What Creates Noticeable Defects on Digital Imagers?

by

Klinsmann Joel Wolfgang Mozart Coelho Silva Meneses

Bachelor in Electrical Engineering, Instituto Federal da Paraiba, 2017

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Applied Science

in the
School of Engineering Science
Faculty of Applied Sciences

SIMON FRASER UNIVERSITY
Summer 2021

Declaration of Committee

Name:	**Klinsmann Joel Wolfgang Mozart Coelho Silva Meneses**
Degree:	**Master of Applied Science**
Title:	**What Creates Noticeable Defects on Digital Imagers?**
Committee:	**Chair:** Ash Parameswaran Professor, Engineering Science
	Glenn H. Chapman Supervisor Professor, Engineering Science
	Israel Koren Committee Member Professor Emeritus Department of Computer and Electrical Engineering University of Massachusetts at Amherst
	Marinko Sarunic Examiner Professor, Engineering Science

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

Hot pixels are defects that are permanent in nature and develop during the camera's lifetime. A common misconception is that a small number of hot pixels will not significantly deteriorate the image quality. In this thesis we show that the accumulating hot pixels will result in image degradation and this increases as pixel shrinks and sensitivity gains. This occurs because the camera's demosaicing and JPEG compression algorithms expand the damage area of each single defect. Worse, two hot pixels within a 5×5 pixel area spread the damage widely and yet are surprisingly probable events. We developed both analytical (birthday problem) and Monte Carlo simulations to estimate the number of hot pixels required to achieve a given probability of having two defective pixels occur in a 5×5 square. Finally, we performed a perceptual assessment to evaluate how detectable the hot pixel degradation is to human eyes in a sequence of images.

Keywords: CMOS; imager defect detection; hot pixel, APS defect rate; demosaicing; image degradation.

Acknowledgements

I would like to thank my thesis committee members – Dr. Glenn Chapman, Dr. Ash Parameswaran and Dr. Marinko Sarunic for their participation in my thesis defense and research. I also want to express special thanks to my supervisor Dr. Glenn Chapman for believing in my potential since our first connection in 2015 and for giving me this opportunity to work in this project. Your years of experience, guidance and patience throughout this project were fundamental elements that helped me to conclude this work and they are much appreciated. Also, I would like to thank Dr. Israel and Dr. Zahava Koren for their assistance in this research.

A special thanks to my research colleagues at SFU – Rohan Thomas and David Yin – as well as all the co-op students that participated and assisted in this project. Additionally, I want to thank my friend Christian Presber for proofreading this paper and his advice throughout the final stages of my education.

Lastly, I want to thank my family and close friends for their help and support while I was working on this degree, especially to my best friend Tina Cusmano who has shown her loyalty and endless support all along the way. The presence of each one of you has made this journey more enjoyable.

Above all, I want to thank Almighty God Jehovah, the Greatest Engineer of all, who created all the finely-tuned incredible laws of physics that make life possible. I would not been able to complete this work without Your help.

Table of Contents

List of Tables

List of Figures

Chapter 1.

Introduction

1.1. The History of Photography

The camera has been one of the most powerful inventions when it comes to influencing the way we interact in our modern society. In fact, since its creation in the 1800s, cameras have been an important tool to record and date important events in human history and also in our personal daily lives. It was only through cameras that important historic events of modern society were recorded and preserved for the future generations to see for themselves

Moving forward into the 21st century, taking photographs is no longer just an act of memory intended to preserve a precious moment, but it is increasingly becoming a tool for showing an individual's identity and communication with others [1]. However, the camera applications go much further than casual or formal photography. Nowadays, imagers are being used in medical devices, security cameras, the industrial manufacturing process, the agricultural field, robotic vision and other areas. In the scientific field, cameras are used to investigate the weather, to analyze and study planets, animal behaviour, vegetation growth, the impact of climate change across the planet and many other fields.

Like most inventions, the camera as we think of it today was developed slowly over time. Cameras evolved from the *camera obscura* (*i.e.* a box shaped device, also known as pinhole camera, that lets light in through a small opening in one side and projects a reversed and inverted image on the other) through different generations of image capture technology – chemical processes such as daguerreotypes, calotypes, dry plates, film, to the modern electronic digital cameras and cell phone cameras [2]. One of the most popular image capture methods was film – a sheet of plastic which has layer of emulsion containing silver halide to be sensitive to light that enables image recording. In fact, film dominated the photography industry for over 100 years. Until the early 2000s, film, and the emulsion coated paper that created the print, was the most popular method of preserving a memory through photography. Companies like Kodak built an empire selling films and cameras. This technology has its owns advantages; until recently most film cameras usually had a higher resolution than what is found in most digital cameras. Also they deliver a higher dynamic range and still retain that advantage. For comparison, the smallest digital pixel available on the market is ~1µm, whereas the average size for the grain on film is ~0.015µm. When it comes to dynamic range, the digital 8 bit (256 levels) is the standard used in digital cameras, while film records a max dynamic range of 20,000 (~15 bits). However, the process of taking pictures with films is long and costly. It involves capturing an image in a film, chemically developing it and then printing it. Moreover, unless the photographer had their own dark room, they would always be dependent on a lab and its availability. Also, film can be easily damaged, it is hard to copy and it has limitations when it comes to post processing an image.

Digital cameras, on the other hand, enable users to instantly review images before they are printed and provide the opportunity to correct any damage or unwanted detail in post processes using software tools. Additionally, the cost of owning a digital camera is lower than film over the entire lifespan of the camera itself. Digital cameras can store thousands of images in single memory card, this causes a drastic reduce on the overall cost of taking pictures. In film, to capture and print an image costs approximately $1. On the other hand, taking a picture in a digital camera can cost less than 0.03 cents, if we consider the price of the camera over its lifetime and the cost to store using an SD card. Additionally, most digital cameras contain built-in software algorithms that detect and reduce visual imperfections like noise, pixel defects, chromatic aberration and others. Moreover, digital cameras enabled features such as sensitivity (ISO) and white balance, that was completely set by the type of film chosen in analog cameras, to be dynamically changed before each picture if desired, or in digital post processing. In fact, digital cameras overcame many of the biggest challenges that characterized film photograph. However, as it will be discussed, this technology had other obstacles that were not present in analog cameras. All these improvements and advantages mentioned above influenced most people to invest in a personal digital camera, and since its biggest burst in the early 2000s, millions of digital cameras and devices that use camera sensors have been manufactured and sold across the planet, making companies like Nikon and Canon the world leaders in digital camera market. Figure 1.1 illustrates the sale profile of cameras (film vs Digital) worldwide from 1979-2019.

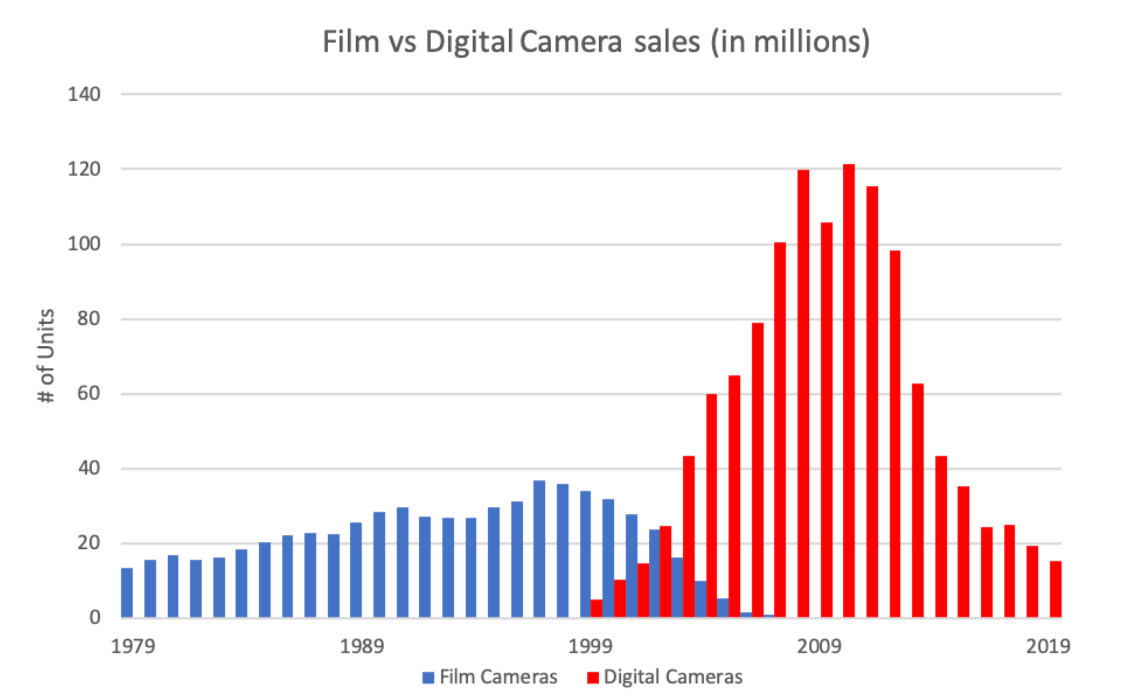


Figure 1.1 Sales profile of film and digital cameras [3,4]

From the statics shown in the graph above, in the year 1999, only 0.05 million digital cameras were sold worldwide in comparison to 33 million of film cameras [3,4]. However, because of the fast progress in technology with digital cameras and their advancing attractive features, the sales in digital cameras started to rapidly increase, and the sales of film cameras to decrease to the point where in 2010, over 121 million digital cameras were sold and the number of film cameras shipped was negligible.

Beginning in the following year, 2011, the dedicated digital camera market experienced the start to a decrease in sales. That time period coincides with the decrease in the cost of digital sensors to where they could be integrated into other devices especially seen by the rise of smartphones with useable cameras, such as the iPhone and Android models. The first touchscreen smartphone arrived stores in 2007; in the same year, the camera industry was having an impressive fiscal year where over 100 million cameras were sold. From the digital camera manufacturers' perspective, they were predicting a steady climb in sales as new features were being added to cameras. However, in the same time period, smartphone companies started to improve their cameras drastically, and that was reflected in consumers' interest in switching to smartphones cameras. This can be seen in the graph displayed in Figure 1.2. This graph shows the comparison between smartphones and camera sales worldwide from 1999 to 2019. According to Statista studies [5], digital camera sales globally dropped by 87% since 2010, while smartphones sales have increased exponentially since 2007.

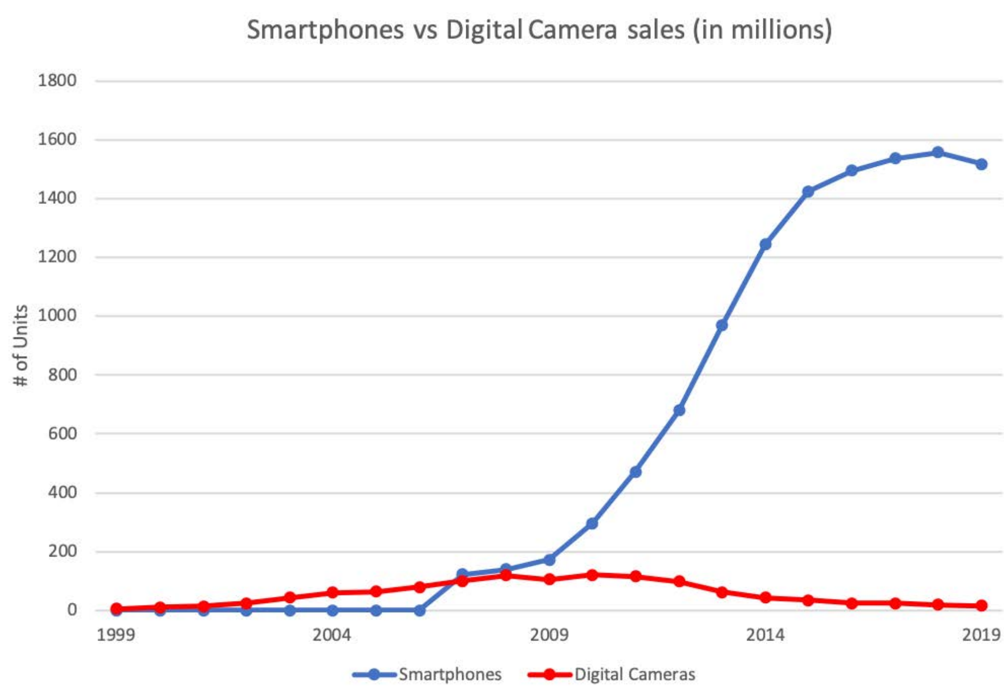


Figure 1.2 Worldwide sales profile of smartphones and digital cameras [4,6]

The largest drive of this trend is the fact that when the smartphone camera manufacturers improved their image quality to levels that equalled the lower end point and shoot cameras, many people no longer saw the need to carry or buy a dedicated photography camera for simple snapshots. While robust high quality cameras such as DSLR (Digital Single Lens Reflex) and mirrorless continue to provide better results to professionals and serious amateur photographers rivalling those of the top end film models, the modern smartphone cameras take pictures that are satisfying enough to the average consumer.

Many photographers believe that for several applications the “best camera is the camera that is always with you”, and in this decade, most people are always with their phones. Taking that into consideration, Apple, the biggest smartphone manufacturer in the world, released in 2019 the first triple-camera system on their new iPhone 11 Pro. This powerful system allows you to zoom from the telephoto camera all the way out to an ultra-wide camera, for an 4x optical zoom range. Also, with the cameras combined, the user can have higher quality images in lower light settings, shoot videos in 4k and edit the images instantly [8]. The appeal to have all of these features built-into one device drives many casual photographers to explore more fully the world of smartphone photography. Nevertheless, the laws of optics mean that the resolution of these will always be below that of semipro DSLRs.

Because smartphones (especially the high end ones such as the new models of iPhones and Galaxies) are now able to take relatively good quality photos, many argue that the end of DSLR reign may soon come to an end. However, photographers are quick to dismiss this, saying that smartphones will never be able to duplicate the quality and technology of DSLRs. This controversy has stimulated the camera developers, both DSLRs and smartphones, to create better features for their cameras. In fact, the smartphone camera industry initiated a competitive race to develop the best built-in camera that will bring the closest results to professional cameras like DSLRs. Many smartphone camera manufacturers are now investing more and more in camera improvements such as high quality cameras, that output impressive images, which are one of the strongest selling points for smartphones.

One of the camera manufacturers’ marketing techniques is to use megapixel counts to impress consumers. It is possible to increase pixel count by increasing the sensor size; however this is very costly and the size of the sensor is limited by other factors, such as the DSLR or smartphone body. Also, increasing megapixels is a much harder task in cell phones than in DSLRs because their sensor size is much smaller. In comparison, a full-frame DSLR camera has a sensor size of 864 mm^2 while smartphone cameras on average have a sensor size of 18-24 mm^2. Since the sensor size in most cases cannot be increased, the only other way to achieve a higher pixel count is by shrinking the size of every individual pixel, allowing for more pixels to fit in the sensor without changing its size. This is a technology currently in use in DSLR, mirrorless and cell phone cameras. As mentioned previously, DSLRs have a much bigger sensor size than smartphones, therefore these cameras can afford bigger pixel sizes. A regular DSLR has a pixel size in the 5-10 µm range, whereas most phone cameras have pixel size in the 1-4 µm range. To achieve the same megapixel count as robust cameras, smartphone manufacturers are developing very small pixel devices whose overall size is reduced to as low as 1 µm. Some adventurous pixel designers are even taking the further risk of developing submicron pixels to surpass the professional cameras megapixel counts. This can be seen in the graphic below (Figure 1.3), which displays the pixel count for the most high-end smartphones of 2020. It is possible to see that there are two phones (Samsung Galaxy S20 Ultra and Xiaomi Mi Note 10) that reach the impressive number of 108 megapixels, which is 2 times more than the pixel count for most DSLR cameras.

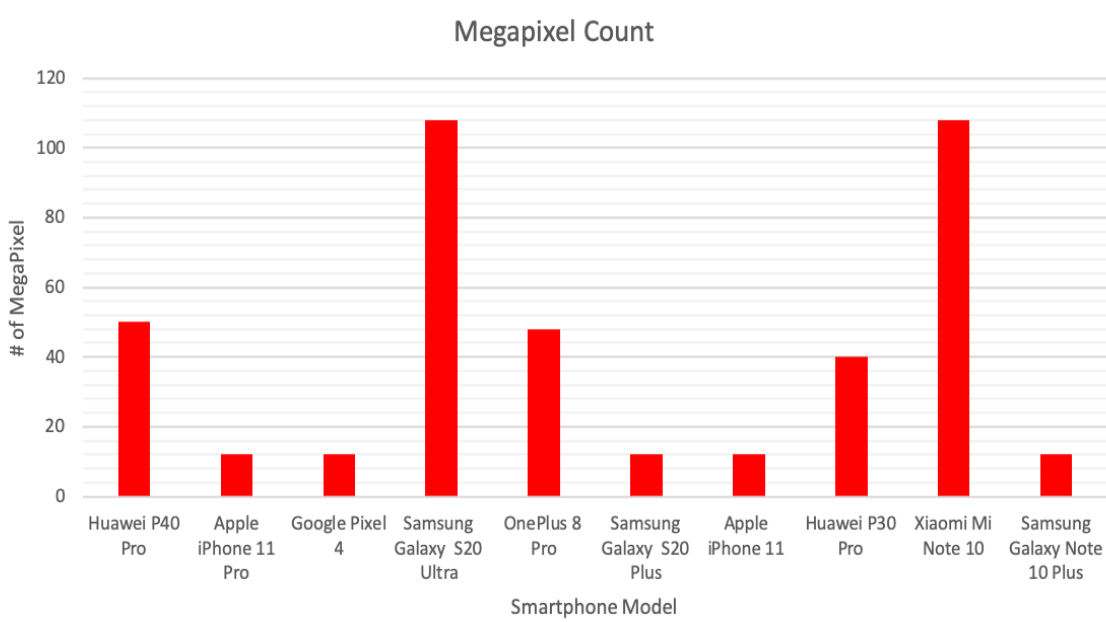


Figure 1.3 Megapixel count for different smartphone models [9]

However, the pixel size shrinkage comes with a trade-off. Although the higher pixel count could be mistakenly seen as image improvement, this is not what has been observed in practice. Smartphone manufacturers have not shown consistent performance improvement or higher quality in image outputs from sensors that have very small pixel sizes. In fact, research has shown that shrinking the overall pixel device results in ineffective pixel sensitivity, cause less noise immunity and produces lower image quality when compared to image sensors with bigger pixels [10]. However, this does not concern the smartphone camera manufacturers because the camera in a smartphone is not the main purpose of the device. The camera in a cell phone is useful and often marketed as a selling point, but the camera is just another accessory of this multitasking device. DSRL cameras, on the other hand, are mainly focused on image quality, having photography and video recording as their main purpose. Therefore, for the smartphone market, selling megapixels is more important than the image quality itself, even if that means using a much smaller sensor that is more prone to display noise and have a shorter life time.

Moreover, smartphones and their cameras, are developed to perform at their best for just a relative short period. The manufactures market these devices as they are meant be upgraded in one year or two, as new features arrive and the technology advances. Whereas, DSLRs are designed to last for several years, and have a complex image system that focuses on long lasting performance. As we will discuss in later chapters, DSLRs have much larger sensors, which allows them to have a more robust and refined light sensor circuitry that is less sensitive to current leakage and better immune to noise. Thus, in this thesis we will focus on high performance cameras that have image quality as their main purpose. We will explore what are the key elements that make a high quality camera system and we will also explore how pixel shrinkage affects the overall quality of images and what is the impact of this trend on the sensor defect development rate. We will mostly analyze DSRL cameras that have pixel sizes between 4-10 μm, however we will also show the results found in a few cell phone cameras with pixel sizes in the 1-2 μm range.

1.2. Defects Arising in Digital imagers

Figure 1.4 Defects in images: (1) A stuck-low defect, (2) partially-stuck hot pixel defect and (3) a hot pixel expended by JPEG

As we have discussed in the previous section, digital cameras have proven to be more effective, user friendly and advanced than film cameras; however, problems such as the reliability of the sensor and the emergence of defects are still significant concerns. In film cameras, a defective film portion would be replaced in the next frame of film or even the next roll. In digital cameras, on the other hand, the sensor permanently remains and accumulates errors over time due to sensor degradation. Although digital sensors can be replaced, the process to change a defective sensor can be very expensive and often it is not viable for highly integrated camera systems. Therefore, excluding software and mechanical failures, the lifetime of a digital camera may be limited by the degradation rate of its sensor.

As it is shown in Figure 1.4, image defects clearly affect the overall image quality. A sensor with faulty pixels will cause a level of disturbance in a sequence of images; in fact, a sensor that has permanent faulty pixels will output defects at the same location in all images. This is not only a problem for professional and casual photographers that are seeking the highest quality in an image, but also for many applications with embedded image sensors. As it has been recently surveyed in Nature [66], there is a considerable interest in the fact that artificial intelligence (AI) systems and deep neural networks (DNNs) can make significant errors with a relatively small number of defects. This study has shown that adding crafted noise to a picture can create an altered image that humans would see as identical, but a DNN sees as utterly different. This can cause serious issues in systems where the camera sensors are an important input for safety, such as self-driving cars. The camera in a self-driving car is used to identify different objects like a stop sign or a speed limit; if the car's onboard AI misread the word "stop" as a specific speed limit, this could cause the car to accelerate into a busy intersection, instead of slowing it down. It is natural to imagine that for this to happen, it would be necessary to add great changes in the stop sign image, however, researchers [65] have demonstrated how easy is to fool an AI system into misreading a stop sign, it only requires a few stickers attached on it. Moreover, for other engineering fields, the reliability of the imager is vitally important. The impact of a small number of pixel defects is not self-evident, because many would argue that a few defects would not cause any detriment in image from a camera with millions of pixels. However, as it was shown, these alterations in the image outputs can cause AIs and DNNs to break in unpredictable ways. Therefore, it is paramount to explore the rise of defects in digital cameras.

Another important reason to look into the sensor degradation is the increased cost of camera systems. For manufacturers, the idea of defect growth could represent more camera replacements in shorter periods, which could generate profit and make consumers interested in upgrading to the next generation of products. This was possibly true about low end commercial cameras such as point and shoot, as when they were popular, it was common to replace them in a shorter time span due to their modest cost and constant emergence of new features. However, the cost of high-end cameras, such as DSLRs and mirrorless, is significantly higher, those consumers are paying for quality and durability. For comparison, a smartphone's camera average cost is $10-20, whereas a DSLR camera system costs from $500-$5000. Hence why professional photographer aim to acquire a camera that provides high quality images for several years. Another important point on this subject is the elevated price of modern smartphones – the camera may be cheap, but the cell phone itself has many features that drive the cost up. This trend has moved most consumers to keep their phones for longer, instead of exchanging them for new models every year. Also, many modern cars use camera systems, and these cars are designed to last many years before they are replaced. Therefore, in all those scenarios the study of defect growth rate is very important since it will determine for how long the device can provide high quality images without developing noticeable defects. This will be greatly explored in Chapters 3 and 4 of this thesis.

When it comes to the cause of sensor defects, the debate in the scientific community is extensive. There are two main sources of defect development: the ones that appear during the manufacturing process and the ones that develop throughout the life of a camera, known as in-field defects. Manufacturing defects are caused by fabrication related process, such as differences across the wafer/chip and material degradation – which is the result of the decay or alteration in the semiconductor structure. These defects seem to have a cluster behavior due to microfabrication process factors and they usually occur at the start of the camera's lifetime. These microfabrication defects are easily trackable and they are usually mapped out of the sensor through calibration processes before the camera is shipped. Also, previous research [11] did not find any traces of post manufacturing caused defects during analysis; proving that these defects are eliminated prior to shipment. Additionally, the research on Dudas [55] has shown that material degradation is not the source of in-field defects. Indeed, previous research [11, 12, 55] discussed that in-field defects are caused by external sources (often cosmic radiation) that damage the sensor. They occur immediately after the fabrication process, throughout the camera's lifetime. This results in the constant development of faulty pixel cells that are visible in the image outputs. These defects are random in nature. The research mentioned previously [11] has indicated that the most common cause of sensor defects are random external sources such as cosmic radiation that hits the sensor causing its degradation. Therefore, the study of defects that develop over time is very important since it is a symptom of the real issue: cosmic ray degradation.

As we will explore in Chapter 6, for photographers, defects that repeat are very noticeable and they become more evident through time as the image quality starts to degrade. In their eyes, defects are mostly seen as another layer of noise that interferes in the final result of photographs. For researchers, these imagers' defects are analysed from the moment the cameras are purchased and, besides the quality assessment, these defects provide valuable information about IC degradation, since camera sensors are a unique type of IC that allow us to identify the location and intensity of defects.

Mainly, we can classify in-field defects into two categories. The first category are the soft errors or transient faults. These are random defects that do not cause permanent damage and do not accumulate over time. Instead they appear in only one image but not in a sequence of images. The second category are the permanent defects. These are the faulty pixels that start to appear soon after manufacture and they accumulate during the lifespan of the sensor. Regardless of their type, defects can be generalized as a malfunction in the pixel original response, or as a result of sensor degradation, causing a detectable error that can be seen in the sensor's image output [12].

1.2.1. Transient Defects – Single Event Upsets (SEU)

The first category of defects that it will be discussed is the temporary or transient defects, most well known as *Single Event Upsets* (SEUs) [13,16]. These defects are in-field faults that randomly appear and disappear in a digital sensor. They are soft errors, short-lived, cosmic rays induced faults in integrated circuits [13, 14, 16]. These defects can cause errors in IC computation by flipping bits in memory or changing results within the process units. Literature [15, 16] has indicated that all these events appear to be caused by cosmic ray particles, especially neutrons, striking the sensor at random times and locations. Cosmic rays are nuclear particles that travel through the space, with extremely penetrating character, they enter the earth's atmosphere from outer space at speed approaching that of light. Different from permanent defects, these cosmic ray collisions merely deposit the charge instead of causing irreversible damage to the IC. Therefore, SEUs analysis is a good indication of how cosmic rays are interacting with the sensor.

SEUs in digital ICs have been extensively studied, especially because of their impact on circuit reliability [12, 16, 17]. The space community, the aircraft industry and the military are especially interested in researching these transient defects as they can affect the performance of their equipment in stronger cosmic radiation environments. Also, it is known that SEUs accumulate faster in higher altitudes and that can cause a great impact in all digital systems (*e.g.* cameras, processors, etc.) on airplanes and spacecrafts. The study of SEUs in regular ICs is very difficult because the transient events are buried within the circuit and their effect is only detected as changes in the output of the chip [62]. Camera sensors, just like any regular IC, are susceptible to the same transient errors. However, as shown in Figure 1.5, these SEU events in digital imagers appear as temporary bright pixels that show up in a single image (within a series of images), but disappear in other images, and cause no permanent damage [14]. Digital cameras can detect SEUs that are not visible in ICs. SEUs are random occurrences that will not appear at the same pixel location in a batch of pictures.

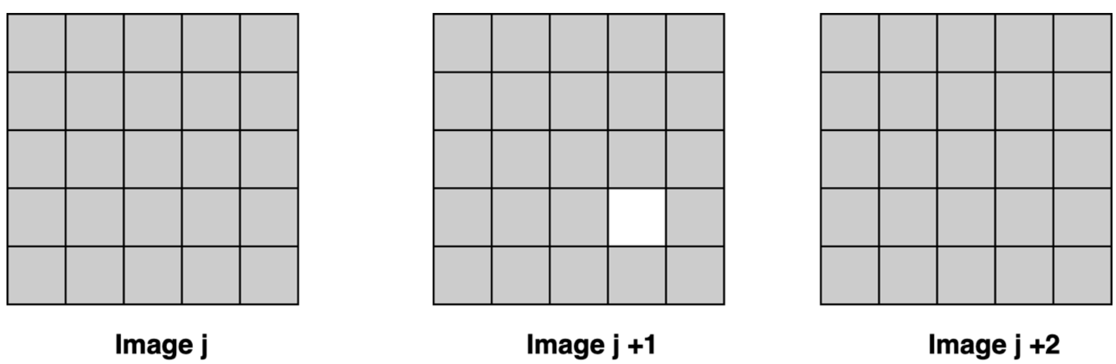


Figure 1.5 SEU event in a sequence of images j, j+1 and j+2, taken with the same parameters

SEUs provide valuable data to researchers. By analysing the dark field images (*i.e.* pictures with the complete absence of light) researchers can gather valueable information about these transient events such as their rate, their location, the charge deposited, and charge area spread. The extensive study on SEUs performed by R. Thomas [62] has shown that digital cameras can be used as radiation event detectors, since digital camera sensors have the ability to retain the charge deposited in dark frame images. They used digital imagers to explore the characteristics of SEUs, as well as to have an understanding of how cosmic radiation impacted the sensors. The data collection of this research was done by capturing a large number of dark-field image datasets (each with 1000 images) from different digital cameras – both DSLRs and cell phones. In one of the DSLRs experiments, this research has found 300 to 500 SEUs per cm^2 in a 1000 image set using 30s exposure time and ISO (or image gain) of 3200. They used software algorithms developed on MATLAB to detect the SEUs in these cameras. This research was able to develop colour noise maps (as shown in Figure 1.6) and a software detection method for SEUs called Pixel Address Distribution [54], that took into consideration the noise behaviour across the sensor while detecting SEUs.

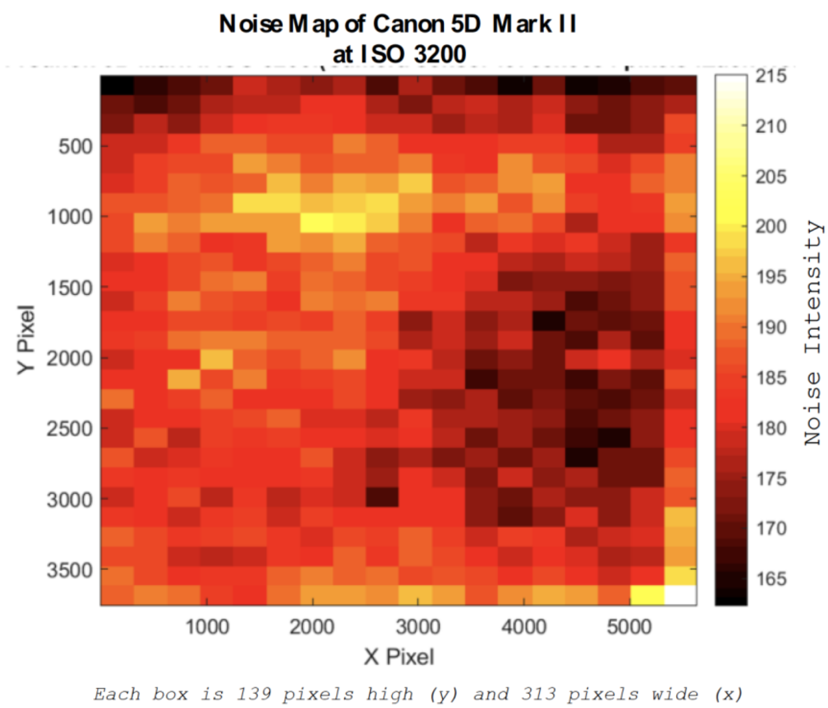


Figure 1.6 Noise Map of Canon 5D Mark II at ISO 3200 (36 mm x 24 mm sensor) (taken from [62])

This method calculates the mean and standard deviation across 1000 images for each pixel address, then it searches and verifies if the value of a given pixel address in a selected image is greater than the same pixel address in the previous and in the next image in the batch. It also checks if the value of this pixel address in the selected picture is greater than the pixel noise mean plus 5 standard deviations, or 1 in $1.7\text{x}10^6$. In other words, this method measures the background noise and then requires the pixel value to be much greater than the noise so it is statically improbable that it is not an SEU. That was a great improvement when compared to previous research that failed to detect SEUs in higher noise environments. In fact, their research was able to detect SEUs and discard false positives in higher ISOs even in small pixel sizes. In addition, they were able to understand charge distribution patterns across the image sensor. One important aspect of this research is that they performed the SEU captures at different elevations to observe the change in the SEU behaviour. Their results indicated a strong rise in SEU rates even when the elevation increases were only around fifty meters. Since they also explored cell phone images, their research was able to predict the impact shrinking IC sizes. In the end, they affirm that the SEU detection system developed on their research paired with a style of crowd would provide an accurate procedure of detecting cosmic rays.

The digital camera research performed on Simon Fraser University, under Dr. Chapman supervision, is a broad study that explored both SEUs and permanent defects, therefore some of the information that is important to both lines of study will be used in this paper. In this thesis we are only going to concentrate on permanent defects, however some of the techniques and concepts that were developed for SEUs in R. Thomas's work [62] will be discussed in this paper. The goal of this section was to deliver an overview regarding the SEUs and how they are reliable measurements of cosmic ray activities for the scientific community. Finally, this section allows us to differentiate them from the permanent defects, which is the main focus of this thesis.

1.2.2. Permanent Defects – Hot pixels

Permanent defective pixels are very different from the transient errors mentioned previously. In this case the cosmic neutron actually changes the crystal creating faults in a particular pixel and thus show up as bright or dark pixels at the same pixel position throughout a sequence of images. Because of the damage, these faulty pixels fail to sense light properly – they constantly add the defective behavior to the regular response of the sensor to the environment, time or camera settings. They can be divided in two different classes: stuck pixels and hot pixels. Figure 1.4, in the beginning of this section, illustrate the different types of permanent defective pixels.

The first class of permanent defects are the stuck pixels. As shown in the literature [18], these are often defects created during the manufacturing time and they fail to respond to light completely; their pixel value will remain the same even if the amount of incident light is changed. As it can be observed in Figure 1.4 (1) and (2), these pixels will be either fully saturated (stuck high), fully dark (stuck low) or partially-stuck (somewhere in between the extremes). Stuck pixels are usually caused by a transistor failure of the CMOS pixel design – the CMOS sensor and its structure will be discussed in detail in chapter 2 of this thesis. In classic stuck pixels, the faulty transistor turns into an open or close circuit that is unable to provide any information out of that pixel. It is important to note that in previous research [13], they have found that these defects occur only at manufacturing time and they do not develop post the manufacturing process. Also, stuck defects can be easily identified at the time of manufacture, and for most DSLRs and commercial cameras, these errors are mapped out and the sensor is calibrated prior to shipment.

The second class of permanent defects and the most important to be explored, are called hot pixels. In camera sensors, hot pixels appear as bright dots in dark images output, and their intensity varies according to the pixels' exposure time and/or the sensitivity gain (ISO). Research in Leung [11] has shown that hot pixels are infield defects that develop continuously after manufacture and, unlike stuck pixels, have been found to increase over time. It appears that the cosmic neutrons damage the pn junction interface creating a leakage current in the pixel. Another difference from stuck pixels, is that hot pixels' responses change according to light intensity variations. The final output is the combination of the pixel illumination and the defective behaviour. This happens because unlike stuck pixels, hot pixel are not caused by transistor failure, instead they are caused by a defective photodiode, that has failed to sense light properly.

Camera parameters such as exposure time and ISO (that will be explained in the next chapter) also have an impact on the hot pixel response. This is shown in Figure 1.7, where it displays the response of two types of hot pixel and a regular pixel in a dark scene.

The intensity of the slope is defined by the ISO and exposure time. It is possible to observe from this Figure that the hot pixel output intensifies as the ISO or exposure time increases. The impact that these camera parameters have and the different type of defects will be covered in more detail in Chapter 3.

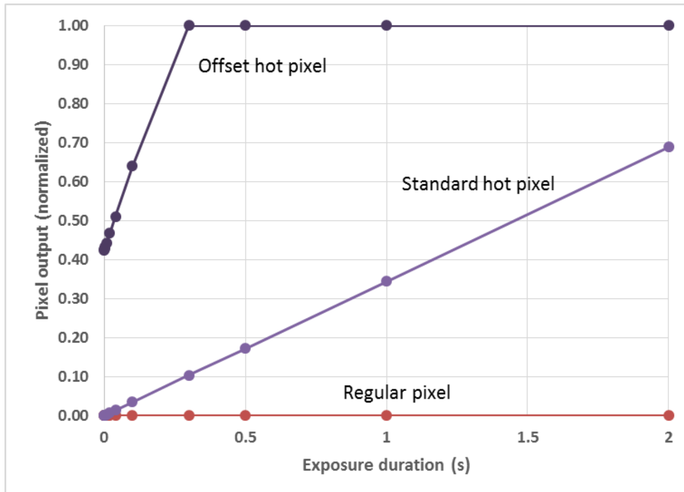


Figure 1.7 The different dark scene Reponses of two types of hot pixels and a regular pixel (taken from Chapman [57])

Figures 1.8 and 1.9 depicts the behaviour of a hot pixel shown in a sequence of images with increasing ISO and exposure times respectively. When the sensitivity gain (ISO) is increased the background pixels are also affected.

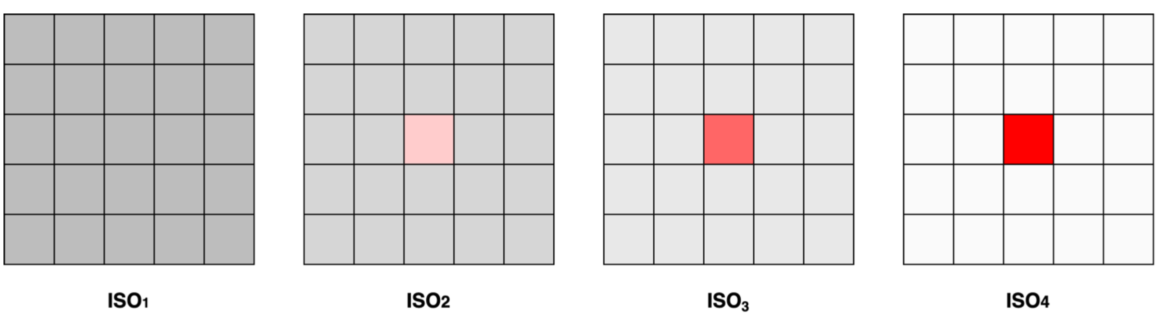


Figure 1.8 Hot pixel behaviour in increasing ISOs ($ISO_4 > ISO_3 > ISO_2 > ISO_1$)

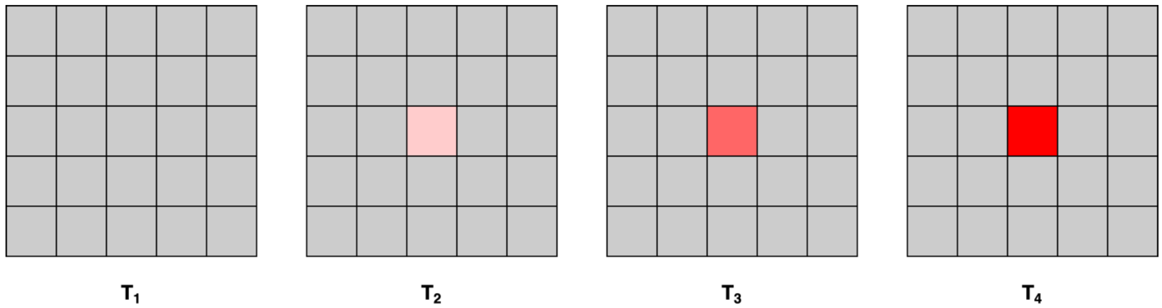


Figure 1.9 Hot pixel behaviour in increasing exposure times ($T_4 > T_3 > T_2 > T_1$)

The literature [20, 21] has shown that cosmic ray particles are the cause of permanent defect growth and development. In fact, just like the transient defects mentioned earlier, they are caused by cosmic ray degradation of the silicon crystal; however in this case, the cosmic ray charge needs to have enough energy to damage the IC substrate and create a permanent defect. Once a hot pixel is formed, it is there permanently. In this context, the study of hot pixels is very important because it shows how the sensor steadily depredates overtime due to accumulation of this damage over the area of the imager. Therefore, it is paramount, to measure how quickly these permanent defects grow in number and how the sensor degradation affects the overall image quality. Previous research in this group [13,22,24] has explored different software algorithm approaches to detect hot pixels. Additionally, those studies progressively discovered the influence that ISO, exposure time and pixel shrinkage have on the hot pixel response. Lastly, they developed a hot pixel rate growth model that related those parameters. Those studies experimentally searched for hot pixels on DSLRs cameras (with pixel size ranging from 6-10 μm) and some cell phone cameras and accumulated data over nearly 15 years. However, for smaller pixel sizes (*i.e.* < 2 μm) the level of background noise was too great to provide accurate numbers, thus the data from this pixel size range was not used in the empirical growth model. This will be covered in more detail in Chapters 3 and 4. Therefore, one of the goals of this thesis is to explore methods to improve the detection process in noisy sensors such as DSLRs in higher ISOs and cell phone cameras. As we will discuss in Chapter 3, this is particularly important for weaker permanent defects in DSLRs that would have been discarded as noise or as an SEU event, and to detect hot pixels in noisy digital camera systems.

In addition, we will explore how the interpolation algorithms used in cameras have a large impact on the spread of the hot pixel defective area. Most people think that that a small amount of hot pixels are not noticeable in images that contain millions of pixels, however because of interpolation algorithms and other processes, these hot pixels become much more visible to the human eye. As we will discuss in chapter 4, these interpolation algorithms are based on demosaicing methods and they are used to create final images in cameras that use colour array patterns. There are different demosaicing algorithms and they have distinctive impacts on the hot pixel. All of these interpolation methods will expand the impact of the defective values into the final images. Figures 1.4 (3) and Figure 1.10 display the effect that the interpolation software has on hot pixels, it is possible to see how the demosaicing method spreads the damage area to the neighbouring pixels. Also, when two or more hot pixels are too close to each other, the interpolation algorithm can cause them to interact and spread the damage area even farther. Additionally, image file compressions can create a larger impact on the final image, especially when photographers use lossy formats such as JPEGs. This chapter will bring light to misconception that requires a large number of defects to affect the overall image quality.

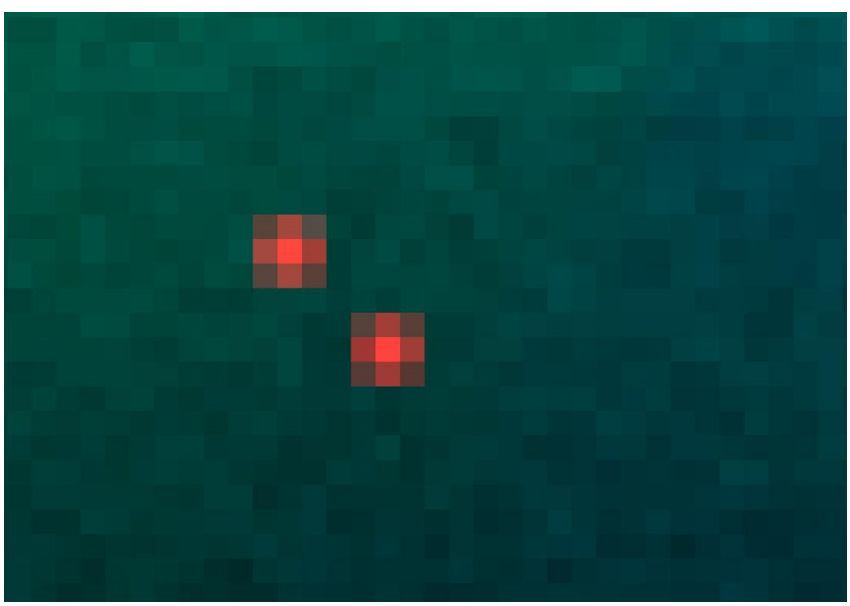


Figure 1.10 Two hot pixels with bilinear interpolation spread

Another important question we will explore in this thesis is how hot pixels affect the overall image quality from the perspective of the human eye. In regular software analysis, a few isolated defective pixels can be easily overlooked in images that contain millions of pixels. In fact, standard metrics such as PSNR (Peak Signal-to-Noise Ratio) often evaluate the impact of these defects as negligible. However, to the human eye and especially to many photographers, an area in the sensor that repeatably contains permanent defects in the same location can become more noticeable in a sequence of pictures even when the scenery changes. Therefore, we wanted to explore how noticeable a small amount of defects are to the average user and how much impact the image file compressions have from the user perspective. Very little research has been done in this area to show how we, as humans, perceive permanent defects in images. Therefore, in chapter 5, we will discuss the methodology we have developed to characterize this data and how we have developed our citizen science project with respect to permanent defects. In the following chapters, we will dive deeply into the hot pixel characteristics and analysis. We will explore this subject in detail from the sensor perspective to the human eye.

1.3. Summary

From what has been discussed in this chapter, it is clear that digital cameras are heavily used in many everyday devices, in industry processes and in the scientific field. Along with this, due to the increased cost of camera systems, DSLR photographers and smartphone users have been keeping their devices for much longer periods than in past years. Therefore, it is paramount to explore and analyze the impact that permanent defects like hot pixels have on sensors and on overall image quality output. Past research has explored hot pixels and transient defects in low noise environments, but failed to show consistent results in higher noise backgrounds. This thesis will discuss the hot pixel detection methodology explored to adapt to these circumstances and it will consider the empirical hot pixel growth model. We will also explore the effect that the demosaicing algorithms and image file compression have on hot pixels, and how we as humans perceive these permanent defects.

This thesis consist of 7 chapters. After this introductory segment, the next Chapter 2 will explore the fundamentals of the technology behind digital camera sensors and digital photography. Then in Chapter 3 we will dive in deeply into the hot pixel characteristics and analysis, summarizing previous work in this field. This chapter will also explain the experimental method used to collect the data from DSLR and cell phones cameras; as well as the hot pixel detection algorithms developed through this research. In Chapter 4, we will discuss the results of this analysis and explore the methodology to create the hot pixel growth model. In Chapter 5, the thesis begins exploring the concept of demosaicing methods. We will take a closer look at how the interpolation software and image file compression impacts the damaged area and why actually creates much more damage than simple intuition of a few hot pixels in a megapixel image commonly suggests. Next, in Chapter 6, we will explore the sensitivity of the human visual system to permanent defects in a series of images. We will discuss the methodology used to develop the prototype of a citizen science project and its results. Finally, Chapter 7 concludes this thesis with a summary of the final thoughts, as well as recommendations for future research on hot pixels.

Chapter 2.

Digital Imagers: From Light, to Sensor to Final Image

2.1. Overview

There are many elements in a digital camera that play a role in forming an image, for example, the lens system, the software features and the user settings will define the quality and the price of a digital camera. However, it is the digital sensor that has the greatest responsibility of converting light into electrical signals. In fact, the sensor impacts all the other components of the camera, hence why it is important to understand its design and behaviour in modern cameras to fully comprehend the sensor influence on defect behaviour. This chapter describes the general operation of digital cameras, from light to the final image. We begin by describing the different types of digital cameras and the image capture pipeline necessary to compose an image. Then we will discuss the sensor and the process behind photo detection. This chapter will explain the digital photography basic settings that are available in most digital and cell phone cameras, such as ISO and exposure time. Finally in the last part of this chapter we will discuss the differences between image file formats and their impact in the defect analysis.

2.2. Digital Cameras

Since the first picture was taken in 1839 the camera has undergone some significant improvements, both in quality and design. However, although many alterations have been made, the camera's principal role remains the same: getting the right amount of light through the lens and into the camera so as to create an image [25]. As discussed in the previous chapter, the microfabrication technology improvements have made digital cameras the most popular way to take pictures. There are four main types of digital cameras available in the market; they are: Point-and-shoot (PS), Digital Single Lens Reflex (DSLR), mirrorless and smartphone cameras. One could argue that smartphone cameras is really a digital camera imbedded in another device, but since it has shown great performance, it has become the common camera for most people, and image quality improvements has lead it to almost displace the Point-and-Shoot (PS) type. In this thesis we will consider cell phone cameras as a separate category.

Initially, due to the difficulty of producing imaging sensors of large area and high pixel count, Point-and-Shoot cameras were very popular before the smartphone era, with most families in developed countries owning at least one PS camera over the last decade. This type of camera targets portability; the goal is to make everything compact such that it can be easily carried around. However, the trade-off is in the imaging performance. A typical PS uses a small sensor (28 mm^2) and, to keep a high pixel count, they have their pixels shrunk to small sizes (usually around 1.5-2 μm). This results in smaller light sensitivity and increase of noise. Figure 2.1 illustrates a Point-and-shoot camera.

Figure 2.1 A typical Point-and-shoot digital camera

Smartphone cameras meet the same personal need (and market niche) as a Point-and-Shoot camera. The goal is to develop a camera that is small enough to be built in a cell phone with the best quality possible. Just like the PS, their sensor is relatively small (around

18-24 mm^2), about the size of Point-and-Shoot cameras, and their pixel size range is 1-2 μm range, with some cell phones going below 1 μm, resulting in similar issues of pixel sensitivity that the point and shoot cameras experience, but having much more thermal noise and other types of noises that come from the various phones' functionality. However, some of the newest high-end smartphones have complex software that treat and reduce much of this inherited noise. They fill the need that often the most important camera may be the one that is easiest to carry and you always have. Figure 2.2 shows the cameras of one of the high-end smartphones of 2020/2021, the Apple iPhone 12 Pro max.

Figure 2.2 The camera of an iPhone 12 Pro Max

For serious amateur and professional photographers, the most used cameras are DSLRs and mirrorless cameras. Figure 2.3 illustrates both cameras. These professional cameras can be very costly, especially when compared to the cameras previously mentioned; however, they allow the photographer to have total control of the camera's parameters (aperture, exposure time, ISO, etc). Also, they offer a much higher image quality, due their refined optics system, sensor and image correction algorithms. The DSLR (digital single-lens reflex) camera get its name from the technique that utilizes a mirror reflection of the light coming through the lens. The light is directed onto a focusing screen and a prism system atop the camera, and then projected onto a viewfinder that enables the photographer to see exactly what the lens sees and approximate visual field that will be exposed to the sensor [28]. To capture an image, the mirror swings upward into a 90° angle creating a path that will allow the light to be exposed onto the sensor. Most DSLRs nowadays also have the option that allows the photographer to utilize the LCD display as a viewfinder. These are still very widely used cameras but being replaced by a related system: the mirrorless DSLR.

Figure 2.3 A Canon 5DSr (DSLR) and a Sony Alpha 9 (Mirrorless)

In this context, a mirrorless camera is a type of digital camera that does not have a reflex mirror inside it or an optical viewfinder; thus, the light is focused directly onto the image sensor. Just like DSLRs, mirrorless cameras are also an interchangeable lens system. These cameras offer an electronic viewfinder display from the sensor which allows photographers to preview their images before they capture them and also allows a photographer to preview the image in a dark environment [31]. These electronic viewfinders are displays that create high quality images that replaces the optical eye piece image. The fact that the image from the sensor is always projected via the display, enable the mirrorless camera to have features that are only available in DSRLs when the mirror is locked up in “Live View” mode. In mirrorless cameras, some of these features include the ability to track face and eyes, to show live depth of field preview and to display how a poorly illuminated subject would look with corrected exposure in real time. One of the biggest advantages of the mirrorless cameras, in comparison with DSRLs, is their size, because they do not require a reflex mirror and the system necessary for the optical viewfinder. As well, their body is much smaller, and often lighter than the usual DSLR camera. However, compared to DSLRs these cameras have a shorter battery life (due to the prolonged used of LED and OLED displays as viewfinders) and some models lack a hand grip design that enables photographers to physically hold their cameras more comfortably. Moreover, mirrorless cameras are often more expensive than DSLR cameras, while the body of an entry level DSLR costs around $400-$500, the entry level mirrorless cameras cost from $1000-$1500. And for more professional models, only the body of the camera can easily cost $8000-$12000.

Those were the four main categories of digital cameras, and within those categories the options of brands, models and price are almost countless. The camera a photographer will choose will depend on their goal, budget and level of expertise. As mentioned before, many amateur photographers are becoming more skilled with their phones, as smartphones are now built-in with progressively more powerful cameras. Still other professional photographers will never relinquish their DSLR or mirrorless cameras!

