing the images with the fixed-threshold method is shown in Figure 5.4c. The majority of the shadows are removed but consequently there are significant portions of scene-change that have been incorrectly detected as shadowed

(a)



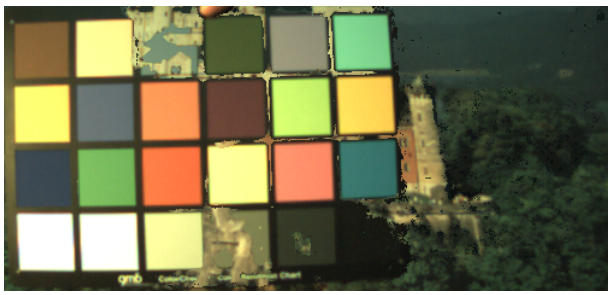
(b)

(c)

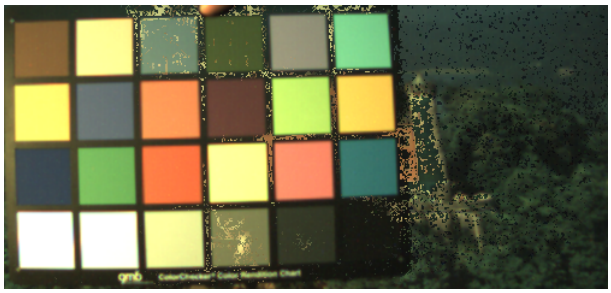


(d)

Figure 5.4 Threshold and noise-limited scene-change detection for image set 'Castle 2': (a) original scene with background; (b) A GMB colour chart with cast shadows placed in front of the background; the masks from (c) fixed-threshold scene-change detection; and (d) noise-limited scene-change detection.



(a)



(b)

Figure 5.5 The reconstructed shadow-free images from Figure 5.4 using (a) the fixed-threshold mask and (b) the noise-limited mask.

pixels, including panels 3 and 22 of the GMB chart and substantial portions of the black regions between and around the GMB colour panels. It is not unexpected that pixels with very small RGB values are classified as shadows, as the signal-to-noise ratio approaches unity making it difficult to distinguish black objects from dark shadows.

Figure 5.4d shows the resultant scene-change mask from processing the images with the noise-limited scene-change algorithm. There are 3 substantial differences between the output masks from the fixed-threshold and noise-limited methods: 1) The shadow-removed areas from the noise-limited algorithm show speckle, whereas the fixed-threshold algorithm tends to have groups of misclassified pixels. 2) The integrity of the GMB chart is substantially improved when the noise-limited algorithm is used. Panels 3, 22, and the black regions around the panels are better detected than by the fixed-threshold algorithm, although there is minor speckle in some areas of the objects. 3) The right-hand column of the castle has been incorrectly detected as shadow by the noise-limited algorithm.

Figure 5.5 shows the resultant shadow-free images when the scene-change masks are used to recombine the original and comparison images. Similar results can be seen when the algorithms are applied to the test set 'Castle' shown in Figures 5.6 & 5.7.

The large areas of castle that were incorrectly determined as shadows by the noise-limited algorithm were analysed further, and the errors can be attributed to the limited

(a)

(b)

(c)

Figure 5.6 Threshold and noise-limited scene-change detection for image set 'Castle': (a) Part of the background image in Figure 5.4 with a pencil sharpener and purple cup casting their shadows onto the background; the masks from (b) fixed-threshold scene-change detection; and (c) noise-limited scene-change detection.

(a)

(b)

Figure 5.7 The reconstructed shadow-free images from Figure 5.6 using (a) the fixed-threshold mask and (b) the noise-limited mask.

number of calibration shadow lines obtained from the GMB chart, where every pixel's shadow-line is required to be approximated by one of the 18 differently coloured panels. This is illustrated graphically in Figure 5.8. Suppose the true scene-change line for a pixel is given by the solid line in Figure 5.8a, and its two nearest-neighbour lines are given by the dashed and dotted lines (shifted to the $n = 1$ point of the true scene-change line). Both of the fixed-threshold scene-change regions for the nearest-neighbour curves significantly overlap the true scene-change line, as shown in Figure 5.8b. Similarly, the scene-change regions for the low P_G-value noise-limited regions significantly overlap, as shown in Figure 5.8c. However, for high P_G values the noise-limited scene-change regions shrink substantially, significantly reducing the amount of overlap of either nearest-neighbour region on the true scene-change line. Therefore any high-intensity pixels whose true scene-change line is not well approximated by its nearest-neighbour line would receive a poor scene-change region, resulting in a high likelihood of error. This effect was observed with other experiments. Figure 5.9 illustrates an example where a shadow is cast over a coloured cloth background. The fixed-threshold method eliminates shadows from below the elephant whereas the noise-limited method performs poorly. In both the castle and the elephant examples, P_{comp} fell inside the fixed-threshold scene-change region, but outside of the noise-limited scene-change region.

Table 5.1 lists the results from the 8 test sets comparing the fixed-threshold and scene-change detection algorithms for each image set. Areas of genuine scene change

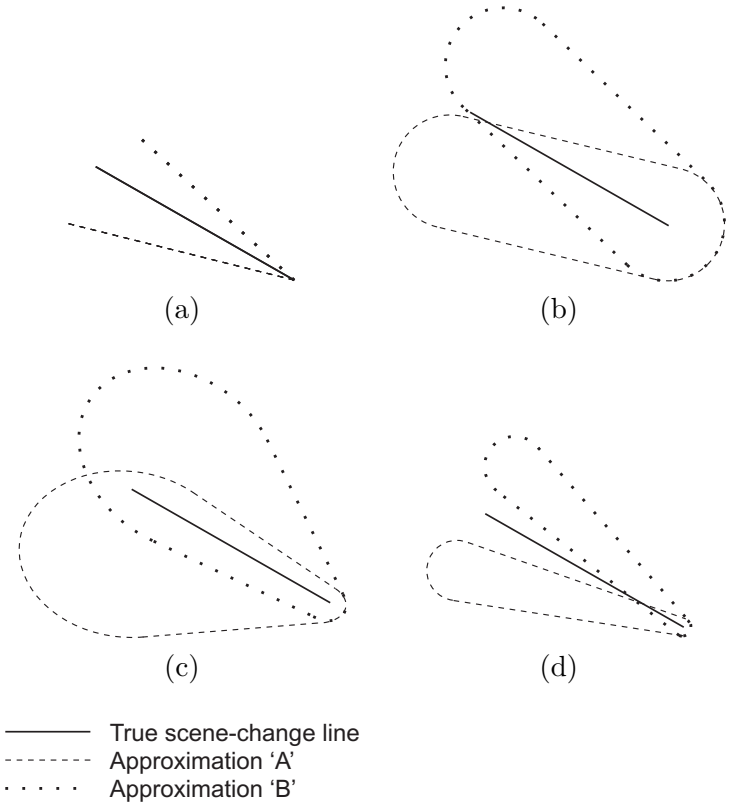


Figure 5.8 Diagram illustrating the effective areas of approximated scene-change regions: (a) the true and approximated scene-change lines, shifted to the $n = 1$ point of the true line, (b) the significantly overlapping regions using the fixed-threshold method, (c) the significantly overlapping regions for low-values of reference pixel P_G using the noise-limited method, and (d) the insignificant overlapping regions for high-values of reference pixel P_G using the noise-limited method, that will lead to misclassification of shadow as scene-change for all but the lightest of shadows.

