High Dynamic Range Imaging

Sensors and Architectures

SECOND EDITION

High Dynamic Range Imaging

Sensors and Architectures

SECOND EDITION

Arnaud Darmont

SPIE PRESS
Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Names: Darmont, Arnaud, author.
Title: High dynamic range imaging : sensors and architectures / Arnaud Darmont.
Description: Second edition. | Bellingham, Washington : SPIE Press, [2019] | Includes bibliographical references and index.
Identifiers: LCCN 2018038150 (print) | LCCN 2018056543 (ebook) | ISBN 9781510622791 (pdf) | ISBN 9781510622807 (epub) | ISBN 9781510622814 (mobi) | ISBN 9781510622784 (softcover)
Subjects: LCSH: Image converters. | High dynamic range imaging–Equipment and supplies.
Classification: LCC TK8316 (ebook) | LCC TK8316 .D37 2019 (print) | DDC 621.36/7–dc23
LC record available at https://lccn.loc.gov/2018038150

Published by

SPIE
P.O. Box 10
Bellingham, Washington 98227-0010 USA
Phone: +1 360.676.3290
Fax: +1 360.647.1445
Email: books@spie.org
Web: http://spie.org

Printed in the United States of America.
First Printing.
For updates to this book, visit http://spie.org and type “PM298” in the search field.

Contents

Acknowledgments

This book uses several images, algorithms, and document reproductions from various authors. I thank all those authors for granting us the right to reuse their material. More details about the ownership of the images can be found in the figure captions.

Special thanks to Martin Wäny, Bart Dierickx, Dirk Hertel, John McCann, Kenton Veeder, Uwe Artmann, and Benoit Dupont for their help, contributions, or manuscript reviews of this edition or the first edition. Also, a special thank you to all of my employees, former employees, students, family members, and friends who helped write this text and gather all of the required documentation.

Thanks are extended to all attendees of my HDR course at SPIE conferences (especially Photonics West) and at other venues for their comments, questions, and remarks.

Arnaud Darmont

Arnaud Darmont (1979–2018) was the founder and CEO of Aphesa, a company that provided image-sensor consultations. He served as an SPIE course instructor and session chair for IS&T/SPIE Electronic Imaging. In December 2017, he joined the European Machine Vision Association as a Standards Manager and introduced two new standards in less than a year. He completed work on this edition before his passing. His enthusiasm and drive as a member of SPIE and the machine-vision community lights the way for others to follow.

Preface

HDR photograph made of exposure bracketing with three shots and post-processed in Photomatrix Pro and Lightroom (courtesy Christian Michaux).

This book collects the knowledge about image sensors, dynamic range, high dynamic range (HDR) image sensors, and HDR applications gained from over 16 years in the image sensor and imaging business as an engineer, project manager, researcher, instructor, business development manager, and consultant.

With my first employer, Melexis, I worked on one of the first HDR global shutter CMOS image sensors and its related reliability and production testing. At that time, very few publications were available on in-pixel HDR imaging and its issues and difficulties, even though several companies (such as FillFactory, IMEC, PhotonFocus, Awaiba, Melexis, and Micron) already had device prototypes and were building the knowledge internally. No company had experience with the full production testing of such devices, and everything had to be made to meet automotive standards.

At Aphesa, several of our custom camera projects or consulting projects were related to or involved HDR. We had several pipe inspection projects, including oil and gas applications, or medical endoscopy applications in which lighting was very difficult to control and the irradiance of the scene was very uneven. The scenes also had some highly reflective parts and lowly reflective parts simultaneously. HDR techniques were implemented in the sensor, in the

camera, or in the host software. Later, we saw more multispectral or hyperspectral projects that had very large differences in the signal levels between bands, causing imaging difficulties similar to what we had encountered in HDR, and therefore techniques derived from HDR were used.

When the first edition of this book was published in October 2012, it was the first comprehensive text about HDR techniques used in pixels and in cameras or software with an engineering level of technical details. Since then, HDR photography has become more popular.

In 2015, I started teaching image sensors and imaging more widely than the SPIE HDR course that inspired this book, and a lot of questions were raised by attendees. It became necessary to update the book with the answers to some of these questions.

In 2017, I joined the European Machine Vision Association as part-time manager of standards, and one of the first actions taken was to extend the EMVA1288 standard to be compatible with HDR image sensors and cameras. There is also an initiative to extend the Gen<i>cam standard to offer HDR pixel formats and compatible data containers and controls.

Since the first edition, things have evolved quickly in the field of HDR because of the use of HDR in consumer markets such as DSLR and mobile, and also because of the development of autonomous vehicles, drones, and the latest generation of camera-based driver-assistance systems.

Therefore, the philosophy of the book has evolved. The explosion of HDR applications has led to a significant increase in the number of algorithms and publications on the topic, so this volume serves as a starting point for exploring HDR imaging by introducing the core concepts with schematics and equations and going deep into the general principles.

This book provides readers with an intermediate-to-advanced knowledge of HDR image sensors and techniques for industrial and non-industrial applications. HDR is increasingly being used in automotive on-board systems, autonomous vehicles, road traffic monitoring, and in industrial, security, medical, and military applications, as well as in photography. It offers advantages such as greater robustness against direct sunlight or reflection of strong lights on metals, and better detection of objects located in shadows. The book is not about the artistic side of HDR images, although some aspects of HDR photography are mentioned, and several photographs are included for illustration. Some aspects of system testing are also introduced. Instead, it describes various sensor and pixel architectures to achieve HDR imaging, as well as software approaches to create HDR images out of lower dynamic range sensors or image sets. Some methods for automatic control of exposure and dynamic range of image sensors are introduced. The most important optical effects are also considered.

Arnaud Darmont
September 2018

Chapter 1
Introduction

1.1 Applications Requiring a Higher Dynamic Range

In industrial processes, illumination is generally a well-controlled element of a system, so that the luminance of a scene matches well with the operation range of a camera. Industrial system developers and integrators have access to a variety of light source types, diffusers, filters, reflectors, and mirrors to design the required light pattern and uniformity. The light intensity can also be easily controlled, and light flashes or pulses can be synchronized with the camera. As the speed of the image sensors and camera increases with each new generation, it is now convenient to take multiple pictures of the same scene with light from different wavelengths or with different directions or intensities. There are, however, some applications where luminance of a scene is not completely (or not at all) under control of the user. These applications are numerous, but most of the time they do not require high dynamic range (HDR) solutions. Some applications, however, are impossible without an HDR imaging system.

HDR applications can be organized into several categories, depending on the following criteria:

- Is there relative movement between the scene and the camera?
- Is the application targeting display or machine vision?
- How much is the required dynamic range?
- Is it a real-time application or not?
- Is dynamic range extension required on the dark side, the bright side, or both?

Welding is a typical case where the main object of the scene is extremely bright (see Fig. 1.1). Exposing the sensor to this bright spot makes the image extremely dark, giving no details except, maybe, within the light arc itself. Exposing the sensor to the largest area of the scene can provide usable information, but the light arc will be totally saturated with no information. In this scene, the area of interest is the welding point, which is located very close to the brightest area, but the point itself is darker. This application is one of

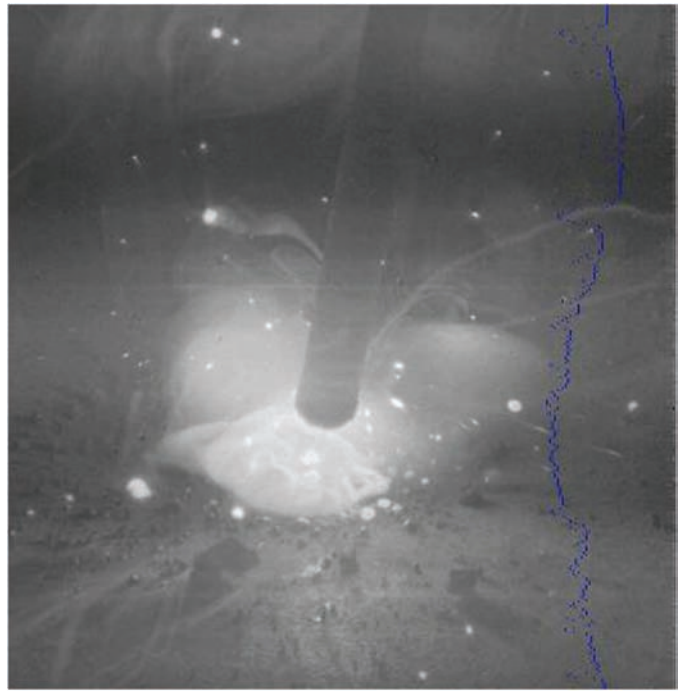

Figure 1.1 140-dB image of a welding application using a logarithmic sensor and off-chip fixed-pattern noise correction. The blue dots are an overlay of histogram data into the image (courtesy of Dierickx [83]).

the most difficult situations. The testing of explosives and electrical fuses results in similar scenes and imaging difficulties. The word "information" is used here because it indicates what is of importance in an image. We do not care much about how nice the picture looks but about how much information can be collected, processed, or analyzed. Sometimes we can accept some saturation in a picture if the saturated areas do not contain much useful information but the saturation improves the image quality around and on some objects of interest. As an example, consider an automotive camera pointed ahead of the vehicle; it is interested in the road signs, road marking, and other vehicles or pedestrians, but not the clouds in the sky.

A similar situation is encountered in automotive on-board applications such as park assist, pedestrian detection, lane departure warning, or traffic sign recognition. In these circumstances, there is no control of the general lighting that can come from various sources such as a low sun at sunset, headlights of oncoming vehicles, or sun reflection on a damp or wet road (see Fig. 1.2). These situations can become more dangerous when the road signs to be imaged are very reflective and illuminated directly by the lights of one's own vehicle, as is often the case in many countries, for example, in France.

Figure 1.2 Example of an HDR automotive front-vision application. The image was acquired in 2004 with an early-model panoramic VGA global shutter automotive image sensor.

Intelligent airbag systems are now measuring the size of the person sitting in the passenger seat to differentiate a child from an adult or a large bag, so that proper airbag deployment is achieved. Such systems make use of synchronized pulsed infrared light. This small amount of infrared light is the signal to be detected; large amounts of sunlight have greater intensity by

Figure 1.3 Image of a lightbulb using a logarithmic sensor with off-chip fixed-pattern noise correction (courtesy of Dierickx [83]).

orders of magnitude. Separating the useful part of the light from the sunlight and other light sources requires HDR techniques.

In road traffic monitoring, the situation is similar to on-board automotive applications, as scene illumination is not under control, and the oncoming vehicles can be oriented with their lights facing the camera. This situation is usually solved in traffic monitoring by placing the camera higher, so that the lights do not shine directly into it. However, when working with tunnel entrances and exits, bright light from the sun outside the tunnel contrasts with the dark, artificial illumination inside the tunnel. If the camera's field of view includes both of these environments, then parts of the image will be useless—either saturated or too dark.

For law enforcement, cameras are usually placed at the front of a vehicle to record chase scenes. Because these images can be used in court as evidence, it is important to be able to see details such as a license plate or a driver's face, even if the camera is positioned toward vehicle headlights or toward the sun. Figure 1.4 shows typical luminance distributions.

In security applications, a situation similar to the tunnel scenario is encountered. A camera located inside a bank, for example, cannot simultaneously view the inside of the room and the outside of the building through the windows on sunny days. A security camera located at an intersection of a well-illuminated street and a shadowed street will only be able to view one of the two streets in a single image (see Fig. 1.5).

These situations can usually be solved by taking two images: one with the exposure adjusted for the bright part of the scene, and another with the exposure adjusted for the dark part of the scene. This approach makes sense in many situations but is not acceptable in others. If information is to be streamed to a control room, it will need to be split on two displays or somehow

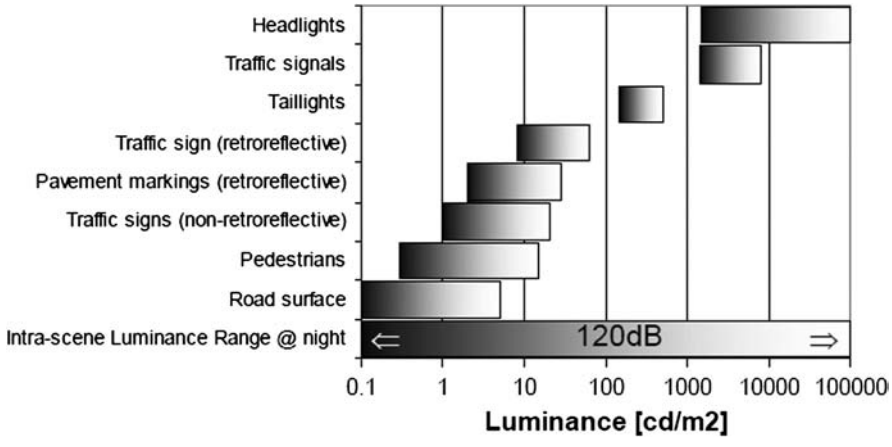

Figure 1.4 Typical luminance distribution in an automotive or law enforcement scene (reproduced from Ref. [67]).

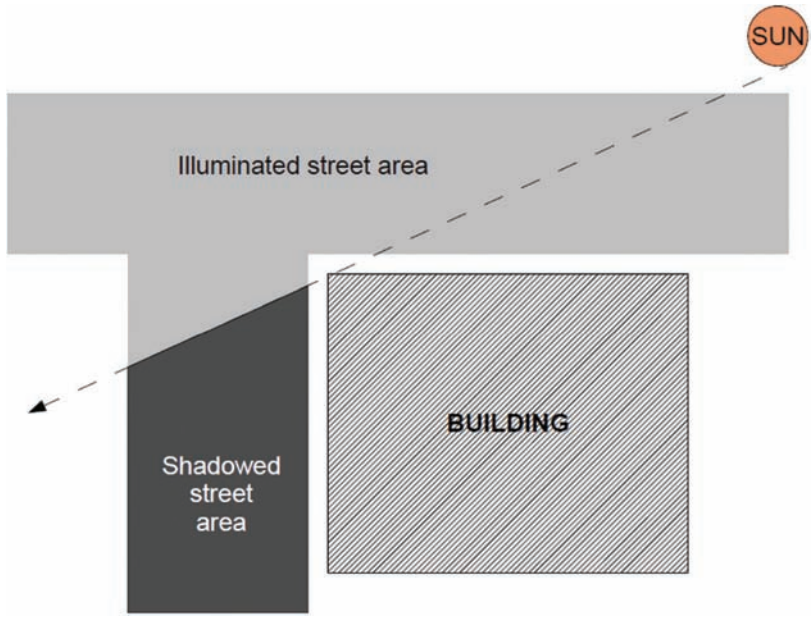

Figure 1.5 Shadowed-street situation.

combined. Splitting the data has a cost; combining it is an HDR technique described later. In other situations such as pedestrian detection, road sign detection, or lane departure warning, basing an algorithm on two images can be too slow for the application. The goals of these systems are meant to offer the highest possible degree of safety, yet at an affordable price. It may be too expensive to increase processing power and work with twice the frame rate. Reducing the overall frame rate affects the decision time, creating conflicts with the highest degree of safety requirement. Moreover, two different exposures can be insufficient for achieving acceptable results. In some situations, eight images or more are required to image the complete scene with enough detail.

Certain medical applications for diagnostic imaging require HDR imaging to see high-contrast images, which enable the detection of small variations in tissue coloration or local contrast. Another medical application that requires HDR is endoscopy because of the difficulties to provide uniform lighting.

Lighting issues are also present in projects related to oil- and gas-well inspections. Because of the nature of the environment and the narrow pipes, as well as the very limited embedded processing capability and communication bandwidth, multiple pictures of the exact scene cannot be made easily from the same vantage point, and so an automotive sensor with an HDR mode can be used.

Many night-vision goggles and other night-vision systems use amplification of light by photomultiplication. Similar results can be achieved by electronic processing. This principle works well in the dark, but saturation

occurs as soon as a light source is present. It is possible to use a photon multiplier together with an HDR sensor to build a better electronic night-vision system.

HDR imaging is also important in certain defense applications such as target acquisition, dynamic target chasing, and missile interception. Future defense systems could involve lasers to blind an opponent's optical targeting or aiming systems. A very high dynamic range camera must be robust against such defensive equipment.

Barcode scanning sometimes suffers from the reflection of light caused by plastic covering the barcode. Although this can be solved by image processing for conventional barcodes, it becomes an issue of higher importance for quick-response (QR) codes.

The figures in this section show examples of luminance encountered in practical scenes. The ranges of luminance are often wider than the range that the film or sensor can image. A lux is the photopic unit of illuminance and emittance. It corresponds to the radiometric unit of W/m^2. The relationship between the lux and W/m^2 is the luminosity function, which is the standardized model of human visual brightness perception over wavelength. 1 lux = 1 lm/m^2 = 1 $cd \cdot sr/m^2$. Candela per meter squared (cd/m^2) is the luminance photometric unit. A few numeric examples of luminance (from Ref. [76]) encountered in everyday situations are listed as follows:

- moonless sky: 3×10^{-5} cd/m^2
- full moon (0.2 lux): 0.06 cd/m^2
- low limit of human color vision: 1 cd/m^2
- living room (50 lux): 12 cd/m^2
- office lighting (500 lux): 125 cd/m^2
- cloudy sky (6000 lux): 1500 cd/m^2
- limit of human visual tolerance: 6000 cd/m^2
- sun: 2×10^9 cd/m^2.

Slide film usually covers a contrast of 8 f-stops or a contrast of around 250:1. Negative film is able to cover contrasts up to around 2000:1 or 11 f-stops. This means that negative film must be used to produce the best pictures of scenes in sunlight. Normal digital cameras range from below 8 f-stops to above 11 f-stops. Film is usually better than a medium-price-range digital camera.

Dynamic range is also an issue for non-imaging optical applications, such as the detection of reflected signals in depth measurement sensors, e.g., lidar and time of flight. Sun-load sensors must also deal with a wide input signal range.

1.2 High Dynamic Range Photography

HDR photography began in 1850 when Gustave Le Cray had the idea of combining a low-exposure negative and a high-exposure negative to reach a

higher dynamic range. In 1945, HDR photography was required for taking pictures of the first nuclear explosions.

In the second and third quarter of the twentieth century, Ansel Adams (1902–1984) created very-high-contrast monochrome images using the zone system, a process he invented for selecting the proper exposure of glass plates (Adams was using large glass plates for their performance and resolution) and for selectively rendering the brightness of the final prints. Adams mostly photographed the American Southwest, more specifically, Yosemite National Park. His most famous photographs include *Monolith, the Face of Half Dome* (1927), *Moonrise* (1941), *Sierra Nevada* (1948), *El Capitan* (1968), *Yosemite and the Range of Light* (1979), and many others. The procedure used by Adams in his developing process, spatial manipulation, predated the multiple exposures and tone mapping used today in digital photography.

Modern commercial HDR imaging started in 1980 with the invention of the RGBE file format. At that time, HDR images were generated using local calculations. It was not until 1997 that images were generated using global calculation by Mann, Picard, and Debevec [76]. This is the approach used today, known as the radiance map or Debevec method (Fig. 1.6). This method produces very high dynamic range data that cannot be displayed, thus the tone-mapping theory compresses HDR data into a low dynamic range (LDR) dataset, which preserves the useful information (Fig. 1.7).

HDR photography is briefly discussed in this book at times because it yields great artistic pictures. However, the main purpose of this book is to

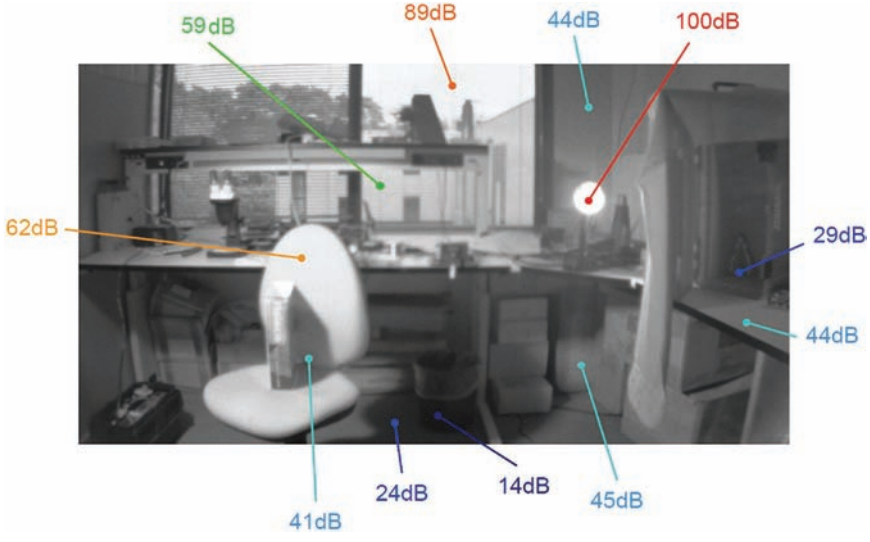

Figure 1.6 Image of an HDR scene and scene luminances relative to the sensor's lowest detectable level (courtesy of Melexis Microelectronic Systems).

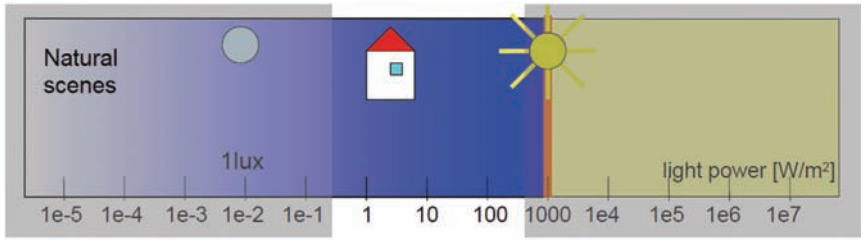

Figure 1.7 Comparison of luminance with the dynamic range of a common low dynamic range (LDR) sensor (courtesy of Dierickx [10]).

learn how to use HDR imaging in industrial machine vision applications or in similar applications that cannot be considered as photographic or artistic. This section shows a few high-quality HDR photographs (see Figs. 1.8–1.10 and 1.12); Fig. 1.10, for example, shows some halo effects that will be explained later in the tone mapping section. All images in the remainder of this book are related to industrial applications, or are used to show a specific effect.

To produce photographic HDR images of still scenes, a camera with a very fast exposure bracketing system (when the trigger is pushed, several images are acquired one after another with different exposure times) or a good tripod is required. See Fig. 1.11 for an example of a bracketing configuration and Fig. 1.12 for a tripod and remote-control shooting setup. A remote control is used to avoid any small changes to the position of the camera when the trigger is pushed or settings are changed. Many cameras today provide a USB or wireless connection to a laptop or smartphone. First, several images must be acquired of the same scene from the same

Figure 1.8 Village overlook at sunset, processed with open-source software and Fattal tone mapping.

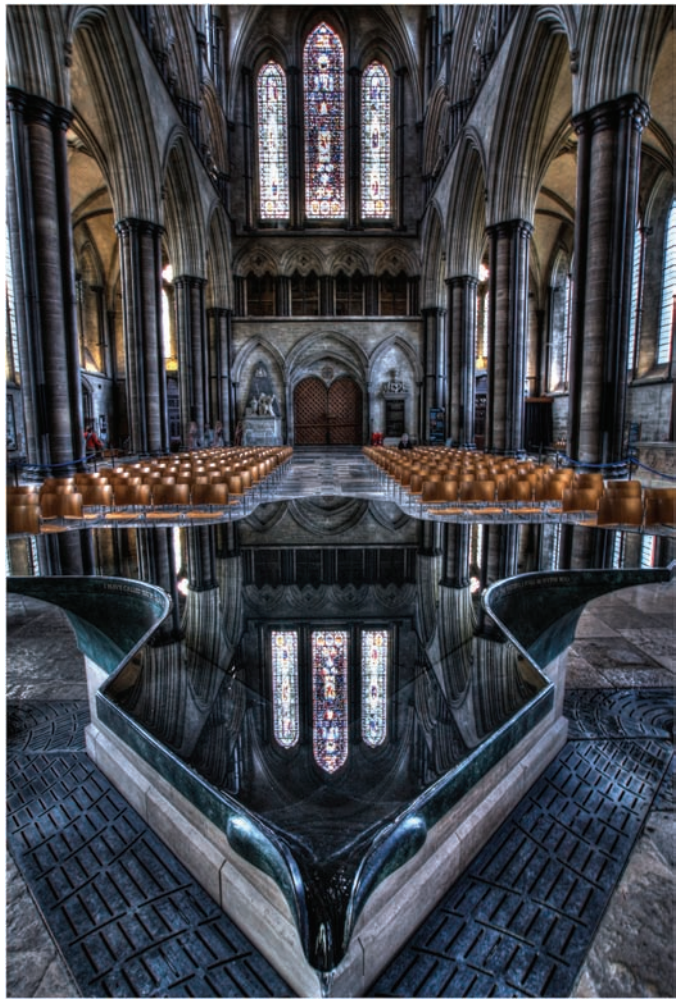

Figure 1.9 Salisbury Cathedral, processed with Photomatrix Pro.

vantage point with several exposures. Then, the images need to be realigned and processed in specific software to extract an estimate of the radiance of the scene. Finally, the images are reduced to an 8-bit format using tone mapping.

1.3 Scientific Applications

Scientific applications often require images with extreme linearity and precisely known camera response to measure absolute and relative values of irradiance and luminance. As will be seen through the examples in this book, achieving this goal with an HDR imaging system is difficult, as the response curve of the sensor can be modified, and the algorithms applied can be

Figure 1.10 HDR photographs by Rachel Santellano (www.pixoto.com/rachels).

Figure 1.11 Exposure menu of the Canon EOS 40D, showing the exposure bracketing settings (AEB).

Figure 1.12 Equipment setup for HDR shooting.

nonlinear or might not preserve the sensor's responsivity. Optical veiling glare is also a significant limitation.

Recently, several CMOS image sensors featuring HDR modes have been specifically developed for scientific applications and tend to replace expensive, high-end CCDs.

1.4 High Dynamic Range, Wide Dynamic Range, and Extended Dynamic Range

Why are there different names? The term "high dynamic range" is generic; wide dynamic range (WDR) and extended dynamic range (XDR) are usually used by manufacturers when describing pixels or sensors. This book uses HDR in general to describe a method or a scene, and XDR to denote the extension of dynamic range compared to a non-HDR (also called low dynamic range) solution. To be complete, there is a last term, high dynamic range rendering (HDRR), used in the opposite process of displaying HDR images. HDRI stands for high dynamic range imaging.

1.5 Reducing the Exposure Time

When an image is overexposed, the exposure time can be reduced until the image looks less saturated. This can be done by changing some sensor registers or camera settings (see Fig. 1.13). In this figure, an image is previewed in the LCD screen of a camera together with its histograms, and the exposure time is adjusted until the histogram no longer saturates. With automatic exposure algorithms, this situation can be solved by using a negative-exposure compensation setting. An exposure compensation of −1 EV would mean a target brightness half that of the default mode.

Figure 1.14 shows the saturation of the image and the corresponding saturation peak of the histogram for the overexposed image. Figure 1.15 shows the clipping of the image in the darkest areas for the underexposed image. This underexposed image has no saturation. Figure 1.16 shows the

Figure 1.13 Adjustment of the exposure of a Canon EOS 40D based on histograms until the image is no longer saturated. The exposure time can be seen in the top-left corner of the screen.

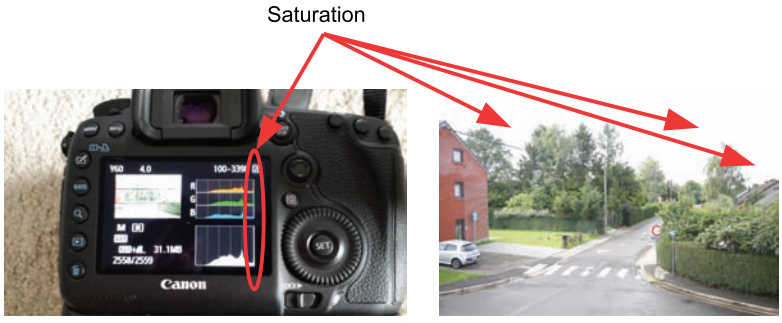

Figure 1.14 Saturation of the overexposed image.

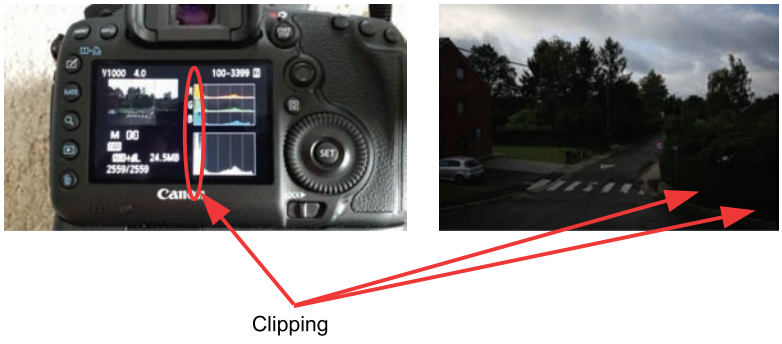

Figure 1.15 Clipping of the underexposed image.

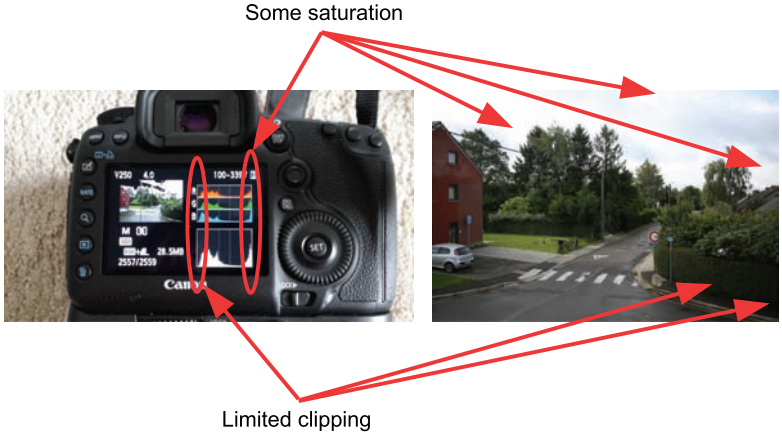

Figure 1.16 Best tradeoff image showing some saturation and limited clipping.

intermediate image that simultaneously has some saturation and limited clipping. It is the best possible tradeoff for this scene and this camera.

The image without histogram saturation clearly has more details in the highlights as the objects now appear, and it may seem that some software

Figure 1.17 Post-processing attempts of the last, non-saturated picture of Fig. 1.13 compared to an HDR version of the same scene taken with the same camera and the same saturation level. Top left: original image. Top right: gamma. Bottom left: strong sharpening. Bottom right: HDR merged and tone-mapped image.

post-processing would recover the dark details that have now disappeared due to the reduced exposure. However, this is not possible due to sensor and camera noise, as well as the analog-to-digital conversion. See Fig. 1.17, where different post-processing solutions are attempted but never produce the image quality of the corresponding HDR image.

1.6 HDR Applications That Do Not Require HDR Images

Sometimes the solution to an HDR problem can be found without using an HDR technique. For example, in lane-departure-warning applications, cameras have to detect road lane markings and estimate the path of the vehicle a few seconds in the future to determine if the vehicle will either stay in lane or cross a lane marking. Unwanted lane changes can happen due to driver drowsiness, and a warning signal (usually vibrations from the steering column) is provided to the driver. Because the road surfaces may reflect the sun and because the sun's location as well as other light sources are unknown, an HDR image sensor is usually used so that all details of the scene remain visible to the image processor, independent of the road condition or lighting situation. Lane markings are typically extracted from the HDR image; then a binary image is generated and processed to filter out everything but the lanes; and then calculations are made to generate a

mathematical model of the lane curvature, the location of the lane center, and the vehicle's position in the lane.

This section will not describe the details of this complete process but rather show how the binary image of edges can be generated without using an HDR image, i.e., without merging exposures and without the use of an HDR pixel. Figure 1.18 illustrates the idea.

The proposed image sensor uses two different exposure times for the odd and even rows. A horizontal edge detection is performed, and the resulting data from the odd and even rows are merged, as if they were located at the same place. If the long-exposure row is saturated, no edges will be detected, but they will be detected in the short-exposure row. Conversely, if an edge is not detected in the short-exposure row because it is too dark and does not offer enough contrast, then the edge is likely to be detected in the high-exposure row.

During readout, the special proposed image sensor will perform the edge detection on the high- and low-exposure rows and will merge the results into a binary image (a single bit per pixel output that indicates whether or not an edge was detected in either of the two lines) by ORing the results of the

Figure 1.18 Road-marking-detection algorithm in an HDR scene without HDR imaging. Top left: even lines, long exposure. Top right: result of a horizontal edge detector on long-exposure lines. Middle left: odd lines, short exposure. Middle right: result of a horizontal edge detector or short-exposure lines. Bottom: combined and filtered binary edge image.

detection on the low-exposure and high-exposure rows. A more advanced approach uses 2D edge detection to reduce the edge noise.

But is this really not an HDR technique? Indeed, an HDR image is not reconstructed, but because either several images at several exposures or a specific sensor with dual exposure is required, it may still be considered HDR.

The same solution could be implemented purely in software out of two images at different exposures, but it would require memory and could suffer from significant motion artifacts due to the relative motion of the road between the two images. The two-image approach also requires a higher-speed sensor, as well as context switching.

There are other cases where an HDR image is not required, although the scene is HDR. For example, if the subject of interest is well exposed and the background saturates, the image will obviously not match the dynamic range of the scene, but this will not affect the result as the subject is well exposed, and no information is required from the background. In some applications, saturating the background can be a solution because it provides more contrast to the object to be inspected.

In many machine vision applications, the dynamic range requirements will be reduced by applying appropriate lighting or filtering techniques, such as spectral filtering, narrow-spectrum illumination, or polarizing filters. For example, barcodes are difficult to read under a plastic film because light strongly reflects on the film and the barcode vanishes into these reflections. The use of a specific light color and a specific type of light source solves the problem (see Fig. 1.19).

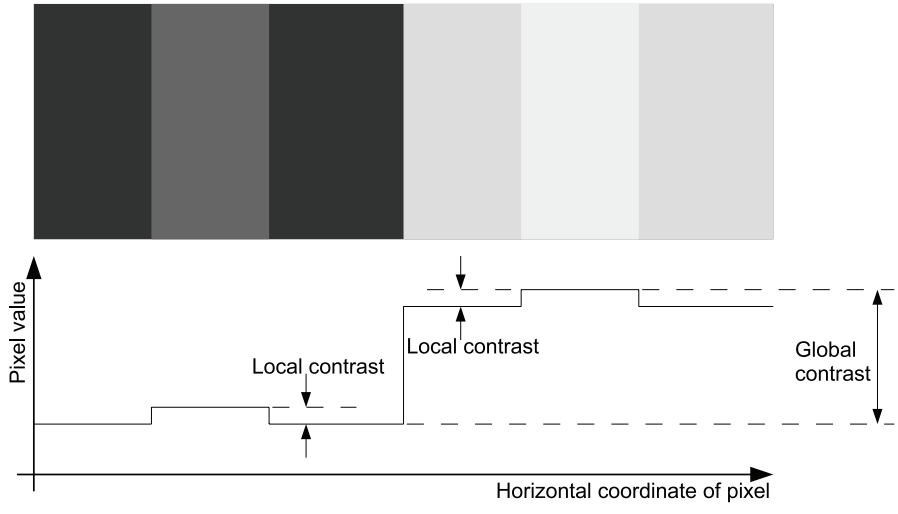

Figure 1.19 The scene's dynamic range of a barcode-reading application is reduced by the application of proper light sources and filters.

1.7 Image Histograms

A histogram is the representation of the distribution of numerical data. It shows the number of occurrences of each possible value of a statistical variable in a set of measured values. For images, the dataset is usually the pixel values of all of the pixels or a region of interest; it can be either the full histogram or only one histogram per color plane. Figure 1.20 shows an example of a histogram for a 5 × 5-pixel monochrome image coded on 3 bits (8 possible values from 0 to 7). Figure 1.21 shows the histogram of a real image. The total area of the histogram equals the number of pixels in the image.

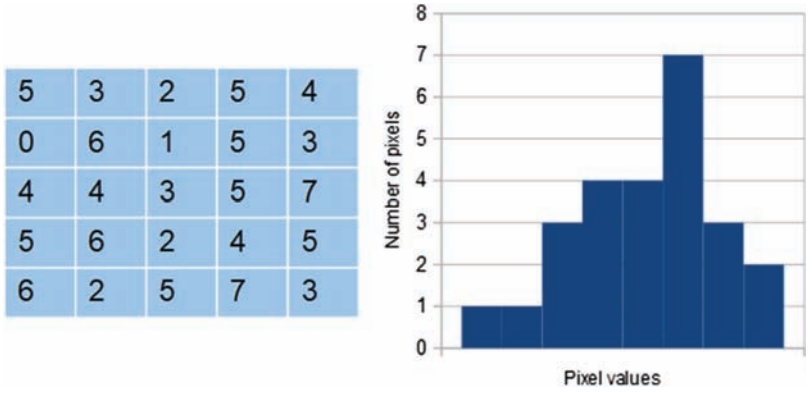

Figure 1.20 Sample 5 × 5-pixel monochrome image coded on 3 bits (left), and the corresponding histogram (right).

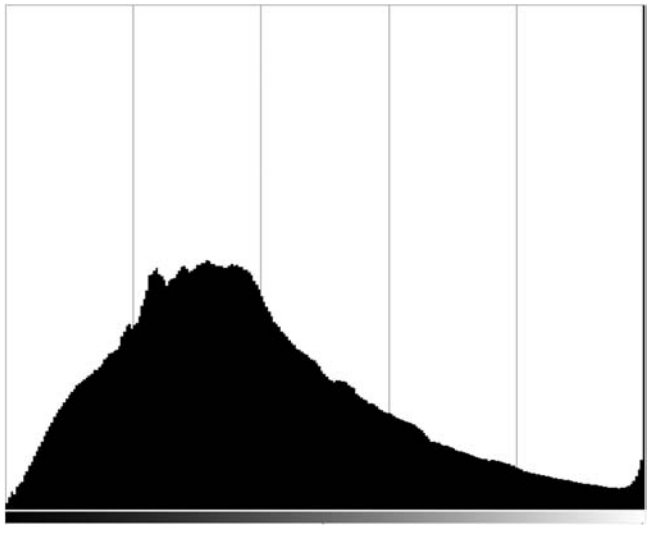

Figure 1.21 Histogram of a real monochrome image (generated with Gimp).

The histogram is very useful to determine if an image is overexposed or underexposed, as can be seen in Figs. 1.13–1.16. The histogram in Fig. 1.21 shows some saturation (the tall peak on the right), but the number of pixels in that peak is only a very small fraction of the total number of pixels, and therefore this saturation is acceptable. Histograms are thus very important in HDR discussions.

1.8 Outline and Goals

After this introduction, Chapter 2 begins with the definition and analysis of dynamic range from a mathematical point of view, and some important notions are introduced. To make this book understandable for most readers, some of the basics are re-explained when necessary.

In Chapter 3, possible system architectures are presented for HDR applications. This chapter primarily discusses the hardware methods used to extend the dynamic range of a camera system. Hardware methods are the methods used inside a pixel or sensor that do not fall into the software category. Any digital implementation of software belongs to the software category, although the software ends up on the chip. Hardware is covered in Chapter 3, and it corresponds to the HDR solutions for many industrial and military applications. The goal of Chapter 3 is to introduce many sensor solutions that exist or have existed commercially (and a few sensors that only existed in laboratories and/or publications) to familiarize the reader with XDR pixel design techniques, XDR performance and issues, and the mathematical techniques used to calculate the performance of such designs.

Chapter 4 introduces various software methods that can be implemented in a camera or on a computer. Due to the strong link between HDR imaging and tone mapping, tone mapping must be discussed, although it is not (strictly speaking) part of HDR extraction for machine vision applications. The world of HDR algorithms and tone mapping is very large, and Chapter 4 only gives a glimpse of it. The goal of this chapter is to familiarize the reader with basic concepts and some common algorithms.