Regardless of the camera chosen, an image from a digital camera is the result of three main steps: the optical image formation through the lens system, the conversion of light into an electrical signal and the image processing operations, such as: demosaicing, white balancing, noise removal, gamma curve adjustment, etc, which will be explained in the next section. Figure 2.4 illustrates a block diagram of a common imaging system image processing flow. This system includes the optical lenses, the colour filter array (CFA) and the image sensor. It also includes the main control systems, such as the automatic (or manual) gain control (AGC), the analog to digital converter (ADC), the auto/manual focus and auto/manual exposure circuitry. The digital signal path is completed with colour and digital image processing, and is finally sent to the baseband for storage, or to the visualization interface [33].

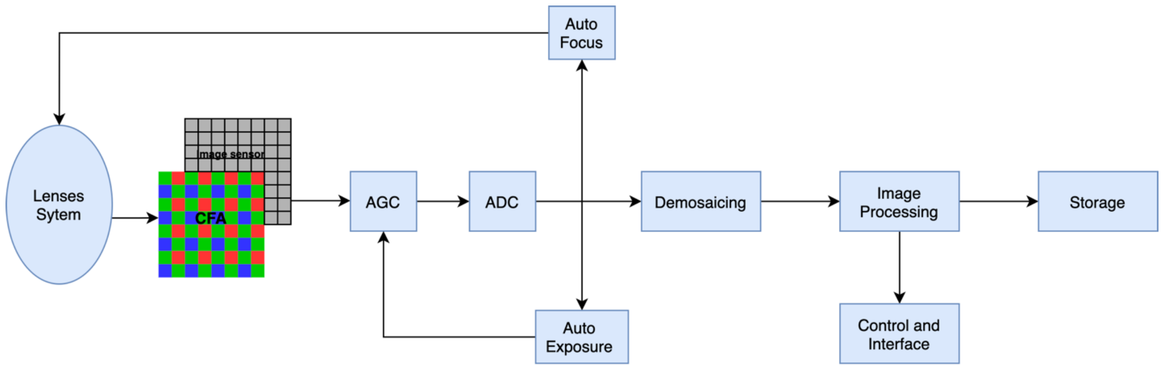


Figure 2.4 General block diagram of a digital camera (after [33])

2.3. Image Sensor

The image sensor of a camera is the most important element of a digital camera system. In fact, the quality of an image sensor is directly related to the quality of the image output and the price of the camera body. The sensor is a large microchip containing an array of pixels responsible for capturing the light that passes through the lenses, converting the light into electronic signals and then transmitting these signals to the camera or the image device processor, which transforms the electronic signals into a digital image. The image sensor comes in different sizes and formats, but it usually has a hot mirror that rejects the infrared light and perform some optical refining. It also has packaging, a sensor chip and wire bonds. Figure 2.5 displays the digital sensor of a DSLR camera.

Figure 2.5 A Canon EOS R5 digital sensor

A sensor is formed using a grid of light sensors known as pixels. Modern cameras have millions of pixels in one sensor. In the next sections we will describe the theory behind the operation of a pixel. We will discuss in detail the main components of this device and how they capture light and convert it to digital signals.

2.4. Theory of Photodetectors

The basis of the pixel are the photodetectors which are devices that convert light into an electrical signal (*i.e.* voltage or current). This process happens at the electron level of a semiconductor. The fundamental particles of light, known as photons, have to interact with the semiconductor. The energy of a photon depends on the wavelength and can be calculated using the Equation 2.1:

$$E_{photon} = \frac{h \cdot c}{\lambda} \tag{2.1}$$

where h is the Planck constant, c is the speed of light and λ is the photon wavelength. The creation of photocurrent will only occur when the photon energy is greater than the energy in the semiconductor band called band gap, E_g, which is the difference between the bands (or energy levels) of the semiconductor. Electrons in a semiconductor can reside in different energy bands, since photodetectors are a type of semiconductor, they also share this fundamental attribute. At the absolute zero temperature, 0° K, electrons will occupy the lowest possible band, known as the valence band. The other energy band is known as the conduction band and that is where electrons typically flow to generate current. However, at 0° K, no electrons are found in the conduction band. The different bands are illustrated in Figure 2.6 below.

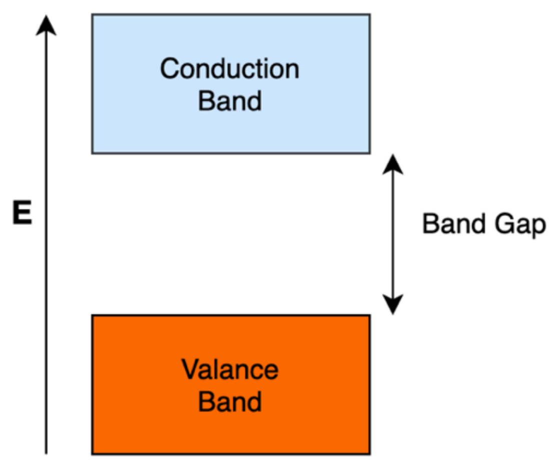


Figure 2.6 Semiconductor energy bands

When photons with energy greater than E_g (measured in electron volts) travel through the semiconductor, they excite the electrons in the valence band causing them to move to the conduction band, creating electrical current flow and leaving a mobile hole behind, this is the creation of electron hole pairs. A representation of this process is shown in Figure 2.7. Every absorbed photon will create an electron-hole pair; however, any photon with insufficient energy will not be absorbed by the semiconductor material.

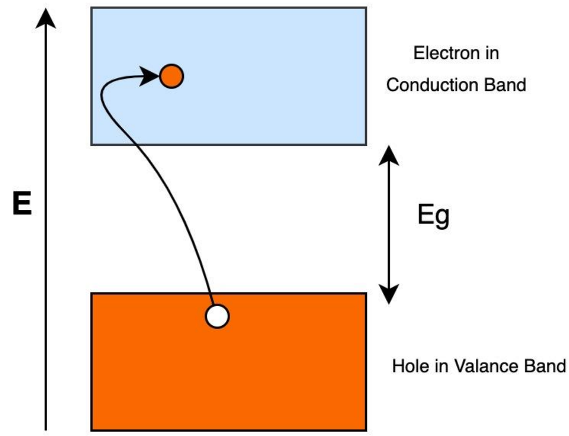


Figure 2.7 Photoelectric process

Every semiconductor material has a cut-off wavelength symbolized by λ_c. This means that any photons below this threshold will not be absorbed by the material. Silicon, for example, has a band gap energy of 1.14 eV and is capable of detecting photons in the visible spectrum (400 – 700nm) but it is not able to absorb the photons in the Infra-Red range (>1012 nm). The band gap energy is the key factor that determines a semiconductor's behaviour. It is possible to easily modify the conductivity of the semiconductor material by a number of processes, such as changing the materials resistance by adding impurities, injecting light or raising the temperature. Materials with insulator characteristics have large bandgaps, in order of 10 eV or greater and semiconductors, by definition, have band gaps that are in the range of where the optical photons have the necessary energy to make the transition (silicon, for example, has a band gap in the range of 1 – 1.5 eV).

When light penetrates the semiconductors, the initial light intensity is reduced due to the interaction between photons and the electric field of the crystal. The Beer-Lambert Law defines the intensity of the light at a given distance *x* below the surface of the semiconductor. The number of photons is decreasing with depth because they are being absorbed by the semiconductor material. This law is represented by Equation 2.2 where α is the material absorption coefficient (in cm^{-1}):

$$I\ (x) = I_0 e^{-x\alpha} \tag{2.2}$$

the absorption profile will depend on the semiconductor material and the photons' energy (wavelength). The absorption length is the average distance travelled by the photons before being absorbed. It is defined as the distance λ into a material where the photon intensity has dropped by a factor of 1/e. A photon with high energy has a larger absorption coefficient, which means that photons will be absorbed in shallower depths when compared to the photons with lower energy. Photons with a low absorption coefficient will penetrate deeply into the semiconductor before being fully absorbed. Figure 2.8 shows the absorption coefficient of silicon for different wavelengths. Note that the wavelength range for the visible light (400 – 700 nm) will fall under that category.

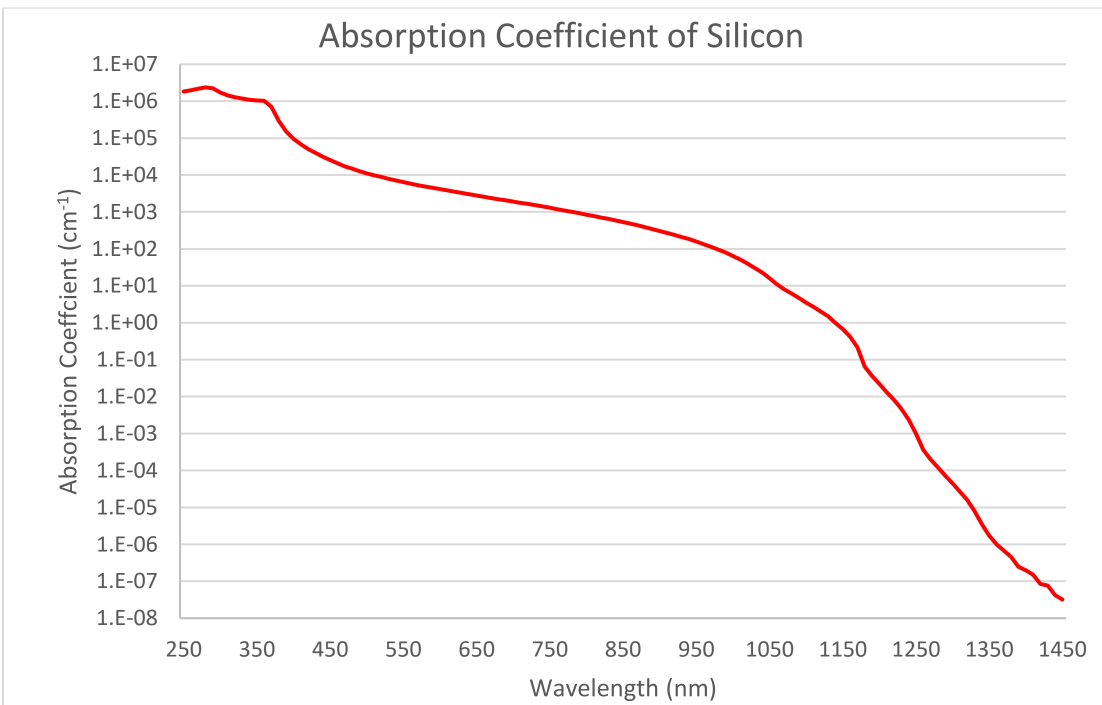


Figure 2.8 Absorption coefficient versus wavelength for single crystal silicon (taken from Green [47])

Photo carries are very transient and have a short lifetime, which makes the electrical measuring of this system complicated. Also, several carrier recombinations occur in parallel with photogeneration. Modern techniques have been developed to prevent recombination and effectively measure the photoelectric output. This technology is used in devices like photodiodes and photogates.

2.5. Theory of Photodiodes

Photodiodes are devices that utilize a P-N junction to collect photocarriers within most digital sensors and are the most common detector used in pixels. Single crystal silicon is often the chosen semiconductor, but germanium and gallium arsenide are also used in scientific devices. A P-N junction is overlapping p- and n-type semiconductors layers within the same substrate. To develop a P-N junction diode, a different impurity is added to each surface of the original semiconductor material to change how many extra holes or electrons are present. The negative (electron) charge region excess is referred the N-type and the positively carrier region is referred to as P-type. The joining of these two regions will form a junction at the interface, known as depletion region. Different from the P-type and N-type region that have positive charge and negative charge correspondingly, the depletion region is an area where almost no charge carries exist. This diffusion creates an electric field in the depletion region, and this electric field separate charges and prevents further recombination of mobile carries. In the event where the potential in the N-type region is lower than the potential in the P-type region, the electrons will flow toward the P-type from the N-type region. Electrons cannot flow from the P-type region back to the N-type region because the field in the depletion region does not allow that flow. Figure 2.9 illustrates what happens to the depletion region when an external potential is applied, this phenomenon is called bias or biasing.

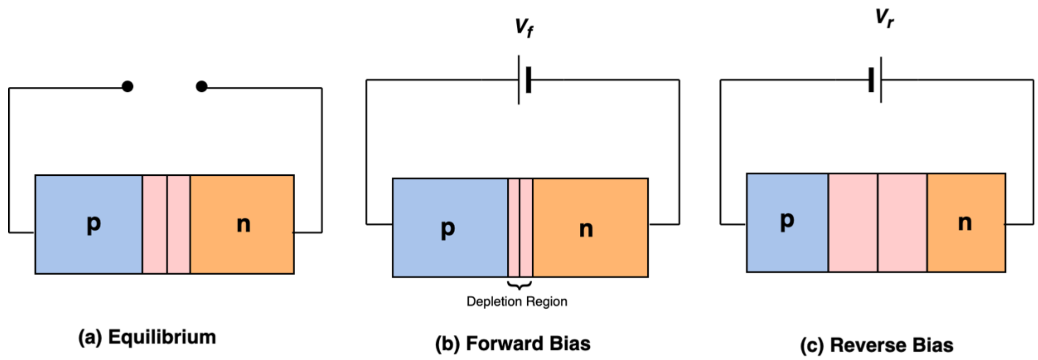


Figure 2.9 P-N junction diode under different bias

As illustrated by Figure 2.9, biasing directly affects the width of the depletion region. When forward bias is applied, the internal potential decreases, hence why the depletion region area decreases (Figure 2.9 (b)). This allows more mobile charges to move across the junction in a 'forward' direction. On the other hand, when the diode is under a reverse bias current or voltage the carriers are pulled away from the junction, hence the width of the depleted region expands (Figure 2.9 (c)). Figure 2.10 illustrates the I/V characteristic curve of photodiode with and without illumination.

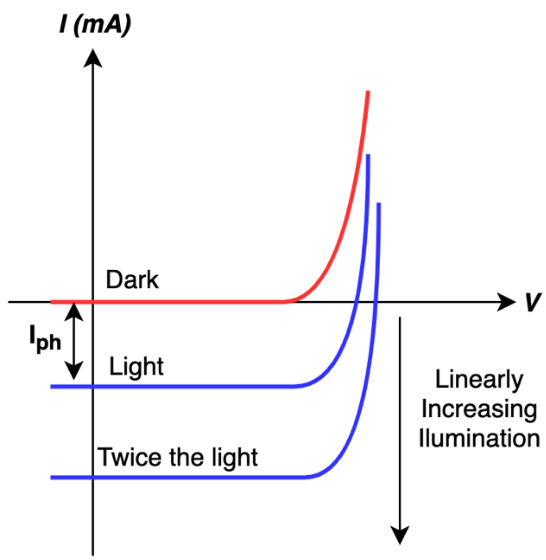


Figure 2.10 I/V curve of a PN junction photodiode (after [48])

From Figure 2.10, we observe that in a photodiode, the current linearly increases with illumination, this is the fundamental behaviour of a solar cell, a type of photodiode. When the light strikes the PN Junction in a solar cell, a photon with sufficient energy can generate electron hole pairs in the depletion region. The electric field in the depletion region drives the electrons and holes out of the depletion region. Then, the concentration of electrons in the N region and holes in the P region becomes so high that a potential difference will develop between them and as soon as a load is connected through these regions, electrons will flow through the load, creating direct current in the solar cell. Hence why the current increases linearly with light intensity. Therefore, in a general concept, a photodiode, such as a solar cell, can be used to generate current and power other devices. Other types of photodiodes can be used to determine light intensity of a scene by measuring the direct current generated by the device. In all the cases, the primary behaviour of a photodiode remains the same described above.

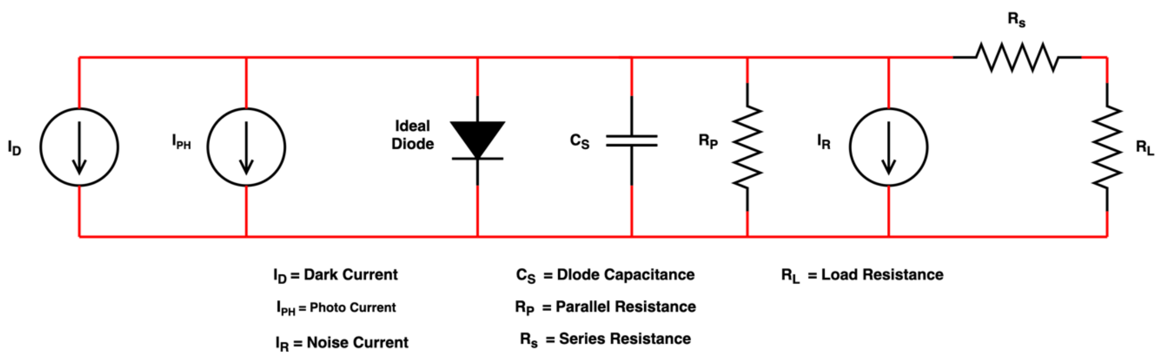


Figure 2.11 Photodiode circuit

Figure 2.11 illustrates the circuit model of a photodiode that represents its operation in terms of current, capacitance and resistance characteristics. When a reverse-biased voltage is applied to this diode, a small amount of leakage current is generated when there is no light coming into the photodiode – due to the thermal electrons going from the valance band to the conduction band. This current is called dark current as there is no light striking the photodiode. Although the dark current is usually small (from ρA to µA), it can become significantly larger when the depletion layer operates in a high temperature, because the dark current is a function of the junction's width and temperature. When there is thermal energy present, then there is a photoelectric current.

The photodiode has clear advantages when compared to other photodetectors but also some shortcomings. The advantages can be listed as the high light sensitivity, low resistance, high frequency and good spectral response. But its main advantage is the fact that they can easily be packed into an array to create a larger photo sensor. On the other hand, the photodiode shortcomings are mainly related to the dark current mentioned previously. As it will be discussed in Chapter 3, it has been found that a substantial dark current is created in photodiode sensors, and this can be aggravated by radiation, since radiation can damage the depletion region enabling the charges to flow from one layer to another, resulting in current leakage. In modern digital cameras, the CMOS sensor uses photodiodes as part of its light capturing circuitry. Therefore, it is important to analyse the CMOS implementation in digital imagers and relate to our hot pixel defect research. The CMOS sensors and its characteristics will be discussed in the following section.

2.6. The CMOS Sensor

The working principle for the CCD (Charged Coupled Device) sensor and the CMOS (Complementary Metal-Oxide Semiconductor) was developed around the same time, in the 1960s, however, the CMOS sensor was not commercialized until later in 1990s because the microfabrication processes were not advanced enough to develop a device that displayed good performance. Technically, it was that the much poorer control of impurities made the CMOS sensor characteristics widely variable where imaging was quite poor. In fact, in the early stages of digital photography's increased popularity, the CCD sensor was the industry's favourite because it was relatively easy to produce and provided good image quality because it was less sensitive to the impurity levels. In contrast, the initial developments of the CMOS sensor displayed poor noise performance, and this restrained its application when it was first introduced to the market. However, as the microfabrication technology advanced, the CMOS sensors gradually replaced the CCD sensors in digital cameras, not only because of the increased progress in the CMOS technology, resulting in better sensor sensitivity and improved noise performance. But also because of the shortcomings of the CCD sensor.

The biggest disadvantage of the CCD sensor is in fact the key difference between the working principle of a CCD sensor and a CMOS sensor. Although they share the same fundamental operation principle (they both use a photosensitive element as the primary means to capture photons), they use different methods to convert the photonic information to the readout circuit. On one hand, the CCD technology uses what is called a "global shutter", this means that the entire frame is captured at once. To achieve this, the CCD constantly uses power to capture and store the charges and shift charge packets to the output. The constant consumption of power during the integration phase makes the CCD sensor use quite a lot of energy and generate heat quickly. On the other hand, the CMOS sensor behaves like a memory chip, its information can be randomly accessed – it does not have to be read out in any particular order. This is achieved by measuring the collected charges at each pixel. Additionally, the sensor only requires power when there is a logical change in the circuit's output. These characteristics results in less energy consumption from the CMOS sensor and less overheating issues. Which makes them more suited for video-recording and cameras with 'live view' functions. Early DSLRs had neither of those features, because they used CCD sensors.

Today, the CMOS technology is prevalent in most digital camera applications. In reality, it is a better cost-effective option when compared to the CCDs. CMOS sensors can be built with standard CMOS technology, which is a very similar fabrication process of a memory sensor, and its design allows the transistors and IC circuits to be built on the same chip. In recent years the applications of the CCD sensor has been limited to the scientific and medical fields. Therefore, taking into consideration current market developments and the cameras tested throughout this research, this thesis will not explore the CCD sensors, photogates and their defect rates, but rather will focus on the modern-day CMOS sensor.

The CMOS sensor, also known as active-pixel sensor (APS) get its name from the technology used to design a pixel cell. The name is an acronym for Complementary Metal-

Oxide-Semiconductor. It shares much of the same structure as the CMOS memory used in computers. However, whereas memory chips use several rows of transistors to record data, the CMOS sensor contains rows of pixels containing photodiodes coupled with individual transistors amplifiers to amplify the electrical signal from the photodiode, plus reset and addressing transistors. This structure not only enables the CMOS sensor to operate with less electrical power than the CCDs, but also promotes speedier and easier readings of electrical charges [50]. Figure 2.12 illustrates the 3T CMOS pixel design. This structure will be discussed in more detail in this section.

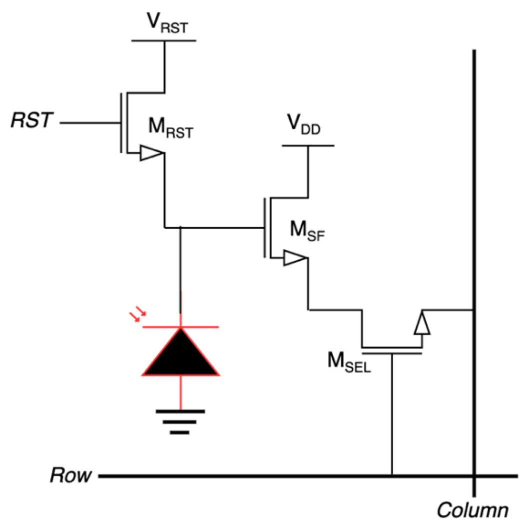


Figure 2.12 3T CMOS pixel structure

The reduced power consumption and faster data writing are the strongest advantages of the CMOS sensor design. The features are possible due to the design characteristics where one of the transistors in the pair is always 'off'. This results in reduced power consumption and higher noise immunity. Also, since CMOS sensors have the same basic structure as computer microprocessors, they can be mass produced using the same well-established manufacturing technology, making their production much less costly than other sensors.

CMOS sensors use active pixel on their design. An active pixel employs an in-pixel amplifier and therefore has a better noise and speed performances when compared to the passive pixel [51]. These primary advantages propelled the active pixel into becoming the primary choice for moderns CMOS image sensors. Looking at the composition of the active pixel, the capacitance is larger than the rest of the circuit. Almost all of the capacitance of the system is on the photodiode, and the photodiode itself takes up around twenty-five percent of the pixel device. Therefore, at first approximation, the total capacitance, namely the total charge collected by the circuitry, is set by the size of the photodiode. However, increasing the size of the photodiode does not make it more sensitive, in fact, the capacitance scales with the size of the pixel. This can be understood by looking at the equations for the voltage in a capacitor and the basic capacitance formula, Equations 2.3 and 2.4 respectively.

$$V = \frac{Q}{C} \tag{2.3}$$

$$C = \varepsilon \frac{A}{d} \tag{2.4}$$

The output voltage of the photodiode (V) is a function of the capacitance (C) of the photodiode and the capacitance is a function of the area (A) of the photodiode. Thus, as the number of photoelectrons (light) collected scales with the area of the pixel, the capacitance (C) also scales with the same photodiode area. Hence, the photocurrent and capacitance varies but the sensitivity stays constant. This is why shirking the size of a pixel in the case of a CMOS sensor does not affect the sensitivity of the device at first approximation, not until the device gets into the micron size pixel where fringing fields become important. Hence this stays true for DSLRs, but not for cell phones pixels in 1 micron range, as for this last case, the noise effects start dominating the pixel. This is the reason why sensors that have pixels in the 1 micron range, such as cell phones, experience high levels of noise even at modest ISO (sensitivity) levels.

Moving forwards into the analysis of the active pixel circuit, we come across the 3T pixel design. Figure 2.12 illustrates the schematics of this type of pixel. The 3T stands for the three transistors that dictate the function of the pixel in this circuit. This is most typical pixel design found in majority of the current CMOS sensors. As it was mentioned, the photodiode takes the most space of the APS pixel design. In comparison, the three transistors used in the 3T structures takes much less space. One point that warrants highlighting is that the photodiode is brought to high in order to cause the incident light to discharge itself. We measure the output of the pixel cell as this pixel discharge. This feature allows the circuit to be protected against outliers that could be present if the measurements were done from low to high.

The 3T pixel sensor uses three transistors to control the pixel operation. As shown in Figure 2.12, its structure is composed of a photodiode PD (which converts the photon energy into electron hole pairs and collects the charge), a reset transistor (RST), a row selector transistor (RS) and a source follower transistor (SF). The first transistor to be considered is the source follower transistor or amplifier. Its function is to collect the output from the diode without removing the accumulated charge, fundamentally working as a temporary buffer. The second transistor to be discussed is the reset transistor. This transistor acts like a switch to bring the circuit in and out of the reset mode. The main function of the reset transistor is to charge the photodiode and the gate of the source follower transistor. The reset raises the voltage of the gate of the transistor and the photodiode to full VDD, and light discharges them. Therefore, the transistor is fully turned on when it is not exposed to light, and when it saturates, it is fully turned off. The reset transistor is also important for when the user powers 'ON' the device. In that state, the circuit will be held briefly in a reset until software initialization is complete. Lastly, the third transistor to be discussed is the row selector transistor. This transistor allows circuits from higher levels to collect data from a specific pixel using row and column selection signals.

Another design that is used in modern CMOS sensors is the 4T CMOS pixel design. This circuitry uses four transistors instead of three. Figure 2.13 illustrates this design.

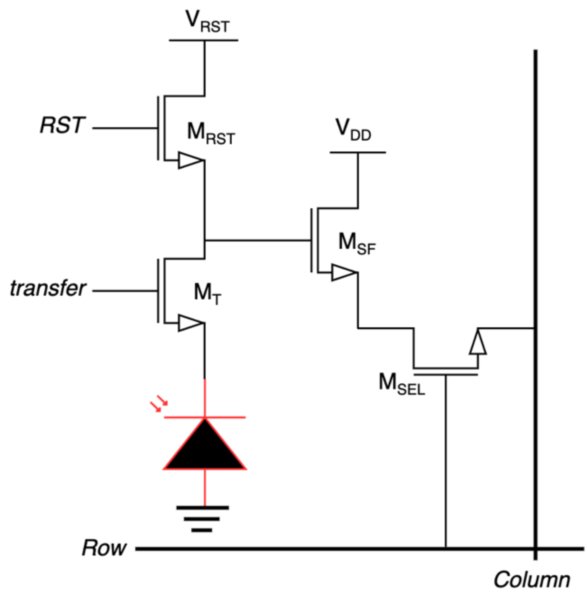


Figure 2.13 4T CMOS pixel design

In principle, the 4T CMOS pixel has a similar design as the 3T CMOS pixel, the difference being the extra transistor added after the photodiode. The fourth transistor is called the transfer transistor, and its role is to add additional speed and accuracy in extracting the total stored charge from the photodiode. There are many other CMOS pixel designs. In fact, as technology progresses and more functionalities are added to camera devices, more complex pixel designs are being developed. High-dynamic-range (HDR) is one of these modern technologies being explored and refined through advanced circuitry design. Some of these devices have transistor components that have increased the dynamic range of the pixel cell and are specifically designed for HDR capture.

The pixel arrangement in a CMOS sensor follows a two-dimensional matrix form structured in an array by rows and columns. Figure 2.14 illustrates the active pixel array in a block diagram form. Each pixel circuit is connected to a row and column selection circuitry in the CMOS sensor, as well as a common reset signal. This configuration allows a higher level of processing circuitry to read out individual pixels output by using a specific row and column signal selector. Moreover, this read out process requires a large multiplexer to collect the pixels information from a given selected row and column. However, having the reset signal shared between the pixels enables the reset information to be switched on and off across the sensor as a unit, instead of resetting individual pixels. One trend implemented in modern CMOS design is the presence of circuitry that will read the pixel's output in groups instead of individually. This results in faster read-out time and increases the overall performance of the sensor. Another important element of APS circuit displayed in Figure 2.14 is the output amplifier connected to the column selector circuit. This column amplifier is directedly related to the gain control on the CMOS sensor; when the gain of the amplifier is changed, the ISO (sensitivity) also changes accordingly. This concept will be covered in more detail in the next section.

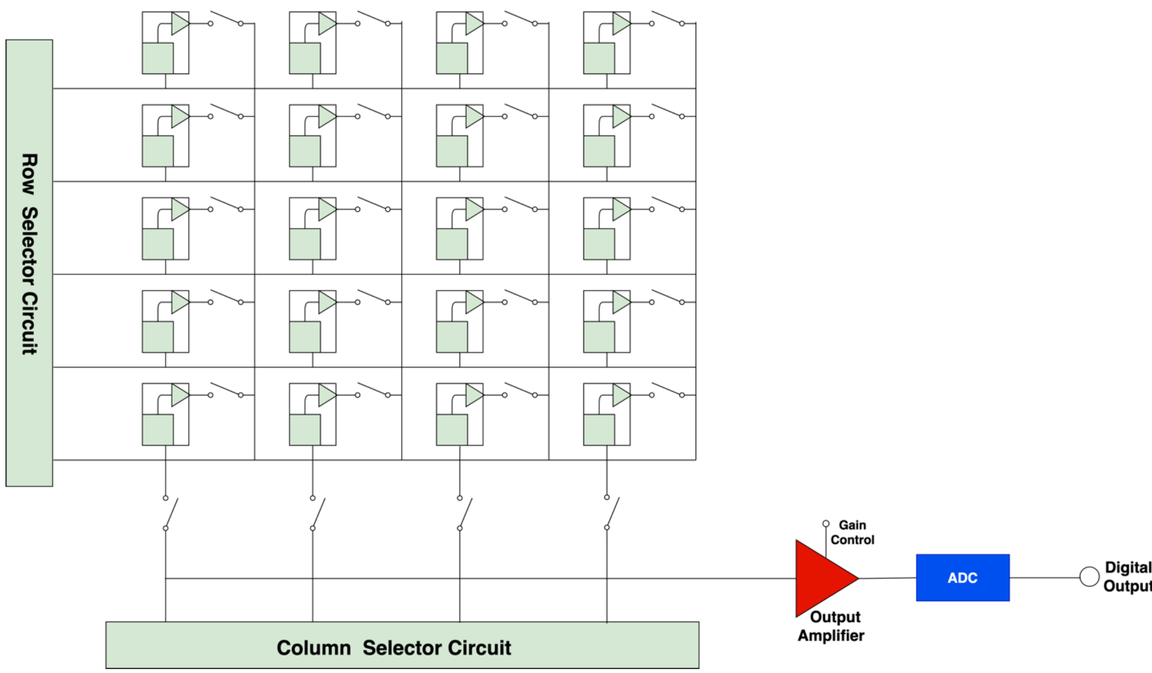


Figure 2.14 Block diagram of CMOS sensor circuit

A notable feature that warrants highlighting is the structure of the pixel. In its design, the majority portion of the pixel area is occupied by the photodetector, which allows the pixel to be more efficient and maximizes light capture. Additionally, the pixel area is also composed of other electronic components, such as the transistors, that also occupy a physical space in the design, and as result, some of the area that could be used for light detection is lost. This ratio between a pixel's light sensitivity area and the total area of the pixel cell itself is called the fill factor of a pixel. Therefore, adding more circuit elements will decrease the fill factor of a pixel, while reducing transistors will increase the fill factor of a pixel. Adding more transistors to the design will improve the overall pixel exposure performance, but there will be a trade-off with respect to fill-factor in more complex designs.

Figure 2.15 illustrates the concept of the fill factor of a standard 3T CMOS pixel cell (8 pixel size). In this particular design, the fill factor is around 50% with a light sensitive area of 4.44 μm x 6.75 μm. M1 is the reset transistor, M2 is the source follower transistor and M3 is the selector transistor. As can be observed, the photodiode takes the majority of the pixel area when compared to the rest of the circuitry. However, in modern camera devices the fill factor became less important because sensor designers started including micro-lenses (shown in Figure 2.16) on each pixel to increase the light detection efficiency. In fact, this feature substantially increased the sensitivity of the pixel. In comparison with a photodiode that does not have micro-lenses, the new modern pixel designs are up to three times more sensitive to light.

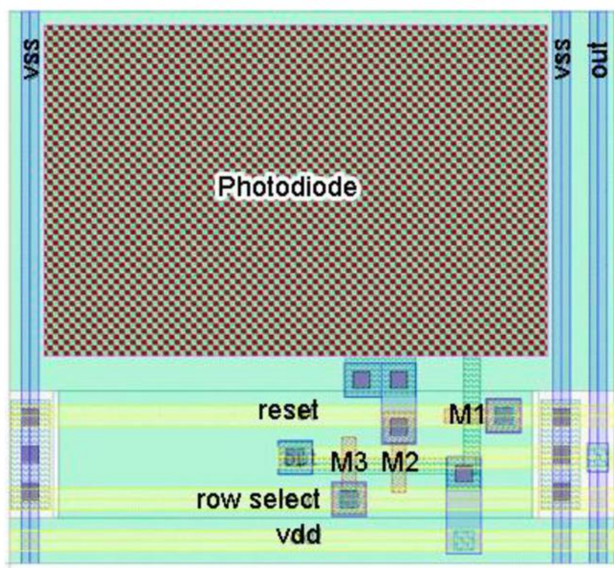


Figure 2.15 3T CMOS pixel design (taken from [12])

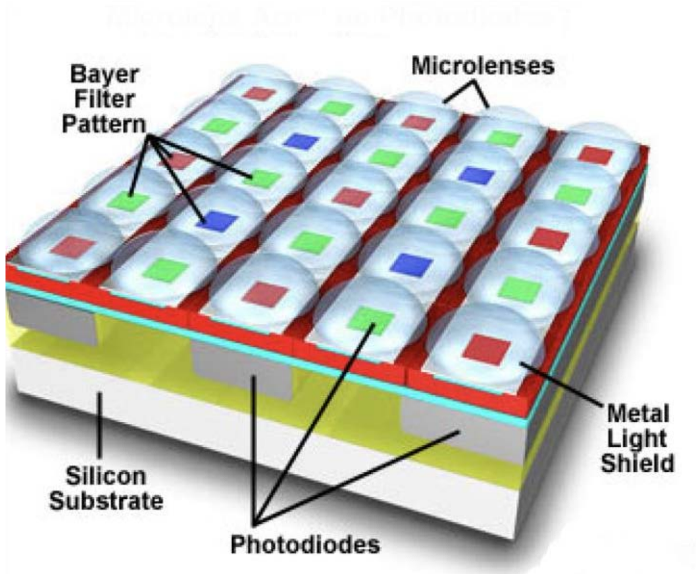


Figure 2.16 Microlens array on photodiodes (taken from [81])

The next section will explore all the important elements to form a digital image. From the definition to pixels and sensor, to the role that the gain system (ISO) and exposure time plays in forming an image. It will also explain the Bayer patter and the impact of choosing RAW or JPEG in analyzing pixel defects.

2.7. Digital Photography

The previous sections explored the photo detection process and the electronics behind the composition of a digital sensor. However, there are other elements that play a role in capturing a digital image. Here we will take a closer look at the gain system (ISO) and exposure time, as well as the colour interpolation methods and the impact that image formats have in the final output of a camera.

2.7.1. ISO

A pixel response is controlled by different factors, and one of them is the ISO. The concept of ISO comes from film photography, where ISO (or ASA) is the indication of how sensitive the film is to the light. The idea is straight forward - the film with a low ISO speed has less sensitivity than a film with a higher ISO speed. As a standard, the ISO 100 was defined by properly exposing an image using an exposure time of 1/60 of a second and an aperture number of F16 in a bright sunny day at noon. In general, for scenes where there is an abundant amount of light, an experienced photographer would use a lower ISO speed. On the other hand, when there is not enough light, a photographer would use a higher ISO speed to balance the final image. Indeed, the ISO is one of the three parameters in the photography triangle (the other two are exposure time and aperture number). These three parameters tools are used by photographers to achieve the perfect image. They will set one of the parameters and balance the other two accordingly. For example, the ISO is set up in way that if the ISO is doubled, the exposure time can be cut in half or the aperture number can be increased by one stop.

In digital image systems, the concept is the same, however the performance is different. ISO in digital sensors is achieved by amplifying the output of the pixels. The actual charge accumulation and collection in the pixel is not affected by the ISO setting itself. To fully understand this concept, please recall Figure 2.14 where we showed the APS circuit, specifically the gain control at the output amplifier of the CMOS design. Gain in a digital camera is the setting that controls the amplification of the signal from the pixel cells. Thus, when the ISO settings are changed, it is the gain in the column amplifiers that is changing, not the charge collection. This means that if the ISO is doubled, the gain in the column amplifier is also doubled. It is important to note that this process amplifies the whole signal and all the pixels, including any associated background noise or defects in the sensor. This causes the background noise at higher ISO speeds to be much more evident. Photographers will often consider this noise trade-off while taking pictures in higher ISO values. This is also important for camera manufacturers and defect analysis [12]. Due to its importance, in the following chapters this thesis will discuss in detail the impact of the ISO amplification on defects in image sensors.

2.7.2. Exposure Time

Exposure time, also known as shutter speed, is the amount of time that the film or the pixel sensor is exposed to the incoming light. In simpler words, it is the time that the shutter takes to close after it has been opened. Therefore, the longer the exposure time, the more light is let into the sensor and the brighter the picture [37]. The exposure time is measured in seconds. It is usually represented in order of fractions of a second such as $1/400^{th}$ of a second, slower shutter speeds such as one or two seconds will be represented as 1" or 2". Modern DSLRs usually have shutter speeds that range from $1/8000^{th}$ of a second to 30 seconds. The sensor designers limit the longer exposures upper range to 30 seconds because of the collection of noise due to thermal electrons.

Similar to the ISO, the exposure time is a crucial parameter in creating a balanced image. When overdone, the image can be overexposed and the pixels could be overly saturated and lose detail. When underexposed, the image can look too dark and it can become difficult to distinguish detail. Figure 2.17 shows an example of an underexposed image, a correctly exposed image and overexposed image. It is important to note that the concept "correctly exposed" is very subjective to each photographer especially as it relates to the final image being sought. Also, shutter speed has an effect on how movement will be captured in an image, with lower speed capturing a blurry movement, and shorter speeds freezing quick movements and capturing them clearly and sharply.

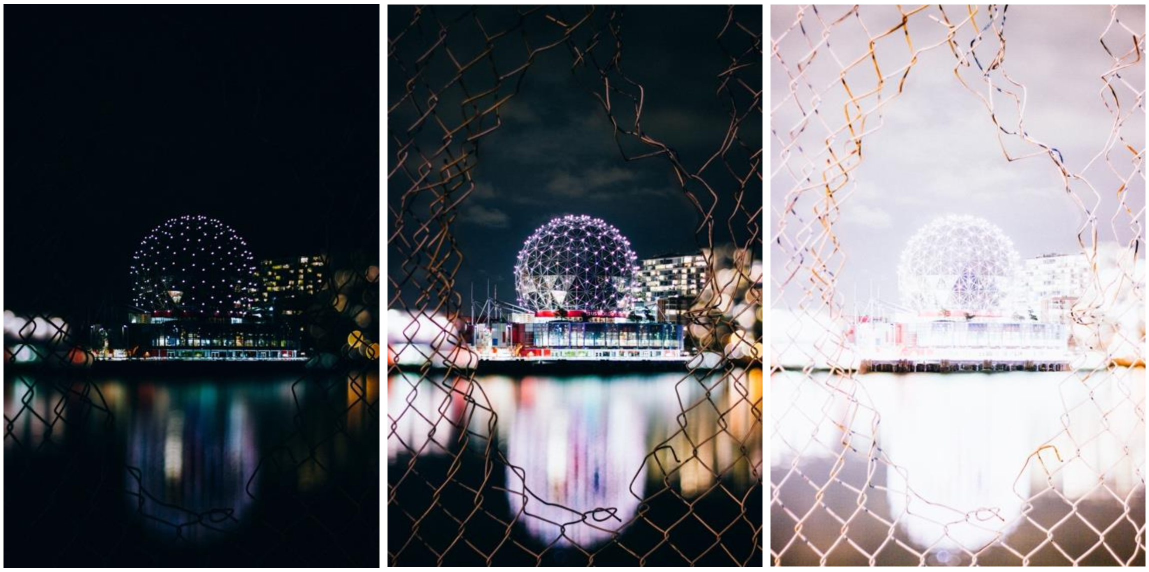


Figure 2.17 An example of an underexposed image – taken at 1/8th s (left), a correctly exposed image – taken at 1” (center) and an overly exposed image (right) – taken at 8” of the Science World in Vancouver

To quantify the exposure, the following equation can be used:

$$I = Et \tag{2.4}$$

where E is the luminance, or total flux, at the surface of the sensor measured in units of lux (lx) and t is the exposure time in seconds (s). The multiplication of these two terms will result in luminance flux measured in units of lx-s [12]. Exposure time plays an important role in characterizing pixel defects. Since the pixel outputs will change according to the exposure time, the defects will have a different behaviour as the exposure time changes. In our experiments and further discussion on defects in this thesis, exposure time will be a very important parameter.

2.7.3. Bayer Pattern

Standard CMOS pixels are sensitive to a wide range of light, from about 400 ηm to 1000 ηm and hence has very little spectral determination ability. To be able to capture colours the sensor is overlaid with something called a colour filter array (CFA) to create pixels with different spectral ranges. This overlay consists of many small filters that cover individual pixels and allow them to render colour information. Typically, they used thin film organic colour filters that pass a selected wavelength band to the pixel. Several types of CFAs have been implemented in the past but the most common one is the Bayer Pattern. Figure 2.18 illustrates the Bayer filter and the Bayer mosaic pattern.

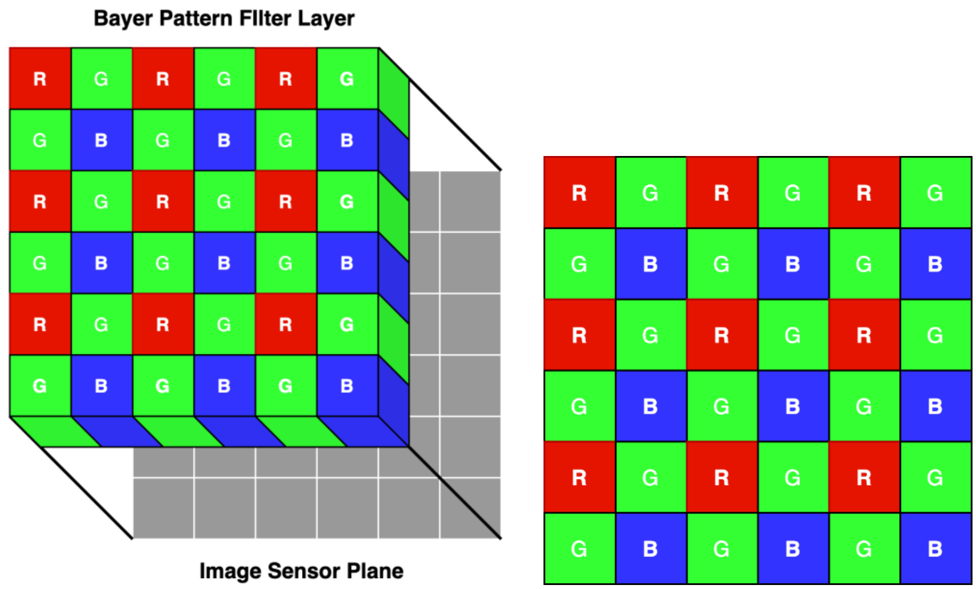


Figure 2.18 CFA sensor

The Bayer Filter was invented in 1974 by Bryce Bayer when he worked for Kodak and consists of an array of two green filter elements (G) for every Red (R) and Blue (B) filter element. This design was chosen because the human retina is more sensitive to green light (actually has twice as many green light sensing cones), and Bayer used this knowledge as an attempt to mimic our visual perception [42]. The Bayer filter works in a way where one of the three colours is positioned above each of the pixel area such that only the desired wavelength range will reach the photodiode. Thus, the CFA sensor will only collect information for that specific wavelength and will block the others. Figure 2.19 illustrates this process. In the blue filter for example, we can see that only the Blue wavelength (440-495 nm) will pass, and the other two wavelengths Red (625-740 nm) and Green (500-565 nm) will not be transmitted, resulting in no charge accumulation for these two other colour frequencies. Note the actual spectral transmission characteristics of the filters varies by manufacturer and camera model.

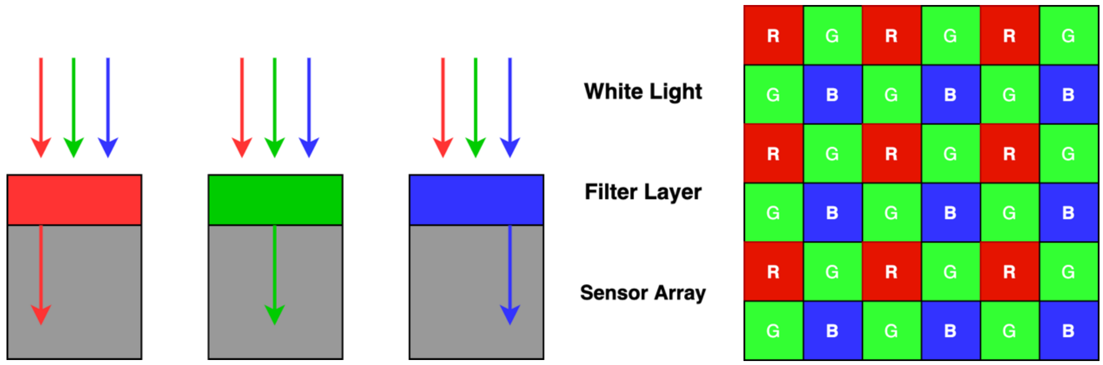


Figure 2.19 Resulting CFA pattern (after [12])

Since the CFA sensor only collects one of the three colours, it is necessary to use interpolation methods to calculate the missing colour information of a pixel, for example the Red and Blue information for Green pixels. These algorithms are called demosaicing, and they estimate the two other missing colours by interpolating the colour information from the surrounding pixels. One of the biggest problems of demosaicing are the distortions caused by interpolation errors of the missing colours (see Chapter 4 for demosaicing detail). Another problem is that current demosaicing methods ignore the presence of faulty pixels in the sensor, resulting in their being treated as normal pixels. Therefore, the presence of a defective pixel in the interpolation process causes a larger interpolation error and widens the defective area, resulting in cases where multiple defects interact with each other. Manufacturers use different demosaicing methods, and also use different attempts to create an in-field defect correction, but interpolation errors and widened defects are observed in all of them [23]. The detailed demosaicing process, the different methods used and the analysis of the impact of demosaicing on defects will be discussed in chapter 4.

2.7.4. RAW vs JPEG

In the early days (~2000) of digital photography memory storage was expensive (initial memory cards were 16MB!) and internet speeds were slow. This problem was even worse with video files. To compensate this issue, image compression techniques became the initial standard for storing/transmitting the image files. Now (2021) memory prices have dropped to 0.1% or less and internet speeds increase by over 4000%. As a result, in the photography community, there is an extensive debate on the advantages and disadvantages of using RAW vs JPEG. On one hand we have the RAW file formats, which by definition are the actual recorded, lossless compressed and unprocessed data files produced by a camera sensor. A RAW file is also known as digital negative, because of its similarities to how negative film store the entire data of the image. After capture, the image has to be processed and treated before it can be seen or used. Also, RAW images capture more bits of data in an image than the other file formats (10 to 16 bits of colour depending on camera model) but at cost of lossless compression, resulting in very large file formats [45]. As such, RAW images are mostly available only in DSLR cameras and are used by professionals who intend to process every image to get maximum resolution, minimum noise and total control over colour balance. RAW files also provide important information about the sensor and the image, information such as the size of the sensor, the colour filter array arrangement, and its colour profile as well as the image metadata such as exposure settings, ISO, camera and lenses used.

Another obstacle that RAW users experience is the fact there are no standardized RAW formats. Every manufacture has a different RAW format that is not open-source and not easily readable. For example, Canon uses .CRW, .CR2 and .CR3, Olympus uses .ORF and Nikon uses .NEF. Also, the camera manufactures change the RAW formats through time, for example, early Canon digital cameras only used .CRW, but as the technology advanced, the newer models started to only use .CR2 and in 2020 .CR3. This bring great complications as through times these older files become unreadable because they are no longer supported by new cameras and software. Therefore, the use of these proprietary RAW files as long-term archival solution carries risk and sharing these different RAW formats across complex workflows is a challenge. One of the solutions developed to overcome this issue was the digital negative (DNG) format created by Adobe. DNG is publicly available archival format that is considered a standard RAW file and it was created to store image data in a generic, highly-compatible format using lossless compression. Unlike the RAW files that are specific to each camera manufacturer, DNG is supported by various digital cameras such as Leica, Hasselblad and Pentax. However, DNG files are still not supported by the main camera manufacturers like Canon, Nikon or Sony.

JPEG, on the other hand, is the image format that uses lossy compression for storing and displaying digital images – this means that the image loses information between its original captured data and the stored file. JPEG stands for the Joint Photographic Experts Group, a committee set up by the International Organization for Standardization (ISO) in 1986 to establish a standard for the sequential progressive encoding of continuous tone grayscale and colour images. As a result of this committee, in 1992 (latest version, 1994) the JPEG standard was created [70]. JPEG files are more correctly described as being JFIF (JPEG File Interchange Format), however, most people do not make this distinction, therefore for the remaining part of this thesis when we refer to JPEG, we are specifically talking about the JPEG file. JPEG is offered in most camera devices and it is widely used due to their high compression rates and small file sizes, with relatively little loss in overall image quality depending on the level of compression chosen by the photographer. The JPEG compression takes advantage of the physical characteristics of the human perception to discard information in an image the eye cannot see. The human eye is able to distinguish the brightness of an image much more finely than its colour information, in other words, our eyes are not very sensitive to detailed differences in colours of the same brightness. Also, the eye does not see the high-frequency changes in an image very well either. In general, the eye is most sensitive to variations in brightness at an angular frequency of 0.1 to 0.2 degrees, typically a few pixels. This means that in an image the luma (brightness) information is much more important a needs higher fidelity than the two chroma (colour) components. Therefore, the JPEG compression uses this aspect of the human perception to discard most of the colour and high frequency information because our eyes will not be able to detect the difference.

The JPEG lossy compression process has two parts. First part is to separate the luminosity of an image (*i.e.* the intensity of every pixel) from the actual colour, so most of the chroma detail can be discarded by downsampling. This step is done by using a colour translation to transform the RGB channels into $Y'C_BC_R$, where Y is brightness, C_B is B – Y and C_R is R – Y. Then, it comes the first step in losing information, the downsample.

This process reduces the chroma values by a given factor. In the JPEG format, there are three options: no downsampling at all, dividing the chroma values horizontally by two or dividing the chroma values both horizontally or vertically by two.

Next, is the second part of the JPEG compression and it is where most data information is discarded. This step starts by dividing the downsampled image into 8×8 pixel blocks. The 8×8 block is the standard size used by the JPEG process, however, these pixel blocks can also vary in size depending on the level of compression, they can be 16×8 for medium compressions or 16×16 for higher compressions. It is important to note that on many occasions, the size of the image will not be a simple multiple of eight pixels in either direction. This can create pixel artifacts along the right and bottom sides of a JPEG picture.

Then, the next step is to perform the DCT (Discreet Cosine Transform) on each block and then quantize the resulting coefficient to discard higher frequency information (smaller than 0.1 degree). Each 8×8 block is converted into another matrix using the DCT. This conversion is performed by applying a set of 64 filters on each block to calculate 64 coefficients. Figure 2.20 display these filters. This is the main lossy part of the algorithm as this stage minimize the higher frequencies over the lower frequencies. The reason for doing this is that the higher frequencies can be minimized or zeroed out making them very compressible, and saving a great amount of data space without losing overall quality. As the human eye does not perceive detail loss as intensely as other lower frequency data.