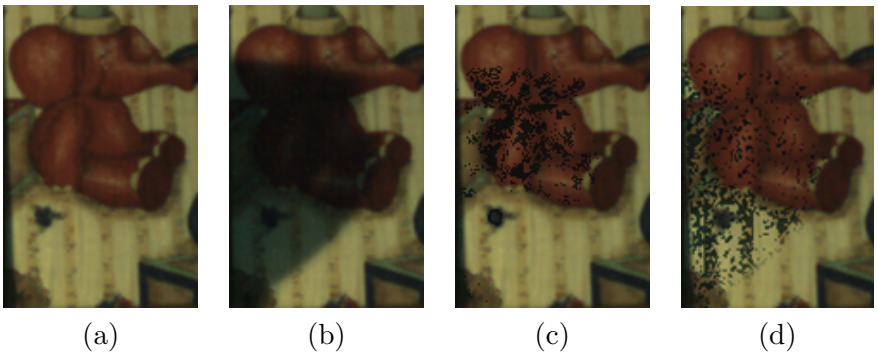


Figure 5.9 Example of differences in performance of the fixed-threshold and noise-limited shadow removal algorithms: (a) original image without shadow, (b) image with cast shadow, (c) shadowed image processed with fixed-threshold algorithm, and (d) shadowed image processed with the noise-limited algorithm.

and areas of shadow were masked by hand so that the performance of the algorithms could be quantified. Two performance criteria were calculated: percentage correct for object detection (the number of genuine scene-change pixels that were correctly classified) and percentage correct for shadow detection (the number of shadowed pixels that were correctly classified). The noise-limited algorithm performed better in 6 of the 8 experiments for detecting scene change, with an overall average of 93.20% correct detection, compared to 91.33% for the fixed-threshold method. Conversely, only 1 out of 8 tests using the noise-limited method demonstrated an improvement in the classification of shadows, with an overall detection rate of 77.65% for the noise-limited method and 83.70% for the fixed-threshold method. The reduction in shadow detection performance can be attributed to the quantization of the true scene-change curves to one of the limited number of curves available from the GMB calibration chart.

In summary, a noise-limited scene-change algorithm has been developed that is free from any empirical internal threshold. The statistically generated scene-change regions show that using a simple fixed-threshold boundary is sub-optimal, as the size and shape of the scene-change regions vary with the specific colour and intensity of the reference pixel. However, experiments using noise-limited regions exposed a deficiency in the limited number of measured shadow lines used in the new method. For the fixed-threshold algorithm, small changes in the angle and length of the line caused by the nearest-neighbour line-selection were hidden by the relatively large fixed radius of the threshold. The noise-limited regions often closely followed the shadow line, especially for higher intensity pixels, that resulted in an increased sensitivity to errors in the approximation of the shadow line by the nearest-neighbour selection method.

Overall, the performances of both the fixed-threshold and noise-limited algorithms are similar, which validates the noise-limited method as a potential alternative to the fixed-threshold method, even with the limitations described above.

Table 5.1 Results of the noise-limited scene-change detection algorithm, showing the percentage areas of scene-change correctly classified as objects or shadows.

Scene	**% Object (noise-limited method)**	**% Object (threshold method)**	**% Shadow (noise-limited method)**	**% Shadow (threshold method)**
Set 'Castle'	93.74	95.86	72.80	82.18
Set 'Hawaii'	96.16	95.25	85.36	72.92
Set 'Castle 2'	95.66	89.10	85.74	95.02
Set 'Hand'	99.79	99.66	86.42	90.6
Set 'Stitch'	95.90	94.43	76.10	84.84
Set 'Stitch 2'	84.67	76.19	78.65	89.91
Set 'GMB'	89.96	89.87	69.08	76.10
Set 'Mattress'	89.73	90.24	67.04	78.01
Average	**93.20**	**91.33**	**77.65**	**83.70**

5.6 INTERPOLATED CALIBRATION DATA

To overcome the limitations of the restricted number of scene-change lines available to the noise-limited scene-change detection algorithm when the GMB chart is used for calibration, an expanded set of lines was created to increase the number of lines available for nearest-neighbour selection.

Delaunay triangulation (Delaunay 1934) was used to generate natural sets of triangles from the $n = 1$ points from the calibration data (shown in Figure 5.1). Interpolated calibration lines were positioned at the centroid of each triangle. The angle and length of each interpolated calibration line was calculated by taking the average of the angles and lengths of the lines located at the triangles vertices. This process was repeated using the expanded calibration set from the first iteration, and an extra 112 calibration lines were generated. This was empirically determined to be a sufficient number to provide satisfactory shadow detection results in our experiments (i.e., a substantial reduction in clusters of misclassified pixels, such as those shown in Figure 5.9d). The full set of 131 lines is shown in Figure 5.10. For each pixel in a reference image the calibration line of the $n = 1$ point closest to the pixel in x/y space of the expanded calibration set (by Euclidean distance measurement) was used to approximate the true scene-change line for the pixel. Though the lines appear cluttered in Figure 5.10, each calibration line is used in isolation from any other line as each RGB triplet has a unique and independent calibration line.

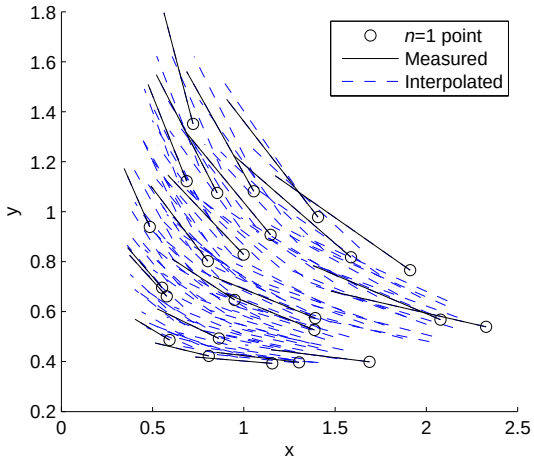


Figure 5.10 The full calibration line data set, generated from the 18 original calibration lines using Delaunay triangulation.

5.7 EXPERIMENTAL RESULTS USING INTERPOLATED CALIBRATION DATA

The 8 image sets listed in Table 5.1 were processed by the noise-limited scene-change detection method using the expanded calibration set. To allow a comparison with the fixed-threshold method which post-filters its scene-change mask, a filter was applied to the resultant scene-change masks of the noise-limited method that assumes small areas of detected scene-change or shadow are erroneous (areas less than that of a 3x3 pixel block). The images for all 8 test sets, their scene-change masks, and reconstructed shadow-free images (filtered) are given in Figures 5.11–5.26.

The experimental results for the fixed-threshold and noise-limited methods are given in Tables 5.2 & 5.3. The performances for scene-change detection is relatively consistent across all methods, with averages ranging from 91.33% to 93.22%, increasing to 95.17% when the size filtering is applied. There are substantial performance differences between the methods for shadow detection however, which ranges from 77.65% to 87.13%, increasing to 95.61% with the filtering. The noise-limited methods used a standard confidence interval value of 2σ for the scene-change detection threshold, which defines a maximum theoretical shadow-detection performance of 95.45% (prior to filtering). The best scene-change and shadow-change detection method is the noise-limited algorithm with interpolated calibration lines.

Table 5.2 Results of the scene-change detection algorithm, showing the percentage areas of image correctly classified as scene-change.

Scene	**Threshold method**	**Noise-bound (19-point calibration)**	**Noise-bound (131-point calibration)**	**Noise-bound (131-point calibration) filtered***
'Castle'	95.86%	93.74%	93.73%	(96.72%)
'Hawaii'	95.25%	96.16%	96.09%	(97.99%)
'Castle 2'	89.10%	95.66%	95.74%	(97.77%)
'Hand'	99.66%	99.79%	99.82%	(99.91%)
'Stitch'	94.43%	95.90%	96.23%	(98.22%)
'Stitch 2'	76.19%	84.67%	84.33%	(86.43%)
'GMB'	89.87%	89.96%	89.86%	(91.22%)
'Mattress'	90.24%	89.73%	89.94%	(93.06%)
Average	**91.33**	**93.20**	**93.22**	**(95.17)**

*areas ≤ 9 pixels in the expanded-calibration set images that were detected as shadows removed, or holes < 9 pixels in objects filled.

Table 5.3 Results of the scene-change detection algorithm, showing the percentage areas of image correctly classified as shadow.