Chapter 5 briefly discusses optical issues encountered in HDR imaging applications. Veiling glare and other optical effects are very important in photography and imaging, and the reader is encouraged to refer to specialized books and publications on this topic.

Chapter 6 reviews algorithms that can automatically control HDR exposure, as more complex sensors have more degrees of freedom and hence require more control logic. Once again, the concepts are only briefly introduced.

Chapter 7 discusses the file formats used to store images with XDR. The user interested in learning more details must refer to published standards.

Chapter 8 discusses the testing of HDR systems and sensors based on the International Organization for Standardization (ISO) and European Machine

Figure 1.22 HDR photgraph by Lofqvist [84].

Vision Association (EMVA) standards; the testing of HDR systems offers unique challenges that are addressed by these standards.

Chapter 9 draws some conclusions, presents a brief discussion of HDR figures of merit, and provides some final questions for students of HDR image sensing.

1.9 Defining a Camera

A camera is typically an imaging device made of an optical element, a sensing element, and some electronic circuitry. For this book, optical elements and electronic circuitry are assumed to be ideal, and the focus remains on the sensor and algorithms. A few very important optical and electronic limitations are introduced. In particular, it is assumed that there is no lens glare, although we will see later that this assumption is almost never correct and greatly affects dynamic range.

Chapter 2
Dynamic Range

Before discussing HDR methods, it is important to first define what dynamic range is. Although the definition of dynamic range seems straightforward, this definition needs to be adapted to further discuss HDR applications.

This chapter will also discuss the signal-to-noise ratio (SNR), a fundamental curve that is most important in the selection of an image sensor or a camera, and contains the dynamic range information (especially when the curve graphs the maximum SNR versus exposure or another HDR expansion technique). Before discussing dynamic range and the SNR, the operation of image sensors, and CMOS image sensors more specifically, will be discussed. Modern CMOS image sensors have reached or exceeded the performance of CCD sensors for most applications and operating conditions. Although some CCDs for specific applications are still the standard, they are less important in HDR applications due to their intrinsic linearity and their difficulty in avoiding high-saturation effects such as smear and blooming. CCD sensors will be briefly addressed.

The following sections only introduce the basic principles of CMOS image sensors and are not intended to replace a full book on image sensors; readers are invited to read specialized literature.

2.1 Image Sensor Theory

Assume for this brief summary of image sensor operation that sensor response is linear and that all noise source stochastic processes are white (or at least wide-sense stationary if not white), and ergodic. This statement can be broken down as follows. A stochastic process is any time development that can be analyzed by probability theory. A white noise is a noise signal that carries the same energy at all frequencies; its frequency spectrum is flat. White noises are always stationary. A stochastic process is stationary if its joint probability distribution is invariant to time and space shifts. It implies that its average and standard deviation are invariant to time and space shifts. A stochastic process is wide-sense stationary (this weaker definition is usually sufficient

for all noise sources that are not white) if at least the first and second moments are invariants to time and space shifts, i.e., $E\{\mu(t)\} = m_x(t) = m_x(t+\tau), \quad \forall\tau \in \mathbb{R}$ and $E\{x(t_1)x(t_2)\} = R_x(t_1, t_2) = R_x(t_1+\tau, t_2+\tau) = R_x(t_1 - t_2, 0), \forall\tau \in \mathbb{R}$, where R_x is the autocorrelation function. The statistical properties of an ergodic stochastic process can be extracted by use of an estimator measured over a single sufficiently long sample. These statements mean, among other things, that the performance of a sensor is independent of the region of interest used on the sensor array to perform the characterization and that the measurement can be repeated in time with the same result under the same conditions. This feature is greatly important for the characterization of sensor performance as only a limited sequence of images and a limited number of pixels are available for statistical calculations.

Although the assumption that sensor response is linear and that all noise source stochastic processes are white is not strictly true and is not always perfectly suited for XDR sensors, it allows for basic equations to be derived that can be used to understand the results of other sensor types. The equations derived with this assumption are in general very close to real results, and therefore the postulate about the noise sources is valid. The next sections consider a digital imaging system step by step, starting at the light source and progressing toward the digital image.

2.1.1 Light source, scene, pixel, and irradiance

A light source emits photons. The photon emission process is a random process characterized by a Poisson distribution. A Poisson process is a stochastic process that follows Poisson's law. Poisson's law (or Poisson distribution) is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed period of time, if these events occur with a known average and occur independently of the time since the last event. If the expected number of occurrences in this interval is ν, then the probability that there are exactly k occurrences, $k \in \mathbb{N}_+$ is equal to

$$P(k;\nu) = \frac{\nu^k e^{-\nu}}{k!}. \tag{2.1}$$

Its average is ν, and its standard deviation is $\sqrt{\nu}$.

Depending on the type of light source, the photons can have the same wavelength (i.e., monochromatic light sources such as LEDs, although LEDs have a narrow spectrum that is usually considered monochromatic for discussion), a discrete set of possible wavelengths (for example, discharge lamps or a combination of multiple monochromatic sources, such as some white LEDs, which are usually made of a stack consisting of a blue LED and a phosphor that converts part of the blue light to green and red light), or a continuous spectrum (for example, the sun or a heated filament behaving as a blackbody). What is of interest is the amount of light that hits one pixel during

the exposure time, and the spectrum of this light. The amount of light is used for dynamic range discussions, and its spectrum is important for color identification and reproduction. As sensor performance is wavelength dependent, the dynamic rage will be wavelength dependent and is easier to measure for a monochromatic source. In real scenes, light is a combination of wavelengths, and the overall dynamic range performance involves the calculation of a weighted integral.

Objects illuminated (the light incident on their surface is called illuminance) by the light source (the illuminant) modify the photon's trajectory, wavelength, polarization, and the number of photons; their surfaces are characterized by their emittance. From the sensor's point of view, each point of a scene is considered to be a light source emitting energy toward one sensor pixel. For a monochromatic light source, the range of these energies is the dynamic range of the scene. For non-monochromatic light sources, it is more complex to define a scene's dynamic range.

The light reflected back from a scene can be of three types, depending on the material of the illuminated objects:

1. direct light (specular reflection): usual reflection and transmission (Snell's law of refraction), acting as a mirror;
2. dispersed direct light (Lambertian reflection or refraction, possibly mixed with some specular reflection): diffuse reflection and transmission; and
3. scattered light: isotropic diffuse reflection and isotropic diffuse transmission.

The corresponding viewing luminances are given by:

- **direct light:** $L = \rho L_i$, where ρ is the reflectance (or transmittance), L_i is the luminance of the irradiating light, and L is proportional to the luminance of the irradiating light.
- **dispersed direct light:** $L = F_s \rho L_i$, where F_s is a factor related to the solid angle, and L is proportional to the luminance of the irradiating light; and
- **scattered light:** $L = (\sigma E/\pi)$, where σ is the scattering rate, E is the illuminance at the point of the object's surface acting as a virtual source and is proportional to the illuminance of the irradiated object.

A scene is the part of the world located in the field of view of the sensor (Fig. 2.1). The field of view is defined by the lens and dimensions of the sensor, and represents a solid cone from the sensor that intercepts a portion of space. If the field of view is wider, then for the same sensor resolution and area there will be less spatial detail in the image because one pixel now represents a larger scene area. If the sensor area is smaller and the lens is unchanged, then only a fraction of the projected lens image will be detected, and the field of view will be narrower.

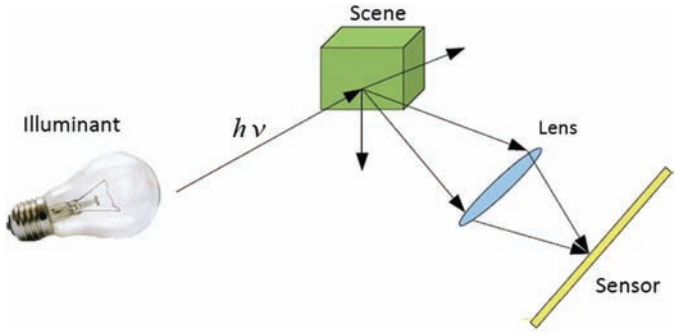

Figure 2.1 Illuminant, scene, optical path, and irradiance.

Let μ_p denote the mean amount of photons that hit a pixel during the exposure time. The number of photons over a short period of time is not constant, as the light source is a Poisson process (Fig. 2.2). The standard deviation of this random variable is $\sqrt{\mu_p}$ and is called photon shot noise. It is also sometimes referred to as "the noise of (the) light."

The value μ_p can be calculated out of the sensor's irradiance, the light wavelength, and the exposure time by

$$\mu_p = \frac{AEt_{\text{exp}}}{h\nu} = \frac{AEt_{\text{exp}}\lambda}{hc}, \tag{2.2}$$

where A is the pixel area, E is the irradiance on the sensor surface in W/m^2 (assuming that the irradiance on each pixel is the same), λ is the wavelength of the light in meters, $c = 299792458$ m/s, which is the speed of light in space

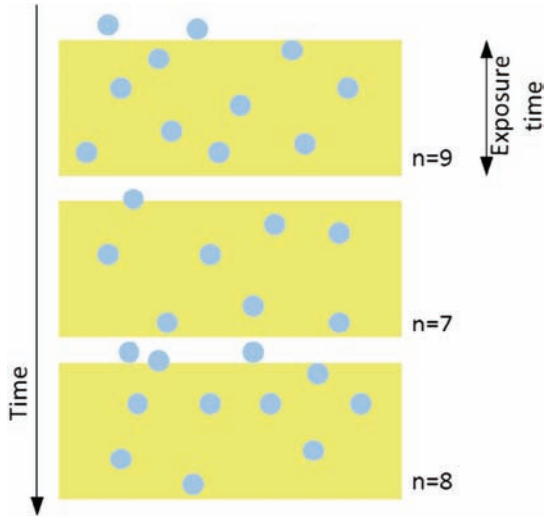

Figure 2.2 A light source is a random Poisson process.

(in meters per second), t_{exp} is the exposure time in seconds, and $h = 6.626068.10^{-34}$ m^2kg/s, which is Planck's constant.

It might be astonishing to hear that light is noisy. As humans see the world, light from a stable light source seems to be constant. This is because we live in a macroscopic world where a very large number of particles are taken into account. Recalling Poisson's law, the signal is proportional to Poisson's parameter, and its uncertainty (or noise) is proportional to its square root. By considering a very large number of particles, as for our macroscopic study of a stable light source, a noise-to-signal ratio of $\lim_{\nu \to +\infty} (\sqrt{\nu}/\nu) = 0$ is attained, which means that the light source seems stable to us. For any smaller number of particles, the noise-to-signal ratio increases until it reaches unity: for a single particle there is as much noise as there is signal. It does not mean that the number of photons is 0, 1, or 2, as noise is understood as a statistical variance; the number of photons for an average signal of one photon can be 100 photons but with a low probability.

Figure 2.2 graphically depicts the photon shot noise process. Each rectangular window represents space (one pixel area, horizontally) and time (exposure time, vertically) for one image; the blue dots are the photons. (It is a simple representation of photons as we know that photons do not behave as "dots".) If viewed from far away, all rectangular areas seem to contain the same number of photons, but a careful counting shows that each area has a slightly different number.

Consider the following experiment. Imagine a very-low-intensity ideal light source that emits exactly one photon every second. If this source were imaged with an ideal image sensor at 10 frames per second, then nine frames would not see any signal, and one frame would get the photon. The source will seem to be noisy.

References to the noise of the light here do not mean the varying intensity over time that some lights exhibit. Due to their electrical properties, some light sources flash at 100 or 120 Hz, or other frequencies. Other light sources randomly vary due to some chaotic mechanisms in gas and chemical structure, e.g., a candle. Other light sources vanish over time due to the depletion of their fuel. The noise, or light variation, mentioned previously is a real random process that happens in any light source — even the most stable — and is related to the quantum-mechanical nature of the light-emission process.

Light can also be described as an electromagnetic wave and calculated using Maxwell's equations, which are the differential forms of Gauss's laws, Faraday's law, and Maxwell's corrected version of Ampere's circuital law:

$\nabla \cdot \vec{D} = \rho$ (Gauss law for electric fields in differential form)

$\nabla \cdot \vec{B} = 0$ (Gauss law for magnetic fields in differential form)

$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}$ (Faraday's law in differential form)

$\nabla \times \vec{H} = \vec{j} + \frac{\partial \vec{D}}{\partial t}$ (Amperes's law in differential form).

In geometrical optics, light is assumed to be a wave and is calculated geometrically without using Maxwell's equations but using rays of light that are perpendicular to the wavefront. The choice of describing light as a ray or as a group of particles depends on the calculations to be performed. In image sensor studies, most of the calculations consider light to be particles (photons). Wave/particle duality is inherent to the fact that light is made of quanta of energies described by quantum mechanics and quantum electrodynamics (QED).

In the case of very small pixels that have a size on the order of the wavelength of light, light behaves more and more like a wave and exhibits diffraction patterns and other quantum-mechanical properties. Therefore, modern pixel simulations imply the simulation of the real behavior of light within the structure of the imaging device.

2.1.2 Sensing node and light–matter interaction

The role of a sensing node is to convert incoming photons into an electrical current so that the signal can be processed by an electronic circuit. The conversion of photons into electrons and holes depends on the sensor's material and structure, and the photon's wavelength. In a semiconductor, some electrons are naturally excited to a conduction band. The conduction band is partially filled by electrons, and the valence band is partially free. The free energy levels in the conduction bands can be used for additional conductivity by free electrons, and the free energy levels in the valence band represent missing electrons in the crystal structure and are therefore called holes. A photon (a quantum of pure energy, the gauge boson of the electromagnetic force) can interact with a semiconductor crystal and transfer its energy. Among the various processes that are possible, one possibility is the excitation of an electron from the valence band to the conduction band by transferring all or part of the photon's energy to the electron. (The energy of the photon is first fully transfered to the electron, but the electron may then lose part of this energy to the crystal.) The electron becomes a free electron and leaves a position vacancy in the crystal lattice. This process is called electron-hole pair generation; it is a quantum-mechanical interaction between three particles, where the hole is a virtual particle. It can be represented by $h\nu \rightarrow e^- + h^+ + \text{heat}$. The hole represents a missing electron and hence holds a virtual positive charge. Note that $h\nu$ is a common way to name a photon as it is the formula for its energy, where h is the Planck constant, and ν is the photon's frequency.

In general, the conversion process can be characterized by a function of the wavelength, called the quantum efficiency (Fig. 2.5) and defined by

$$\eta(\lambda) = \frac{\mu_e(\lambda)}{\mu_p(\lambda)}, \tag{2.3}$$

where μ_e is the average number of electrons being generated during the exposure time for μ_p incoming photons. To be correct, it should be added that this explanation is valid if there are no recombinations. In the case of recombinations, the quantum efficiency represents the net conversion only, taking positively into account the generated electrons and holes, and taking negatively into account the recombined electrons and holes. This quantity is also a Poisson-distributed random variable, and its standard deviation is $\sqrt{\mu_e}$. At high-energy wavelengths, deep in the UV band, multiple electron–hole pairs can be generated for one incoming photon. In general, the image sensor theory described here is valid for wavelengths in the near-UV, visible light, or longer-wavelength spectra.

A detailed study of the behavior of a photodiode (or of a more complex sensing structure such as a photogate) is outside the scope of this book. There are many different types of photodiodes and photogates. The photon-to-electron conversion process of the PN photodiode is briefly given here. A similar process occurs in most structures. Figure 2.3 shows the structure of a PN photodiode constructed as an N-well in a P-substrate.

Photons that fall on silicon collide with the crystal lattice and give energy to its atoms. They then free electrons and leave holes behind. This process is random, and its probability depends on the wavelength of the incoming photon. Long-wavelength photons have a lower probability of becoming involved in this process, and a thick silicon is required to collect the photon. If the wavelength is too long (longer than 1125 nm for silicon), the process becomes impossible because the photon does not have enough energy to free an electron (bandgap energy), even if it collides with an atom. The result is that silicon becomes transparent to long-wavelength light.

As the absorption of a photon is a probability that is constant in time and space (it does not depend on the fact that the photon has previously not been absorbed), the net number of photons being absorbed in a slice of thickness *dz*

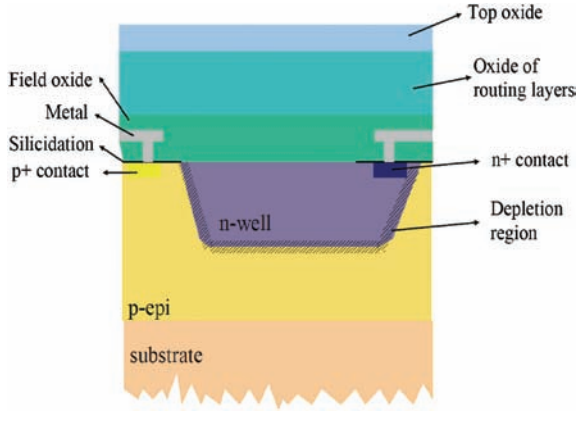

Figure 2.3 N-well in P-substrate photodiode.

is a fraction of the photons still present $\chi(z)\mu_p(z)$. Therefore, the absorption of photons is modeled by the differential equation

$$\frac{d\mu_p(z)}{dz} = -\chi(z)\mu_p(z).$$

It solves as

$$\frac{d\mu_p(z)}{\mu_p(z)} = -\chi(z)dz \Leftrightarrow \int \frac{d\mu_p(z)}{\mu_p(z)} = -\int \chi(z)dz$$
$$\Rightarrow \ln \mu_p(z) = \ln \mu_p(0) - \int \chi(z)dz.$$

In the case of a uniform material for which $\chi(z)$ is constant, such a solution has a solution of the form $\ln(\mu_p(z)) = \ln(\mu_p(0)) - \chi z$, or an exponential solution of the form $\mu_p(z) = \mu_p(0)e^{-\chi z}$. It is linear in a logarithmic plot.

Some electrons are recombined (i.e., rejoin the crystal lattice) and are not able to contribute to the current flow. This occurs in regions 1 and 3 in Fig. 2.4 and explains the wavelength dependency of the quantum efficiency. The photodiode process can be optimized to operate at a specific wavelength where the performance, and therefore the dynamic range, will be maximum. Such optimizations involve changes in doping concentrations and depth of the junction. These changes can also influence the full well capacity of the pixel, the response time, and lag.

If a panchromatic light is illuminating the sensor, then the number of charge carriers generated by each wavelength of the continuous spectrum needs to be summed to calculate the total number of carriers generated (Fig. 2.5). This

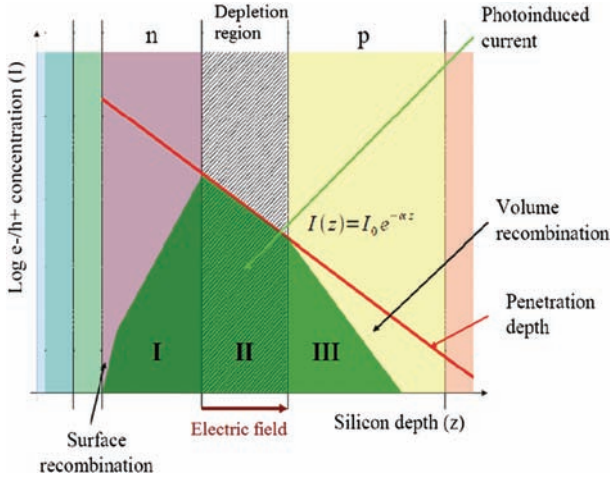

Figure 2.4 Net generation of electron–hole pairs in a PN photodiode.

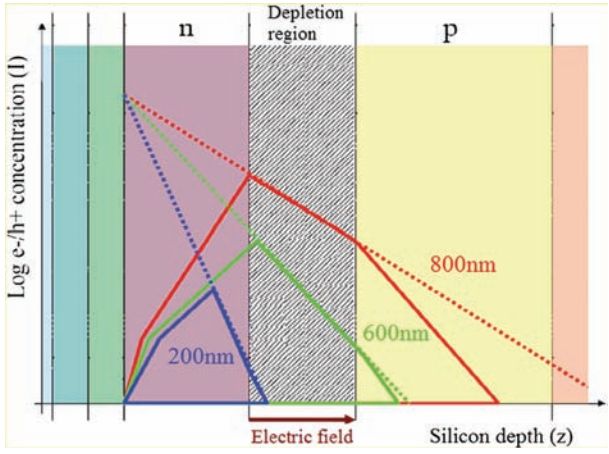

Figure 2.5 Effect of wavelength on quantum efficiency.

calculation is often made with a formula such as $\mu_e = G_m \int_0^\infty P_\lambda S(\lambda)\, d\lambda$, where G_m is the maximum sensor spectral responsivity in electrons per photon, P_λ is the irradiation in spectral radiant quantities (radiant unit per nanometer), and $S(\lambda)$ is the relative spectral sensor responsivity.

A common structure for a real photodiode is the pinned structure, where an additional highly doped positive implantation (p+) is added near the surface of the silicon to reduce the effect of the surface states, and to allow for total extraction of the stored charges, in turn reducing image lag. The pinned layer will also reduce the sensor's performance on blue light. A pinned photodiode should not be confused with a PIN photodiode, which introduces an intermediate intrinsic (undoped) region between the P and the N regions and is usually used for high-speed optical communications. The pinned photodiode is usually realized by extending the diffusion of a transfer gate transistor over the photodiode, which will improve the dark current as it allows for CCD-like charge transfer operation, i.e., without a metal connection.

2.1.3 Pixel

A pixel is made of a sensing element and its associated readout circuit, as well as some control signals and routing. Common active complementary metal-oxide semiconductor (CMOS) pixels usually contain a reset transistor, an amplifier, and a row-select transistor to multiplex the readout of the sensor lines over a column bus. Pixel operation is discussed in greater detail in Chapter 3.

Although each sensor differs in pixel design and technology, all sensors can be characterized similarly. The most important characteristics of a pixel are its geometrical fill factor, full well size, shuttering capabilities, response control, image lag, leakage, conversion gain, and noise. This book focuses on full well size, leakage, conversion gain, noise, and response control because

these factors affect the dynamic range of the sensor. Details about how a pixel, sensor, or camera can be characterized or modeled are found in the EMVA1288 standard (see Ref. 62). This standard has been widely adopted worldwide and is the fundamental standard for machine vision cameras. However, the standard still needs several adaptations for use with HDR sensors and cameras.

The noise of a single pixel belongs to the temporal noise category. Temporal noise cannot be corrected, it can only be reduced by averaging out the pixel data across many samples. This is only possible if the scene is stationary and if enough time is available to acquire the same image several times. There are also requirements on memory. Averaging n images of the same stationary scene will reduce temporal noise by a factor of $1/\sqrt{n}$ if it is assumed that the noise is white, ergodic, and at least wide-sense stationary. If the temporal noise represents 10% of the signal, at least 100 images will be required to bring it below 1%.

A pixel, pixel array, or readout circuit can include antiblooming and anticorona circuits. If the frame rate, sensor resolution, and processing capability permit it, the correction can also be made in a processor. In some applications where more advanced correction is required, the camera does not make any corrections, and the correction and calibration are made in the host computer. Some temporal noise sources can be corrected in the sensor, as will be seen later with the CDS circuit.

2.1.4 Pixel array

Pixels are arranged in an array, typically rectangular. The pixels are never perfect, and each of them differs slightly from the others. This is represented by sensor parameters such as dark signal non-uniformity (DSNU) [Fig. 2.6(a)] and photo response non-uniformity (PRNU) [Fig. 2.6(b)]. These noise sources belong to the category of spatial noises; they do not vary with time, so they can be corrected using a field-programmable gate array (FPGA) and

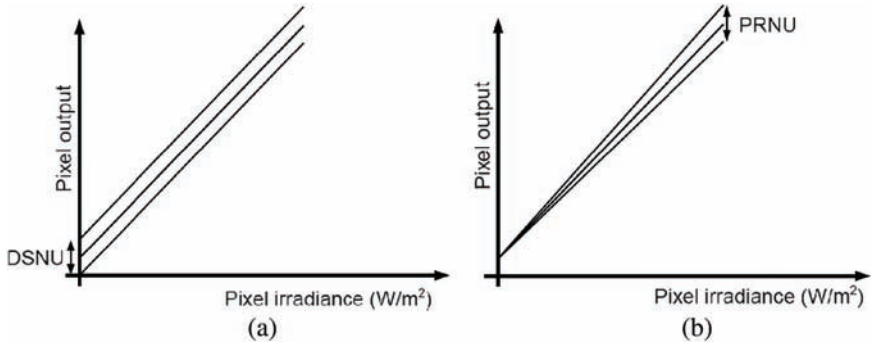

Figure 2.6 (a) DSNU and (b) PRNU.

non-volatile calibration memory if enough calibration memory and processing power are available inside the camera. DSNU depends on temperature and on the age and life history of the sensor. It can also have slow varying components over the sensor array, especially on very large sensors due to small variations in the wafer foundry process across the wafer or temperature gradients within the sensor.

DSNU is a spatially random offset component of zero average that affects each pixel, but its variation in time is deterministic, as it only depends on the exposure time, readout time, and temperature. For common CMOS pixels, the DSNU includes the dark current process (random) and the variation of transistor threshold voltage (fixed). DSNU is an absolute measurement in digital numbers (DNs).

PRNU is a spatially random gain component of unity average that affects each pixel, but its variation in time is deterministic. PRNU is a relative measurement in percentage, referred to a signal level (DN of PRNU/DN of signal). PRNU is wavelength dependent and is usually calibrated in machine vision applications for the wavelength of interest. If a non-monochromatic light source is used, then the PRNU is calibrated together with the target light source but will not be as perfect if the objects in the field of view have a non-uniform reflectance over the light spectrum (which is usually the case). PRNU also depends on the incidence angle of the light on the pixel.

Pixel saturation non-uniformity (PSNU) is the variation in saturation level from pixel to pixel. This noise is normally not visible as gain is usually applied inside the sensor to reach analog-to-digital converter (ADC) saturation before reaching pixel saturation.

2.1.5 Readout circuits

When leaving the pixel array, an image signal is processed by some analog circuits to perform various corrections and conversions (Fig. 2.7). The most

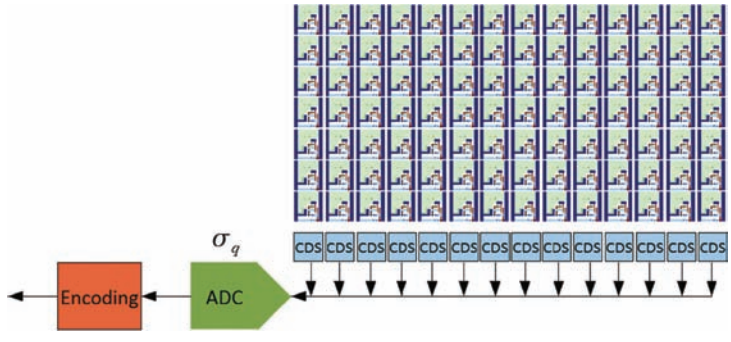

Figure 2.7 Conceptual diagram of a CMOS image sensor. This example considers that the analog data are converted by a single ADC; however, modern CMOS designs have one ADC per column.

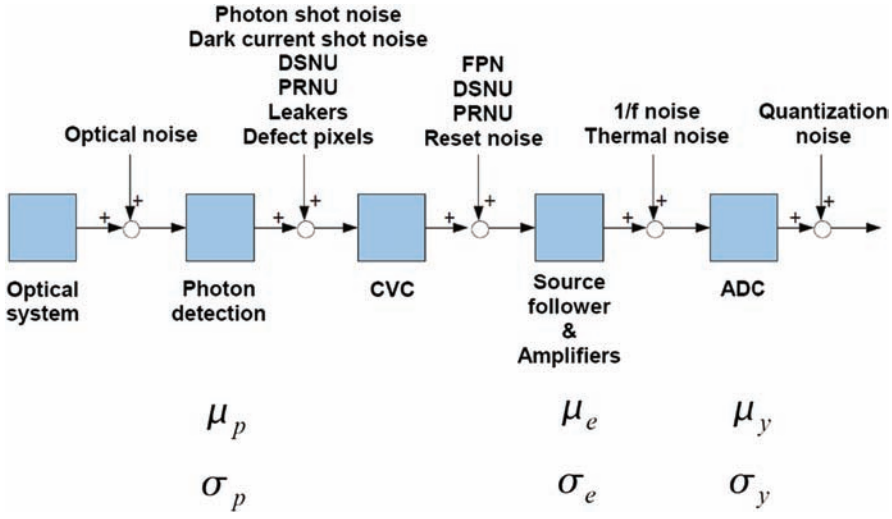

Figure 2.8 Signal chain block diagram of an image sensor with noise sources and symbols for the image mean and image variance at several locations on the signal chain [derived from G. C. Holst and T. S. Lomheim, *CMOS/CCD Sensors and Camera Systems*, Second Edition, JCD Publishing, Winter Park, FL (2011)].

important of these circuits is the ADC, which converts the analog pixel signal to a digital signal usable in the imaging system. Another important circuit is the correlated double sampling (CDS), which will be discussed later.

An ADC has a limited word width (for example, 12 bits). This means that the maximum number of different signal values that can be converted by this ADC is $2^{12} = 4096$ values, ranging between 0 (darkest) and 4095 (brightest). This reduction of signal precision is modeled by a noise source called quantization noise (see Fig. 2.8), and its standard deviation is denoted by σ_q, which always equals $\delta/\sqrt{12}$ digital numbers. Its value is constant per digital number but depends on the ADC resolution, and is therefore lower for a higher-resolution ADC if it is considered in the analog voltage domain. A calculation of this value is provided later. In modern CMOS image sensor designs, one or more ADCs per column are provided so that all columns are converted in parallel (column-parallel ADC). High-speed image sensors use more than one ADC per column so that multiple rows are also converted at the same time.

2.1.6 Image encoding

The signal is then processed in the digital domain. Some additional corrections are performed and the response of the sensor can be adapted, for example by a gamma curve or an S-curve. It is also possible to perform color reconstruction, color improvements, contrast enhancement, or other algorithms.

2.2 Low-Light Imaging Limitations

2.2.1 Noise sources summary

Noises and artifacts can be classified into five categories: optical, temporal, signal level, spatial effects, and defects, which are introduced next.

Optical effects arise in the optical system (lenses) due to imperfections of materials, surfaces, or assembly. Temporal noise sources are, for example, thermal sources such as reset noise, $1/f$ noise, power supply rejection ratio, dark current shot noise, quantization noise, etc. Signal level effects include the device's linearity and photon shot noise. Spatial effects and defects are the dark fixed pattern noise (FPN), the DSNU, the PRNU, high leakage, and contaminated and defect pixels.

Optical effects that arise due to imperfections in the lenses, coatings, materials, and assembly include aberrations and veiling glare. This discussion considers only noise that belongs to the image sensor and electronics. It is understood that optical effects must be taken into account when discussing a complete camera. This is the subject of a later chapter.

Reset noise (also called kTC noise) is the uncertainty of the amount of charge on a capacitor after charging that capacitor through a resistor or the channel of a transistor. It can be calculated by

$$\sigma_{e,reset} = \frac{\sqrt{kTC}}{q}, \tag{2.4}$$

where $k = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant, T is the absolute temperature in Kelvin, C is the capacitance in Farads and $q = 1.6 \times 10^{-19}$ C is the elementary charge (one quantum of charge is the charge of one electron, considered positive). $\sigma_{e,reset}$ is an number of electrons. The equivalent noise voltage for a resistor-capacitor (RC) circuit (using $Q = CV$ and $Q = qN$, where N is an amount of electrons) is

$$V_{e,reset,rms} = \sqrt{\frac{kT}{C}} \tag{2.5}$$

showing that a large pixel capacitance is advantageous regarding reset noise. (It has been demonstrated that, in many designs, because of the subthreshold operation of the reset transistor, kTC noise is actually about half of what this simple formula states. However, in these cases the reduction in reset noise is accompanied by an increase in image lag for very dark signals.)

Reset noise can be corrected by correlated double sampling (potentially on-chip for CMOS devices), especially for pinned photodiode sensors. Double sampling is a measurement of the difference between two samples. With correlated double sampling, one sample is the pixel in its reset state just before

the start of exposure, and the other is the pixel in its partially discharged state at the end of exposure (see Fig. 2.9). The difference between the two samples is only due to pixel discharge and not reset. Correlated double sampling is usually found in rolling-shutter CMOS sensors. With uncorrelated double sampling, the reset sample is taken before the start of the next exposure and is therefore not correlated to the sample at the end of the current exposure. Uncorrelated double sampling only partially corrects reset due to the lack of correlation; it only corrects for offset errors and does not correct any reset noise (it can even increase noise, depending on its implementation, as two uncorrelated samples are taken). A global shutter pixel with correlated double sampling exists; in this pixel, the reset level is stored in an additional capacitor within the pixel. This solution allows for reset noise reduction at the expense of a lower fill factor.

Note that the term "correlated double sampling" is often abusively used for double sampling (uncorrelated). For this reason, real correlated double sampling is sometimes called true correlated double sampling (or real correlated double sampling).

Thermal or Johnson noise is the noise generated in resistors and in the channel of MOS transistors. It is defined as

$$V_{Johnson,rms} = 2\sqrt{kTRB}, \tag{2.6}$$

where R is the resistance and B is the bandwidth. This shows that optimizing bandwidth, cooling the device, and reducing resistance (i.e., increasing the width of the transistors) is favorable for CMOS imagers. However, increasing the width will usually reduce conversion gain. The Johnson noise, also called thermal noise or Nyquist noise, is caused by the random motion of electrons within a resistor, wire, or transistor channel due to their vibration energy at non-zero temperatures.

Flicker noise or $1/f$ noise in MOS transistors is inversely proportional to the frequency and area of the transistor's channel. It is again favorable to design with large transistors. A possible countermeasure is multiple sampling.

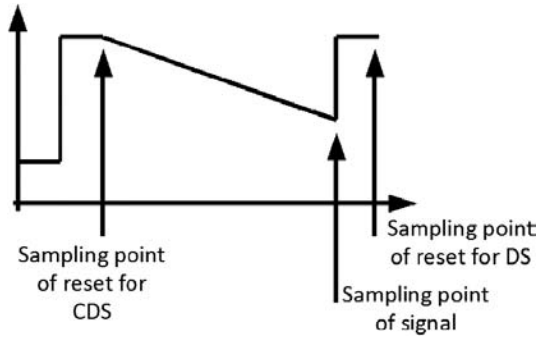

Figure 2.9 Principle of correlated double sampling.

Power supply rejection ratio (PSRR) is the inverse of the sensitivity of a device to external noise sources. If a noise is present in the device's supply, especially the analog supply or other supply lines by coupling, then part of this noise will be transferred to the signal. A high PSRR is desirable. Power supply noises can be highly visible in the image if they are in phase with the pixel clock, row clock, or frame clock. This noise can be analyzed by the spectrogram method with additional phase information. The spectrogram method is defined in the EMVA1288 standard.

The noise sources previously mentioned are encountered in the pixel, but also in all analog circuitry present in the device.

Quantization noise (Fig. 2.10) is a mathematical representation of the limited resolution of the ADC. Signal levels that fall in-between two ADC codes cannot be distinguished from each other. This is represented by a noise of $\sigma_q = \delta/\sqrt{12}$. This noise can be calculated as follows. The analog signal that falls between two ADC codes is uniformly distributed in the range $[\alpha - \frac{1}{2}; \alpha + \frac{1}{2}]$, where α is the closest digital number representation. The variance of such a distribution can be calculated as

$$\sigma^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} x^2 f(x) dx = \frac{1}{\frac{1}{2} - (-\frac{1}{2})} \int_{-\frac{1}{2}}^{\frac{1}{2}} x^2 dx = \left(\frac{x^3}{3}\right)_{-\frac{1}{2}}^{\frac{1}{2}} = \frac{1}{24} + \frac{1}{24} = \frac{1}{12}. \quad (2.7)$$

The ADC resolution should be appropriate (i.e., the ADC step size should be compatible with the signal noise so that quantization noise is negligible). Most characterization measurements must also have a quantization noise much smaller than the noise to be measured so that the measured digital samples provide enough resolution on the distribution of the noise. This is

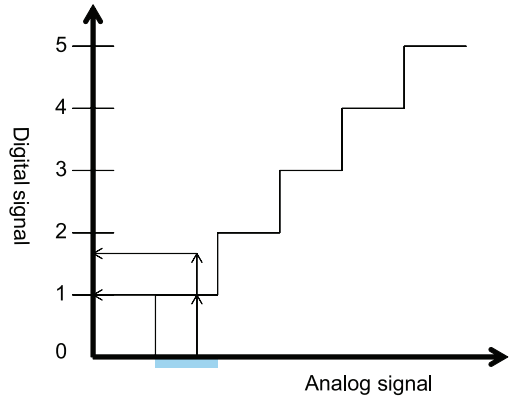

Figure 2.10 Quantization noise: the range of analog levels represented by the light blue area will give the same digital code, 1. Other ADC implementations are offset by half a code and therefore have a centered error.

especially important for low-light conditions because they have the lowest noise.

Dark-current shot noise is the uncertainty of the generated dark-current electrons. As this process is Poisson distributed, for N dark-current electrons there is an uncertainty (or noise) of $\sqrt{N}$ electrons. This noise can be reduced in cooled sensors and is strongly layout dependent. Larger pixels have more layout freedom (for a given technology), and can thus be more optimized for a low dark current, which implies low dark-current shot noise. Dark-current shot noise should not be confused with dark noise, which includes dark-current shot noise as well as other noise sources. Dark-current shot noise is often one of the least important contributors to dark noise. The dark-current shot noise becomes one of the dominant noise sources in high-temperature applications, such as automotive on-board cameras or oil- and gas-well inspections.

Dark current is caused by the leakage currents of diodes and other devices. The photocurrent (the current induced by light in the photodiode), when very small, is primarily affected by and can vanish within the dark current. Dark current is made of a constant component (exponentially affected by temperature) and a variable component, i.e., the dark-current shot noise. Only this variable component is a temporal noise. The constant component can vary from pixel to pixel and is one of the contributions to DSNU; it can be corrected by calibration.

The importance of dark current has been reduced by three orders of magnitude in the past 30 years, and its root cause of foundry contamination is mostly resolved. Therefore, new phenomena are now the main causes of dark current, which has also changed the way dark current behaves over pixel position and temperature. It is not uncommon to now see a multi-normal distribution of dark current over a CMOS image sensor, corresponding to different sources of dark current. Each of these causes has its own temperature dependence, and thus the temperature evolution of dark current is different for each pixel. Considerable nonlinear calibration is required for applications with large changes in operating temperature.

Photon shot noise is the uncertainty of the amount of electrons generated for a given exposure. It is a Poisson distributed process originating from two sources: the uncertainty of the amount of photons hitting a pixel for a given exposure, and the uncertainty of the net amount of electrons converted out of these photons. For N photons, the noise is $\sqrt{N}$. As this is a signal-dependent noise, it is not present in dark images and barely present in low-light images.

Next, spatial noise sources are examined.

Dark FPN is a combination of spatial variations of dark-current generation (DSNU) and the transistor threshold voltage spatial variations. Although spatial noise is usually random, it is often noticed that some lines or columns have a different offset. It is common to also see gradients or rings in a

dark image. These effects arise from process variations across the wafer, stress produced by wafer dicing or sensor assembly, or hot spots in the sensor that cause a temperature gradient across the pixel array. This comes from variations in line drivers or column amplifiers, or some other line or column defects. For a given temperature and exposure time, this noise can be corrected off-chip (or on-chip for some smart area sensors and line scan sensors). Most cameras have embedded FPN correction based on calibration images and temperature sensors. Double sampling or correlated double sampling also reduce part of the FPN.