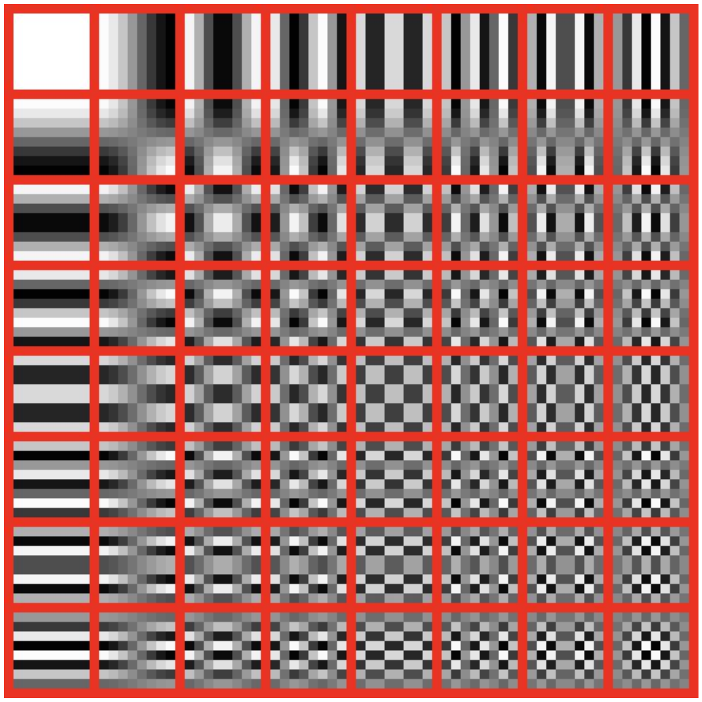


Figure 2.20 DCT Filters (taken from [71])

After the calculation of the DCT coefficients, the JPEG process removes the coefficients that are not important to recreate the image, this is known as removing the high frequency data quantization. This step is performed by dividing every one of the DCT coefficients by the corresponding quantization value in a specific quantization table and round it to the nearest integer (another lossy operation). The quantization table chosen depends on what level of quality was selected by the user. Different quality levels will use different tables that will preserve or discard more high frequency information. Other software might have their own quantization tables, Adobe Photoshop for example has 12 quality settings and they use different quantization tables for most of those settings and different sampling frequencies.

Finally, the resulting quantized matrix is encoded using the Huffman compression. This compression step is performed in a zig-zag pattern, starting from the top left corner moving the bottom right corner. This means that the zero cells in the quantized matrix tend to appear at the end of the zig-zag chain and therefore can be easily compressed by the Huffman encoding.

The multiple lossy compression steps that is performed by the JPEG process can create image defects, as there are always loss of detail because JPEG throws away high frequency data. Especially in high compressions this might result in posterization issues as well as visible artifacts. In fact, high quality JPEG files are quite successful in fooling the human eye in believing that there is no change in terms of overall quality. However due to its lossy detail nature, JPEG is not suited for images with many edges and sharp variations, or in the medical and scientific field or where the image needs to reproduce the exact data as captured [46]. Additionally, all of this complex compression process requires a high computational effort from the camera system. This results in performance limitations of digital cameras that are not designed with powerful processors that are able to handle this high computational effort. Also, the JPEG process uses a significant amount of battery power to complete all the compression steps. The loss of high frequency data, colour aberrations that results from the all the compression calculations and loss of detail are some of the main reasons why professional photographers choose RAW format files instead of JPEGs.

One problem that both RAW images and JPEG compression have is the image noise. Image noise is a random variation of brightness or colour information in the images captured. This means that pixel values are not representing the colour or the exposure of the scene correctly. Noise is caused by the degradation in the image signal. As we have discussed in Chapter 1, this degradation of the image sensor is caused by external sources such as the cosmic rays. Also, there is the thermal noise that is naturally produced by the camera while it is being used. Noise is always present in digital images, from acquisition, to coding, to transmission and to processing steps. In fact, the lossy compression process such as JPEG can change and enhance the noise levels in an image. RAW images can also display significant levels of noise depending on the photography setting. This happens because there is always an inherent amount of native noise produced by the camera, however RAW images usually have less noise content because there is no processing step that spread or aggravate the native noise. Although the noise filtering techniques and digital camera circuitry have improved throughout the years, modern cameras cannot entirely remove the noise distortion from RAW images and JPEG files because removing noise is a very complex process.

In terms of image file format, they all share a similar processing pipeline, as shown below in Figure 2.21. As can be observed, there are two noise reduction steps. These processes reduce the thermal noise effect, but change the original pixel value. These noise reduction algorithms are not open source and they vary between each camera manufacturer, making it a challenge to analyze image defects even for RAW images, since the original pixel value has changed. Also, during this process pipeline, demosaicing algorithms are applied to reconstruct a full colour image from the sensor data using the colour scheme set by the CFA pattern. All current JPEG images undergo this process during their image creation. These demosaicing algorithms interpolate the pixel's neighbours to retrieve the missing colour information and this can result in conversion noise and image artifacts, therefore manufacturers have to create complex software solutions to preserve image resolution.

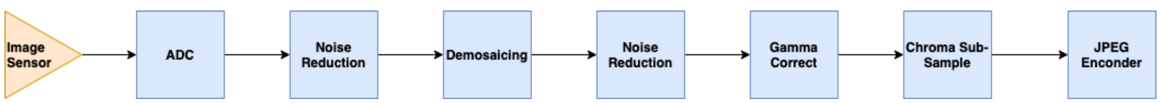


Figure 2.21 Image processing pipeline

As can be observed, the choice of the image file format has a great impact in the image sensor output. When it comes to the defect analysis, RAW images are the format chosen because they preserve most of the original image data and have the least modification. This information is necessary when analysing defect behaviour. JPEG conversions can spread these defects around, due to the heavy compression and interpolation algorithms, making it difficult to analyze the isolated defect. Therefore, in this research the analysis will be done based on RAW images collected from the imagers tested. However, the impact of JPEG compression will be also discussed to access the impact of the interaction of multiple defects and the creation of artifacts.

2.8. Summary

This chapter explored the various types of digital cameras and the technology behind the photon detection process. The CMOS sensor and the active pixel design is today's industry favourite for most digital cameras available in the market. This chapter also explored digital photography basics such as ISO, exposure time and the Bayer pattern. The impact of the image picture formats was outlined as well. The next chapters will look into defects in imagers, especially the Hot Pixels defects in greater detail as well as the detection and analysis methods used in DSLR and cell phones.

Chapter 3.

Hot Pixels in Digital Cameras

3.1. Introduction

Earlier in the thesis we mentioned the most common types of defects that are found in image sensors. In this research, we have focused on the permanent defects known as hot pixels. The literature [15,16] has shown that hot pixels are most likely caused by cosmic ray damages that impact the sensor. Dr. Chapman's ongoing research on different types and models of cameras has found that the hot pixels were the dominant type of defect that developed post fabrication, therefore hot pixels will be the area of expertise of this work, since, as will be presented in later chapters, it has a large impact on the overall image quality.

In this chapter we will focus on characterizing the hot pixels and their detection process. Past research has pursued the analysis of hot pixels in DSRLs and has started to explore the hot pixel detection in cell phones, however it failed to accurately find these defects in sensors where there is a high noise environment. We will discuss the experimental procedure to capture dark-frame images used to reveal hot pixels. Then, we will explore the detection software algorithms that have been used in the past, such as the threshold method, and the algorithms we are proposing to address the pixel noise issue. In the final part of this chapter, we will discuss the concept of defect growth rate and the curve fitting techniques used in the results.

3.2. Permanent Defects

The pixel, like all electronic devices, can develop defects through time. As of the latest work [62], hot pixels are the only type of permanent defects that we have seen develop after fabrication. However, to understand the hot pixel behaviour and its different categories, we first need to consider a non-defective pixel response. Figure 3.1 illustrates the ideal response of a perfect pixel device to a high level of illumination. Notice how the response is characterized by the shutter speed, also known as exposure time, the sensitivity of the device itself and the amount of illumination.

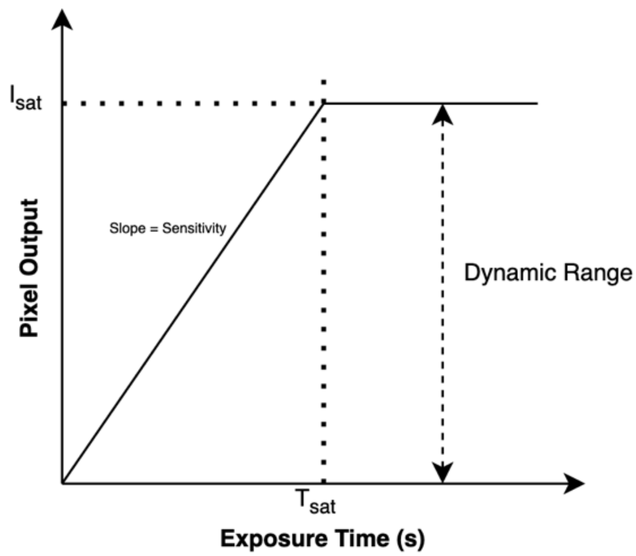


Figure 3.1 Ideal pixel response to constant illumination

As the plot shows, an ideal non-defective pixel will output 0 when there is no exposure to a light source; however, when the pixel is exposed to light, its response will increase linearly upwards with exposure time until the saturation level is reached if the illumination is strong enough. Relating this to the pixel circuit that was discussed in Chapter 2, I_{sat} is when the pixel becomes fully discharged, at the lowest gain, the saturation is the point where the electrons have been removed from the diode and the gate of the source follower transistor. Additionally, the slope inclination of Figure 3.1 is set by the gain (ISO) selection. This means that, when the ISO is doubled, the slope will become twice steep and the pixel output will reach the saturation point two times quicker as seen at the imager output and A/D converter. This can be observed in Figure 3.2 where the ISO effects on the pixel response is displayed. However, in reality when the ISO is doubled, the pixel cell itself still collects only half of the number of the electron pairs to reach saturation point. This means that for higher ISOs the diode and gate are not fully discharged when they reach the saturation. As we will discuss in this Chapter, when the ISO is increased, the hot pixel leakage current remains the same, but the hot pixel appears to become brighter, and the signal to noise ratio decreases.

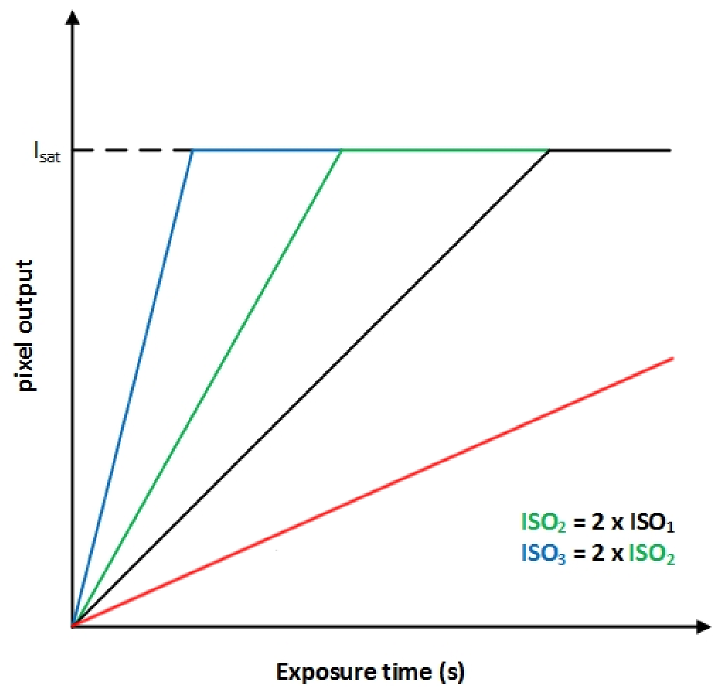


Figure 3.2 ISO effects on pixel response (adapted from [12])

For this thesis we will use a scale where "0" represents a completely dark pixel and "1" is a pixel at full saturation; however, in a real 8-bit colour pixel system, "0" represents complete darkness and "255" is a completely saturated white pixel. The mathematical model of a non-defective pixel response (I_{pix}) follows Equation 3.1, where m is the numerical slope set by the ISO gain, R_{pix} is the incident light rasponse, T_{exp} is the exposure time and T_{sat} is the exposure time where the pixel reaches the max saturation level (I_{sat}).

$$I_{pix} = \begin{cases} m[R_{pix} \cdot T_{exp}] & for\ T_{exp} < T_{sat} \\ I_{sat} & for\ T_{exp} \geq T_{sat} \end{cases} \tag{3.1}$$

Defective pixels, on the other hand, do not have the same response behaviour demonstrated by the equations and plot above; in fact they will fail to sense light properly. Figure 3.3 illustrates a hot pixel defect in a sensor array. In a perfect sensor, the entire 6x6 grid should be showing as a uniform green background, with no other colours. However, it is possible to see a red pixel in this array and this pixel is considered a defect.

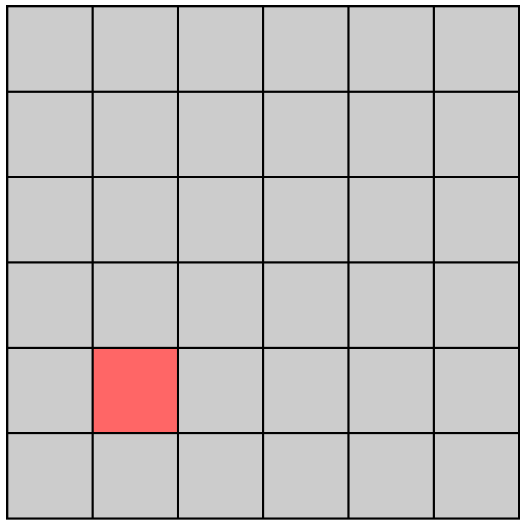


Figure 3.3 An illustration of a hot pixel defect where the illumination is a uniform colour and the defect is a red pixel in the CFA

There are different causes for defective pixels, such as electrical and material degradation, but the literature [54] has shown that cosmic rays are the main cause of defect creation because they generate the leakage current that we see in the hot pixels. Permanent faulty pixels can be categorized by three different types: stuck defective pixels, standard hot pixels and hot pixels with an offset, Figure 3.4 illustrates them. The next sections will explore these three different categories.

Figure 3.4 An image with a standard hot pixel (a), a stuck defective pixel (b) and a hot pixel with an offset (c)

3.3. Stuck Defective Pixels

Pixels that fail to respond to light completely are called stuck defects; in other words, these types of pixels retain a fixed intensity value regardless of the exposure time. Pixels categorized as fully-stuck pixels will either be stuck-high (fully saturated) or fully dark (stuck-low); and they would appear on images as a white bright spot and a black dark spot, respectively. An example of a stuck-low defective pixel can be seen in Figure 3.4 (b). A stuck pixel has an output function described in Equation 3.2, where b is a constant value.

$$I_{pix} = b \quad for\ all\ T_{exp} \tag{3.2}$$

Previous research [55] has been conducted to find and analyse stuck defective pixels, however, in the thousands of camera tests performed they have found no true stuck pixels. In fact, what this research found were hot pixels with special conditions, either an ultra-high leakage current or a large offset. Stuck pixels are easily trackable during the fabrication process, and eliminated prior to shipment through calibration and software techniques, hence why no true stuck pixels were found in previous research. Another notable discovery that warrants highlighting is the fact that stuck pixels have not been observed to develop over time in DSLR cameras [56], indicating that these defects only occur at manufacturing. Considering the results found in the above-mentioned research, stuck defective pixels will not be explored in this thesis, but it was necessary to differentiate them from hot pixels since this category of defects are infield defects.

3.4. Hot Pixels

Hot pixels defects are the main category of permanent defective pixels that will be discussed and analysed in this thesis. Different from stuck pixels, this type of defect is sensitive to light, however hot pixels fail to respond to the incident light in a similar way that a non-defective pixel does. The Equation 3.3 models the output of a hot pixel defect; it is noticeable that this formula is similar to Equation 3.1, however there are other variables inserted that adjust to the hot pixel behaviour.

$$I_{pix} = \begin{cases} m\left[R_{pix} \cdot T_{exp}\right] + m\left[R_{dark} \cdot T_{exp} + b\right] & for\ T_{exp} < T_{sat} \\ I_{sat} & for\ T_{exp} \geq T_{sat} \end{cases} \tag{3.3}$$

The mathematical model above indicates that hot pixels have an extra leakage current component in their response due to cosmic ray damage that increases linearly with exposure time (T_{exp}). This leakage component is known as dark current and it is represented by R_{dark}. Also, there is an additional offset, b, that represents the dark offset – possibly related to either transistor damage or what happens to the leakage during the rest cycle. The analysis of hot pixels, as we will discuss in later sections, is done by using images that are captured under no illumination, called dark-frames. In this context, an ideal good pixel as a unit will have a response of zero when observed in a dark-frame image. The absence of any light, R_{dark} and b will have a value near zero for a non-defective pixel. However, hot pixels will have a significant value for those variables although no incident light is present.

Therefore, the output of hot pixels will always be higher than the non-defective pixels for the same illumination intensity. There are two categories of hot pixels: standard and partially-stuck hot pixels. Figure 3.5 illustrates the pixel intensity values under no illumination (*i.e.* dark response) for a good pixel, a standard hot pixel, and a partially-stuck hot pixel (offset hot pixel).

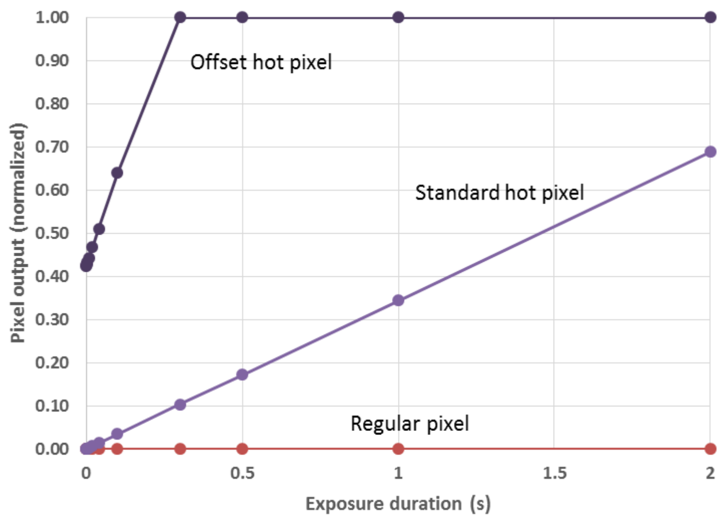


Figure 3.5 Comparing the experimental dark response of different pixels: a regular pixel, a standard hot pixel and an offset hot pixel (taken from [57].

The graphic above consolidates what was considered previously: in a dark-frame image, with no light present, an ideal non-defective pixel will have an output value of "0", because no dark current effects are disturbing the pixel behaviour. Under no illumination, an ideal good pixel should always be nearly black (0) over any exposure range; however, in practice, as we will discuss further in this chapter, there is an inherent noise in camera systems, which means the pixel's response will correspond to the noise minimum level and will not exactly equal zero. Another point to make is that hot pixels are not affected by the colour of pixel, in fact, it is the post processing combined with the Bayer filter that makes the hot pixel look like a single colour pixel. For example, in Figure 3.3, the only reason that hot pixel is displayed as red is because there is a red filter on top of that pixel cell and the camera interprets any brightness in that location as red illumination.

From Figure 3.5 we can also observe the behaviour of a standard hot pixel. This is the classic hot pixel discussed in the literature. Its linear progression through time is similar to a good pixel being exposed to light. However, in the case of a hot pixel under no illumination, what is increasing is the R_{dark} component and it is independent of the particles of light. This type of hot pixel, like any other pixel, may reach a saturation point, but often is just brighter than the background. Standard hot pixels are most visible at longer exposures as the dark current effects intensify through exposure time, thus weaker standard hot pixels could be mistaken by regular noise.

Partially-stuck defects behave similarly to a standard hot pixel; its R_{dark} component will also increase linearly through time even though the device is not being exposed to a light source. However, they have a constant offset b that adds to their response. This offset will affect the pixel output as it contains a significant initial intensity value even when the exposure time is zero.

After defining the different types of hot pixels, we can adapt Equation 3.4 to a dark response scene which the light intensity rate component R_{pix}, goes to zero. The new equation (3.5) now becomes:

$$I_{pix} = \begin{cases} m[R_{dark} \cdot T_{exp} + b] & for\ T_{exp} < T_{sat} \\ I_{sat} & for\ T_{exp} \geq T_{sat} \end{cases} \qquad (3.5)$$

In addition to the mathematical model, it is also important to understand the hot pixel behaviour in an image set. As we mentioned, the detection of hot pixels is done by capturing a set of images under no illumination called dark-frame images. In a picture with no illumination, only defective pixels will have a response that is different from "0". In dark-frame images is also possible to observe transient defects such as SEUs and noise components. However, as we have defined in this section, a hot pixel is a permanent defect with a linear crescent behaviour as exposure time increases. The defect behaviour will be observed in all images of the set. It is important to differentiate noise from hot pixels. Noise outputs have random behaviour, whereas hot pixels response display a steady leakage current pattern. As we have discussed, the hot pixel is an actual damage on the device, and it has a similar behaviour to an exposed pixel. While noise is a random process coming from other operations that are inherited to the camera system, such as thermal noise, flicker noise, etc. This is illustrated in Figure 3.6. As can be seen in Figure 3.6 (a) the hot pixel starts off with an initial defective offset and its intensity grows as the exposure time increases. This is a typical behaviour of a true hot pixel. In comparison, as seen in Figure 3.6 (b), transient defects like SEUs or noisy pixels will only be overserved in one image of the set, but not in all of them; therefore, it will be discarded from the hot pixel list. Note in this figure that exposure times are increasing.

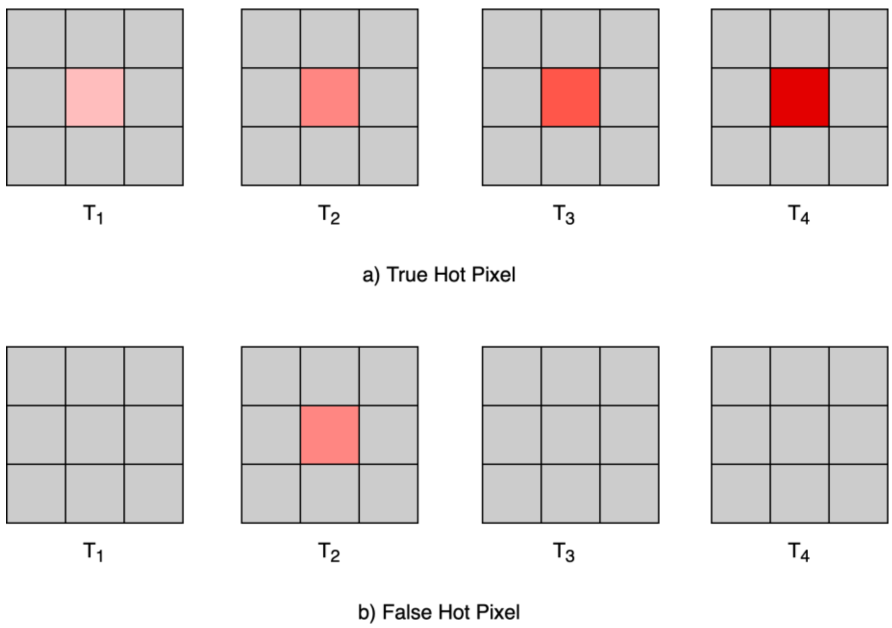


Figure 3.6 Examples of a true and a false hot pixel ($T_1 < T_2 < T_3 < T_4$)

Previous research performed in SFU has discovered that in commercial digital cameras, partially-stuck hot pixels were the most prevalent permanent defects found in image sensors. In fact, approximately 70% of these faults were partially stuck hot pixels and 30% were standard hot pixels [58]. However, these results could be the shadow of a limited detection algorithm that is not sensitive to weaker standard hot pixels. A different detection algorithm approach that is sensitive to these weaker faults could potentially improve the hot pixel detection analysis.

3.5. Hot Pixel Detection Techniques

Our detection of hot pixels is a state-of-the-art that has evolved over many years of experiments into a process that integrates many steps. This section will discuss in detail the experimental method developed to detect hot pixels in DSLRs and cell phone cameras using dark frame techniques. A dark frame or a dark-field image is the result of a method used in research where the picture is taken in the complete absence of light. Thus, this technique is mainly used to identify any bright effects. In fact, this is the method used to detect SEUs, hot pixels or any partially-stuck pixels with offset components depending on the detection algorithm you use.

There are a few advantages in using the dark frame method over other methods of defective pixel identification, the main one being the ability to record the linear nature of the pixel response. In our research, we captured multiple dark-field images in a range of exposure times (from 1/125th of a second to 4 seconds), this process is fundamental to analyse the hot pixel behaviour and its defective growth rate. It is important to mention that in previous work Dudas [55], different methods were tested. The research on Dudas [55] extracted hot pixels from a sequence on images. In their experiments they used a Bayesian type method that took in the consideration the hot pixel model behaviour. Their research took a sequence of pictures and used a Bayesian model to determine the prior and the posterior probability of the same pixel being a hot pixel in a sequence of images. This method showed consistent results when it was executed; however it is harder to preform and takes higher computational power. Another advantage of the dark frame technique is the experimental process itself; the capturing procedure is easy to reproduce as the surroundings are entirely controllable by the researcher. Experiments can be performed by simply covering the image sensor in dark room or in a sealed box. Considering the advantages of this method, this thesis will use dark frame techniques for detection and characterization of hot pixels and their behaviour, which is the similar behaviour displayed in Figure 3.6 (a) and Equation 3.5. The hot pixels captured in the dark frame have steady leakage current that will increase according to the exposure time, as displayed in the graph.

3.5.1. General Hot Pixel Experimental Method for DSLRs

DSLR cameras are the main category of image devices that were used to analyse hot pixel defects in this research. One of the biggest reasons for choosing to use DSLRs in this analysis is the fact that these cameras enable the users to manually control all the photography settings involved in capturing an image. The manual setting of these cameras allows for full control of the ISO speed, aperture, exposure time, white balance and file format. Most importantly, these cameras can output images at the highest quality of RAW format. As mentioned before, RAW images are the direct response from the sensor, and permanent defects have not been altered by compression or other image processing algorithms. Also, distinct from JPEGs, RAW images do not intensify or spread the pixel defect among their neighbours.

Throughout our research, the dark frames images from DSLRs were captured by covering the sensor (generally using a sensor cover provided by the manufacturer) in a dark illumination situation – *e.g.* in a photography dark room, or in a dark covered box. This methodology ensures that no unwanted light will leak through the camera's viewfinder which becomes important for long exposures as the ISO increases. Therefore, since the pixels are unilluminated, only sensor noise, SEUs and hot pixels will be recorded in the dark-frame image. Then the settings of the camera were adjusted to only record the highest quality format of RAW images, we also disabled any optional noise reduction and correction algorithms, flash, picture rotation and decreased the brightness of the LCD screen of the camera. The purpose of these configuration steps is to maintain the output pixel value at their most unprocessed state.

The images were taken over a wide range of increasing exposure times that were available in the DSLR cameras were selected. The first exposure time was usually 1/125th of a second. It was decided to start with this exposure time because partially-stuck hot pixels and standard hot pixels can be detected at this setting – only very strong hot pixels, with a large offset, can be observed in exposure times shorter than 1/125th of a second. The exposure time range is also increased by a factor of 2, resulting in a range of up to 4 seconds, in photography this is known as a one-stop change – *i.e.* 1/125, 1/60, 1/30, 1/16, etc. It is possible to observe that some of these shutter speeds are not exactly double of the previous, as the cameras devices only have standard exposure times available. However, we used the closest setting available to what would be the double of the previous exposure. For this experiment, one up to ten images were taken at each exposure time. Also, we used a series of different ISOs. The ISO range started at ISO 100 and increased by a factor of 2 up to ISO 25600 – *i.e.* 100, 200, 400, etc. As we increased the ISO we were able to detect weaker hot pixels.

Additionally, there is a 30s waiting period that is observed between shots to allow the camera and thus the sensor to cool down to near room temperature. The camera only produces heat when the image is being read out, processed and saved, the device does not generate heat during the exposure time. Therefore, taking in consideration these features, 30 seconds was found experimentally to be enough time to allow the system to cool down to near room temperature. Additionally, there is a 2 min stop between ISOs for the same purpose. Allowing the sensor to cool down between shots limits the thermal noise generated by the camera. Figure 3.7 illustrates this methodology. In this representation the ISO settings increases from left to right. This procedure not only enables us to verify the existence of hot pixels, but also allow us to study the hot pixel intensity growth as exposure time increases for a given ISO selection.

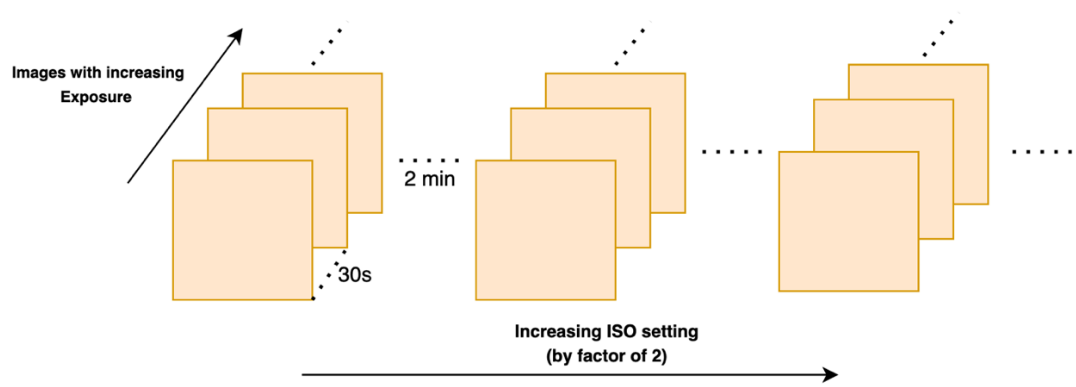


Figure 3.7 Hot pixel experimental methodology

It is important to mention that in previous research, the upper limits for the ISO Speeds and exposure times were determined by the amount of noise present in the image.

Although DSLR cameras generally have effective noise suppression processes, RAW dark-field images captured under longer exposure still contain a high quantity of noise. The previous software algorithms developed were not able to separate the excessive amount of noise and SEUs from hot pixels in higher ISO speeds. Thus, images with longer exposure times were simply discarded. Additionally, what was considered a high ISO would depend on the camera system. Older cameras such as Canon T2i, the highest ISO was 3200; for the Canon 5D Mark ii, it was ISO 12800. Newer cameras such as Canon R5 can reach up to 102,400 ISO. For cell phones these numbers are much more modest, as we will consider, ISO 800 was the highest ISO speed we could test before the pictures were overly noisy. Therefore, in this paper we will discuss the previous algorithms developed and the challenges observed. Also, we will explore the new methodologies proposed that enabled us to detect hot pixels in longer exposures.

As described previously, the experimental method of capturing dark-field images for hot pixel analysis is very dynamic; it requires constant manipulation of different camera settings. In the early stages of this research, the calibrations were done manually. However, now we use the Canon EOS Utility software – a tool that allow us to control Canon DSLR cameras remotely through an USB connection. The biggest advantage of using the EOS Utility tool is the fact that it eliminates the need for a photography dark room, as we are able to place the camera inside of a covered box and control it using a laptop or a desktop computer. Another advantage of this software is that we can store the images directly in the computer without the need of a memory card. Which removed the problems of memory cards filling and the need to constantly replace them. Also, since most of the processing is performed by the computer, the camera uses less energy and create less heat than when it's directly writing on memory cards. Therefore, this feature not only speeds up the processing times and preserves battery charge, but also significantly increases the amount of memory we can use to store the images allowing us to collect a full set of images in every run. Figure 3.8 displays the EOS Utility software. There are other software alternatives such as the Astro Photography Tool (APT). This software allows the user to store and capture the images to the computer, but also enables them to create specific routines of image shooting that fully automate the calibration process.

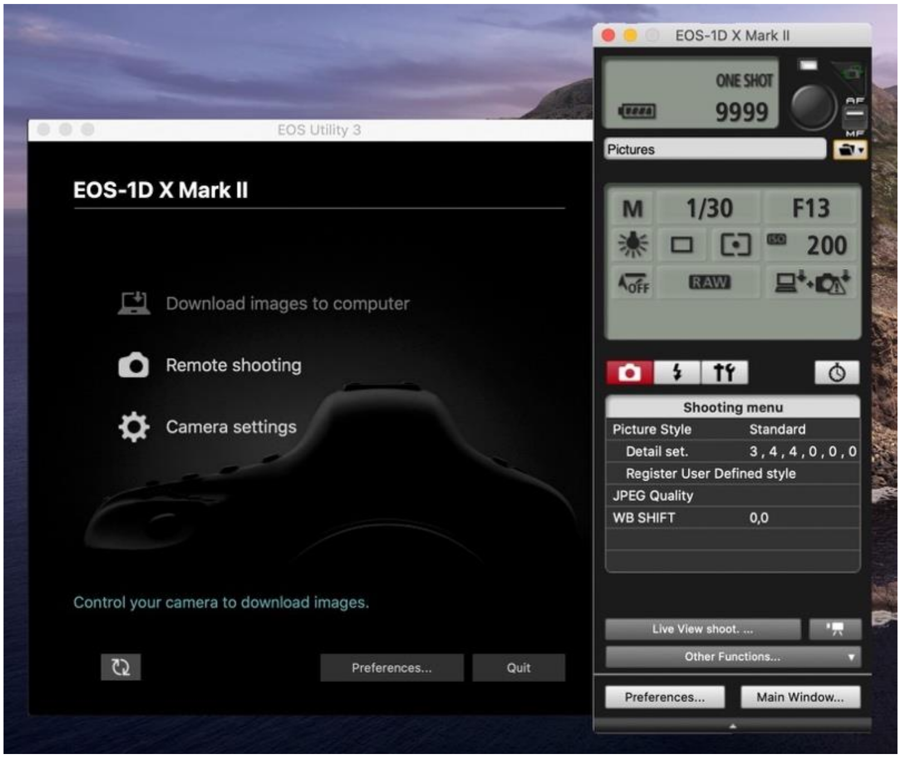


Figure 3.8 Canon EOS Utility

The process described in this section is referred to as dark-frame calibration. When completed, a calibration process will be saved with the timestamp that informs us when the images were captured. Multiple calibration processes are performed over time to produce image sets with the same ISO and exposure time ranges. This enables us to study the camera's hot pixel behaviour and its sensor degradation rate as time progresses. Once a dark-frame calibration is finished for a given timestamp, the entire image set is analysed by a software algorithm that provides the information on the hot pixel behaviour and count.

3.5.2. Dark Frame Experimental Set Up for Cell phones

As we have discussed previously, the increased popularity of smartphones has caused these devices to become the number one choice for taking pictures for most people. Many photographers agree that the best camera may be the one that they always have –

which for the causal photographer is often the cell phone camera. Therefore, we were particularly interested in studying and analysing how permanent defects affect the image quality on those cameras. These cameras, however, are structurally very different from DSLRs and the experimental method we described in the previous section had to be adapted for cell phones. Throughout the calibration process for these devices, we faced new obstacles that were not present in DSLR cameras. First, the quality of the smartphone camera sensor itself was a challenge. The cell phone camera manufacturers do not invest in pixel image quality as much as the manufacturers of DSLR and mirrorless cameras. For comparison, on average a cell phone camera sensor costs around 10 dollars at manufacture time while a DSLR camera system can cost anything from 300 to 5000 dollars. In general, cell phone camera sensors are not as well refined as the imagers in other camera devices and they have a modest quality optical system; this impacts greatly the response of the final image and the sensor performance. Also, the size of the lenses in a cell phone is much smaller than the diameter of lenses used in a DSLR system. The classic lens diffraction limit formula, shown in Equation 3.6, help us understand the impact of smaller lenses:

$$\theta = 1.22\frac{\lambda}{D} \tag{3.6}$$

where θ is the angle between point sources, λ is the wavelength of the light and *D* is the diameter of the lens system. By definition, detraction limit is the maximum sharpness limit of a lens due the laws of physics. This criterion determines the resolution of a system, larger lens diameters can focus light to smaller spot sizes with tighter focus and reduced depth of the focus than smaller apertures. Cell phone lenses usually have 3mm diameter size , on the other hand, the average DSLR lens have 3 cm diameter size. Therefore, the diffraction limit of the lens system in a cell phone is in 10 times smaller than a regular camera. Although there is a misconception that a higher number of pixels equals to a higher resolution, what happens in reality is that the resolution is defined by the diffraction limits of the lens system. This means that, when the lens has reached its diffraction limits, the image will not get any sharper regardless of how many pixels a camera has.

Additionally, the smartphone camera sensors' sizes are much smaller than a DSLR camera. A cell phone camera has a sensor size range of 18-24 mm^2 while a full frame DSLR camera has a sensor size of 864 mm^2. In addition, to accommodate a higher pixel count in such small sensors, the smartphone cameras manufacturers had to push the shrinkage of the pixel size to a very small range - from 1.4 µm down to 1.2 µm. The reduced sensor size and shrunk pixels limit the phone incoming light exposure area, causing it to be less sensitive to darker scenes, and creating challenges to capture dark-field images.

The smaller hardware size intensifies another problem found in cell phone images: the high level of noise. This is in fact the greatest challenge of analyzing permanent defects in cell phones images. There are some factors that cause cell phones to have greater noise content than the DSLRs. One of them is the design structure and fabrication process of the pixel cell. The very small pixel design mentioned above is less immune to the system noise and consequently accumulates more noise than DSLR cameras. Additionally, cell phone cameras do not employ the complex and effective noise suppression algorithms that are built-in in most DSLRs. Cell phone cameras have other unique limitations that are not present on DSLR cameras. For example, the biggest source of noise in a cell phone is the thermal noise that comes from the heat generated by the phone's other functions. In fact, all the communication systems, the LCD and the constant use of other phone applications create a considerable amount of heat. This was a new challenge we encountered during this experimental process because dedicated cameras do not generate thermal noise at the same intensity and pace as cell phone devices, since DSLRs cameras do not have any applications running in parallel they only have camera related processes. Indeed, only when the image is being captured and being read out is when a dedicated camera is dissipating heat. For DSLRs, adding a pausing period between shots was sufficient to minimize these thermal noise impacts. However, for cell phone cameras we had to develop a new methodology that would reduce the heat dissipated by the device. This experimental process will be described in this section.

Another major limitation found in cell phone cameras is the extraction of the true digital RAW image. The output of RAW images has just become available in Apple iPhones in their latest model, the iPhone 12 and it is only supported for newer Android models. It is fundamental to have access to the true digital RAW of an imager to be able to identify and analyze its permanent defects. Another obstacle found in the RAW version of smartphones is the lack of noise suppression software that reduces some of the inherent noise. Hence, identifying and analyzing hot pixels in the RAW format of cell phone cameras is a non-trivial task.

Another major drawback is the modest exposure times and ISO speeds available in cell phone cameras. During this research, we found that smartphone images captured using exposures times longer than 2 seconds and using ISOs higher than 800 contain too much noise, making it impossible to search for permanent defects. This sensitivity is significantly smaller in comparison to the DSLRs that are capable of capturing images in very high ISO speeds of 25600 or greater and taking images of 30 seconds or longer. These manufacturing limitations not only show that cell phones cameras must undergo a great amount of improvement to be compared to DSLRs, but that also, according to our research, it limits the range of ISO and exposure times combinations we can analyze in these devices.

Therefore, to overcome all these manufacturing limitations and effectively detect and analyze hot pixels, we have developed an adapted methodology to capture cell phone dark-field images. The first step is to attempt to reduce the largest amount of thermal noise possible. As mentioned before, cell phones create heat rapidly because of their multiple functionalities. Thus, prior to the image capturing process, we used the 'airplane mode' feature where all the communications systems on the cell phone are turned off and blocked. This stops the communication processes that generate more heat. Then, we decreased the brightness of the cell phone to the lowest level available because one of the main sources of battery drain and heat dissipation in a cell phone is the usage of the OLED (Organic Light-Emitting Diode) displays. We only lowered the brightness to the lowest level because the option of turning the display completely off was not available on the cell phones we tested.

These steps will guarantee that the phone does not produce more heat as pointed in our previous work [53, 54, 57]. However it was necessary that we minimize the heat already existent in the device before commencing the tests. Therefore, after turning off the cell phone communications and reducing the brightness, we placed the phone in a refrigerator that has an inside temperature of approximately 4º C for 10 minutes. This step is essential because it lowers the internal temperature of the cell phone to approximately 16º C. After allowing the phone to cool for 10 minutes before the test, we kept the phone inside of the refrigerator to perform the whole image capture process. This step ensures minimal heat will be generated during the image capture process and give us an environment with no incoming light. After the phone has been cooled and before starting the tests, it was necessary to cover the phone's screen and the lenses of the camera with dark covers. This is crucial as the small amount of light from the screen could enter the camera sensor during image capture and affect the dark-field image reliability. It is important to note that we chose 4º C for the internal refrigerator temperature because any lower temperatures will bring the phone to a freezing level. And, because of the phone's hardware protection features, if the device reaches close to freezing level, the phone will automatically shut down since the battery is not allowed to freeze as it will lose power and stop working, and that will interrupt the experimental process, which is undesirable.

Another difficult challenge of the experimental process in cell phones is the image capture itself. As you may recall, the calibration process for DSLRs was done manually in a photography dark room or inside a covered box with waiting periods to prevent the camera from generating heat. However, this methodology is not transferable to cell phones because the device is inside a refrigerator for the whole capturing process. We searched for solutions to work-around this challenge and we developed two different methods. The first one was to use the Vysor app to trigger the phone camera remotely via a USB connection from an external computer. The Vysor application [59] is a Google Chrome extension for computers along with an Android app. It enables the user to display and interact with their Android smartphone from their personal computer. The operator can use the camera, apps, games, and even type using their computer keyboard. In our research, we used this app to control the cell phone camera. This enabled us to keep the device inside the refrigerator throughout the whole image capture process. Figure 3.9 illustrates the Vysor app connected to an Android phone and being controlled by a computer.

Figure 3.9 The Vysor App controlling a smartphone

The experimental methodology with the Vysor app has the following workflow:

1. Download and Install the app in the Android phone that is being analyzed.

2. Download and Install the Google Chrome extension on the computer that will be used to control the cell phone.

3. Enable the USB debugging option on the cell phone settings.

4. Connect the device via USB to the desktop.

5. Once connected, launch the Vysor extension from the Google Chrome apps button.

6. When the Vysor app opens, click on “find device” and select the cell phone that is being analyzed.

7. When the Android home screen from the phone opens, click on the camera icon

8. Select on the camera settings to output RAW image files.
9. Select the first exposure time and ISO.
10. Capture the image.
11. Wait 30 seconds.
12. Select a new exposure time.
13. Repeat the process until all the exposure times and ISOs are finished.

This experimental process allowed us to capture all the dark-frame images necessary for the calibration set without removing the phone from the refrigerator or adding any extra heat in the device. We captured dark-frame images using a range from ISO 100 to ISO 800, using exposure times from 1/125 of second going up a photography step (*i.e.* doubling the previous exposure time) until we reached the 2 seconds shutter speed. The pictures will be stored in the phones internal storage and when finished, they will be transferred to a computer to be analyzed. Overall, this procedure enables us to capture permanent defects in dark-frame images for cell phone cameras. However, since it is done entirely manually, it can take a few hours until completion. In recent camera apps, the developers started to enable intervalometer features, but they usually do not support RAW formats.

As mentioned previously, cell phone cameras are now the number one choice for the majority of the population when it comes to choosing a photography device. Also, smartphone manufacturers are starting to invest more time and resources to improve cell phone camera sensors, as the quality of a camera has become one of the most important selling points in the smartphone market. Thus, it is fundamental to understand the hot pixel behaviour in those devices and how it affects the overall quality of the images captured using cell phone cameras. Additionally, smartphone cameras enable us to comprehend the impact of pixel shrinkage and reduced sensor size on the behavior of hot pixels.

In summary, dark-frame images are the experimental method chosen by this research to collect the necessary data to explore the formation of hot pixels. However, to detect and analyze the behaviour of these defects, it is necessary to have an automated process that can properly identify them. The next section will discuss the hot pixel software algorithms developed in this research that we have used to detect and analyze them. Also, it will explore the solutions found to detect permanent defects on images with high noise content, as noise is a constant problem for both DSLRs and cell phone cameras.

3.6. Hot Pixel Detection Algorithms

So far in this chapter we have discussed the foundation of permanent defects and hot pixel theory. We have explored the definition of a hot pixel and the differences between the various types of hot pixels and other image sensor defects. We also have discussed the experimental process to collect dark-frame images for both DSLR and cell phone cameras. In this section, we will discuss the computer algorithms developed to detect hot pixels. As we have discussed, modern cameras have a high pixel count; they range from 12 million to 50 million pixels, and analyzing such a large amount of pixels manually can be very difficult, nearly impossible. Therefore, throughout this research we have developed and tested different approaches for software-based algorithms to detect hot pixels effectively. This section will explore the algorithms developed in the past and their efficiency. We will also discuss the reasons for changing and refining previous algorithms that did not provide the most accurate results and the new approaches that we have developed. One of the main reasons is, as we have discussed, the high amount of noise present in both DSRLs, at higher ISO levels, and cell phone cameras at modest ISO levels.

To create these computer algorithms, we decided to use MATLAB. Its native image toolkit provides the users the ability to input RAW images data and read/manipulate their pixel values. Another advantage of using MATLAB is the coding structure of the software. We took advantage of MATLAB's built-in parallelism while developing the hot pixel detection software; this feature increases processing performance and supports a smoother work-flow when compared to the implementation of nested loops or conditional statements. Also, the native functions in MATLAB decrease the algorithm processing time - usually, it took just a few seconds to run the defect detection analysis for a given ISO set. Thus, taking into consideration the advantages mentioned, we historically have been using MATLAB to develop the detection software algorithms for this research.

The three main algorithms we have explored are (from oldest to newest): The Threshold Method [12] and The Nearest Noise Criterion Method [61]. The analysis results of each method will be compared and discussed in detail towards the end of the chapter.

3.6.1. The Threshold Method

The detection of hot pixels is a process that has been refined throughout this research. We have taken in consideration different aspects of the imagers and the detection methodology to develop the best approach to properly identify them. One of the first methods explored early on this thesis, and most obvious, was the threshold method [11,12,60]. The threshold metric is a software algorithm developed in MATLAB, it separates and categorizes false hot pixels, regular pixels and noise and only selects and extracts the value of true hot pixels by using a nominal threshold. The software also does a regression fit to validate true hot pixels, generates a list of the detected hot pixels and for each candidate, hot pixel intensity vs. exposure time plots. Additionally, the information about the pixel colour type is also provided.

The threshold method performs in a sequential manner. First, each image calibration set has to be divided into ISO groupings. As you may recall, for each ISO set there are 20 to 100 images darkfield with exposure time ranging from 1/125th of a second to 4s. At a high level, the threshold method algorithm reads the values of each individual pixel in an image, then it applies a fixed threshold as a first test to separate noise and good pixels from potential hot pixels. Afterward, it uses statistical methods of linear regression fit to confirm if the candidate pixel is a true or false hot pixel. We perform this algorithm in these two steps because computationally we cannot apply the linear regression analysis in all the 50 megapixels, it would take a long time to do it. Therefore, we explored a way to identify a filter that would only pass a few thousands of potential hot pixels that we would do more detailed analysis on.

In more detail, the software works as follows. The first step is the extraction of all the pixel values in a set of *N* images. The software will read picture K_i (*i* ranges from 1 to *N*) and then will start extracting the pixel value for each pixel address *P(x,y)*. Next, the software starts collecting the pixel values of the next image K_{i+1}, the reading process is done until all the pixels from all the images of the set are read (image $K_{i+1 = N}$).

After the reading process, the software does a linear regression analysis of intensity vs. exposure time for each pixel address across the sensor. Then, at the end of the curve generation, the algorithm will apply a nominal threshold (usually 5% of the saturation value) on all pixel curves. The 5% of the saturation threshold was defined based on previous experiments [12, 19, 20]. This is the reason why this algorithm is called the threshold method, because this software simply filters out pixels that fall below this threshold, as they could be either good pixels or noise variations. The software discards the majority of the pixels and the rest are considered hot pixel candidates.

Next in the algorithm pipeline is the identification of true hot pixels in the candidates' list. As we have discussed, hot pixel defects have a different output behaviour when compared to transient defects, noise or good pixels. Recall Figure 3.6, where we have displayed the difference between a hot pixel and a transient defect. From that image we can observe that the same hot pixel address will display an increasing output value with longer exposure time, whereas noise and transient defects will show a defective behavior in only one or two images of the sequence but not in the entire exposure set. The software analysis will look for true hot pixels by applying this fundamental principle. For all the hot pixel candidates, the algorithm identifies true hot pixels and discard false ones by using statistical methods of linear regression fit to the hot pixel model that is described in Equation 3.3. For a better understanding of this procedure, we applied the regression fit to the raw intensity values of a true hot pixel and a false one in Figures 3.10 and 3.11.

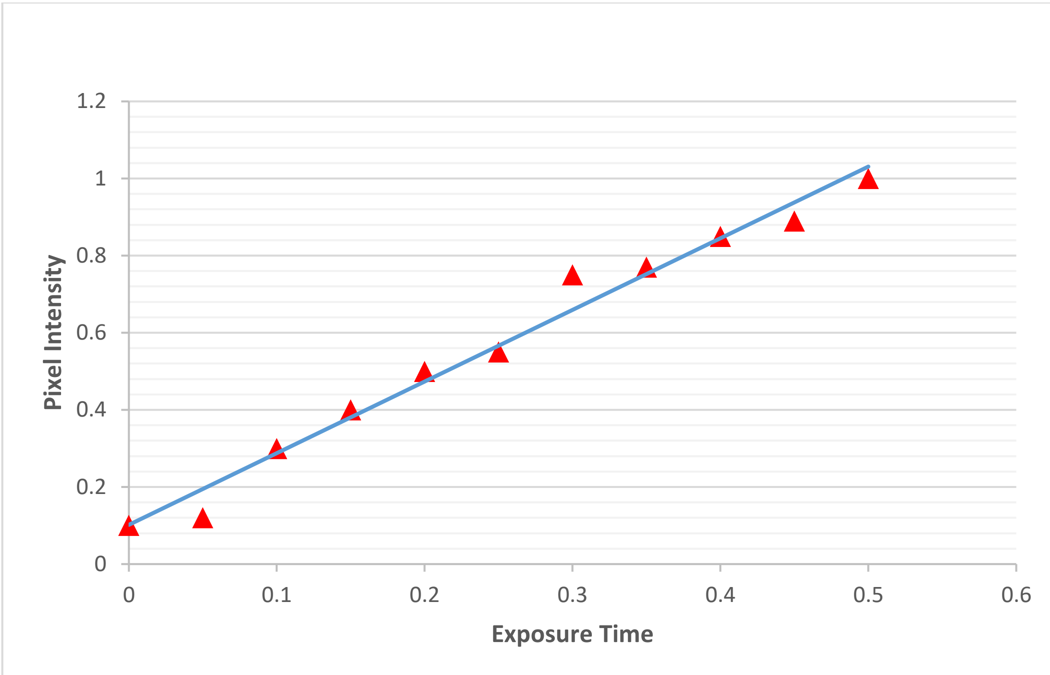


Figure 3.10 Fitted intensity curve of a true hot pixel ($y = 0.1018 + 1.8581x$ $R^2 = 0.9805$)

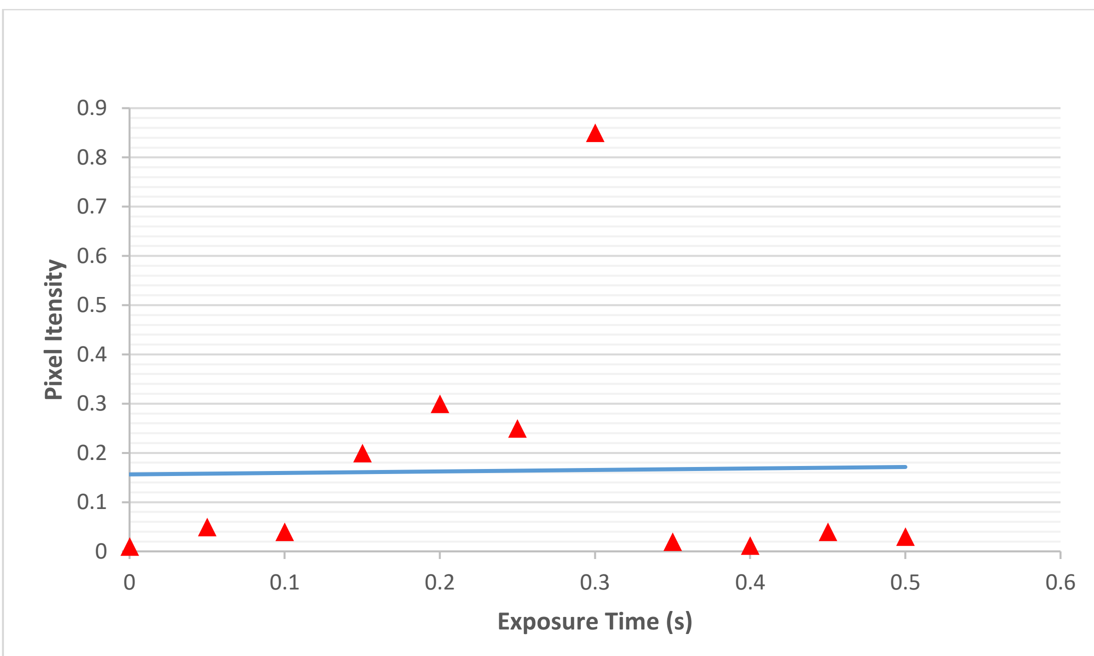


Figure 3.11 Fitted intensity curve of a false hot pixel ($y = 0.1562 + 0.03018x$) $R^2 = 0.0004$)

The graphs above allow us to visualize the difference in the data patterns for a true (Fig. 3.10) and a false hot pixel (Fig. 3.11). In Figure 3.10 we can observe the raw values and the regression fit for a true hot pixel produce a linear growth (slope) with exposure and an initial offset analogous to the one represented in Figure 3.6(a). On the other hand, when we analyze Figure 3.11 we can observe that one of the data points are random in exposure time with some points distant from the curve's slope and offset. This example would be considered a false hot pixel, similar to the one illustrated in Figure 3.6 (b), as it is clear that the majority of the fitted slope relies on the same pixel output range and only one outlier has a strong intensity value. The pattern depicted in Figure 3.11 is similar to the behavior of an SEU event or a noisy pixel.