Scene	**Threshold method**	**Noise-bound (19-point calibration)**	**Noise-bound (131-point calibration)**	**Noise-bound (131-point calibration) filtered***
'Castle'	82.18%	72.80%	82.24%	(92.14%)
'Hawaii'	72.92%	85.36%	91.34%	(98.87%)
'Castle 2'	95.02%	85.74%	90.17%	(98.30%)
'Hand'	90.65%	86.42%	90.81%	(97.63%)
'Stitch'	84.84%	76.10%	87.08%	(96.25%)
'Stitch 2'	89.91%	78.65%	87.43%	(95.66%)
'GMB'	76.10%	69.08%	85.38%	(93.79%)
'Mattress'	78.01%	67.04%	82.57%	(92.23%)
Average	**83.70%**	**77.65%**	**87.13%**	**(95.61%)**

*areas ≤ 9 pixels in the expanded-calibration set images that were detected as shadows removed, or holes < 9 pixels in objects filled.

(a)

(b)

(c)

Figure 5.11 Images from image set 'Castle'. (a) the background image - a large printed photograph on a desk; (b) the comparison image with the added objects (pencil sharpener and cup) and their respective cast shadows; (c) a difference mask showing any change in pixel value greater than 2 standard deviations of the reference pixel noise value, highlighting changes between the two images.

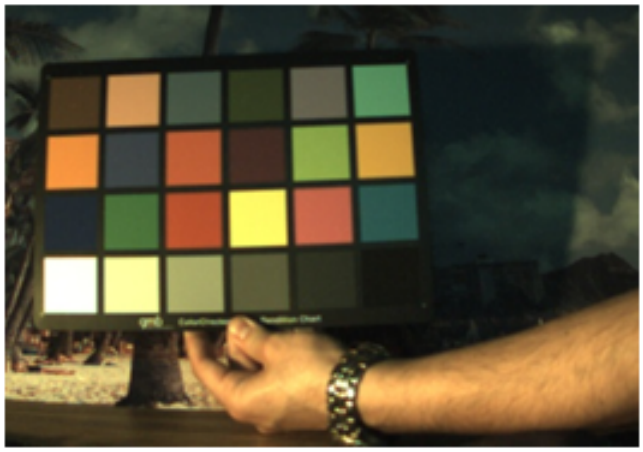
(a)

(b)

(c)

Figure 5.12 The shadow-detection masks for image set 'Castle'. (a) the result of the noise-limited scene-change algorithm, (b) a filtered mask of (a) with arbitrary-sized small holes filled and small objects removed (< 9 pixels), demonstrating an example final mask that could be used for further processing in an object detection process, (c) reconstructed shadow-free image obtained by combining the images from Figures 5.11a & 5.11b using the mask in 5.11c.

(a)



(b)

(c)

Figure 5.13 Images from image set 'Hawaii'. (a) is the background image - a large printed photograph on a desk; (b) is the comparison image with the added objects (arm and colour chart) and its associated cast shadow; (c) a difference mask showing changes in pixel value greater than 2 standard deviations of the reference pixel noise value, highlighting changes between the two images.

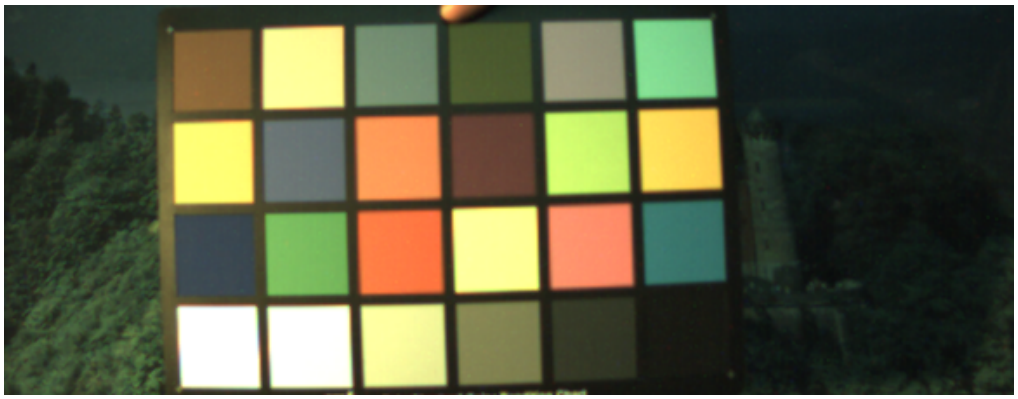
(a)

(b)

(c)

Figure 5.14 The shadow-detection masks for image set 'Hawaii'. (a) the result of the noise-limited scene-change algorithm, (b) a filtered mask of (a) with arbitrary-sized small holes filled and small objects removed (< 9 pixels), demonstrating an example final mask that could be used for further processing in an object detection process, (c) reconstructed shadow-free image obtained by combining the images from Figures 5.13a & 5.13b using the mask in 5.13c.

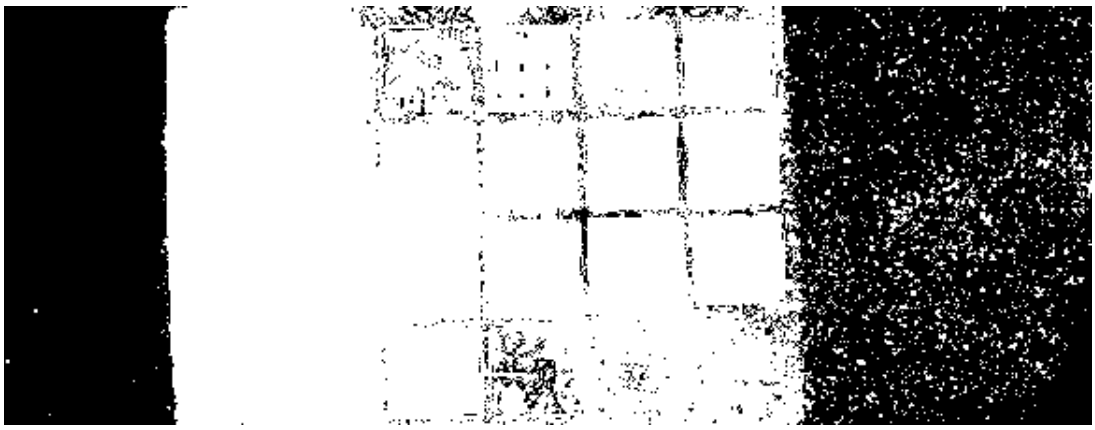
(a)



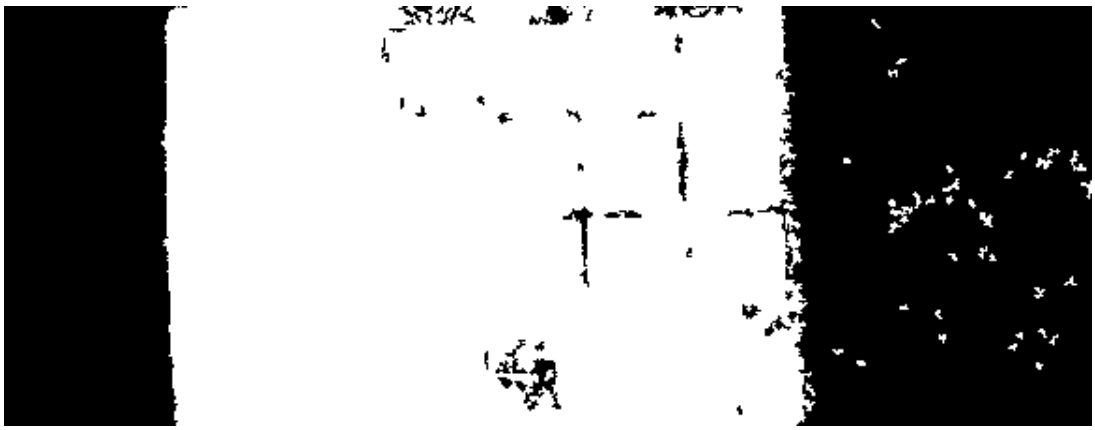
(b)