When looking at a logarithmic dark-image deviation histogram (the logarithmic histogram of the absolute deviation from the mean), three regions can be identified (Fig. 2.11). In the leftmost region, the log histogram follows a parabola due to dark-current distribution. Some pixels show a deviation due to various variations, including the dark FPN. Finally, there are some pixel outliers that have more important defects (defective and contaminated pixels) or an excessive dark current (leakage). The log-histogram exhibits a parabola because a Gaussian (or normal) distribution in a linear plot becomes a parabola in a logarithmic plot.

Note that contaminated pixels are not necessarily dead or dark. It all depends on the type of particle contaminating the pixel, its size versus the pixel size, and its location on the pixel. Contaminated pixels can also sometimes be brighter than normal due to an optical effect that makes the contamination behave like a small lens.

There are many causes of defect pixels, including process defects, crystal lattice defects, cosmic rays, contaminating atoms, dirt from the packaging process, and dust. The behavior of defect pixels can range from completely non-functional pixels stuck as black or white, or even some intermediate gray level, to partially functioning pixels clipping at some intermediate level, or blinking

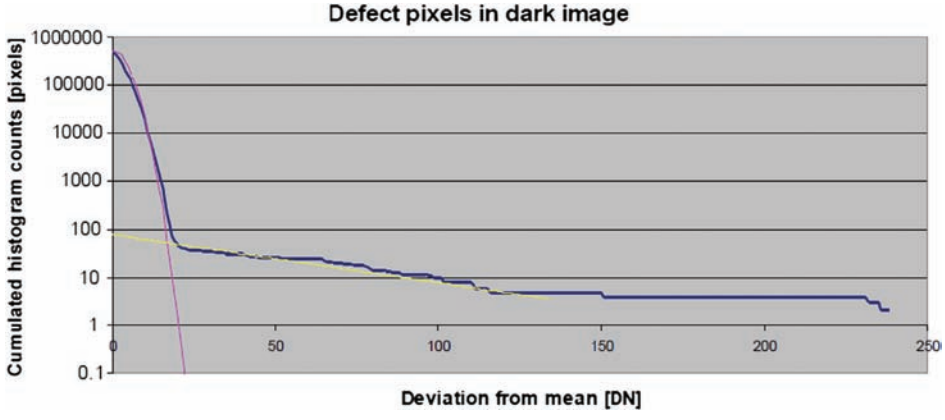

Figure 2.11 Logarithmic histogram of a dark image of a CMOS sensor, filtered to remove temporal noise contributions.

pixels that change from white to black at a random rate from frame to frame. Some pixels can behave properly in some frames and be incorrect in others. Another type of defect involves random telegraph noise (RTN) pixels, which have a variable offset from time to time that can only take a few discrete values and will change randomly after a random number of frames. Other defects include pixels that have a behavior very different from most of the other pixels, including errors in linearity, response, noise amplitude, or lag.

Photo response non-uniformity is due to various gain errors in a pixel or readout circuits. It is layout dependent and can be corrected off-chip by means of subtraction and division by constants, if the device has been calibrated for each pixel. PRNU can be wavelength dependent, and therefore the calibration depends on the type of illuminant. PRNU also depends on the angle of incidence of light onto the pixel. It may be different for +X, −X, +Y, and −Y directions.

Note that thermal noise, dark current, and dark-current shot noise increase exponentially with temperature. That means that dynamic range decreases exponentially with temperature, because (as will be seen) the lower limit of dynamic range is the noise floor.

Dynamic range decreases with the increase of exposure time as more dark-current shot noise is accumulated.

Temporal noise also increases with pixel frequency (because readout noise increases and because the device heats up) but may decrease with frequency (because there is less dark noise during readout). This is of higher importance in global shutters than in rolling shutters. There is an optimum pixel frequency at which the dynamic range is maximum. This optimum frequency depends on the device's temperature.

Noise equivalent image lag or incomplete pixel reset can also be included here. If the sensor is a charge-coupled device (CCD), blooming and smear should be considered, as this effect also reduces dynamic range by destroying the signal of neighboring pixels in cases of oversaturation of a sensor area. There are also spatial and signal level limitations of dynamic range.

2.2.2 Lowest detectable limit

The previous paragraphs have shown that an image sensor's performance is limited by various kinds of noise. The sum of all of these noises defines the lowest detectable light power, which corresponds to the darkest light level that can be detected by the sensor in the given configuration. This darkest light level is defined as the irradiance for which the signal-to-noise ratio (SNR) is SNR = 1, i.e., the signal is equal to the noise. (Of course, if the image must be displayed to a user, another limit on SNR can be defined, and then the minimum reasonable SNR for a decent image must be discussed.)

This level cannot be practically measured due to the potential clipping of some pixels close to dark. To maintain noise whiteness at this level, the

EMVA1288 standard recommends offsetting the analog signal so that no significant number of pixel clips and statistics can be extracted from the data. The standard also recommends interpolation techniques through many illumination levels, for example, the use of the photon transfer method with many measurement points to properly interpolate the photon response plot until the SNR = 1 point [61]. This is the lowest bound of the dynamic range, denoted by DR_{Lo}, where DR stands for dynamic range.

This limit can be calculated from the noise formulas if it is assumed that all types of noise sources are uncorrelated. In this case, the total noise is given by $\sigma_{tot,dark} = \sqrt{\sum_i \sigma_i^2}$, where each of the noises is expressed in digital numbers at the output of the device.

2.3 Bright-Light Imaging Limitations

2.3.1 Saturation

A definition of saturation encountered in many books, datasheets, and presentations is as follows: saturation is defined as the largest possible output signal. This definition can cause some practical difficulties with real sensors, even after FPN, DSNU, and PRNU correction, due to shot noise and readout noise. A more practical definition is: saturation is the light level for which the image variance is maximum. Increasing the irradiance causes the variance to decrease, meaning that a significant number of pixels start to clip. As no information can be retrieved from saturated pixels (but a small amount of saturated pixels does not greatly affect the overall image information), the maximum variance criteria is a good practical definition of saturation for most machine vision applications (see Ref. 61 for more information). Depending on the target application, other definitions of saturation can be required. For example, saturation occurs as soon as at least one pixel saturates or when all pixels are saturated.

Note that the maximum of the variance does not correspond to the first clipping pixel but occurs when a significant amount of pixels are clipped. It also does not correspond to the saturation of the average image, as this is reached only when all pixels have clipped. This definition is also very useful for sensor characterization, as the variance versus exposure plot (Fig. 2.12) is linear for a linear sensor up to this level (just defined as saturation). This is the root of a measurement method called the photon transfer method (Figs. 2.13 and 2.14). The photon transfer method is also used as a key measurement procedure in the EMVA1288 standard.

Figure 2.15 shows a simulation of a very small sensor image with PRNU and no temporal noise besides the photon shot noise for four different light levels. The first image does not have any saturated pixels and is properly exposed as its histogram is centered. The second image has one saturated pixel,

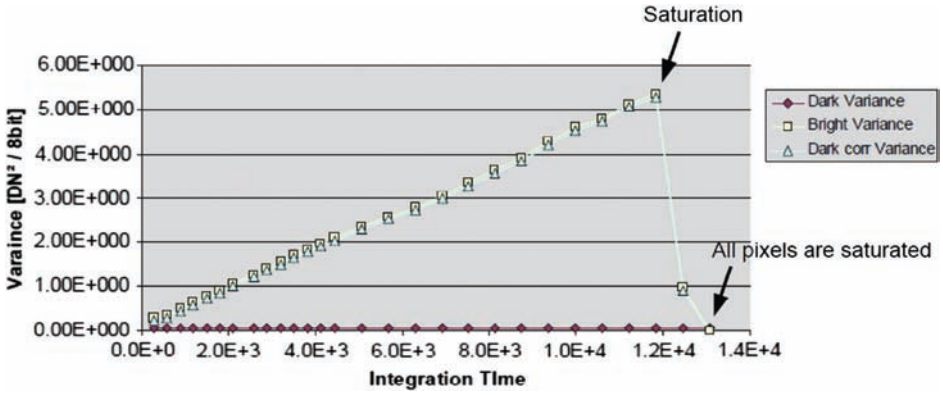

Figure 2.12 Image variance vs. exposure (measurement): photon transfer curve.

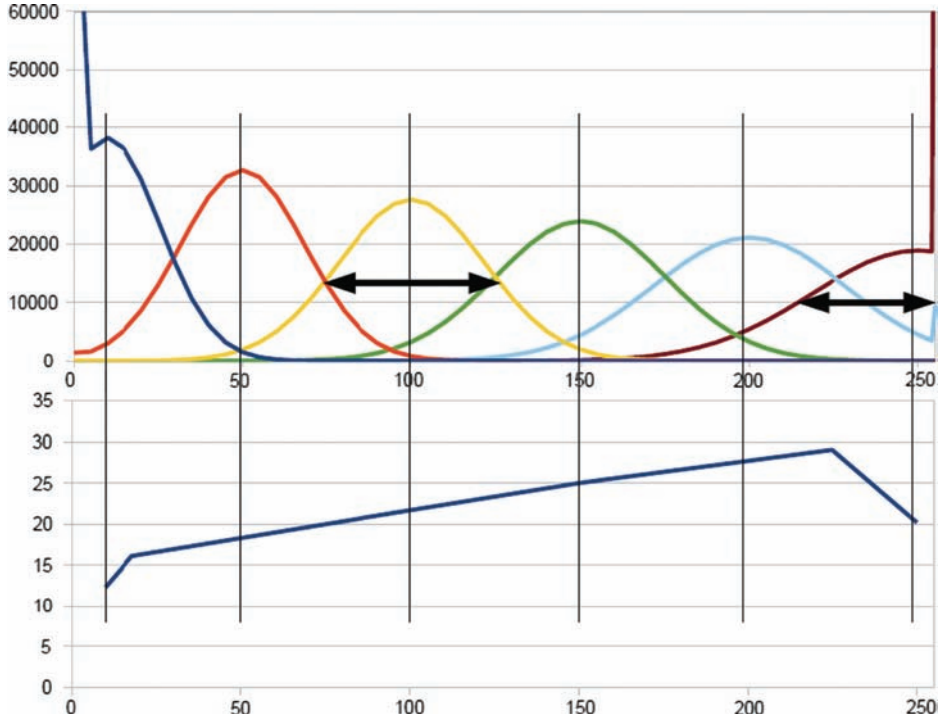

Figure 2.13 Top: simulated, ideal histograms for various illuminations showing spread and saturation. Spread increases with light. Bottom: corresponding photon transfer method (plot of variance vs. mean [61]) showing limits due to saturation.

and the third image shows the image that corresponds to our definition of saturation. The last image has all but one pixel saturated. The pattern of the pixels is clearly visible up to the third image but is no longer visible in the last one, which shows that the definition of saturation as the level where all pixels

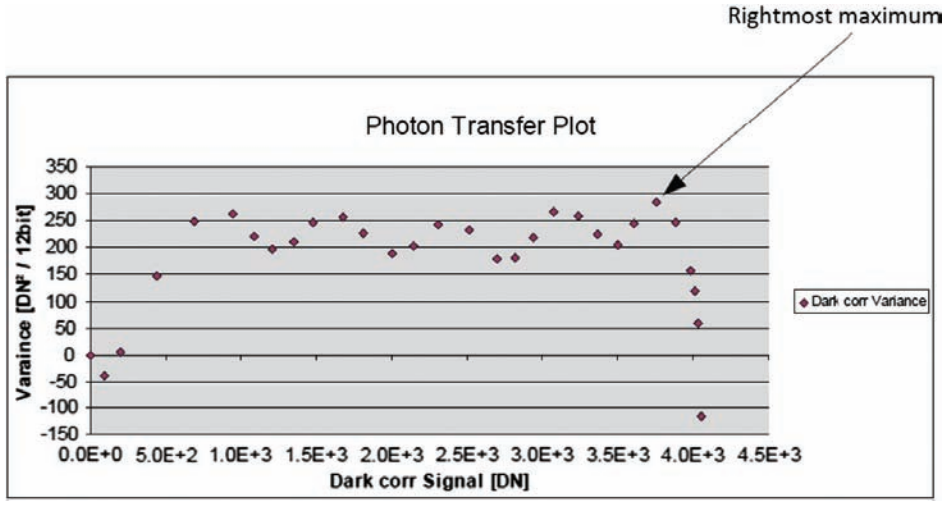

Figure 2.14 Photon transfer curve of a piecewise linear sensor (measurement).

have saturated (i.e., when the average of the image reaches saturation) is not appropriate because the image no longer contains information. It also shows that defining the saturation limit as the light level at which one pixel saturates is too pessimistic because useful information is still available in the image.

The difference in terms of dynamic range expressed in decibels is very small between the two definitions. Assume that the saturation level per the variance definition is 90% of the total pixel range; the dynamic range is thus only reduced by a little more than 10%, or about −1 dB.

For nonlinear sensors, if there are several maxima in the variance curve, it is the rightmost maximum that must be considered. Other maximums are considered later when discussing the dynamic range gaps. Note that most HDR sensors are nonlinear.

It is important to notice the difference between a saturated pixel and the image saturation. A saturated pixel is a pixel that has reached its maximum level in an image. Due to photon shot noise and readout noise, it might not be saturated in all images. Saturation of the entire image is reached when there are significant number of pixels that are saturated. From one image to the next, the saturated pixels are not necessarily the same, but the average remains the same due to the large number of pixels considered in a complete image.

From this definition, it can be concluded that a sensor with a larger readout noise will have a lower saturation level. As this sensor also has a higher minimum detectable level, it is limited in performance on both sides of the dynamic range. Minimizing readout noise is therefore key for HDR image sensors. If they cannot be corrected, PRNU and DSNU will also reduce saturation and impact dynamic range. The photon transfer curve is plotted after DSNU and PRNU correction; only the temporal part of the noise is used.

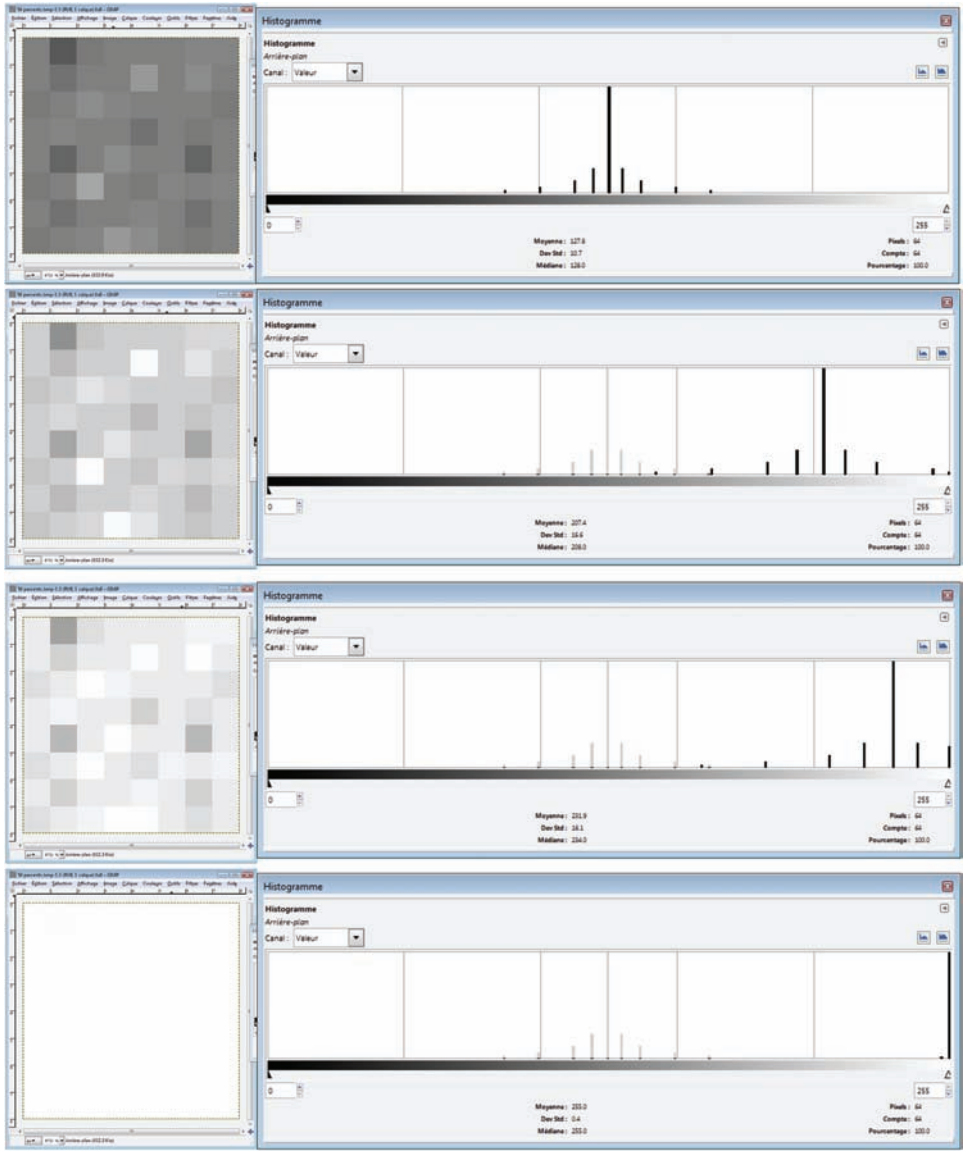

Figure 2.15 Explanation of saturation in the presence of noise using a small resolution image.

It can also be concluded from this definition that a higher photon shot noise affects the saturation level and therefore reduces dynamic range. However, as will be seen, a higher SNR is the goal, which involves a tradeoff in sensor design parameters.

Note that saturation (as just defined) is always lower than saturation defined as the maximum output signal. Note also that some sensors have almost no saturation at all. This is the case for some sensors that exhibit a

logarithmic behavior, and it is also the case for silver halide photographic film. It is almost impossible to burn highlights with a film camera. In this case, it is necessary to define a specific estimate of saturation based on the application requirements or based on the incremental SNR (iSNR). The ISO standards propose some definitions for the saturation of photographic films.

2.3.2 Highest detectable level

The highest detectable light power (which corresponds to the brightest level detected by a sensor in a given configuration) is the saturation level. This light power defines the highest boundary of the dynamic range, denoted by DR_{Hi}.

2.4 Signal-to-Noise Ratio

Figure 2.16 summarizes the noise sources, SNR, and dynamic range in a sensor and image (including optical veiling glare). For the remainder of the calculations, optical veiling glare is not considered and is explained later.

SNR is defined as the ratio between the signal level and the total noise level, and is a function of pixel irradiance. SNR can be calculated for a linear sensor with the following formula (plotted in Fig. 2.17 on a logarithmic scale):

$$SNR(\mu_p) = \frac{\eta\mu_p}{\sqrt{(\sigma_d^2 + DSNU^2) + PRNU^2\eta^2\mu_p^2 + \frac{\sigma_q^2}{K^2} + \eta\mu_p}}. \tag{2.8}$$

K is the overall system gain taking into account the in-pixel gain and the readout gain. A proof of this formula, a formula that is used in older revisions of the EMVA1288 standard, will be given later.

The limiting cases are of interest. In the high-photon range where

$$\eta\mu_p \gg \sigma_d^2 + DSNU^2 + PRNU^2\eta^2\mu_p^2 + \sigma_q^2/K^2, \tag{2.9}$$

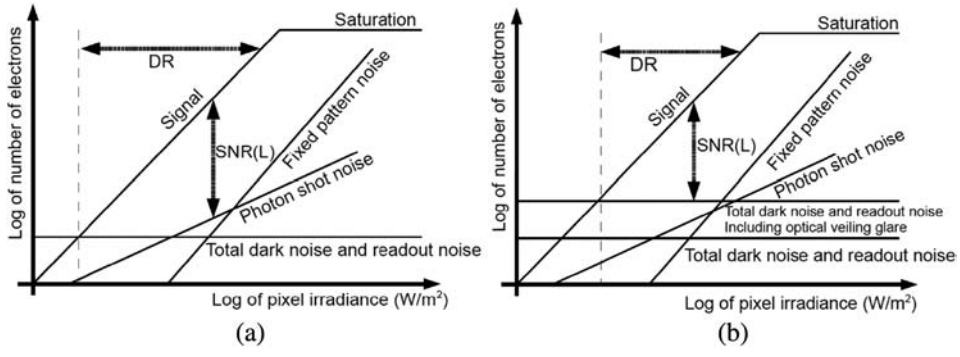

Figure 2.16 Summary of noise sources, SNR, and dynamic range of (a) a sensor and (b) an image.

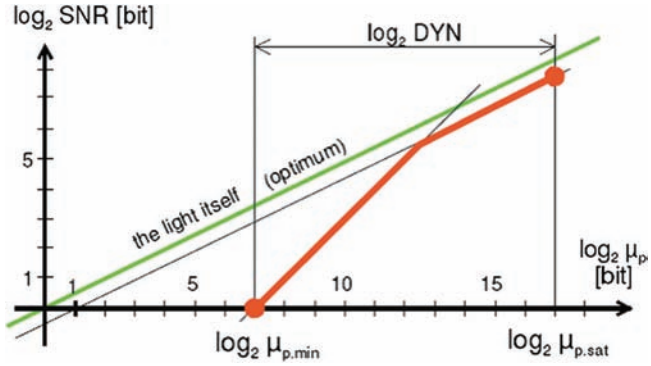

Figure 2.17 SNR of the "light itself" and a sensor (using an approximated shape for the sensor's SNR curve). The SNR and sensor level are both expressed in bits (reproduced from Ref. 63).

the SNR increases only with the square root of the amount of photons while in the low limit, where

$$\eta\mu_p \ll \sigma_d^2 + DSNU^2 + PRNU^2\eta^2\mu_p^2 + \sigma_q^2/K^2, \tag{2.10}$$

the SNR increases linearly as

$$SNR(\mu_p) \approx \frac{\eta\mu_p}{\sqrt{\sigma_d^2 + DSNU^2 + \frac{\sigma_q^2}{K^2}}}, \tag{2.11}$$

which is linear. These two limit cases can be used to draw the simplified SNR curve of Fig. 2.17.

The SNR plot is the most important description of an image sensor as it shows on a single plot the sensitivity, the saturation, dynamic range, quantum efficiency, and SNR. An important point in the SNR plot is where the slope 1 curve joins the slope 1/2 curve. For light intensities larger than the one corresponding to this point, the sensor is said to be photon shot noise limited, and this is where the sensor optimally works. The SNR plot is the most important plot of the EMVA1288 template datasheet, but unfortunately it is usually not provided in non-EMVA1288 compliant datasheets.

2.5 Dynamic Range Gaps

2.5.1 Response curve

A response curve is a plot of the sensor output (in the digital domain) against the pixel irradiance. It is usually measured as the average output of the sensor (in DN) versus an average irradiance (in photons or in W/cm^2).

It is the shape of the response curve that defines the type of sensor response: linear, logarithmic, S, lin-log, piecewise linear, etc. (see Figs. 2.18 and 2.19).

Note that although the vertical axis is the digital output, the curve is usually not plotted as a staircase. This is due to the averaging that is performed over a large part of the array. Individual pixels can only have discrete values, but the average value of an image is a real number.

The first derivative of this curve is called the incremental gain:

$$g = \frac{d\mu_y}{d\mu_p} = \frac{\sigma_y}{\sigma_p}. \tag{2.12}$$

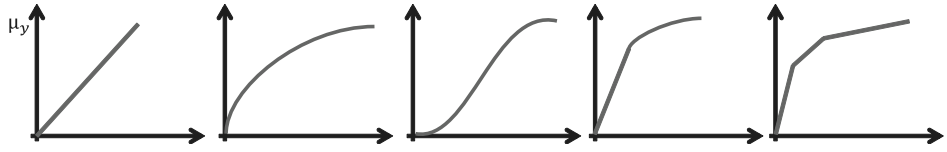

Figure 2.18 Response curve of various pixels: (a) linear, (b) logarithmic, (c) S, (d) lin-log, and (e) multiple slopes (piecewise linear response, also called multiple-slope response or multiple-segment response).

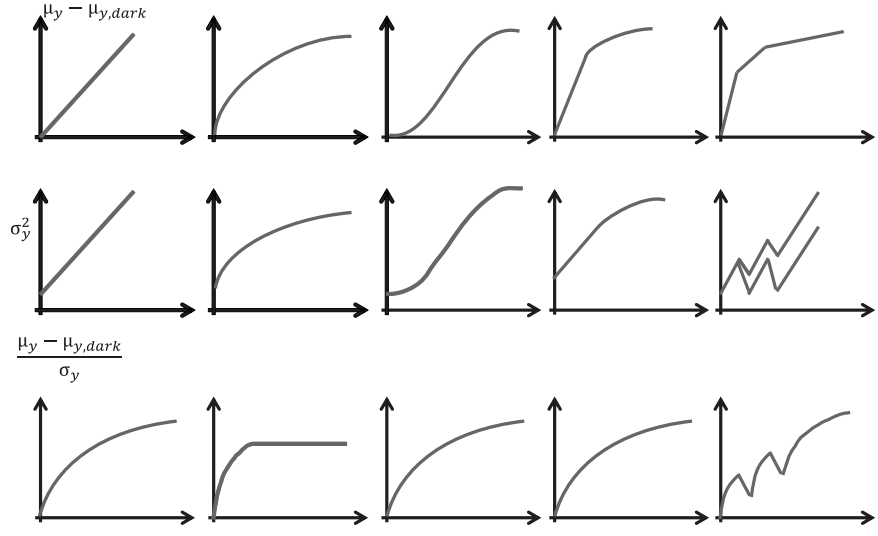

Figure 2.19 (a) Dark corrected response curve, (b) total noise, and (c) SNR of various pixels: linear, logarithmic, S, lin-log, multiple slopes (piecewise linear response). The noise curve for logarithmic sensors depends on the sensor design: some sensors have a logarithmic noise curve, some have a linear noise curve, and some have an almost-flat curve (temporal noise that is almost independent from the light level). The multiple-slope pixel has two different noise behaviors depending on how the pixel is designed.

2.5.2 Dynamic range gaps

The signal variation is considered the carrier of information. Fine details of the image are carried by small variations of light levels. Scene information can only be detected reliably if the image capture system can detect the small signal variations that carry that information. These small variations are present at any point of the dynamic range, and it is then possible to miss details of the scene at other locations than the brightest and darkest areas of the scene.

Similarly to the darkest detectable level DR_{Lo}, the mathematical condition that translates this criterion is

$$SNR = \frac{\mu_y - \mu_{y,dark}}{\sigma_y} \geq 1. \tag{2.13}$$

When this condition is not verified, the dynamic range is said to have a gap. For a linear sensor, the expression can be rewritten as

$$SNR = \frac{\mu_y - \mu_{y,dark}}{\sqrt{K^2\sigma_d^2 + \sigma_q^2 + K(\mu_y - \mu_{y,dark})}} \geq 1, \tag{2.14}$$

using Eq. (2.8), with $\mu_y - \mu_{y,dark} = K\eta\mu_p$ and neglecting DSNU and PRNU. The noise's expression must be modeled or measured for every pixel type.

There are no gaps for linear sensors or sensors that have a smooth response curve, but nonlinear sensors for which the slope of the response curve changes abruptly can have dynamic range gaps at the transition points. Dynamic range gaps of multiple-slope sensors are discussed in Chapter 3. Reference 62 describes dynamic range gaps in terms of incremental SNR (iSNR) based on the ISO15739 incremental gain formula. In this reference, dynamic range gaps are called SNR holes. Dynamic range gaps are expressed in the horizontal axis as a gap in the dynamic range; SNR holes are expressed in the vertical axis for SNR < 1.

If the derivative of the response curve is too close to zero, no information is transferred to the image, although saturation is not yet reached. This is also a dynamic range gap.

Figure 2.20 illustrates the effect of dynamic range gaps in a 100-dB automotive scene. In Fig. 2.20(a), the sensor is running in linear mode, giving no details at all in the bright and dark areas (respectively because of saturation and noise limit). The dynamic range of the sensor in this case is about 60 dB.

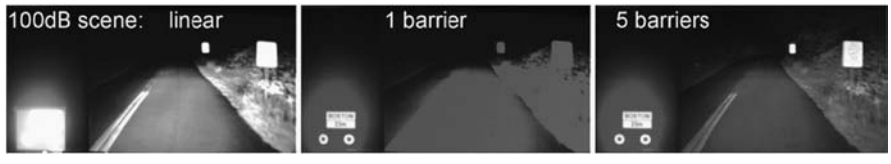

Figure 2.20 Illustration of dynamic range gaps in a typical automotive scene (reproduced from Ref. 62).

In Fig. 2.20(b), the dynamic range of the sensor is 100 dB, but the strong nonlinear and non-smooth response curves used (piecewise linear response with two segments, with two segments equaling one barrier, as there is one barrier for each kneepoint) generate a large gap in the dynamic range. Although details of the dark and bright areas are visible, there are no details in the mid-gray tones.

In Fig. 2.20(c), the sensor still provides 100 dB of dynamic range, but this time the response curve features six segments (five barriers), which does not generate any dynamic range gap. The full 100 dB of the range are usable.

2.5.3 Presence function of dynamic range gaps

This function is used to localize the dynamic range gaps (Fig. 2.21). It is defined as

$$DRG(\mu_p) = \begin{cases} 1 & \text{when } SNR \geq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.15}$$

A zero value means that a dynamic range gap is present. A dynamic range gap does not need to be present at every sudden drop of SNR, as seen in Fig. 2.22.

Figure 2.21 Illustration of dynamic range gaps in a typical automotive scene with gaps around lights. Reflective signs and one person's details are lost.

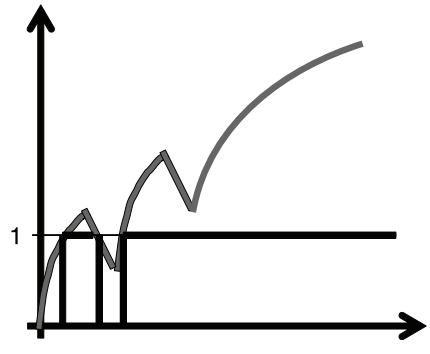

Figure 2.22 Dynamic range gap presence function.

2.6 Dynamic Range

2.6.1 Definition

Dynamic range (Fig. 2.23) is then defined by

$$DR = \frac{DR_{Hi}}{DR_{Lo}} \quad \text{or} \quad DR[dB] = 20 \log \frac{DR_{Hi}}{DR_{Lo}}. \tag{2.16}$$

2.6.2 Remark

Dynamic range can be misleading in datasheets. Some datasheets report dynamic range as a typical value for one configuration; other datasheets report the maximum intraframe dynamic range, which is the maximum reachable dynamic range among all possible configurations; other datasheets report a reasonable value for dynamic range that is reached by most sensor configurations; and some datasheets report a purely hypothetical value.

For example, one of the XDR CMOS image sensors available on the market has approximately 60 dB of dynamic range in linear mode, 154 dB of maximum intrascene dynamic range, around 180 dB of maximum reachable dynamic range over several images, and 130 dB of dynamic range in most reasonable configurations.

When working with sensor and camera dynamic range, it is important to always check the formula definition and the measurement conditions. When working with cameras, it is important to check if the published dynamic range refers to the sensor or to the camera, because the camera's dynamic range is reduced by optical veiling glare. It is surprising to see that some cameras are not painted in black or anodized inside their lens mounts, making them possible sources of reflections.

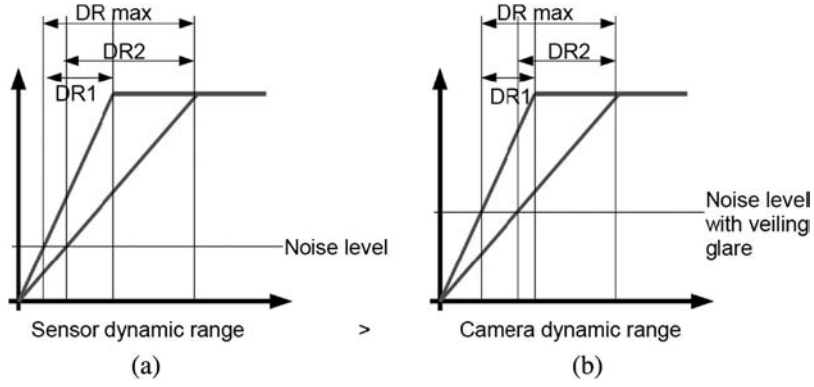

Figure 2.23 (a) Dynamic range of a sensor for short exposure time (DR2) and long exposure time (DR1), and the sensor's maximum reachable dynamic range (DR max); (b) dynamic range of a camera for short exposure time (DR2) and long exposure time (DR1), and the camera's maximum reachable dynamic range (DR max).

2.6.3 Relative measurement

Dynamic range is a relative measurement (Fig. 2.24) between a saturation level and a noise floor:

$$DR_{Hi,2} = \alpha DR_{Hi,1} \Leftrightarrow \log\, DR_{Hi,2} = \log\, \alpha + \log\, DR_{Hi,1}, \tag{2.17}$$

and similarly for the lowest bound.

$$\log\, DR = \log \frac{DR_{Hi,2}}{DR_{Lo,2}} = \log \frac{\alpha \mathrm{DR}_{Hi,1}}{\alpha DR_{Lo,1}} = \log \frac{DR_{Hi,1}}{DR_{Lo,1}}. \tag{2.18}$$

Scaling both elements will always yield the same dynamic range. As discussed earlier, there is a trade-off between performance in the dark areas and performance in bright areas, hence the need for parameter scaling.

2.7 Image Information

The incremental gain defined above is a measurement of the amount of information that is transferred from the scene into the image. When the incremental gain is high, a small variation of luminance results in a large variation of the sensor's output, which means that a lot of details, i.e., a lot of information, are transferred from the scene into the image. When the incremental gain is low, a small variation of luminance results in almost no variation of the sensor's output, which means that almost no details, i.e., almost no information, are transferred from the scene into the image. When a dynamic range gap is present, there is no information at all transferred from the scene into the image.

The function

$$I(\mu_p) = g(\mu_p) \cdot \mathrm{DRG}(\mu_p) \tag{2.19}$$

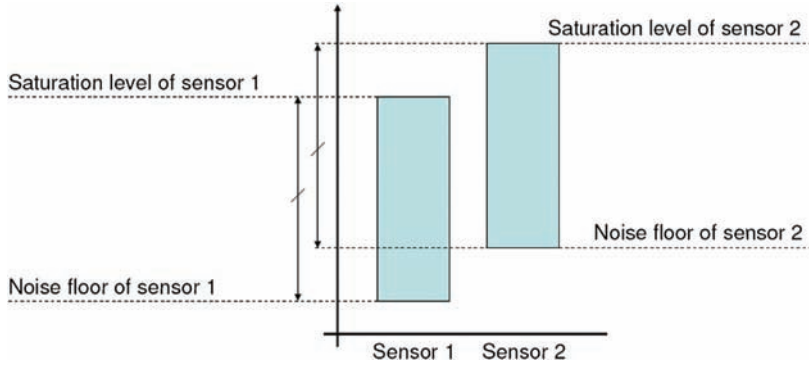

Figure 2.24 Dynamic range is a relative measurement (logarithmic vertical axis).

is a representation of the amount of information that can be transferred between the scene and the image, and is called the information transfer function. If

$$I(\mu_p) = \frac{d\mu_y}{d\mu_p} \cdot \mathrm{DRG}(\mu_p) \approx \frac{\Delta\mu_y}{\Delta\mu_p} \cdot \mathrm{DRG}(\mu_p) < \frac{ADC_{step}}{\Delta\mu_p}, \tag{2.20}$$

then no information can be retrieved from the variations in a scene smaller than $\Delta\mu_p$, as shown in Fig. 2.25. (Due to noise, if several images are averaged to obtain a floating point representation of the pixel value, it is possible to obtain information from an irradiance variation of less than $\Delta\mu_p$. Averaging images is the same as having less noise and therefore a better SNR curve; a better SNR curve is further from the SNR = 1 limit.)

A noiseless linear sensor operating on the same ADC would transfer at most (if the dynamic range gap close to total dark is ignored) an information quantity of

$$I_{linear}(\mu_p) = \frac{ADC_{Range}}{DR_{Hi}}. \tag{2.21}$$

One can compare the information transfer of any sensor to this ideal linear sensor using the information gathering ratio, defined as

$$i = \frac{I}{I_{linear}}, \tag{2.22}$$

or its average:

$$\bar{l} = \frac{\frac{1}{DR_{Hi}} \int_0^{DR_{Hi}} I(\mu_p) d\mu_p}{\frac{ADC_{Range}}{DR_{Hi}}} = \frac{1}{ADC_{Range}} \int_0^{DR_{Hi}} I(\mu_p) d\mu_p, \tag{2.23}$$

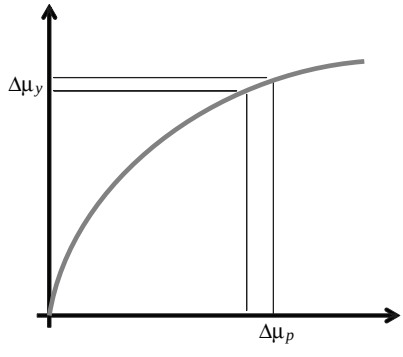

Figure 2.25 Relationship between irradiance variation and sensor output variation.

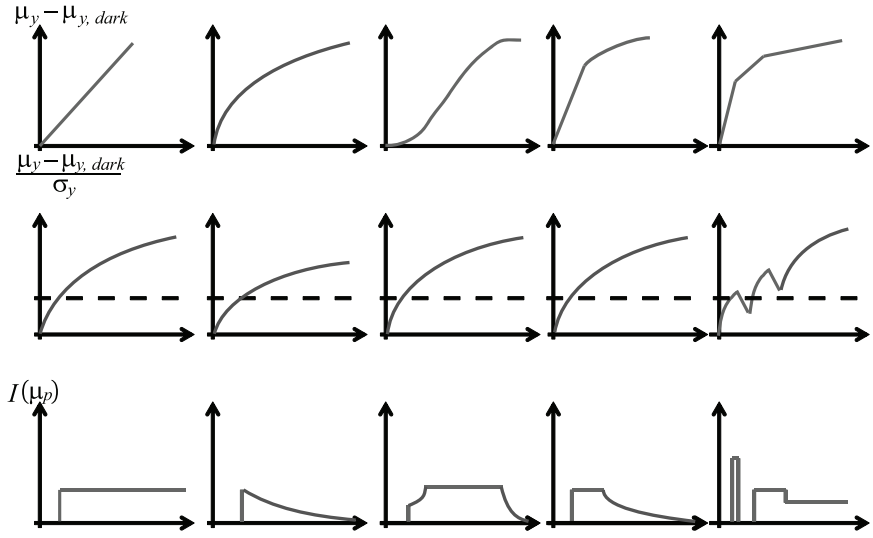

Figure 2.26 Information transfer functions for several sensor-response curves.

where the DR_{Hi} value is the same for the XDR and the linear sensors. It is also possible to define a similar formula for the case where the linear sensor does not have the same saturation limit as the XDR sensor. In this case, an integral of the information curve of the linear sensor that takes into account the saturated part up to the saturation point of the XDR sensor shall be used.

Figure 2.26 shows examples of information transfer functions for several sensor response curves [remember that $(d/d_x)\log_a(x) = 1/(x \ln a)$]. There is no information transferred when $SNR < 1$ (dynamic range gaps) and the information transfer is proportional to the derivative of the response curve, where $SNR \geq 1$.

2.8 Image Information of a Real Scene

When dealing with real scenes, the brightnesses of objects are not uniformly distributed over the range of luminances between the dark limit and sensor saturation. Some of them are darker than the lowest detectable level, some of them are brighter than the saturation limit, and between these two boundaries the other values have a non-uniform distribution.

It is therefore important to have a good information transfer function for the parts of the histogram that contain a lot of data, and it is acceptable to use an inferior information transfer function for the parts of the histogram that contain less data. This principle is used in histogram-equalization tone mappings, explained later in this book.