In the software, the statistical identification of a true hot pixel is performed by analyzing linear regression fit calculated using the format expressed by Equation 3.6:

$$y = a + bx \tag{3.6}$$

The software uses the linear regression fit of the raw data. The algorithm then checks the standard deviation (covariance) of the raw values on the fit parameters *a* and *b* to evaluate whether the potential defective pixel is a hot pixel or not. Hot pixels can be identified by measuring the slope's curve and assessing its intensity. In this case, a potential hot pixel is validated when the regression fit for this pixel indicates a curve's slope value, *b*, greater than 3 times of the fitted slope error, Δb. The same 3 standard deviation criteria is applied for the offset, a. Mathematically, we performed this step by calculating the ratio between the slope and the slope error, also known as the t-ratio. This is expressed in Equation 3.7 and it gives a 99.7% confidence level that the slope/offset is statistically significant.

$$\frac{|b|}{\Delta b} > 3 \tag{3.7}$$

Using the true hot pixel example in Figure 3.10, we can observe the effectiveness of the curve fitting method. The linear regression curve for the data in Figure 3.10 is y = 0.1018 + 1.8581x, and the error of the slope (Δb) is 0.08744 and the error calculated for the a coefficient (Δa) is 0.02586. Applying these values in Equation 3.7, we have found the results expressed in Equation 3.8 and Equation 3.9. For the b coefficient, we found a t-ratio of 21.25, and for the a coefficient we found a t-ratio of 3.94. It is clear that the regression fit for this example is above 3 standard deviations and therefore have a strong significant statistical value. For a confidence interval of 21σ there is a 6.6×10^{-98} % probability of this being random and for 4σ it is a 6.7×10^{-5} % probability of occurring randomly. Therefore, this regression fit is considered a true hot pixel. Note a pixel is considered hot if either the slope b or offset a pass the t-ratio test.

$$\frac{|1.8581|}{0.08744} = 21.25 \tag{3.8}$$

$$\frac{|0.1018|}{0.02586} = 3.94 \tag{3.9}$$

Conversely, a hot pixel candidate will be discarded if the regression fit reveals a curve slope value b, and an offset a, that is less than 3 times the fitter sloped error, Δb, or the offset error Δa. This is represented mathematically by Equation 3.10 and is 3 standard deviations or 0.3% probability of occurring randomly. For the false hot pixel example illustrated in Figure 3.11, the regression fit is expressed by y = 0.1562 + 0.03018x, the error of the slope (Δb) is 0.50323 and the error calculated for the a coefficient (Δa) is 0.14886. Applying these values into the t-ratio formula, we have found for the b coefficient, a t-ratio of 0.0599, and for the a coefficient we found a t-ratio of 1.0498. Therefore, in this case, as the t-ratio for both coefficients are far from 3 standard deviations, the probability of the slope R_{dark} being zero or not having a significant statistical value is too high for the pixel to be considered a true hot pixel. Thus, it will be discarded from the hot pixel list. Also in that figure, it is clear that the slope curve does not have a great statistical significance, as you can see the data points are randomly distributed. In summary, the algorithm monitors the slopes values to look for statistically significant slopes with exposure or offsets. The point here is that the purely noise related will not show this linear relationship with exposure time but rather a much more random relationship.

$$\frac{|b|}{\Delta b} < 3 \tag{3.10}$$

The threshold method follows the flow chart described in Figure 3.12. This work flow has the advantage of being a unified method that can be used for any camera that has RAW images output options. This was important in our research, since we have tested different cameras models from a selection of different manufacturers and devices ranging from high end cameras to consumer level DSLRs. The same flow was proven to be successful in all these cameras that varied in pixel size, ISO rages, sensor quality, age, and other parameters. Therefore, it eliminates the need of having different detection flows for each camera model or manufacturer. Another benefit of this flow is that it can be implemented as the foundation for future algorithms developments that adapts to the cameras' needs as technology advances. We expect this process flow to work on different cameras because noise generally have the same behaviour illustrated in Figure 3.10.

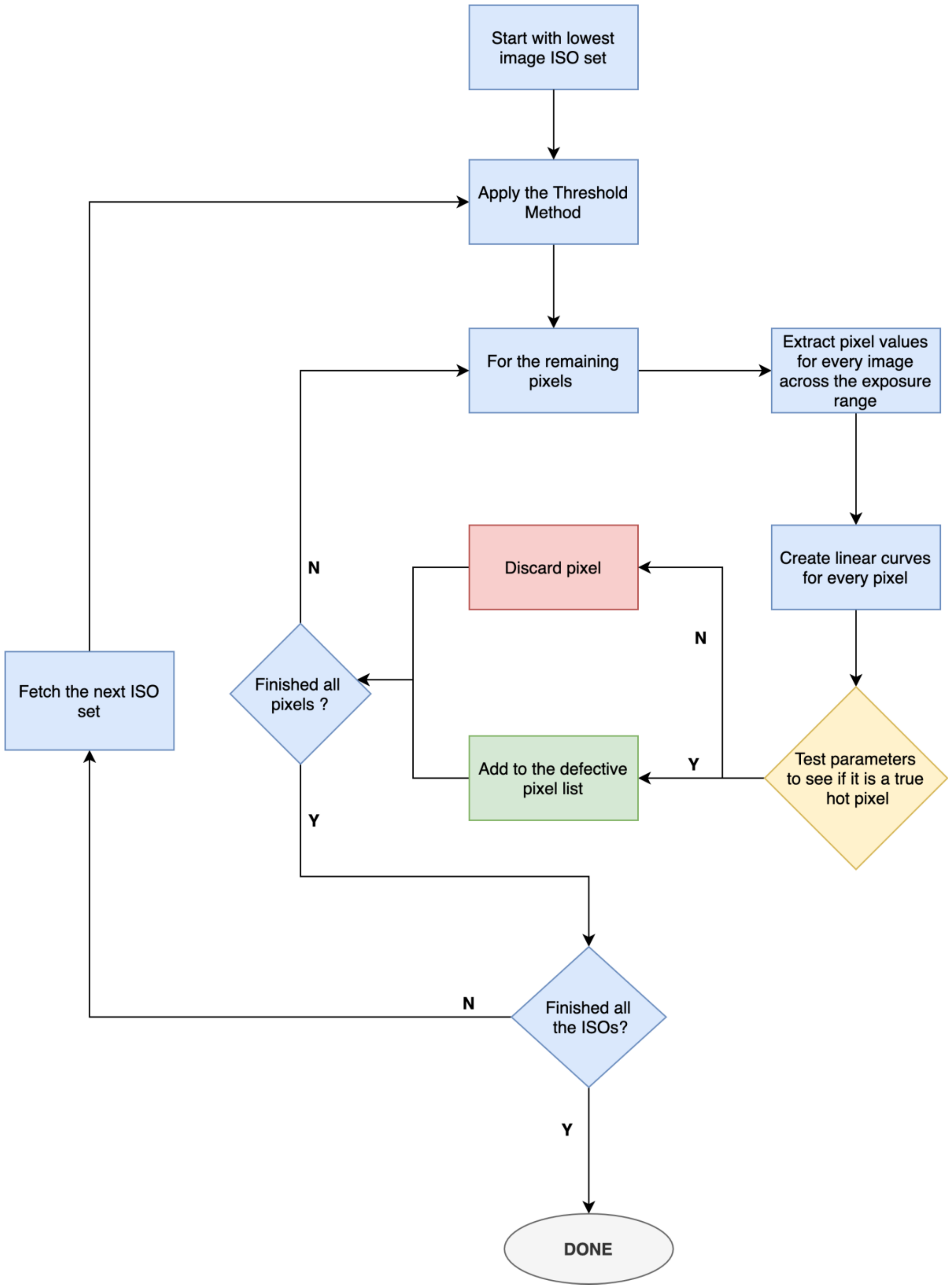


Figure 3.12 The algorithm detection flow

In summary, the threshold method has been used in this research as an attempt to take into account the noise floor of the image sensor. However, because of the pixel output fluctuation, due to the random behavior of the noise across the sensor, the amount of data filtered varies significantly. Also, we would need to apply a different threshold for higher ISOs where the noise content is greater. Although this method has shown successful results in detecting hot pixels in initial experiments, it is not refined to detect them in complex noise environments such as cell phone cameras or DSLRs in higher ISO speeds. Therefore, we tested different methodologies that are more sensitive to the noise across the sensor. One of these algorithms adaptations was the nearest noise criterion detection method that we will discuss in the next section.

3.6.2. The Nearest Noise Criterion Method

The nearest noise criterion method [61] was our first attempt to explore hot pixels where the threshold method was not successful – *i.e.* those image where the noise was too high. This algorithm takes in consideration the local noise distribution around a given pixel. As it has been discussed, the noise in a camera sensor is one of the major challenges for the hot pixel identification. From Figure 3.13, one important characteristic of noise that we can observe is that it is not uniform across the sensor, in fact, the noise components vary strongly across the image set. Therefore, to enhance the detection algorithm, we decided to compare a potential hot pixel to the local noise of its surrounding pixels.

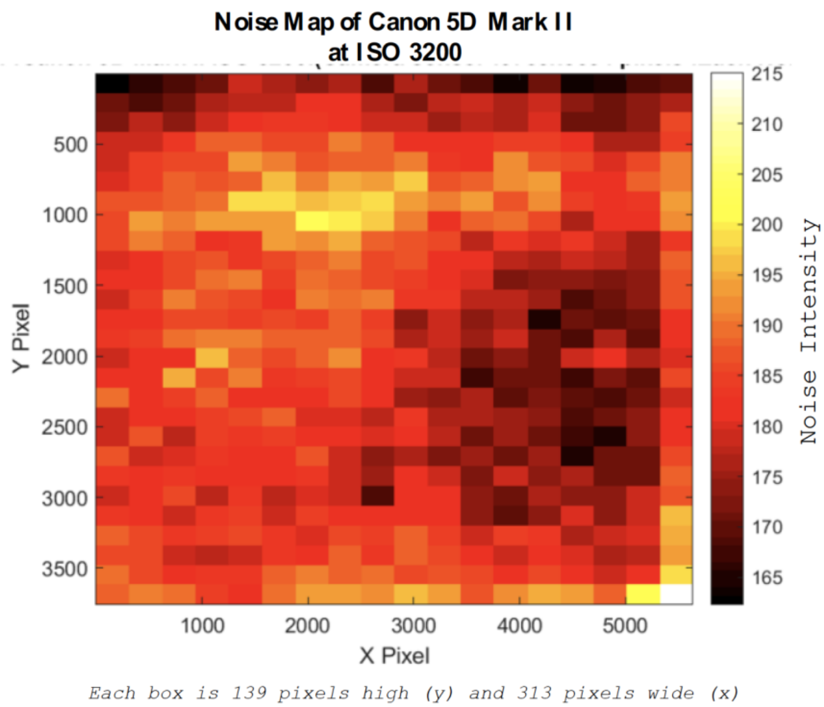


Figure 3.13 Noise map of Canon 5D Mark II at ISO 3200 (36 mm x 24 mm sensor) (taken from [62])

The nearest noise criterion method starts differently from the simple threshold method. First, instead of analyzing and extracting the pixel information of all the images in the batch at once, the algorithm only analyses a set of three images per time of the same exposure setting for the entire exposure range. The decision to investigate and analyze calibration sets of just three images instead of the entire batch at once is to understand how the local noise behaves around each pixel being tested, as noise variations can be observed in more detail across a smaller image set.

Another key difference from the threshold method is the detection criteria itself. The software does not apply a nominal threshold to eliminate the majority of the pixels, instead it evaluates the noise around each pixel and applies specific conditions to identify a potential hot pixel. However, before exploring these criteria, consider Figure 3.14 which diagrammatically displays the neighboring pixels in a 5×5 square around a potential hot pixel.

$P_{(-2,2)}$	$P_{(-1,2)}$	$P_{(0,2)}$	$P_{(1,2)}$	$P_{(2,2)}$
$P_{(-2,1)}$	$P_{(-1,1)}$	$P_{(0,1)}$	$P_{(1,1)}$	$P_{(2,1)}$
$P_{(-2,0)}$	$P_{(-1,0)}$	$P_{(0,0)}$	$P_{(1,0)}$	$P_{(2,0)}$
$P_{(-2,-1)}$	$P_{(-1,-1)}$	$P_{(0,-1)}$	$P_{(1,-1)}$	$P_{(2,-1)}$
$P_{(-2,-2)}$	$P_{(-1,-2)}$	$P_{(0,-2)}$	$P_{(1,-2)}$	$P_{(2,-2)}$

Figure 3.14 5x5 square of the near neighbors of the hot pixel $P_{(0,0)}$.

The noise criterion method, like the threshold method, was implemented on MATLAB. The software starts by collecting the first 3 images of the set and extracts all of their pixel values. Then, the software analyses every pixel address and its neighbors across the 3 images set looking for hot pixel candidates. To decide whether the pixel being tested is a potential hot pixel or not, the algorithm executes the following steps:

1. It selects a pixel address to be tested. Then it stores the values of the same pixel address across the 3 images, resulting in 3 central pixels P(0,0) as illustrated in Figure 3.14.

2. It creates 5×5 pixel boxes around the central pixels being tested and analyses the noise statistics on the 24 neighboring pixels around each central pixel across the 3 images (P(x,y) for image K_i where x = y ≠ 0, i =1…3). This results in total of 72 data points for the noise information. Values were gathered as 16bit (i.e. 65,536 max).

3. It measures the mean (denoted by η) and the standard deviation (denoted by σ) of the neighboring pixels excluding the central P(0,0) pixels.

4. After gathering the information about the 3 central pixels and their neighbours, the algorithm checks if the pixel address satisfies the following rules:

    a) The 3 central pixel values have to be higher than all the 72 surrounding pixels. This is represented by Equation 3.9.

$$P_{(0,0)}\ (K_i) > \ P_{(x,y)}\ (K_i)\ \ for\ x = y\ \neq 0, \qquad \text{(3.9)}$$
$$for\ all\ i = 1..3$$

    b) The 3 central pixel intensities have to be greater than 3 standard deviations (denoted by σ) plus the mean of the noise of neighbouring pixels (denoted by η). This is represented by Equation 3.10.

$$P_{(0,0)}\ (K_i) > (\eta + 3\sigma), \qquad for\ all\ i = 1..3 \qquad \text{(3.10)}$$

5. If both conditions (4a) and (4b) are satisfied, the pixel location P(0,0) is considered a hot pixel candidate but unverified. The pixel address P(0,0) is added to the potential hot pixel list to be yet validated. However, if only one of the conditions is not satisfied or neither of them are, the pixel location P(0,0) is rejected.

6. Repeat the 5 steps above for all the pixel addresses.

7. When all the pixel addresses of the 3 images are finished, the algorithm gathers the next 3 pictures of the set with increasing exposure times and repeats all the 6 steps until all the images of the calibration set are finished.

After finishing the nearest neighbour criterion for all the pixels, the algorithm will validate the potential hot pixels found during the selection process. However, for this part the software will look into the calibration image set as whole, instead of just 3 pictures per time. The verification step is performed identically to the statistical methods applied in the threshold method. The software creates curve fit (intensity vs. exposure time) for every hot pixel candidate in the list, then it applies the same statistical linear regression criteria to the hot pixel curves explained in the previous sections. In order to validate if the pixel address being tested is rather a true or a false hot pixel, the software checks if the suspected hot pixel have a dark current (R_{dark}) and an offset *b* with sufficient statistical significant (*i.e.* a curve slope intensity greater than 3 times of the fitted slope error, Δa). Finally, the software adds to the final list all the hot pixels that passed the nearest neighbor noise criterion and were validated by the statistical analysis, but rejects all pixels the did not satisfied both detection verifications. Figures 3.15 and 3.16 are the plotted noise distribution for the near neighbors around hot pixels of two different devices. Note that the pixel values are expressed in a 16-bit range.

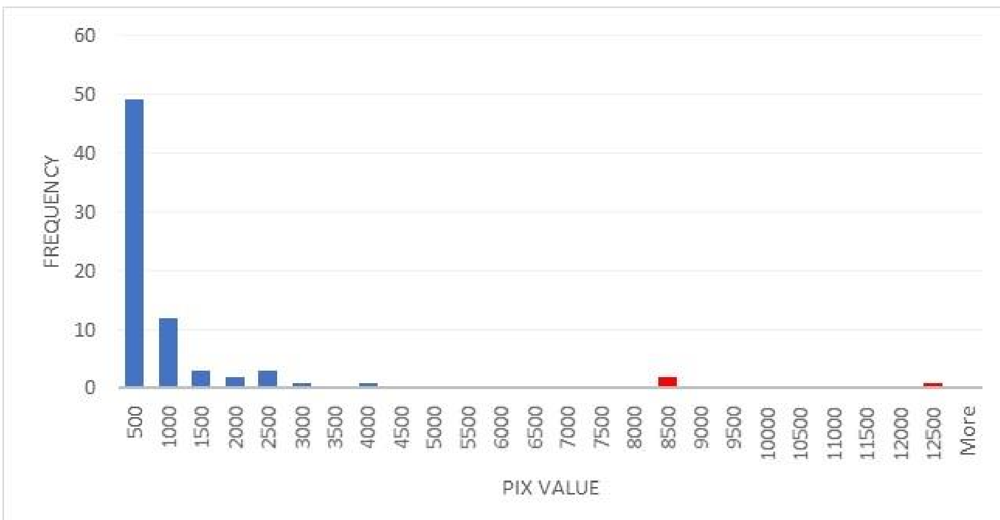


Figure 3.15 Near neighbor noise distribution (blue) with standard deviation of 745 and hot pixel $P_{(0,0)}$ (red) of typical minimum pixel (12.8σ) separation

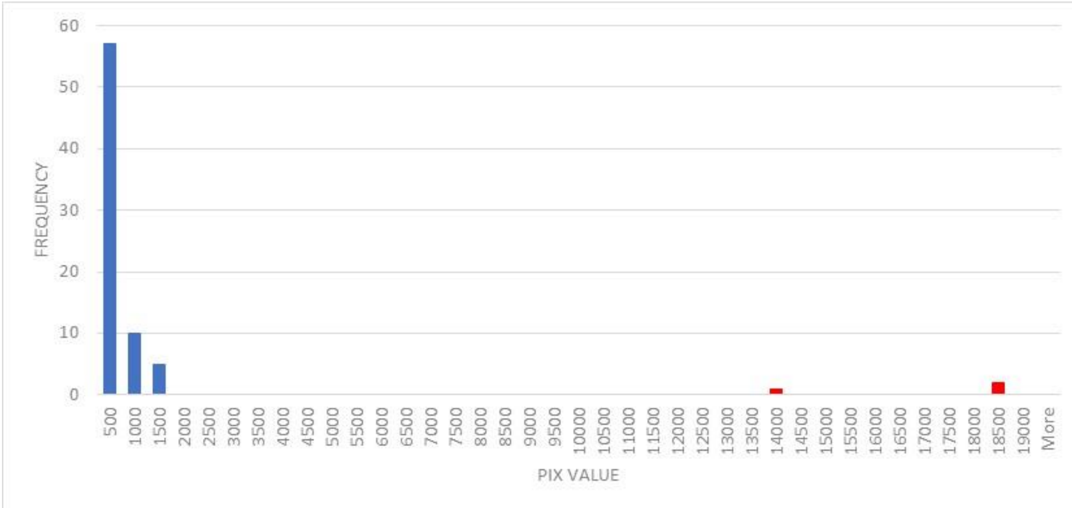


Figure 3.16 Near neighbor noise distribution (blue) with standard deviation of 369 and hot pixel $P_{(0,0)}$ (red) of typical minimum pixel (45.2σ) separation

The Figures above illustrate what we expected to observe when analysing the final results: a significant statistical separation of the hot pixel from the mean values of the neighboring pixels. The noise distribution (in blue) is much further apart than the hot pixel intensities (in red), this the behaviour of a true hot pixel, where its value is much greater than the values of the neighbouring pixels. In fact, for the camera device used in Figure 3.15 we can see the minimum statistical separation of 12.8σ and for the device used in Figure 3.16 observed a statistical gap of 45.2σ. All the results we found displayed hot pixels with very high statistical significance, pointing that considering the noise each pixels allows the software to be more sensitive to the noise across the sensor.

Although the nearest noise criterion method showed results that pointed out to the right direction, it still had its limitations. First, the software only worked in longer exposure times (2-30 seconds). This limits the overall comprehension of the image noise across the exposure range. Second, the software was not able to reliably detect weaker hot pixels. It only worked on images in the longer exposure range that contains a great amount of noise information, only hot pixels with strong statistical significance would be detected, excluding weaker hot pixels that we detected using the Threshold method. Thus, after analysing the results from this method we concluded that it is paramount to investigate and take into consideration the noise information around each pixel to properly classify if it is rather a hot pixel or not. However, the method was not able to adapt to the entire exposure range and was unable to search for weaker hot pixels across the sensor. Other students are exploring different areas in the hot pixel detection methods that will adapt to the different noise intensity levels found across the image sensor.

In conclusion, we did not find that the nearest noise criterion delivered superior results and for all the detection methods we tested, the Threshold method was the one that showed the most consistent and accurate results, that comprehended the entire exposure range. Thus, this was the method used to create the hot pixel growth model, which we will explore in the next chapter.

3.7. Summary

This chapter has discussed the definition of permanent defects, as well as it has outlined the methodology to detect hot pixels that we have developed throughout the years. We explored in detail the behaviour model of defective and non-defective pixels. The chapter focused on exploring the most common type of permanent defect, which is the hot pixel with an offset. Then we outlined the different methodologies used to collect dark-frame images from DSLR cameras to cell phones. One of the highlights of the chapter is the various detection algorithms developed to detect the hot pixels in the dark-field images, especially the Threshold method that has been the most effective detection algorithm that we tested throughout this research. The next chapter will explore the methodology used to analyse the hot pixel detection results and how the results has been refined through the years, also we will provide the hot pixel growth model for DSLR and cell phone cameras.

Chapter 4.

Hot Pixels Results and Analysis

4.1. Introduction

In our ongoing research, we have captured a series of image sets of more than 30 cameras over the past 15 years, each involving 100's to 1000's of pictures. These test cameras range from 7 μm pixel size (DSLR's) with large sensors (860 and 364 mm^2) down to 1.12 μm pixels in cell phone cameras with small 24 mm^2 imagers. The data extraction from these images was analyzed using the computer algorithm methods discussed in the previous chapter. We then we used this information to develop an empirical model for the defect growth rate of hot pixels. In this research, the two main parameters explored are the ISO amplification and pixel size S of the sensor array. This chapter will discuss the methodology used to develop the defect growth model. We will also walk through the results found in previous experiments and provide latest model using the results from Threshold method.

4.2. Experimental Methods to Measure the Relationship Between ISO, Pixel Size in Different Cameras

In Chapter 3 we described the methodology we used to both detect and extract the hot pixels data from an individual camera. After running these experiments in multiple cameras, we have collected a great amount of measurements that has to be statically analyzed in order to provide information on the hot pixel growth rate. In fact, in this part of our research we want to examine the relationship between various cameras parameters, in particular the ISO and pixel size, and the rate of which defects develop.

Hence, we have tested a wide range of cameras with different pixel sizes and ages. We have also used a broad range of ISOs. Additionally, we included cell phone cameras, as they provide valuable data for the range of small pixel sizes. This data has been built over time (~15 years) through the research performed on Simon Fraser University, under Dr. Chapman supervision. Table 4.1 displays a list APS sensor used in this research, with their pixel size and release year. Note that the release year on Table 4.1 refers to the year the camera model was introduced to market, does not necessarily mean that was the year it was acquired.

Table 4.1 List of APS imagers with pixel size

Camera Type	APS Camera	Pixel Size (in µm)	Release Year
DSLR	Canon EOS 10D	7.38	2003
	Canon EOS 300D	7.38	2003
	Canon EOS 350D	6.41	2005
	Canon EOS 5DMarkII	6.41	2008
	Canon EOS450D	5.19	2008
	Canon T1i	4.68	2009
	Canon T2i	4.30	2009
	Canon T3i	4.30	2010
	Canon T5i	4.29	2013
Cell phone	Nexus P6	1.55	2015
	Nexus 5	1.4	2012
	OnePlus One a0001	1.12	2014
	Samsung S6	1.12	2016

One important aspect showed in Table 4.1 is how the pixel size decreases drastically through time, especially in the cell phone category. In our research, depending on the age of the camera, we have taken anywhere from 1 year to 10 years' worth of data. Performing these tests in a broad range of cameras with different parameters and ages, allows us to obtain the data to analyze the relationship between pixel size, ISO and the rate of hot pixels increasing over time. This information enables us to stablish the hot pixel development rate. The next section will describe the statistical techniques we used to analyse this data.

4.3. Curve Fitting Methodology

Hot pixel defect rates are very important for predicting what is the effective useful lifetime of an imaging sensor. Since there is no accepted theoretical model making predictions, we have focused on using experimental imaging data to discover the trends in defect developments rates. In this research we have used statistical methods to develop the empirical equations that describe the hot pixel behaviour. Previous research found that the best method was to use linear least square approach in regression analysis to relate the hot pixel growth to the imager parameters. This research employed the techniques discussed in [75]. However, the issue with such an empirical approach is that there are many possible equation models. In this section we show the statistical process by which we identified the empirical equations that give the best fits.

These techniques focus on the analysis of residuals to assess the effectiveness of a curve fit. Residuals are defined as the geometrical distances calculated in the y-direction (vertical distances) between the observed data and the fitted regression line. A residual R is mathematically expressed by Equation 4.1:

$$R = Y_i - F(X_i) \tag{4.1}$$

where $F(X_i)$ is the expected value of the data point Y_i. The positive residual values indicate an underestimation of the fitted parameters. Conversely, negative values suggest that prediction in the model is too high for the observed data.

Residuals play an essential role in regression diagnostics, as they are a measure of how well a regression model fits a given data set. The easiest way to analyse residuals is through visual examination for model accuracy; this can be performed by plotting a graph of the residuals versus the predicted values. Figure 4.1 shows a regression fit for the data of a pixel output over a range of exposure times (hot pixel behaviour). Figure 4.2 displays the residual plot for this linear fit. Visually, the distance from the residual point to the *x*-axis indicates the divergence between the observed and the predicted values. Furthermore, to be considered an effective fit, the gap between the expected and observed values must not be predictable.

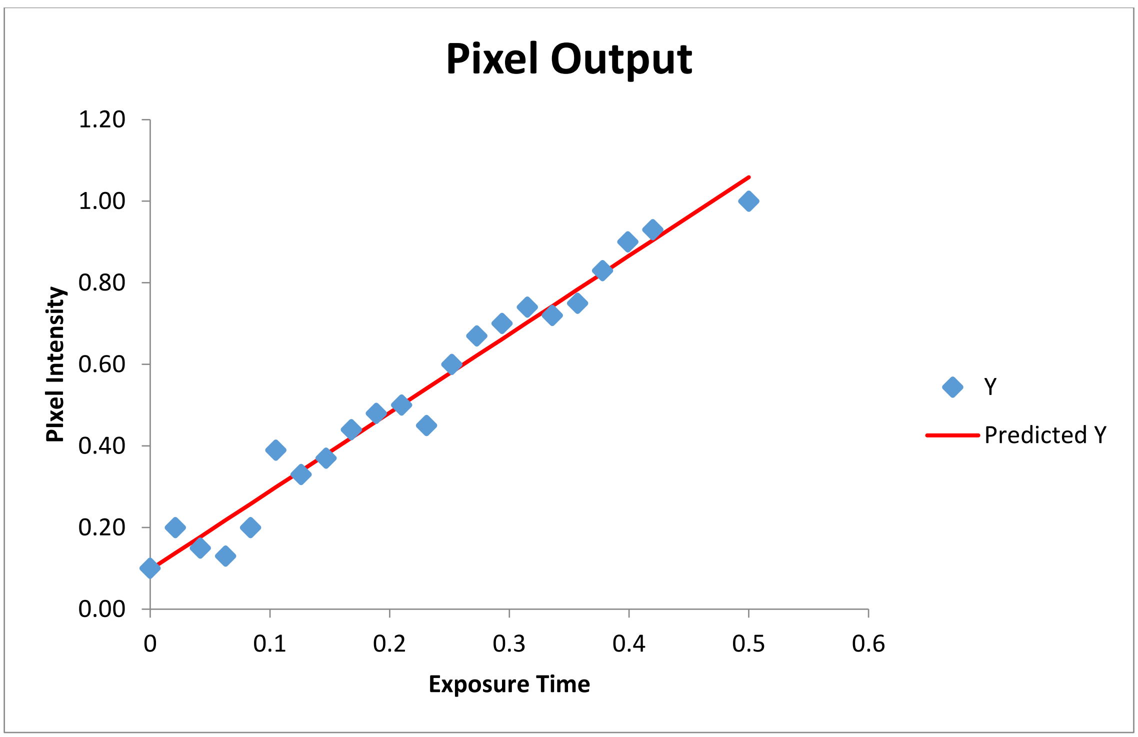


Figure 4.1 Pixel intensity vs Exposure time plot

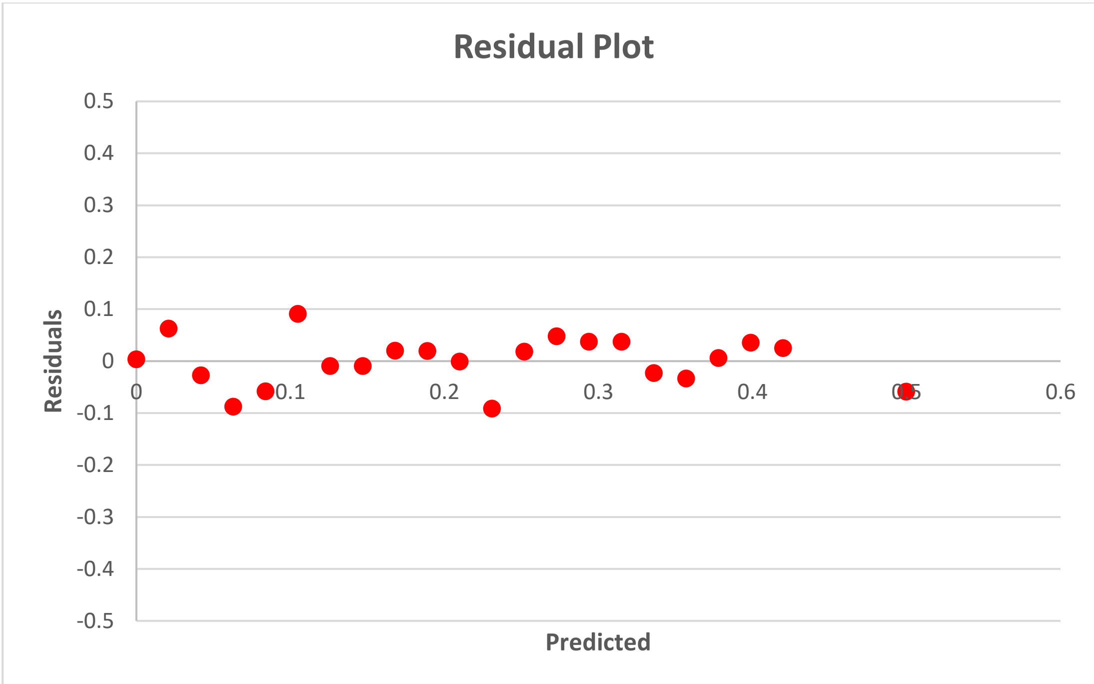


Figure 4.2 Residual plot for the pixel output regression fit

The examination of the residual plots enables us to determine whether the residuals are consistent with a random behaviour – that is there must be no pattern relating the fitted parameter (x) and to the residual value if the fit is correct. Furthermore, the general form and pattern of the residual plot can strongly indicate the overall effectiveness of a regression fit and the statistical significance of a prediction model. In general, a good linear regression fit is indicated by the following three characteristics:

- Residuals are randomly spaced in a gaussian distribution around the vertical axis with no indication of clear patterns;

- They show low error as they cluster around lower residual values, not large values. Ideally, the probability of the error is inversely proportional to the square of the distance.

These characteristics indicate that the developed linear regression model is appropriate to use with this data. Figure 4.2 illustrates an ideal residual plot that follows these characteristics. Conversely, the example in Figure 4.4 displays the residual plot of a regression fit of a parabolic data pattern (see Figure 4.3). Figure 4.4 illustrates a problematic residual plot – in this case a first order fit to a process that follows a second order equation. These types of residual plots are a warning that linear regression might not be the best fit for this particular data. Generally, problematic residual plots present the following characteristics:

- The residuals display a clear pattern, with a large collection of adjacent residuals on the same side of the curve;
- Plot is not randomly distributed on both sides of the x axis;
- Contains evident outliers – points far from the curve.

To develop the proper growth model for hot pixels, our research takes into consideration the fundamental definitions considered above. Our prediction models have been highly analysed for residuals with clear patterns, heavy dependence on either x or y axis and a great amount of outliers. Statistically we used a method called run errors that measures the probability of the data being clustered at the negative and positive sides of the plot, as opposed in being randomly distributed as shown in Figure 4.2. It is highly improbable that an effective fit will show a clear pattern such as the one illustrated in Figure 4.4.

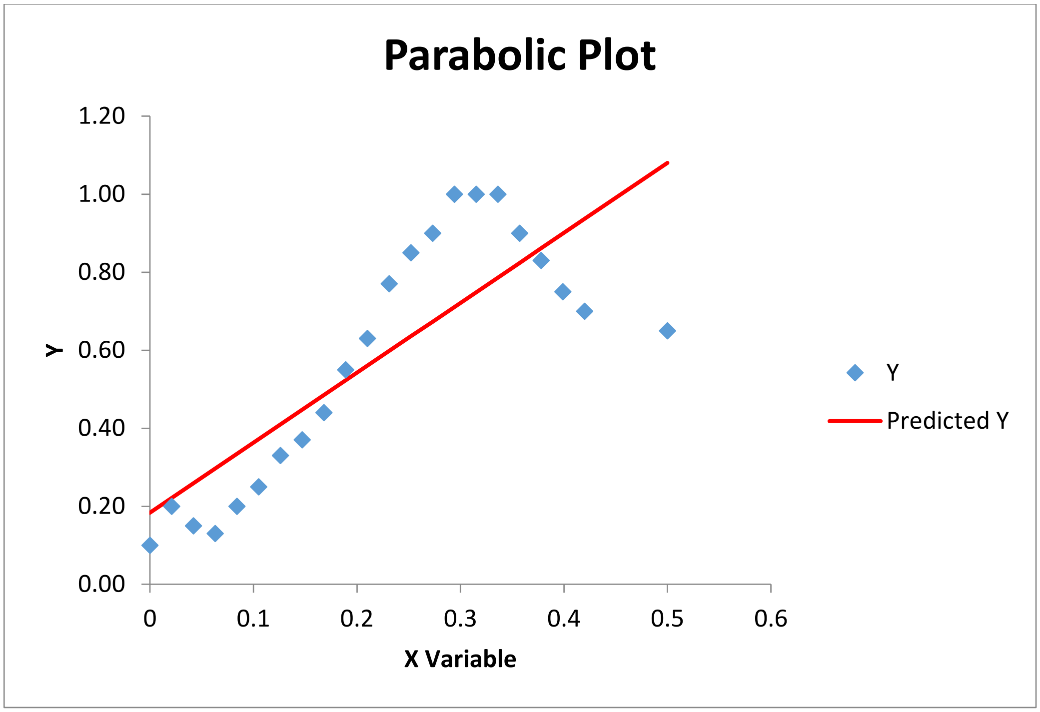


Figure 4.3 Example of a parabolic data pattern with regression fit

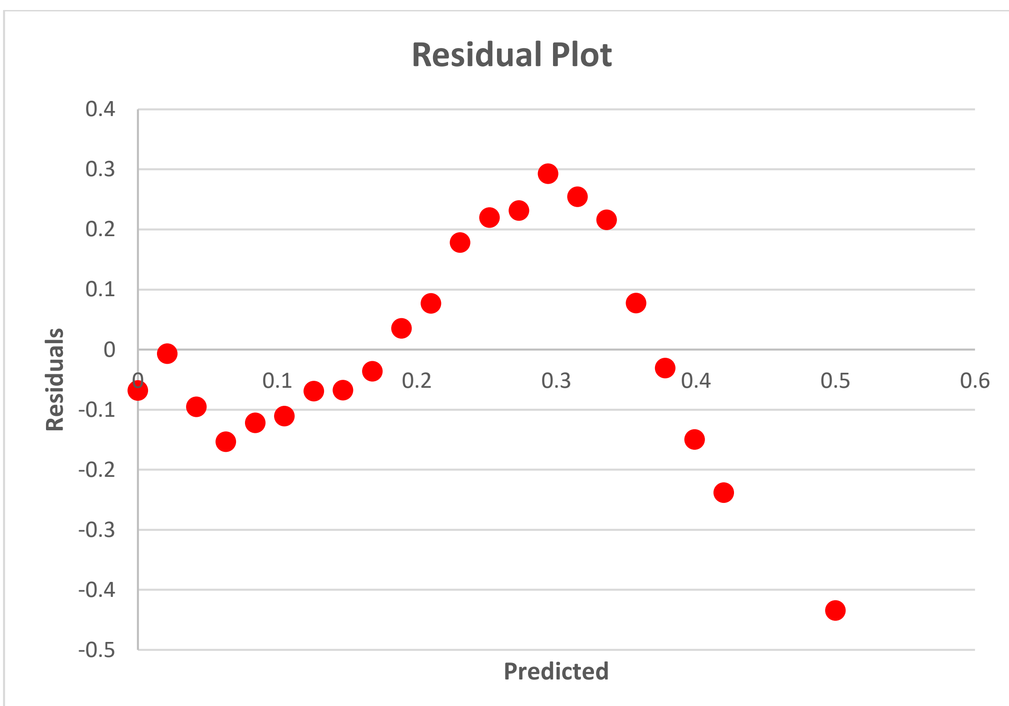


Figure 4.4 Residuals plot for the parabolic curve in Figure 4.3

Another important characteristic of residual plot diagnostics is the analysis of the actual size of the residuals. In a scatterplot, residuals that have random behaviour and contain low values indicate that the linear regression elaborated is a good fit, as the error is small. Figure 4.5 illustrates the residual plot of a good fit. On the other hand, Figure 4.6 displays a scatterplot with residuals that contain large values. Although this plot displays the random behaviour expected from residuals, its regression fit indicates that this data has a greater amount of noise because of the large error values. While developing our regression models, we analysed the residuals taking into consideration the pattern, the trend as well as the residual values themselves. Additionally, we applied the same methodology described in Chapter 3 to assess the quality of the fit. Therefore, we used the residual behaviour and the statistical significance of the parameters (*i.e.* the number of standard deviations of the fit parameters) to determine which fit is the correct one.

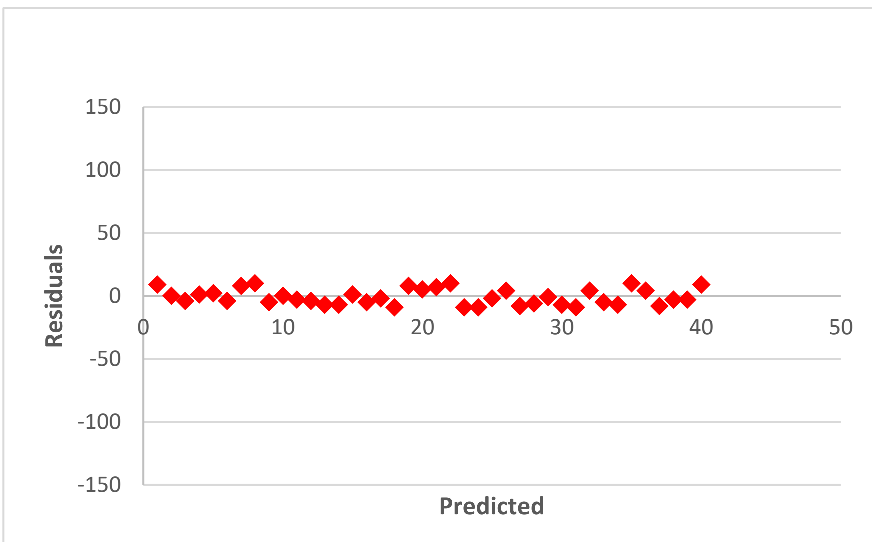


Figure 4.5 Residual plot with low error

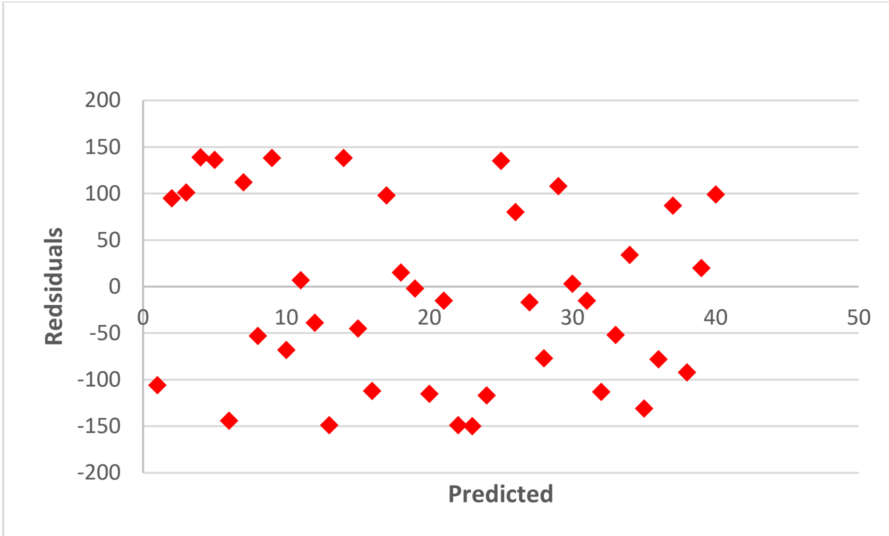


Figure 4.6 Residual plot with high error

4.4. Hot Pixel Defect Growth Model

This section will explore how the hot pixel defect growth model was developed and refined through the years. The statistical methods and tools discussed previously are employed when generating the defect growth model.

4.4.1. The Relationship Between ISO and Pixel Size

Over the past 15 years, the research on hot pixels has tested several cameras to develop an accurate model for the hot pixel defect growth. The cameras varied in age, sensor and pixel size, ISO speeds and sophistication. This large comprehensive data enables us to develop an accurate empirical defect growth model based on pixel size and ISO. The main goal is to derive a relationship that models the defect growth in terms of defects/years/mm^2. We established this relationship by using the fit method described in Chapter 3 which empirical equation best describes the hot pixel growth rate. These models have been refined through the years as new data has been collected and different detection algorithms have been developed. The initial research in 2012 [77] focused on generating a defect relationship between a fixed ISO and varying pixel sizes as displayed on Figure 4.7. The results showed a linear relationship in the growth model.

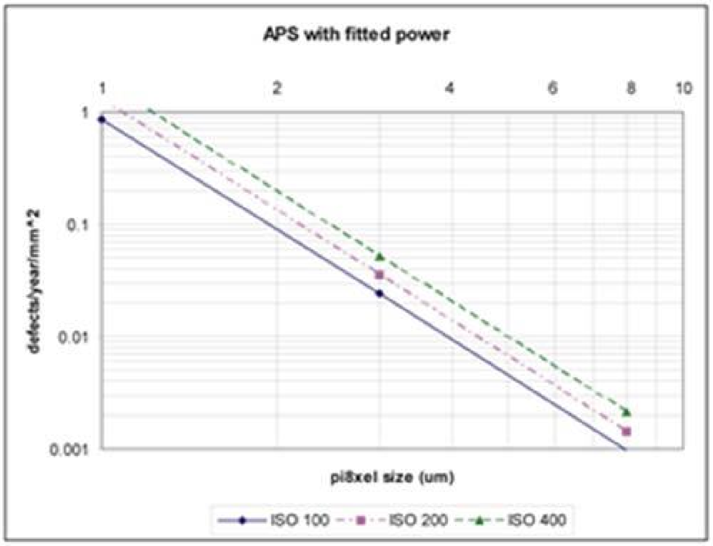


Figure 4.7 APS Defect rate/year/mm^2 vs pixel size for fixed ISO (taken from [77])

As we can observe from the Figure 4.7, the defect growth has a linear relationship with pixel size, as we will show it is inversely to about the 3rd power and increasing the ISO created the same slope but with different offset. This strongly suggests that there is a relation between the ISO and pixel size with the defect growth rate. Also, this was giving the clear message that as the pixel shrink, the defect rate increases much faster than the number of total pixels in a sensor. Therefore, these parameters are taking in consideration when developing the model for the defect growth rate.

4.4.2. Power Law

Previous research [12] used Datafit [76] to find the best fit relation for the hot pixel growth model. Datafit is a curve fitting software developed by Oakdale Engineering that suggests the best fit relation given input parameters, such as ISO and pixel size. The software suggested a power law form for the hot pixel growth model; a generic power law is shown as in Equation 4.2:

$$\mathrm{D(X,Y)} = 10^{\mathrm{A}}\mathrm{X}^{\mathrm{B}}\mathrm{Y}^{\mathrm{C}} \tag{4.2}$$

X and Y are general variable members of a power law that in this research has become the ISO and S (pixel size) variables for the growth model. The terms A, B and C are constants that are defined based on the nonlinear regression model. The term linear here refers to the equation being linear in its parameters (*i.e.* parameters can be separated). A check for linearity can be done by taking a derivative with the respect to the parameters of the expression in question. In this context, the main challenge of the power law form is that it is not linear and very hard to analyse because the parameters cannot be easily separated. Nonlinear curve fitting is prone to fit failures. Importantly, Equation 4.2 can be converted to a linear form by taking the logarithm of the equation. Using Equation 4.4 the parameters in Equation 4.3 can be separated and the general hot pixel model takes the formula expressed in Equation 4.4.

$$\mathrm{D(S,ISO)} = 10^{\mathrm{A}}\mathrm{S}^{\mathrm{B}}\mathrm{ISO}^{\mathrm{C}} \tag{4.3}$$

$$\log\big(\mathrm{D}(S,\mathrm{ISO})\big) = \mathrm{A} + \mathrm{B}\ \cdot \log(S) + C \cdot \log(ISO) \tag{4.4}$$

The resulting formula in Equation 4.4 is now in a form that can be used for linear regression and curve fitting. Residual analysis is also easier in this logarithmic form. This empirical formula predicts the defect density *D* (defects per year per mm^2 of sensor area) based on ISO and the pixel size S. Then, Microsoft Excel was used to develop the data fits.

4.4.3. The results for the Hot Pixel Growth Analysis Combining Pixel Size and ISO

We used Equation 4.4 in a least square fit to calculate the constants to be applied in the growth model. After exploring different parameters, we have found that we wanted a relationship between the pixel size and the ISO. We have performed measurements in cameras with different pixel sizes and in a wide range of ISO to properly stablish this model. Additionally, we have refined the defect growth model over time. Initial research [75] focused on larger pixel sizes (5-10 μm) from mainstream DSRLs to find the values of the constants for the growth model, and as we added more data from smaller pixels and different ISOs, the parameters were refined. However, as we show in this section they did not change considerably. The research in [75] developed separate relationships for CMOS and CCD sensor types; they are shown in Equations 4.5 and 4.6 respectively and the constants found in the fitted models with the error bounds are displayed in Table 4.2:

$$D_{APS}(S, ISO) = 10^{-1.13} S^{-3.05} ISO^{0.505} \tag{4.5}$$

$$D_{CCD}(S, ISO) = 10^{-1.849} S^{-2.25} ISO^{0.687} \tag{4.6}$$

Table 4.2 Power Law fitted constants with error bounds

Constant	APS	CCD
A	-1.13 ± 0.26	-1.849 ± 0.22
B	-3.05 ± 0.25	-2.25 ± 0.17
C	0.505 ± 0.081	0.687 ± 0.086

It is important to mention that Equations 4.5 and 4.6 were developed when the research was limited to DSLRs cameras with pixel size of (5-10 μm). However, this earlier model predicted a drastic increase of the defect rate when the pixel size falls below 2 microns. From Equation 4.5 we see that the defect rate grows inversely to approximately the third power of the pixel size, thus as the pixel size gets smaller the defect rate rises rapidly. Also from Equation 4.5, we see that when higher ISOs are used, the defect rate increases with the square root of the ISO. The research in [75] predicted a defect development rate of 12.5 defects/year/mm^2 at ISO 25,600. Figure 4.8 shows the plot for fitted power law distribution developed for the CMOS sensor. One important characteristic of this plot is that the curve at lower pixel size (<2 μm) and higher ISO levels are purely projected values, because at the time of capture (circa 2012) cameras with smaller pixel size did not support RAW images and/or had higher ISO with great noise content that could not be properly analysed. Therefore, additional research was performed to include the testing within smaller pixel sizes as the technology advanced.