(c)

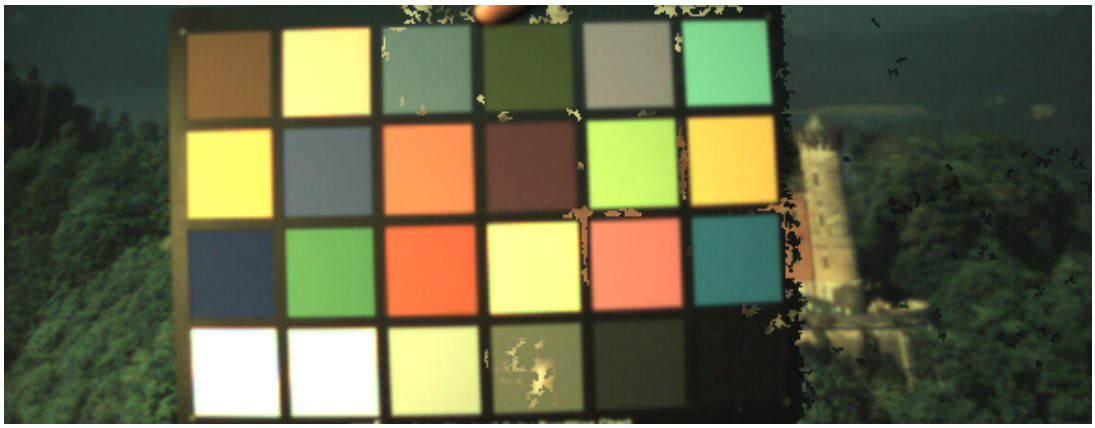
Figure 5.15 Images from image set 'Castle 2'. (a) is the background image - a large printed photograph on a desk; (b) is the comparison image with the added objects (GMB colour chart) and its associated cast shadow; (c) a difference mask showing changes in pixel value greater than 2 standard deviations of the reference pixel noise value, highlighting changes between the two images.



(a)



(b)



(c)

Figure 5.16 The shadow-detection masks for image set 'Castle 2'. (a) the result of the noise-limited scene-change algorithm, (b) a filtered mask of (a) with arbitrary-sized small holes filled and small objects removed (< 9 pixels), demonstrating an example final mask that could be used for further processing in an object detection process, (c) reconstructed shadow-free image obtained by combining the images from Figures 5.15a & 5.15b using the mask in 5.15c.

(a)

(b)

(c)

Figure 5.17 Images from image set 'Hand'. (a) is the background image - a large printed photograph on a desk; (b) is the comparison image with the added object (arm) and its associated cast shadow; (c) a difference mask showing changes in pixel value greater than 2 standard deviations of the reference pixel noise value, highlighting changes between the two images.

(a)

(b)

(c)

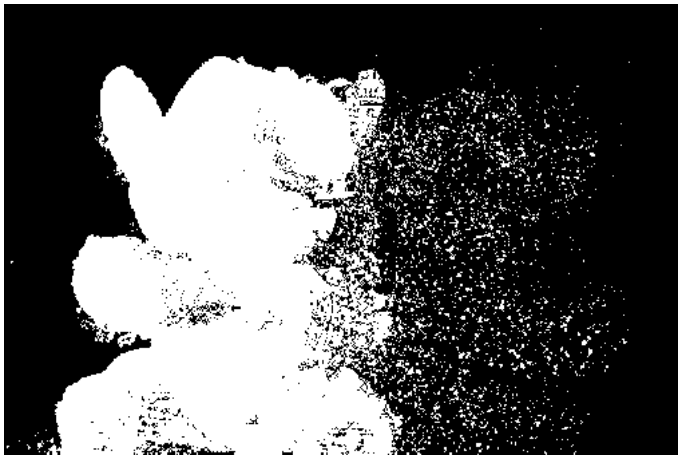
Figure 5.18 The shadow-detection masks for image set 'Hand'. (a) the result of the noise-limited scene-change algorithm, (b) a filtered mask of (a) with arbitrary-sized small holes filled and small objects removed (< 9 pixels), demonstrating an example final mask that could be used for further processing in an object detection process, (c) reconstructed shadow-free image obtained by combining the images from Figures 5.17a & 5.17b using the mask in 5.17c.

(a)

(b)

(c)

Figure 5.19 Images from image set 'Stitch'. (a) is the background image - a large printed photograph on a desk; (b) is the comparison image with the added objects (coloured stuffed toy) and its associated cast shadow; (c) a difference mask showing changes in pixel value greater than 2 standard deviations of the reference pixel noise value, highlighting changes between the two images.



(a)

(b)

(c)

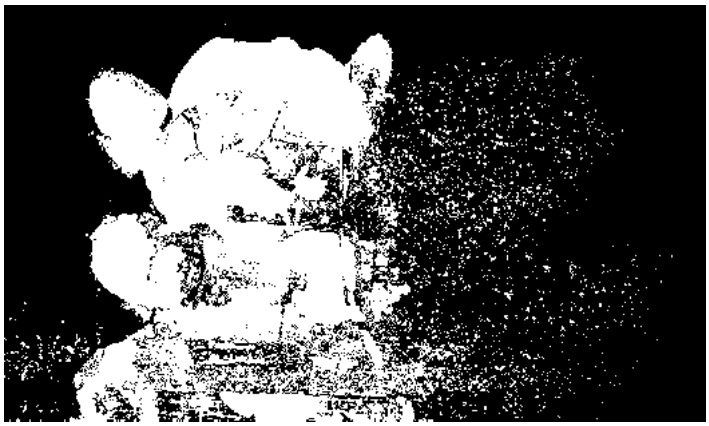
Figure 5.20 The shadow-detection masks for image set 'Stitch'. (a) the result of the noise-limited scene-change algorithm, (b) a filtered mask of (a) with arbitrary-sized small holes filled and small objects removed (< 9 pixels), demonstrating an example final mask that could be used for further processing in an object detection process, (c) reconstructed shadow-free image obtained by combining the images from Figures 5.19a & 5.19b using the mask in 5.19c.

(a)

(b)

(c)

Figure 5.21 Images from image set 'Stitch 2'. (a) is the background image - a large printed photograph on a desk; (b) is the comparison image with the added objects (coloured stuffed toy) and its associated cast shadow; (c) a difference mask showing changes in pixel value greater than 2 standard deviations of the reference pixel noise value, highlighting changes between the two images.



(a)

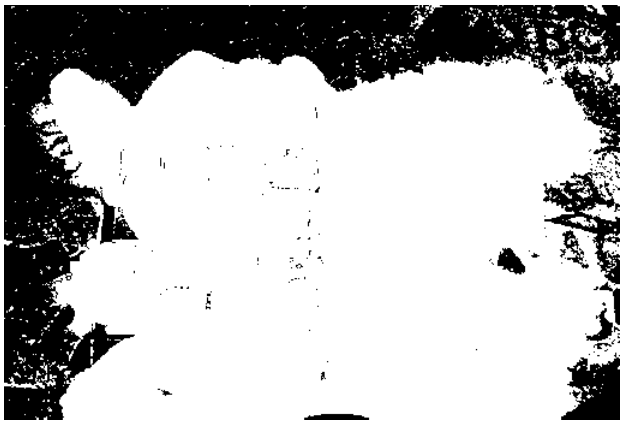
(b)

(c)

Figure 5.22 The shadow-detection masks for image set 'Stitch 2'. (a) the result of the noise-limited scene-change algorithm, (b) a filtered mask of (a) with arbitrary-sized small holes filled and small objects removed (< 9 pixels), demonstrating an example final mask that could be used for further processing in an object detection process, (c) reconstructed shadow-free image obtained by combining the images from Figures 5.21a & 5.21b using the mask in 5.21c.