The histogram represents the probability to find a pixel with a given brightness in an image. It can be plotted as a function of the amount of

incoming photons and scaled so that it has an area of unity. Then each point of the histogram can be multiplied with the corresponding point of the information transfer function to obtain a plot of the image information histogram. The total area of this new histogram is representative of the information actually acquired by the imaging system for the given scene.

Unfortunately, the histogram of the real scene is usually unknown.

2.9 Human Vision System and Its Dynamic Range

Human vision is not covered in detail here due to its extreme complexity. However, it is important to understand that the eye behaves very differently compared to a camera, and the human eye, if not an ideal sensor, is only a good image sensor that is familiar to us.

2.9.1 General properties of human vision

Rods are present in large quantities across the entire surface of the retina (wide field of view). They have high sensitivity but a low resolution. Vision based on rods is called scotopic, and the spectral response is called the scotopic curve.

Cones have a significantly lower sensitivity but higher resolution (it is said that their resolution is comparable to a 10-megapixel sensor). They are more sparsely populated in the periphery, and most are found around the center of the complete field of view of the rods. Vision based on cones is called photopic, and the spectral response is called the photopic curve.

Cones exist in three types, each sensitive to a different part of the spectrum: red, green, and blue. However, their spectral response is not strictly red, green, and blue; this is just what they are called. But what is red? Green? Or blue? This limitation to three spectral bands produces a phenomenon called metamerism, whereby a pair of spectrally different specimens match under one illuminant but not under other illuminants. Because of the limited spectral detectors available in the human eye due to a color space limitation of the human eye, metamerism can occur. Some animal species have more-advanced multispectral-imaging systems, similar to specialized image sensors.

Metamerism can be explained as follows. Light made of two wavelengths falls on a detector (here, an image sensor) and is detected partly by each of the two color channels. For example, the first wavelength generates a signal of 100 DN on the green channel and a signal of 25 DN on the blue channel, and the second wavelength generates a signal of 50 DN on the green channel. In total, the signal is 100-DN green and 75-DN blue. Now imagine that a different light source made of two other wavelengths but the spectral response of the sensor is such that the first wavelength gives a green value of 50 DN, and the second wavelength gives a green value of 50 DN and a blue value of 75 DN, we also have the same total signal of 100 and 75. The example shows

that the resulting signal is the same for two different stimuli (combination of two wavelengths), and therefore the colors cannot be distinguished from each other. Adding more color channels could allow one to separate the two sets of wavelengths and reduce the risk of metamerism. This technique is used in multispectral and hyperspectral sensors.

The human eye has greater sensitivity to relative luminance levels than to absolute levels and has a nonlinear relationship between luminance and subjective brightness perception (demonstrated on incremental thresholds by the Weber–Fechner logarithmic law). The visibility threshold often increases in the vicinity of edges; this effect is sometimes called sensitivity masking.

2.9.2 Dynamic range of the human eye

It is often stated that the human vision system has a large dynamic range of around 120 dB. However, there is confusion between the intrascene dynamic range and the maximum dynamic range. In the presence of bright light, the eye has difficulty seeing details in the shadows and its dynamic range is about 40 dB. In the dark, the eyes need time to adapt before they can see reasonably well. There is a large ratio between the bright objects that the eyes can see in bright sun and the dark scenes that it can see in the dark after adaptation, but it won't be able to see both levels in the same scene.

The retina has a contrast ratio of around 100:1 (40 dB, maximum intrascene dynamic range*). Initial dark adaptation can take place in about four seconds; complete dark adaptation caused by the regeneration of rhodospin in rods takes place in about 30 minutes but can be affected by tobacco or alcohol. This process allows the human eye to reach a total dynamic range of about 1000000:1 (120 dB, maximum reachable dynamic range across multiple scenes). This adaptation to light is independent of the adaptation of the pupil.

2.9.3 Noise perception

In a still image, the random noise becomes FPN because it is just one sample of the temporal noise. For video, random noise is usually averaged by human vision and is less critical. Correlated noises (noises that form a regular repeated spatial or temporal structure) are usually more easily noticed than purely random noises. Random temporal noise is assumed invisible if the SNR is higher than 30 dB.

*Other phenomena, adaptations to the scene, and brain processing can improve this value.

Chapter 3
Hardware Methods to Extend the Dynamic Range

High dynamic range images can be generated by several imaging system architectures. This chapter focuses on sensor design and does not cover scene-dependent optical veiling glare or other optical aspects. The methods in this chapter use a special pixel design or a special pixel control, said to belong to the "hardware methods" category. Several of the described methods have been patented.

The other method uses a standard sensor (linear or nonlinear) or camera and processes the data to generate HDR images. This method, referred to here as the "software method," works with any type of sensor but is usually used with S-curve or linear sensors (see Chapter 4 for more details).

Several new sensors use a hybrid method that enables two exposures simultaneously, using two sets of pixels or combining overlapped exposures. Such approaches are very similar to the software method and will also be addressed in Chapter 4.

3.1 Introduction: Integrating Linear Pixels

Before introducing various HDR pixels, the response, noise, SNR, and DR of a traditional linear pixel will be addressed. The linear pixels considered here are said to be active and integrating. Active means that transistors are used for signal amplification or to perform other active tasks in the pixel. Integrating means that the pixel output is an integral function of the photons falling onto the pixel during the exposure time. These photons are converted to electrons (i.e., a current) and then to a voltage by means of a capacitor. This type of pixel is also said to have CVC (current-to-voltage conversion).

3.1.1 Rolling-shutter pixel architecture

In rolling-shutter sensors, not all pixels are exposed at the same time, although they have the same exposure time. Because they have no storage element,

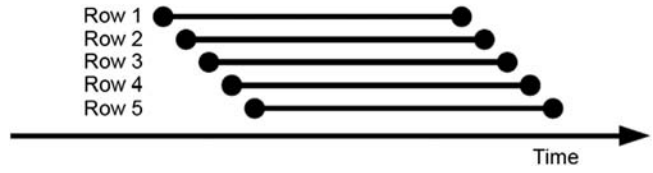

Figure 3.1 Line exposure window for five consecutive lines of a rolling-shutter sensor.

these pixels read the output value as soon as the exposure is stopped. In order to maintain the same exposure time for all pixels, they must start exposure at different moments (Fig. 3.1). This process is similar to the movement of the shutter lids in a photo camera with a focal plane shutter. One lid moves away to start exposure and then the other one covers the sensor to end exposure.

Rolling shutters have the least-complex pixels with typically only three transistors. They have low noise and high fill factor, but suffer from shutter artifacts in the presence of moving objects because each row of the sensor images the scene at a different moment. Figure 3.2 shows the imaging process of a horizontally moving bar by a rolling shutter with vertical (row-by-row) readout. Similar artifacts occur with a vertically moving bar or any general object movement; they also exist in the presence of flickering light (squared sine wave over time) or flash light. This issue can be solved by setting the exposure time equal to an integer amount of periods of the light source or by using special pixel-design techniques for LED flicker mitigation. Rolling shutters are thus not practical for use in some applications, but because the artifacts are not visible in most of the scenes, they are used in all mobile phone cameras, camcorders, photo cameras, and in many industrial and security applications. Figure 3.3 shows an image with the typical rolling-shutter artifact of object distortion, e.g., a plane's propeller.

Figure 3.4 shows a typical 3T rolling-shutter pixel and the corresponding charge integration model that will be used for the calculations. The capacitor used for integration is dominated by the capacitance of the photodiode itself. The total capacitance is the capacitance of the photodiode, the capacitance of the gate of the readout transistor and the capacitance of the implantation of

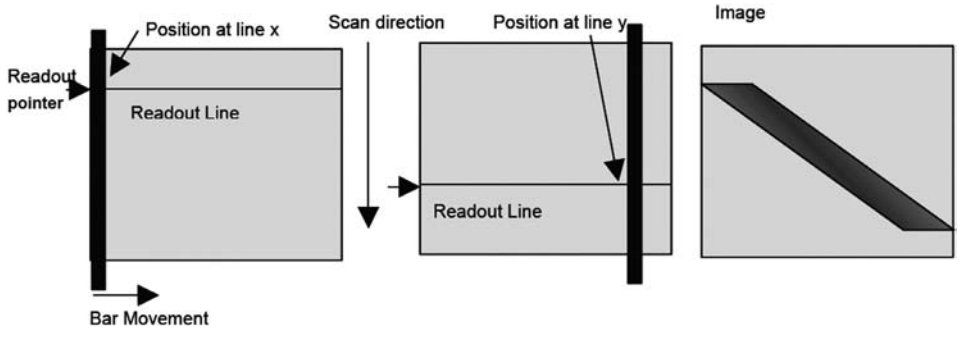

Figure 3.2 Rolling-shutter artifact when imaging a moving object (courtesy Martin Wäny).

Figure 3.3 Plane propeller distorted by a rolling-shutter artifact.

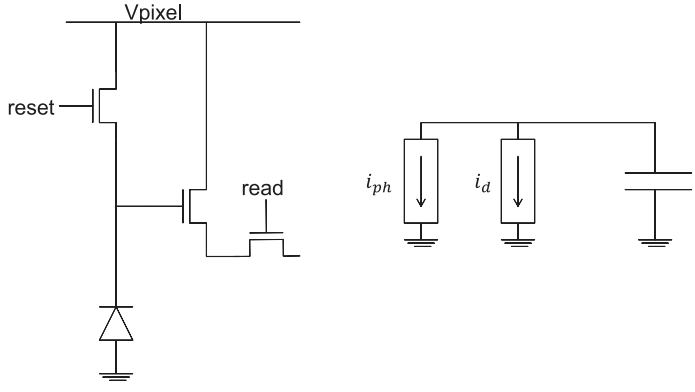

Figure 3.4 3T rolling-shutter pixel and corresponding integration model.

the reset transistor. A photodiode has capacitance because it has area and perimeter. This capacitance is assumed constant for most of the calculations, but it is, in general, a function of the charge that it stores. The depletion depth of the photodiode varies when stored charges vary, which changes the total

capacitance of the photodiode, because the capacitance of the depletion region is one component of the total capacitance of the photodiode. This affects the pixel's linearity. Other nonlinearities arise from the transistor used to amplify the pixel signal to drive the output bus. Nonlinearities are not considered in the SNR and DR study that follows.

Other designs than 3T exist for rolling shutters. For example, a 4T rolling-shutter is often used; this pixel has a transfer gate in series with the photodiode, avoiding the metal contact to the photodiode, which is a major cause of dark current. The charge transfer from the photodiode is then similar to a CCD transfer. Such a design uses a pinned photodiode. As the transfer gate acts like a CCD, it is common to represent the pixel schematic half with transistors and half with a silicon cross-section showing the structure of the photodiode and the transfer gate (Fig. 3.5). To understand charge transfer, a voltage barrier and charge diagram is used (Fig. 3.6). In such a diagram, the flow of electrons follows the waterfall analogy.

The 4T pixel can also be used in a transistor-sharing fashion (Fig. 3.7), a common method to reduce the pixel area and share one or more transistors between pixels. The timing and behavior of the pixel is changed, but its functionality remains similar. As the 4T shares the reset and readout circuit with its neighbors, a total of 4 pixels use 7 transistors, and each pixel is said to be a 1.75 T.

3.1.2 Global-shutter-pixel architecture

Global-shutter pixels have a storage mechanism inside the pixel so that the exposure can be started and stopped at any time, independently of the time of readout. These pixels are used in the presence of fast-moving objects or in

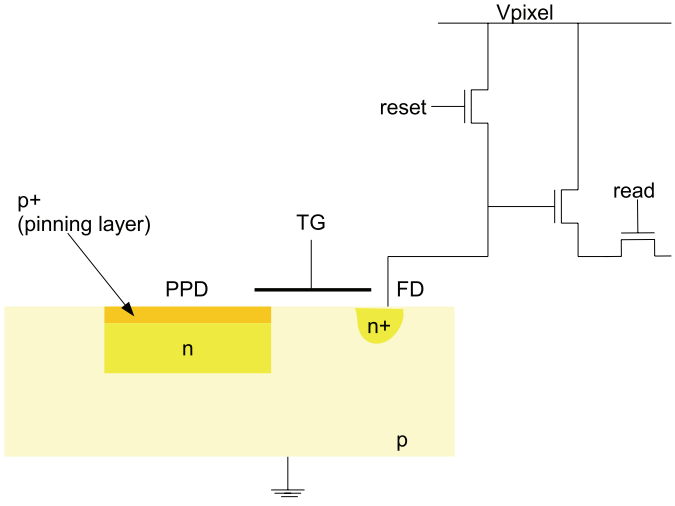

Figure 3.5 4T rolling-shutter pixel.

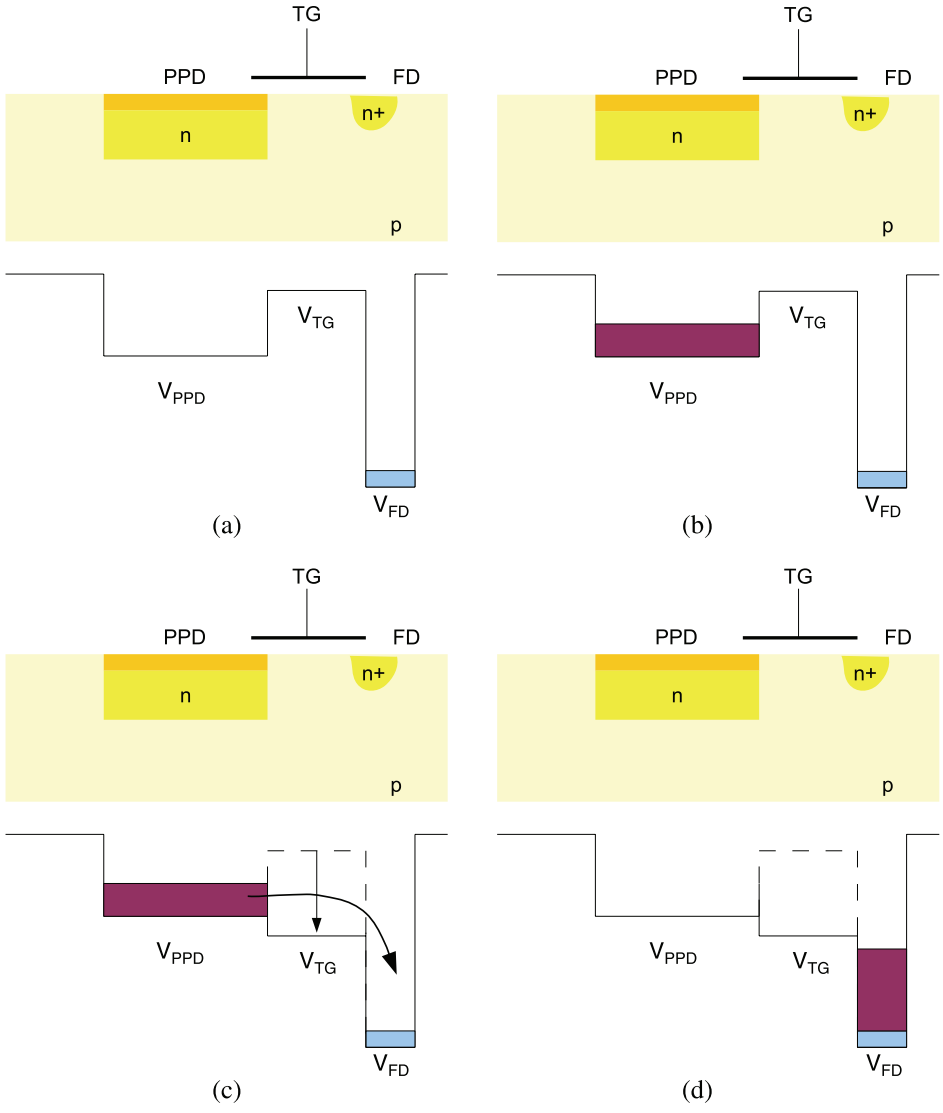

Figure 3.6 Voltage barriers and transfer of charges out of the pinned photodiode: (a) before exposure, (b) at the end of exposure, (c) barrier lowered, and (d) at the end of transfer.

high-speed imagers; they usually suffer from a narrower signal swing, lower fill factor, and additional noise sources and nonlinearities, but have the great advantage of fewer motion artifacts.

There are several global-shutter pixels with various transistor counts. Figure 3.8 shows an example of a 5T pixel based on the 4T pixel and pinned photodiode previously discussed. There are also pixels with six (6T), seven (7T), eight (8T), or even more transistors, depending on the performance and the features embedded into the pixel.

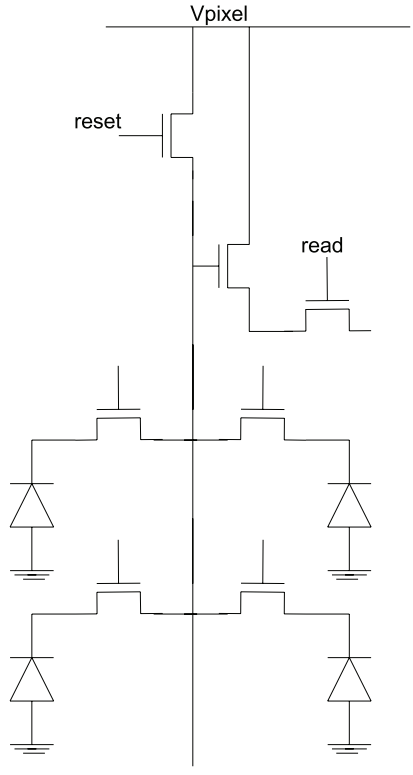

Figure 3.7 1.75T pixel quad that shares a single reset transistor.

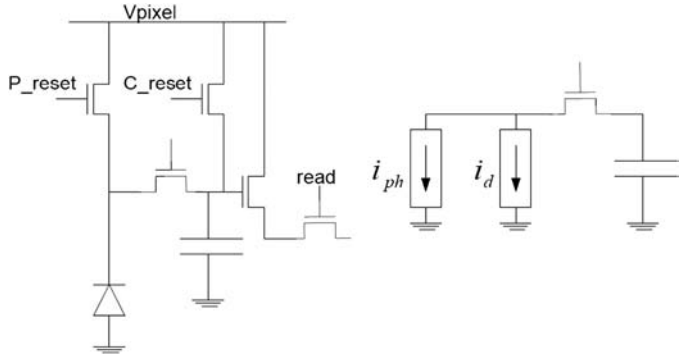

Figure 3.8 Global-shutter pixel and corresponding integration model.

Global shutter pixels are usually not compatible with correlated double sampling and therefore also exhibit a higher reset noise. One patented pixel with eight transistors has an in-pixel CDS circuit and offers real CDS operation against a slightly less-optimal pixel layout.

3.1.3 SNR and dynamic range study

The Yang & El Gamal method [64] is used here to calculate SNR and DR based on the integration models of the rolling-shutter and the global-shutter pixels. The calculations made here are valid for both types of pixels even though the order of magnitudes for the noises and saturation levels are not the same.

The calculations assume that the charges are integrated into the capacitor, but it is known that the capacitors are actually discharged. This simplification avoids the need to carry the reset level and the charge variations throughout the calculations.

In order to calculate $DR = i_{\max} = i_{\min}$, we need to find the expressions of $i_{\max}$ and $i_{\min}$ and the response curve f.

The signal and noise model is illustrated in Fig. 3.9. The average value of the photocurrent is denoted as by i_{ph}; photon flux is considered constant so that it can be decomposed into an average value and its shot noise. Photocurrent is the useful signal. The average value of the dark current is denoted as i_d; dark current is decomposed into its average value and its shot noise. Shot noise of the photocurrent and shot noise of dark current are grouped into a single current source $I_s(t)$, modeled as a white noise with a power spectral density of $q(i_{ph}+i_d)$. The total current is $I(t) = I_s(t) + i_{ph} + i_d$. Integrating the current during t_{int} builds up a total charge of

$$f = \min\left\{\int_0^{t_{int}} I(t)dt, q_{\max}\right\}, \tag{3.1}$$

where $q_{\max}$ is called the full well capacity or the saturation capacity, an upper limit to the pixel response (the charge that corresponds to saturation). Because the full well capacity is related to the saturation point, it is lower than the maximum charge that can be stored in the pixel. In this case of a linear pixel, f is a linear function until saturation. For nonlinear pixels, this function will model the pixel's nonlinearity; $q_{\max}$ is the maximum charge that the capacitor can store. The resulting voltage is then amplified, and an additional voltage source Q_r is added to represent the readout noise and the quantization noise. It has zero mean and a variance (or average power) of σ_r^2. Fixed pattern noise, DSNU, and PRNU are neglected in this discussion.

In order to calculate the dynamic range, one needs the maximum non-saturating signal and the minimum detectable signal. The maximum non-saturating signal is given by

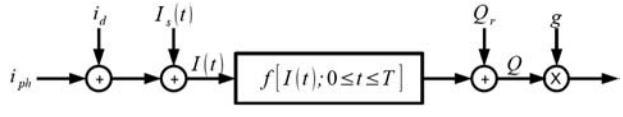

Figure 3.9 Simple pixel model.

$$i_{\max} = \frac{q_{\max}}{t_{int}} - i_d. \tag{3.2}$$

The minimum detectable signal $i_{\min}$ is equal to the standard deviation of the input-referred noise at the minimum detectable level (almost no signal). A new random variable will represent the sum of all noise sources referred to the output (what is measurable); it has zero mean and a variance of $\sigma_N^2 = q(i_{ph} + i_d)t_{int} + \sigma_r^2$.

When all noise is assumed to be present at the output, f becomes $f_0(i) = \min\{it_{int}, q_{\max}\}$, where $i = i_{ph} + i_d$ is the total current (see Figs. 3.10 and 3.11).

Assuming that the noise is small compared to the signal, the input equivalent noise N_i can be derived from the output referred noise N using

$$f_0(i + N_{in}) \approx f_0(i) + N_{in} f_0'(i) \Longleftrightarrow N_{in} = \frac{Q - f_0(i)}{f_0'(i)}. \tag{3.3}$$

It follows that

$$\sigma_{N_{in}} = \frac{\sigma_Q}{f_0'(i)} = \frac{\sqrt{qi_d t_{int} + \sigma_r^2}}{t_{int}}, \tag{3.4}$$

and the sensor's dynamic range is

$$DR = \frac{i_{\max}}{i_{\min}} = \frac{q_{\max} - i_d t_{int}}{\sqrt{qi_d t_{int} + \sigma_r^2}}. \tag{3.5}$$

The SNR is given by

$$SNR(i_{ph}) = \frac{i_{ph} t_{int}}{\sqrt{q(i_{ph} + i_d)t_{int} + \sigma_r^2}} \tag{3.6}$$

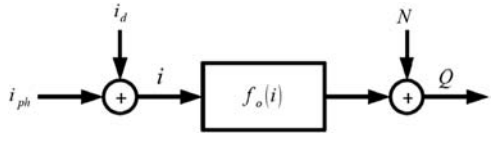

Figure 3.10 Simple pixel model with all noise assumed to be referred to the output.

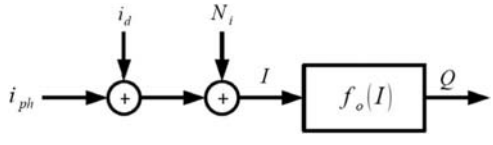

Figure 3.11 Simple pixel model with all noise assumed to be referred to the input.

up to the saturation point at which SNR goes progressively to infinity as the pixel values clips. This formula is very close to Eq. (2.8), as both approaches consider a linear pixel; however, this case only introduces actual design components instead of externally measurable parameters and neglects fixed pattern noise and photo response non-uniformity. Because SNR is a monotonic increasing curve, sensors with a larger DR also have a larger maximum SNR. This is not necessarily the case for sensors with dynamic range enhancement (XDR). The fact that the SNR curve will go to infinity as a significant number of pixels clip at saturation also explains why the saturation capacity is measured shortly before total saturation, in a part of the sensor's operational envelope where the SNR curve does not start to go to infinity yet.

Figures 3.12–3.16 show the effect of all of the parameters on SNR and DR for realistic sensor parameters. The plots are generated from simulations of the simple models introduced in this section. In Fig. 3.12, the readout noise is varied (three values of readout noise yielding to tree curves in each plot). Readout noise mostly affects the sensor close to dark, and it reduces SNR and DR. In Fig. 3.13, it is the dark current variation that is simulated. The plots in Fig. 3.14 are generated by simulating a variation of full well size. The plots in Figs. 3.15 and 3.16 analyze SNR versus a variation of photocurrent or exposure time.

The response curve is of the form $y = ax + b$, its incremental gain is $g = a$. The information transfer function is $I = a \cdot DRG(x)$, which is a as soon as the photocurrent is above $i_{\min}$ and 0 before. The information transfer function is constant within the dynamic range of the device.

3.2 Multilinear Pixels

3.2.1 Principle

Multilinear pixels are linear in nature, but their control waveforms are modified during exposure time so that they show nonlinear behavior. Most of the difficulties of the nonlinear pixels, as well as the controllability of the pixel response, can be solved, but at the expense of more circuitry and possibly additional artifacts.

The multiple-segment method (also called piecewise linear response or multiple slopes) is a way to approach a logarithmic response with an integrating device. The logarithmic response is approximated by a piecewise linear curve with usually three segments (Fig. 3.17), but more segments are often required for very high dynamic range. The method is widely used, and various implementations exist, all of which yield a similar controllability and a similar response curve. The multiple-segment method has several advantages compared to logarithmic response pixels: global shutters are possible; there is very limited image lag; the pixel is programmable in dynamic range

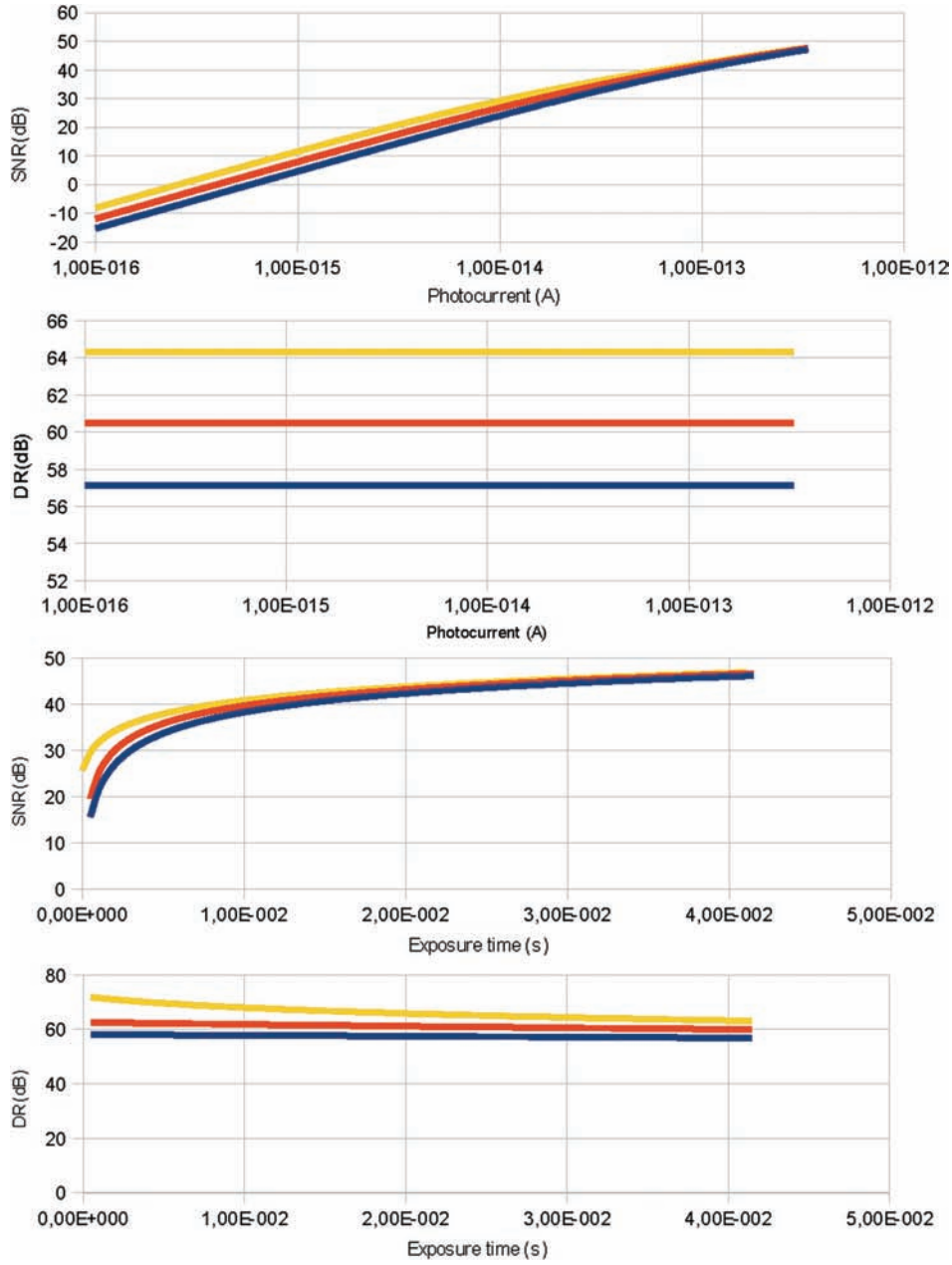

Figure 3.12 SNR and DR of a linear pixel vs. readout noise.

and response curve shape, and there is a linear response within each segment of the curve.

To understand how the various methods work, it is necessary to explain in further detail how a linear pixel works. Typically, a storage element (i.e., a

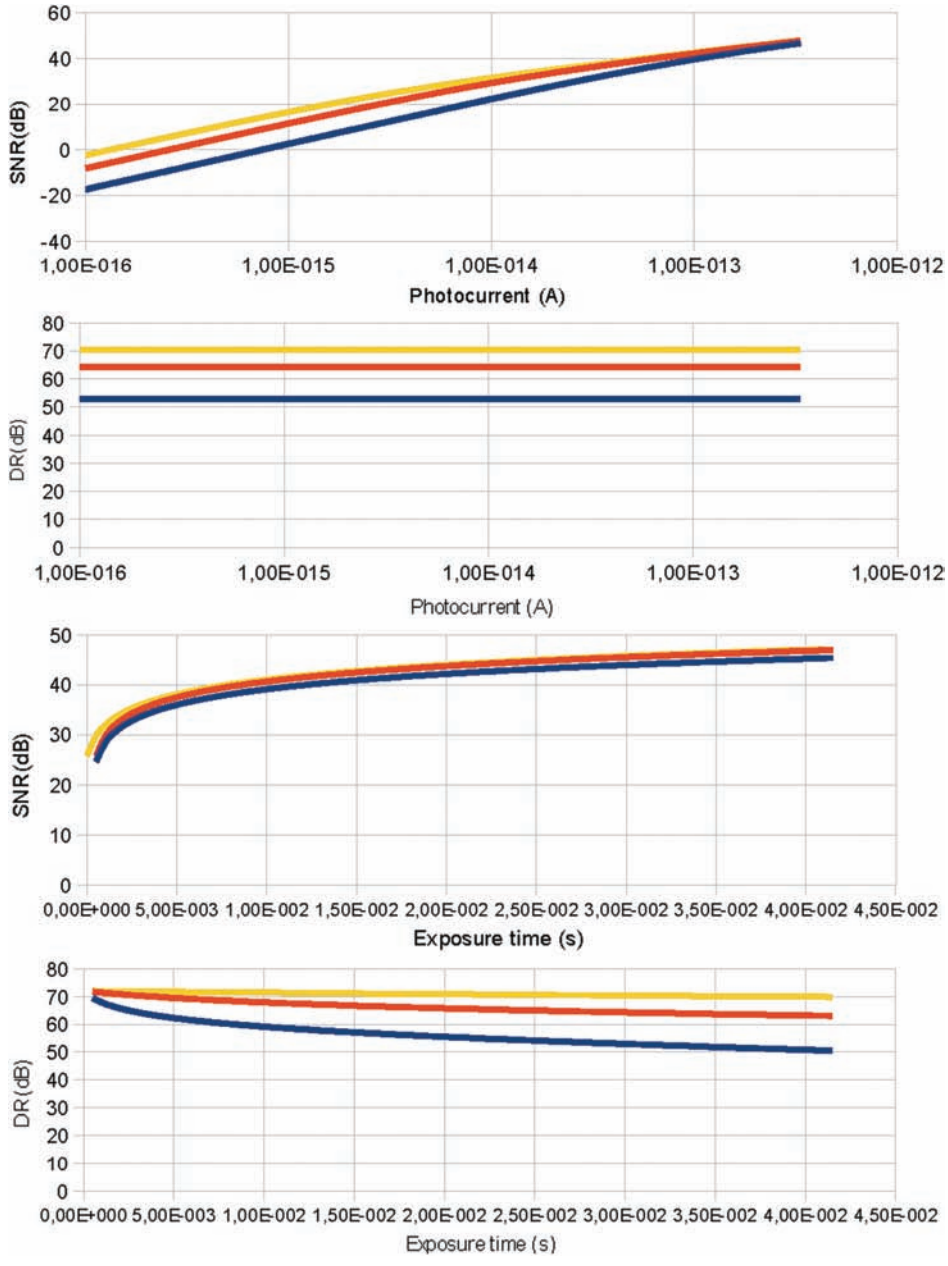

Figure 3.13 SNR and DR of a linear pixel vs. dark current.

capacitor), which can be either a real component or just the parasitic capacitance of the photodiode and transistors, is charged to a known level by a so-called reset circuit in a reset phase. The storage element is then discharged by the photocurrent generated by the photodiode during the integration

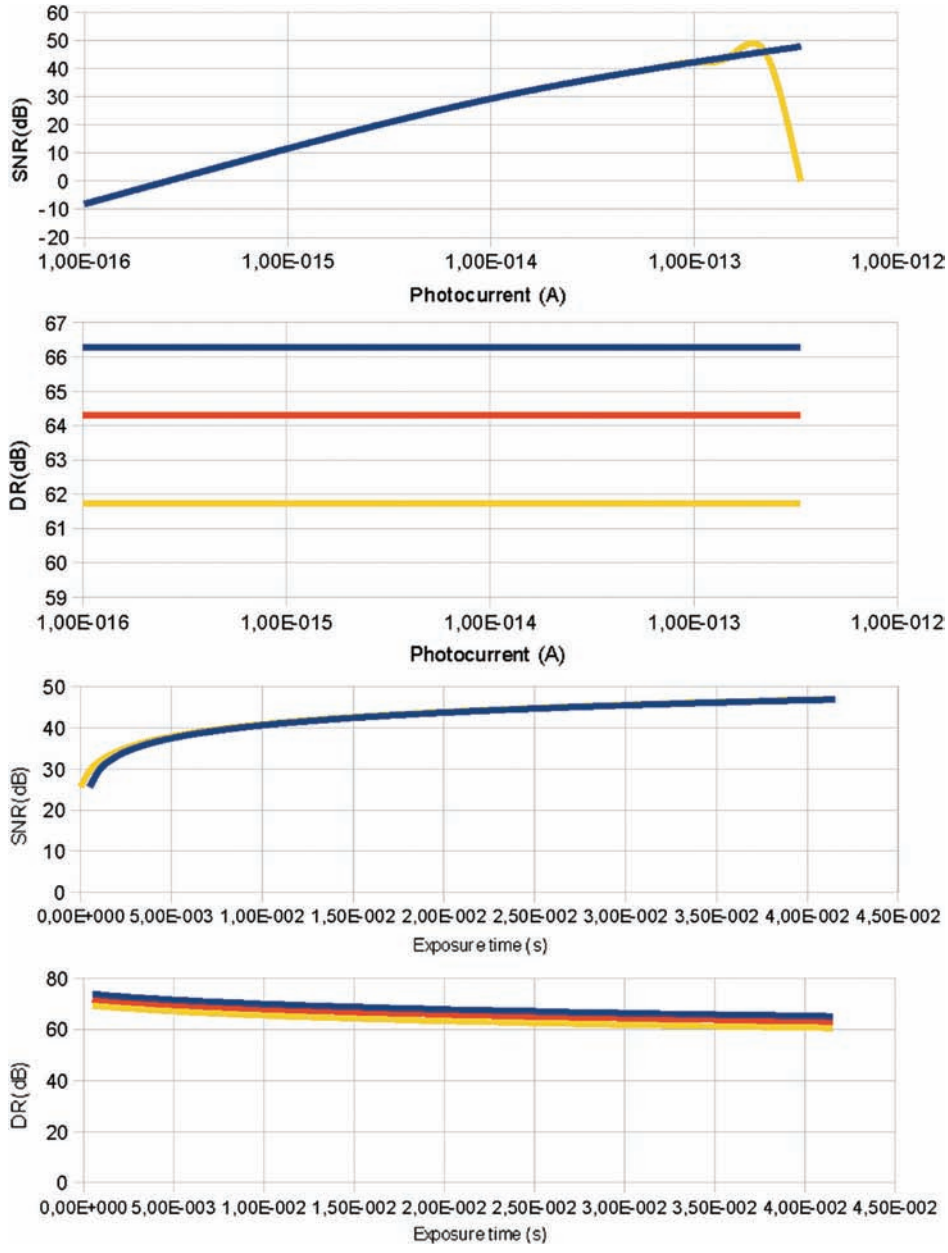

Figure 3.14 SNR and DR of a linear pixel over full well size; the drop in SNR for the yellow curve corresponds to the saturation. It is a calculation artifact—SNR should go to infinity if saturation is properly simulated.

period. The saturation level corresponds to the discharge of the storage element below the lowest readable level of the pixel (or to this level if anti-blooming and anti-corona circuits are present). The relation between the output, integration time, and photocurrent is given by

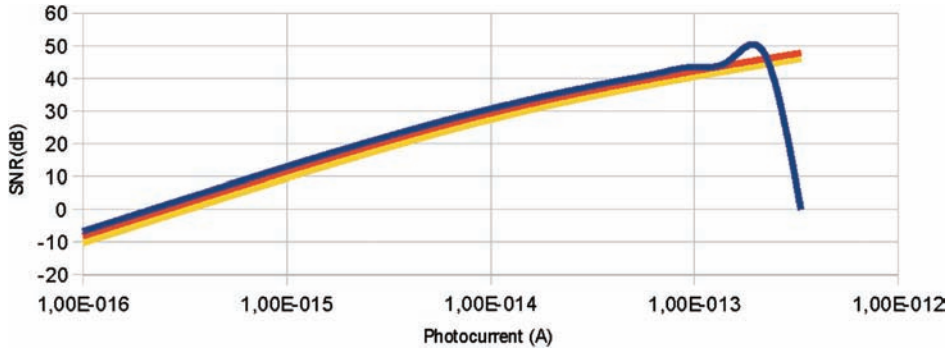

Figure 3.15 SNR of a linear pixel over photocurrent (fixed exposure time, variable photocurrent); the drop in SNR for the blue curve corresponds to the saturation. It is a calculation artifact—SNR should go to infinity if saturation is properly simulated.

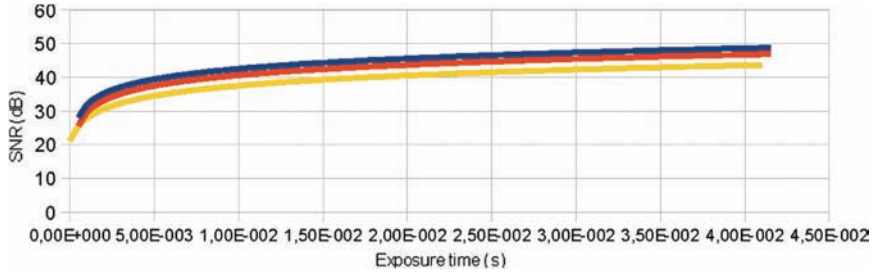

Figure 3.16 SNR of a linear pixel over exposure time (fixed photocurrent, variable exposure time).