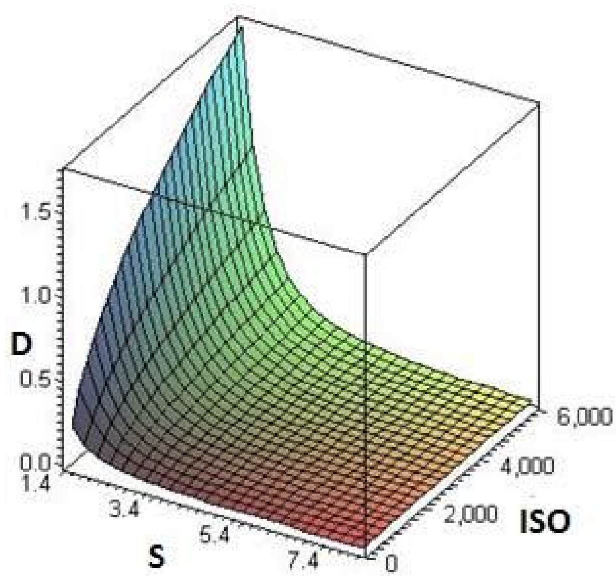


Figure 4.8 Fitted power law for APS defect density (D=defects/year/mm^2) vs. pixel size S (μm) and ISO (I) [74]

In 2016, research developed by R. Thomas and Dr. G. Chapman [12] expanded the ISO range and pixel size significantly. Their research included the previous camera calibration images collected and the new data sets captured from newer cameras. They were able to enlarge the research data in all ranges, especially in the 4 micron range at higher ISOs, such as ISO 1600 to 3200. However, the main contribution of the 2016 research [12] compared to the research in 2012 [75] was the inclusion of cell phone images in the hot pixel analysis. The previous research described two main issues in regards to cell phone testing: first, a small amount of phone models supports RAW outputs and the phone images contained a great amount of noise. To attenuate the noise content, they used a similar methodology described in Chapter 2. They were able to test cell phone cameras with pixel sizes that ranged from 1.5 to 1.1 μm, and ISOs from 200 to 800. This was a greater expansion compared to the research in 2012.

Additionally, they used the Threshold Method described in Chapter 3 and applied the statistical techniques we discussed previously to detect true hot pixels in cell phones. After identifying permanent defects that had regression line coefficients with strong statistical significance, they were able to update the defect growth model that included the pixel sizes in the 1 μm range. Figure 4.9 displays the expanded data fitted results for the CMOS sensor in the 1 – 2.5 μm pixel range. The analysis of the results from the extended data demonstrated that the defect development rate (defect/year/mm^2) for pixels with sizes in the 1 μm range is 100 time higher than the defect growth rate found for DSLRs in the 4 μm range. The results confirmed the predictions found in the initial research [75] and expressed in Equation 4.5. In fact, the updated model provided in the research performed by R. Thomas [12] showed that the constants of the defect rate curve only changed slightly from Equation 4.5. The refined model demonstrated that the predictions for the 1 – 2.5 μm range in the initial model trended in the right direction. The updated defect rate relationship is expressed in Equation 4.7 and Figure 4.10.

$$D(S, ISO) = 10^{-1.12} S^{-3.15} ISO^{0.522} \tag{4.7}$$

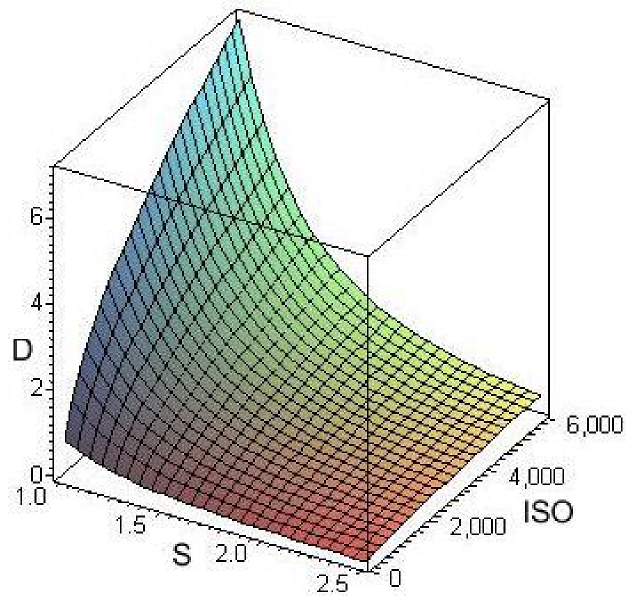


Figure 4.9 Expanded data fitted power law for APS in the 1 to 2.5 um pixel range: defect density (D=defects/year/mm^2) vs. pixel size S (um) and ISO (I) (taken from [12])

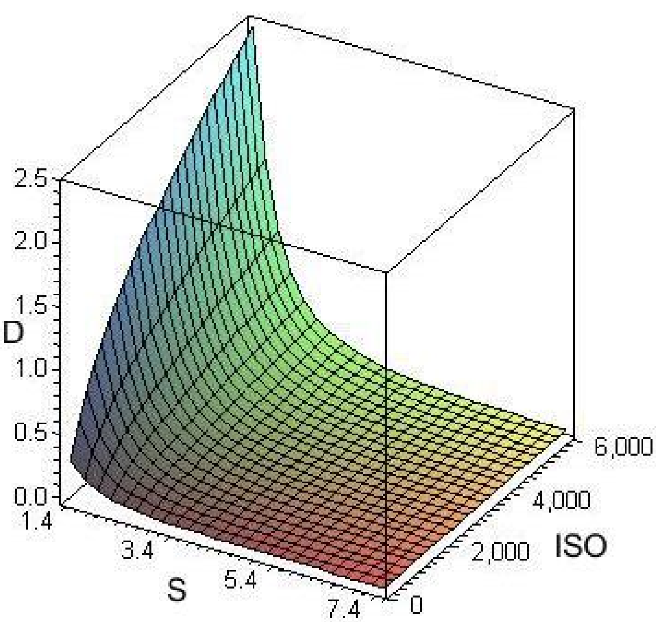


Figure 4.10 Expanded data fitted power law for APS defect density (D=defects/year/mm^2) vs. pixel size S (um) and ISO (I) (taken from [12])

In this research, the goal was to add additional data sets and refine the fit parameters, especially in the higher ISO area. We continuously performed the experimental capture and analysis to improve the hot pixel growth model. One of the venues we worked on was a new detection algorithm that used noise distribution to select hot pixel candidates, as opposed to the static Threshold filtering. This algorithm had two main steps: first, it calculated the mean and standard deviation of a camera in a 1000 image dataset of same exposure and ISO and then, using the dark frame exposure range, calculated the combined mean and combined standard deviation of the 8-neighbours of every pixel to select as potential hot pixel or noise. However, by the time of the conclusion of this thesis, we were still having problems with this software. In fact, we found that some of the known hot pixels were missing from the final list. Therefore, we did not use the results from this algorithm.

The updated hot pixel growth defect discussed in this chapter brings important implications for sensor designers, especially as they push the pixel size to submicron range. Additionally, as we have discussed, this is further exacerbated by using higher ISO speeds. Our defect growth model data indicates that defect numbers will become significant at small sensor areas (15 – 25 mm^2) within the few years of the lifetime of a typical cell phone. Furthermore, with imaging systems becoming part of many other products such as self-driving cars that have longer lifetimes (some car manufacturers have targets of up to 20 years in field usage). This has the potential to create other reliability issues. For example, it was discussed in Chapter 1 that A.I programs can be easily fooled with corrupted or noisy pixels. If hot pixels accumulate to the point that can affect the driving automation process, they will have a significant impact in performance, reliability and safety. For DSLRs, where sensors are much larger than the typical phone or what is applied in embedded systems, moving the pixel sizes towards 2 microns will significantly increase their defect rates even when noise suppression algorithms are applied. Additionally, the much longer lifespan of a DSLR camera combined with the user sensitivity to defects, make this a significant issue for them. The user sensitivity to defects and how other processes in the camera exacerbate the hot pixel damage area is what we will discuss in the next two chapters of this thesis.

4.5. Summary

In this chapter we discussed the statistical methodology we use to analyze the hot pixel data outputted from the detection computer algorithms described in Chapter 3. We also outlined the progress our research has made through the years in providing a refined empirical hot pixel growth model that includes smaller pixel sizes. In the next chapter, we will discuss how the demosaicing algorithms and the image file compressions such as JPEG will spread and exacerbate the hot pixel damage area.

Chapter 5.

How Demosaicing and Compression Algorithms Enhance Hot Pixels effects on Images

5.1. Introduction

In the previous chapters we discussed various concepts of hot pixels in digital cameras and defined the new methodology used to effectively detect and analyse hot pixels in higher noise backgrounds. Additionally, towards the end of the Chapter 4, we discussed the results found from the new algorithm and provided an updated hot pixel growth model with the new data. This chapter will continue to explore hot pixels. However, instead of analysing them from the sensor perspective, we will focus on how two steps of the image processing pipeline impacts the hot pixel damaged area. The two processes we will discuss in this chapter are demosaicing and the image compression and how they really enhance the damage created by hot pixels.

As was mentioned in Chapter 2, demosaicing algorithms are the method implemented to create coloured images in sensors that use CFAs (Colour Filter Arrays). Now, at this stage of the thesis, it is important to note that the CFA sensor does not differentiate hot pixels from normal pixels; consequently, the faulty pixel will distort the final image output. There is a common misconception (asked at conferences) that a few hot pixel defects in modern 10 to 50 Megapixel images do not affect the overall image quality and are not really noticeable by the end-user. However, as we will show in this chapter the demosaicing algorithm and the JPEG compression processes combined with the occurrence rate of the hot pixel creates a significantly noticeable degradation in the image. We will show that the demosaicing algorithm expands the hot pixel damaged area and the use of image files that compress the original data, such as JPEG, further exacerbates this problem to the point where this becomes considerably noticeable. This chapter will show that these processes can create defect clusters that make the hot pixel damaged area more visible than just a single pixel defect, and if two or more hot pixels are in closer proximity, the defect cluster can increase significantly. Therefore, it is paramount to discuss how these two processes affect the hot pixel damaged area and what their impact is on the overall image quality.

The goal of this chapter is to provide an overview of the most popular demosaicing algorithms, as well as to explore their impact on the hot pixel output. Additionally, it will discuss the effect that lossy file compressions such as JPEG have on the hot pixel damaged area. Lastly, the chapter will examine how two hot pixels in closer proximity can affect the overall image quality and what is the statistical probability at the level where it can occur in many cameras.

5.2. Demosaicing algorithms

In Chapter 2 we discussed how the standard CMOS pixels are sensitive to a wide range of light, from about 400 ηm to 1000 ηm, but to be able to capture colours the sensor is overlaid with what is called the Colour Filter Array (CFA). Using this system, each pixel can (ideally) only detect one of the three colours of the RGB by using thin film colour filters that pass a selected wavelength band to each pixel. While in an ideal sensor every pixel would have the ability to sense all the three RGB colours components, the reality is that most sensors cannot do this, and they rely on the CFA to be able to interpret the colour information of a scene. This was not an issue for the colour films as they were able to sense all the colour components across the film area. Some advanced pixel designs, such as the Foveon sensor, can replicate that, as they have three vertically stacked photodiodes to efficiently detect the true colour measurement for all the red, green and blue levels, but few cameras have this. However, for most imaging sensors, each pixel location is only able to sense one of the three colours as per the CFA layout. Figure 5.1 calls to mind the Bayer pattern mentioned in chapter two; this pattern is the most common type of CFA used in modern digital cameras.

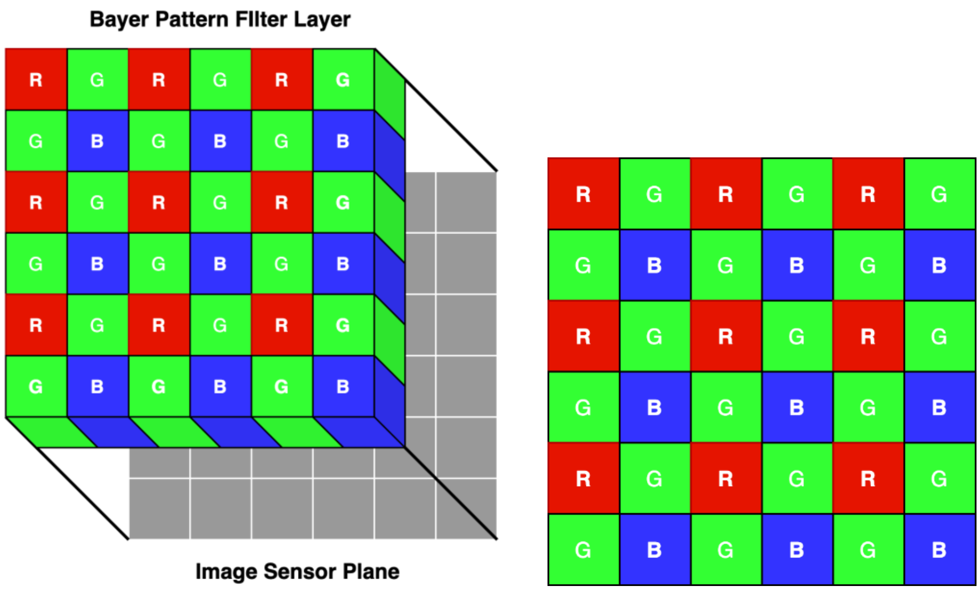


Figure 5.1 The CFA layout

Therefore, since the CFA layout has only one colour information per pixel, the digital camera systems use demosaicing algorithms that calculate the other two colour components of the RBG spectrum by interpolating the values from the neighbouring pixels. The demosaicing algorithms are a necessary step in producing colour images using a CFA sensor and enhancing the camera resolution. This process enables a digital camera to estimate a close approximation of the levels of RBG for every pixel in a captured image. The demosaicing is the first step in the image processing pipeline; its performance will impact the subsequent processes and it will affect the overall quality of the final image.

Each camera manufacturer has its own demosaicing algorithm and they are not open-source. However, most of the demosaicing algorithms discussed in the literature can be categorized into three types: Bilinear interpolation, statistical and adaptive. This section will explore one algorithm from each of the three categories and their impact on the presence of hot pixels. From Leung [11] we see they are:

- The Bilinear Demosaicing Algorithm
- The Median Demosaicing Algorithm
- The Kimmel Demosaicing Algorithm

One important point to mention is that these are not perfect filters; they are an attempt to replicate a coloured image that is as similar as possible to the original scene. One of the reasons these algorithms are not perfect is the fact that the colour response from the camera filters overlap with each other, as it does the human eye. An example of this behaviour is displayed in Figure 5.2 where we see the colour transmission curves for a Canon T3i. Therefore, these algorithms will result in quality loss, artifacts and other issues such as the moiré pattern. The moiré pattern in photography is interference that creates artifacts on edges or specific patterns that did not exist in the original scene. Another critical issue related to CFA that follows the Bayer pattern is the fact that the red and blue light has less information available than the green. As mentioned in Chapter 2, the RGGB Bayer pattern captures twice the information of green light in comparison to red and blue, which results in information loss for those colours. Later in this chapter, we will explore how the demosaicing algorithms affect the hot pixels. But before diving any deeper into the impact that these methods have on permanent defects, we will first discuss how these three demosaicing algorithms work.

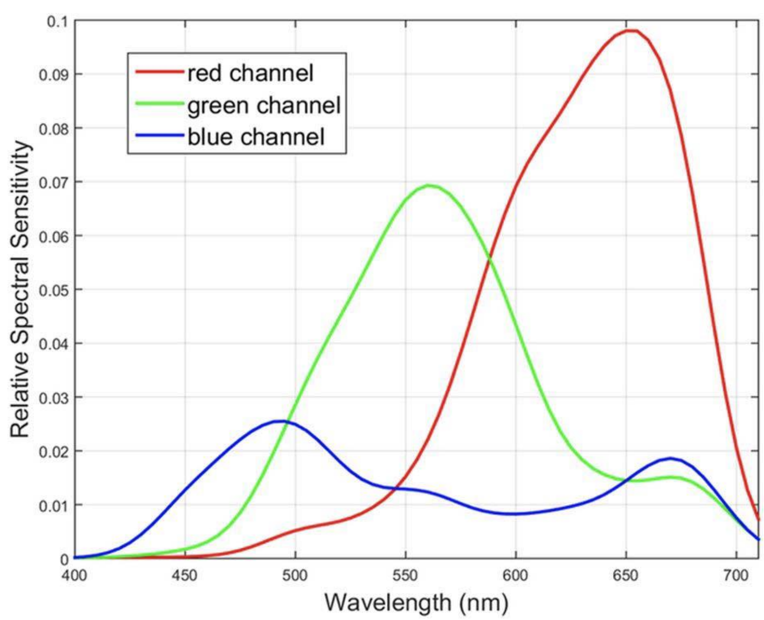


Figure 5.2 Canon T3i colour response (taken from [78])

5.2.1. The Bilinear Demosaicing

The Bilinear interpolation method is the simplest demosaicing algorithm and it is often used as the first step for other more advanced methods. As per the CFA layout, each pixel location is able to detect only one colour channel and will require the calculation of the two other colours. The Bilinear algorithm estimates the value of a missing colour in a given pixel by interpolating the colour values from the neighbouring pixels. It operates on the red, green and blue colours independently, therefore the calculation of each colour plane is a separate process. This demosaicing method has low complexity and can produce quick results; however, it suffers from quality loss and moiré effects. In fact, the Bilinear demosaicing is better suited for smooth areas where little variation is present because it generates significant artifacts across edges and other high-frequency content. Figure 5.3 displays the colour kernels used to interpolate the missing colour in a pixel.

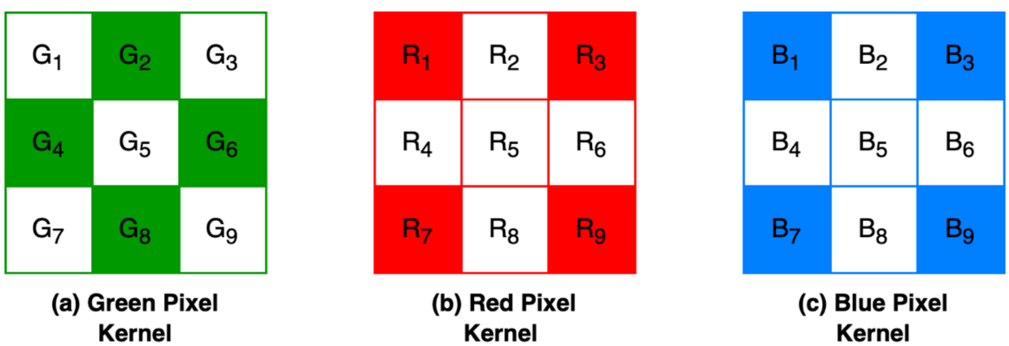


Figure 5.3 Bilinear interpolation of (a) green, (b) red and (c) blue pixels

In the Bilinear interpolation, the missing colours for each pixel of the entire sensor can be calculated using the following equations:

Green:
$$G_5 = \frac{G_2 + G_4 + G_6 + G_8}{4} \tag{5.1}$$

Red: $$R_2 = \frac{R_1 + R_3}{2};\ R_4 = \frac{R_1 + R_7}{2};\quad R_5 = \frac{R_1 + R_3 + R_7 + R_9}{4} \tag{5.2}$$

Blue: $$B_2 = \frac{B_1 + B_3}{2};\ B_4 = \frac{B_1 + B_7}{2};\quad B_5 = \frac{B_1 + B_3 + B_7 + B_9}{4} \tag{5.3}$$

As previously discussed, the Bayer filter has twice of the number of green pixels when compared to the red and blue pixels. As a result, the calculations of the missing green pixel will always include the 4 neighbouring pixels as can be seen by the green colour kernel in Figure 5.3 (a) and Equation 5.1. The estimation for the red and blue pixels, however, will most often involve only two neighbouring pixels. In the cases where the missing pixel is in the center (*e.g.* R_5 and B_5) the calculations will include the four neighbouring pixels as can be seen in Figures 5.3 (b) and (c) and Equations (5.2) and (5.3). Consequently, this method of interpolation will always be more sensitive and show more accurate results for the green channel than the red and blue pixels. As mentioned, the Bilinear interpolation is a simple algorithm with quick results; however, as we will show in section 5.3, this interpolation method creates large numbers of defects on the edges. The Bilinear interpolation works well for uniform areas where there are no sudden variations of colour but creates problems for any areas that contain patterns. Also, it generates false colour artifacts and decreases the overall image quality. The false colour artifacts become very noticeable. Consider the simple case illustrated in Figure 5.4, we can see that a gray image with different edges (on the left) gets converted to parallel alternating lines of colours (on the right), in a plaid pattern. Because these artifacts are so noticeable, more complex demosaicing methods are vital to make clear images, even though they use much more computational power. Figure 5.4 illustrates well the reason why more complex methodologies were developed to attempt to reduce these negative effects. The next two demosaicing algorithms we will discuss are: The Median and The Kimmel Demosaicing.

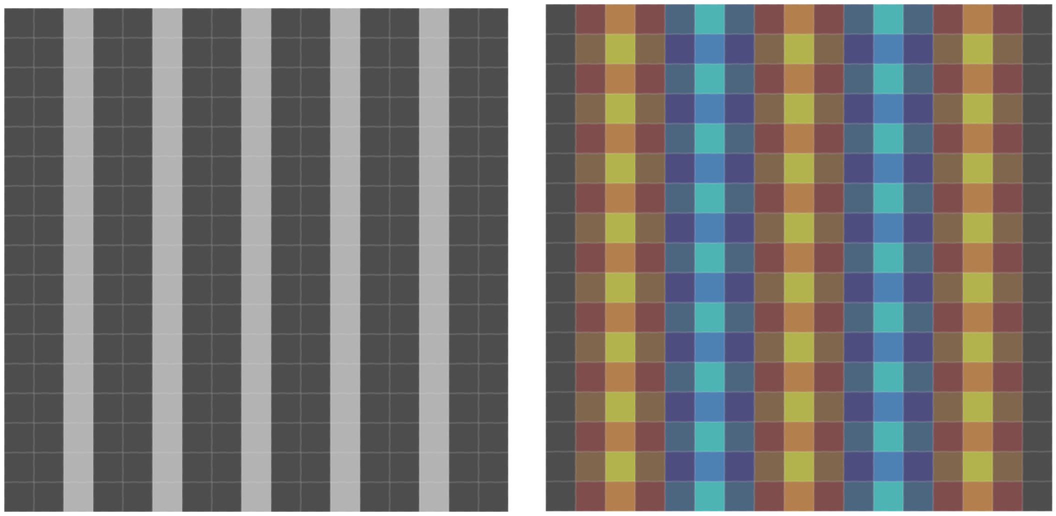


Figure 5.4 Bilinear edge artifacts, which the left image is the original and the right image in the Bilinear results

5.2.2. The Median Demosaicing

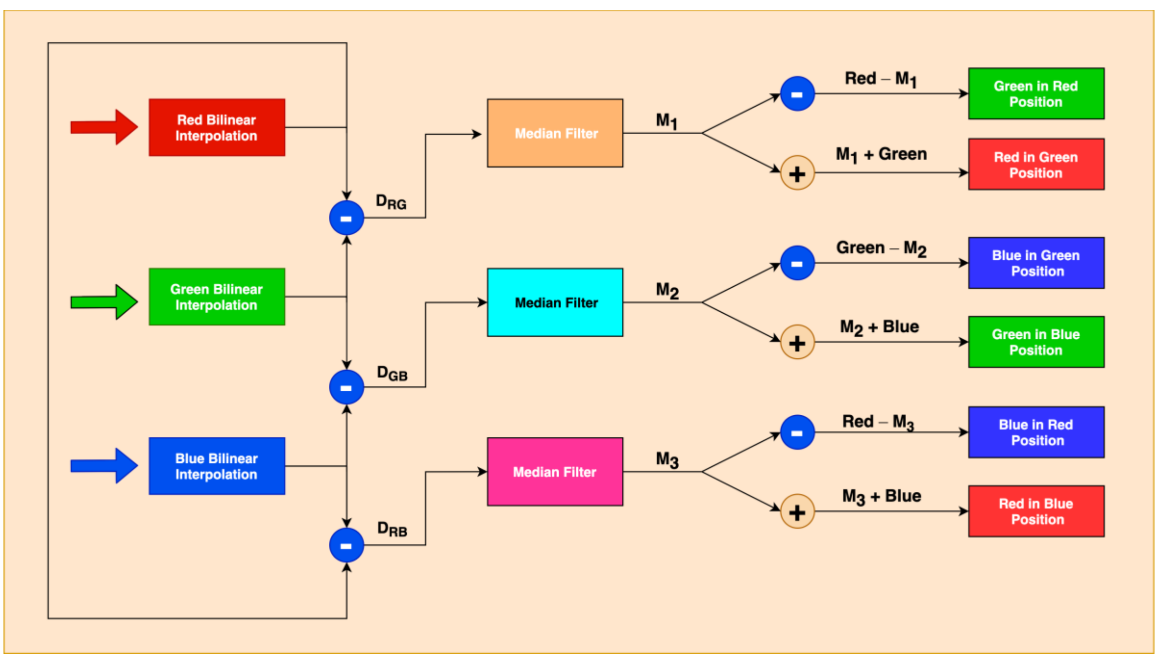


Figure 5.5 The Median demosaicing algorithm developed by Freeman (after [63])

The second demosaicing algorithm we will discuss is the statistical method proposed by Freeman [52], which attempts to minimize the Bilinear demosaicing artifacts using local colour statistics. This is the Median demosaicing method and it builds on the Bilinear demosaicing by taking into consideration the other colour planes when estimating a missing colour of a pixel. The Median method, like other more complex algorithms, is not an alternative to replace the Bilinear demosaicing, instead they all start with the Bilinear interpolation and extend or correct it. This algorithm is executed in four steps, Figure 5.5 illustrates the interpolation process of the Freeman Median filter. The Median demosaicing begins with the Bilinear interpolation to estimate the missing colours of each photosite. Then, it creates three functions, f_R (x,y) for the red channel, f_G (x,y) for the green and f_B (x,y) for the blue. These functions represent the pixel value at a specific location x, y for the corresponding red, green and blue colour plane. The next step, as we see in the Diagram 5.5, it is to calculate the difference between each interpolated colours at every pixel using Equations (5.4) to (5.6).

$$D_{RG}(x,y) = f_R(x,y) - f_G(x,y) \tag{5.4}$$

$$D_{GB}(x,y) = f_G(x,y) - f_B(x,y) \tag{5.5}$$

$$D_{RB}(x,y) = f_R(x,y) - f_B(x,y) \tag{5.6}$$

It warrants noting that there are different approaches to the Median algorithm; for example, the method proposed by Cok that uses ratios instead of differences [63]. But in this section, we will only explore the Freeman Median demosaicing.

The third step, after calculating the differences, is to apply a median filter to the results encountered in D_{RG}, D_{GB} and D_{RB} to find the values for M_1, M_2, and M_3. Figure 5.6 help us understand the median filter.

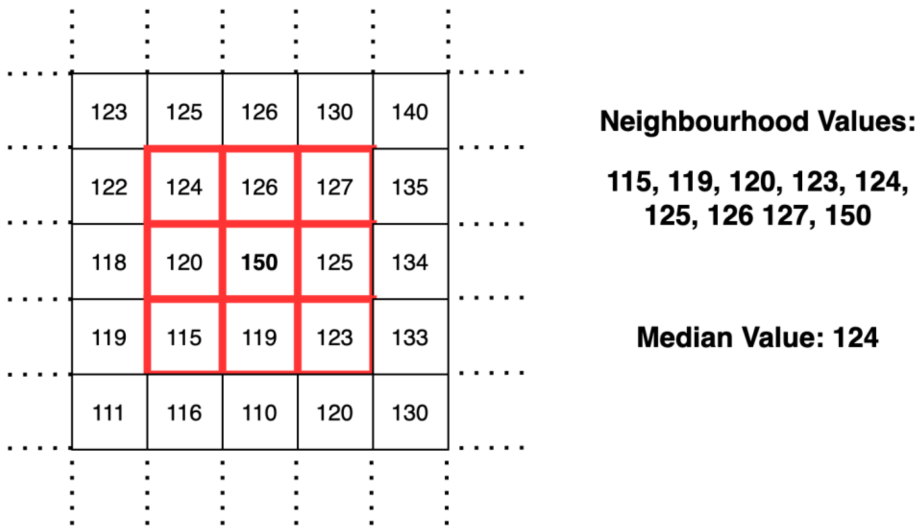


Figure 5.6 The Median filter (after [79])

The Median filter considers each pixel in the image and checks the nearby neighbours to decide whether this pixel is a representative of its surroundings [79]. If the central pixel value is far from the neighbours' values, the pixel is replaced with the median value of the surrounding pixels. In the example in Figure 5.6, a 3×3 pixel square is used here, the central pixel value of 150 is unrepresentative of the surrounding pixels and is replaced with the median value of 124. The goal of the Median filter is to eliminate any large divergences between colour planes based on the neighbouring pixels. Generally, the Median filter takes the information from the local area in a 3×3 or a 5×5 pixel kernel, larger kernels will produce more severe smoothing.

After the Median filter is applied, the final stage is a correction step where the results from the median filter (M_1, M_2, and M_3) are used to correct the values of red, green and blue estimated by the Bilinear interpolation, as we can see in Figure 5.5. The Median method helps to minimize artifacts on objected edged regions in the picture. In fact, this method can reduce the Bilinear interpolation errors and faulty patterns in the final image. The Median works well for great or large areas of colour with reduced errors at edges. However, as we will show in section 5.3, not all the edge artifacts in the images can be corrected by the Median demosaicing as the edge orientation becomes important.

Therefore, a more complex iterative method, that can calculate the interpolation adapting to the scene, was developed.

5.2.3. The Kimmel Demosaicing

It was discussed that the Median method does not correct all the edge artifacts since it is just a simplistic filter applied across the image. Therefore, an adaptive demosaicing algorithm was developed to take into consideration the local structure of the image. The adaptive approach we will discuss is the Kimmel demosaicing algorithm. This is an advanced process of interpolation that is far more complex than the Bilinear or Median methods, Leung [11]. This algorithm uses various steps to find the edges and use the local information to correct them. The Kimmel algorithm requires more computational power as it takes into consideration mathematical modelling of the local area to estimate the best approximation of the pixel colour. The Kimmel demosaicing is executed in three different steps and it incorporates various methods such as Bilinear, weighted-gradient and colour ratio interpolation, to adapt to the local structure of the image. To understand how this algorithm works, recall the green pixel kernel that is displayed in Figure 5.7.

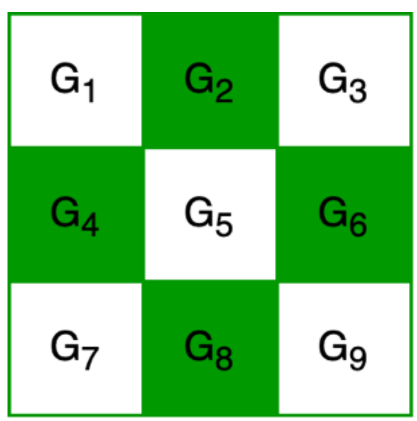


Figure 5.7 Green pixel kernel

The green pixel kernel will be used in the first step of the Kimmel algorithm. Like the Median method, Kimmel starts by calculating the missing colours using the Bilinear interpolation. But there is a difference, the Kimmel method uses a weighted version of the

Bilinear interpolation to calculate the missing green components of each pixel. Equation (5.7) is used to estimate the weighted Bilinear for the central green pixel (G_5) that will be used in the processing pipeline.

$$G_5 = \frac{E_2 G_2 + E_4 G_4 + E_6 G_6 + E_8 G_8}{E_2 + E_4 + E_6 + E_8} \tag{5.7}$$

The weight factor E_i observed in the Equation (5.7) is used to adjust the Bilinear interpolation in the direction of the local edge and can be calculated using the Equation below:

$$E_i = \frac{1}{\sqrt{1 + D^2(P_5) + D^2(P_i)}} \tag{5.8}$$

The gradient function *D* introduced in the previous Equation (5.8) is calculated using Equations (5.9) – (5.12). Figure 5.8 helps us to understand the direction of the gradient function *D* that is being calculated. One important characteristic of these equations is that the calculations remain the same for all the RGB colours regardless of the colour at the pixel P_5, therefore it is appropriate to use the notation P_i instead of R_i, G_i and B_i.

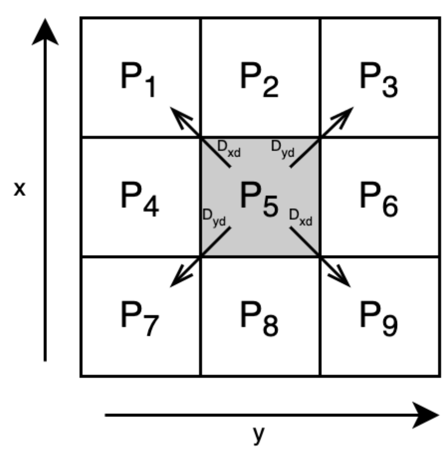


Figure 5.8 Kimmel gradient mask

Equations (5.9) and (5.10) are used to calculate the vertical and horizontal gradients respectively. And the equations (5.11) and (5.12) are used to estimate the *D* values for the +45° and -45° diagonals. This is an advanced step compared to the Bilinear method, as it does not take into consideration the values of the diagonal of the central pixel.

$$D_x(P_5) = \frac{P_2 - P_8}{2} \tag{5.9}$$

$$D_y(P_5) = \frac{P_4 - P_6}{2} \tag{5.10}$$

$$D_{xd}(P_5) = max\left\{\frac{P_1 - P_5}{\sqrt{2}}, \frac{P_9 - P_5}{\sqrt{2}}\right\} \tag{5.11}$$

$$D_{yd}(P_5) = max\left\{\frac{P_3 - P_5}{\sqrt{2}}, \frac{P_7 - P_5}{\sqrt{2}}\right\} \tag{5.12}$$

Next is the second stage of the Kimmel algorithm. This step is the estimation of the red and blue components using a ratio interpolation method. After this process, the ratio between colour planes is assumed to stay the same within the image scene. Equations (5.13) and (5.14) allow us to calculate the red and blue information separately. Both equations are based on the ratio related to the green components since the Bayer patterns have two times more green information than the red and blue colour planes.

$$R_5 = \frac{\left(E_1 \cdot \frac{R_1}{G_1}\right) + \left(E_3 \cdot \frac{R_3}{G_3}\right) + \left(E_7 \cdot \frac{R_7}{G_7}\right) + \left(E_9 \cdot \frac{R_9}{G_9}\right)}{(E_1 + E_3 + E_7 + E_9)} \times (G_5) \tag{5.13}$$

$$B_5 = \frac{\left(E_1 \cdot \frac{B_1}{G_1}\right) + \left(E_3 \cdot \frac{B_3}{G_3}\right) + \left(E_7 \cdot \frac{B_7}{G_7}\right) + \left(E_9 \cdot \frac{B_9}{G_9}\right)}{(E_1 + E_3 + E_7 + E_9)} \times (G_5) \tag{5.14}$$

Then, the third and final step is a correction procedure. The goal of this stage is to ensure the colour ratio in the image stays constant. In the Kimmel algorithm, this is achieved by recalculating the green components estimated in the first stage, using the computed red and blue values obtained from the second stage. Equations (5.15) through (5.17) show this process.

$$G_{R_5} = \frac{\left(E_2 \cdot \frac{G_2}{R_2}\right) + \left(E_4 \cdot \frac{G_3}{R_4}\right) + \left(E_6 \cdot \frac{G_6}{R_6}\right) + \left(E_8 \cdot \frac{G_8}{R_8}\right)}{(E_2 + E_4 + E_6 + E_8)} \times (R_5) \tag{5.15}$$

$$G_{B_5} = \frac{\left(E_2 \cdot \frac{G_2}{B_2}\right) + \left(E_4 \cdot \frac{G_3}{B_4}\right) + \left(E_6 \cdot \frac{G_6}{B_6}\right) + \left(E_8 \cdot \frac{G_8}{B_8}\right)}{(E_2 + E_4 + E_6 + E_8)} \times (B_5) \tag{5.16}$$

$$G_5 = \frac{G_{R_5} + G_{B_5}}{2} \tag{5.17}$$

After correcting the green components, the red and blue colour planes also must be recalculated. This is achieved by using Equations (5.18) and (5.19) respectively.

$$R_5 = \frac{\sum_{i=1}^{9}\left(E_i \cdot \frac{R_i}{G_i}\right)}{\sum_{i=1}^{9} E_i} \times G_5, \qquad i \neq 5 \tag{5.18}$$

$$B_5 = \frac{\sum_{i=1}^{9}\left(E_i \cdot \frac{B_i}{G_i}\right)}{\sum_{i=1}^{9} E_i} \times G_5, \qquad i \neq 5 \tag{5.19}$$

This correction step is typically repeated three times to achieve the best results, then the Kimmel Algorithm is finalized. This method of demosaicing incorporates a weighted Bilinear interpolation, a colour ratio interpolation and a correction step to adapt to the scene of an image. Introducing these features to the demosaicing process has proven to reduce artifacts that are found in other methods such as the Bilinear and Median algorithms. However there is a trade-off; the Kimmel demosaicing method requires high computational performance, which can have an impact on the price of the camera and uses much of the camera's power, decreasing the battery life significantly.

There are many other, much more complex, demosaicing algorithms that are being used in modern camera models. Bilinear, Median and Kimmel algorithms are just three examples. These methods are used whole or in part or entirely to develop final images in sensors that use the CFA pattern. However, these algorithms are not perfect, and they create what is called demosaicing Artifacts. Also, none of these methods can track and eliminate permanent defects prior to the interpolation process and this results in the amplification of the damaged area of a defective pixel. This is what will be explored in the following sections of this chapter.

5.3. Demosaicing Artifacts

Figure 5.9 (a)Original image (b)Demosaicing image by Bilinear interpolation with visual artifcats in the fance and the wall on the left (taken from [69])

While demosaicing algorithms generally improve image resolution, they inherently develop artifacts near edges or sudden changes in colours/intensity when calculating the interpolation of the missing colours in the CFA pattern. These errors lead to visual artifacts that decrease the overall image quality. In Figure 5.9 (b) we see the artifacts caused by the Bilinear interpolation in the fence area on the right and the wall on the left. Observe that while the fence is nearly white, blue and red colour errors are created that were not present in the original image, Figure 5.9 (a). The original image does not suffer from interpolation artifacts because it was taken in a film camera and then scanned to a computer, where no demosaicing was needed. In our research on permanent defects it is paramount to understand the errors caused by the demosaicing process because when the interpolation calculations are combined with a hot pixel, which is a sudden change in intensity on a pixel, these errors become even more evident. This happens because the algorithm calculates the interpolation using the faulty pixel as a regular pixel. As has been mentioned previously, hot pixels are often not mapped out before the demosaicing process.

To better understand the demosaicing artifacts introduced by the Bilinear interpolation, research performed by Chang et al. [68], created an artificial image with a vertical edge, processed with the bilinear interpolation and displayed the results. This can be seen in Figures 5.10 (a) – 5.10 (f).

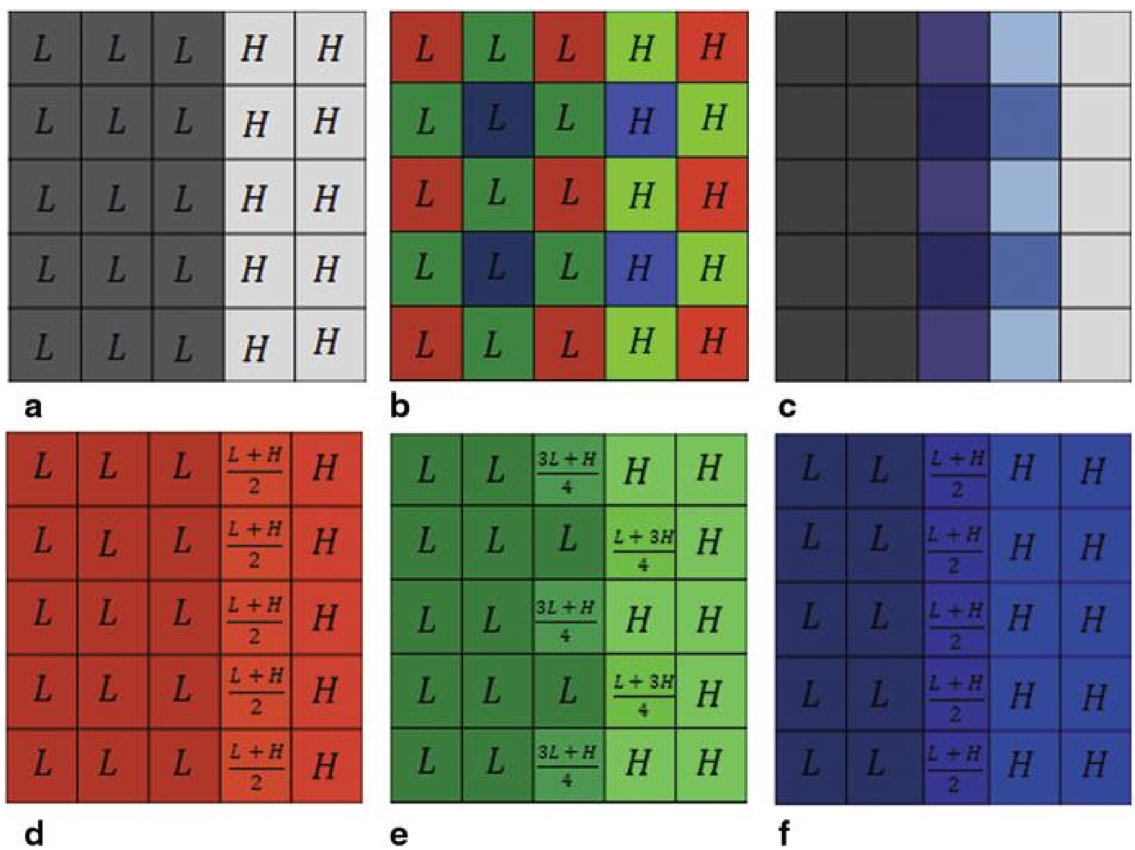
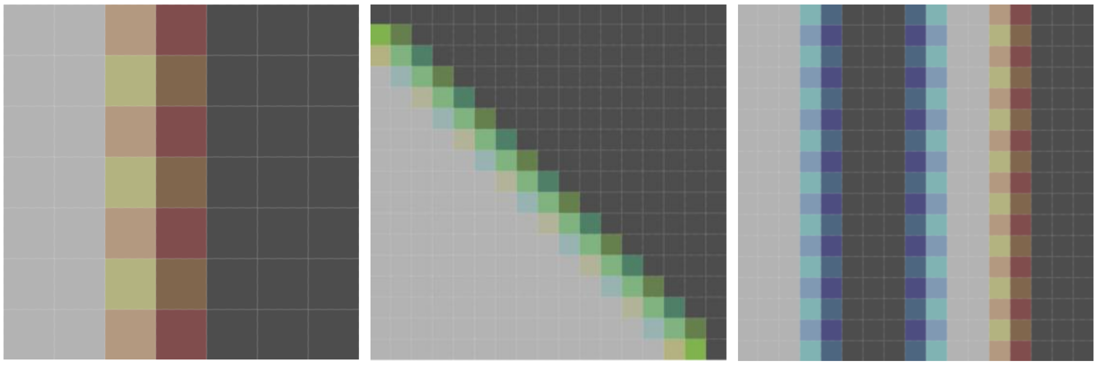


Figure 5.10 (a) Artificial gray image, (b) CFA samples, (c) Bilinear interpolation results, (d) Bilinear interpolation red plane, (e) Bilinear interpolation green plane, (f) Bilinear interpolation blue plane (taken from [69])

The artificial image (Figure 5.10 (a)) was created with two homogenous areas with different gray levels *L* (Low) and *H* (High), and the three colour components in each gray area are equal (i.e. no colour present). Figure 5.10 (b) shows the Bayer CFA pattern used to create the image and Figure 5.10 (c) shows the final output. Figures 5.10 (d) – 5.10 (f) display the results of the Bilinear interpolation for each colour plane. By analysing the final image and the three separate colour channels, we can see that the three interpolated colour planes suffer from various errors due to their sampling patterns. From Figure 5.10 (d) to Figure 5.10 (f) we see that the grid error pattern is produced by the green sampling and that red and blue planes create an intermediate intensity level between the L and H levels. These are two examples of visual artifacts generated by the demosaicing interpolation. One is the grid effect that creates a pattern of alternating colours. Looking at Figure 5.10 (c) we can see that the demosaicing creates for this example light and dark blues artifacts, and there is variation in the blue colour intensity. This type of artifact is also known as the zipper effect. The second example is the noticeable colour errors, such as the bluish tint shown in the final image, known as false colour [69].

In our research, we performed the same test methodology to explore the sampling errors caused by the Bilinear interpolation. We recreated the original tests described previously, but also developed new artificial images with different edge locations. Figure 5.11 displays a collection of those results. It is possible to see that all of them suffer from zipper effects and false colour distortions, confirming they are an inherited side effect of the demosaicing interpolation. Also, looking at Figure 5.11 we see that the colour distortion created depend on where the edge is: a red column edge will create different colour distortion compared to a blue column edge. In this context, it is worth noting that the diagonal edge brings a completely different colour pattern. Therefore, from these tests, we see that the demosaicing will not only create colour artifacts, but the colour pattern will vary depending on the edge location and orientation. We will not explore these results in further detail because they follow the same interpolation errors described previously. The purpose of these tests was to show that demosaicing artifacts will happen regardless of the location and orientation of the edge.



(a) (b) (c)

Figure 5.11 Demosaicing artifacts in gray images with different edge patterns: (a) single edge, (b) diagonal egde and (c) repeated pattern

Returning to Figure 5.9 (b), in the beginning of this section, we can see a clear case of the zipper effect. The zipper effect can be defined as an abrupt or unnatural change of intensity over several neighbouring pixels; it manifests as an “on-off” pattern in regions around the edges [68]. The zipper effect is primarily caused by improper averaging of neighboring colour values across edges. For this reason, many proposed demosaicing algorithms are edge-sensitive. Figure 5.12 displays a closer look at the fence bars from the previous image. It is possible to see that the original colour values from the fence bars have been corrupted by the zipper effect as a side effect of the Bilinear interpolation. A closer look in 5.12 (b) we observe that there are oscillating blues and oscillating reds where the edge of the fence is, and that is the same behaviour we saw in Figure 5.11 (c).

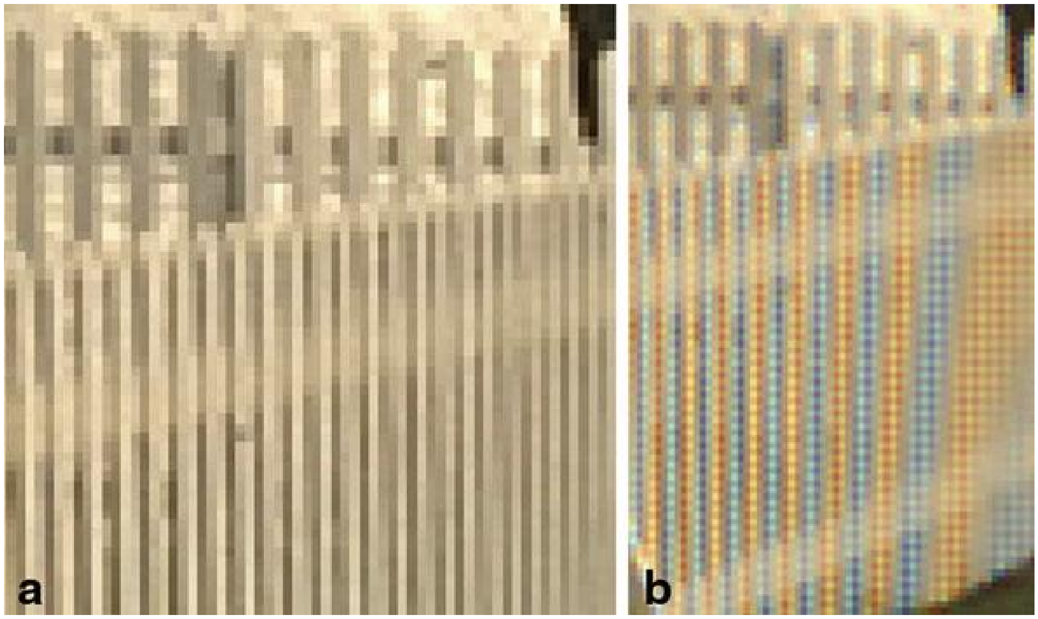


Figure 5.12 Zipper effect. (a) Fence bars in the original image, (b) Fence bars in the bilinear interpolated image [taken from (69)]

The second type of demosaicing artifact that we have discussed are the false colours distortions. False colours are adulterated colours that are not present in the original image. Figure 5.13 (b) displays this type of artifact across the numbers and the edge above them. False colours appear as sudden hue changes due to inconsistency among the three colour planes. This inconsistency usually results in large intensity changes in the colour difference planes [68].

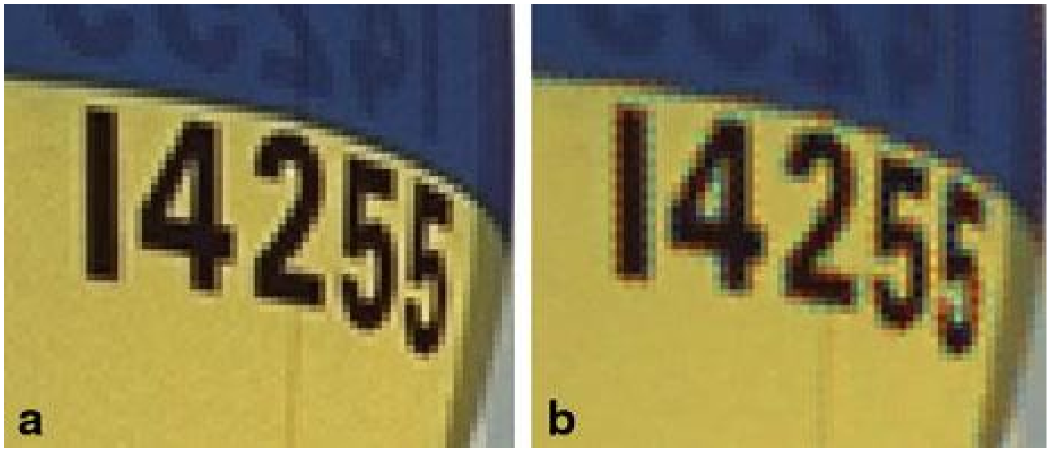


Figure 5.13 False colour. (a) Numbers in the original image, (b) Numbers in the bilinear interpolated image (taken from [69])

In general, both the zipper effect and false colours are referred to as misguidance colour artifacts mainly caused by inaccurate interpolation direction. However, the zipper effect are repeated patterns, whereas the false colours are changes in colours at edges, and they do show any clear pattern. These artifacts mostly affect regions with high-frequency content. Zhen et al. [69] argue that even with the correct interpolation direction and using the ideal demosaicing algorithm, the reconstructed image may still contain several errors called interpolation artifacts, which are associated with limitations in the interpolation process.

The Bilinear algorithm is not the only interpolation method that creates artifacts. More complex demosaicing methods, such as the Median and the Kimmel algorithms, were introduced to reduce these effects, but in fact they create new types of errors. In the literature these algorithms are used to improve the Bilinear artifacts, however, both processes are well-known to fail under some conditions. This can be observed in Figure 5.14, where we have processed a simple image using the three demosaicing algorithms we have discussed previously: The Bilinear, the Median and the Kimmel. In Figure 5.14(a) we see that the original image has one defined edge and that creates problems for the demosaicing algorithms. As we can see in Figure 5.14(b) the Bilinear creates false colour patterns that did not exist in the original image, and that is also observed by the added colour peaks shown in the Bilinear histogram.