(a)

(b)



(c)

Figure 5.23 Images from image set 'GMB'. (a) is the background image - a large coloured and patterned background with a GMB chart; (b) is the comparison image with the added object (coloured stuffed toy) and its associated cast shadow; (c) a difference mask showing changes in pixel value greater than 2 standard deviations of the reference pixel noise value, highlighting changes between the two images.

(a)

(b)

(c)

Figure 5.24 The shadow-detection masks for image set 'GMB'. (a) the result of the noise-limited scene-change algorithm, (b) a filtered mask of (a) with arbitrary-sized small holes filled and small objects removed (< 9 pixels), demonstrating an example final mask that could be used for further processing in an object detection process, (c) reconstructed shadow-free image obtained by combining the images from Figures 5.23a & 5.23b using the mask in 5.23c.

(a)

(b)

(c)

Figure 5.25 Images from image set 'Mattress'. (a) is the background image - a large coloured and patterned background; (b) is the comparison image with the added objects (hand and coloured stuffed toy) and their associated cast shadows; (c) a difference mask showing changes in pixel value greater than 2 standard deviations of the reference pixel noise value, highlighting changes between the two images.

(a)

(b)

(c)

Figure 5.26 The shadow-detection masks for image set 'Mattress'. (a) the result of the noise-limited scene-change algorithm, (b) a filtered mask of (a) with arbitrary-sized small holes filled and small objects removed (< 9 pixels), demonstrating an example final mask that could be used for further processing in an object detection process, (c) reconstructed shadow-free image obtained by combining the images from Figures 5.25a & 5.25b using the mask in 5.25c.

5.8 DISCUSSION

The theoretical maximum performance of 95.45% for shadow detection arises from the statistical variations in image noise, for a value of 2σ. However, object detection performance is dependent upon the colouring of objects, illumination, and backgrounds. There will be instances where the colours of objects will be similar to the shadowed background, which will cause a reduction in performance of scene-change detection. For example, in set 'Stitch 2' (Figure 5.21) the blue colouring of the added object appears to be the same colour as the white clouds and sky when in shadow, and is detected as such, resulting in an object integrity of just 84%. Conversely, conditions can arise when there are significant colour differences resulting in much higher performance such as in set 'Hand' (Figure 5.17) which saw a 99.82% correct detection of scene change. This is not a surprising result as the hand is highly reflective and has a substantial colour-difference from the background. Another interesting observation in the results of set 'Hand' (Figure 5.18), is that there are thin lines of apparent scene-change extending from the tips of several fingers, as well as an apparent growth in the webbing of the hand between the thumb and forefinger. Closer inspection of the source images show that there is in fact a change in colour underneath those areas due to obscuring of the ambient light from the hand, colouring the background with a red tinge. This demonstrates a limitation of the experimental conditions where the approximation of ambient light by a fluorescent tube fails. However, it does illustrate the effectiveness of the noise-limited method for detecting very small changes in illumination.

Shadow boundaries are well detected and removed using the noise-limited method of shadow removal, as illustrated in the experimental results. Unlike many of the existing methods of shadow removal, this method is purely colour-based and has no reliance on the detection of edges.

The shadow-detection performances of the algorithms vary considerably more than the scene-change detection performances, with the initial 19-point calibration for the noise-limited method performing worse than the fixed-threshold method. The reasoning for this, provided in Section 5.5, is because of the limited number of calibration lines derived from the original GMB chart data. The expanded calibration set increases the performance to beyond that of the fixed-threshold method with a 26% relative improvement in shadow detection, reducing the shadow-detection error from 16.3% to 12.8%.

It is not surprising that there is very little improvement in the object detection performances when the expanded calibration set is used, as the error in object detection arises from the incorrect classification of genuine scene-change as shadows due to metamerism. Given a constant illumination, the likelihood of an object being categorised as shadow is solely dependent on the relative colour of an object compared to the background, and so a significant reduction in metamerism was not expected from

using an increased number of calibration lines.

It is possible that the fixed-threshold method could provide an improvement in performance if the expanded calibration set was used. However, the aim of the work described in this chapter was to develop a robust and viable method whose performance was described by the objectives; specifically, improvements were required to the original fixed-threshold method to remove the necessity for arbitrary thresholds. This was achieved by integrating knowledge of noise into the algorithm, and the same initial calibration data that was available to the fixed-threshold method (the original 19 calibration lines) was used. The use of interpolated calibration points in the noise-limited method does not affect the performance of the fixed-threshold algorithm described in Chapter 4.

The scene-change detection performances for the fixed-threshold method are set by the parameters inside the noise filter and in the value of t. The value of t was empirically determined by observation of the scene-change masks, which resulted in a very close match to the performances of scene-change detection by the noise-limited method, where performance is characterised by the scene-change confidence interval (2σ) used. The reduction in performances of the noise-limited method from the theoretical maximum shadow-detection of 94.95% can in part be attributed to the approximation of the true calibration line by its nearest-neighbour. A slight variation in angle or length could result in misclassifications of shadowed areas.

The application of simple filtering for small blobs (areas less than that of a 3-by-3 block) improved the scene-change detection performance, at the expense of detecting scene-changes smaller than the size of the removed blobs. This leveraging of non-colour based information increases the performance of the scene-change detection (to 95.17%) and shadow detection (to 95.61%) for the image sets tested.

There are three approaches which could further improve the performance of the noise-limited scene-change method: extended interpolation of the calibration set; calibration with a larger number of colour panels; and modelling of the calibration lines which could be used to accurately generate the true shadow line for any pixel in the reference image. These could form the basis for future work.

The noise-limited method for scene-change detection has been applied to outdoor images, but the limited dynamic range of the two cameras trialled could not capture sufficient shadow information without saturation of pixel data for non-shadowed areas. High dynamic-range cameras would provide a solution to the saturation problem.

All objectives described in Section 4.1 have been met by the new scene-change detection algorithm. Table 5.8 compares the new algorithm and Finlayson et al.'s algorithm with the objectives. The significant difference between the methods is not in their performance, but in their properties: Finlayson et al.'s method will simply not work without the objective conditions in Table 5.8 being met, hence there is no benefit

in providing a direct comparison of performance between them.

Table 5.4 A comparison between the new scene-change detection algorithm and Finlayson et al.'s shadow removal algorithm, with regard to the objectives outlined in Section 4.1.

Objective	**Finlayson et al.'s method**	**Noise-bound method**
Independent of object shape	×	✓
Independent of additional shadows cast on the scene	✓	✓
Operates on both indoor and outdoor environments	×	✓
Does not require manual interaction	×	✓
Free of arbitrary internal thresholds	×	✓
Does not require significant training	×	✓

5.9 CONCLUSION

A novel method of scene-change detection has been presented that is independent of object shape or any other *a priori* scene information, and only requires the setting of one parameter (the confidence interval for noise separation). It is the first that is based upon two arbitrarily coloured illuminants, with subsequent integration of camera noise to remove dependencies on empirical thresholds. This is a significant advance in image preparation for machine processing, as there are few methods in any branch of image processing that are independent of arbitrary or empirical thresholds.

With calibrated information on camera noise and a confidence interval of 2σ (94.45%), an average object detection of 93% percent and shadow detection of 87% was achieved on experiments with 8 image sets (*a posteriori* simple filtering raised these figures to 95%). The scene-change detection method is based on colour change only, and therefore requires careful calibration of illumination change for the direct plus ambient to ambient illumination only. For application to outdoor images a camera with a high dynamic range is required, as the intensity of the sun can be several orders of magnitude greater than the intensity of cast shadows, resulting in loss of colour information with the majority of currently available cameras.

Part IV

Discussion and conclusions

Chapter 6

CONCLUSION AND FUTURE WORK

6.1 CONCLUSIONS AND ACHIEVEMENTS

This thesis investigates the composition of noise in standard industrial digital cameras, develops methods for measuring camera noise, and integrates noise characterstics into a new scene-change detection algorithm. Each is discussed in the following noise analysis and scene-change sections.