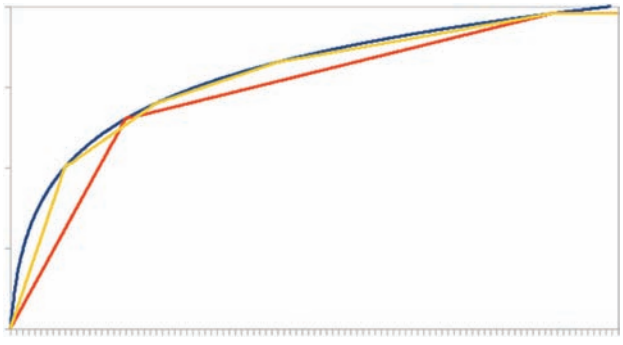

Figure 3.17 Approximation of a logarithmic curve by a two-segment curve and a four-segment curve. The last segments of the curves are horizontal due to saturation.

$$V_{(k,l)} = V_{(k,l),reset} - \frac{1}{C}\int_0^{t_{int}} [i_{photo,(k,l)}(t) + i_{d,(k,l)}(t)]dt \tag{3.7}$$

for pixel (k, l). If the parasitic (dark) current is neglected, the relation becomes

$$V_{(k,l)} = V_{(k,l),reset} - \frac{1}{C}\int_0^{t_{int}} [i_{photo,(k,l)}(t) + i_{d,(k,l)}(t)]dt$$
$$\approx V_{(k,l),reset} - \frac{1}{C}\int_0^{t_{int}} i_{photo,(k,l)}dt = V_{(k,l),reset} - \frac{\bar{i}_{photo,(k,l)} \cdot t_{int}}{C}, \quad (3.8)$$

and the discharge of the storage element is almost linear. The photocurrent in the last expression has to be understood as the average photocurrent over the integration time. The two parameters that can be modified by the user or by the digital circuit that drives the pixel are C and t_{int}. A variation of C yields to the well-sizing method discussed later. The variations of integration (exposure) time will be discussed first.

As seen in Fig. 3.18 and in Eq. (3.8), the slope of the response curve versus light intensity is proportional to the integration time (or exposure time). With a short integration time, the slope of the response curve is low, the saturation level is high (far to the right), and the sensitivity (ability to see dark scenes) is low (high light intensity required). With a long integration time, the slope of the response curve is high, the saturation level is low (not far to the right) and the sensitivity is high (low light intensity required). Creating a piecewise linear-response curve with a high slope for the low light intensities and a low slope for the high intensities will combine the advantages of a high sensitivity and a high saturation level—hence, a high dynamic range.

This two-segments approach usually gives a too-abrupt change in pixel behavior, and a multiple-segment approach (Fig. 3.19) is preferred for smoothness of the response curve and to avoid dynamic range gaps. The illumination is assumed constant during the exposure time. If this is not the case (e.g., in the case of pulsed illumination), then this approach can lead to image artifacts.

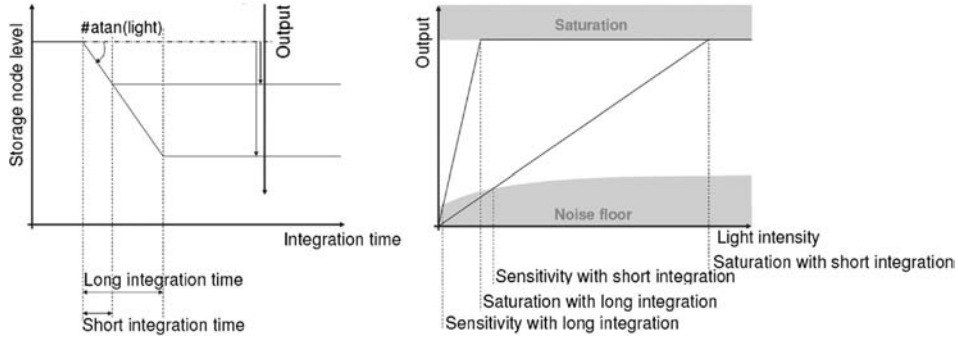

Figure 3.18 Linear pixel response according to the relation in Eq. (3.8) for two integration times. Left: evolution of the pixel voltage over time for a constant photon flux. The pixel voltage starts from the reset voltage and decreases. The pixel value is sampled at the end of exposure time and remains constant until readout occurs (global-shutter example). Right: response curves for the two same exposure times; for a longer exposure time, the difference between the reset voltage and the output voltage of the pixel is larger, and thus the response curve is higher.

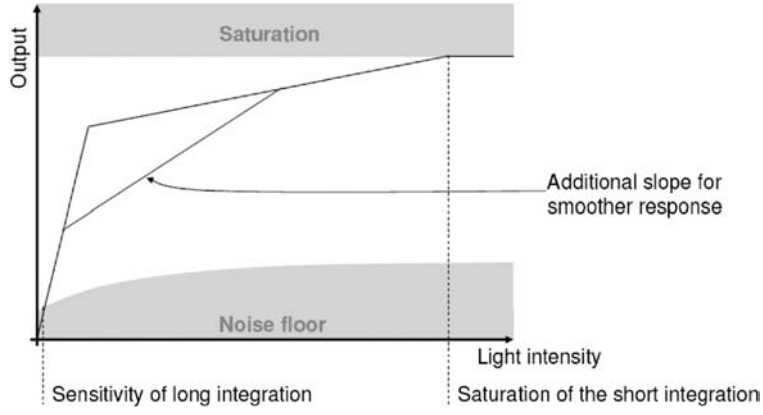

Figure 3.19 Multiple-segment pixel response.

3.2.2 How can multiple segments be realized practically?

The first method to realize multiple segments resets the pixels to a given level after an intermediate integration time $t_{segment}$. The reset only occurs for pixels that are already integrated below this level and are potential saturation candidates. This easy method generates the proper multiple-segment response but has the disadvantage of having a potentially non-uniform and noisy reset.

An improved method consists of clamping the pixel discharge at a certain level by compensating for the photocurrent until $t_{segment}$. The latter idea was actually invented first, back in 1983 (see Ref. 7). This text will refer to this method as pixel clamping—some authors call it a dynamic well capacity adjustment, but we will use this term for methods that scale the capacitor size and not the maximum capacitor charge.

These two methods only need some waveforms to be applied to some transistors of the pixel (Fig. 3.20); they do not require additional pixel circuitry. The difficulty lies in driving the large parasitic capacitance connected to the clamping voltage for high-resolution image sensors, especially when short integration times are used, i.e., a relatively large voltage

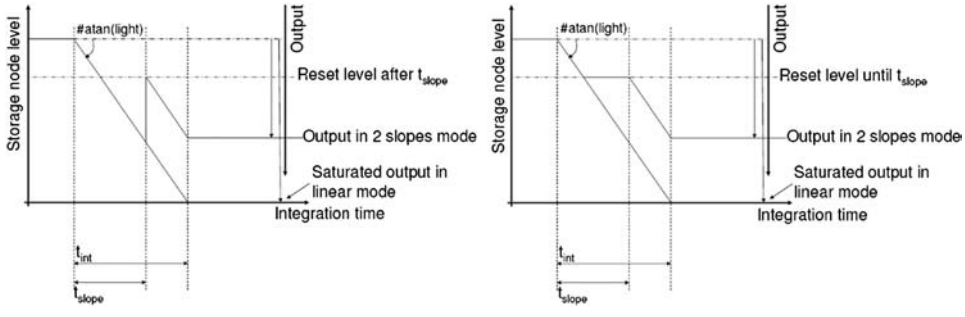

Figure 3.20 Pixel reset method (left) and pixel clamping method (right).

swing of the clamping voltage in a short time, which requires high-slew-rate DAC converters. This factor is probably the main reason why clamping after its first publication was not used until recently. It is important to notice that the pixel level, in case of reset or clamping, will always be less than for a normal linear exposure because it corresponds to a shorter exposure time. By these methods, the slope of the response curve can only be decreased.

Other methods based on partial charge transfer or integration capacitor leaking yield similar curves. One method consists of a leaking integration node with adjustable leaking threshold (lateral overflow). The generated curves are similar to the ones of the clamped pixel. Some of these techniques can cause artifacts when the photocurrent varies fast or in large limits over the integration time.

Another very important point is that the reset voltage or clamping voltage is directly related to the gray level at which the kneepoint has to be present in the response. One can imagine that the histogram can be used to set the kneepoint positions for a specific scene. Similar to statistics, the histogram is the graphical representation of the frequencies of the pixel values. It is a plot between 0 and $ADC_{\max}$ of the number of pixels in the image that have a given value. If a given level is present a lot in the image, it will correspond to a high peak in the histogram. Its area is the total amount of pixels present in the image or unity if the histogram is normalized. Kneepoints (Fig. 3.21) are usually chosen with equal spacing (half range for a single kneepoint, 1/3 and 2/3 for two kneepoints, 1/4, 2/4, 3/4 for three kneepoints, etc.).

The segment's slopes are directly proportional to the exposure times; it is thus possible to shape the response by tuning the exposure length for each part of the response. Due to the pixel principle of multiple segments, each segment of the curve has less slope than the previous one. A smooth, logarithmic-like response is usually generated by keeping a constant ratio between the slopes of two consecutive segments. For example, the exposure ratio between segment *i*

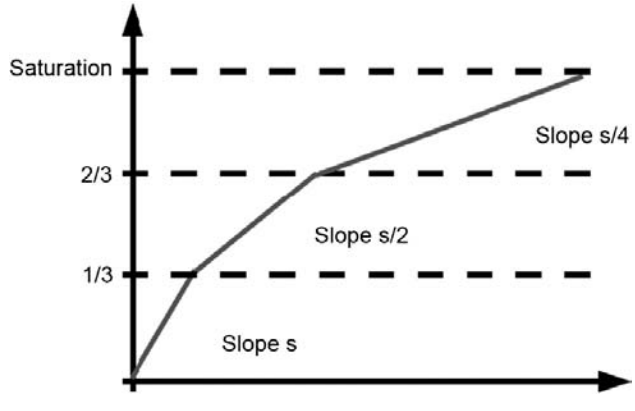

Figure 3.21 Typical settings for a multiple-segment curve.

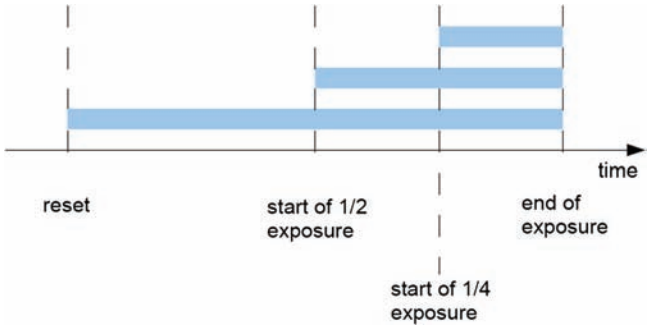

Figure 3.22 Overlapped exposures of long, medium, and short length in multi-segment response HDR mode. All exposures occur during the time window of the longest exposure.

and segment $i+1$ is kept constant at α. This is not the best approximation of a logarithmic curve, but it has proven to give good results with simple formulae.

It is important to stress that there is only one total exposure time and that all pixels are exposed at the same time. Timing changes affect the response curve but not the total exposure time or the start of exposure. This is depicted in Fig. 3.22.

3.2.3 Equations of the 3T pixel: reset and multiple-segment HDR exposure

Figure 3.23 depicts a simple model of the reset transistor, the reset signal, and a photodiode with a constant photocurrent in a 3T pixel. A similar calculation

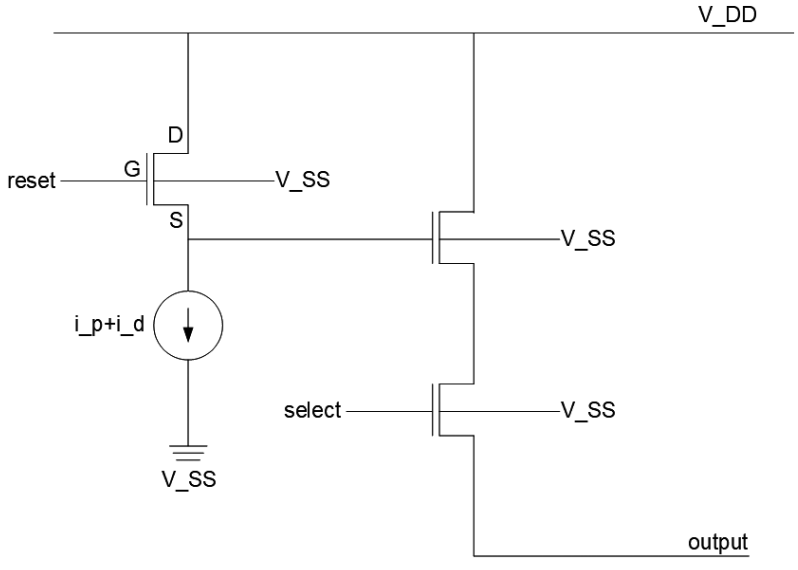

Figure 3.23 Design of a 3T pixel.

can be made for any pixel design. The 3T pixel is the easiest pixel design to explain. This section and the following section will derive some equations of the basic 3T pixel.

When the reset voltage is too low, the NMOS reset transistor is cut-off and the photodiode floats, being discharged by its own photocurrent. Too low a position in this case means $V_G < V_S + V_{Th}$, where V_{Th} is the transistor's threshold voltage. This happens when V_G is low or when V_S is high, i.e., if the photodiode has not been significantly discharged. Therefore, a lowly discharged photodiode will not be affected by a reset pulse if that reset pulse is not at least V_{Th} higher than the photodiode's voltage. Practically, V_G is not set to ground but remains slightly above V_{Th} so that the reset gate becomes an anti-blooming gate, operating as described in the next paragraph for extremely discharged photodiodes.

If $V_G > V_S + V_{Th}$ and $V_G < V_{DD}$, then a soft reset of the pixel occurs. An increase of V_G should increase the drain current, but as the drain current is forced to be the photocurrent, it results in an increase of V_S. The small signal voltage gain of a source follower is

$$A_V = \frac{g_m R}{g_m R + 1} \approx 1,$$

and so the source voltage will follow the gate voltage; this common drain amplifier is a source follower. It happens during the reset phase when V_G is high enough or during intermediate HDR resets because V_S is low enough due to photodiode discharge. As the bias of the common drain amplifier depends on the photocurrent and the dark current, the source will not exactly follow the gate. The actual equation is

$$V_s = V_G - V_{Th} - \sqrt{\frac{2\mathrm{L}(i_p + i_d)}{W\mu C_{oc}}}.$$

However, as the currents are at least two to three orders of magnitude larger than $k = \mu C_{ox}$ and W/L is not smaller than 1, the square root is at most 150 mV, which is much less than typical threshold voltages. This factor can produce nonlinearities or other imperfections in the pixel, and under very strong light it is one of the causes of black sun.

Practically, as the body voltage of the NMOS is not tied to the source because there is no room for a P-well isolation in pixels, the body voltage causes a back bias that affects the performance of the transistor. One of the effects is a reduction of the voltage gain to

$$A_V = \frac{g_m R}{(g_m + g_{mb})R + 1}.$$

Another effect is a variation of the threshold voltage that depends on the photodiode's voltage according to $V_T = V_{T,0} + \gamma(\sqrt{\phi + V_s} - \sqrt{\phi})$. This, together with the channel length modulation, affects the linearity of the circuit. Nonlinearity is a fundamental issue of CMOS APS sensors.

If $V_G > V_S + V_{Th}$ and $V_G > V_{DD} + V_{Th}$, then a hard reset of the pixel occurs, and the photodiode's voltage is forced to V_{DD}.

Figure 3.24 shows pixel operation for a low photocurrent, i.e., a pixel that will be exposed during the long exposure time and is not affected by the intermediate reset pulse.

Figure 3.25 shows the pixel operation for a high photocurrent and a higher intermediate reset pulse, i.e., a pixel that will be exposed during the short exposure time. The decision to use the short or the long exposure time is made within the reset transistor acting as reset and comparator to a photocurrent threshold, and therefore there is no additional circuitry in the pixel to perform this HDR operation.

A higher photocurrent or reset pulse voltage, or a shorter exposure for the short-exposure time (larger delay before the reset pulse), results in a pixel that operates on the shorter exposure time.

3.2.4 Equations of the 3T pixel: readout

Figure 3.26 shows the schematics of the readout bus circuitry of the 3T pixel of the previous section. The vertical bus is shared among all the pixels of a

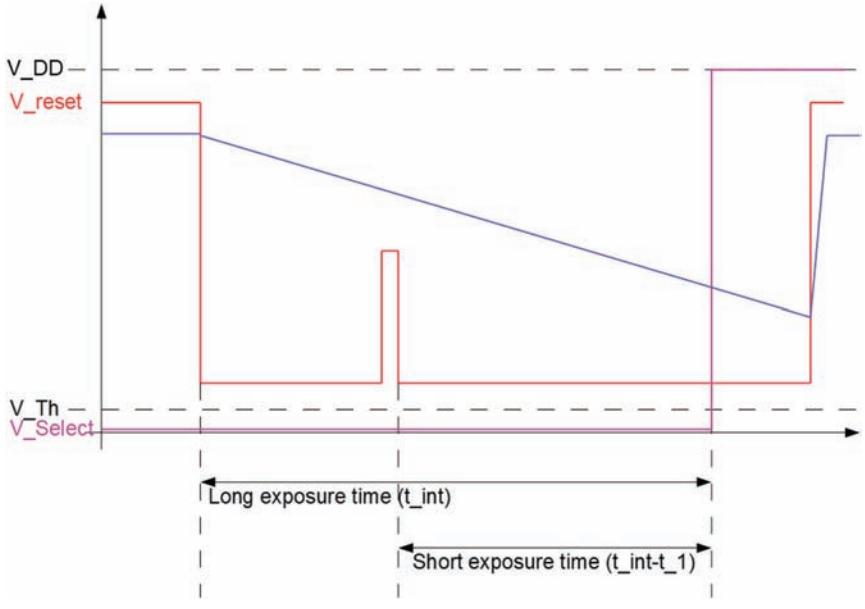

Figure 3.24 Conceptual waveforms of a 3T pixel with two-segment HDR in the case of a low photocurrent.

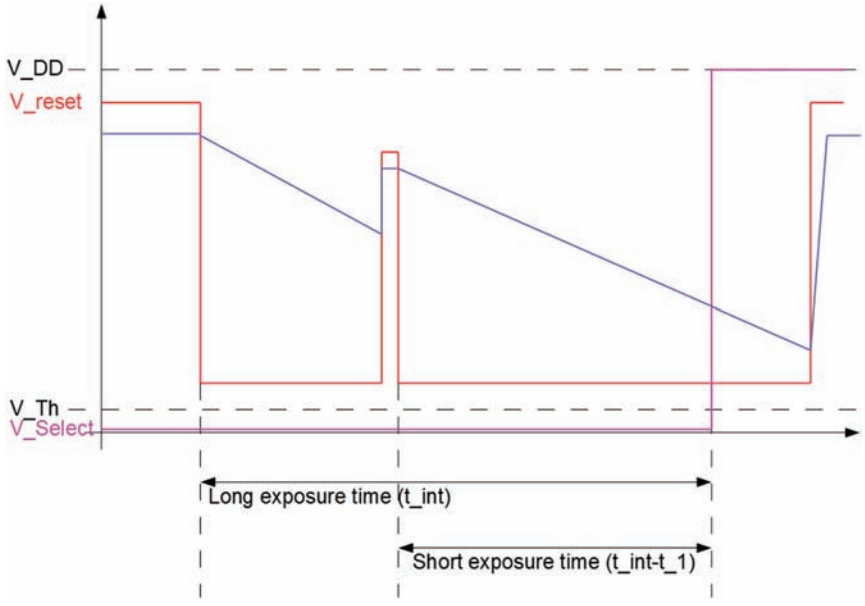

Figure 3.25 Conceptual waveforms of a 3T pixel with two-segment HDR in the case of a high photocurrent.

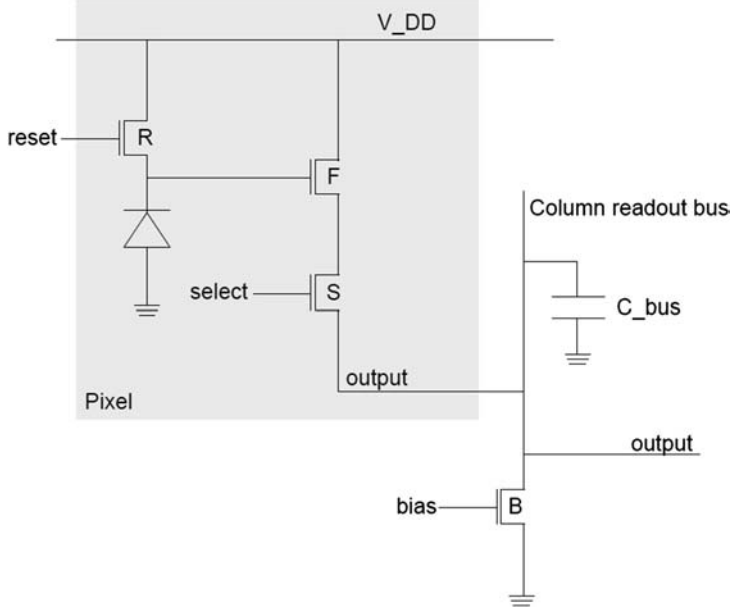

Figure 3.26 A 3T pixel and the shared readout bus with a constant current load.

column and has a constant current load to bias the pixel's readout devices. The bias current is constant and defined by the bias voltage. Because this bus is long, its capacitance cannot be neglected and must be taken into account when calculating the readout speed.

Assuming that the photodiode is reset to V_{reset} and that a charge Q has been accumulated during exposure, the voltage across the photodiode at the end of exposure is

$$V_{end} = V_{reset} - \frac{qQ}{C_{diode}},$$

where q is the charge unit, and C_{diode} is the virtual diode capacitance made up the various capacitances within the diodes as well as the parasitic capacitances to the surrounding devices. This capacitance is also a function of the charge stored into the photodiode and is another cause of nonlinearity.

The output voltage of the pixel is $V_{output} = V_{end} - V_{GS,F} - V_{DS,S}$, where $V_{GS,F}$ is the gate-to-source voltage loss across transistor F, and $V_{DS,S}$ is the channel loss across the select transistor. In steady state, the effect of the select transistor and the resistance of the readout bus can be neglected. Therefore, V_{output} is also the voltage at the foot of the column. The same calculation is valid for the sampled reset voltage.

After double sampling, the analog pixel value will be $\Delta V_{output} = V_{end} - V_{reset}$, and the effect of the F transistor cancels out. The double sampling stage is not represented in this figure. A concept of a double sampling stage is represented in Fig. 3.27. Much more advanced designs of double sampling circuits exist, typically switched capacitor subtractors. There is also often a current mirror stage between the readout column and the CDS. The programmable gain amplifier (PGA) can be an actual amplifier to match

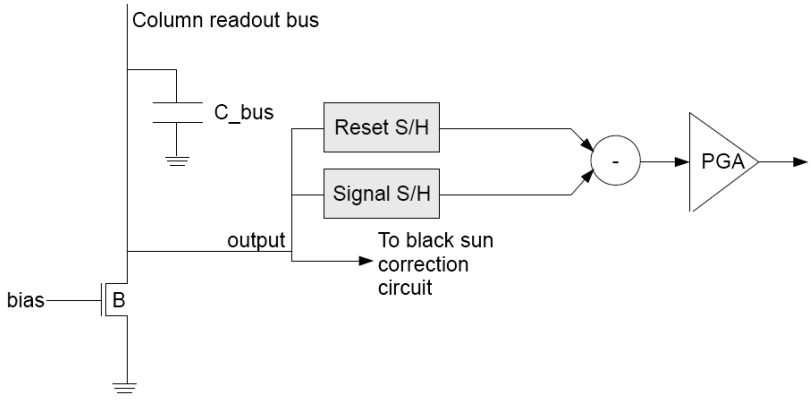

Figure 3.27 Simple conceptual example of a double sampling circuit.

the analog signal with the input of the ADC or a programmable range for the ADC to match its range with the analog signal.

The sensor's conversion gain is q/C_{diode} and is expressed in volts per electron or, more practically, microvolts per electron.

Another important parameter of the pixel and its readout circuit is the voltage swing. There is a minimum voltage below which the output bus cannot work. This minimum voltage is related to the subsequent double sampling circuit and also a minimum drain-source voltage required on the bias transistor B to keep it in saturation.

The maximum output voltage is reached when $Q = 0$ and approximately equals $V_{output} = V_{reset} - V_{GS,F}$. As in steady state, the bias current will flow from the supply, through the F transistor, and then through the B transistor. A simple two-transistor circuit can be used to calculate the gate-to-source voltage drop across the F transistor:

$$i_{bias} = \frac{1}{2}\mu C_{ox}\frac{W_F}{L_F}(V_{GS,F} - V_{Th,F})^2,$$

and therefore

$$V_{GS,F} = V_{Th,F} + \sqrt{\frac{2L_F}{\mu C_{ox} W_F} i_{bias}}.$$

The output swing is

$$V_{swing} = V_{reset} - V_{Th,F} - \sqrt{\frac{2L_F}{\mu C_{ox} W_F}} - V_{output,\min}.$$

It depends on the bias current. A higher bias current provides a lower voltage swing and a higher power consumption. In order to reduce power consumption, the bias voltage can be set to zero to cut off the bias transistor during exposure and idle states.

The condition to keep the bias transistor in saturation is $V_{output,\min} = v_{bias} - V_{Th,B}$. If this is the limiting condition, then the bias voltage will affect both the maximum voltage and the minimum voltage, but the conclusion that a higher bias provides a lower swing remains valid.

Practically, the full voltage swing across the diode cannot be realized because the diode needs to remain sufficiently reverse biased. This is also guaranteed by the minimum output voltage. The voltage swing of small-technology pixels is less than 1 V and can be much less in even smaller technologies. The technology nodes used today for image sensors are still usually large, typically 75, 90, 110, or 180 nm. If smaller technologies are used, the transistors in the analog part of the device, including the pixel, are usually not of minimum size. In particular, the source follower is made much wider in order to

reduce noise. A larger voltage swing helps to improve the dynamic range as the dynamic range starts in the pixel as voltage swing divided by dark noise.

Another important performance parameter is the readout speed of the circuit. When the S transistor is switched on, the current flowing through the F and S transistors must provide the bias current through the B transistor and the loading current of the distributed capacitance of the readout bus and of the sample and hold capacitor. An intermediate current mirror is usually added to limit the capacitance to the readout bus only. The following current equation is then used to calculate the readout speed:

$$i_{bias} + C_{bus}\frac{dV_{output}}{dt} = \frac{1}{2}\mu C_{ox}\frac{W_F}{L_F}(V_{GS,F} - V_{Th,F})^2.$$

Because $V_{GS,F} = V_{pixel} - V_{output}$, the equation can be rewritten as

$$i_{bias} + C_{bus}\frac{dV_{output}}{dt} = K(V_{pixel} - V_{output} - V_{Th,F})^2, \tag{3.9}$$

where

$$K = \frac{1}{2}\mu C_{ox}\frac{W_F}{L_F}.$$

Defining the row access time as the time required for the output signal to reach within half a bit of the ADC,

$$V_{output}(t_{row}) \geq V_{output}(\infty) - \frac{V_{swing}}{2^{k+1}}.$$

Equation (3.9) can then be solved by separating the variables to get

$$t_{row} = \frac{C_{bus}}{2\sqrt{2\mathrm{K}i_{bias}}}\ln\left(\frac{A}{B}\right),$$

where

$$A = \left(V_{output}(\infty) - V_{output}(t_{row}) + 2\sqrt{\frac{i_{bias}}{2\mathrm{K}}}\right)(V_{output}(\infty) - V_{output}(0))$$

and

$$B = (V_{output}(\infty) - V_{output}(t_{row}))\left(V_{output}(\infty) - V_{output}(0) + 2\sqrt{\frac{i_{bias}}{2\mathrm{K}}}\right).$$

This readout time is typically less than 1 microsecond but must be repeated for each row of the sensor and is therefore critical for the readout time.

3.2.5 Equations of the 3T pixel: power supply rejection ratio

Differential designs are strong against common-mode bias errors, bias variation, and noise. However, pixel designs are not differential and may therefore have a significant sensitivity to power supply noise. The power supply rejection ratio (PSRR) is the capability of the design to make the output signal insensitive to power supply variations and noise, estimated by the transfer function of the power supply to the output of an analog circuit.

The source follower of the 3T pixel is sensitive to such supply noise as the source voltage of that transistor depends on the bias voltage, which may vary with the supply noise. Also at the end of the pixel's response, the source voltage will also depend directly on the pixel's supply. Feed-forward design approaches of the bias circuit can be used to reduce the effect.

In a rolling shutter operating in hard reset, the power supply noise can also be transferred to the photodiode during reset.

In-pixel or in-column PSRR usually results in row and column temporal noise. It is also possible to correct this noise off-chip using deconvolution image-processing techniques. The spectrogram method of the EMVA1288 standard helps characterize such noises. Of course, PSRR can also affect other parts of the image sensor, but there is more freedom for optimum differential design outside the pixel array.

3.2.6 Charge injection and clock feedthrough

Turning off a transistor of a switched capacitor circuit introduces errors called charge injection and clock feedthrough. They limit the performance of CMOS APS pixels and other analog circuit, such as ADCs and DACs.

Charge injection is the distribution between the drain and the source of the charges present in the channel when the transistor is switched off. As the photodiode is connected to the source of the reset transistor in most pixels, about half of the charges present in the channel of the reset transistor will be released into the photodiode at the end of reset, causing a small voltage offset. This is more important on small photodiodes as the small capacitance produces a higher voltage offset.

Clock feedthrough is the capacitive coupling between the gate and the source that produces an offset on the capacitor connected to the source when the gate voltage changes. The combination of charge injection and clock feedthrough can reduce the dynamic range. Both errors are difficult to model or simulate, especially for smaller technologies.

3.2.7 Multiple-segment method based on well sizing

Because image generation is based on the integration of a photocurrent in a capacitor (using the general relation of the previous section), it is obvious that the slope of the pixel response can be adjusted by scaling the integration

capacitor. The larger the capacitor is, the smaller the responsivity. This technique has the advantage of providing an in-pixel gain that can increase the low-light response and provide a high saturation level. Although the response curve looks similar to the multiple-segment response, one should note that the well sizing method increases the sensitivity in the low-light parts of the scene where the multiple slopes method decreases the sensitivity in the high-light parts of the scene. The threshold level is usually programmable. Although this method increases the sensitivity, it also decreases the saturation level, and the dynamic range is only shifted to the low side.

The lateral overflow pixel is able to receive charges until the potential of the integration nodes reaches the level of a potential barrier or until the potential barrier is raised. This behavior allows one to limit the amount of charge that a particular step of integration can store and hence produce multiple segments. The additional charges are taken away from the pixel, which "scales" the capacitor and is then called a dynamic well adjustment [also called skimming (see Fig. 3.28)]. Although this method is easy in principle, just sizing the well over time does not work. Additional circuitry is required to make it work properly. The well sizing method can give for each segment higher or lower response than the corresponding normal linear exposure.

3.2.8 Dynamic compression

Figure 3.29 explains the dynamic compression of a large dynamic range by a triple-segment response. For example, a typical >100 dB of dynamic range scene can then be mapped into a 60-dB word (10 bits). The green response (lowest curve) is the best-matching reproduction of the scene, but the details of the first two peaks are not well reproduced (linear response with low gain and/or low integration time and/or low sensitivity). The red response (highest curve) gives a lot of details for the first peak, but the third peak and one half of the second peak are lost by saturation (linear response with higher gain and/or longer integration time and/or higher sensitivity). The blue curve (middle) is the response for a triple-segment response fitted to the scene. The first peak is reproduced as with the red line, the second peak is reproduced as with the

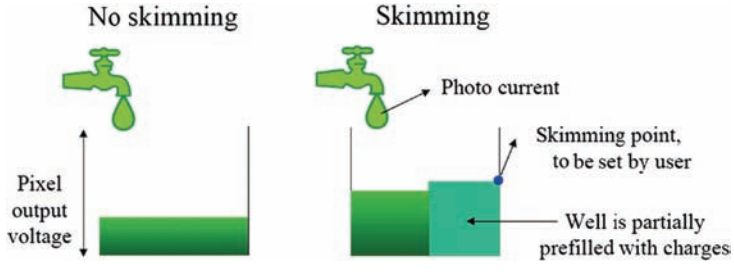

Figure 3.28 Water-bucket analogy to the skimming principle.

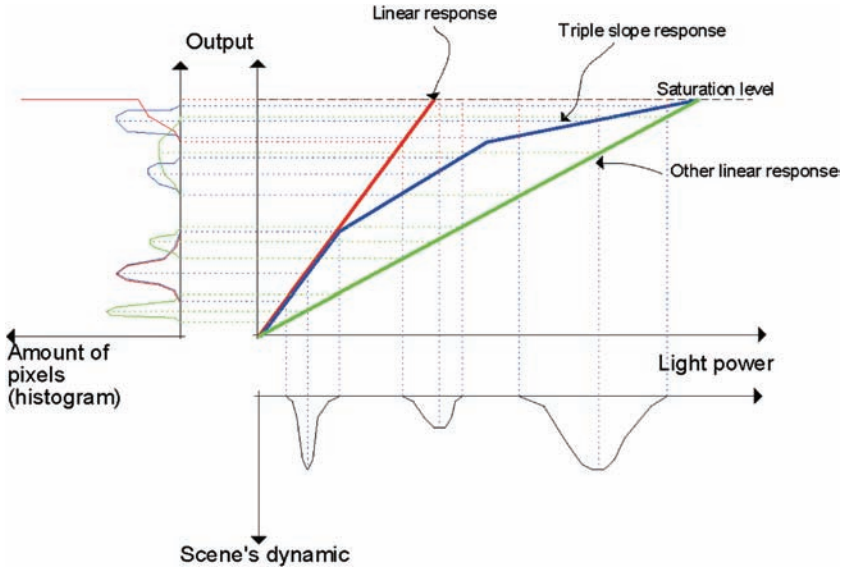

Figure 3.29 Histogram-based example of multiple-segment dynamic compression.

green curve (but with an offset), and the third peak is still present. The advantage of the triple-segment response in this example is to keep all of the details of the first peak without having saturation for the other parts of the image. The loss of information when compression occurs can be seen by the reduction of the third peak's width when going from the green curve to the blue one. The information gain can be seen in the modification of the saturation peak into a Gaussian when going from the red curve to the blue one. Together, there is no change or almost no change in the total amount of information encoded.

3.2.9 SNR and dynamic range study

Similar to the linear pixel of the previous section, the SNR and DR of the piecewise linear-response pixel will be computed based on clamping. A similar calculation can be made for the piecewise linear-response pixel based on charge reset by adding an additional noise component for the uncertainty of the reset level. This section shows where in the developments this extra noise source should be inserted, but the detailed developments are not discussed.

Let us begin by making the developments for a response curve with two segments; later, we will examine how to extend to multiple segments. Define the charge level at which the response switches from one responsivity to the other; with $q_{\max}$ being the full-well size, this kneepoint level is defined as $q_k = q_{\max}\theta$, with $0 < \theta < 1$. Using the same notations as previously, f can be defined as

$$f = \begin{cases} \int_0^{t_{int}} I(t)dt & \text{if } 0 \le i_{ph} < \frac{q_k}{t_1} - i_d \\ q_k + \int_{t_1}^{t_{int}} I(t)dt & \text{if } \frac{q_k}{t_1} - i_d \le i_{ph} < \frac{q_{\max}(1-\theta)}{t_{int}-t_1} - i_d \\ q_{\max} & \text{otherwise} \end{cases} \tag{3.10}$$

and σ_q as

$$\sigma_Q^2 = \begin{cases} q(i_{ph} + i_d)t_{int} + \sigma_r^2 & \text{if } 0 \le i_{ph} < \frac{q_k}{t_1} - i_d \\ q(i_{ph} + i_d)(t_{int} - t_1) + \sigma_r^2 & \text{if } \frac{q_k}{t_1} - i_d \le i_{ph} < \frac{q_{\max}(1-\theta)}{t_{int}-t_1} - i_d. \end{cases} \tag{3.11}$$

One can then calculate $f_0(i)$ and $f_0^{'}(i)$.

In this expression, σ_r also includes the kTC noise of the overflow gate. It may also include the reset noise in the case of a charge-reset-based pixel; in this case it cannot be considered constant, as it will be lower in the first segment of the curve than in the other due to a better reset noise after correlated double sampling. Correlated double sampling is only applicable to the initial reset:

$$SNR(i_{ph}) = \begin{cases} \frac{i_{ph}t_{int}}{\sqrt{q(i_{ph}+i_d)t_{int}+\sigma_r^2}} & \text{if } 0 \le i_{ph} < \frac{q_k}{t_1} - i_d \\ \frac{i_{ph}(t_{int}-t_1)}{\sqrt{q(i_{ph}+i_d)(t_{int}-t_1)+\sigma_r^2}} & \text{if } \frac{q_k}{t_1} - i_d \le i_{ph} < \frac{q_{\max}(1-\theta)}{t_{int}-t_1} - i_d. \end{cases} \tag{3.12}$$

At $i_1 = (qk/t_1) - i_d$, SNR(i_{ph}) is not continuous, and there is a dip of a factor of about $DIP_1 = \sqrt{1 - (t_1/t_{int})}$ (i.e., a small dip factor means a large dip).

Knowing σ_Q and $f_0^{'}$, σ_{Ni} can be calculated, and thus

$$DR = \frac{i_{\max}}{i_{\min}} = \frac{\frac{q_{\max}(1-\theta)}{t_{int}-t_1} - i_d}{i_{\min}} \approx \frac{1-\theta}{1-\frac{t_1}{t_{int}}} DR_{linear} = DRF \cdot DR_{linear}, \tag{3.13}$$

where

$$DRF = \frac{1-\theta}{1-\frac{t_1}{t_{int}}} \tag{3.14}$$

is called the dynamic range enhancement factor and is a fundamental factor for characterizing a response curve of a piecewise linear pixel.

As $DIP_1 = \sqrt{(1-\theta)/DRF}$, a large dynamic range enhancement means a small dip factor, i.e., a large dip. If this dip is too large, the SNR will decrease too much (below $SNR = 1$) and there will be a dynamic range gap (SNR hole). The dip and the dip factor are explained in Fig. 3.30. Here, the SNR is represented on a linear axis instead of the usual dB axis.

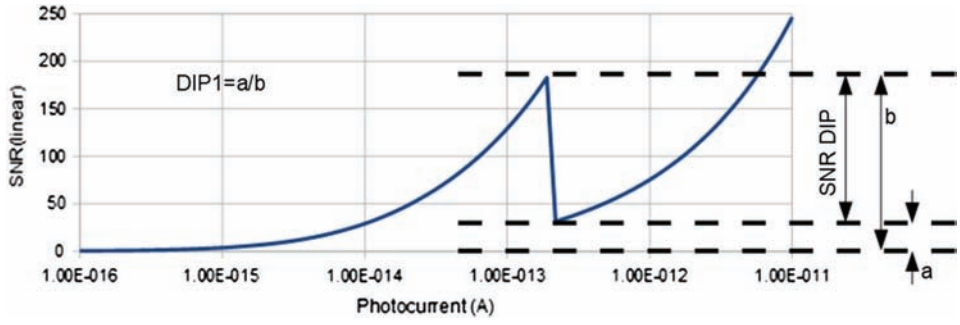

Figure 3.30 SNR dip and dip factor explained. The vertical axis is linear so that the relationships are valid and the horizontal axis is logathmic for better readability.

As σ_r may be different in the first and second segments of the curve, and i_d may not be neglectable, the dip factor also depends on these three noises, and the formula for DIP1 is significantly more complex.

Assuming a constant photodiode current during exposure, the response of the sensor is in the form $y = ax + b$ in the first segment and $y = a[(t_{int} - t_1)/t_{int}]$ $x + b'$ in the second segment. The information transfer is then proportional to a in the first segment and $a[(t_{int} - t_1)/t_{int}]$ in the second segment. There is then more information transferred in the first segment than in the second one.