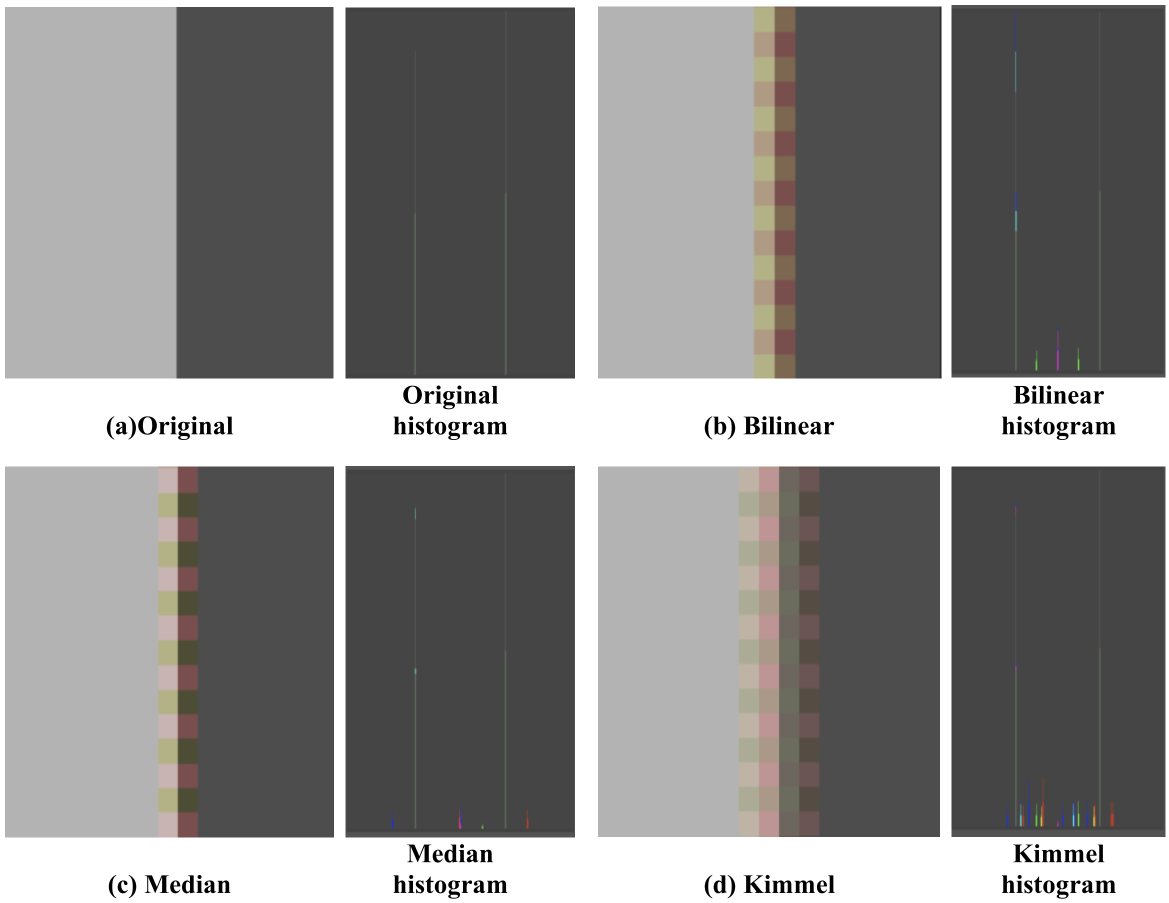


Figure 5.14 An image containing one simple edge processed by the Bilinear (b), Median(c) and Kimmel(c), with histograms

Similar behaviour is seen in more complex demosaicing methods. As we can observe, the Median method in 5.14 (c) slightly reduces the appearance of the artifacts, the colours are less intense. However, analysing the Median histogram, we see that it creates more false colours when compared to the Bilinear results. For the Kimmel algorithm in 5.14 (d) we observed a new behaviour, this method reduces the intensity of the Bilinear artifact, however it spreads the false colours to the neighbouring pixels around the edge creating a larger artifact. This happens because both the Median and Kimmel leverages the information from other colour planes while calculating the missing colour and they use larger pixel areas when compared to the Bilinear demosaicing. Additionally, the Kimmel results show that this method creates more false colours than the previous algorithms, and it spreads the false colours into the colour channels. This is better observed on the histograms displayed for each algorithm.

The histograms were created using Photoshop. They display the intensity levels for the RGB in an image, and where the RBG levels overlap is an indicator that the colours are added together. In the histogram for the original image, Figure 5.14 (a), we see two pure gray levels: where the red, green and blue are at the exactly same intensity. When we start processing these images with different demosaicing methods we see differences in the histogram results. Figure 5.14 (b) displays the results for the Bilinear interpolation. In the Bilinear histogram we see that there are other colours, greens, blues and reds, that are happening outside of the intensity values of the two gray levels. This shows that the creation of additional colours in an inherent side effect of the Bilinear interpolation. Then, in Figures 5.14 (c) and 5.14 (d) we see that both the Median and the Kimmel interpolation further exacerbate this problem, the histograms show that more false colours are created in these complex demosaicing methods.

As we can see from this experiment, neither the Median nor the Kimmel algorithm cannot remove the false colours entirely. It is true that more complex algorithms make the Bilinear artifact less intense. However, it is often disregarded that these demosaicing methods are adding a wider range of colours to the original values, and because they take into consideration the neighbouring pixels, they will spread the false colour effects to larger areas of the image. The results from this experiment show us that all demosaicing methods will create false colours, and the histogram analysis showed us that they will spread the false colours into other colour planes. Additionally, when the image has more complex edges these problems are further exacerbated. This can be seen in Figure 5.15 which we created an image with three edges and processed them with the Bilinear, Median and Kimmel algorithms.

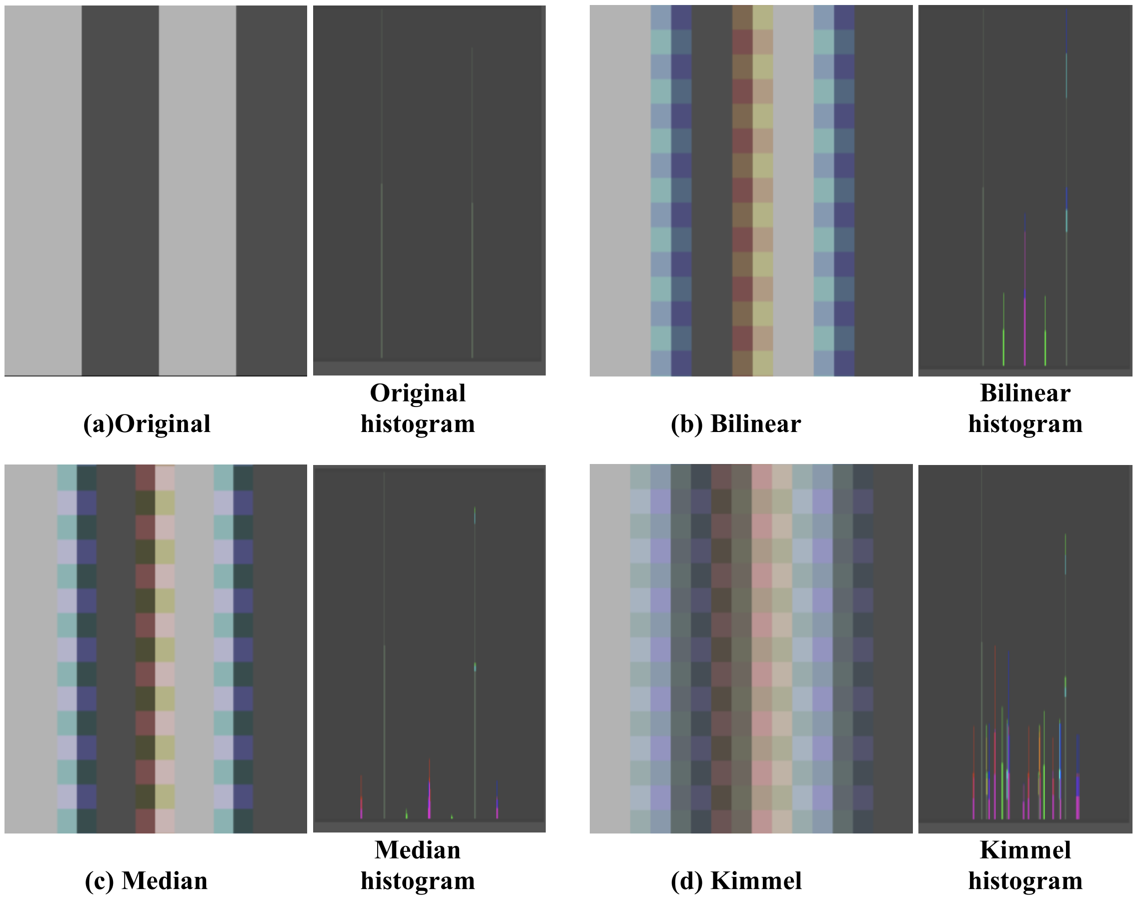


Figure 5.15 An image containing three edges processed by the Bilinear(b), Median (c) and Kimmel(d), with histograms

For the tests in Figure 5.15, we created an image with two different gray levels (30% and 70%) of 3 pixels wide. The results for this slightly more complex pattern (in comparison to Figure 5.14) shows a new effect: we see multiple colours appearing because we have a close edge interaction. Then, as we added the Median and Kimmel demosaicing, the patterns that are in close enough proximity start to interact even more because these methods calculate the missing colours over an expanded area. This effect appears in the histograms as a huge increase in the number of false colours, and in the intensity of those colours. As we will show in section 5.5, this is a similar behaviour we have found when two hot pixels are in close enough proximity to interact.

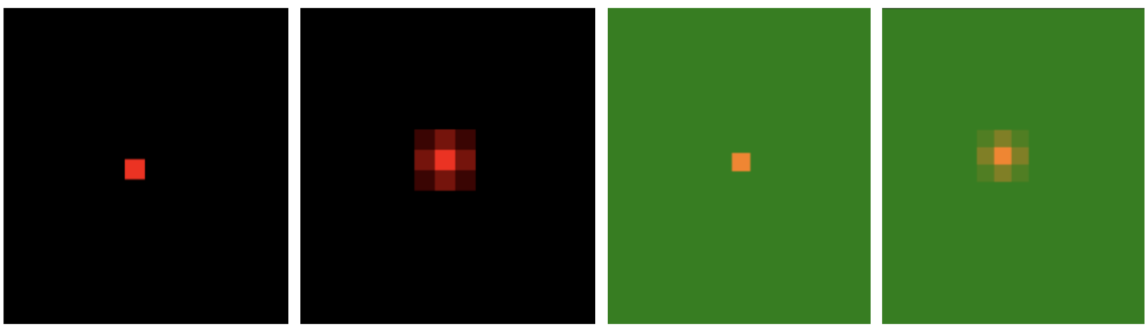
Therefore, we conclude that, although the Median and the Kimmel demosaicing attempt to correct the false colour effect created by the Bilinear interpolation, they are not able to achieve it and in fact, will spread the false colour to further pixels. However, in this thesis what we are not interested how good are the demosaicing methods at correcting the errors from the Bilinear in a typical image. Instead our focus is how much error each method generates when they interact with a hot pixel type defect, which we explore in the next section.

5.4. Analysing Permanent Defects in Demosaiced and Compressed Images

There are several studies that analyse the inherited errors caused by the demosaicing process. However, there appears to be nothing in the literature other than the work of Dr. Chapman's group that explored the impact that interpolation algorithms have on permanent defects. It is important to highlight that the CFA sensor does not differentiate hot pixels from normal pixels; consequently, the faulty pixel will have an impact on the final image output which is larger than expected. In fact, as we will discuss in this section, the hot pixel's damage area expands due to two important processes: the demosaicing, discussed in the next section and the JPEG compressions (in 5.4.2). All these processes are executed before the final displayed image is reconstructed, hence why it is important to understand their effect on the hot pixel behaviour in the final image. In this section and the next we will show that both the demosaicing and the JPEG compressions impact the hot pixel damaged area, affecting the neighbouring pixels in an image. Therefore, it is an error to make the ofte- stated assumption that a single hot pixel would be negligible when inserted into modern 10-50 megapixel images, especially if displayed at lower resolutions. Moreover, in section 5.5 we will also explore how when two defects are in close proximity, the area impacted by the two hot pixels becomes much larger than that impacted by a single pixel.

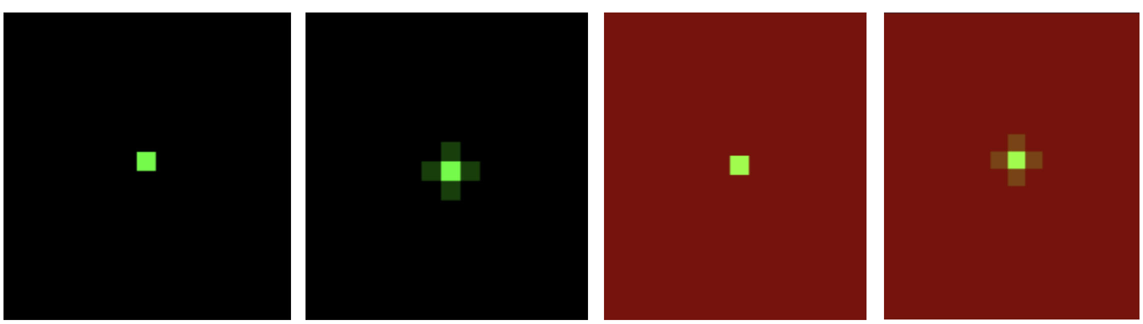
5.4.1. The Impact of the Demosaicing Algorithms in Single Hot Pixels

To start this analysis, consider how a single red and a single green hot pixel is affected by the simplest demosaicing algorithm, the Bilinear interpolation. Figure 5.16 displays a single red hot pixel defect in a black and green background, and the visual representation of the artifacts created by Bilinear interpolation. On the other hand, Figure 5.17 shows the artifacts created by the Bilinear process using a green hot pixel in a black and red background.



(b) Original (b) Bilinear (c) Original (d) Bilinear

Figure 5.16 Single red hot pixel defect in a black background and green background processed by the Bilinear interpolation

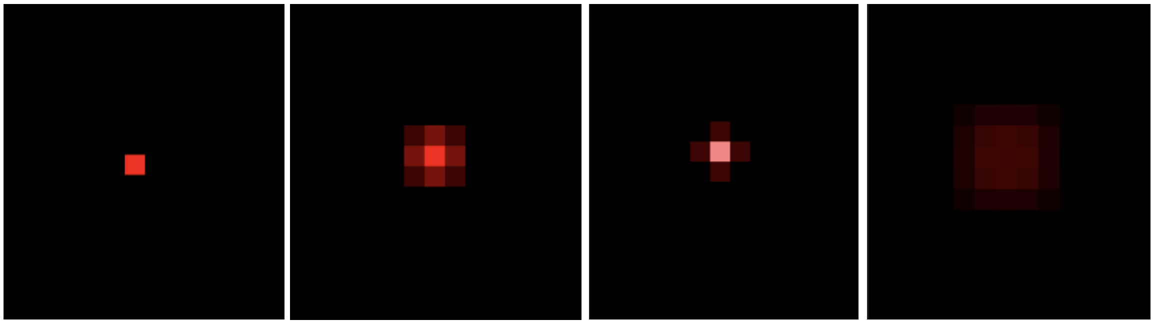


(a) Original (b) Bilinear (c) Original (d) Bilinear

Figure 5.17 Single green hot pixel defect in a black background and red background processed by the Bilinear interpolation

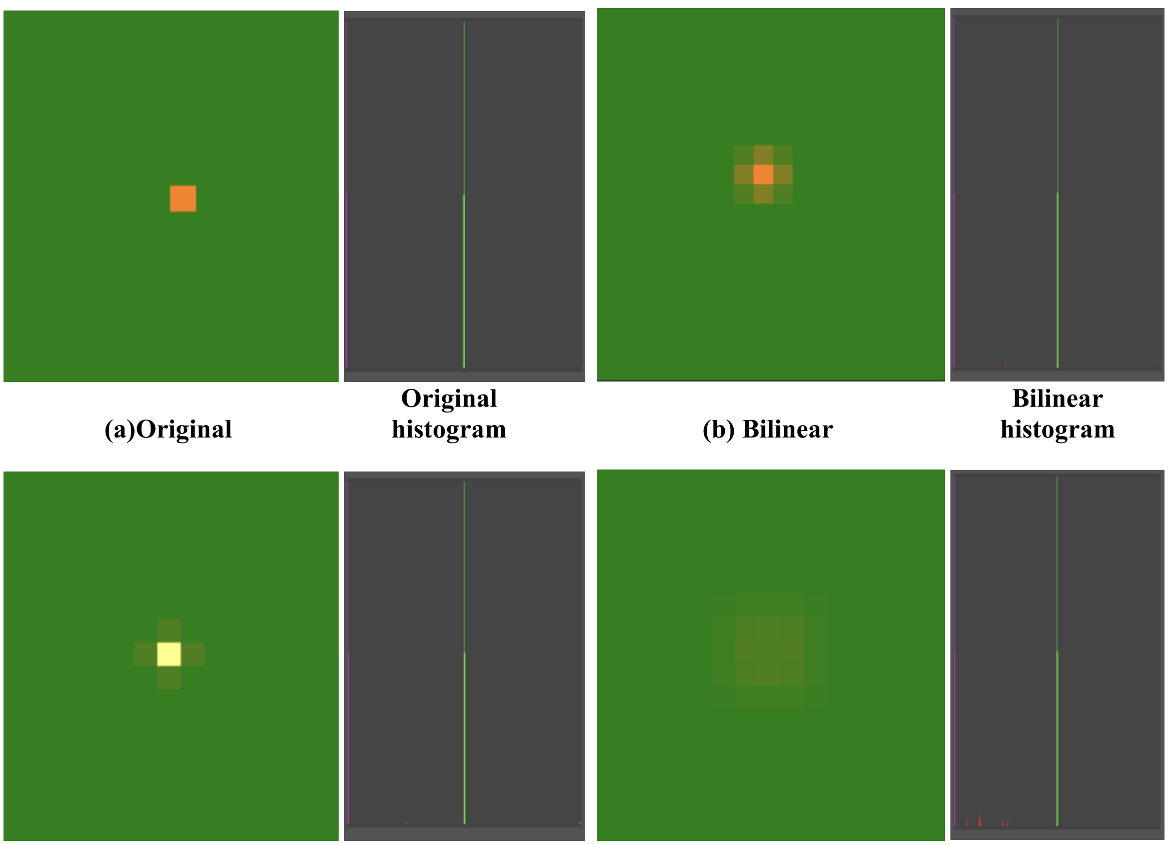
The resulting images in Figures 5.16 and 5.17 provide us valuable information on different artifacts created by the Bilinear interpolation process. Starting with the red hot pixel in Figure 5.16, we see that the Bilinear algorithm creates visual artifacts that spread the hot pixel damaged area to the nearest neighbouring pixels. In Figures 5.16 (a) and (b) the red hot pixel has an offset of $I_{offset} = 1$ above a black background and in Figures 5.16(c) and (d) the red hot pixel is above a green background with 0.5 intensity. Using Equation 5.2, the Bilinear interpolation creates a 3×3 array of red pixels with intensities of 0.5 in the vertical/horizontal pixels and 0.25 on corners, expanding the 1 hot pixel effect to 9 pixels in total area. The red colour of the hot pixel is spread by the demosaicing algorithm to the R values to neighbouring G and B pixels creating a local colour shift. This is most clearly observed in Figure 5.16 (d) where the red hot pixel appears to have orange colour shades on a green background. Also, the Bilinear algorithm adds the red colour as an increased signal to the neighbouring pixels without decreasing its own value, effectively increasing the local intensity effect, *i.e.* the total brightness added to the area, from the hot pixel by 4 times that created by the hot pixel. Since the Bilinear interpolation calculations for red pixels and blue pixels (Equation 5.2 and Equation 5.3 respectively) are the same, we see that a blue hot pixel will display similar behaviour to that illustrated in Figure 5.16.

However, green pixels are calculated differently in the Bilinear interpolation as expressed in Equation 5.1. For green pixels the Bilinear interpolation averages 4 surrounding pixels, instead of only 2. Hence why the Bilinear interpolation creates a smaller artifact for the green hot pixel compared to the red hot pixel. Figure 5.17 displays the results for the tests using a green hot pixel. The Bilinear expands the 1 hot pixel effect to 5 pixels and creates green pixels with intensities of 0.5 in the vertical/horizontal directions but does not create any false colours on the corners. Additionally, similar to what is observed for the red hot pixel, the Bilinear algorithm adds the green colour as an increased signal to the neighbouring pixels without decreasing its own value, effectively increasing the intensity effect, from the hot pixel by 2 times. The results from the simplest Bilinear interpolation shows that the red hot pixel is the one colour defect to create the strongest artifacts.



(a) Original (b) Bilinear (c) Median (d)Kimmel

Figure 5.18 Single red hot pixel defect the impact of different demosaicing algorithms



(a)Original — Original histogram — (b) Bilinear — Bilinear histogram

(c) Median — Medium histogram — (d) Kimmel — Kimmel histogram

Figure 5.19 Single red hot pixel defect on a green backgroung and the impact of different demosaicing algorithms with histograms

Next, we decided to extend these tests to the more complex demosaicing algorithms using a single red hot pixel on a black background and on a green background with 0.5

intensity. Figure 5.18 and Figure 5.19 display the results. Previous research by Chapman et al. [57] have performed tests on a single red hot pixel on the simplest black background to understand the impact of the other two more advanced demosaicing methods. However, in this research we have refined the calculations and the software algorithms for these demosaicing methods and extended the tests to include a red hot pixel in a half-intensity green background. To show the impact of all the demosaicing methods in spreading the hot pixel image to the other colours, we added the colour histograms (from Photoshop) for each algorithm. From the results, we observe that the histogram of the original image in Figure 5.19 (a) shows only a sharp line for the green values, and very little colour spreading. In Figure 5.19(b) we see the results for the Bilinear interpolation, if we look closely at the Bilinear histogram (we suggest magnifying the histograms of this section using a digital copy of this paper), we start to observe red levels that do not exist in the original histogram. Then, in Figures 5.18(c) and 5.19(c), we observe the Median algorithm results. The Median method creates a smaller damage area compared to the Bilinear interpolation, from 1 hot pixel, this algorithm affects another 4 neighbouring pixels. However, it causes the pixel to be brighter as it adds other colours, changing the hot pixel towards a pinkish tint in the simplest black background rather than the bright red pixel we have in the first image. From the green background results, (Figure 5.19(c)), it is clear the there is a colour shift as the original red hot pixel turns into a white-yellow tone after the Median process, and effectively displaying a higher intensity. This happens because the Median builds on the Bilinear demosaicing by taking into consideration the other colour planes when estimating a missing colour of a pixel. Therefore, the Median process will spread the hot pixel effect as a colour shift into other colour planes and the neighbouring areas. This can be seen by taking a close look into the Median histogram, where we observe the addition of red levels that did not exist in the original image. Hence why the Median method tends to intensify the brightness of the hot pixel by adding additional colours in the interpolation process. As can be seen from the resulting images, the Median filter does not correct the defects. In fact, this method will enlarge the defect onto all three colour planes making the defect appear as a white spot when the pixel is at or near saturation.

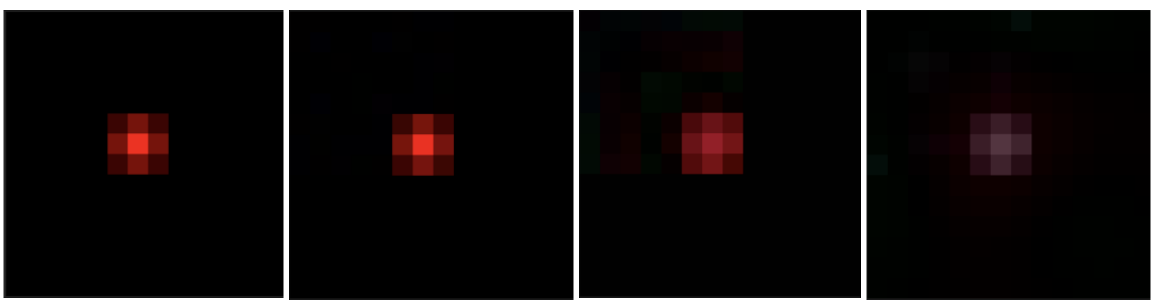
Lastly, we have the Kimmel algorithm. At first glance in Figures 5.18(d) and 5.19(d), we see that the Kimmel results spread almost uniformly the single hot pixel intensity in both the black background and uniform green background. A closer look shows that it is noticeable that this algorithm enlarges the damaged area the most. From the one single hot pixel in the original image, the Kimmel method spreads the colour error to a 6×6 damage area in different intensity levels of red, affecting 36 pixels in total. This is much larger than the 3×3 region measured from the Bilinear and Median demosaicing. Also, the colour ratio interpolation used in the Kimmel functions enhances the expansion of the error values on colour planes that do not contain defects. A close look into the Kimmel histogram we observe a much stronger spreading of the red levels compared to the other demosaicing methods. However, different from the Median demosaicing, the Kimmel algorithm will not cause the red hot pixel to appear as a white spot.

At the end of this experiment, it was clear that all of the demosaicing algorithms will impact the hot pixel and will have a trade-off between the spread of the damaged area, intensity of the defect and accuracy. Although the Median and the Kimmel demosaicing attempt to reduce the impact of the Bilinear interpolation, they are not able to achieve it and in fact, will spread the hot pixel damaged area to the neighbouring pixels. Additionally, as we will show further in this chapter, the impact of the demosaicing methods are further exacerbated when two hot pixels in near proximity are involved.

5.4.2. The Impact of JPEG Compressions

The next step of our research is to understand the impact that another image processing step, the JPEG compressions, have on hot pixels after they have been spread by the demosaicing. Most professional and serious amateur photographers use digital RAW file formats to capture images. As we have discussed in Chapter 2, digital RAW images do only perform losses compression processes. JPEG files, on the other hand, use lossy compression for storing and displaying digital images, affecting the quality of the final image. Additionally, when associated with the interpolation algorithm, the JPEG compression can further enlarge the damaged area created by the hot pixel and the demosaicing sampling. Although RAW images have much higher quality and provide the closest results to the original scene, they usually are large files and need post-processing to be displayed and printed. Conversely, JPEGs are often the final files ready to be used and their sizes are much smaller than the RAW files. Also, many camera systems, such as majority of cell phones models, do not have the option to shoot in RAW. Therefore, most of consumers tend to use JPEG file formats and their different compression levels. Finally, almost all web images are JPEG to reduce file size.

Since JPEGs are widely used across many digital imaging systems, it is paramount to understand the impact that different levels of JPEG compressions have on hot pixels, and how this process can create stronger artifacts that affect the overall quality of the image. More importantly, as we will explore in this section, this file format has a great impact in expanding the hot pixel defective behaviour. To begin discussing this issue, we will first take a look at Figures 5.20 and 5.21.



(a) Tiff (b) JPEG 12 (Low loss) (c) JPEG 7 (Medium loss) (d) JPEG 4 (High loss)

Figure 5.20 The impact of JPEG low loss to high loss compressions in a red hot pixel, Bilinear

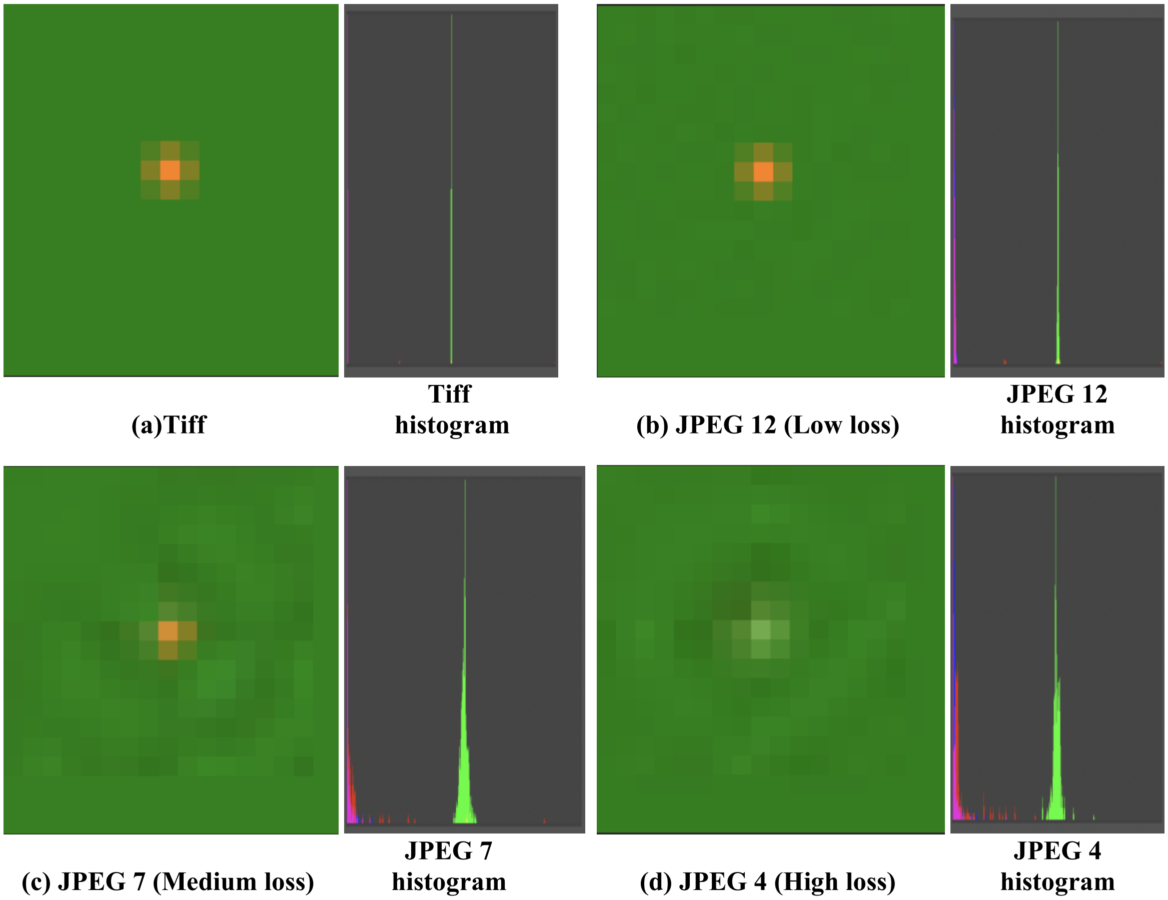


Figure 5.21 Bilinear algorithm results using a red hot pixel in a uniform green background processed with different JPEG compression levels, with histograms

Figure 5.20 displays the impact that different JPEG compressions have on a single red hot pixel on a black background that has undergone the Bilinear interpolation and how the defective area enlarges as higher compressions are applied. In Figure 5.21 we see the same analysis in a uniform green background, with their corresponding histogram. We have performed these tests using a lossless image file format, TIFF and different levels of JPEG compressions. The TIFF format does not have any loss of information, details or changes in colour since it uses lossless compression. However, it does apply the Bilinear demosaicing as we can see by the 3×3 defective area around the single hot pixel in the centre. In general, there are 12 levels of JPEG compressions, where JPEG 12 is the compression that shows the highest quality as it has the least loss of information and JPEG 1 is the one that displays the poorest results, with a greater amount of data loss. JPEG 12 is not as widely used as other higher compression levels because, while it is smaller than Tiff, it has larger file sizes when compared to other JPEG levels. On the other hand, JPEG 1 is rarely used because it creates strong artifacts and has a very low image quality. Adobe Photoshop groups the different levels of compression in terms of quality. JPEGs 1 – 4 are the low-quality range, JPEGs 5 – 7 the medium quality range, JPEGs 8 – 9 the high quality and JPEG 10 – 12 are the maximum quality range.

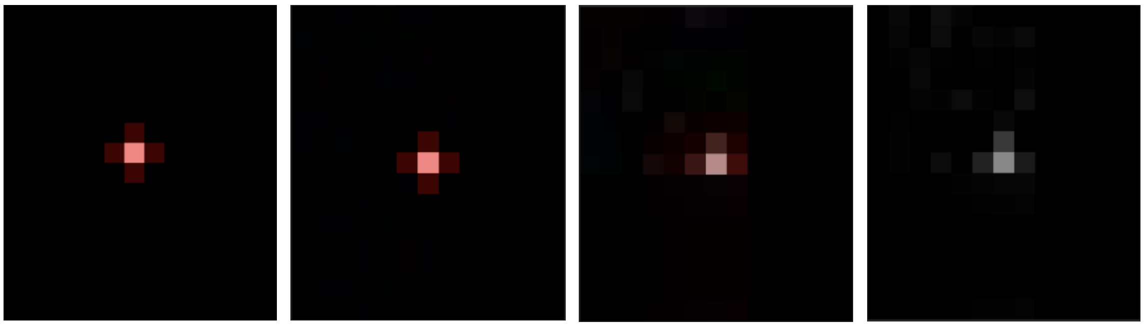
In our experiments, we have decided to explore the impact of the JPEG compressions choosing one level of each range. Displayed in Figures 5.20 (b) and 5.21(b) is JPEG 12, the lowest level of compression available. The results of JPEG 12 are similar to the TIFF format, however with less overall resolution. Also, a new effect was observed. As we have discussed in Chapter 2, the JPEG process divides the image into 8×8 pixel tiles before doing any additional compression in the individual tiles. This is performed for all images and compression levels. However, it is necessary to look closely to Figure 5.21 (b) to observe defect area spreading and colour distortion, as at JPEG 12 this effect is not so evident. Also, when analysing its histogram, notice that there is a slight broadening of the green and red levels at 128 (or half max) levels, showing that this JPEG format has the lowest loss of detail, colour changes and high frequency data information.

Next, we observe JPEG 7 in Figures 5.20(c) and 5.21(c). At JPEG 7, the side effects of the 8×8 pixel tiles process become much more evident as the compression gets stronger. In this particular example, the defect area from the two hot pixels spreads to 4 of the nearest 8×8 blocks, affecting a total area of 256 pixels. From the histogram we see the changes spreads in colour, the green colour plane changes from a sharp line at 128 to a much wider width of different intensities. The green has spread the colours from 0 to 27 colour levels. Also, we observe an increase of the red and blue colours levels added to the image in the range from 64 and below, see the histogram. The red change from 0 to 29 levels and the blue is spread from 0 to 8 levels. These colour shifts observed on the histogram are a solid indication that there are changes in both intensity and in colours channels and those changes are seen as noticeable changes in the image as an impact of higher JPEG compression levels expanding on the hot pixel damaged area. Also, visually we observe a significant and noticeable to the viewer colour distortion of the artifact, changing from a bright orange pixel to a dull-looking orange tone. The results found in JPEG 7 are an excellent example of what happens in most cameras systems that output JPEG images: this process step will create artifacts and will change the colour information across the image.

At the low-quality range of the JPEG compressions using JPEG 4: the results are displayed in Figures 5.20 (d) and 5.21(d). The results of JPEG 4 showed a worsening behaviour seen in the previous JPEG results. This compression level expanded the initial 3×3 hot pixel damaged area to a 16×16 area. It appears as a whole range of colours being spread through the damaged area. Additionally, we observed that this compression level also creates a greater colour distortion and spreading in the histogram as this was the highest compression used in our experiments. For this particular compression, the damage from the single red hot pixel expands to 4 of the 8×8 tiles affecting 256 pixels in total area. Analysing the histogram, we notice that a wide range of green has greatly expanded 71 colour levels of higher intensities down to lower intensities. Additionally, the red levels are almost a full band between 0 and the half-point intensity, expanding to 46 colour levels. Lastly, we also observed a significant increase of the blue levels in the histogram, changing from 0 to 25 colour levels. Hence, from the compression's levels shown in Figure 5.14, JPEG 4 was the compression level that most expanded and distorted the original hot pixel damage area and created the most amount of artifacts, affecting the overall image quality. In this context, it is important to note that even though JPEG 4 displays very poor quality, it is commonly used by casual photographers that take pictures in simple cameras system such as point and shoot cameras and cell phones.

The results found in this experiment have shown that the JPEG compressions create artifacts, colour distortions, decrease image quality and resolution and will spread the damaged area of the hot pixel that has been processed by the Bilinear interpolation. Showing that when combined, the Bilinear demosaicing and the JPEG compressions have a substantial role in exacerbating the damaged impact of a hot pixel.

The initial research performed by Leung [12] explored the side effects of JPEG compression in hot pixels that have been processed by the more complex demosaicing algorithms. In this research, we have expanded these tests to include the refined software algorithms for the demosaicing methods and integrated the tests in a uniform green Background. We will start with the results found for the Median algorithm, they are displayed in Figures 5.22 and Figure 5.23.



(a) Tiff (b) JPEG 12 (Low loss) (c) JPEG 7 (Medium loss) (d) JPEG 4 (High loss)

Figure 5.22 The impact of different JPEG levels on the Median demosaiced image with one red hot pixel defect

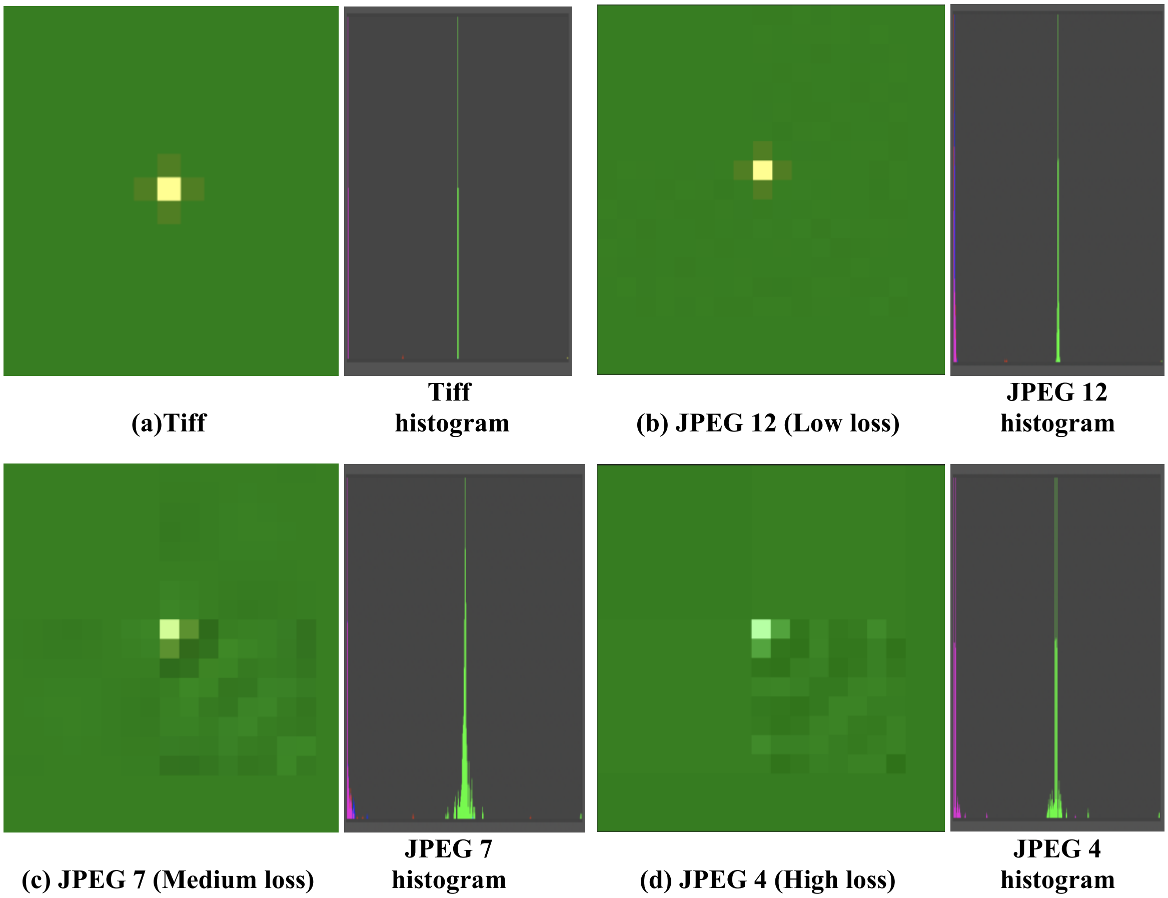
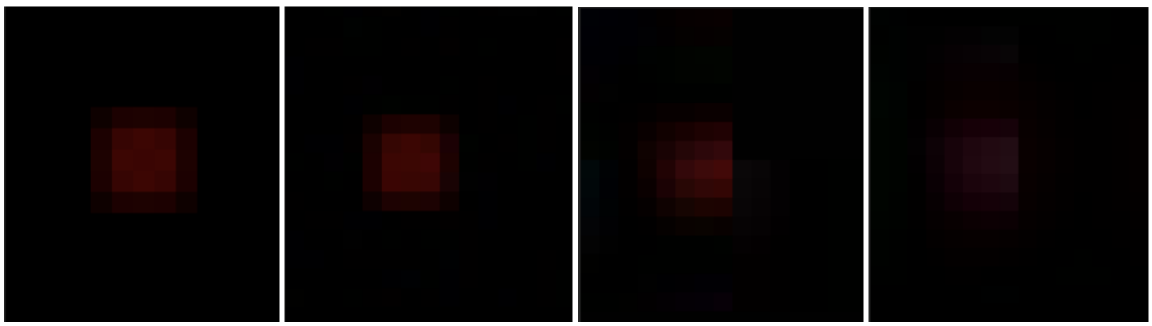


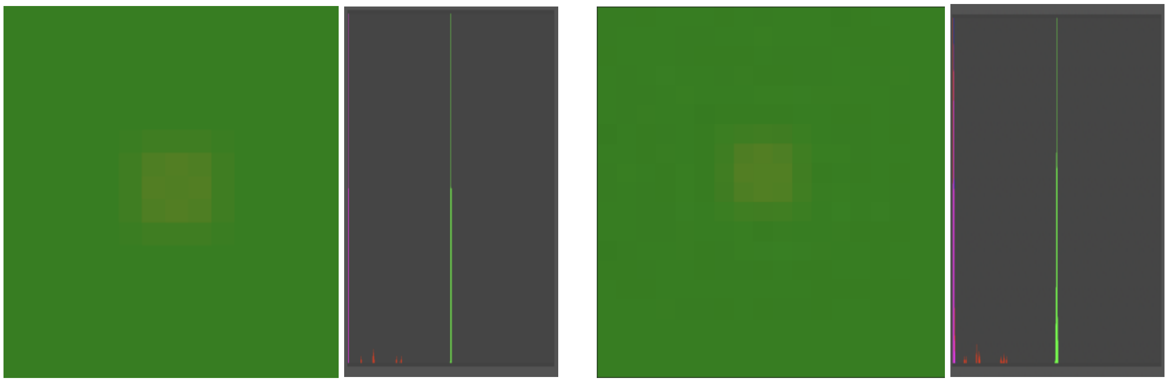
Figure 5.23 Median algorithm results using a red hot pixel in a uniform green background processed with different JPEG compression level

The examination of Figures 5.22 and 5.23 shows that the Median demosaicing is also greatly affected by the JPEG compressions. As we have discussed previously, we see that in Figures 5.22 (a) and 5.23 (a) the Median algorithm makes the red hot pixel appear as a bright white pixel surrounded by the neighbouring pixels with a weaker red and orange tone. In Figures 5.22 (b) – 5.22 (d) and Figures 5.23(b) – 5.23(d) we see displayed the side effects of different levels of JPEG compressions when combined with the Median demosaicing. We will focus on the JPEG 7 because this is the most common compression level used in simple camera systems. From the JPEG 7 histogram we see that the green channel changes from a sharp line in the TIFF format to a much wider range of 42 colour levels of different intensities. We also see a slight increase of red and blue levels in this compression level, the red changes from 0 to 14 levels and the blue plane spreads to 12 levels. This is further aggravated as higher compressions are used, such as JPEG 4. Additionally, JPEG 7 strongly affects 1 of the 8×8 tiles, but also slightly spreads the defect onto the other 2 of the tiles affecting a total area of 192 pixels. While the effect is general, the exact distribution varies according to the location of the hot pixel compared to the edge of the 8×8 tile of the JPEG process.

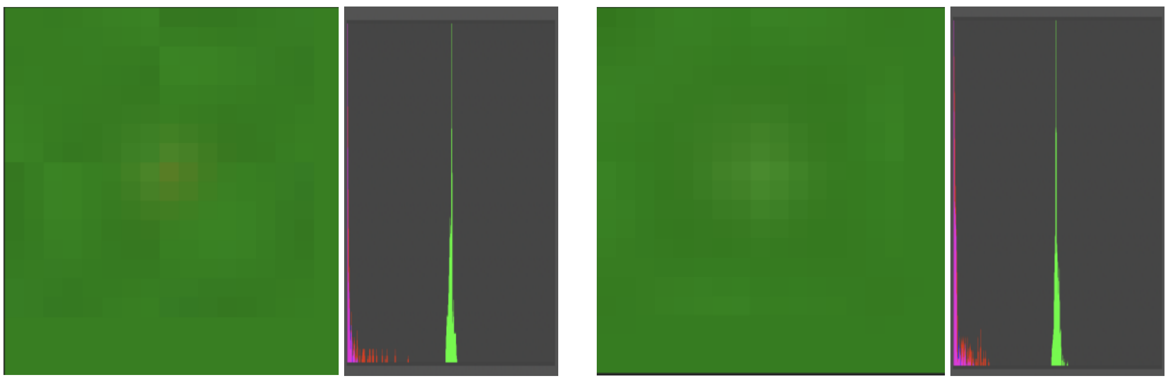


(a) Tiff **(b) JPEG 12 (Low loss)** **(c) JPEG 7 (Medium loss)** **(d) JPEG 4 (High loss)**

Figure 5.24 The impact of different JPEG levels on the Kimmel demosaiced image with one red hot pixel defect



(a)Tiff | Tiff histogram | (b) JPEG 12 (Low loss) | JPEG 12 histogram



(c) JPEG 7 (Medium loss) | JPEG 7 histogram | (d) JPEG 4 (High loss) | JPEG 4 histogram

Figure 5.25 Kimmel algorithm results using a red hot pixel in a uniform green background processed with different JPEG compression level, with histograms

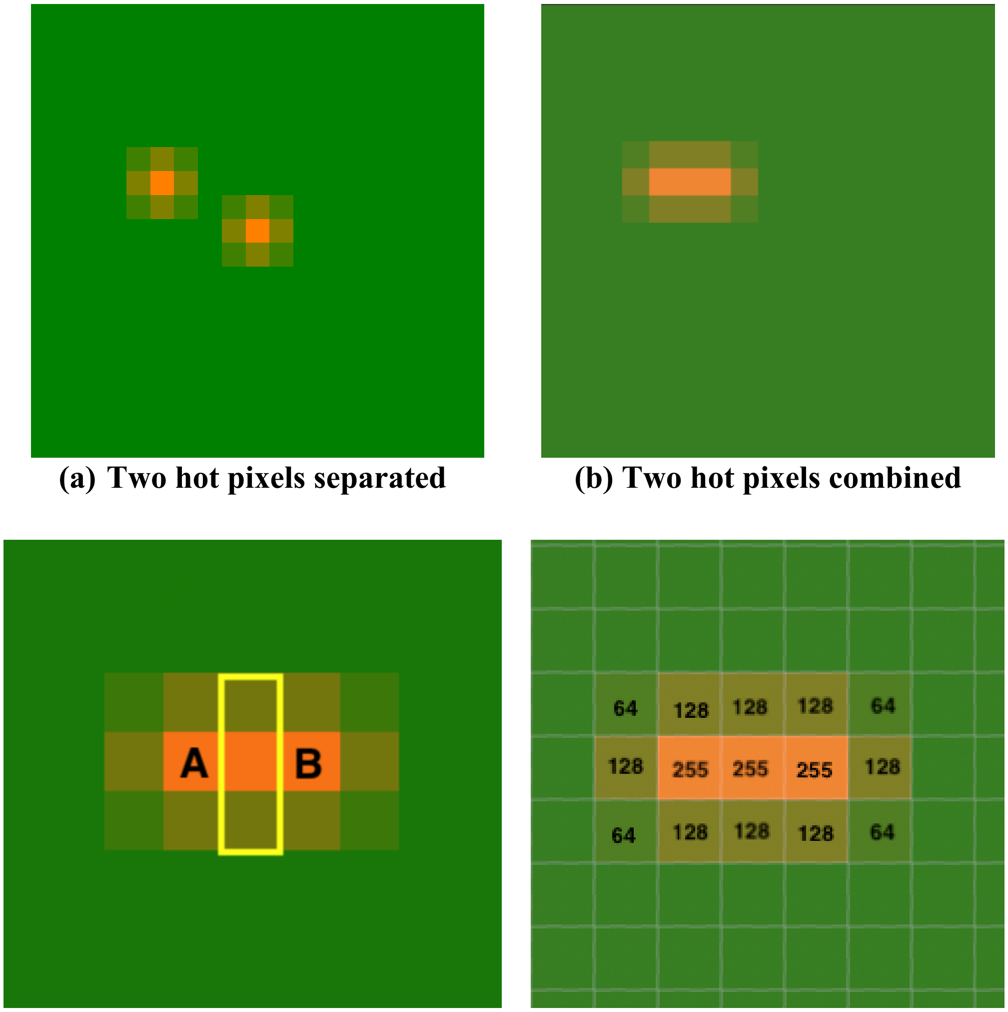
In Figure 5.24 and 5.25 we see displayed the results of the red hot pixel processed by the adaptive method, Kimmel. Different from the Median demosaicing, the correlation of the colour planes in the Kimmel interpolation will not cause the hot pixel to appear as a white pixel. In fact, from all the algorithms, the Kimmel methods shows the smoothest results with less intensity. However, as we have shown previously, the Kimmel algorithm inherently creates a defective area of 6×6 pixels without any compression. Furthermore, because the Kimmel algorithm takes into consideration the other colour planes, this is the interpolation method that spreads the false colours to the largest areas of the image. From the histograms we see that, even at the lowest JPEG compression, JPEG 12, there are added red information that did exist in the original image. This effect is greatly enhanced as the compression levels increase. At JPEG 7, we see that the red level has now spread to 51 levels. Additionally, the green levels change from 0 to 18 levels and the blue plane spreads the false colours to 12 levels. Furthermore, we observed that the single hot pixel defective area is spread to 16×16 area in all the compression, creating a significant colour distortion in the artifact.

From the tests we have performed it is clear that all JPEG compressions will spread a single defective area into its neighbouring pixels, regardless of the demosaicing process chosen. In fact, medium and higher compression levels such as JPEG 7 and JPEG 4, will not only enlarge the hot pixel damage area by two or three times, but it will also cause great colour distortions in the hot pixel original colour. Furthermore, this data compression strategy will discard and average many pixel areas in 8×8 or 16×16 pixel blocks, and when combined with the demosaicing interpolation, it will calculate the hot pixel values as a regular pixel, spreading the defect to the neighbouring pixels. The results found in this research are the reflection of the complex lossy compression process that is performed by the standard JPEG file used in vast majority of the digital camera systems. The important point in all of these is that expanding the defect to much larger areas and creating colour shifts, the change becomes much more visible to the viewer. Moreover, this expansion suggests that hot pixels that are nearby, but not actually touching, will interact as we show in the next section.

5.5. The Impact of Two Close Hot Pixels: Demosaicing and JPEG

So far in this chapter, we have discussed that the demosaicing process and the JPEG compressions spread the damage of one single hot pixel into much wider areas. This raises in the question – how close do two hot pixels need to be before this spreading causes the image damage to interact. We will start this analysis by taking a look at what the demosaicing does when two hot pixels are in close proximity, then in subsection 5.5.2, we will explore how this behaviour is aggravated by the JPEG.

5.5.1. Demosaicing



(a) Two hot pixels separated (b) Two hot pixels combined

(c) Combined area marked (d) Combined area with intensities

Figure 5.26 Two hot pixels in an image processed by Bilinear interpolation

In the previous sections, we have seen that the Bilinear demosaicing algorithm spread the hot pixel damaged area to the nearest neighbouring pixels creating a more noticeable visual artifact. Given this single defect expanded area now consider the 2 hot pixels in Figure 5.26 (a), where we look at how the expanded areas interact. The red hot pixels are separated by 5 pixels horizontally and 3 pixels vertically. We used Photoshop to create these images and to analyse the results. As we will in Chapter 6, in section 6.2.1, Photoshop, allow us to add the hot pixels to the background the same way the Bilinear does. In the analysis, we observed the demosaicing spreads the red colour to the neighbouring G and B pixels creating a colour shift 9 times larger. In addition, the Bilinear algorithm added the red colour as an increased signal to the neighbouring pixels without decreasing its own value, showing that regardless of the background the side effects of the Bilinear interpolation adds the value to the background. This is why both hot pixels have an orange tone instead of the original red, because Bilinear demosaicing spreads the values which when combined with the background results in a colour change.