The results of this work provide a step forward in the area of vision processing. The understanding and measurement of noise is critical for developing highly robust vision-processing algorithms, as the signal-to-noise ratios are fundamentally limiting what information that can be extracted from any data. The output from the scene-change detection methods developed in this thesis can be used for higher level processing of image content.

6.1.1 Noise analysis

New models for commercial CCD- and CMOS-based cameras have been developed. By analysing the image capture process, new comprehensive noise models have been created that incorporate all known sources of noise for both CCD and CMOS cameras. Subsequently, a method to validate the CCD model by analysis of output images alone has been developed. The CMOS model is more complex and is difficult to validate solely from output images due to the extra circuitry surrounding each pixel.

Methods for measuring spatial, temporal, and combined noise have been developed that are applicable to any imaging device. Using these methods an investigation into the effect of environmental temperature on base-line camera noise has been completed, with results showing surprising variation in noise with temperature. All objectives outlined in Section 3.1.4 have been achieved:

- Comprehensive models for commonly available CCD and CMOS cameras have been developed;

- A method of validating the CCD model using the camera's output images only has been derived and tested experimentally, with results demonstrating a good fit between the measured and modelled data;
- Methods for measuring the combined noise for any camera, for use in subsequent image processing applications, have been developed; and
- An investigation into the effects of environmental temperature on camera noise has been completed.

The outputs from the noise analysis section of work have provided interesting observations. Firstly, quantization is frequently acknowledged by researchers but normally unaccounted for during analysis of noise in image data; there are no reported experiments on real image data that have incorporated the effects of quantization. The effect of including quantization noise into each component in the validation of the CCD noise model (Section 3.4) is significant, as the model fit is poor without it. Secondly, the CCD cameras' noise-response to temperature variation is not predictable (Section 3.7). It is assumed that the supporting electronics required for driving the CCD element is responsible for this unpredictability, as measurements show that the CMOS sensor responds in a more predictable manner.

6.1.2 Scene-change detection

Part III describes the development of a new scene-change detection model, the first of its kind that is based upon two arbitrarily coloured illuminants, with subsequent integration of camera noise to remove dependencies upon empirical thresholds. This in itself is an achievement, as there are few methods in any branch of image processing that are independent of arbitrary or empirical thresholds.

The development of the scene-change detection method was a two-part process. The initial dual-illuminant model was developed to meet the majority of objectives in 4.1, and the experimental results validated the dual illuminant approach. The integration of the noise research from Part II followed, which highlighted deficiencies in the original number of calibration curves due to the increased sensitivity of the algorithm. Once this was overcome by interpolation of the calibration data, the noise-integrated method completed the list of objectives by removing the necessity for empirical thresholds, and also showed improved results compared to the initial method. The noise-limited scene-change detection algorithm correctly classifies 93% of scene-change pixels and 87% of cast-shadow pixels using colour change alone. When simple post-analysis size-filtering is applied these figures increase to 95%. Therefore, all objectives outlined in Section 4.1 have been met. In summary:

- A method of scene-change detection was developed using a dual-illuminant model for illumination change, that classified areas of image changes as scene-change or shadow;
- The scene-change detection method is independent of object shape, suitable for operating in both indoor and outdoor environments, and does not require manual interaction or significant training;
- Integration of noise knowledge into the method removed the need for empirical thresholds.

One unanticipated side-effect of the developed scene-change detection method is that it can be used to remove all cast shadows from stationary outdoor scenes. By analysing a series of images captured over the course of a day, integration of shadow-free regions of a scene will effectively eliminate cast shadows and will eventually result in an intrinsic image of the scene.

6.2 SUGGESTIONS FOR FUTURE WORK

There are several avenues for research that can follow the work presented in this thesis that could potentially improve the results and methods presented. Future work is split into camera-noise and scene-change detection sections.

6.2.1 Further camera-noise research

Further research into quantization noise effects in images is warranted, as its effect is neither simple nor linear. Ideally, a parameterised model of quantization noise for increasing pixel-data variation would provide more accurate quantization noise error compared with using measured data, as was done in this work.

The analysis of the impact of environmental temperature on camera noise provides a measure of the base-line noise response for several cameras. However, image noise increases with increased image-sensor irradiation, and an investigation into the effects of environmental temperature on illumination-dependent noise (PRNU and photon shot-noise) would reveal new insights into the dynamics of these cameras.

The development of a stand-alone program that could capture and analyse image noise would result in a useful tool. At present the capture of images is separate from the analysis, which uses MATLAB code and manually programmed calibration points on a GMB colour chart. Such a tool would allow for a fast and easy method for camera calibration that would be useful to both application developers and users.

The integration of temperature into the camera noise models would significantly broaden the application of the models to environments susceptible to temperature

changes. For example, an application could use a thermocouple or other simple temperature sensor as an input to the model to provide an accurate estimation of the noise present at a specific pixel within a particular image. This would also allow the stability of noise characteristics over the life of the camera to be analysed.

Validation of the CMOS model is difficult due to the potential manufacturer-specific design of the active pixel elements. However, further research could uncover new methods for measurement of noise components by developing iterative techniques to predicting the likely magnitudes of the noise components.

Other image processing applications could benefit from integration of noise knowledge into their algorithms. Chapter 5 of this thesis has demonstrated that the benefits can be significant as empirical or arbitrary thresholds can be removed, as well as other broad noise filtering that an algorithm may employ to overcome the detrimental effects of image noise. The integration of noise knowledge may also increase performance and robustness of the algorithm, as demonstrated with the new scene-change removal algorithm discussed in the next section.

6.2.2 Further scene-change detection research

As a camera's SNR improves, the amount of allowed error in calibration of the illuminant colours for scene-change detection is reduced, as the scene-change regions reduce in size making them more sensitive to error. Small changes in skylight colour can occur over the course of a day, with potentially large changes at sunrise and sunset. In these conditions a dynamic method of calibration would be required to track changes in illumination colour to ensure the dual-illumination model was accurate.

Similarly, the error due to the selection of calibration curves by nearest-neighbour interpolation will increase as SNR improves. The changes in calibration line angle and length could be modelled for a particular pair of illuminants, which could then be used to generate accurate scene-change regions for any pixel colour. This would require a new approach for generating the regions: the MATLAB code used in this thesis to generate the 131 scene-change regions in Chapter 5 took over 20 hours of computational time on a 2.33 GHz Core 2 Duo processor.

A natural extension to the dual-illuminant scene-change detection method is to apply the process to outdoor images taken with a high dynamic range camera. The standard 8-bit intensity per colour channel used in the cameras tested is not of sufficient dynamic range to capture shadow detail while retaining detail of objects in direct sunlight. Hence all experiments were performed on images captured in an environment with controlled lighting.

As the noise-bound method is dependent upon image noise statistics, it inherits the noises dependence upon temperature. An extension of the noise-bound method to

dynamically update its noise model with temperature would increase the robustness of the method in applications where environmental temperature is expected to change.

An analysis of the effect of varying of the single parameter σ using a variety of cameras and illumination conditions and scenes would provide an objective measure for optimising a value for σ, for a given environment.

An interesting application of the scene-change detection method is that it could be used to dynamically update a background image over the course of a day. In principle, this would provide a background image that is free from any cast shadow, resulting in an intrinsic image of the scene.

One of the constraints for the scene-change detection method is that the camera must be stationary, which is an acceptable condition in many industrial image processing tasks (e.g., traffic monitoring or process line monitoring). It would be interesting to see a research project attempt to incorporate camera motion into the scene-change detection method. A direct pixel-to-pixel mapping would be difficult unless accurate camera movement is provided to the algorithm, but an approach that allowed for slight misalignment of the background and comparison images may result in an acceptable solution, with little reduction in accuracy.

REFERENCES

Anaxagoras, T., Guerrini, N., Turchetta, R. & Allinson, N. M. (2005). High dynamic range sensor active pixel sensor, *Proc. of the First International Conference on Sensing Technology* pp. 448–453.