To complete the information transfer analysis, the DRG presence function must be computed. This can be found by solving $SNR(i_{ph}) > 1$ for values of i_{ph}. The equation to solve is a simple second-order equation in i_{ph} after taking the square of both sides. In the first segment, $SNR(i_{ph}) > 1$ is realized if

$$i_{ph} > \frac{q}{2t_{int}} \quad 1 + \sqrt{1 + \frac{4i_d t_{int}}{q} + \frac{4\sigma_r^2}{q^2}}\Bigg). \tag{3.15}$$

In the second segment, the condition is realized if

$$i_{ph} > \frac{q}{2(t_{int} - t_1)} \left(1 + \sqrt{1 + \frac{4i_d(t_{int} - t_1)}{q} + \frac{4\sigma_r^2}{q^2}}\right), \tag{3.16}$$

which is always true if

$$\frac{q}{2(t_{int} - t_1)} \left[1 + \sqrt{1 + \frac{4i_d(t_{int} - t_1)}{q} + \frac{4\sigma_r^2}{q^2}}\right] < \frac{q_k}{t_1} - i_d. \tag{3.17}$$

This last equation is then the condition to have no gap.

These formulae can be extended to the multiple-segment response. In this case, with $1 \le i \le k$, $0 < \theta_i < \theta_{i+1} < 1$, and $0 < t_i < t_i + 1 < t_{int}$, it can be demonstrated that

$$DRF = \frac{\frac{q_{\max}(1-\theta_k)}{t_{int}-t_k} - i_d}{\frac{q_{\max}}{t_{int}} - i_d} \approx \frac{1-\theta_k}{1-\frac{t_k}{t_{int}}} \tag{3.18}$$

and that

$$SNR(i_{ph}) = \begin{cases} \frac{i_{ph}(t_{int}-t_0)}{\sqrt{q(i_{ph}+i_d)(t_{int}-t_0)+\sigma_r^2}} & \text{if } 0 \le i_{ph} < \frac{q_1}{t_1-t_0} - i_d \\ \frac{i_{ph}(t_{int}-t_1)}{\sqrt{q(i_{ph}+i_d)(t_{int}-t_1)+\sigma_r^2}} & \text{if } \frac{q_1-q_0}{t_1-t_0} - i_d \le i_{ph} < \frac{q_2-q_1}{t_2-t_1} - i_d \\ \cdots & \cdots \\ \frac{i_{ph}(t_{int}-t_{k-1})}{\sqrt{q(i_{ph}+i_d)(t_{int}-t_{k-1})+\sigma_r^2}} & \text{if } \frac{q_{k-1}-q_{k-2}}{t_{k-1}-t_{k-2}} - i_d \le i_{ph} < \frac{q_k-q_{k-1}}{t_k-t_{k-1}} - i_d \\ \frac{i_{ph}(t_{int}-t_k)}{\sqrt{q(i_{ph}+i_d)(t_{int}-t_k)+\sigma_r^2}} & \text{if } \frac{q_k-q_{k-1}}{t_k-t_{k-1}} - i_d \le i_{ph} < \frac{q_{\max}-q_k}{t_{int}-t_k} - i_d. \end{cases} \tag{3.19}$$

Two important remarks. First, the sum of the SNR dips is always greater than the DRF, and increasing the dynamic range will always cause dips. Second, the SNR at saturation is always lower than the SNR at saturation in the linear mode. These two remarks demonstrate that optimizing the sensor timings for the largest dynamic range will not give the highest SNR, and therefore the HDR settings must be used with care. Sometimes some saturation is acceptable if the SNR, i.e., the image quality, around the objects of highest interest improves.

3.3 Multiple Sampling

A linear pixel can be sampled several times during its exposure without destroying the inner charge. This process can be used to extend the dynamic range, as some pixels that would normally saturate can be sampled at least once before saturation. The darkest parts of the images are taken from the long exposure image (last sample), and the brightest parts of the image are taken from the short exposure image (first sample).

Figure 3.31 is the result of the simulation of a pixel according to the previous theoretical formulae for the following device and operating conditions:

- Saturation capacity: 80000 electrons
- Dark noise: 20 electrons
- Dark current at ambient temperature: 11 electrons/s
- Operating temperature: 80 degrees Celcius (junction)
- Dark current doubling temperature: 7 degrees
- Dark current at operating temperature: 62422 electrons/s, or 10 fA

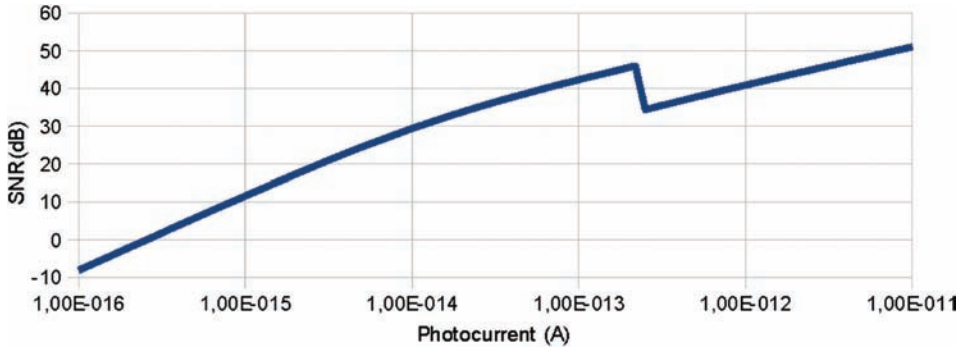

Figure 3.31 SNR plot of a two-segment piecewise response pixel.

- Exposure time: 30 ms (long exposure)
- Second exposure start delay: 28 ms (i.e., the short exposure is 2 ms)
- Second exposure start level: 50% (theta)

With this simulation, a dynamic range of 82 dB is reached with a peak SNR of a bit more than 46 dB before HDR saturation. The DRF is 7.5 and yields a virtual full-well size of 600*k* electrons regarding saturation. The part of the curve beyond 46 dB does not have a physical meaning as the pixel is then saturated.

Consider that the exposure times of two samples are t_{int} and t_{int}/a for $a > 1$. Similar to the previous two models, it can be calculated that

$$f = \begin{cases} \int_0^{t_{int}} I(t)dt & \text{if } 0 \le i_{ph} < \frac{q_{\max}}{t_{int}} - i_d \\ \int_0^{\frac{t_{int}}{a}} I(t)dt & \text{if } \frac{q_{\max}}{t_{int}} - i_d \le i_{ph} < \frac{aq_{\max}}{t_{int}} - i_d, \\ q_{\max} & \text{otherwise} \end{cases} \tag{3.20}$$

$$\sigma_Q^2 = \begin{cases} q(i_{ph} + i_d)t_{int} + \sigma_r^2 & \text{if } 0 \le i_{ph} < \frac{q_{\max}}{t_{int}} - i_d \\ q(i_{ph} + i_d)\frac{t_{int}}{a} + \sigma_r^2 & \text{if } \frac{q_{\max}}{t_{int}} - i_d \le i_{ph} < \frac{aq_{\max}}{t_{int}} - i_d, \end{cases} \tag{3.21}$$

$$SNR(i_{ph}) = \begin{cases} \frac{i_{ph}t_{int}}{\sqrt{q(i_{ph}+i_d)t_{int}+\sigma_r^2}} & \text{if } 0 \le i_{ph} < \frac{q_{\max}}{t_{int}} - i_d \\ \frac{i_{ph}\frac{t_{int}}{a}}{\sqrt{q(i_{ph}+i_d)\frac{t_{int}}{a}+\sigma_r^2}} & \text{if } \frac{q_{\max}}{t_{int}} - i_d \le i_{ph} < \frac{aq_{\max}}{t_{int}} - i_d, \end{cases} \tag{3.22}$$

$$DRF = \frac{\frac{aq_{\max}}{t_{int}} - i_d}{\frac{q_{\max}}{t_{int}} - i_d} \approx a. \tag{3.23}$$

As for the piecewise linear response (Figs. 3.32–3.35), dynamic range can be extended at the expense of SNR dips, but it can be shown that the dips are

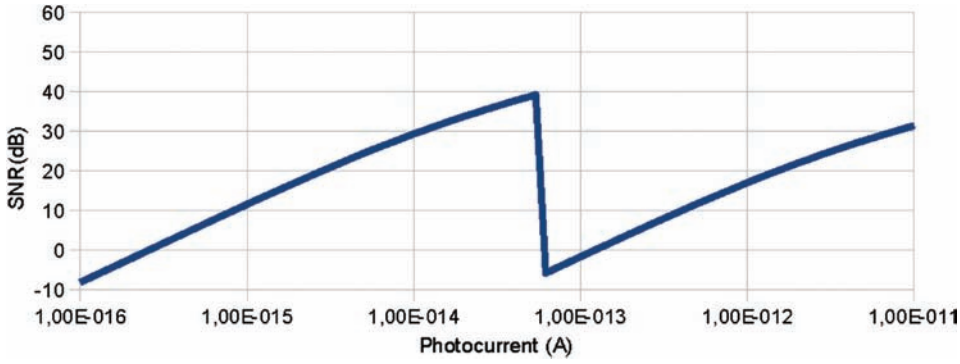

Figure 3.32 SNR plot of a two-segment piecewise response pixel with a dynamic range gap.

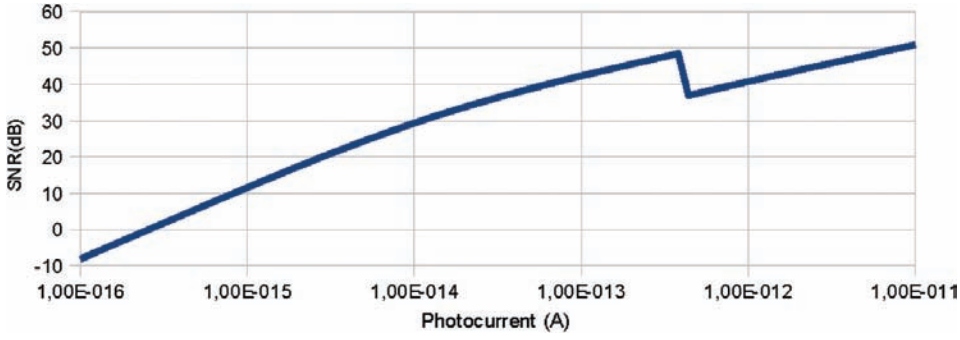

Figure 3.33 SNR plot of an imager response based on two samples of linear exposure. [There are actually three samples because one is used for the dark level by the (correlated) double sampling.]

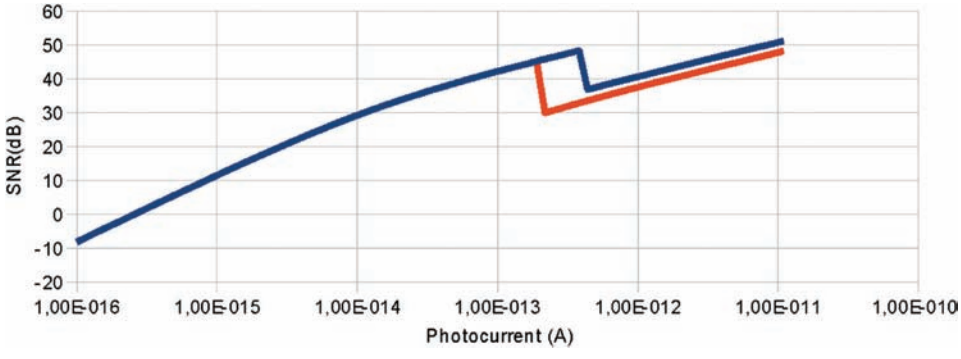

Figure 3.34 Comparison of the SNR plots for an imager based on two samples of linear exposure and the same sensor operated in piecewise linear response (PWLR) under the same conditions and parameters, both reaching a dynamic range of 88 dB. The lowest (orange) curve is the piecewise linear response and the upper (blue) curve is the two samples response.

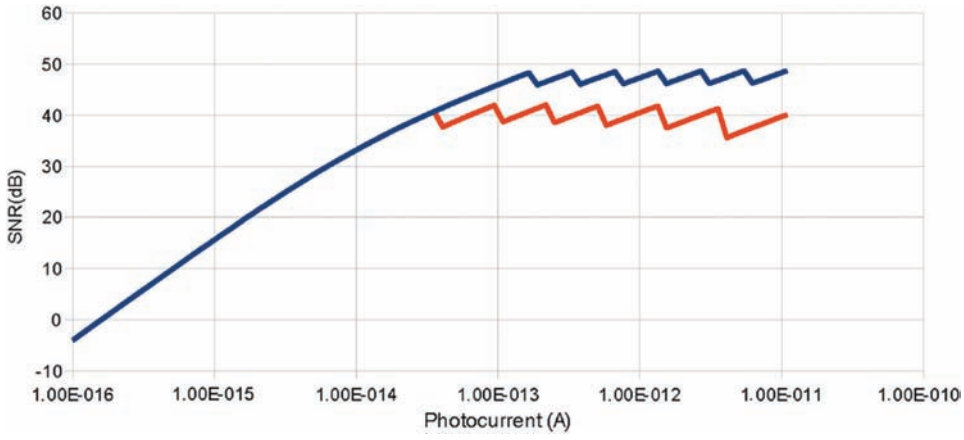

Figure 3.35 Comparison of SNR curves for eight adjustments of well size and nine captures for multiple captures, both for the same total dynamic range (reproduced from Ref. 28).

smaller than in the piecewise linear response for the same DRF. Moreover, the final SNR at saturation always equals the peak SNR of the linear mode.

It is also possible to analyze a combination of more than two samples. The general case is not of practical interest, but consider the formulae for the selected exposure times in a 2's power geometric progression. This case corresponds in photography to a 1-EV (exposure value) exposure difference between images. In photography, the exposure value is defined as $EV = \log_2(F^2/T)$, where F is the aperture's f-number, and T is the exposure time. Doubling the exposure time is the same as increasing the exposure value by 1. (This is interesting to put in parallel with the software methods that will be discussed in Chapter 4.) There are also in-pixel multiple-sampling solutions that use this approach as, for example, one bit of the final word is used per sample.

For $k+1$ samples at $\frac{t_{int}}{2^k}, \frac{t_{int}}{2^{k-1}}, \ldots, \frac{t_{int}}{2}, t_{int}$,

$$DRF = \frac{\frac{2^k q_{\max}}{t_{int}} - i_d}{\frac{q_{\max}}{t_{int}} - i_d} \approx 2^k \tag{3.24}$$

and

$$SNR(i_{ph}) = \begin{cases} \frac{i_{ph} t_{int}}{\sqrt{q(i_{ph}+i_d)t_{int}+\sigma_r^2}} & \text{if } 0 \le i_{ph} < \frac{q_{\max}}{t_{int}} - i_d \\ \frac{i_{ph}\left(\frac{t_{int}}{2}\right)}{\sqrt{q(i_{ph}+i_d)\left(\frac{t_{int}}{2}\right)+\sigma_r^2}} & \text{if } \frac{q_{\max}}{t_{int}} - i_d \le i_{ph} < \frac{2q_{\max}}{t_{int}} - i_d \\ \cdots & \cdots \\ \frac{i_{ph}\left(\frac{t_{int}}{2^{k-1}}\right)}{\sqrt{q(i_{ph}+i_d)\frac{t_{int}}{2^{k-1}}+\sigma_r^2}} & \text{if } \frac{2^{k-2}q_{\max}}{t_{int}} - i_d \le i_{ph} < \frac{2^{k-1}q_{\max}}{t_{int}} - i_d \\ \frac{i_{ph}\left(\frac{t_{int}}{2^k}\right)}{\sqrt{q(i_{ph}+i_d)\frac{t_{int}}{2^k}+\sigma_r^2}} & \text{if } \frac{2^{k-1}q_{\max}}{t_{int}} - i_d \le i_{ph} < \frac{2^k q_{\max}}{t_{int}} - i_d. \end{cases} \tag{3.25}$$

The response curve, information transfer, and presence of dynamic range gaps can be calculated in a similar way as in Section 3.2.

Most of these devices fall in one of the following three design categories: (1) Very-high-bandwidth output that allows for all intermediate samples to be read out before the next sample occurs, and the image is reconstructed outside the image sensor or pixel array. (2) Devices with in-pixel signal reconstruction like that in Fig. 3.36. (3) Devices with in-pixel analog storage of the samples, and one or multiple readout taps for reconstruction are outside the pixel array.

The motion artifacts that this technique exhibits are very similar to the those of the multiple-segment response because in both cases multiple exposures are made within the same long-exposure period. Motion artifacts can be worse if the multiple samples are replaced by multiple images that do not overlap their exposure over time (as will be seen in Chapter 4).

Such complex pixels have a large pitch and small fill factor, and they are therefore not suitable for traditional CMOS image-sensor implementations. However, they make sense in line scan sensors or 3D stacked structures as the pixel electronics are no longer within the pixel area.

3.4 Multiple-Sensing Nodes

The image sensor is made of two interlaced matrices of pixels with two different responsivities. The sensor then produces two images with two different exposures almost simultaneously, from which the user can get information about the whole dynamic range. In some designs, a single sensitive element connects to two or more storage nodes to store the multiple images. From a mathematical point of view, because each of the sensitive elements is linear, this process is similar to the merging of two images acquired with two different exposure times; the calculations are then the same as in Section 3.3.

3.5 Logarithmic Pixels

It is possible to extend the dynamic range of a sensor without changing its voltage swing by giving the pixel response a nonlinear shape. The most

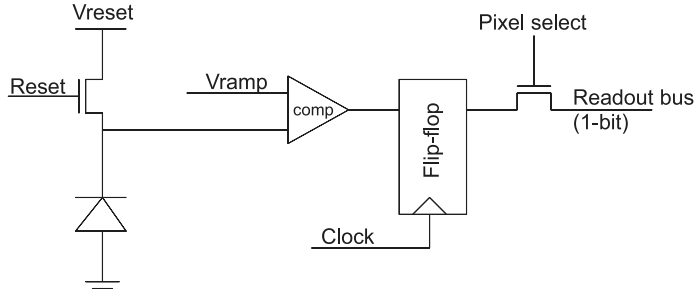

Figure 3.36 A multiple-sampling pixel, similar to the one proposed in Ref. 6.

common shape is the logarithmic curve that gives high sensitivity to low light and low sensitivity to high light.

Logarithmic sensors' responsivity (i.e., the variation of the user signal for a given variation of the light intensity) is inversely proportional to the incoming light intensity. These sensors have the advantage of being able to match extreme dynamic ranges because they do not require an integration time, but they suffer from fixed pattern noise (which needs calibration and external correction) and "response delay." Fixed pattern noise is significantly higher than for other types of sensors, and corrections such as correlated double sampling are not possible, as there is no way to collect the data for a dark reference image in a real situation. FPN can be corrected off-chip out of calibration data. FPN values of up to 40% before correction have been reported. The main source of FPN is the large variations of the MOSFET parameters when operating in weak inversion (subthreshold).

Response delay means that a long acquisition time is required for the dark parts of the response, which makes this type of sensor unusable at high frame rates. The speed variation corresponds to the variation of RC constant due to the variation of the drain current through the load transistor. The resistance involved in this RC constant is the small signal resistance of the MOSFET, approximately given by $(nkT/qi_{ph}) \approx (25\ \text{mV}/i_{ph})$; the capacitor of the RC constant is a combination of the photodiode's capacitance and the parasitic and gate capacitances of the transistors. This time constant is simulated in Fig. 3.37 for a total capacitance of 100 fF and a variation of photocurrent in the range of 0.1–10 pA. For 500-nm light, 100% quantum efficiency and a pixel size of 1.5 μm^2; this range corresponds to a range of irradiance of approximately 12 mW/m^2 to 1.2 W/m^2. The time constant increases very quickly at low light and is already 25 ms for 0.1 pA.

A logarithmic sensor also has a strong temperature dependency. Because this sensor measures the photocurrent and not an integrated voltage, it is not a

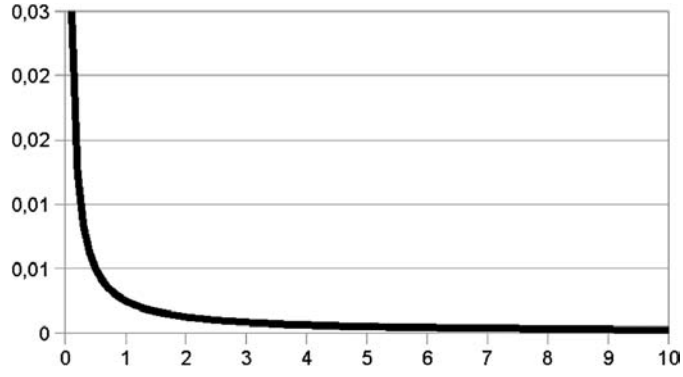

Figure 3.37 Simulation of the time constant of a logarithmic pixel with 100-fF capacitance for a photocurrent between 0.1 and 10 pA (horizontal axis); the results are in seconds (vertical axis).

snapshot sensor, nor does it have a rolling shutter. Its artifacts are similar to the ones of the rolling shutter. The pixels usually require a high current bias (function of the light level), and the device's power consumption can be very high for high resolutions. The snapshot version of these sensors usually has a low performance and has an even more severe image lag.

As shown in Fig. 3.38, a logarithmic pixel can be made out of a 3T rolling-shutter linear pixel by maintaining the reset transistor connected to the pixel's supply. It is therefore also possible to operate a 3T rolling-shutter pixel by maintaining its reset level at a constant voltage that equals the pixel's supply; however, the reset transistor cannot be optimized for both use cases.

At low photocurrent, the load transistor (modeled by a variable resistor, a function of the current) operates in weak inversion, but at high photocurrents the source-gate voltage comes close to the threshold voltage. Transistor operation in this range is well modeled by the Enz–Krummenacher–Vittoz (EKV) equation:

$$I_{DS} = \frac{W}{L}\mu C'_{ox} 2n\phi_t^2 \left[\ln\left(1 + e^{\frac{V_{GS}-V_t}{2n\phi_t}}\right)\right]^2, \tag{3.26}$$

where W and L are the transistor's dimensions, $I_{DS} = i_{ph} + i_d, C'_{ox}$, is the effective capacitance per unit area, μ is the effective surface mobility, V_t is the threshold voltage, ϕ_t is the thermal voltage, and n is the subthreshold slope parameter. Solving for V_{GS}, this equation can be put in the form

$$y = a + b\,\ln[e^{\sqrt{c(i_{ph}+i_d)}} - 1], \tag{3.27}$$

which follows the experimental measurements over at least six decades of photocurrent [65].

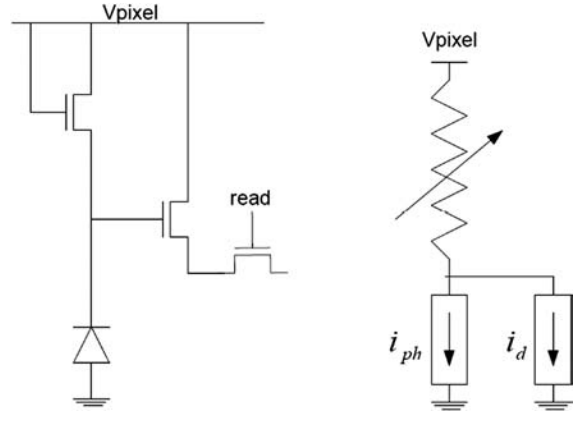

Figure 3.38 Schematic of a logarithmic pixel and corresponding model.

It is not obvious that Eq. (3.27) gives a logarithmic response. The response is only purely logarithmic in the weak-inversion region of operation of the load transistor. In this region, the transistor's V-I formula is approximately

$$i_{ph} + i_d = I_{ds} e^{\kappa \frac{V_{GS} - V_t}{V_t}} = I_{ds} e^{\kappa \frac{V_{pixel} - V_{out} - V_t}{V_t}} = I_{ds} e^{\kappa \frac{V_{pixel} - (V_{pixel} - y) - V_t}{V_t}} = I_{ds} e^{\kappa \frac{y - V_t}{V_t}}, \tag{3.28}$$

where I_{ds} is the drain current at saturation, and

$$\kappa = \frac{1}{1 + \frac{c_{dep}}{c_{ox}}}. \tag{3.29}$$

Solving for y gives an equation of the form

$$y = a + b \ln(i_{ph} + i_d) = V_t \left(1 - \frac{\ln I_{ds}}{\kappa}\right) + \frac{V_t}{\kappa} \ln(i_{ph} + i_d), \tag{3.30}$$

which is purely logarithmic. Without going into detail, with this simpler formula, it is possible to calculate

$$DR = \frac{i_{\max}}{i_{\min}} = \frac{I_{DS} e^{\frac{(V_{pixel} - V_{\max})\kappa}{V_t}} - i_d}{I_{DS} e^{\frac{(V_{pixel} - V_{\min})\kappa}{V_t}} - i_d} \approx e^{\kappa \frac{V_{\min} - V_{\max}}{V_t}} = e^{\frac{\left(V_{swing} - \sqrt{\frac{kT}{2\kappa C_{pd}} + \frac{\sigma_r^2}{C_{pd}^2}}\right)\kappa}{V_t}} \tag{3.31}$$

and

$$SNR(i_{ph}) = \frac{i_{ph} \frac{V_t}{\kappa}}{(i_{ph} + i_d)\sqrt{\frac{kT}{2\kappa C_{pd}} + \frac{\sigma_r^2}{C_{pd}^2}}}. \tag{3.32}$$

This curve tends to become flat as the photocurrent becomes significantly larger than the dark current. The information transfer is given by $I = V_t/[\kappa(i_{ph} + i_d)]$ as soon as $SNR > 1$.

As SNR and dynamic range only depend on process and design parameters, the dynamic range is fixed by design and is not always appropriate. For LDR signals, a gain is used between the pixel array and the ADC so that the low-voltage swing signal matches the ADC's input range. Another approach implements a programmable ADC input range, typically by changing the gain or the reference current of the ramp generator of a column-parallel ramp ADC.

3.6 Logarithmic Photovoltaic Pixel

Introduced in 2010, this pixel (Fig. 3.39) is an improvement of the usual logarithmic pixel, producing higher performance by resolving the problems of image lag and higher fixed-pattern noise. The advantages of using a photodiode in photovoltaic mode (unbiased versus photoconductive mode, such as in a solar panel) are numerous: excellent response uniformity versus temperature, better performance in ultra-low light due to a better dark current, and no dependency on a bias current of the photodiode. The hard-reset operation completely removes the image lag and gives a physically accurate dark reference that allows for highly efficient FPN compensation on-chip by a simple double-sampling stage. Besides these interesting advantages, the main disadvantage is the impossibility to compensate the kTC noise related to the hard reset of the photodiode. Therefore, the FPN level cannot match the one of the recent 4T pixel designs, where the kTC noise can be largely compensated by a true CDS. The sensor is expected to have much less image lag than other logarithmic sensors (Fig. 3.40).

Compared to the usual logarithmic design, the logarithmic photovoltaic pixel does not have a biasing resistor or an additional MOS device operating in subthreshold mode. In this solar-cell mode, the logarithmic response is directly given by the photodiode itself (working in photovoltaic mode) and does not depend on any additional components. Therefore, the logarithmic response is very predictable and does not require any control on the photodiode or the amplifier. The readout sequence is then performed by a DC-isolated buffer coupling with an FPN reduction circuit. The power consumption is also reduced, as much less current is drawn from the pixel's supply.

This drastic reduction in FPN is another major achievement of the logarithmic solar-cell sensor. The FPN reduction implementation can be considered as a CDS process, which is most common in linear CMOS sensors.

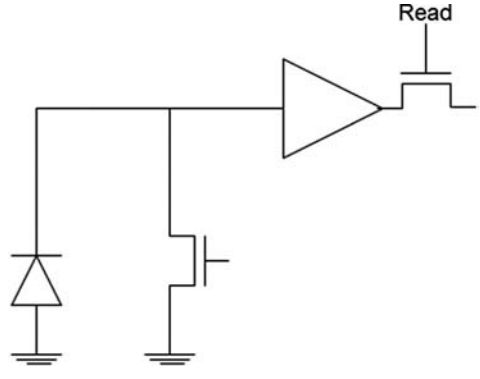

Figure 3.39 Logarithmic photovoltaic pixel.

Figure 3.40 Images from a logarithmic photovoltaic sensor. Top: image of a color test chart with a lightbulb between the camera and the sensor; Bottom: image of a color test chart with a 330-W flood light (courtesy of NIT).

The absence of image lag is also obtained thanks to the FPN readout circuitry by setting a zero voltage on each photodiode, which provides a virtual black image that can be used to cancel the mismatch of the readout circuitry and flushes all charges present in the photodiode.

The dynamic range of such a pixel is fixed. As for the logarithmic pixel, it has no exposure time and no response settings. Its dynamic range is usually too large for many scenes, and the output image may need tone remapping to improve the visual impression. Dynamic ranges up to 120 dB have been reported, but this technology can achieve higher results (up to 140 dB) because of the absence of full-well capacity.

The proposed pixel is depicted below for rolling shutter operation. A derived version can be operated as a global shutter at the expense of more transistors. The buffer can be a simple source follower, but a P-channel metal-oxide semiconductor (PMOS) design with a bias source may be required due to the low voltage on the photodiode. A PMOS design requires additional well implantation that will increase the pixel size. The transistor in parallel with the photodiode can be used to provide a dark reference by short-circuiting the photodiode during reset sampling.

The response curve is not exactly logarithmic, as the response of a diode tends to be linear at very-low-light intensities. There is also a small peak of shot noise at the transition between the linear and the logarithmic behavior. This noise is negligible compared to other noise sources.

3.7 Time to Saturation

The time-to-saturation pixel measures the time at which the photodiode reaches a certain level instead of the level reached in a certain time. The main disadvantage is the larger amount of circuitry required in the pixel. It is another example of a pixel that makes sense in line scan designs or 3D stacked designs.

The photodiode's voltage in each pixel is monitored during integration. Once the photodiode discharges to a certain threshold, the integration is stopped for this pixel and the photodiode is sampled. The end time of integration is also stored in the pixel. At readout, both the integration time information and the integration value are read out. Both values are fused to a floating point representation.

The pixel in Fig. 3.41 works in the following way. The two switches on the left are closed, the photocurrent starts to integrate on the signal capacitor, and the reference current $V_{ramp\ timestamp}$ starts to integrate on the timestamp capacitor $C_{timestamp}$. When the pixel voltage reaches V_{comp} or at the end of the exposure time, the switches are open so that the voltages on the timestamp capacitor and signal capacitor remain constant. The Select switches are then closed for readout. V_{Comp} is constant during exposure but is changed at the end of exposure and during reset. If saturation is not reached, then the comparator does not trigger, and the pixel behaves like a regular linear pixel.

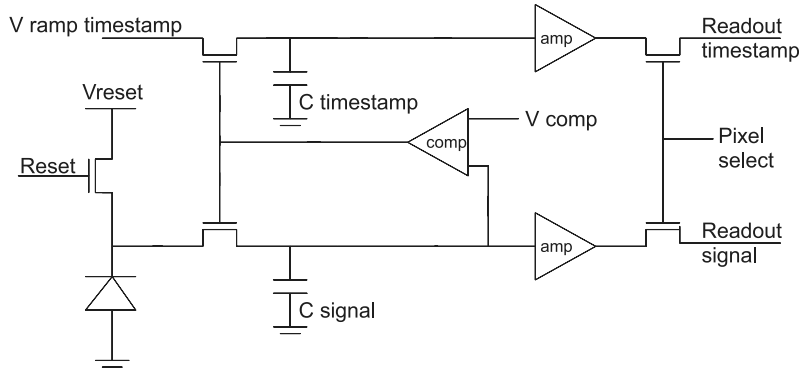

Figure 3.41 A time-to-saturation pixel circuit combined with a regular integration circuit (floating point output), similar to that proposed in Ref. 3.

It can be demonstrated that

$$DR = \frac{\frac{q_{\max}}{q}\left(\frac{t_{int}}{\sigma_{sat}}\right) - \frac{i_d}{q}t_{int}}{\sigma_{dark\ noise}} = \frac{q_{\max} - i_d\sigma_{sat}}{\sigma_{dark\ noise}}\left(\frac{t_{int}}{\sigma_{sat}}\right) \tag{3.33}$$

and that

$$SNR(i_{ph}) = \begin{cases} \frac{i_{ph}t_{int}}{\sqrt{q(i_{ph}+i_d)t_{int}+\sigma_r^2}} & \text{if } i_{ph} \le \frac{q_{\max}}{t_{int}} - i_d \\ \frac{q_{\max}/q}{\sqrt{\frac{q_{\max}}{q}+\left[\frac{(i_{ph}+i_d)\sigma_{sat}}{q}\right]^2+\sigma_r^2}} & \text{if } i_{ph} > \frac{q_{\max}}{t_{int}} - i_d. \end{cases} \tag{3.34}$$

3.8 Gradient-Based Image

Gradient-based detectors detect the spatial gradient of the signal instead of the signal itself. The spatial gradient is a vector representing the derivative of the pixel intensities in the vertical and horizontal directions of each pixel. It represents the rate of variation to the above, below, left, and right neighbors of each pixel. The relative luminance map can then be computed as an image; because the gradient is a derivative, the real image can be computed by the calculation of an integral. The additive constant that appears in the calculation can be chosen so that all pixel values are positive or can be calculated if the luminance of one point of the scene is known. The detector uses the fact that the luminance of the scene does not often swing from the maximum level to the minimum level over a single pixel and that the encoding of this variation requires less dynamic range than the encoding of the absolute value. These pixels transfer information to each other; they are very complex, the resulting image is highly sensitive to defect pixels, and any pixel influences the value of its neighbors. Each pixel cannot be considered as an independent element.

3.9 Light to Frequency

Frequency modulators provide an output signal of constant amplitude but with a variable frequency directly proportional to the photocurrent. A dynamic range of 120 dB has been reported on 2D matrices, but due to the need for PMOS and NMOS in the same pixel and more than ten transistors per pixel, the technique is usually only used on line-scan sensors or single-point light sensors. In low light, the generated frequency requires a long exposure time (or readout time) in order to see at least one period. In high light, the bandwidth of the readout must be very high. Such a pixel is read in two ways: At low frequencies, the time between reset and the first pulse is measured. At higher frequencies, the pulses are counted over a fixed time.

An example light-to-frequency pixel is presented in Fig. 3.42, with its operating waveform shown in Fig. 3.43. The signal on the photodiode triggers

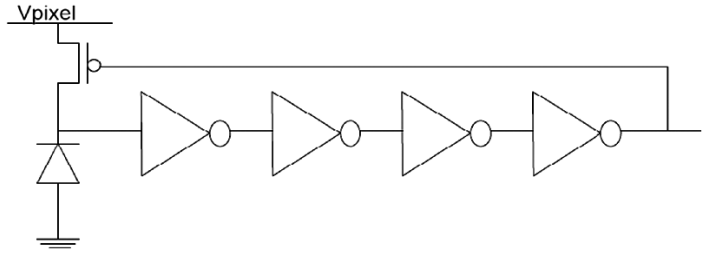

Figure 3.42 A light-to-frequency pixel.

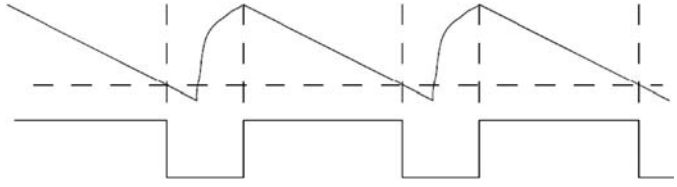

Figure 3.43 Operation waveform of the pixel in Fig. 3.42.

a chain of inverters that will reset the pixel with some delay. The reset of the photodiode again toggles the states of the inverters to end the reset. The chain of inverters defines the reset period. The reset signal is the digital frequency output signal.

3.10 Multiple Readout Gains

Multiple readout gains is a CMOS pixel-readout technology that provides a feature similar to a two-slope companding ADC. The pixel is read out simultaneously by two ADC channels. The first one has a high-gain preamplifier so that the pixel signal saturates quickly, but the low-light signal is well into the ADC's range, and the complete ADC range is used to image the darkest areas of the scene. The second ADC has a low preamplifier gain so that it is well suited to cover the complete range of the signal until pixel saturation, but it is not well suited to convert fine details of the dark areas of the scene. When remapping to a single digital word, this ADC gives a result similar to a companding ADC with exceptional noise performance.

Because the sensor behaves linearly for each ADC, it is quite easy to prove that

$$DR = \frac{q_{\max} - i_d t_{int}}{\sqrt{\sigma_{HG}^2 + q i_d t_{int}}} \tag{3.35}$$

and that

$$SNR(i_{ph}) = \begin{cases} \frac{i_{ph}t_{int}}{\sqrt{q(i_{ph}+i_d)t_{int}+\sigma_{HG}^2}} & \text{if } 0 \le i_{ph} \le \frac{q_{\max,HG}}{t_{int}} \\ \frac{i_{ph}t_{int}}{\sqrt{q(i_{ph}+i_d)t_{int}+\sigma_{LG}^2}} & \text{if } \frac{q_{\max,HG}}{t_{int}} \le i_{ph} \le \frac{q_{\max,LG}}{t_{int}}. \end{cases} \tag{3.36}$$

Note that strongly amplifying a pixel signal does not significantly increase the signal-to-noise performance. All pixel noise is amplified together with the pixel signal. The only advantage is that the noise of the ADC and the additive noise of the amplifier itself are not amplified so that the signal-to-noise can be increased.

The readout over two ADCs allows for less bits per ADC and therefore a lower clock frequency. As a result, each ADC has less $1/f$ noise than a single high-resolution ADC.

The two independent non-destructive readouts can be further combined in order to reduce the noise.

This approach is used in so-called scientific CMOS image sensors. It is one of the few HDR methods that increases the dynamic range both in terms of saturation level and sensitivity.

It is also possible two implement the two conversion gains inside the pixel.

3.11 Other Methods

The LinLog™ sensor is a snapshot sensor with photocurrent integration that behaves linearly at low light and logarithmically in bright light. It has the disadvantage of requiring a complex pixel and a pixel bias. Due to its large pixel area and limited noise performance, it is not used in production sensors.

The logarithmic compression with feedback solves the response-speed issue of the logarithmic pixel by shifting the pole of the RC low-pass filter to a higher frequency; it has higher static noise due to the extra transistor matching and higher temporal noise due to the feedback loop. It also has even larger power consumption.

Some photoconversion layers have a nonlinear response to light intensity. They can be used on almost any device, but this text does not consider them as CMOS (or CCD).

The high-current-gain-amplifier sensor uses a current mirror with a large current-gain ratio to amplify the photocurrent. A very high responsivity with PWLR can be realized, but at the expense of an amplification of the photodiode's dark current and a higher DSNU and PRNU. Furthermore, similar to the logarithmic sensor, the loading MOSFET reduces the response time and is responsible for a very large image lag for fast acquisition speeds. See Fig. 3.44.

AC pixels (pixels that are only sensitive to the variation of the signal over the exposure time and not its average value) can be used to detect small signal variations in a large-DC dynamic range scene. AC is the variable part of a

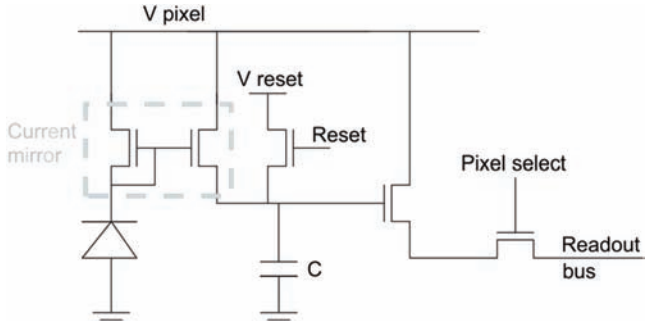

Figure 3.44 High-current-gain-amplifier pixel, similar to that proposed in Ref. 5.

time-varying signal, and DC is the continuous part of this signal. Any time-varying signal can be decomposed in the sum of its average (DC) and its variable components (AC). In this case, the dynamic range of the DC part of the image is not an issue, as only the AC part of the signal is acquired (see Ref. 66). This type of pixel is similar to what is used in time-of-flight ranging cameras.