Next, in Figure 5.26 (b) we move the two hot pixels are in close enough proximity (1 column apart) that their damage area starts to overlap. At first observation the visual artifact becomes more evident, as the area of the hot pixels combine and becomes a larger defect, changing from two 3×3 pixel areas to one 5×3 combination. In Figure 5.26(c) the black marked letters are the hot pixels, yellow is where the neighbouring pixels from the hot pixels have overlapped and all the remaining orange coloured pixels are the side effect resulting from the Bilinear interpolation. In this example, a 5 pixel wide by 3 pixels tall united area is created, from the right side of the 3×3 neighbour area of hot pixel A and 3 pixels from the left side of the 3×3 neighbour area of hot pixel B overlap. The yellow area represents the 6 hot pixels that have overlapped. The combined area has increased in intensity in the R colour channel. One would expect a larger area, a brighter defect is much more noticeable. Figure 5.26 (d) shows the numerical results for the red channel of each pixel. The results have shown that when separated, hot pixels A and B have pixels values that range from 64 to 128 in the red plane (in an 8-bit colour range, where 0 is complete darkness and 255 is at full saturation). Conversely, when combined, the overlapped area increases their R channel intensity, ranging from 128 to 255. This confirms that when two hot pixels are in close proximity, the demosaicing process will cause their damage area to interact, combining the intensity values of the hot pixel colour and consequently creating a much larger artifact.

Give the random process of hot pixels many would argue that in modern megapixel cameras, having two hot pixels in close proximity would be a rare event, but, as we will show later in section 5.6, it only requires a modest number of hot pixels to have a significant probability of this event happening.

5.5.2. JPEG

Now consider how adding image compression expands this effect. We have shown in section 5.4.2 that the demosaicing algorithms when combined with the JPEG compressions can enlarge a single hot pixel defective area to a 16×16 pixel square. Therefore, we decided to test how JPEG compressions expand the damage area of two hot pixels in close proximity. From section 5.5.1 it is seen that the Bilinear demosaicing results in the damage area being separated by one pixel or less if the two hot pixels are placed anywhere in a 5×5 box (see figure 5.27).

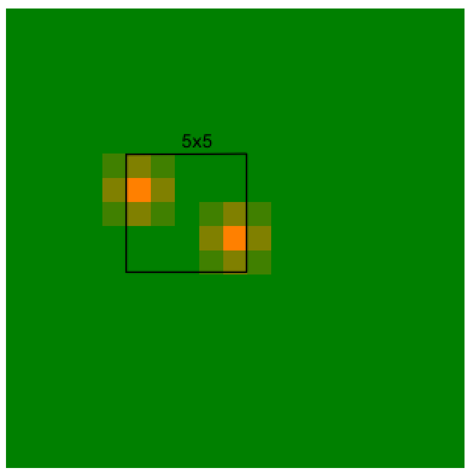
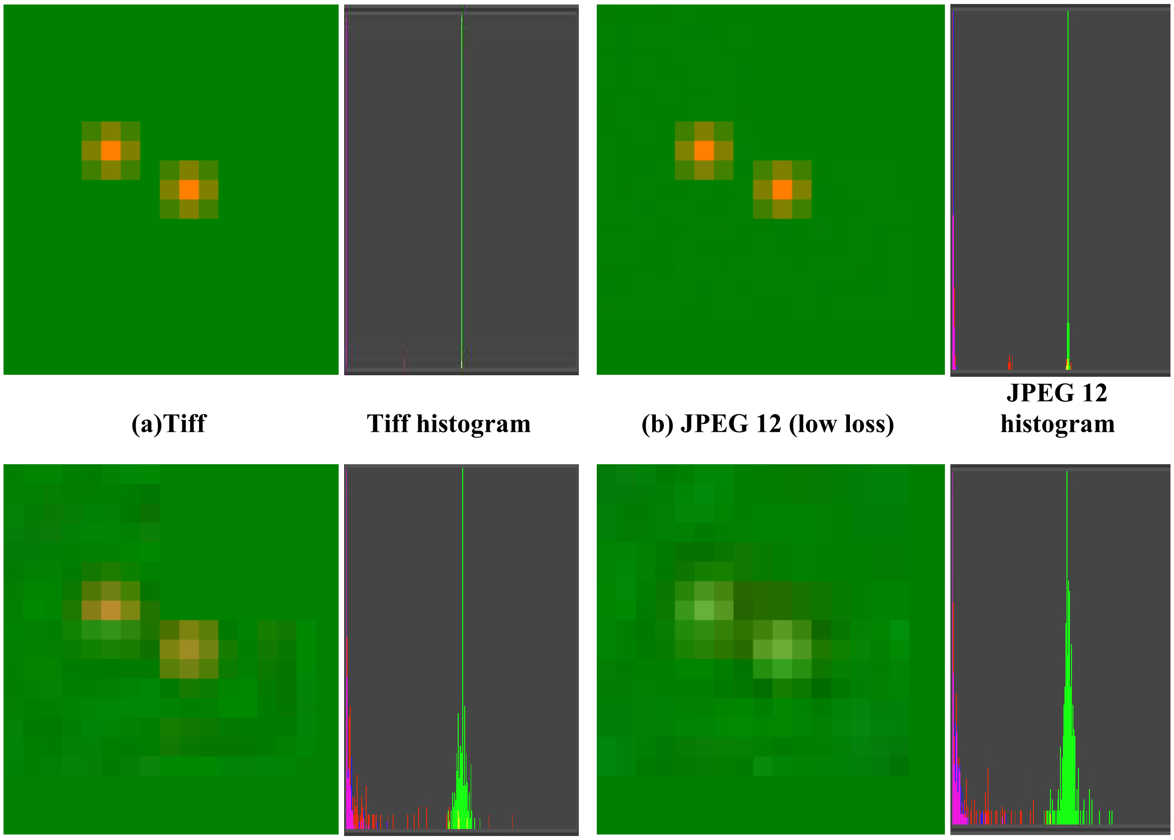


Figure 5.27 Two hot pixels within a 5x5 box after demosaicing

Again, using a uniform green background with 0.5 intensity allows us to better visualize the impact of the JPEG compressions on the multiple hot pixels' defective area. Also, the two hot pixels are separated by 5 pixels horizontally and 3 vertically. As in section 5.4.2, we use different image file compressions, from the lossless compression format, Tiff, to the high loss, JPEG 4. Figure 5.28 displays for each compression level the resulting image and its colour histogram, which show the expansion of hot pixels into adjacent colours within this region.



(a)Tiff | Tiff histogram | (b) JPEG 12 (low loss) | JPEG 12 histogram

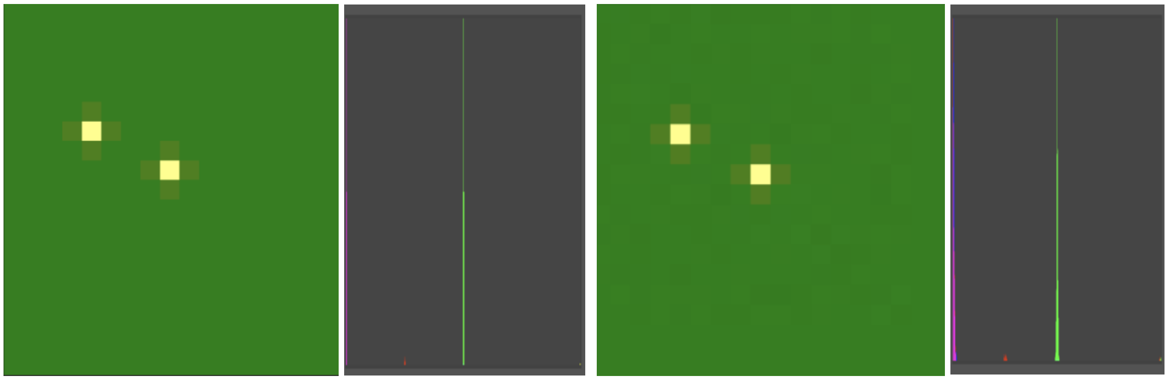
(c) JPEG 7 (medium loss) | JPEG 7 histogram | (d) JPEG 4 (high loss) | JPEG 4 histogram

Figure 5.28 Two red hot pixels in a 5x5 square with uniform green background processed with different compression files, with histograms

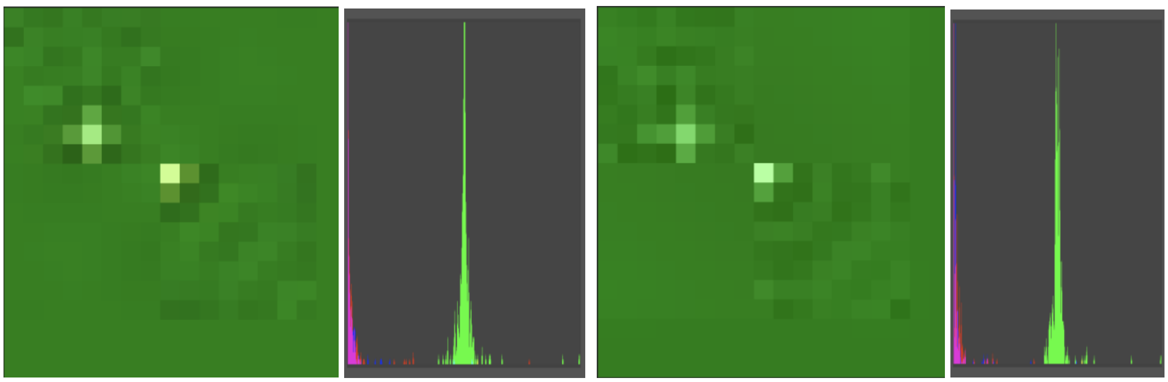
From Figure 5.28, we see how much damage each compression causes. The lossless Tiff in Figure 5.28 (a) enlarges the Bilinear damage from 2 pixels to 18 pixels in the area and the R colour plane is spread to 3 levels, according to the histogram. Again, the effects are intensified when we begin applying the JPEG because it spreads the damage both in area and colour. Figure 5.28 (b) displays the JPEG 12 compression level. Now that hot pixel damage is enlarged from 2 pixels to 24 pixels in an area with colour/intensity shifts. Also, the histogram indicates that the green colour channel is spread to different levels (slightly widening its histogram area) and the red plane shows an increase of added levels. The results are similar to that we observed from the JPEG 12 in a single pixel.

However, for JPEG 7 (see Figure 5.28(c)) we start to see interactions. The defect area from the two hot pixel is enlarged to 3 of the 8×8 blocks discussed in section 5.4.2, affecting a total of 192 pixels in area. The changes in colour are also observed. In the single hot pixel experiments, we showed that the green has spread to 27 colour levels, the red change from 0 to 29 levels, and the blue is spread from 0 to 8 levels. For the two hot pixel experiments, we see from the JPEG 7 histogram from the histogram, we see that the Green colour plane changes from a sharp line to a much broader area of 32 colour levels, the Red colour plane change from 0 to 191 levels and the Blue plane changes is affected ranging from 0 to 48. These colour shifts observed on the histogram are a solid indication that there are changes in both intensity and in colours channels when two hot pixels are in a5×5 box. Those changes are what are seen as noticeable defects on the image. Finally, for the lowest compression we tested, the JPEG 4 (Figure 5.28 (d)), we see that the damage from the two hot pixels expands to 4 of the 8×8 tiles affecting a 256-pixel area in total. Analysing the histogram, we observe more intense colour shifts in the all the RGB. The Green channel expanded to 80 levels, the Red channel from 0 to 191 and the Blue channel from 0 to 60.

One important point to make is regarding the location of where the hot pixels are relative to the JPEG tiles. In the examples of Figure 5.28, the hot pixels were placed in different tiles, which is the most common case observed in pictures. However, even when two hot pixels were in the same 8×8 block, the expansion of the demosaicing process combined with the JPEG compression still results in multiple tiles being affected. Additionally, the side effects of these image processing steps will also be observed when more complex demosaicing algorithms are applied.



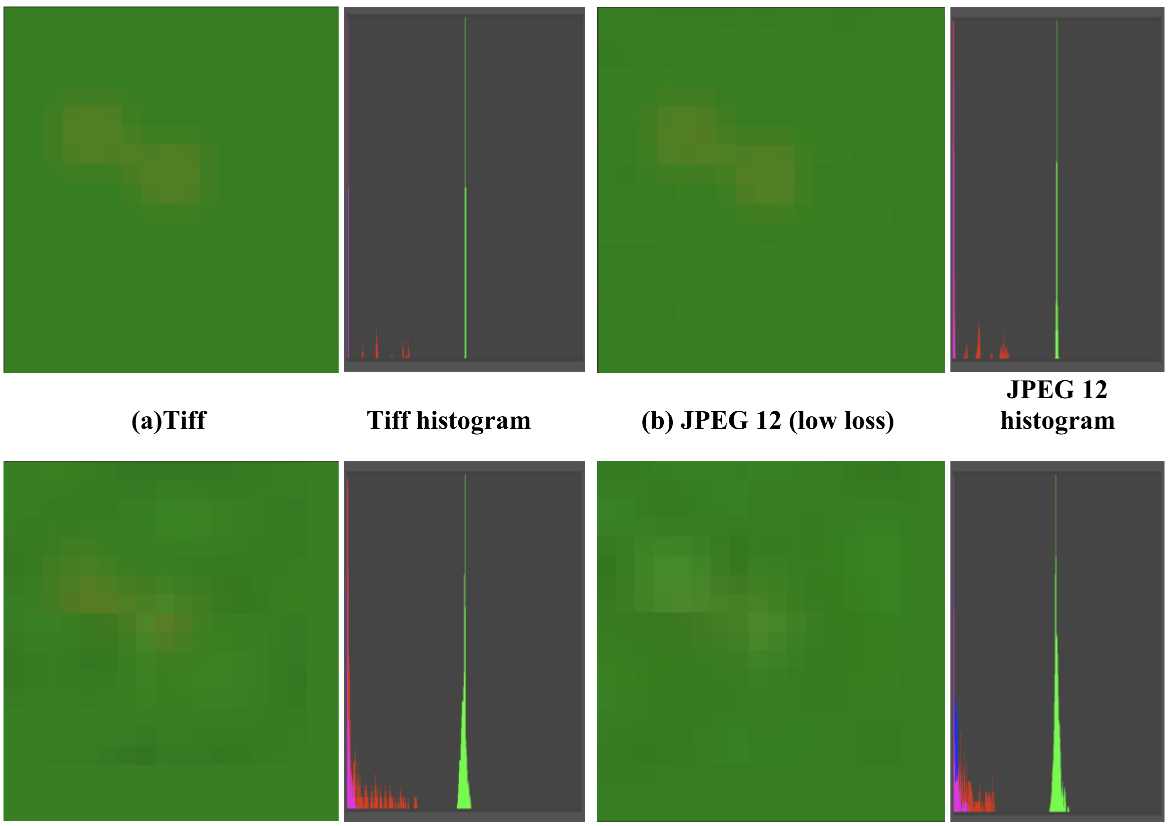
(a)Tiff | Tiff histogram | (b) JPEG 12 (low loss) | JPEG 12 histogram



(c) JPEG 7 (medium loss) | JPEG 7 histogram | (d) JPEG 4 (high loss) | JPEG 4 histogram

Figure 5.29 Two red hot pixels in a 5x5 square with uniform green background processed by the Median demosacing using different compression files, with histograms

In Figures 5.29 (b) – 5.29 (d) we see displayed the side effects of different levels of JPEG compressions when combined with the Median demosaicing. As we have discussed previously, the Median algorithm makes the red hot pixel appears as a bright white pixel surrounded by the neighbouring pixels with a weaker orange tone. From the JPEG 12, we see that the Median slight spreads the damage from the 2 hot pixels to 4 of the 8×8 tiles, affecting a total area of 256 pixels. In Section 5.4.2, we discussed that the Median strongly spreads the single hot pixel defect to 1 of the 8x8 tile and mildly affects the other two tiles at JPEG 7, impacting a total area of 192 pixels. For two hot pixels, we see that the Median algorithm processed by JPEG 7 strongly impacts two of the 8×8 tiles, and modestly impacts the other two neighbouring tiles, affecting a total of 256 pixels. Additionally, colour artifacts are spread onto all three colour planes. We saw from the single hot pixel experiments that the Median combined with the JPEG spreads the G channel to 42 colour levels, the R to 14 levels and the B to 12 colour levels. For the two near hot pixel pixels experiments, the JPEG 7 histogram shows that green levels change from a sharp line to a to 74 colour levels of various intensities. Also, there is a higher increase of red and blue information, these colour channels change from 0 to 32 colour levels and 0 to 25 colour levels, respectively. This shows that this algorithm combined with the JPEG compression creates greater colour distortion in comparison to the Bilinear interpolation.



(a)Tiff **Tiff histogram** **(b) JPEG 12 (low loss)** **JPEG 12 histogram**

(c) JPEG 7 (medium loss) **JPEG 7 histogram** **(d) JPEG 4 (high loss)** **JPEG 4 histogram**

Figure 5.30 Two red hot pixels in a 5x5 square with uniform green background processed by the Kimmel demoisaicng using different compression files, with histograms

Next, in Figure 5.30 we see the results for the adaptive method, Kimmel. As we have shown previously, the Kimmel algorithm inherently creates a defective area of 6×6 pixels without any compression. Therefore, the two hot pixels in a 5×5 box will start interacting before any JPEG is applied. This can be observed in Figure 5.30(a), which displays the TIF. For the higher compressions, Figures 5.30 (b) – 5.30(d) show that this interpolation method creates the falsest colours, as displayed in the histogram. From the histograms, we see an increased number of red levels across all the compression range we tested. There are more colour shifts of red for the two hot pixels results than the single hot pixel experiments, where the red colour was spread to 51 levels at JPEG 7. For the two hot pixels experiment, we see from the JPEG 7 histogram that the red plane has expanded to 83 levels, while the green spreads to 17 levels and the blue channel to 16 levels. This effect is greatly enhanced as the compression levels increase. Furthermore, as observed from the single hot pixel experiments, the JPEG mildly spreads the two hot pixels defective area is to 4 of an 8×8 tiles, affecting an area of 256 pixels in all the JPEG levels, creating a significant colour distortion in the artifact.

In summary, after analysing the impact of the JPEG process in single and multiple hot pixels, we concluded that a small number of hot pixels will affect much greater areas in the image. All the demosaicing methods we have experimented combined with the JPEG process will enlarge the defect area of two hot pixels to up to 256 pixels, creating visual artifacts and significantly decreasing the overall image quality. The results we have found prove that the idea that a small number of hot pixels will not affect the image is wrong, as they will strongly impact the neighbouring pixels after the image processing steps. However, another question remains unanswered: in a camera with a given number of hot pixels, what are the probabilities of two hot pixels being in the same 5×5 pixel box? This is what we will consider in the next section.

5.6. How Often Do Two Hot Pixels Interact

Previously in this chapter we have shown that when two hot pixels are within a 5×5 pixel box, the demosaicing algorithm combined with the JPEG compressions will cause their defect areas to interact creating a larger damaged area, resulting in a much more visible artifact. Many would argue that in a camera with 10 to 40 megapixels, it would be very unlikely that two hot pixels could be within a 5×5 pixel box (see Figure 5.31). Therefore, we decided to explore how many hot pixels are necessary to have a significant probability of this event happening in megapixel cameras.

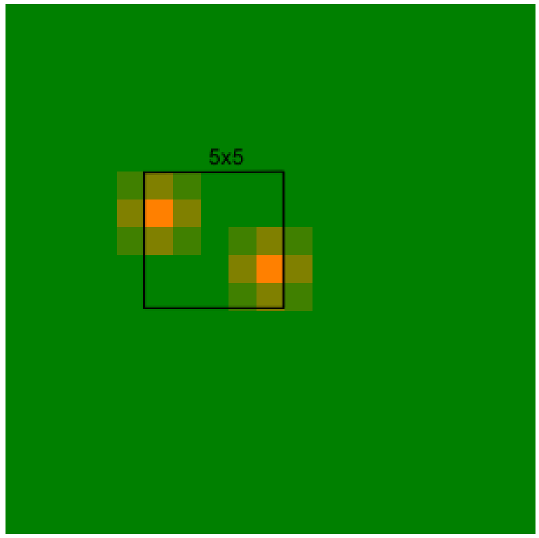


Figure 5.31 Two hot pixels within a 5x5 box after demosaicing

To obtain an initial estimate of this probability, we tiled the sensor with 5×5 squares and asked how often 2 defects appear in that square as the number of hot pixels increase. To this end, we noted that this situation is similar to the generalized birthday problem. In probability theory, the birthday problem (also known as the birthday paradox) deals with the probability that, in a set of n randomly chosen people, at least two of them share the same birthday within an m = 365 day year. The surprising characteristic about this paradox is that the likelihood grows quite fast: for n = 23 the probability of two people sharing the same birthday is 50%, for n = 47 the probability of a shared birthday is slightly greater than 95%. Figure 5.32 shows how quickly this probability grows.

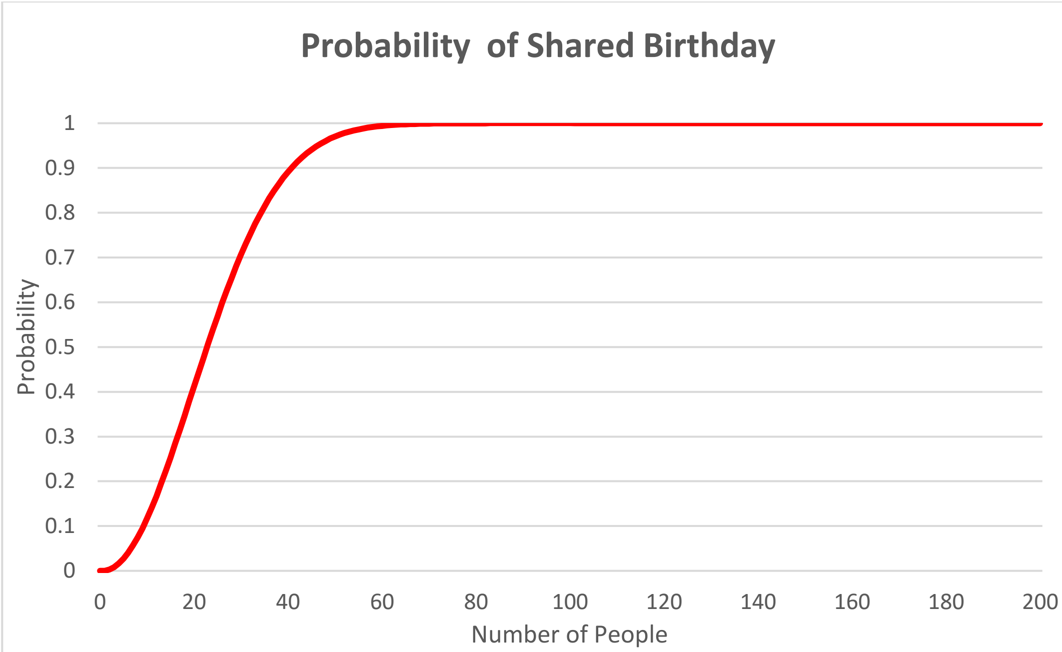


Figure 5.32 The probability of the birthday problem

Directly calculating this probability can be a difficult task; however, it is possible to calculate this probability using another statistical artifice: the probability of the complement of this event. Namely, we calculate the probability of no pairs of people sharing the same birthday (which is the complement of the situation we are interested in) and subtract this probability from 1. From Hill [73] we can derive a generalized birthday paradox formula that can be used in hot pixel probability calculations [57]. This formula is expressed in Equation 5.20, where n is the number of hot pixels necessary to have a probability P of two hot pixels occurring in the same 5×5 box, in a sensor with m 5×5 pixel square boxes ($m = Mpix/25$).

$$P = 1 - \exp\left(-\left[\frac{(n-1)n}{2m}\right]\right) \tag{5.20}$$

Solving Equation 5.20 for the number of hot pixels (n) necessary for a given sensor size m to have a probability P (for large m) of this event happening, we found Equation 5.21:

$$n = \sqrt{-2m \times \ln(1 - P)} \tag{5.21}$$

Therefore, using Equation 5.21 we can calculate for a given sensor size, the number of hot pixels required to have different P probabilities of cameras sensors showing this phenomenon. After developing the equations, we performed this test using Microsoft Excel. In Table 5.1 we show the results we have obtained using the Birthday problem. We calculated it for a range of probabilities and different megapixel counts. In this table we display the results for 1%, 8% and 20%, using 10, 20, 40 and 100 megapixels.

Table 5.1 Results found for the generalized birthday problem for hot pixels

Sensor (MPix)	Square (Pix)	Hot Pix (P=1%)	Hot Pix (P = 4%)	Hot Pix (P = 8%)	Hot Pix (P=20%)
10	5x5	90	180	259	423
20	5x5	128	259	350	598
100	5x5	284	572	817	1337

The results presented in Table 5.1 show that the number of hot pixels necessary to have this event occurring is quite modest. For example, in Table 5.1 we can observe that in 10 megapixel cameras, it only requires 90 hot pixels occurrences to have 1% of the cameras having hot pixels that will overlap. To put this into perspective, this means that in a group of 100,000 10 megapixel cameras, 1000 of them will display this behaviour when their sensor degradation reaches at least 90 hot pixels. When we increase the number of megapixels in a sensor, the number of hot pixels required to have a 1% probability will also increase but with a square root relation. This means that as the camera sensor increases in pixels, the number of hot pixels necessary for this event increases far more slowly. From Equation 5.21 we see that n grows as a square root of the logarithm of one minus the probability and the logarithm function grows far more slowly than a linear function. Therefore, when the square root of the logarithm is applied, it results in an even slower probability growth. This shows that, as we see in Table 5.1, increasing the number of megapixels does not make it less likely that two defects in a 5×5 square will happen, because the number of hot pixels that are necessary for this event grows really slowly. Similarly, because it is the square root of the logarithm, increasing the probability by, for example 8 times, only means a bit more than just doubling the number of hot pixels necessary and that is irrespective of the number of megapixels.

The results from the generalized birthday problem allowed us to obtain the first estimates of the probabilities and the range of hot pixels required to observe this event happening in digital cameras. However, the results we have found are the lower limit on this probability. The birthday problem does not take in consideration when two hot pixels are in different adjacent 5×5 squares, but still within the 5×5 distance to interact. This statistical approach we tested only searches for multiple hot pixels in the same 5×5 pixel box. Therefore, to truly investigate this problem, we used another mathematical technique: The Monte Carlo Simulation. The Monte Carlo simulation, also known as the Monte Carlo Method, is used to estimate the possible outcomes of an uncertain event. This technique uses a set of outcomes based on an estimated range of values instead of a set of fixed input values. In other words, the Monte Carlo Simulation builds a model of possible results by calculating a probability distribution, such as a normal or uniform distribution, for any variable that has intrinsic uncertainty. Then, it recalculates the results repeatedly, each time using a different set of random numbers between the minimum and maximum values [64]. Usually, in Monte Carlo experiments, this process is repeated thousands of times to produce a large number of possible outcomes.

For our experiment, we developed a Monte Carlo Simulation program in MATLAB to estimate the number of cameras that would show this behaviour. First, this algorithm creates random x and y addresses that would correspond to the hot pixel's locations. As has been considered, hot pixels are caused by random cosmic ray striking the digital sensor. Therefore, the addresses of the hot pixels have a random nature and do not follow any particular order. Then, for a given set of hot pixels addresses we asked if any two hot pixels are in a 5×5 square. We performed this experiment 10,000 times for a range of camera sizes and probabilities. Each simulation and given set of hot pixel address is considered a different camera. Then, after the experiments have been concluded, we asked how many cameras in the 10,000 simulations have at least 1 case of two hot pixels in a 5×5 box for a given number of hot pixels. In addition, to understand the relationship between the results from the Birthday problem and the Monte Carlo simulation, we used the number of hot pixels that has been predicted for different probabilities using the Birthday problem. This means that, for example for P = 1%, the Monte Carlo simulation populated a 10 Megapixel sensor with 90 hot pixels. For an 8% probability, this software populated the same sensor size with 259 hot pixels, and so on. An excerpt of the results is shown in Table 5.2.

Table 5.2 Results found for Monte Carlo Simulation

Sensor (MPix)	Square (Pix)	Number of Hot Pixels	Birthday Problem (%)	Monte Carlo (%)	MC/BD
10	5x5	90	1.0	4.46	4.46
	5x5	180	4.0	17.36	4.40
	5x5	259	8.0	33.00	4.12
	5x5	423	20.0	65.89	3.29
20	5x5	128	1.0	4.49	4.49
	5x5	259	4.0	17.34	4.24
	5x5	520	8.0	30.66	4.17
	5x5	679	20.0	66.16	3.31
100	5x5	284	1.0	4.53	4.53
	5x5	572	4.0	17.73	4.43
	5x5	817	8.0	33.40	4.18
	5x5	1337	20.0	65.71	3.28

Analysing the results from the Monte Carlo simulations displayed in Table 5.2, we can see that the accurate probability of having two hot pixels in a 5×5 box is much higher than that predicted by the Birthday problem in Equation 5.21. For the 90 hot pixels in a 10

Megapixel sensor, the Monte Carlo results show that 4.46% of the cameras will have two hot pixels within a 5x5 square. If the number of hot pixels are doubled to 180, this probability is more than quadruples, going to 17.36%, whereas for the same 180 hot pixels, the Birthday problem calculated a 2.01% probability. This is also observed in bigger camera sensor sizes, as the results for the 20 and 100 megapixels show.

One aspect of the results that caught our attention was the ratio between the Monte Carlo simulation and the birthday problem probabilities, expressed in Table 5.2 by the MC/BD column. For example, the 1% probability for the Birthday problem, we see that the Monte Carlo simulations estimated an of occurrence of ~4.5% for all sensor sizes. This steady ratio was also observed for all the other Birthday probabilities. Therefore, the results from the Monte Carlo simulation and the Birthday problem, allows us to predict how many hot pixels are necessary to generate a given occurrence (for example, 4.5%) for every camera size. This observation drove us to develop a fitted Equation that would categorize the relationship between the Monte Carlo simulation and the Birthday problem. Using the same linear regression methodology, we explored in Chapter 3, we have found a quadratic regression that best fits the relationship between the MC/BD ratio vs the probability P of the Birthday problem. Mathematically, the fitted curve can be expressed by Equation 5.22.

$$y = 4.112x^2 - 7.509x + 4.651,\ R^2 = 0.9971 \qquad (5.22)$$

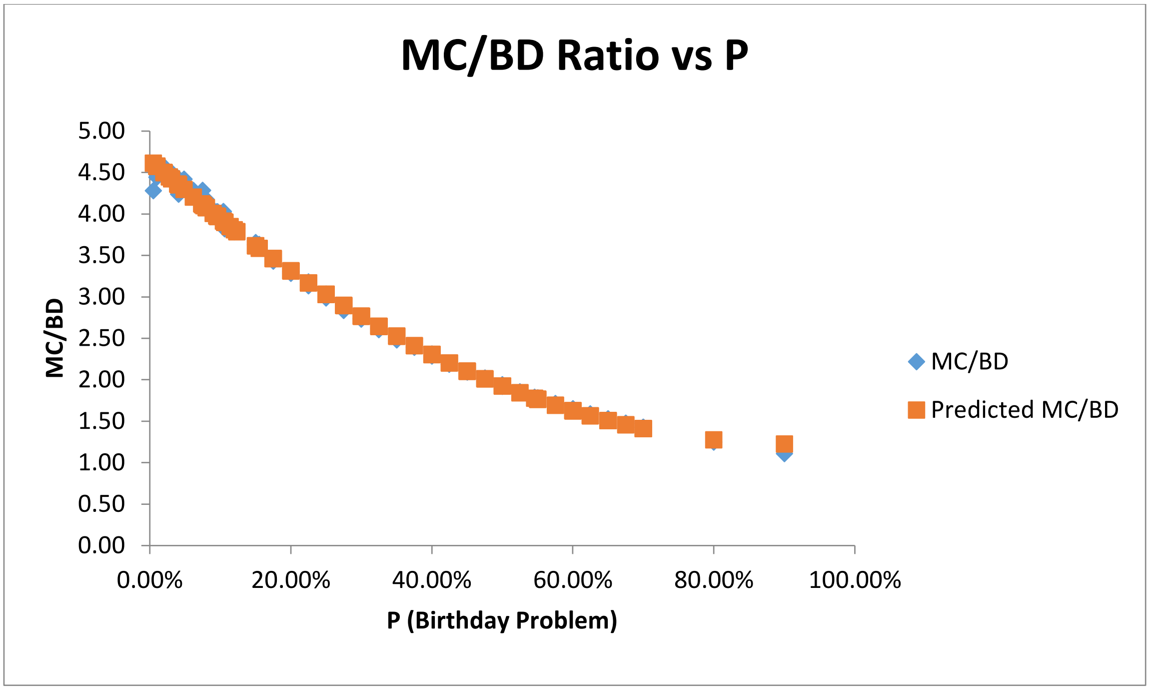


Figure 5.33 The fitted curve for the MC/BP ratio vs the probabilities of the Birthday Problem ($y = 4.11x^2 - 7.5095x + 4.651$ $R^2 = 0.9971$)

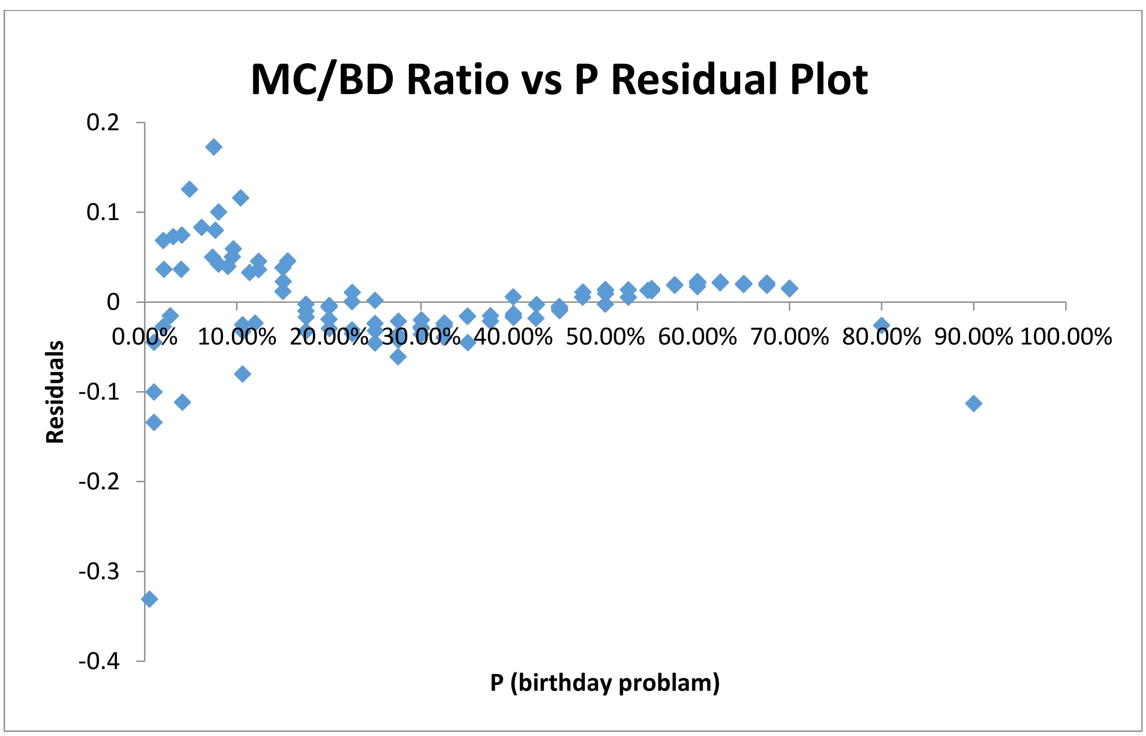


Figure 5.34 Residuals plot of the MC/BD fitted curve

Table 5.3 Standard Error and t-Stat for Equation 5.22

	Coefficients	Standard Error	t-Stat
X^2	4.112	0.112	36.53
X	7.509	0.085	-88.13
Offset	6.651	0.013	355.02

Figures 5.33 shows the Equation 5.22 plot and Figure 5.34 its residuals. In Figure 5.33, we observe that fitted values in orange are almost on top of all the actual original values marked in blue. Also, the residuals' plot (Figure 5.34) indicates that the fitted curve has a strong statistical significance because almost all the error on the ratio is in the 0-10% range and as we increase the Birthday problem probability, the error becomes nearly insignificant. The reason why of most of the error is in the low percentage is due to the fact that the accuracy is not very high when only 90 pixels are being used in a 10 megapixel camera, but the accuracy is good enough for the fit we developed. Additionally, in Table 5.3 is displayed the standard error and the t-Stat for all the coefficients of Equation 5.22. The small errors and high t-Stat values indicate a strong fit and small average errors. Therefore, by applying the probability P for a given hot pixel number in the x variable, we can use Equation 5.22 to find the ratio between the Monte Carlo and the Birthday problems, and consequently, find the probabilities of the Monte Carlo Simulation without having to run complex software.

One of the reasons we have found an MC/BD ratio of ~4 for the smaller probabilities, is related to the fact that the birthday problem does not take into consideration when two hot pixels are in different adjacent 5×5 squares, but still within the 5×5 distance to interact and the Monte Carlo simulations do. However, as observed in Figure 5.33 and Table 5.1, the MC/BD ratio decreases as the probabilities grow because the Birthday problem probability rises rapidly and then saturates (see Figure 5.32). Therefore, this means that when we see the MD/BD ratio decreasing it is because the Birthday problem probabilities are moving into the saturation portion of the curve.

The results from the Monte Carlo Simulations, the generalized Birthday and Equation 5.22 gives us a reliable way to predict the probability of two hot pixels occurring in 5×5 squares, and the number of hot pixels necessary to have this event happening regardless of the sensor size. Additionally, the results show that only a relatively modest number of defects in the camera sensor is required to have two hot pixels in close enough proximity to interact. As we have discussed in Chapter 4, it requires a fairly short amount of time for a camera sensor to develop a significant number of hot pixels. To put this into perspective, consider a typical DSLR level camera with 20 Megapixel sensors using a 4.3 micron pixels. From Table 4.2 we see that this event occurs at a 4.49% rate (1 in 22 cameras) when 128 hot pixels are present in the sensors. Using our current data and defect growth model (Equation 4.7 in Chapter 4) we can get estimates of when 128 hot pixels will occur in function of the image ISO. This suggests that for this camera, at ISO 6400, this would happen in 1.4 years. At ISO 3200 (preferred for lower noise levels) it would be 3.2 years for the 128 hot pixels occurrences. Therefore, we see that there is a 1 in 22 chance that two hot pixels damages will exhibit a noticeable effect in current DSLR cameras after only 3 years of use. These are not unreasonable times for these expensive cameras to be used for, as DSLRs that are built to last for many years. Additionally, as we have described in the previous section, the demosaicing and the different JPEG compression levels will enlarge the effective size of the interaction of multiple defects significantly, creating visual artifacts and decreasing the overall image quality.

Note there is an important implication of these 2 hot pixels interacting equations and the probabilities they predict. The assumption that you can treat hot pixels (defects) as not being close has allowed manufactures to assume they could just mask out such defects with simple interpolations from neighbour pixels. However, these predications show that the probability of close hot pixels is actually high in reasonable periods. The attempt to mask out near (within 5x5) hot pixels does not work when the two are that close and creates even larger defects.

5.7. Summary

This chapter has examined the demosaicing interpolation process that is used to create final images in CFA sensors. There are many demosaicing algorithms that manufacturers may use. However, in this chapter, we have explored three demosaicing algorithms that are used in most digital cameras: Bilinear, Median and Kimmel. This chapter has also discussed the moiré patterns and artifacts created by the interpolation method. We have shown that all three demosaicing algorithms tested inherently create visible artifacts. Then, we explored how the demosaicing process combined with the JPEG compressions significantly impact the hot pixel damaged area spreading the defect to the neighbouring pixels. Lastly, we discussed the probability of two hot pixels being in a 5×5 square, and we have shown that it requires a modest number of hot pixels to have a significant probability of this event happening. The conclusions we have drawn from the analysis performed in this chapter is that a modest amount of hot pixels will greatly affect the overall image quality. The next chapter will explore how the human eyes perceive permanent defects and if we are as sensitive to hot pixels as our research suggests.

Chapter 6.

Perceptual Identification of the Hot Pixel Damage Area

6.1. Introduction

In previous chapters we have shown that hot pixels are an inherent problem of modern digital cameras and that their damage is further aggravated by the multiple processing steps that are performed by the camera. As we have discussed, there is a common misconception that a few hot pixel defects do not affect the overall image quality and are not really noticeable by the end user. In common statistical metrics, a few isolated defective pixels in 10’s of megapixels do not change the metrics and can be easily overlooked in images that contain millions of pixels. In fact, standard metrics such as PSNR (Peak Signal-to-Noise Ratio) often evaluate the impact of these defects as negligible. Indeed, part of the problem is that researchers just insert the defect as single pixels in the final image, rather than at the raw image level as we did in Chapter 5.

We have shown in the previous chapter that a modest number of hot pixels will affect the local image quality of an image, creating visible artifacts. However, the question that remains is: can the human visual system actually detect artifacts created by a small number of hot pixels? By its very nature, the human visual system (*i.e.* a combination of the brain, nerves and eyes) is tuned to spot things that are unusual in a scene. Humans in general can easily detect unexpected, repeated differences in a local area of an image. However, this is not what happens in common statistics metrics analysis, because these metrics consider the overall image and disregard smaller changes in local areas. In this context, little is known about how sensitive the human visual system is to images that contain permanent defects. To the end users, especially to many photographers, an area in the sensor that contains permanent defects in the same location can become more noticeable in a sequence of pictures even when the scenery changes. Therefore, we wanted to explore how noticeable a small amount of defects are to end users. In this chapter, we will discuss the methodology we have developed to help answer this question and report the results of the perceptual experiments we designed to explore the human visual system sensitivity to permanent defects. The goal of this chapter is to lay down the foundation for a future extensive project that can potentially be uploaded to a citizen science platform, and be performed by thousands of subjects, to provide a comprehensive answer to this question.

6.2. Methodology

This section discusses the methodology used to create the images and the test procedure we designed to perform the experiments. One inspiration for this research was the experiments discussed in P. Longère et all [80]. In this article they discussed the methodology and results of perceptual experiments designed to real people evaluate the efficacy of different demosaicing algorithms. In their research they asked the subjects to select which picture within an image set was the most pleasing to them. There are a number of perceptual criteria that can be used to evaluate how humans see permanent defects in images. For example, subjects could be asked which image from the ones displayed had the best quality from their perspective. In this context, using subjects that are familiar with photography and image processing, researchers could ask subjects which image was visually the closest to the true full-colour image. However, the most important question to our research is to evaluate how sensitive the human eye is to a small number of hot pixels. Hence, this was the area we were interested in exploring. Therefore, we concluded that for our experiments the appropriate perceptual question to ask the subjects was: “Do you see any hot pixel defects in this image?” and if the answer is positive, then the next question would be: “Can you draw a circle around them?”. The following paragraphs describe our experimental methods. We will start by describing how we created the images for this experiment and then how the test was performed.

6.2.1. Creating the Images

We decided to use real images to develop this test. Creating artificial images or using a uniform coloured background would not accurately represent what most users see in their pictures. Real images have complex layers and there are many colour variations and different edges. Therefore, we decided to use images that have already been captured. Those image sets were captured by the writer in British Columbia and Brazil, under various daylight conditions, using a Canon T3i DSLR camera and a 50 mm lens or an 18-55 mm lens. The native resolution of the Canon T3i is 5184×3456 pixels (~18 Megapixels). This DSLR interface software allows access to the RAW images, where no compression, no colour correction, no gamma correction and no tone mapping have been applied to them. Therefore, we collected 24 RAW images that have not undergone any sort of processing and separated them into 4 different sets of 6 images.

The next step was to crop the images into 1920×1080 pixels sections. We decided to use this rectangular size because it corresponded to the typical resolution of most monitors (1080p) and when the image is displayed in full-screen mode, each pixel in the image will correspond to one pixel on the monitor. Figure 6.1 displays this step. For this experiment we used the Adobe Photoshop image editing software to develop all the images for this test. We chose Photoshop because it is the industry standard software for photo editing and manipulation. It is widely used among photographers, designers, artists and in the scientific field. Photoshop has a wide variety of editing tools that allows the user to crop images, retouch, enhance colours, eliminate imperfections, blend layers and many other features. Photoshop is a very powerful photo editor that enabled us to generate the desired images for this experiment and allows us to add the hot pixel in a way that reflects the camera processing.

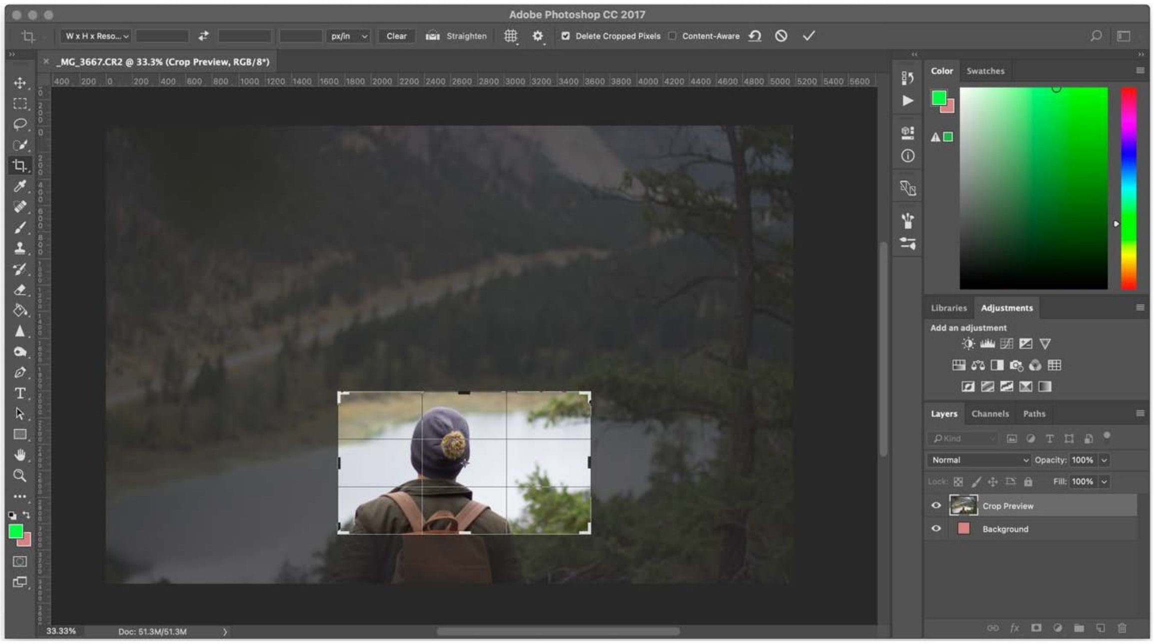


Figure 6.1 Cropping a 1920 x 1080 section of a RAW image

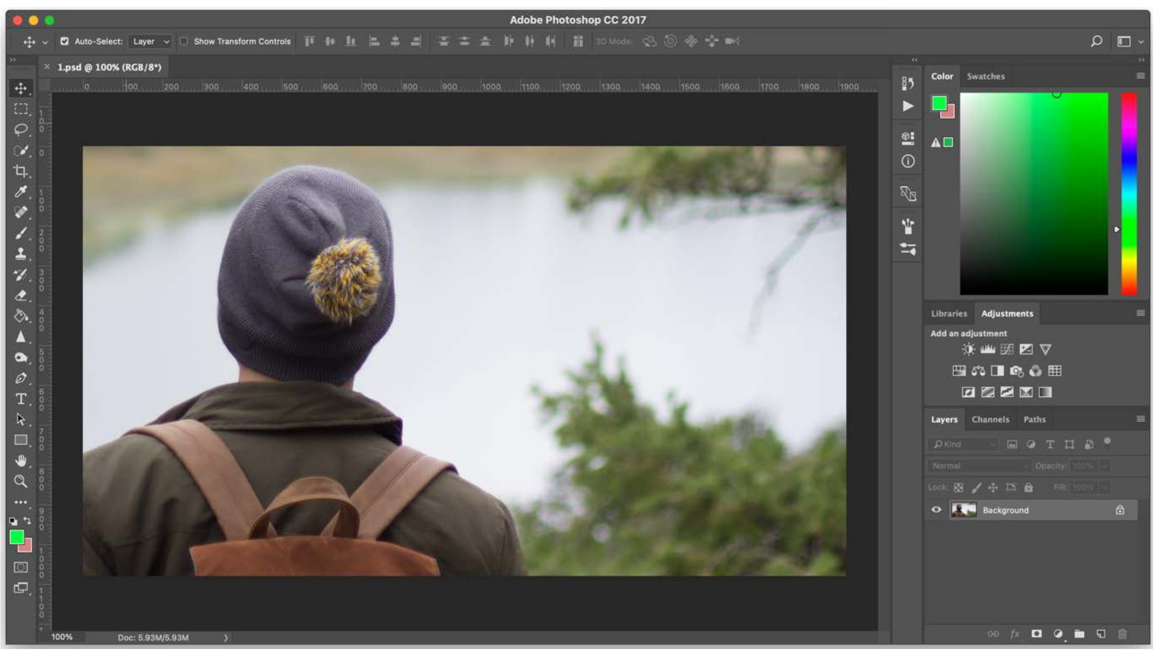


Figure 6.2 1920 x 1080 cropped image

After a 1920×1080 section was cropped from the image, the next step was to insert a pair of artificial hot pixels in the right locations. We decided to start with only two hot pixels because we wanted to explore the human eye sensitivity to a very small number of defects. It is important to note that once the hot pixels' location is selected for one image, this location will remain the same for all the other images in the same batch. This means that in a batch of 6 different images, all of them will have the two hot pixels at the exact same position. This was established because permanent defects from the same digital sensor do not change location in a set of images. They are static defects that will display the defective behaviour regardless of the scenery being captured. However, every batch had a different hot pixel location, because we wanted to test the sensitivity of the human eye to hot pixels in different locations of an image sensor.

The correct location of each hot pixel depends on its colour, and as we have discussed in previous chapters, the CMOS sensors use the Bayer filter on top of each pixel cell to produce a coloured image. The hot pixel itself does not have a particular colour, however, in the interpolated image it is interpreted as red, green or blue because there is a colour filter on it. After analysing the specific Bayer Filter used in a Canon T3i, we were able to distinguish which RBG colour filter the pixel cell has by checking its position. We concluded that if the pixel location is in an odd column and odd row, it is a red pixel. On the other hand, when the pixel is located in an even column and even row it is a blue pixel. Lastly, if this pixel is positioned in an even column and odd row, or odd row and even column it is a green pixel. For the first tests of this experiment, we decided to create two red hot pixels in a 5×5 box. The two hot pixels are separated by 5 pixels horizontally and 5 vertically. This step is depicted in Figure 6.3.

At the outset, we performed all the colour interpolation calculations manually, and then we edited the results in the picture. This was performed according to the following steps. First, we located two red-layered pixels in a 5×5 square that were positioned in an odd column and an odd row. Then we used the drop tool on Photoshop to collect all the RGB colour values on those two pixels. Second, we changed the value of the red channel to 255 (*i.e.* full saturation), but kept the blue and green colour channels intact. Third, we manually calculated the damaged area caused by the Bilinear demosaicing using the colour values we had collected from the previous step and the red hot pixel. We then edited the results in the neighbouring pixels. As we have considered in Chapter 5, the demosaicing interpolation is performed by all cameras to create the final coloured image. Also, the hot pixel damage area is spread by the demosaicing process, therefore we added these side effects to properly replicate what is seen by the end user. For this experiment, we have chosen to use the Bilinear demosaicing because this interpolation method is the first step for many other demosaicing processes and from the tests we explored in Chapter 5, this method is the one that creates the strongest colour shifts around the neighbouring pixels but has the smallest area effect. Figure 6.3 and Figure 6.4 depict these steps in a relatively uniform background. Figure 6.5 displays the hot pixel addition in a busier background.