Arbel, E. & Hel-Or, H. (2007). Texture-preserving shadow removal in color images containing curved surfaces, *IEEE Conference on Computer Vision and Pattern Recognition* pp. 1–8.

Baher, H. (1990). *Analog and digital signal processing*, John Wiley & Sons Ltd.

Barnard, K., Cardei, V. & Funt, B. (2002). A comparison of computational color constancy algorithms. i: Methodology and experiments with synthesized data, *IEEE Trans. Image Processing* **11**(9): 972–984.

Bhattacharyya, G. K. & Johnson, R. A. (1977). *Statistical concepts and methods*, Wiley series in probability and methematical statistics, John Wiley & Sons, Inc.

Billmeyer, Fred W., J. & Saltzman, M. (1966). *Principles of Color Technology*, John Wiley & Sons.

Blanc, N. (2001). CCD versus CMOS - has CCD imaging come to an end?, *Photogrammetric Week '01'* .

Boie, R. & Cox, I. (1992). An analysis of camera noise, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **14**(6): 671–674.

Brouk, L. & Nemirovsky, Y. (2004). CMOS SOI image sensor, *Proc. of the IEEE International Conference on Electronics, Circuits and Systems* **11**: 156–159.

Catrysse, P. B., Wang, M. & El Gamal, A. (1999). Comparative analysis of color architectures for image sensors, *Proc. of SPIE Sensors, Cameras, and Applications for Digital Photography* **3650**: 26–35.

Chang, C.-J., Hu, W.-F., Hsieh, J.-W. & Chen, Y.-S. (2002). Shadow elimination for effective moving object detection with gaussian models, *Proc. 16th International Conference on Pattern Recognition* **2**: 540–543.

Chen, T., Catrysse, P. B., El Gamal, A. & Wandell, B. A. (2000). How small should pixel size be?, *Proc. of SPIE Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications* **3965**: 451–459.

Chuang, Y.-Y., Goldman, D., B., Curless, B., Salesin, D. H. & Szeliski, R. (2003). Shadow matting and compositing, *ACM Transactions on Graphics* pp. 494–500.

CIE (2009). `http://www.cie.co.at`.

Costantini, R. & Süsstrunk, S. (2004). Virtual sensor design, *Proc. of SPIE Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications V.* **5301**: 408–419.

Crow, F. C. (1977). Shadow algorithms for computer graphics, *Proc. 4th Annual Conference on Computer Graphics and Interactive Techniques* pp. 242–248.

Dalsa (2006). CCD technology primer, dalsa inc, available: `http://www.dalsa.com`.

Delaunay, B. (1934). Sur la sphere vide, *Izvestia Akademii Nauk SSSR, Otdelenie Matematicheskikh i Estestvennykh Nauk* **7**: 793–800.

Ditchburn, R. W. (1976). *Scattered and filtered solar UV measurements*, third edn, Academic Press, London.

Diwa (2007). Row and column noise. `http://www.diwa-labs.com`.

Drew, M. S., Chen, C., Hordley, S. D. & Finlayson, G. D. (2002). Sensor transforms for invariant image enhancement, *10th Color Imaging Conference* pp. 325–329.

Eisemann, E. & Durand, F. (2004). Flash photography enhancement via intrinsic lighting, *ACM Trans. Graphics* pp. 673–678.

El Gamal, A. & Eltoukhy, H. (2005). CMOS image sensors, *IEEE Circuits and Devices Magazine* **21**(3): 6–20.

El Gamal, A., Fowler, B. A., Min, H. & Liu, X. (1998a). Modeling and estimation of FPN components in CMOS image sensors, *Solid State Sensor Arrays: Development and Applications II* **3301**: 168–177.

El Gamal, A., Fowler, B., Min, H. & Liu, X. (1998b). Modeling and estimation of FPN components in CMOS image sensors, *Proc. of SPIE* .

Erbou, S. G., Sorensen, H. B. D. & Stage, B. (2005). Detection of cast shadows in surveillance applications, *Den 14. Danske Konference i Monstergenkendelse og Billedanalyse* .

Farrell, J. E., Xiao, F., Catrysse, P. B. & Wandell, B. A. (2004). A simulation tool for evaluating digital camera image quality, *Image Quality and System Performance* **5294**: 124–131.

Finlayson, G. D., Hordley, S. D. & Drew, M. S. (2002a). Removing shadows from images, *Proc. 7th European Conference on Computer Vision-Part IV* pp. 823–836.

Finlayson, G., Drew, M. & Lu, C. (2004). Intrinsic images by entropy minimization, *Proc. 8th European Conference on Computer Vision* pp. 582–595.

Finlayson, G., Hordley, S. & Drew, M. (2002b). Removing shadows from images using retinex, *Proc. 10th Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications* .

Finlayson, G., Hordley, S., Lu, C. & Drew, M. (2006). On the removal of shadows from images, *IEEE Trans. Pattern Analysis and Machine Intelligence* **28**(1): 59–68.

Flory, R. (1985). Image acquisition technology, *Proc. of the IEEE* **73**(4): 613–637.

Fossum, E. (1997). CMOS image sensors: electronic camera-on-a-chip, *IEEE Trans. Electron Devices* **44**(10): 1689–1698.

Fowler, B., El Gamal, A. & Yang, D. (1994). A CMOS area image sensor with pixel-level A/D conversion, *1994 IEEE International Solid-State Circuits Conference* pp. 226–227.

Funka-Lea, G. & Bajcsy, R. (1995). Combining color and geometry for the active, visual recognition of shadows, *Proc. IEEE 5th International Conference on Computer Vision* pp. 203–209.

Funt, B., Barnard, K. & Martin, L. (1998). Is machine colour constancy good enough?, *5th European Conference on Computer Vision* pp. 445–459.

Funt, B. & Finlayson, G. (1995). Color constant color indexing, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **17**(5): 522–529.

Gerritsen, F. (1975). *Theory and Practice of Color*, Studio Vista.

Gershon, R., Jepson, A. D. & Tsotsos, J. K. (1987). From [R, G, B] to surface reflectance: Computing color constant descriptors in images, *Proc. International Joint Conference on Artificial Intelligence* pp. 755–758.

Gregory, R. L. (1988). *Eye and brain*, fifth edn, Oxford University Press, Oxford.

Gunturk, B. K., Glotzbach, J., Altunbasak, Y., Schafer, R. W. & Mersereau, R. M. (2005). Demosaicking: Color filter array interpolation (exploring the imaging process and the correlations among three color planes in single-chip digital cameras), *IEEE Signal Processing* **22**(1): 44–54.

Healey, G. & Kondepudy, R. (1994). Radiometric CCD camera calibration and noise estimation, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **16**(3): 267–276.

Henderson, S. T. (1977). *Daylight and its spectrum*, second edn, Adam Hilger Ltd, Bristol.

Hernandez-Andres, J., Lee, R. L. & Romero, J. (1999). Calculating correlated color temperatures across the entire gamut of daylight and skylight chromaticities, *Applied Optics* **38**(27): 5703–5709.

Horowitz, P. & Hill, W. (1989). *The Art of Electronics*, second edn, Melbourne.

Horprasert, T., Harwood, D. & Davis, L. S. (1999). A statistical approach for real-time robust background subtraction and shadow detection, *Proc. IEEE ICCV* pp. 1–19.

Huang, W.-C. & Wu, C.-H. (1998). Adaptive color image processing and recognition for varying backgrounds and illumination conditions, *IEEE Trans. Industrial Electronics* **45**(2): 351–357.

Hynecek, J. (1990). Spectral analysis of reset noise observed in CCD charge-detection circuits, *IEEE Trans. on Electron Devices* **37**(3): 640–647. 0018-9383.

Ibrahim, M. M. & Rajagopal, A. (2007). Shadow detection in images. US Patent No. 2007/0110309 A1.

Iqbal, M. (1983). *An Introduction to Solar Radiation*, Academic Press Canada, Ontario.