Designs based on beam splitters (optical components that split a beam of light in two) allow for the same image to be projected on several sensors (see Fig. 3.45). This arrangement produces results similar to those made by 3CCD cameras. If each of the sensors has a different exposure time or a different dynamic range, is covered by a different uniform, neutral-density (ND) optical filter, or if the splitting ratio of the beam splitter is not 50%, then an HDR image can be reconstructed in a way similar to the software method or the multiple-sensing-nodes method. This process allows global shuttering and moving scenes to be imaged in HDR using the software method with standard linear sensors, which can also be combined with the use of HDR sensors; it is called a hybrid (hard-and-soft) method. In general, beam splitters have a splitting ratio that depends on the wavelength, thus producing color aberrations in such a design.

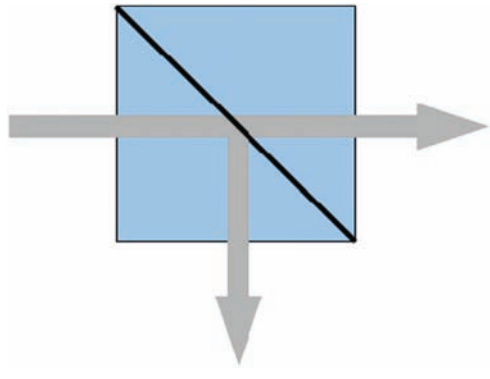

Figure 3.45 A beam-splitter implementation.

In locally adaptive image sensors, the responsivity of a pixel is adapted as a function of the mean intensity detected in the neighborhood of the pixel, and the resulting image is computed externally with considerable effort. Up to 90 dB has been reported with a programmable neighborhood size. This method of local adaptation is similar in concept to the tone-mapping algorithms based on local operators. One of the designs uses a hexagonal resistor network to measure an analog average of the surrounding pixels (similar to a Gaussian average filter) in order to define a value to subtract from the current pixel.

It is of course also possible to design a double camera with two sensors near each other and synchronized so that a long and a short exposure are acquired at the same time. (This approach will be discussed in Chapter 4 as it is very similar to two exposures of the same sensor but with fewer motion artifacts and more positioning artifacts. A variant of this approach uses two synchronized sensors with the same exposure time and a lens with the same field of view but a different F/# or a different sensor responsivity.

3.12 Multiple-Exposure Windows

For image processing applications, especially some front-vision automotive applications, it is not necessary to have the same response curve over the whole pixel matrix. This has already been investigated by the "tiled gain" feature that allows the user to set a different gain in different regions (tiles) of the pixel array. Because the brightness increase is based on a gain, there is no increase of dynamic range. Because the regions are tiled, it is difficult to set a different brightness for one particular object.

Multiple-exposure-window sensors feature a multiple-integration areas-of-interest mode whereby the integration parameters of each area can be set almost independently of the others. The regions can also be defined by their dimensions and the coordinates of the top-left corner (see Fig. 3.46). This sensor allows for optimum settings within a region of the image without taking into account the requirements of another region; Fig. 3.47 shows an example. This feature is especially useful for object tracking while extracting global image properties or for tunnel situations where a given part of the image (usually in the center) is much brighter or darker than the remaining parts. The feature can also be used in night-vision applications to draw the driver's attention to a particular part of the image by making it brighter. Another example involves applications where a single sensor must observe two or more independent regions through a periscope or wide-angle lens.

Road-sign detection suggests another useful application. If road signs are imaged far from a camera, they will be too small for a good reading with a limited sensor resolution. When imaged too close to a vehicle, they will be affected by motion blur due to the relative movement of the vehicle during

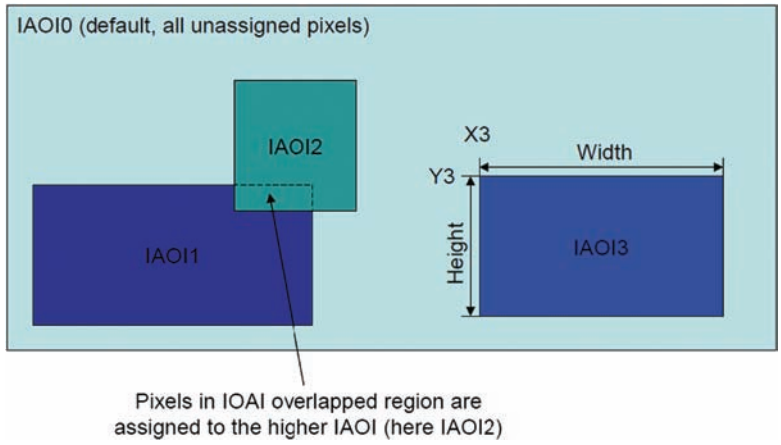

Figure 3.46 Principle of multiple-exposure windows. IAOI stands for independent areas of integration (MIEW, a synonym, stands for multiple independent exposure windows).

Figure 3.47 Use of multiple-exposure windows to extend the intrascene dynamic range (courtesy Melexis).

the integration time. A longer integration time is usually required for the scene in front of the vehicle, but a shorter integration time is desirable for the sign because it often has sufficient contrast. Using multiple regions allows the whole scene to have its optimum integration time and dynamic range, and a small area following the relative movement of the road sign will have a shorter integration time to avoid blurring and allow for a correct detection.

The method can be combined with multiple-segment response, multiple sampling, multiple exposure, well sizing, or any other XDR method with integrating pixels.

3.13 Combined Methods within One Sensor

It has already been mentioned that several methods can be combined in one design by changing the control signals of the pixels. For example, a 3T pixel

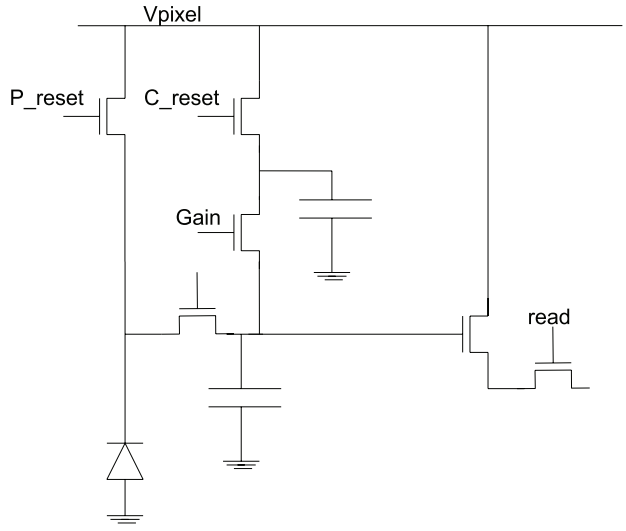

Figure 3.48 5T with a 6T pixel.

with a multiple-segment response based on changes of the reset signal over the exposure time can also be used as a logarihmic sensor by connecting the reset signal to the supply voltage at all times. The multiple-exposure-windows approach can be combined with nonlinear pixel control.

Another design that combines multiple methods is the 5T with a 6T pixel of the Pyxalis HDPYX sensor (see Fig. 3.48). In this pixel, an additional transistor is present in series with the reset transistor to allow for a gain switch within the pixel; it can reach a dynamic range of 92 dB with a SNR dip of 6 dB. The pixel can operate in global-shutter and rolling-shutter modes.

The image sensor also features the capability to provide a different exposure time for the odd and even lines so that they can be recombined by software during readout and reach a dynamic range of up to 114 dB. Both methods combined achieve 120 dB.

Such a pixel could additionally operate as a regular 5T multiple segment XDR pixel.

3.14 Summary

Figure 3.49 shows several images from the same scene acquired by several XDR methods, including the software method that will be discussed in Chapter 4. The light used is a 150-W bulb pointing to the printed circuit board and to the sensor; a standard fluorescent light is lighting the scene from above; the sky is approximately 10000 lux; and the poles on each side of the window are matte black and do not receive any direct light. The objects on the right are metallic and highly reflective. The table is light gray and reflective. A cheap 60-deg FOV in M12 and without IR blocking filters is used. Lens flare

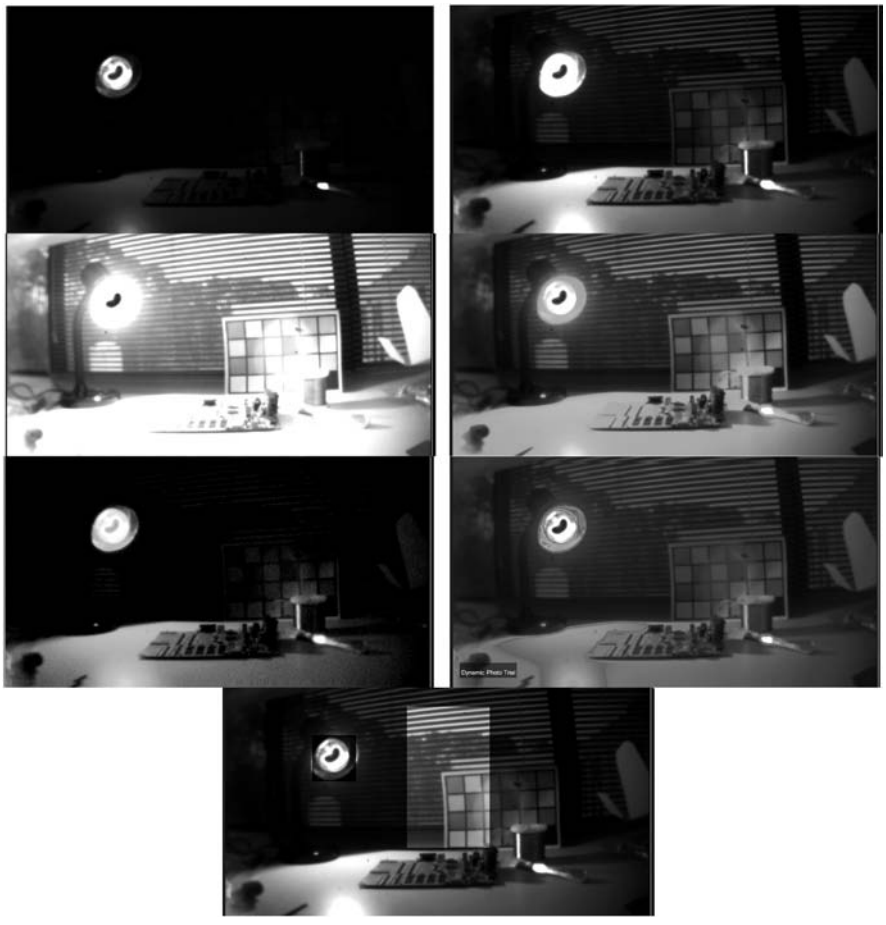

Figure 3.49 Test scene comparing various methods out of the same test sensor (real images). From top left to bottom right: linear low-integration-time image, linear medium-integration-time image, linear long-integration-time image, multiple-segment HDR image, snapshot logarithmic image, multiple-exposure HDR image (software reconstruction), and finally a multiple-integration areas-of-interest image (courtesy Melexis).

can be seen on the test chart and in the bottom left. The sensor has a panoramic VGA resolution and is sensitive to near-IR and visible light.

Many of the depicted methods can be combined in a pixel design with little overhead to achieve maximal dynamic range performance. Table 3.1 summarizes the advantages and disadvantages of the various techniques discussed in this chapter. Figure 3.50 provides a simplified comparison of XDR methods.

3.15 Companding ADCs

A 120-dB linear dynamic range requires a 20-bit data bus, which is not always suitable for high-speed imaging, data transmission, or data storage.

Table 3.1 Comparison of HDR methods

Category	Method	SNR	DR	Other Advantages	Other Disadvantages
Multiple gains	Multiple gains	– Very good – SNR dip between high and low gain is small	– Limited	– Pixel can have large fill factor and excellent low-light performance – Linearity is good – High speed	– Limited extension of dynamic range
Nonlinear pixel response	Logarithmic	– Limited	– Large	– Instantaneous current measurements – Simple pixel – High fill factor – Easy to operate	– Pixel response time is a decreasing linear function of photocurrent; low-light-level imaging is not possible at video rate – Very high FPN – Fixed dynamic range – No exposure
	Logarithmic photovoltaic	– Good	– Large	– Instantaneous current measurements – Simple operation	– Fixed dynamic range – No exposure
	Multiple slopes by clamping, skimming, or well sizing	– Large SNR dips, potentially causing dynamic range gaps	– Large	– Can be used with simple, small, high-fill-factor, rolling-shutter pixels – Can be used with global-shutter pixels – Simple operation	
Variable exposure	Multiple samples (in-pixel)	– Small SNR dips (can even be almost monotonic with sigma-delta pixels) – Good SNR	– Good	– Excellent linearity	– High power dissipation – Large pixel – Low fill factor – Motion blur – Low-light performance can be poor – High-speed readout is necessary
	Multiple samples (off-chip)	– Small SNR dips – Good SNR	– Good	– Excellent linearity	– Motion blur – High-speed readout is necessary
Time to saturation	Time to saturation	– Good only for mid-range photocurrents	– Large		– Large pixel – Low fill factor – High power dissipation
Local	Multiple-exposure windows	– As good as linear	– Good, but not global	– Can be combined with other methods	– Requires *a priori* information about the scene – Only usable in some specific applications
	Locally adaptive	– Acceptable	– Average		– Complex pixel – Low fill factor – Difficult image reconstruction
Gradient-based	Gradient-based	– Good	– Large		– Complex pixel with low fill factor – Very sensitive to defect pixels

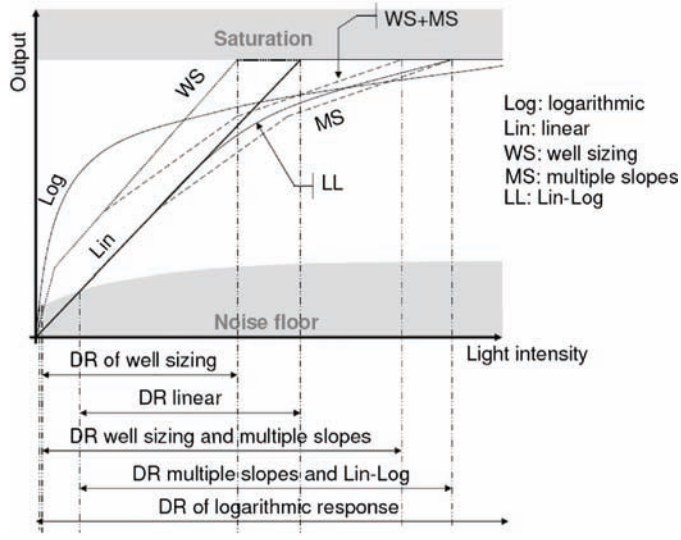

Figure 3.50 Simplified theoretical-pixel-response comparison for various XDR methods.

Moreover, the amount of information contained in a 20-bit word is much smaller than 20 bits. Close to dark, most of the bits are 0, and the amount of information is limited to the few least-significant bits. In all other situations, the least-significant bits only contain the shot noise of the light. For example, just before saturation, a 20-bit linear image sensor can only have 10 bits of useful information. See Fig. 3.51 for an example.

To avoid using all 20 bits with this limited information, a technique called companding ADC is often used (see Fig. 3.52). A companding ADC changes its quantization step with the signal so that an appropriate quantization step is used at every level of the response. A small quantization step is used close to dark so that fine details of the scene can be imaged. As the signal and the shot noise increase, the quantization step is increased so that the word width is

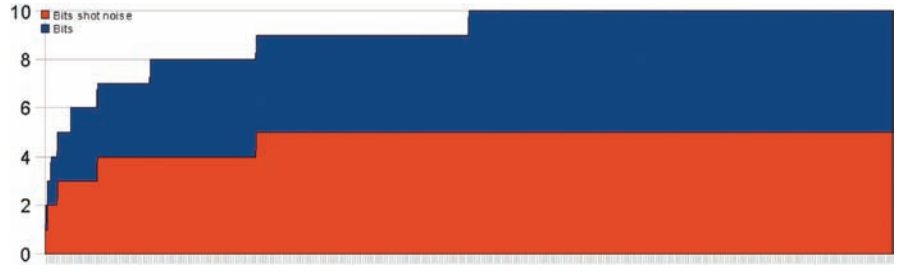

Figure 3.51 Non-zero bits of a 10-bit sensor vs output value in the range [0:1023] (upper area) and amount of bits statistically containing only shot noise (lower curve). Only up to 5 bits out of 10 are useful in this example.

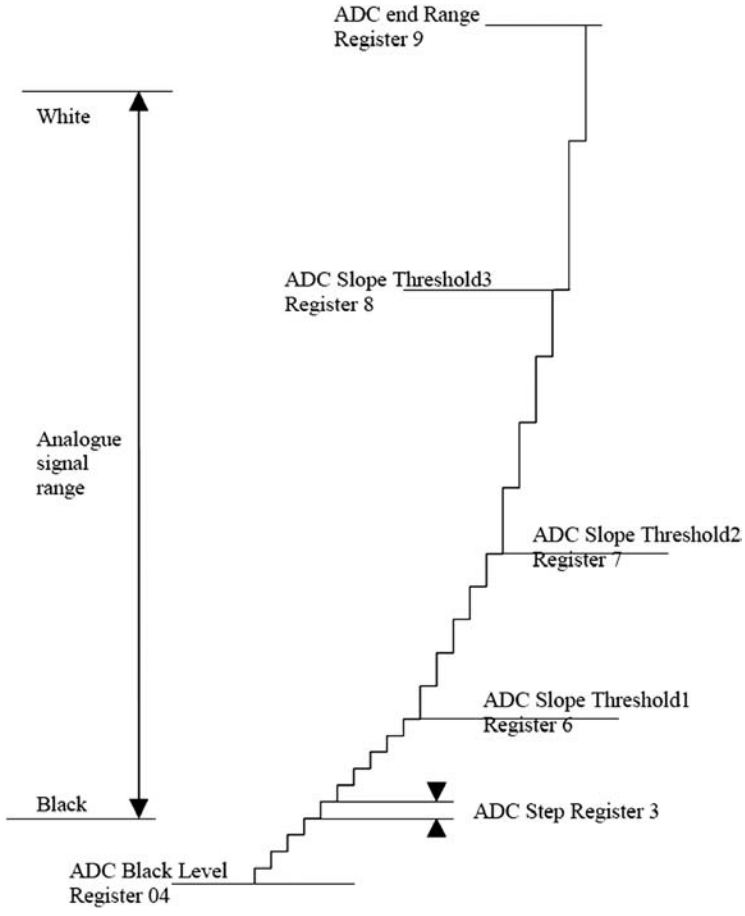

Figure 3.52 Companding-ADC response of an image sensor (courtesy Awaiba).

reduced to only its useful range. This technique can, for example, be used to compress 12-bit pixel data to 10 bits without losing nearly any information, as almost only shot noise is lost. This process is also called a shot-noise-adapted quantization.

The pseudo-code in Fig. 3.53 shows an example reconstruction of 12-bit image data out of a 10-bit companding data source. To avoid missing codes in the decompressed representation, uniform digital noise can be added to the signal, depending on the signal level.

```
If (DNin < ADCT1) DNout = DNin
elseif (DNin < ADCT2) DNout = ADCT1 + 2*(DNin – ADCT1) + 1 bit of noise
elseif (DNin < ADCT3) DNout = ADCT1 + 2*(DNin2 – ADCT1) + 4*(DNin – ADCT2) + 2 bits of noise
else DNout = ADCT1 + 2*(ADCT2 – ADCT1) + 4*(ADCT3 – ADCT2) + 8*(DNin – ADCT3) + 3 bits of noise
```

Figure 3.53 Reconstruction of a linear signal out of a companding ADC signal.

3.16 Extended-Dynamic-Range Color Imaging

In imaging, color is usually made out of the combination of three channels: red, green, and blue (though other means exist). The three channels are present for each pixel so that a single pixel of the image is represented by three numbers, Red/Green/Blue, or a mathematical combination of these three channels (for example, YUV, YCrCb, etc.). The three channels define a color space, which will be briefly discussed in Section 7.1.

At the pixel level, a pixel is modified by a color filter so that it can only be sensitive to a portion of the color spectrum. These portions are almost red, almost green, and almost blue. These pixels, sensitive to "almost" a color channel, are then combined with their neighbors to obtain "almost" three color channels per pixel. This process is called demosaicing. A very common pixel arrangement process is the Bayer pattern, represented in Fig. 3.54 with one corresponding demosaicing principle (other formulae exist).

The resulting values are then modified by color-reconstruction and color-enhancement algorithms in order to provide better color and correct RGB values for each pixel. This is often done using a color-correction matrix, such as

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \tag{3.37}$$

combined with spatial information about each channel. A similar matrix is used for color-space conversions.

On a percent scale, pure red is 100/0/0, pure green is 0/100/0, and pure blue is 0/0/100. For a non-pure color such as 100/50/26, the sensor must be linear to keep the same color content when the light intensity is changing. In the case of a linear sensor, a change of light intensity of 50% will give 50/25/13, and the ratios Red/Green, Red/Blue, and Green/Blue are kept constant. In the case of

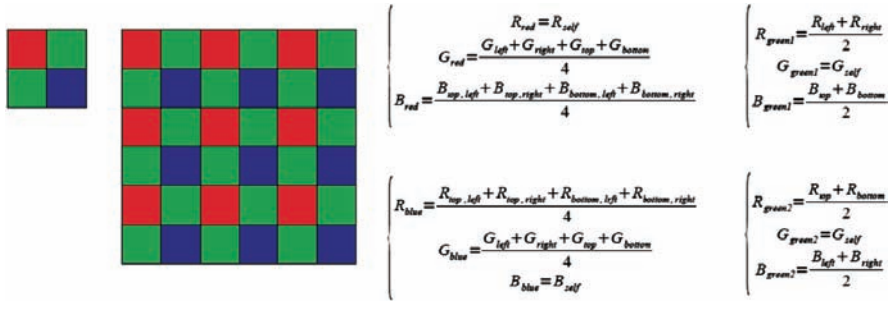

Figure 3.54 The Bayer pattern. Left: the Bayer 2 X 2 base element made of four pixels. Middle: a pixel matrix with Bayer pattern. Right: the most-common corresponding demosaicing formulae (Green1 is the top-right green of the 2 X 2 box, and Green2 is the bottom-left green).

a nonlinear pixel response, these ratios will vary, and the color will be altered. Such ratios are similar to chrominance channels.

Assume a color sensor in linear mode, and an object uniformly colored and uniformly illuminated with all wavelengths in the visible range. The perceived color of the object should not change when the intensity of the illuminant is varied. Unfortunately, when using nonlinear response or ADC companding (if not relinearized on-chip), the output level of the brightest color channel is reduced, and the color is altered. This is illustrated with a two-segment response curve (see Fig. 3.55).

The figure represents the same color at three different brightnesses. In the first three plots, the response curve is linear, and the ratio between the color channels is constant. In the fourth plot, a response curve made of two segments is used. The color ratio is obviously altered, and the color reproduction is wrong, as shown by the color patches on the side of each plot. In this example, the color patch becomes too red due to the relative reduction of the blue intensity. This can be solved by linearization and separation of the luminance and chrominance information. Luminance is kept compressed, and the color operations are performed in a linear space.

Most image sensors suffer from metamerism, as does the human eye. Metamerism means that several light spectra have the same integrated color triplets because the sensor is unable to distinguish between wavelengths (there are only three types of receptors). Multispectral imaging can solve this limitation.

Moreover, in the case of color imaging, each channel has a different responsivity, and analog gains may be applied separately to each channel. This means that the dynamic range and SNR must be defined for each channel (this is, for example, the case in the EMVA1288 standard for image sensor measurements) and that saturation may not be reached on some of the channels.

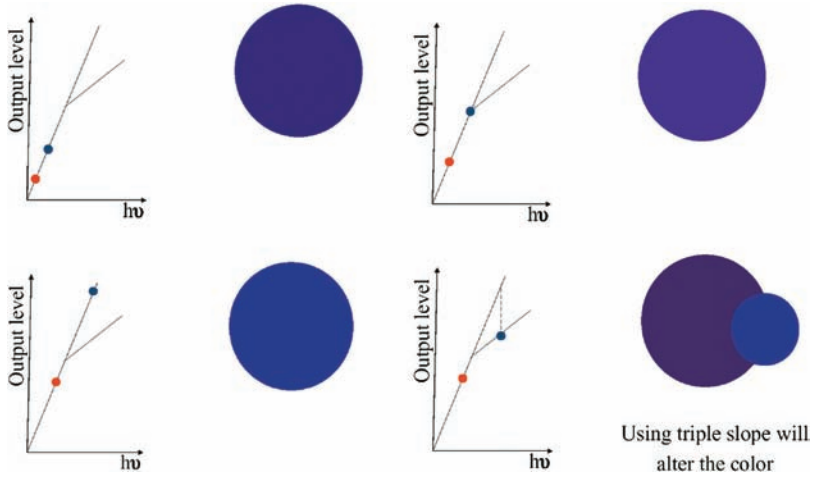

Figure 3.55 Effect of response curve compression on color images.

For example, if there is a bright, red object in a scene, saturation will be reached on the red channel first, and the other channels will be underexposed.

3.17 LED Flicker Mitigation

Recent developments in automotive image sensors include LED flicker-mitigation techniques. LEDs flicker because they are pulsed and therefore may not be on during the exposure time of the sensor. Therefore, there is no guarantee that LEDs will be seen, which is important for red lights, car stop lights, and other LED signs.

With global shutters, the LED is on, off, or dimmed over the whole image. For rolling shutters, the effect changes from line to line.

Various pixel designs for LED flicker mitigation have been developed with either time or space multiplexing of the exposure, or a combination of low and high sensitivities. The goal is to guarantee that the LED is captured at least once.

For example, instead of a 10-ms exposure time in a 30-ms frame, suppose the exposure is split over 20 short exposures of 0.5 ms within the 30 ms. If the LED on time is at least 1.5 ms, then it will be seen by at least one of the short exposures. The brightness of the LEDs will also be similar in all pixels.

Such pixel designs or pixel timings have a tendency to reduce the performance and therefore the dynamic range. However, the additional complexity of the LED flicker mitigation can be combined with HDR techniques to make a XDR pixel with LED flicker mitigation at a reasonable area and performance cost.

Details regarding LED flicker-mitigation designs are outside the scope of this book.

3.18 Sensors Used in Applications

Most of the sensors used in XDR applications use a multiple-readout gain approach or a multiple-segment approach. Increasingly more sensors use multiple independent exposures, as will be considered in the next chapter. There are several companies that design and produce such sensors, each of which has a different design approach to create the nonlinear response of the sensor. Each design is different, and the tradeoffs in performance vary from manufacturer to manufacturer or with the CMOS technology. The decision to use a given type of sensor or a given supplier is very application-dependent; it will sometimes be necessary to request a customization of a sensor in order to reach the best match for a specific application.

3.19 3D Stacking

For many years, image sensors have been made of a single silicon device, sometimes with a complex package, sometimes as chip on board. However,

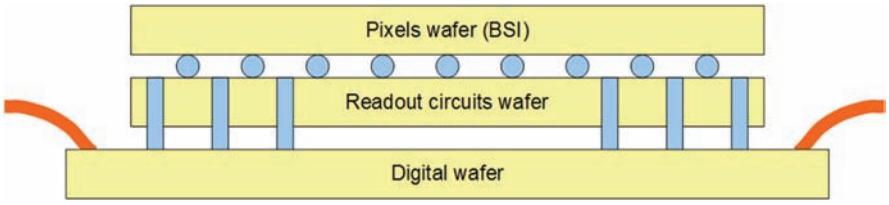

Figure 3.56 Example of wafer stacking with a pixel wafer, a readout wafer, and a digital wafer.

quantum infrared detectors have long used a hybrid structure where the detector wafer, e.g., InGaAs, is bonded to a silicon readout circuit wafer. This has been possible due to the large pixel size of infrared detectors. Recent advances in wafer stacking allow for the same technique to be used for visible-light silicon detectors with a pixel wafer in a silicon process optimized for pixels and one or more readout and processing wafers (each produced in a process optimized for their purpose). Figure 3.56 shows such an example.

In this example, a pixel wafer is processed in a CIS foundry using a backside illumination (BSI) process, for example, 90 nm with pinned photodiodes. This wafer is then bonded to a readout wafer, also processed in 90 nm and optimized for analog circuitry. The bonding in this case uses one ball per pixel to connect the readout node. Connecting only the photodiode would involve a metal to implant a connection that is a significant source of dark current and capacitance. The second wafer is then bonded to a 45-nm digital wafer through silicon vias (TSV). The digital wafer has the bonding wires to the package's leadframe.

This approach allows for the design of complex pixels without compromising the pixel size and fill factor because the complexity is in the third dimension. It opens the door for many future XDR developments with complex circuitry. For example, one could imagine a different exposure time for each pixel or circuits with a very good time to saturation. It is also possible to develop large full-well sensors with external pixel wells. Some stacked image sensors are already available on the market.

One of the important issues to consider when stacking wafers, besides yield, is heat dissipation. Pixels are very sensitive to temperature, and two problems arise when stacking wafers. First, the thermal resistance from the pixel to the package is inferior due to the additional thermal resistance from the other wafers in series. Second, the heat generated by the other wafers also heats the pixel wafer.

3D stacking is one of the most important areas of research for the future of image sensors and especially HDR image sensors.

3.20 Packaging Issues

Packages are required to protect the sensor and the bond wires. The factors of merit for packages are, e.g., their capability to extract heat from the image sensor; the maximum speed of the IO bonding; and the robustness in shocks, stress, sudden temperature changes, and cycles of temperature changes.

Most image sensor packages use a glass window on top of the pixel array to protect the sensor and bond wires. Such a glass is an issue for reliability if a cavity exists between the glass and the sensor. It is also an issue for dynamic range.

The glass adds one or two additional optical interfaces to the system. These interfaces will contribute to veiling glare, transmissivity loss, and blur. The glass material may also contain impurities that will cause PRNU or defect pixels. The glass material may also change the spectral response of the sensor. Therefore, the design of a package with glass should be taken into account for HDR applications.

It is usual to remove the glass for UV applications and some HDR applications. Some packages are glassless or glue the glass directly on the silicon to minimize its effects.

Packages have one additional goal. Light from the lens is projected as a circular spot on a rectangular matrix of pixels. Therefore, light reaches areas of the device that should not receive light. This peripheral circuitry is still slightly sensitive to light and can cause additional noise or incorrect operation of the image sensor. This can be demonstrated in, for example, an OBIRCH test. It is common to cover the non-pixel area of an image sensor with the blue filter of the bayer array to provide a first shield to light. Some package designs are such that the package itself provides additional shielding.

Chapter 4
Software Methods to Extend the Dynamic Range

In addition to the hardware methods discussed in Chapter 3, software methods have also been developed (note that there are algorithms other than the ones presented here; many are proprietary). These methods are of great interest in photography, computer graphics, and scientific applications that require extreme dynamic ranges but need to image a static (or almost static) scene.

Software methods can give much better-looking results than hardware methods, but they have several disadvantages. The first and most important disadvantage is that they require several images of the same scene, which is obviously not possible if the scene contains moving objects or variable illuminations (for example, numerous lights are a good example of varying illumination because their intensity is related to the frequency of the electricity). Flickering LEDs are a problem for most rolling shutter sensors and can also cause problems for software-based HDR methods if one or more of the exposures does not collect information from ON LEDs.

Secondly, the computation time, even for reasonable image resolutions and on powerful computers, is long. High-definition (HD) images are typically processed in several seconds. It is not realistic to think about a fast and efficient embedded solution using the software method described in this chapter. However, for certain applications, there are specific software methods based on very few exposures that can be used in a powerful embedded system.

Finally, software methods create better pictures for people to look at, but they fail to record accurate scene data. It can be demonstrated that, with the exception of very low-glare scenes, multiple exposures fail to increase the range of accurate scene measurement.

Several image sensors implement methods that require exposure fusion similar to multiple exposures. Several such approaches will be presented here.

4.1 General Structure of a Software Approach

Figure 4.1 depicts the general concept of a software approach. The last step (tone mapping) is only necessary if the image data must be compressed with loss in a visually efficient way, or if the image must be displayed to a person using a default display (which always has less dynamic range than the image data). An HDR image acquisition method stops at the radiance map, yet some principles of tone mapping are explained in the next section.

First, a set of images of a scene should be acquired from the same vantage point but with different exposure times. It is important to mention that to obtain perfectly comparable images, exposure time is changed, and not the lens opening. Varying the opening would also vary the depth of focus, and therefore, some of the algorithms presented here would be no longer usable. It is alternatively possible to use neutral density filters. "From the same vantage point" means that the camera and the scene should not move at all throughout the complete acquisition process. The light should also not change between each shot. For a photographer, this means acquiring images of a static scene with a tripod and remote control under sunlight or, DC (continuous) light. Many common light sources are not DC, although they look constant to the human eye. For example, they flicker at 100 or 120 Hz, depending on the region of the world where they are used. LEDs typically flicker between 60 Hz and a few kHz.

4.2 High Dynamic Range Image Data Merging

4.2.1 Ideal case

In theory, the set of images and exposure times to be used are limitless, so long as they cover enough range and as long as the gaps between each exposure are small enough to avoid dynamic range gaps similar to the ones encountered with hardware methods. In photography, exposures are usually chosen so that there is a step of 1 EV between two consecutive images. A step of 1 EV means doubling the exposure time. Figure 4.2 shows how a complete dynamic range can be reconstructed from images.

Each image covers part of the dynamic range. Long-exposure images capture more detail of the dark areas but are mostly saturated. Short-exposure images capture more detail in the bright areas but are dominated by noise in the dark areas.

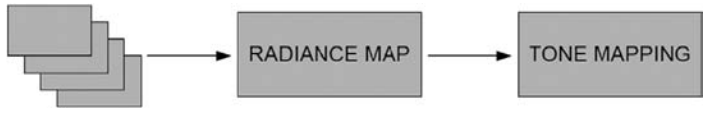

Figure 4.1 General structure of a software approach.

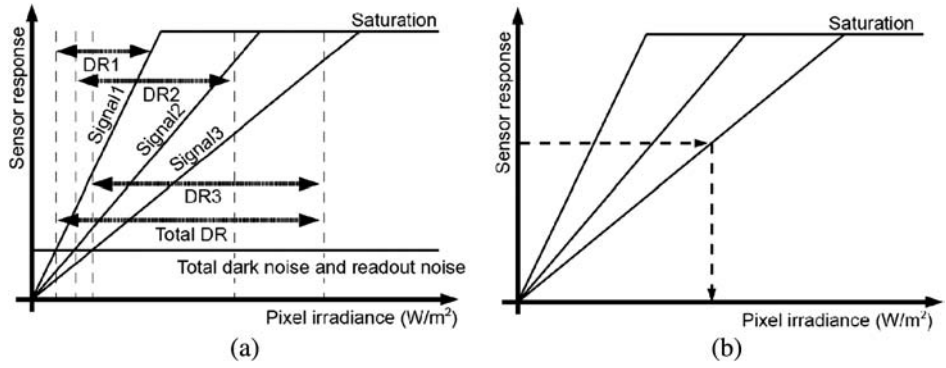

Figure 4.2 Radiance map reconstruction from photographs: (a) dynamic range combination and (b) method to calculate pixel irradiance.

Combining the information from several images achieves more details across the complete range of irradiance. However, due to the change of quantization of the light when exposure time is changed, the rendition of the areas will look better, but an accurate radiance measurement will be sacrificed.

Figure 4.2(b) shows how pixel irradiance can be calculated by knowing the output image of the sensor and its responsivity (the slope of the segments for a given exposure). The absolute value of the irradiance is not important here, and it is understood that the result is affected by a constant scaling factor. This means that the only characteristics of the curves are their relative slopes, which are given by the relative exposure times between the images. Unfortunately, this technique cannot be used to retrieve exact scene radiance, as pixel radiance is the sum of scene radiance and glare.

Overlapping the dynamic ranges of each image helps with noise reduction, as the resulting irradiance value can be computed as the average of the values provided by each curve when this value is not falling into the noise or saturation parts of the curve. The resulting irradiance (light on the sensor) is almost proportional to the radiance of the scene (light from the scene) for the bright and mid-level parts of the image. Although irradiance E and radiance L are proportional for a given pixel, this proportionality factor can be different at different places on the sensor due to the lens design. The resulting data are usually called the relative radiance map. If the responsivity of the sensor and the behavior of the optical setup is known, it is possible to compute an absolute radiance map, although this map will not be correct for dark areas.

4.2.2 Real sensors and cameras

In the previous discussion it was assumed that the sensors were linear, i.e., that sensor response was proportional to $E\Delta t$, where E is the irradiance, and Δt is the exposure time. This also means that doubling the irradiance and halving the exposure time results in the same sensor output; this

property is known as the reciprocity principle. This principle holds true for a wide range of exposures.

Although the response is linear for CCD sensors and almost linear for CMOS sensors (it is more linear if there is enough readout gain), it is clearly not linear for consumer or professional digital cameras and for some silver halide films. Digital cameras apply a nonlinear filter to the image to mimic the response of silver halide films. The response of silver halide films is an S-curve, which is under-responsive close to dark and close to saturation, and over-responsive in the middle range, where it is almost linear. This typical response curve is partially due to the film itself and partially due to the film development process, which also behaves nonlinearly. Digital cameras are specifically designed to have this property.

If the exact overall response curve f is known, the same principle as in the linear case can be applied so that the irradiance value and therefore the radiance map can be reconstructed using the mathematical inverse function of the response curve f^{-1}. The response curve can be known from datasheets or after an experimental measurement. In ISO standards, the response curve is obtained using a calibrated test chart. The chart consists of patches of known radiance when illuminated (usually back-illuminated) with a standard illuminant and is usually a plot of the camera output versus the log luminance. In the EMVA1288 standard, the response curve is obtained by measuring the sensor response for several monochromatic irradiances and is a linear plot of the sensor output versus the linear irradiance.

If, however, the response curve is not known, the method will only work if the response curve can be estimated from the radiance map reconstruction process. This algorithm is known today as the Debevec method (the original paper was published by Debevec and Malik) [76]. There are other possible algorithms as well.

4.2.3 Debevec's algorithm

Although this algorithm is applicable to any image, it is described here using only one wavelength to simplify the math involved. To consider a real scene, some integrals involving the spectral response of the sensors should be introduced into the calculations.

For a set of exposure times T for which the elements are Δt_j, a set of image pixels for exposure j (noted Z_j) for which the elements are Z_{ij}, and if the pixels receive the irradiance E_i, $\forall i$, the response curve (assuming that the reciprocity principle holds true) can be rewritten as $Z_{ij} = f(E_i \Delta t_j)$. The pixel response, although nonlinear, is only a function of the exposure $E_i \Delta t_j$.

As the response curve is strictly monotonic, this equation can be inverted. Taking the logarithm of both sides and defining $g = \ln f^{-1}$ gives $g(Z_{ij}) = \ln E_i + \ln \Delta t_j$. For i running over I pixels and j running over J exposure times, there are a total of IJ such equations, with unknowns being

the E_i values and the g function (or actually the finite values that it takes across the domain of Z).

This set of values is computed by minimizing a goal function in the least-square sense. Although Debevec used singular value decomposition (SVD) to solve this overdetermined problem, any method should work, as the system of equations is robust. The chosen goal function is

$$O = \sum_{i=1}^{I} \sum_{j=1}^{J} [g(Z_{ij}) - \ln E_i - \ln \Delta t_j]^2 + \lambda \sum_{z=Z_{\min}+1}^{Z_{\max}-1} g''(z)^2, \tag{4.1}$$

where $g''(z)^2 = g(z-1) - 2g(z) + g(z+1)$is a numeric approximation of the second derivative of g. This minimization over the second derivative forces the response curve to be smooth.