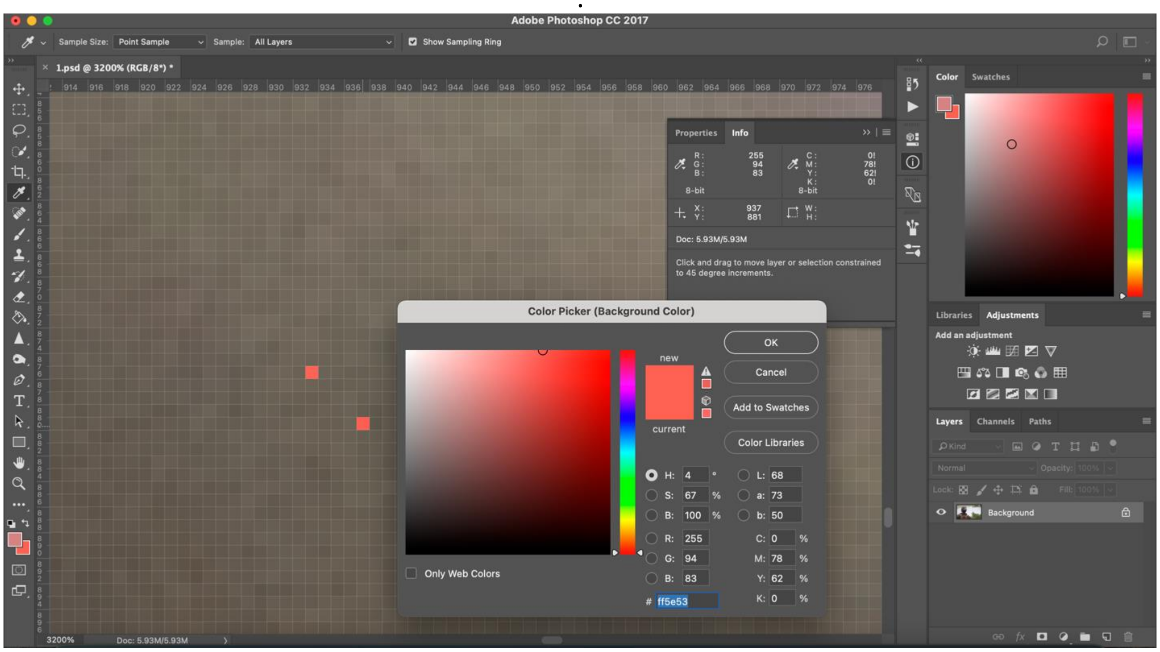


Figure 6.3 Creating two artificial hot pixels in an image

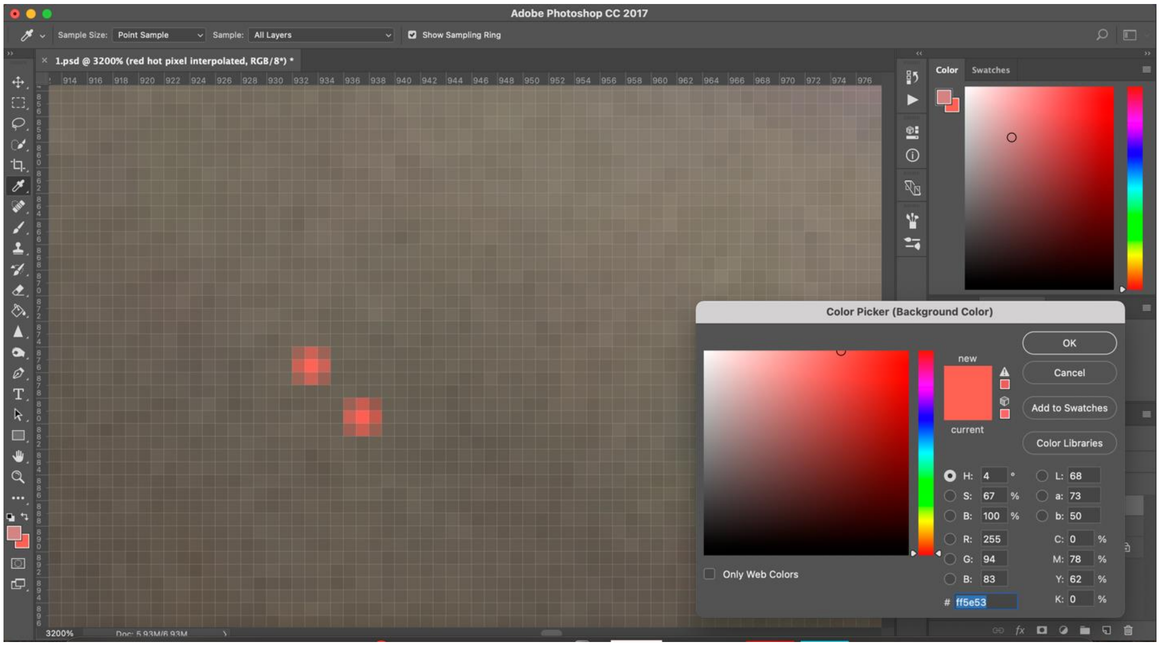


Figure 6.4 The Bilinear interpolation artifact

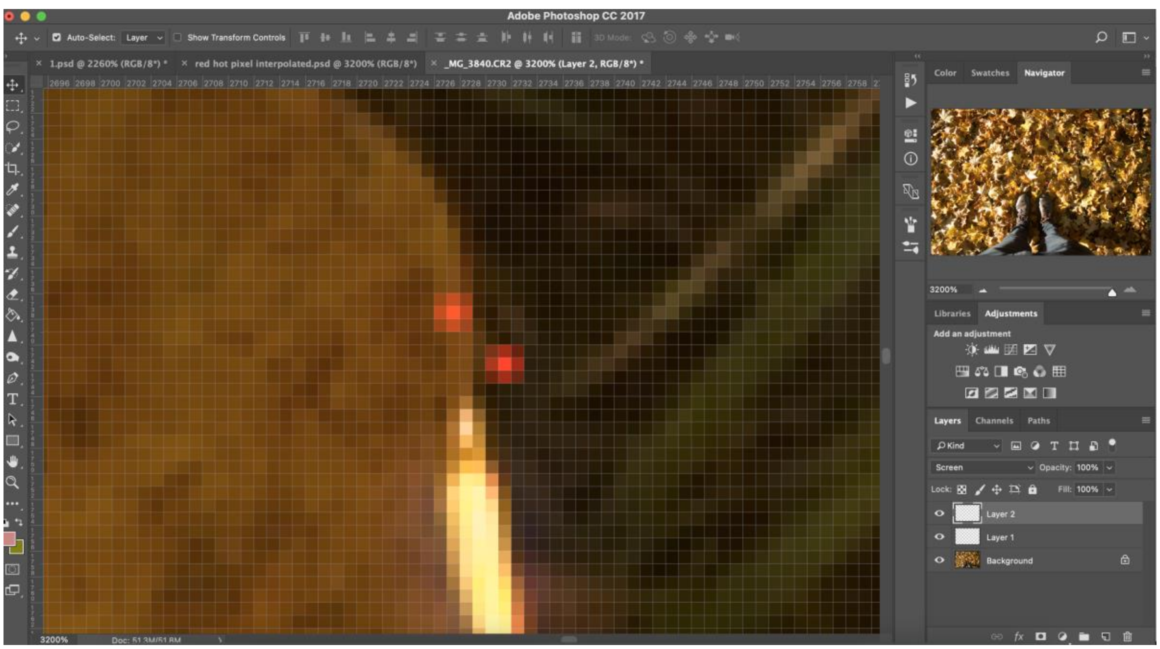


Figure 6.5 The Bilinear interpolation artifact in a busier background

Manually creating these hot pixels and calculating the bilinear interpolation to be added requires a lot of time as it is a detailed process. Therefore, we developed a different methodology that accelerated the insertion of the two red hot pixels in the image. In a separate file, we created a single red hot pixel in a uniform, pure black background and added the Bilinear artifact to the neighbouring pixels. This file was a 3×3 image that contained a single red hot pixel with the Bilinear interpolation artifact. The red hot pixel was set at 255, and the neighbouring pixels followed the Bilinear demosaicing calculations. Then, taking advantage of the native blending modes of Photoshop, we placed this file in a layer on top of the image we were editing and used the “screen” blending mode to add the red hot pixel in the image. The screen blending mode looks at each channel’s colour information and multiplies the inverse of the blend and the base colours [72]. Using the screen blending mode with anything that is pure black will produce no changes and it will disappear from the view. However, using it with pure white will keep the white value unchanged, and any shades that are between pure black or pure white will always result in a lighter colour. The effect is similar to projecting multiple photographic slides on top of each other. In Figures 6.6 and 6.7 we display the processes described in this paragraph and the results.

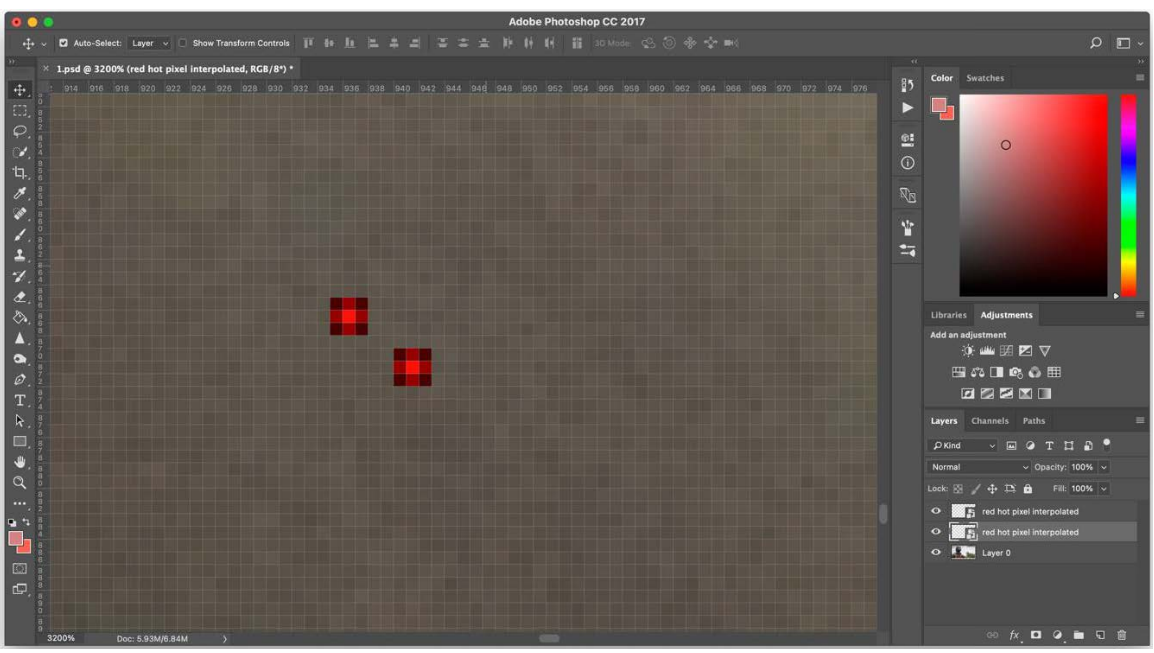


Figure 6.6 Two hot pixels being added

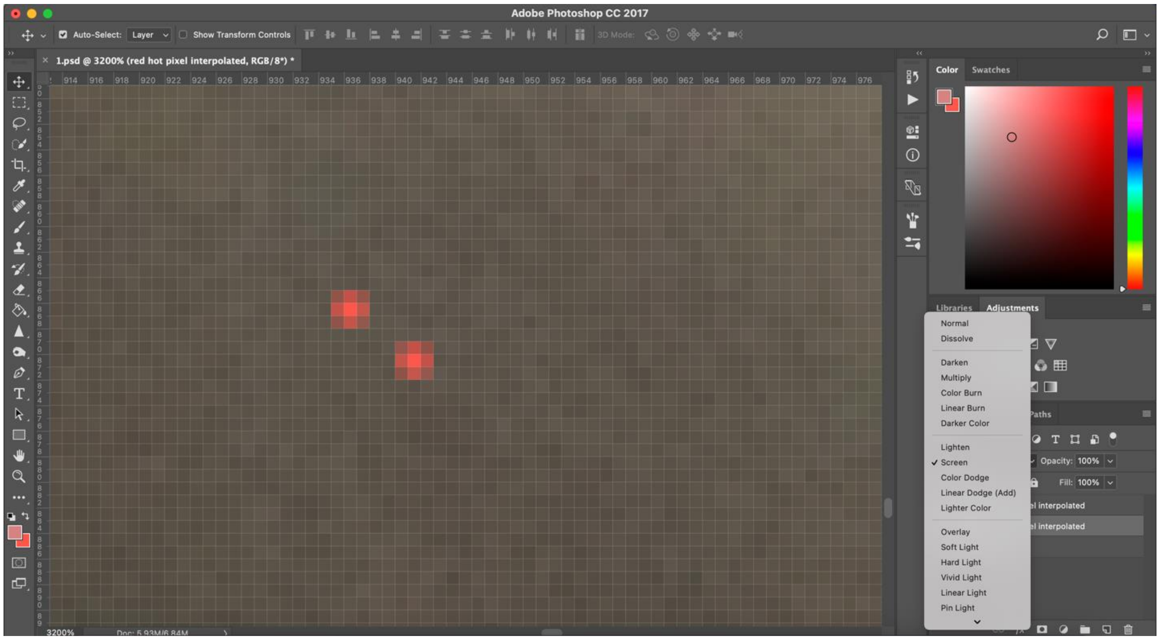


Figure 6.7 Two hot pixels added after the blending mode

Finally, after adding the hot pixels using the screen blending mode, we used the drop tool to collect the values for all the colour channels of the 18 pixels and compared it to the ones created manually. The results of this confirmation step have shown that the final values after the blending mode are the same as seen from the manual calculations, showing that this methodology is efficient and saves time. Therefore, this process was used as the final methodology to create the hot pixels in all the images of this experiment. Additionally, using this methodology we created a few images using Green and Blue hot pixels to explore human perception of these colours. However, the main hot pixel colour tested in this experiment was red. Figure 6.8 shows one image used in our tests and two different zoom levels in the red hot pixels damage area. Finally, after preparing 18 images in 3 different batches, we distributed them among the subjects. It is important to mention that all the images in the batch contained two red hot pixels, there were no images with no hot pixels. Figure 6.9 shows the whole image set we used for this experiment.

Figure 6.8 An example of an image used in our experiments with two different zoom levels

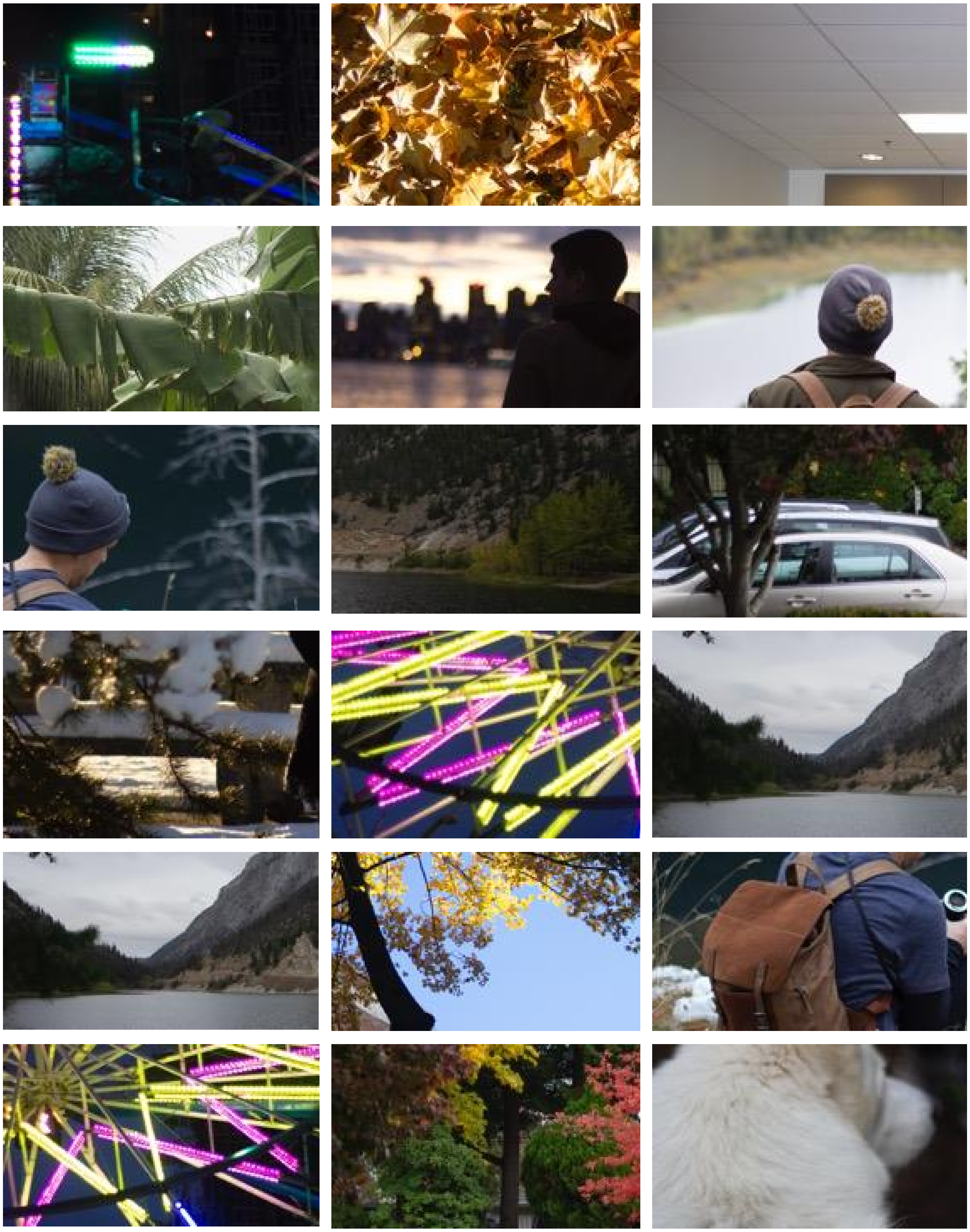


Figure 6.9 Hot pixel image set

6.2.2. The Test Methodology

As described above we created different sets of images, each batch contained 6 pictures, all the pictures were different from each other, but they had two hot pixels in the same location. Each subject received an online link to download the 4 batches and instructions to perform the test remotely. The setting for the tests was established to be remote for two main reasons: first, to comply with the provincial safety guidelines for the COVID-19 pandemic and second, this is how these tests would be performed if we wanted it to be accessed by many people in the future.

The different sets of images were uploaded in a storage service and sent to the observers. The images were uncompressed and sent at full quality using the TIFF format. We used a total 5 subjects. All the subjects reported they had normal colour vision, and they had normal or corrected to normal visual acuity. The subjects were volunteers from the Simon Fraser University that were working with Dr. Chapman during the Fall Semester of 2020. All of the subjects were informed of the purpose of the experiment and they all had an entry-level understanding of what hot pixels are and what type of artifacts they can possible create. They were also knowledgeable in image processing and understood photography. However, all were naive as to which colours were the hot pixels and at which position they were located; thus the subjects would not purposely focus on any particular colour or area of the image.

The observers were instructed to view the images presented according to their natural viewing conditions; that could be a room illuminated from the ceiling light, natural sunlight or any other form of illumination that was the standard for the subject. These viewing conditions were chosen because we wanted to get closer to the natural conditions under which many users would view digital photographs. The subjects were instructed to display the image in full screen and to not use any zoom features, or to try to enhance any aspects of their natural viewing conditions. Additionally, to participate in this experiment the subjects needed to have one monitor with least 1920×1080 pixels of resolution. As we have considered in the previous section, this is the corresponding size of the images created and when the image is displayed in full screen, every single pixel in the monitor is equivalent to every single pixel in the original image.

The subjects were allowed to view the images for 1-2 minutes. However, in the actual operation, the subjects identified it so quickly that the actual time spent was measured in seconds, not in minutes. The subjects were asked two questions: “Are there any hot pixels in this image?”. If they answered yes, the question that followed was “Can you draw a circle around each of them?”. For the second question the subjects were allowed to zoom in as they already had spotted the hot pixels. Also, they could use any marking software tools that were available to them. Then, after answering both questions for the first image, they would move to the next image in the batch. That image would be completely different from the one previously displayed, but the hot pixel addresses would be the same. As we have discussed in previous chapters, hot pixels are permanent defects on the sensor and they will have the same address across all the images captured by the camera. Finally, when all the images from the batch were completed, they would return their results for analysis and start the next batch, until all batches were completed and returned.

6.3. Results and Discussion

The results of the experiment are displayed in Table 6.1. They are expressed as the percentage of how many hot pixels the subjects were able to properly identify and locate in each image. Thus, each subject was assigned a score between 0% and 100%. The higher the percentage, the more accurately the subject had located and marked the two hot pixels in each of the images in the batch.

Table 6.1 Results of human identification of images with hot pixels

Subjects	Number of Pictures Seen	Identification rate
S1	18	100%
S2	18	100%
S3	18	100%
S4	18	100%
S5	18	100%

The results displayed in Table 6.1 were surprising. It shows that all the subjects were able to visually detect and precisely locate the two hot pixels in an image that contained 2 Megapixels. This shows an impressive consistency across the subjects. Also, the subjects identified the defects in a much shorter time than the initial plan of 1-2 minutes, in fact, they reported they were able to identify the hot pixels in seconds. This tells us, at first approximation, that the human eye is very sensitive to outliers in digital images. Note that these are the demosaicing hot pixels in pairs. The important point here is that the demosaicing spread pixels, especially in a 5×5 area, are highly identifiable because they are the real results to be expected, not the single pixel added onto an image that most people have assumed. However, it is important to highlight that these results are from a small number of people and this group of subjects was knowledgeable and aware of the concepts of hot pixels and image artifacts. It remains possible that less consistent results could have been observed if we had used a more heterogeneous population of subjects (*e.g.,* a wider range of cultural backgrounds or ages), a bigger group of people and subjects that have little or no information about hot pixels.

Two aspects of the experimental results caught our attention. First, the subjects were able to detect the hot pixels regardless of their colour, however, the subjects informed us that when the hot pixel was blue, it was harder to identify. Second, the subjects reported that when they found the hot pixels in the first image, they were able to more quickly identify the hot pixels in the other images of the same batch. This demonstrates that a permanent defect is very visible to the human eyes when we visually analyse them in a sequence of pictures.

The results from this experiment are just an initial step to explore this question deeply. This data has shown us that a small number of permanent defects is greatly noticeable by the human eye, and that regardless of the colour of the hot pixel or the image background, we will perceive these hot pixels as image artifacts. However, this was performed with a small number of subjects, but the results were so strong, that we plan to expand this experiment to larger group of subjects, with different backgrounds and that are unknowledgeable with respect to hot pixel characteristics. This will allow us to answer this question properly and draw solid conclusions.

In future research we plan to use a citizen science methodology in this project. Citizen science is the practice of public participation and collaboration in scientific research to increase in scientific knowledge [82]. Collaboration in citizen science involves scientists and researchers working with the public. Volunteers have different backgrounds and various levels of expertise. The history of citizen science started in astronomy, when scientists counted on the willingness of the public to participate in identifying the shape of different galaxies. A large number of people volunteered to participate and the scientists had phenomenal results and found new discoveries that they did not expect. Modern advances in technology make citizen science more accessible today than ever before, online platforms such as Zooniverse allow people of all ages and backgrounds to participate in real research.

Moreover, in this bigger project, we would investigate different aspects of the image degradation caused by hot pixels. Additional questions could be asked in this experiment, such as: “Which image do you find most pleasing?”, “Do you see a decrease in quality?”, “Which hot pixel for you is more evident, red, green or blue?”. These questions would help us answer not only how humans perceive hot pixels but also how the hot pixels affect the overall image quality according to the end user perspective. Additionally, we could perform the standard metrics, such as PSNR (Peak Signal-to-Noise Ratio), in those images and compare the results to the ones we have found from the perceptual assessment. This would allow us to understand the sensitivity of the human eye compared to software metrics and draw proper conclusions if we perceive permanent defects with better accuracy than a computer. We would also expand beyond the simple demosaicing to include the JPEG effect.

In summary, we conclude that these preliminary results justify opening these tests to a much larger group of subjects to verify the hypothesis explored in this chapter (*i.e.* the average human is more sensitive to the damage caused by hot pixels than the standard mathematical analysis). If the results from a wider set of subjects show that this hypothesis is true, this would strongly suggest that the standard mathematical analysis used to identify the impact caused by defects is not accurate and new mathematical metrics would need to be obtained.

6.4. Summary

This chapter has shown the experimental methodology we have created to acquire a first idea of how sensitive the human eye is to permanent defects and the results we have found. It is a common misconception that a modest number of hot pixels would not be detectable by the end user, but as we have shown in this chapter, this is not the case. These preliminary results have shown that the human eye is very sensitive to hot pixels as there are permanent defects in the same location in a sensor and they become more noticeable when analysing them in a sequence of images. However, in order to have a comprehensive answer to this question, it is necessary to expand the scope of this experiment with other tests and to perform it with a greater number of subjects with different backgrounds.

Chapter 7.

Conclusion

7.1. Summary

The early chapters of this thesis explored many aspects of the digital sensors in great detail. First, we discussed the emergence of digital cameras to current days where it is an essential feature of many everyday devices. In Chapter 2, this thesis outlined the digital sensors with emphasis on the CMOS design. The CMOS sensor is the current industry's favourite pixel design for most digital camera applications. We also discussed the technology behind the photon detection process, allowing us to understand the active pixel design at a low-level. The current market trend is to develop smaller pixels. However, the reduction in the overall pixel size comes with trade-offs, such as increased noise and lower resolution.

The main purpose of this thesis was to explore what creates noticeable artifacts in digital imagers. As we have discussed previously, the phenomenon of cosmic ray particles hitting the sensor can create defects in these devices. This thesis has focused on the permanent defects, known as hot pixels. Previous work has explored hot pixels and transient defects DSLR and cell phone cameras. However, the literature has not deeply investigated the impact that image processing steps such as the demosaicing algorithms and the JPEG compressions have on hot pixels, and how sensitive the human visual system is to permanent defects.

As explored in Chapter 3, hot pixels were the main focus of this research. According to the literature leveraged for this thesis, these faulty pixels are generally caused by cosmic rays that strike the sensor with enough density to create permanent damage. These hot pixels do not follow the standard pixel response, as they will have significant dark current components in dark field images. Different types of hot pixels were discussed in this paper: the stuck, partially stuck and standard hot pixels, however emphasis was put on the standard hot pixel as it is the most common type of permanent defect found in camera systems. One of the goals of this research was to detect, analyse and understand the growth rate of the standard hot pixel. In this context, this thesis has discussed the dark-frame capturing methodology developed for both cell phone and DSLRs. We have also explored the automated software intervalometers that have the ability to capture the various calibration datasets with minimal user interaction. Additionally, this thesis discussed the hot pixel detection algorithms tested through the years. We tested different algorithms such as the Nearest Noise Criterion that aimed to take into consideration the local noise while doing the first filtering of the hot pixels. However, the software was not able to reliably detect weaker hot pixels. Also, we worked on a new detection algorithm that uses noise distribution to select hot pixel candidates, as opposed to the static Threshold filtering. However, by the time of conclusion of this thesis we were still having problems with this software. Hence, we did not use the results from this algorithm. Therefore, from the different detection algorithms explored, the Threshold method was the one that showed the most consistent and accurate results, that comprehended the entire exposure range. Thus, this was the detection method used to collect the data to build the hot pixel growth model.

In Chapter 4 we explored the methodology we used to create the empirical hot pixel growth model. In this research we continuously gathered data by performing the experimental capture and analysis to refine the hot pixel growth model. The results continued from previous trends. The latest defect growth rate seen in Equation 4.7 indicates that if the pixel size is shrunk by a factor of 2, then the defect rate will increase by 8.9 times. Similarly, if the ISO is doubled, the defect rate will increase the power of 0.522. The results from the hot pixel growth rate has driven us to explore how these defects affect the overall image quality from the end user perspective taking into consideration the different image processing steps that occur before the final image is displayed to the consumers.

In Chapter 5, this thesis walked through the different demosaicing interpolation process that are used to create final images in CFA sensors. In this research we explored three demosaicing algorithms that are used in most digital cameras: Bilinear, Median and Kimmel. Additionally, this research discussed the moiré patterns and artifacts created by the interpolation methods. It was shown that all demosaicing algorithms inherently create visible artifacts. Then, we explored how the demosaicing process combined with the JPEG compressions significantly impact the hot pixel damage area spreading the defect to the neighbouring pixels. We have shown that when these processes combine, they can spread the defect area of one or two hot pixels to up to 256 pixels, creating visual artifacts and significantly decreasing the overall image quality. Moreover, the thesis discussed the probability of two hot pixels being in a 5×5 square, and the results from the mathematical techniques such as the birthday problem and the Monte Carlo simulation have shown that it requires a modest number of hot pixels to have a significant probability of this event happening. For example, the results from these experiments have shown that in 4.49% (1 in 22 devices) of 20 Megapixel cameras will have two hot pixels within a 5×5 square if they have at least 128 hot pixels. Using our current data and defect growth model from Chapter 4, we estimated that for this camera, at ISO 6400, this would happen in 1.4 years. At ISO 3200 it would be 3.2 years for the 128 hot pixels occurrences. Also, the ratio between the birthday problem approximation and the Monte Carlo simulation drove us to create a highly accurate mathematical model, as seen in Equation 5.22, that allows us to generalize the percentage of cameras for a given number of defects and megapixels that will have two hot pixels within a 5×5 square. The conclusions we have drawn from this analysis performed in Chapter 5 have shown that a modest number of hot pixels affects the overall image quality.

Lastly in Chapter 6, this thesis explored the sensitivity of the human eye to identify permanent defects. Very little research has been performed in this area and we wanted to first acquire an idea of how sensitive the human visual system is to permanent defects because there is a common misconception that a modest number of hot pixels would not be detectable by the end user. These preliminary results have shown that in a sequence of images our visual system is very sensitive to hot pixels after demosaicing. These faulty pixels are permanent defects in the same location in a sensor and they become more noticeable when analysing them in a sequence of images. The subjects used in this experiment detected the hot pixels at a 100% accuracy level. However, it was also discussed that these are just preliminary results as we have used a small number of subjects.

7.2. Suggestions for Future Work

This thesis has explored, in depth, the standard hot pixel and how they can create noticeable defects when combined with demosaicing and JPEG processing. However, as we have discussed, there is a significant amount of noise in DSRLs at high ISO levels and in cell phones at modest ISO levels. This noise component creates a significant obstacle for the hot pixel detection process using the Threshold Method, because the noise can be falsely detected as hot pixels and weaker hot pixels can be discarded if they are mistaken as noise. Additionally, noise is not evenly distributed across the camera sensor. Therefore, it is paramount to develop a new detection method that replaces the threshold filtering, and instead adapts to local noise of a sensor at high and modest ISO levels.

Moreover, in this research only a small number of Android cell phones were tested as very few support RAW outputs. As well, Apple has just started to include RAW support for their native camera app in the latest iPhone model (*e.g.* iPhone 12 Pro). Additionally, given the time of this thesis (circa 2021) – the global COVID-19 pandemic made it nearly impossible to share and test different cell phones. Therefore, this future research should be expanded to cover a wider selection of cell phones that includes iOS and Android models. Moreover, as the cell phone manufacturers start to introduce more effective noise suppression algorithms, it would be insightful to analyse how these noise correction techniques will impact the permanent defects. Related to more refined camera systems, as professional and serious amateur photographers continue to migrate to mirrorless models, this research should expand its dark frame analysis in this type of camera, as well as include data from newer DSLR models.

Another important avenue of further expansion is how hot pixels are directly related to the overall image quality. Although we have explored how different demosaicing algorithms combined with the JPEG process will spread the hot pixel damage area, it has a lot of room for growth. For the demosaicing tests, we mainly focused on three demosaicing methods: Kimmel, Median and Bilinear. The future work should be extended to test more advanced demosaicing methods that are used in cameras to provide a better understanding of how other interpolation methods impact the hot pixel's damage area. Also, this research should explore what happens to the probabilities of the Birthday Problem and the Monte Carlo Simulation when we spread the two hot pixels further than a 5×5 area. It should also discuss what are the consequences of this bigger separation to the hot pixel damage area and image quality. Additionally, different experiments in the JPEG compressions should be performed to include hot pixels that present a variety of colours and in non-uniform backgrounds. In terms of image quality, this research should be expanded to analyse how SEU events, when combined with hot pixels, can create stronger artifacts. Using the SEU occurrence results developed in Rohan's research [62], this proposed thesis should develop an updated Birthday Problem and Monte Carlo simulations to express the probabilities of an SEU event occurring in closer proximity to hot pixel defects. This would help us understand how both transient and permanent defects have a role in decreasing the overall image quality.

Lastly, the perceptual assessment of the hot pixel damage area should be greatly expanded to a bigger group of subjects, with different backgrounds and that are unfamiliar with as to hot pixel characteristics. Also, this experiment should be uploaded in a citizen science platform, such as Zooniverse, to reach a wide variety of people. While our results are good with 5 subjects, expanding this to hundreds or thousands via citizen science would add strength to the point that human detection is the standard we must work with and the mathematical metrics we use must reflect that reality. After all, it is the photographer who is going to be upset when their expensive camera starts showing defects.

We suggest building an interactive platform that can assess the many different aspects of how hot pixels relate to image quality. Also, additional questions that we have not covered in the previous experiment should be asked in this broader test, questions such as: "Which image is the most pleasing to you?", "Do you see a decrease in quality?", "Which hot pixel for you is more evident, red, green or blue?". These questions would help us answer not only how humans perceive hot pixels but also how the hot pixels affect the overall image quality according to the end user perspective. Also, we would be able to explore many other different aspects of the effects of different demosaicing and JPEG levels in hot pixels. Additionally, this research should perform the standard metrics, such as PSNR (Peak Signal-to-Noise Ratio), in those images and compare the results to the ones found from the expanded perceptual assessment. This would be fundamental to understand the sensitivity of the human eyes compared to software metrics, and draw proper conclusions if we perceive permanent defects with better accuracy than a computer.

7.3. Closing Thoughts

This thesis has explored in detail the various aspects of permanent defects, putting emphasis on the standard hot pixels. We have experimented and analysed a large set of dark field images to detect and understand the hot pixel behaviour. Additionally, we discussed the algorithm tool (*i.e.* the Threshold Method) used to extract the hot pixel data from the dark field images and provided the latest hot pixel growth model. Moreover, this research also discussed how hot pixels have a significant impact on the overall image quality. Notably, our experiments have shown that when the hot pixel defect is processed by the demosaicing interpolation and the JPEG compressions, it will create noticeable artifacts. In this context, through statistical techniques we also have shown that it only requires a modest number of hot pixels to have a significant percentage of cameras having multiple hot pixels in closer proximity. Finally, we have also shown through preliminary results that the human visual system is much more sensitive to hot pixels than one might have expected, and a small amount of defects can be easily spotted by viewers. Such observations further illustrate the need for greater research on the impact of hot pixels on the image quality and how these defects are noticeable to the end user.

References:

[1] J. Van Dijck, "Digital photography: communication, identity, memory", Visual Communication, 7(1), 57-76. 2008. [Online]. Available: https://doi.org/10.1177/1470357207084865. [Acessed January 2020].

[2] T. Gustavson, "Camera: a history of photography from daguerreotype to digital", New York: Sterling Signature, 2009. Print.

[3] CIPA Report, "Total shipment of digital cameras", 2019. [Online]. Available: http://www.cipa.jp/stats/documents/common/cr300.pdf. [Acessed February 2020].

[4] CIPA Report, "Total shipment of film cameras", 2008. [Online]. Available: http://www.cipa.jp/stats/documents/common/cr400.pdf. [Acessed February 2020].

[5] F. Richter, "Digital Cameras dropped 87% since 2010", Statista Article, [Online]. Available: https://www.statista.com/chart/5782/digital-camera-shipments/. [Acessed February 2020].

[6] S. O'Dea, "Number of smartphones sold to end users worldwide from 2007 to 2020", Statista Article, [Online]. Available: https://www.statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007/. [Acessed February 2020].

[7] F. Richter, "What Smartphones Have Done to the Camera Industry", Statista Article, [Online]. Available: https://www.statista.com/chart/5782/digital-camera-shipments/. [Acessed May 2020].

[8] Apple (CA), "iPhone 11 Pro", [Online]. Available: https://www.apple.com/ca/iphone-11-pro/. [Acessed June 2020].

[9] B. Kronfli, J. Peckham, "The Best Camera Phone in 2020", Techradar, [Online]. Available: https://www.techradar.com/news/best-cameraphone. [Acessed June 2020].

[10] G.H. Chapman, Parham Purbakht, Peter Le, I. Koren, and Z. Koren, "Exploring Soft Errors (SEUs) with Digital Imager Pixels ranging from 7 to 1.3 µm", in *Proc. 2017 Int. Sym on Defect and Fault Tolerance*, Oct. 2017.

[11] J. Leung, "MASc Thesis: Measurement and Analysis of Defect Development in Digital Imagers", Simon Fraser University, Burnaby, Canada, 2011.

[12] R. Thomas, "MASc Thesis: Enhanced Digital Imager Defect Analysis with Smaller Pixel Sizes", Simon Fraser Uni, Burnaby, BC Canada, Aug. 2016.

[13] S. Mukherjee, “Architecture Design for Soft Error*s*”, *Elsevier, Inc.*, 2008.

[14] G.H. Chapman, R. Thomas, R. Thomas, K. Meneses, T. Yang, I. Koren, and Z. Koren, “Single Event Upsets and Hot Pixels in Digital Imagers”, in *Proc. 2015 Int. Sym on Defect and Fault Tolerance*, pp. 41-46, Oct. 2015.

[15] A.J.P. Theuwissen, “Influence of terrestrial cosmic rays on the reliability of CCD image sensors. Part 2: experiments at elevated temperature”, *IEEE Transactions on Electron Devices*, Vol. 55 (9), pp. 2324-8, 2008.

[16] Z. Török and S. P. Platt, “Application of imaging systems to characterization of single-event effects in high energy neutron environments”, *IEEE Trans. Nucl. Sci*., vol. 53, pp. 3718–3725, 2006.

[17] F. Firouzi, M. Salehi, F. Wang, and S-M. Fakhraie, “An accurate model for soft error rate estimation considering dynamic voltage and frequency scaling effects”, *Microelectronics Reliability*, 2011.

[18] S. Mitra, T. Karnik, N. Seifert, and M. Zhang, “Logic soft errors in sub-65nm technologies design and CAD challenges”, *Proc. of the 42nd Design Automation Conference*, DAC '05, pp. 2-4, 2005.

[19] J. Dudas, L. M. Wu, C. Jung, G. H. Chapman, Z. Koren, I. Koren, “Identification of in-field defect development in digital image sensors”, *in Proc. Electronic Imaging, Digital Photography II*, V6502, 6502Y1-0Y12, San Jose, Jan. 2007.

[20] J. Leung, G.H. Chapman, I. Koren, and Z. Koren, “Statistical Identification and Analysis of Defect Development in Digital Imagers” in *Proc. SPIE Electronic Imaging, Digital Photography V*, v7250, 742903-1 – 03-12, San Jose, Jan 2009.

[21] A.J.P. Theuwissen, “Influence of terrestrial cosmic rays on the reliability of CCD image sensors. Part 1: experiments at room temperature”, *IEEE Transactions on Electron Devices*, Vol. 54 (12), 3260-6, 2007.

[22] A.J.P. Theuwissen, “Influence of terrestrial cosmic rays on the reliability of CCD image sensors. Part 2: experiments at elevated temperature”, *IEEE Transactions on Electron Devices*, Vol. 55 (9), 2324-8, 2008.

[23] J. Leung, G. H. Chapman, I. Koren, Z. Koren, “Tradeoffs in imager design with respect to pixel defect rates”, *in Proc. of the 2010 Intern. Symposium on Defect and Fault Tolerance* in VLSI, 231-239., Kyoto, Japan, Oct 2010.

[24] G.H. Chapman, R. Thomas, I. Koren, and Z. Koren, “Empirical formula for rates of hot pixel defects based on pixel size, sensor area and ISO”, *in Proc. Electronic Imaging, Sensors, Cameras, and Systems for Industrial/Scientific Applications XIII*, v8659, 86590C-1-C-11 San Francisco, Jan. 2013.

[25] I. E. L. Rekaa, "Master's Thesis: The Usability of DSLR Cameras for Amateur Photographers – An Empirical case on Design". University of Oslo. Oslo, Norway. May, 2010.

[26] Canon, "PowerShot ELPH 180 Overview", [Online] Available: https://www.canon.ca/en/product?name=PowerShot_ELPH_180&category=/en/products/Cameras/Compact-Digital-Cameras/Point-and-Shoot-Cameras. [Acessed July 2020].

[27] A. Ismall, "The Best Camera Phones in 2020*",* [Online] Available: https://www.tomsguide.com/us/best-phone-cameras,review-2272.html. [Acessed July 2020

[28] I. E. L. Rekaa, "Master's Thesis: The Usability of DSLR Cameras for Amateur Photographers – An Empirical case on Design"*,* University of Oslo. Oslo, Norway. May, 2010.

[29] Canon, "EOS 5D Mark IV Overview"*,* [Online] Available: https://www.canon.ca/en/product?name=EOS_5D_Mark_IV&category=/en/products/Cameras/DSLR-Cameras/Professional. [Acessed July 2020].

[30] SONY, "α9 Feautirng full-frame stacked CMOS Sensor", [Online] Available: https://www.sony.ca/en/electronics/interchangeable-lens-cameras/ilce-9. [Acessed July 2020].

[31] P. Hiranchaowiwat, "Master's Thesis: A Study of Factor Influencing The Purchase of Mirrorless Cameras", Thammasat University, May of 2019.

[32] A. H. Incekara, D. Z. Seker, A. Delen, A. Acar, "Investigating The Suitability of Mirrorless Cameras In Terrestrital Photogrammmetric Applications"*, in ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume IV-4/W4, 2017 4th International GeoAdvances Workshop, 14-15 Ocotober 2017.

[33] H. Phelippeau, M. Akil, B. Dias Rodrigues, H. Talbot, S. Bara, "Bayer Biletaral Denoising on TriMedia3270", in *Proceedings of SPIE*. February 2009.

[34] S. Fumihiko, T. Asaida, "Image Defect correcting curting for a solid state imager", U.S. Patent 5144446, Sep. 1, 1992..

[35] LUCID Vision Labs, "Understanding The Digital Image Sensor", [Online]. Available: https://thinklucid.com/tech-briefs/understanding-digital-image-sensors/. [Acessed July 2020]

[36] T. Bergmuller, L. Debiasi, A. Uhl and Z. Sun, "Impact of sensor ageing on iris recognizion", Authentic Vision, Austria.

[37] I. E. L. Rekaa, " Master's Thesis: The Usability of DSLR Cameras for Amateur Photographers – An Empirical case on Design", University of Oslo. Oslo, Norway. May, 2010.

[38] R. Heald, "How Cosmic Rays Cause Computer Downtime", IEEE Rel. Soc. SCV Meeting, 2005.

[39] A. Ponting, "The Aperture and Exposure", [Online] Available: https://steemit.com/photography/@apteacher/the-aperture-and-exposure. [Acessed July 2020].

[40] E. Vlietinck, "Introduction to Colour Management for Printing" [Online] Available: https://www.clipstudio.net/how-to-draw/archives/156866. [Acessed July 2020]

[41] G. H. Chapman, "Bilinear Interpolation", SFU, ENSC 895: Optical Engineering and Laser Applications, Sep-2018 [Online]. Available: http://www.sfu.ca/~gchapman/e895/e895l11.pdf.

[42] ARRROW, "Introduction to Bayer Filter*s*", [Online] Available: https://www.arrow.com/en/research-and-events/articles/introduction-to-bayer-filters. [Acessed June 2020]

[43] D. Groom, "Cosmic Rays and Other Nonsense in Anstronomical CCD Imagers", *Lawrance Barkeley National Laboratory*, pp. 301-311.

[44] B. A. Nasi, S. Saadoon, A. Adbullabas and H. K. Easa, "Power Electronic", *Power Electronic*, p. 5, May 2018.

[45] Z. H. A. Azyan, "Master's Thesis: An Investigation into the Practicality of Using a Digital Camera's RAW Data in Print Publishing Applications", Rochester Institute of Technology. May 2005.

[46] H. V. Dwivedi, "Bachelor's Thesis: Design of JPEG Compressor", National Institue of Technology, Rourkela.2009.

[47] M. A. Green, "Optical Properties of Silicon," PVEducation, [Online]. Available: https://www.pveducation.org/pvcdrom/materials/optical-properties-of-silicon. [August 2020].

[48] Opto Electronics, "Photovoltaic Devices and Applied". February 2021. [Online] Available: https://opticforelectronic.wordpress.com/2015/01/08/12-photovoltaic-devices-and-applied-part-2/. [Acessed August 2020].

[49] U. Jain, "MASc: Characterization of CMOS Image Sensor", Delft University of Technology, Delft, Netherlands, August 2016.

[50] CANON, “CMOS Sensors”, [Online]. Available: https://global.canon/en/technology/s_labo/light/003/05.html. [Acessed August 2020].

[51] G. Meijer, M. Pertijs, K. Makinwa, “Smart Sensor Systems: Emerging Technologies and Applications”, *John Wiley & Sons. Inc*, pp. 173-189, 2014.

[52] W. T. Freeman, “Median Filter for Reconstructing Missing Colour Samples”, U.S Patent, 2724395, 1988

[53] G. H. Chapman, R. Thomas, K. J. Meneses, I. Koren and Z. Koren, “Analysis of Single Event Upsets Based on Digital Cameras with Very Small Pixels”, in *IEEE Int. Symposium on Defect and Fault Tolerance*, 2018.

[54] G. H. Chapman, R. Thomas, K. J. Meneses, I. Koren and Z. Koren , “Using digital imagers to characterize the dependence of energy and area distributions of SEUs on elevation”, *in Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT) IEEE International Symposium*, 2020.

[55] J. Dudas, “MASc: Characterization and Avoidance of In-Field Defects in Solid-State Image Sensors”*,* SFU, Burnaby, CA, 2008.

[56] J. Leung, J. Dudas, G. H. Chapman, I. Koren, Z. Koren, “Quantitative Analysis of In-Field Defects in Image Sensor Arrays”, *Proc. 2007 Intern. Sym on Defect and Fault Tolerance in VLSI, 526-534*, Rome, Italy, Sept 2007.

[57] G. H. Chapman, R. Thomas, K. J. C. S. Meneses, I. Koren and Z. Koren, “Image degradation from hot pixel defects with pixel size shrinkage”, *Electronic Imaging,* vol. 2019, no. 9, Burlingame, USA, January 2019.

[58] G. Chapman, R. Thomas, I. Koren, and Z. Koren, “Improved image accuracy in Hot Pixel degraded digital cameras”, *Proc. of the 2013 IEEE Intern. Symposium on Defect and Fault Tolerance in VLSI Systems*, 172-177, New York, NY, Oct. 2013.

[59] VYSOR. “The Vysor app”, [Online] Available: https://www.vysor.io/ [Acessed December 2020]

[60] G. H. Chapman, R. Thomas, R. Thomas, K. J. Meneses, T. Q. Yang, I. Koren and Z. Koren , “Increases in Hot Pixel Development Rates for Small Digital Pixel Sizes”, *IS&T International Symposium on Electronic Imaging –Image Sensors and Imaging Systems (EI)*, pp. 1-6(6). San Francisco, California, Feb. 2016.

[61] G. H. Chapman, R. Thomas, K. J. Meneses, P. Purbakht, I. Koren and Z. Koren, “Exploring Hot Pixel Characteristics for 7 to 1.3 micron Pixels”, *IS&T International Symposium on Electronic Imaging –Image Sensors and Imaging Systems (EI), pp 401-1-401-(6)*. Burlingame, California, Jan. 2018.

[62] R. Thomas, “MASc: Detection and Analysis of Single Event Upsets in Noisy Digital Imagers with Small to Medium pixels”. Burnaby, CA, 2020.

[63] G. Zapryanov and I. Nikolova, “Comparative Study of Demosaicing Algorithms for Bayer and Pseudo-Random Bayer Color Filter Arrays”, *International Scientific Conference Computer Science,* 2008.

[64] IBM, “Monte Carlo Simulation”. [Online] Available: https://www.ibm.com/cloud/learn/monte-carlo-simulation. [Acessed February 2021].

[65] K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, C. Xiao, A. Prakash, T. Kohno, and D. Song. “Robust Physical-World Attacks on Deep Learning Visual Classification”*, Conference on Computer Vision and Pattern Recognizion*, Salt Lake City, Utah, 2018. Available: https://arxiv.org/abs/1707.08945

[66] Nature. “Why deep-learning AIs are so easy to *fool”.* [Online] Available: https://www.nature.com/articles/d41586-019-03013-5. [Acessed February 2021]

[67] G. H. Chapman, “Freeman Median Demosaicing”, SFU ENSC 895: Optical Engineering and Laser Applications, September 2018. [Online] Available: http://www.sfu.ca/~gchapman/e895/e895l12.pdf

[68] Chang, L., Tan, Y-P. “Hybrid color filter array demosaicking for effective artifact suppression” *J. Elec. Imaging* 15(1), 013003–013003, 2006.

[69] R. Zhen and R. L. Steven, “Image Demoisaicing in Colour Image and Video Enhancement”, *Springer*., 2015

[70] ISO/IEC, “Information Technology: Digital Compression and Coding of Continuous-tone still images: Requirements and Guidelines”, 10918-1:1994.

[71] G. H. Chapman, “JPEG Image Compression”, SFU ENSC 895: Optical Engineering and Laser Applications. September 2018. [Online] Available: http://www.sfu.ca/~gchapman/e895/e895l17.pdf

[72] Adobe. “Blending Modes”*,* [Online] Available: https://helpx.adobe.com/ca/photoshop/using/blending-modes.html. [Acessed March 2021]

[73] J.E. Hill, “Birthday Paradox Calculations and Approximation”. [Online]. Available: https://www.untruth.org/~josh/beta/birthdayparadoxapproximation.pdf. [Acessed April 2021]

[74] Draper, N, “Applied Regression Analysis”, 3rd Edition. New York, NY. *Wiley*. 1998.

[75] G.H. Chapman, R. Thomas, I. Koren, and Z. Koren, "Empirical formula for rates of hot pixel defects based on pixel size, sensor area and ISO", *Proc. Electronic Imaging, Sensors, Cameras, and Systems for Industrial/Scientific Applications XIII*, v8659, 86590C-1-C-11 San Francisco, Jan. 2013.

[76] Oakdale Engineering, "DataFit". [Online] Available: https://www.petromehras.com/petroleum-software-directory/miscellaneous-software/datafit. [Acessed March 2021]

[77] G. H. Chapman, J. Leung, R. Thomas, A. Namburete, Z. Koren, I. Koren, "Projecting the rate of in-field pixel defects based on pixel size, sensor area, and ISO", *in Proc. Electronic Imaging, Digital Photography II*, V8298, 82980E, San Jose, Feb 2012.

[78] G. H. Chapman, "True Camera response", SFU ENSC 895: Optical Engineering and Laser Applications, September 2018. [Online] Available: http://www.sfu.ca/~gchapman/e895/e895l8.pdf

[79] R. Fisher, S. Perkins, A. Walker, E. Wolfart, "Median Filter", [Online] Available: https://homepages.inf.ed.ac.uk/rbf/HIPR2/median.htm#:~:text=Like%20the%20mean%20filter%2C%20the,the%20median%20of%20those%20values. [Acessed June 2021]

[80] P. Longère, X. Zhang, P. B. Delahunt, D. H. Brainard, "Perceptual Assessment of Demoisaicing Algorithm Performance"*, in Proceeding of the IEEE*, vol. 90, NO 1. January 2002.

[81] G. H. Chapman, *Micrlens and APS sensor*, SFU ENSC 895: Optical Engineering and Laser Applications, September 2018. [Online] Available: http://www.sfu.ca/~gchapman/e895/e895l8.pdf. [Acessed July 2021].

[82] National Geographic, *Citizen Science*. [Online] Available: https://www.nationalgeographic.org/encyclopedia/citizen-science/ [Acessed July 2021].