Irie, K., McKinnon, A. E., Unsworth, K. & Woodhead, I. M. (2006). A comparison of noise in CCD and CMOS image sensors, *Proc. Image and Vision Computing New Zealand* pp. 43–48.

Irie, K., McKinnon, A. E., Unsworth, K. & Woodhead, I. M. (2007a). Measuring digital camera image noise, *Proc. 2nd International Conference on Sensing Technology* .

Irie, K., McKinnon, A. E., Unsworth, K. & Woodhead, I. M. (2007b). Shadow removal for object tracking in complex outdoor scenes, *Proc. Image and Vision Computing New Zealand* pp. 25–30.

Irie, K., McKinnon, A. E., Unsworth, K. & Woodhead, I. M. (2008a). An investigation into noise-bound shadow detection and removal, *Proc. Image and Vision Computing New Zealand* pp. 1–6.

Irie, K., McKinnon, A. E., Unsworth, K. & Woodhead, I. M. (2008b). Measurement of digital camera image noise for imaging applications, *Sensors and Transducers* **90**: 185–194.

Irie, K., McKinnon, A. E., Unsworth, K. & Woodhead, I. M. (2008c). A model for evaluation of noise in CCD digital-video cameras, *Measurement, Science, and Technology* **19**.

Irie, K., McKinnon, A. E., Unsworth, K. & Woodhead, I. M. (2008d). A technique for evaluation of CCD video-camera noise, *IEEE Trans. Circuits and Systems for Video Technology* **18**(2): 280–284.

Javed, O. & Shah, M. (2002). Tracking and object classification for automated surveillance, *Proc. 7th European Conference on Computer Vision-Part IV* pp. 343–357.

Jiang, H. & Drew, M. (2003a). Shadow-resistant tracking in video, *Proc. 2003 International Conference on Multimedia and Expo* **3**: III–77–80 vol.3.

Jiang, H. & Drew, M. S. (2003b). Shadow-resistant tracking in video, *Proc. IEEE International Conference on Multimedia and Expo* **3**: 77–80.

Kodak CCD Primer KCP-001 (1999). *Technical report*, Eastman Kodak Company - Microelectronics Technology Division.

Kremens, R. (2001). Image processing system applications: from barbie cams to space telescopes, *Technical report*.

Leone, A., Distante, C. & Buccolieri, F. (2006). A shadow elimination approach in video-surveillance context, *Pattern Recognition Letters* **27**(5): 345–355.

Litwiller, D. (2001). CCD vs. CMOS: Facts and fiction, *Photonics Spectra* pp. 154–158.

Liu, F. & Gleicher, M. (2008). Texture-consisten shadow removal, *Proc. European Conference on Computer Vision* pp. 437–450.

Loose, M., Meier, K. & Schemmel, J. (2001). A self-calibrating single-chip CMOS camera with logarithmic response, *IEEE Journal of Solid-State Circuits* **36**(4): 586–596.

Lu, C., Drew, M. & Finlayson, G. (2006). Shadow removal via flash/noflash illumination, *8th IEEE Workshop on Multimedia Signal Processing* pp. 198–201.

Lynch, D. K. & Livingston, W. (1995). *Color and Light in Nature*, Press Syndicate of the University of Cambridge, Cambridge.

Marchant, J. A. & Onyango, C. (2000). Shadow-invariant classification for scenes illuminated by daylight, *Journal of the Optical Society of America A* **17**: 1952–1961.

Matsushita, Y., Nishino, K., Ikeuchi, K. & Sakauchi (2004). Illumination normalization with time-dependent intrinsic images for video surveillance, *IEEE Trans. Pattern Analysis and Machine Intelligence* **26**(10): 1336–1347.

Mendoza, D. (1997). An analysis of CCD camera noise and its effect on pressure sensitive paint instrumentation system signal-to-noise ratio, *Record International Congress on Instrumentation in Aerospace Simulation Facilities* pp. 22–29.

Nadimi, S. & Bhanu, B. (2004). Physical models for moving shadow and object detection in video, *IEEE Trans. Pattern Analysis and Machine Intelligence* **26**(8): 1079–1087.

Nakamura, J. (2006). *Image Sensors and Signal Processing for Digital Still Cameras*, CRC Press.

Newton, I. (1730). *Opticks or a treatise of the reflections, refractions, inflections and colours of light*, fourth edn, Oxford University Press, London.

Ott, H. W. (1988). *Noise Reduction Techniques in Electronic Systems*, second edn, Canada.

Parisi, A. V., Sabburg, J. & Kimlin, M. G. (2004). *Scattered and filtered solar UV measurements*, Kluwer Academic publishers, Netherlands.

Porikli, F. & Thornton, J. (2005). Shadow flow: a recursive method to learn moving cast shadows, *Proc. IEEE 10th International Conference on Computer Vision* **1**: 891–898.

Poynton, C. (1996). *A Technical Introduction to Digital Video*, John Wiley & Sons.

Ramachandran, V. S. (1998). Perceiving shape from shading, *Scientific American* **259**: 76–83.

Ramanath, R., Snyder, W. E. & Bilbro, G. L. (2002). Demosaicking methods for bayer color arrays, *Journal of Electronic imaging* **11**(3): 306–315.

Salvador, E., Cavallaro, A. & Ebrahimi, T. (2001). Shadow identification and classification using invariant color models, *Proc. International Conference on Acoustics, Speech, and Signal Processing* **3**: 1545–1548.

Sekular, R. & Blacke, R. (1988). *Eye and brain*, second edn, McGraw-Hill, New York.

Stauder, J., Mech, R. & Ostermann, J. (1999). Detection of moving cast shadows for object segmentation, *Proc. of IEEE Trans. on Multimedia* **1**: 65–76.

Stone, M. C. (2003). *A Field Guide to Digital Color*, A K Peters Ltd.

Tian, H. & El Gamal, A. (2000). Analysis of 1/f noise in CMOS aps, *Proc. SPIE* **3965**: 168–176.

Tian, H., Fowler, B. & Gamal, A. (2001). Analysis of temporal noise in CMOS photodiode active pixel sensor, *IEEE Journal of Solid-State Circuits* **36**(1): 92–101.

Trussell, H. J., Saber, E. & Vrhel, M. (2005). Colour image processing (basics and special issue overview), *IEEE Signal Processing* **22**(1): 14–22.

Tsin, Y., Ramesh, V. & Kanade, T. (2001). Statistical calibration of CCD imaging process, *Proc. IEEE 8th International Conference on Computer Vision* **1**: 480–487.

Wang, J., Chung, Y., Chang, C. & Chen, S. (2004). Shadow detection and removal for traffic images, *Proc. IEEE International Conference on Networking, Sensing and Control* **1**: 649–654.

Weiss, Y. (2001). Deriving intrinsic images from image sequences, *Proc. 8th IEEE International Conference on Computer Vision* **62**: 68–75.

Withagen, P., Groen, F. & Schutte, K. (2007). Ias technical report ias-uva-07-02 shadow detection using a physical basis, *Technical report*, Intelligent Autonomous Systems, University of Amsterdam.

Woo, A., Poulin, P. & Fournier, A. (1990). A survey of shadow algorithms, *Computer Graphics and Applications, IEEE* **10**(6): 13–32.

Wu, T.-P. & Tang, C.-K. (2005). A bayesian approach for shadow extraction from a single image, *Proc. IEEE 10th International Conference on Computer Vision* **1**: 480–487.

Xiao, M., Han, C.-Z. & Zhang, L. (2007). Moving shadow detection and removal for traffic sequences, *International Journal of Automation and Computing* **4**: 38–46.

Xu, D., Liu, L., Li, Z. & Tang, X. (2005). Insignificant shadow detection for video segmentation, *IEEE Trans. Circuits and Systems for Video Technology* **15**: 1058–1064.

Yadid-Pecht, O. & Fish, A. (2004). *Active Pixel Sensor Design - from pixels to systems*, Kluwer.