Solutions are only up to a single scale factor, which is chosen by the additional constraint,

$$g\left(\frac{Z_{\min} + Z_{\max}}{2}\right) = 0. \tag{4.2}$$

Finally, as g has a steep slope near $Z_{\min}$ and $Z_{\max}$, the results fit the actual data points poorly near these extremes. To compensate for this and to focus on the central (more important) data points, a weighting function $w(Z_{ij})$ is introduced. It is a triangle-shaped function with its maximum at

$$\frac{Z_{\min} + Z_{\max}}{2}, \tag{4.3}$$

and zero at extremes $Z_{\min}$ and $Z_{\max}$. Using this weighting function, the goal function finally becomes

$$O = \sum_{i=1}^{I} \sum_{j=1}^{J} \{w(Z_{ij})[g(Z_{ij}) - \ln E_i - \ln \Delta t_j]\}^2 + \lambda \sum_{z=Z_{\min}+1}^{Z_{\max}-1} [w(z)g''(z)]^2. \tag{4.4}$$

Practically, the method is not used in this way, as it would involve solving an overdetermined system of several tens of millions of equations for a sensor with only a few megapixels. The solution is then to only consider a couple of pixels to compute the response curve. This response curve is then used to compute the radiance map of the complete image. The condition for having a sufficiently overdetermined system of equations is $I\ (J-1) > Z_{\max} - Z_{\min}$. The I pixels should be chosen so that they have enough distribution across a range of possible pixel values; this is so that the response curve has sufficient data points all along the curve. The data points should also be spatially

distributed sufficiently within the image. These pixels are usually selected in reasonably uniform areas of the image, where the optical blur is minimum.

A radiance map is finally reconstructed using the formula $\ln E_i = g(Z_{ij}) - \ln \Delta t_J$. For robustness, all exposure values for a particular pixel are used. As accuracy is better when the pixel is closer to the center of the response curve, these available pixel exposure values are weighted for the average calculation so that more weight is given to the central pixels. Using the same weighting function as before, the radiance map reconstruction formula becomes

$$\ln E_i = \frac{\sum_{j=1}^{J} \{w(Z_{ij})[g(Z_{ij}) - \ln \Delta t_j]\}}{\sum_{j=1}^{J} w(Z_{ij})}. \tag{4.5}$$

This method works for non-monochromatic light and for color devices if each color channel is treated separately. Unfortunately, due to the additional constraint applied separately to each channel, the white balance of the image becomes altered. It is then necessary to perform a separate white balance, or to include additional constraints, knowing the spectrum of the light source and the spectral response of each color channel of the sensor or camera. A mismatch in color response can occur; this is more important closer to dark and closer to saturation.

This technique creates better-looking pictures, but it cannot be used for a precise radiance measurement of the scene. Evaluation of scene capture must be done with glare-free measurements of scene radiance versus digital values.

Figures 4.3–4.6 show various linear-mode images (in 8 bits per pixel and per color channel) of several scenes and the corresponding reconstructed image. To print the reconstructed images, it was necessary to reduce their bit width (data range) to 8 bits per color channel using tone mapping. Tone mapping is briefly discussed in the next section.

4.2.4 Alternate method: Mann and Picard

The advantage of this method is estimating the camera's response function using only two images of the same scene, captured with different exposure times. The algorithm starts with a dark pixel in the short-exposure image. This pixel is related to the irradiance by $Z_{iL} = f(E_i \Delta t_L)$. The same pixel in the long-exposure image is given by $Z_{iH} = f(E_i \Delta t_H) = f(E_i k \Delta t_L)$, where the exposure time is k times longer. Then a pixel is sought in the short-exposure image that has the same value as Z_{iH}, and the same process is repeated until the saturation value is reached. At the end of this process, a set of pairs is generated. If the response of the camera is assumed to be a power function with $f(0) = 0$, then it can be retrieved using a regression method through the

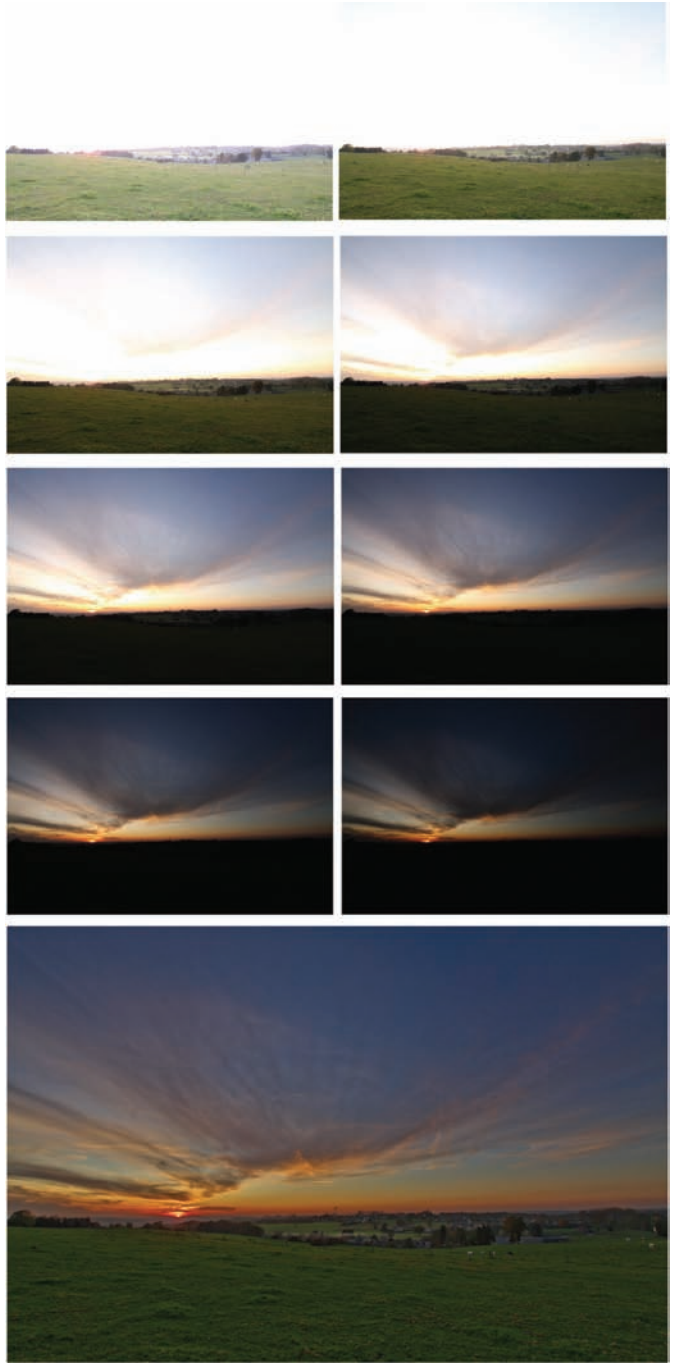

Figure 4.3 HDR image constructed out of eight photographs with exposure steps of 1 EV. This reconstructed image shows a beautiful sunset with warm cloud colors and (at the same time) the correct colors for the cows, trees, sky, and grass, with still enough detail around the sun and within the small foggy area in the center.

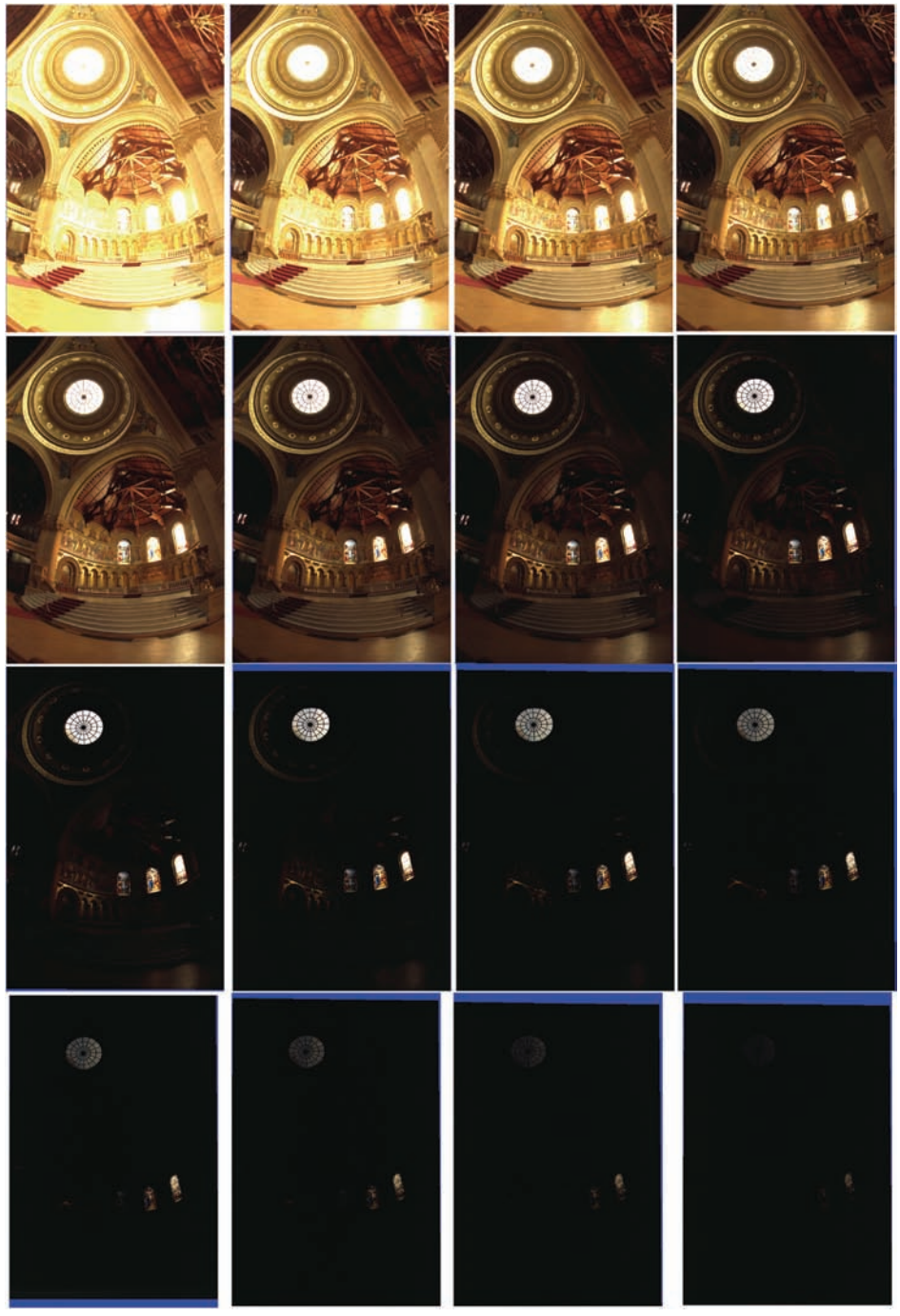

Figure 4.4 Pictures of the Stanford memorial church taken at different light levels [76].

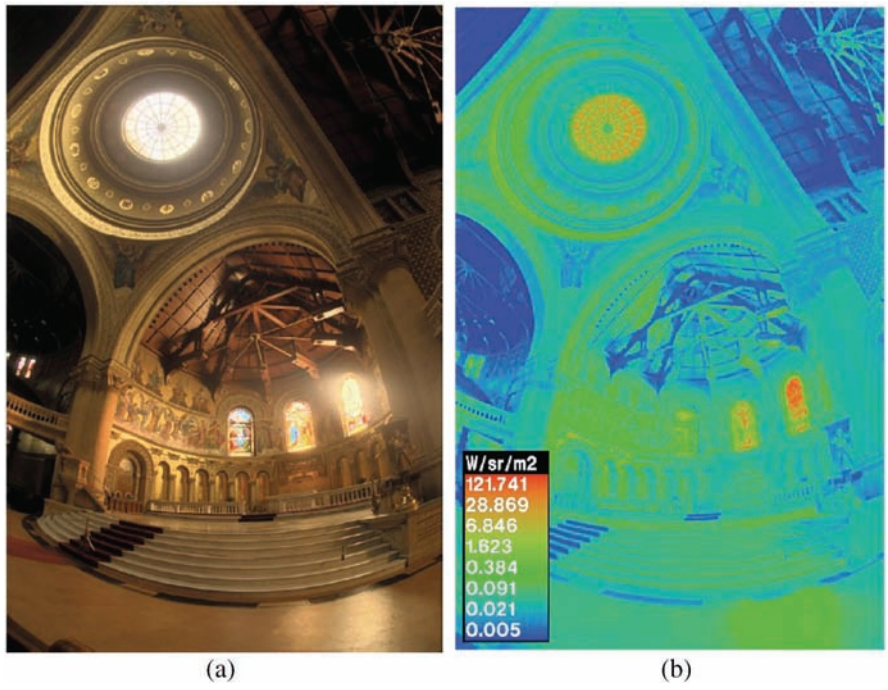

Figure 4.5 (a) HDR image reconstruction of the famous Stanford memorial church image from Debevec (reproduced from Ref. 76) and (b) corresponding radiance map in false color. These images were scanned from film and realigned.

known set of points. When the response function of the camera is finally known, the method is the same as in the algorithm proposed by Debevec and Malik. The Mann and Picard calculation method uses much less computation power than the Debevec and Malik method.

4.2.5 Alternate method: Mitsunaga and Nayar

In this method, the camera's response is modeled by a polynomial of order K, where K and the $K+1$ coefficients are unknown. With this method, the response curve can be estimated in the presence of noise in the input images and without a precise knowledge of the exposures. This method is heavier in computation requirements, as it involves a goal function similar to the Debevec and Malik method but also involves the off-axis angle and aperture size. Moreover, several iterations of the algorithm are required to select the correct order for the polynomial. When the response function of the camera is finally known, the method is also the same as in the algorithm proposed by Debevec and Malik.

Fig. 4.6 HDR image reconstruction out of seven photographs separated by 1 EV, with very long night exposure (up to 8 sec) and two different tone-mapping algorihms (last two images).

4.2.6 Alternate method: Robertson *et al.*

This method proposes modifying the weighting function to include the square of the exposure time because longer exposures provide a higher SNR. Another main difference in method compared to Debevec's algorithm is the choice of a cubic spline function with almost zero derivative at 0 and 100% of saturation as an estimate of the camera's response curve.

4.3 Noise Removal

4.3.1 Temporal pixel noise

It has been shown that SNR is lower in short-exposure images. With the reconstruction process proposed in Debevec's algorithm, this noise is further amplified with the division by the exposure time, which is of small value. This noise is temporally varying. To improve the noise of reconstructed HDR images, it is necessary to average several images at each exposure, with the amount of required images being larger for short exposures than for long exposures. To avoid the acquisition of additional images, and assuming that the primary source of noise is the photon shot noise, it is possible to combine noise removal with the HDR reconstruction process.

4.3.2 Ghosts and misalignments

Ghosts are created when a scene is moving during the image acquisition process. Ghosting can be reduced by using optical flow algorithms, deconvolution, or other methods. Misalignments occur when the camera is moving during the image acquisition process. The images can be realigned by translation, rotation, and scaling if the alignment error is known. In this process, the edge of the image is usually lost because it is not covered by enough images of the set to be reconstructed. Misalignment of the images can be estimated by detecting edges and corners and by performing a search for matching points between the images. This is similar to the internal calibration of sensors used in 3D stereovision algorithms.

The problems with ghosts are illustrated in the following pictures. Figure 4.7 is the properly produced HDR image (bottom right) out of five exposures without motion. The image is correct except for some artifacts due to the open-source software used to generate this example.

Figure 4.8 shows the incorrect reconstruction out of two exposures with one moving car. Most of this image is correct, but a positive ghost and a negative ghost are clearly visible due to the excessive relative motion of the car in the picture.

Figure 4.9 shows the incorrect reconstruction out of five exposures with one moving car. The ghosts here are less visible as each of them roughly represents one-fifth of the data instead of one-half, as in the previous example. In this example, the car almost disappears, and using more exposures could make it disappear completely.

Figure 4.7 Five linear images with different exposures times and the corresponding reconstructed HDR image (bottom right); no motion.

The ghost may look different if the relative motion of the object is more limited. In the case of camera movement, there is more correlation in the changes, a correction is possible (without significant artifacts if the motion is limited), and there is a reasonable processing power. The features of each image of a pair of exposures are extracted in each image and randomly matched between the images. The features can be corners, for example, the result of a Harris–Stephens detector. The matching is first random and then improved in multiple passes of a random sample consensus algorithm (RANSAC). When all candidates have either found a match or been discarded, a model of the motion between the two pictures can then be established, usually as a combination of zoom, pan, translation, and rotation. The model can then be applied to one image to realign it with the other and will result in part of

Figure 4.8 Two linear images with different exposures times and the corresponding reconstructed HDR image (bottom) with ghost motion artifact; one car moves.

images being cut, as in Fig. 4.4 (blue areas correspond to missing data). The correction involves the recalculation of pixel locations using a 3×3 matrix and then advanced interpolation into the gaps. After realignment, the images can be merged to a radiance map.

In the case of mobile phones, data from the accelerometers and gyroscopes can be used as an educated guess for the initial matching pairs of features, which reduces the number of iterations required in a RANSAC.

4.4 Tone Mapping

Tone mapping is a technique used to map one set of colors to another. Each color gamut represents a subset of the existing colors. Tone mapping is required to transfer the image data acquired with one type of color representation to another (Fig. 4.10) (color spaces are briefly discussed in a later chapter). Each color gamut also has a different dynamic range and precision related to the image acquisition system, image encoding and storage systems, and image display system.

Figure 4.9 Five linear images with different exposures times and the corresponding reconstructed HDR image (bottom right) with ghost motion artifact; one car moves.

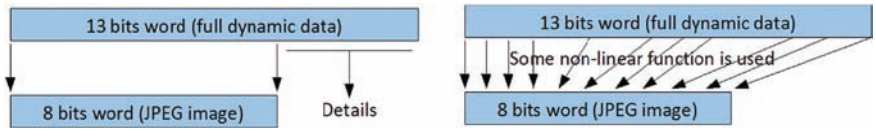

Figure 4.10 Why is tone mapping required?

Three typical applications of tone mapping are the conversion of a film negative to a color photoprint by a chemical process; display of an image on a computer screen that has limited dynamic range and color performance; and the conversion of an HDR image into a JPEG image (or other non-HDR format). In general, the goals of tone mapping greatly vary from application to application.

Various tone-mapping operators have been developed, but they can all be divided into two categories: global or local.

- Global (or spatially uniform) operators: these are nonlinear functions based on image parameters. Once the nonlinear mapping function has been determined, it is applied in the same way to all pixels in the image using a look-up table and does not take into account the neighborhood of each pixel. These operators are fast and usually simple, but they can cause a loss of contrast.
- Local (or spatially varying) operators: a new value is calculated for each pixel depending on its neighborhood, then a global map is applied. The operators are usually more complex and produce unrealistic images. Some operators also generate artifacts such as halos around dark objects (if the contrast is high), ringing, a cartoon-like appearance (due to a lack of large global contrasts), or highly saturated colors. For example, some operators are based on the contrast (gradient) of the images and are called gradient-domain methods.

A mix of both types of operators is often used in HDR photography. Many people prefer the local operators, as they produce better-looking images, but the overall parameters of an image should not be altered too much, so that the image still looks somewhat realistic. Tone-mapping operators are often based on models of the human vision system. Typical global operators are gamma correction and histogram equalization.

The images in Fig. 4.11 are generated from the same HDR image using various tone-mapping operators. Each of these operators has settings that were chosen to produce a decent image, with the exception of linear mapping, which is included to show what happens if an HDR image is displayed without proper tone mapping. Figures 4.12 and 4.13 illustrate a cartoon-like (or painting-like) artifact and halo artifact, respectively.

A review of tone-mapping operators can be found in Ref. 81. This reference also introduces tone-mapping operators that depend on time. Some of these operators also mimic the perception change of the human vision system over time, as they take into account visual adaptation.

It is important to remember that the primary goal of tone mapping here is to reduce the bit depth of an image for storage and display. This is not necessary for machine vision applications, where it is usually more efficient to directly process a wide-bit-depth radiance map. Bit depth here is no longer correlated with scene radiance and dynamic range.

4.5 Software Methods Applicable to Certain Image Processing Applications

In some situations, a simple software method can be used to reach the goals of a computer-based vision system without going through a complete

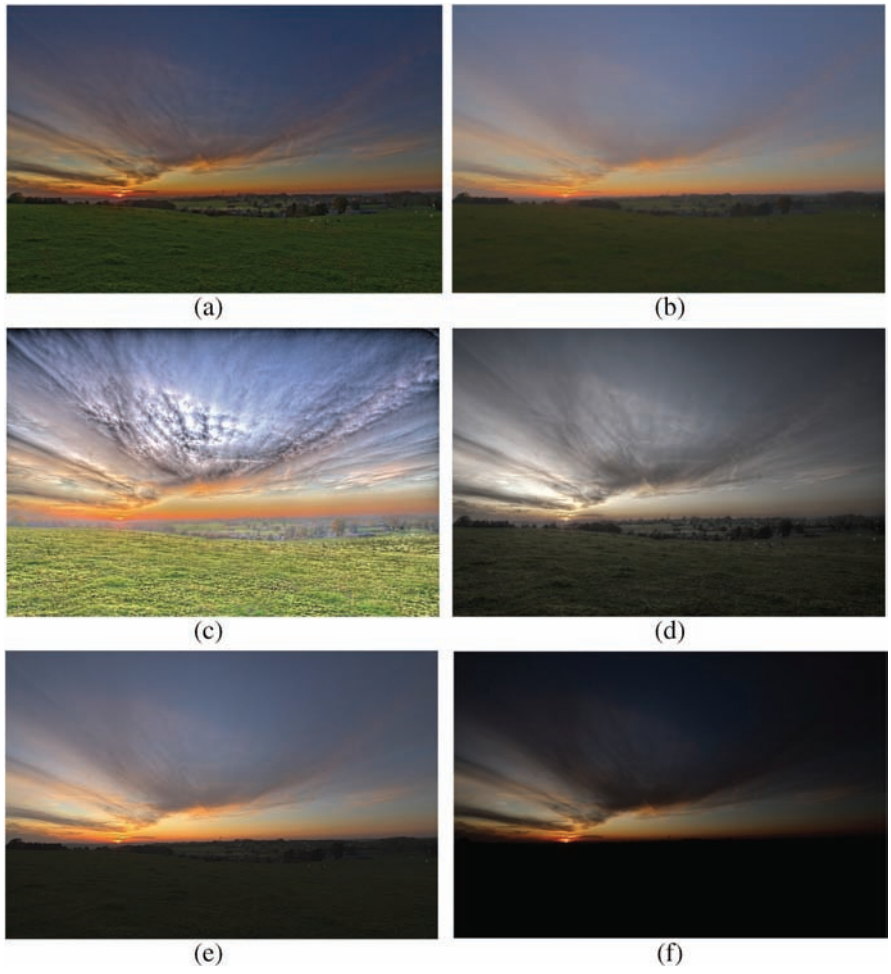

Figure 4.11 The HDR image represented on this LDR page uses various tone-mapping operators: (a) HDR Photostudio 2 from Unified Color Technologies; (b) Ashikhmin's operator [77]; (c) Fattal's operator [78]; (d) Mantiuk's operator [79]; (e) Pattanaik's operator [80]; and (f) linear mapping.

reconstruction process. A good example of this is the automotive application of lane departure warning (LDW). In this application, the only important objects to detect are the lane markers. Their luminance can vary greatly depending on road conditions, illumination, shadows, and tunnels.

This algorithm can be implemented in different ways. One method uses horizontal scan lines to detect vertical edges, and then connects the edges to extract the lane boundary using a modified Hough transform. Another method detects all edges and applies a feature-based filter to define marker candidates before performing the Hough transform. Other methods also exist.

Figure 4.12 Cartoon-like artifact.

Figure 4.13 Halo artifact (around the trees at the interface with the sky).

In both cases (as edge information is the only interest), it is possible to operate in HDR using several LDR images without performing any reconstruction. In each LDR image, the non-saturated part is filtered to extract edges. Then the resulting partial edge information is merged from each LDR image into the resulting image. This produces almost the same edge image as would be produced if it was calculated directly out of a reconstructed HDR image.

4.6 Sensors with Integrated Processing

Some sensors are available with on-chip implementation of a software method. Such sensors acquire two images and merge them on chip, providing a single image containing the merged data. Some of these sensors also include local and global tone-mapping algorithms. The sensors are controlled by programming the desired exposure for each image, or the desired exposure for one image and the ratio of exposure for the second image. The sensors can actually acquire and store the two images or use smart readout pointers to sample the same exposure twice.

Figure 4.14 Simulated HDR images (left) and corresponding original images (right).

There are less-advanced sensors that provide register banks for control of two images, and that automatically alternate the register banks so that a stream of alternating long- and short-exposure images can be obtained without reprogramming the sensor for each frame. There are more-advanced sensors that also implement motion compensation on chip.

Some sensors also provide multiple exposures at the same time, with either time or space multiplexing. A rolling shutter sensor can be operated so that multiple shutters roll through the pixel array at the same time, with two or more reset pointers and two or more read pointers. Therefore, the short and long exposure (two-pointer case) follow each other in very short succession and reduce the motion artifact. The sensor's output can also be made smart enough so that the readouts for the short and long exposures occur right after each other and ease the recombination with limited memory buffering. This is an example of a time-multiplexed multiple exposure.

Space-multiplexed multiple exposures use, for example, different exposure times for the odd and even lines, so that the HDR data can be reconstructed from the readout of two consecutive lines at the expense of the loss of half of the vertical resolution. Another approach uses a random arrangement of multiple exposures within a small block of pixels (3×3, 4×4, 5×5, etc.) so that there is more randomness in the spatial distribution of the exposures, resulting in less resolution loss (or at less lowered human perception of the loss of resolution).

Both time-multiplexing and space-multiplexing methods involve more pixel-control signals or more transitions of the control signals, and therefore a lower intrinsic pixel performance due to signal coupling.

4.7 Simulated High Dynamic Range Images

There are certain algorithms that simulate the look of an HDR image by processing a single image. Such algorithms are available for smart phones as applications, and in some image processing or photo-editing software packages as filters. Although such images look like HDR images because some filters and tone-mapping algorithms are applied, they usually contain less information than the original image and are therefore not HDR images.

The images in Fig. 4.14 were made with a smartphone and an open-source application. It can be seen that the saturated areas and areas that were too dark remain unchanged.

Chapter 5
Optical Limitations

In some situations, engineers focus on the sensor as the key element of a vision system but forget the importance of optics. It is very important to note that there are currently few optical systems that reach the performance of a sensor in terms of dynamic range. The necessity to race for dynamic ranges higher than 150 dB must be questioned if no lens is able to reliably transfer this high dynamic range from a scene to a sensor.

5.1 Lens Glare

Two important artifacts that arise when lenses have a limited dynamic range are flare and ghosting. Ghosting and flare occur when an object is visible several times in an image, or when the shape of lenses or apertures appear in the image (see Figs. 5.1 and 5.2). These objects are usually bright spots that reflect multiple times in the optical path because no lens has perfect transmissivity, even with multiple layers of antireflective coatings. They can also come from parasitic light paths due to non-perfect absorption by the lens barrel, camera housing, or lens mount.

It is incorrect to think that optics used in professional photography are not affected by lens artifacts. The photograph in Fig. 5.1 was acquired using a Fuji bridge still camera. This camera was among the best bridges of its time. The photograph in Fig. 5.2 was acquired using a Canon EOS 40D with a good-quality 11-mm Tamron EF lens. Similar artifacts can easily be created with expensive professional equipment.

It is also obvious that an HDR lens should be fast (i.e., an efficient collector of light) to lower the detection limit close to dark, as more collected light allows the signal to be farther away from the sensor's noise floor. Fast lenses also allow for a reduction of exposure time, also reducing camera movement and the effects of dark current (hence the name fast lenses).

Ghosting and flare are only highly visible effects of lens glare. Lens glare is almost always present but usually not detectable. Because of lens glare, the actual image formed on a sensor is the sum of multiple images created in the

Figure 5.1 Lens flare (linear image).

Figure 5.2 Lens flare in an HDR image (software method, using HDR Photostudio 2 from Unified Color Technologies).

optical path. For an optical device with n elements, the total number of images that sum up on the sensor is given by

$$N = (2n)^2 - n, \tag{5.1}$$

which means that for a 15-element lens, there are 885 images summed. One is of interest here, while the other 884 are parasitic. The formula does not include the effects of the barrel, the reflections on the sensor, and the walls of the camera. The image formed on the sensor is therefore the sum of the expected image and many parasitic images. Each of them is a modified, scaled, shifted, and attenuated version of the expected image.

Glare is scene dependent, which makes accurate radiance map capture impossible for darker areas. Light captured by the pixels in such areas is mostly glare from other parts of the scene, and not light originating from the corresponding scene location (Fig. 5.3). Scene capture is usually impossible below 1.5 log units of dynamic range. Due to parasitic veiling glare, there is no longer a direct ratio (for linear response sensors) between the digital values produced by the sensor and the luminance of the scene, as shown by Fig. 5.4.

Veiling glare is independent from sensor noise and is therefore a noise that adds up with sensor noise to reduce the SNR and dynamic range of a real camera. Veiling glare does not affect the SNR and dynamic range of the sensor.

Figure 5.3 Lens flare in an LDR image of an HDR scene (courtesy of A. Mazzoli).

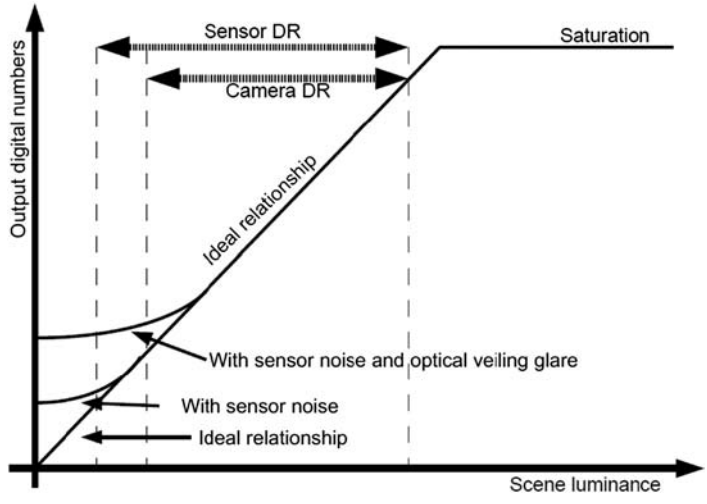

Figure 5.4 Relationship between sensor digital numbers and scene luminance.

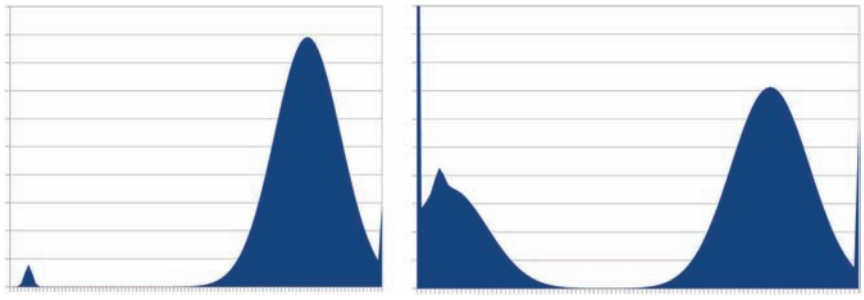

Figure 5.5 Veiling glare.

This additional noise cannot in general be considered white noise, as it is correlated with the scene. However, as reflections are usually random, it can be approximated by white noise in the dark areas of the image if the optical setup has enough elements and a large number of parasitic images sum up. Parasitic images are a dominant issue in HDR imaging.

Figure 5.5 illustrates veiling glare. It represents the histogram of an image that contains the Sun and a dark object on a black background. On the left is the histogram without veiling glare. The curve is the histogram of the dark object. On the right is the histogram with veiling glare. The peak from the Sun has less energy and significant signal, and noise is presently close to dark. As a result, the dark object becomes almost invisible within this noise.

5.2 Modulation Transfer Function

In addition to dynamic range, an important parameter to consider is the modulation transfer function (MTF) of the complete system, which includes the well-known MTF of the lens and the pixel-to-pixel crosstalk of the sensor (optical and electrical). The importance of these parameters is usually underestimated by the end customer. For example, most HDR image sensors are unable to move from their saturation level to their minimum detectable level in less than five pixels. Here again it is advantageous to have larger pixels. Also, the race for smaller pixels does not take into account the diffraction limit of the lens. The diffraction limit is a function of the *f*-number of the lens, which is a function of the aperture of the lens (the limit is the Airy disk, defined for an aperture size of $D = 1.22\lambda$, where λ is the wavelength). For good HDR imaging, large pixels and large-aperture lenses are required. Large apertures can only be realized with larger lenses.

5.3 Conclusions

HDR lenses must have wide-spectrum antireflective coatings on both sides of every lens element to reduce lens glare. The lens barrel should be totally

non-reflecting inside so that it absorbs all deviated rays. The sensor should also be coated, and the first element of the lens should have a UV cut and an IR cut filter to make sure that useless wavelengths are filtered before entering the optical system. Due to the Airy disk limitation, parasitic wavelengths longer than the maximum useful wavelength (i.e., usually NIR light) require a larger lens aperture. In automotive HDR applications, to achieve exceptional HDR results, the price of the optics is usually in the range of one to five times the price of the sensor.

Chapter 6
Automatic High Dynamic Range Control

6.1 Automatic Exposure of Linear Sensors

6.1.1 Principle

In most linear response sensors, there are mechanisms for selecting the correct exposure and correct gain to provide the best exposure and noise tradeoff. These controllers act on exposure and gain settings to target a certain average image brightness for a complete image or for a region of an image (or a weighted average of several regions, sometimes called tiles). If the image is too dark, exposure time or gain can be increased. If the image is too bright, exposure time or gain can be decreased. The decision to work on gain or exposure depends on the situation and on the sensor. High gains should be avoided because of the resulting noise amplification and reduction of precision in exposure time settings due to enlarged steps. Exposure time variation is preferable when possible, i.e., until exposure reaches a maximum value, defined by the type of scene and frame rate requirements. Still scenes allow for long exposures, while moving objects require short exposures and a higher gain. Compared to photography, exposure time is the shutter time, and gain is the ISO speed.

An example of the control process follows:

```
if (brightness < target)
        if (current exposure time < maximum exposure time)
                increase exposure time,
        else
                increase gain,
else
        if (current gain > default gain)
                decrease gain,
        else
                decrease exposure time.
```

Another possible method is to predict the required exposure or gain to reach the target. In this case, a normal proportional-integral-derivative (PID) controller is used, and the theory of linear control applies. It is also possible to combine both techniques—for example, predicting the correct exposure when image brightness is far from the target, and using the incremental adjustment when the actual brightness is close to the target. This solution improves speed and stability.

6.1.2 Brightness calculation

The easiest method for estimating image brightness is to consider the average brightness of the image. Using the average is a time-consuming calculation, but good estimates can be obtained by using only the most significant bits of the pixel values and a subsampled version of the image.

An alternative is to use the image median brightness. The median can easily be estimated using a two-bin histogram, i.e., a histogram that only counts pixels that are above and below 50% (only the most significant bit of pixel data). Although this method of estimating the median is computationally more efficient and less memory demanding than estimating the average, it has the disadvantage of not being compatible with a predictive algorithm, as an unbalanced single-bin histogram will not indicate how far from the center the median actually is.

6.1.3 Filtering and stability for machine vision

There are two different cases in machine vision requirements. If an application requires several consecutive images to be acquired with the same exposure, the exposure time or gain changes in a sequence must be avoided. Oppositely, applications that are based on a single image can have variations of exposure time at every frame, the changes can be fast, and stability is less important. "Less important" means that small oscillations are not a problem (of course, the exposure should be correct and stable).

6.1.4 Filtering and stability for display

The human eye is very sensitive to brightness variation. Oscillations in image brightness (flickering) must be avoided. Controllers should not be too fast (so that stability can be maintained) and should be filtered for typical illumination frequencies due to the 50- or 60-Hz frequency of the supply (100- or 120-Hz flickering). A possible technique is to use the average brightness over several frames as input to the control algorithm, instead of the brightness of the last acquired image (low-pass filter).

6.1.5 Guard-band-based filtering

To avoid oscillations around the target position, guard bands are usually added around the target. Guard bands are similar to a Schmitt buffer (i.e., an electronic digital buffer with hysteresis).

For example, exposure time increases if the brightness is sufficiently lower than the target, and decreases if the brightness is sufficiently higher than the target. If there is a small variation of image brightness around the target brightness, this will not trigger an increase or decrease of the exposure time.

Guard bands are of high importance when the step size of the control signal is large (for example, the smallest possible change of the exposure time makes a large change in image brightness). They are of less importance if the control is nearly continuous.

6.2 Automatic Exposure of High Dynamic Range Sensors

Some HDR sensors do not require control at all, for example, the direct-current sensor that uses the logarithmic response. Other sensor responses have only one parameter to control and can therefore be controlled based on target brightness only. Difficulties arise when the response curve must be controlled by several parameters (several degrees of freedom). A two-parameter curve is common with, for example, multiple-segment responses.

Response curves that depend on more than three parameters are very difficult to control. The degrees of freedom can usually be reduced to two or three parameters by introducing relationships between the parameters so that control becomes easier. For example, multiple-segment responses have two degrees of freedom per segment: the slope and the kneepoint position. In multiple-segment response sensors (Fig. 6.1), the distance between the kneepoint is usually constant, and the slope ratio between segments is also kept constant. The degrees of freedom are then the total exposure, the number of segments, and the exposure ratio between two consecutive segments.

With a single degree of freedom, the state vector of the controller reduces to a single value: the average (or median) image brightness. With two degrees of freedom, the state vector has two elements: the average (or median) image brightness and the number of saturated pixels. The level over which the pixels are considered saturated is known from the sensor's

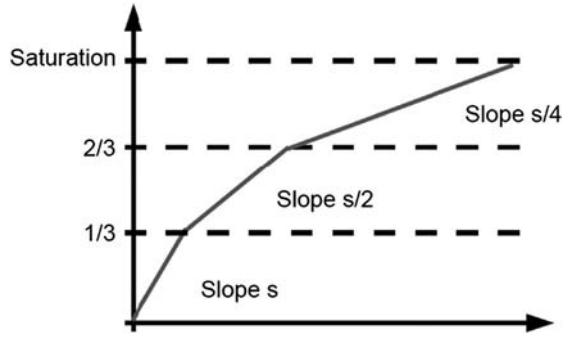

Figure 6.1 Reduction of the degrees of freedom in a piecewise linear regression sensor response curve with three segments.

characterization or for practical reasons is chosen arbitrarily as a certain percentage of the ADC range.

With three degrees of freedom, the state vector has three elements. The first two are the same as in the previous case, and the third one is an estimate of the total size of the dynamic range gaps (see Eq. (3.16)). The first two elements control the total exposure and total dynamic range, and the last one controls the number of segments of the curve. Using more segments for the same dynamic range provides a less-steep angle between two consecutive segments of the response, hence reducing the dynamic range gaps (SNR holes). Adding or removing a segment to the curve, even if the total dynamic range is preserved, may make the image look very different. Therefore, there are challenges in controlling the number of segments that can be partly solved by using a large hysteresis.

A major difficulty in the control of systems where multiple inputs affect multiple outputs is that the control inputs and outputs are all linked. Separating the variables so that one input affects only one output helps to reduce the complexity of the control loop and the stability analysis. Using the median of an image as input to the total exposure time control has the advantage of the median being almost independent of the response compression (dynamic range extension) because the compression of the bright pixels affects the average of the image, but (almost) not the median. The median is affected if there are kneepoints with strong compression located in parts of the pixel response that are lower (i.e., darker) than the median. The change in the median is always less than the change in the average.

6.3 Offset Control

Most sensors and cameras provide an analog offset control. This offset is either a summing amplifier and a DAC or a programmable lower level of the ADC. The offset is intended to maintain a black image regardless of any change in temperature or operating conditions. The offset is also required to sample the dark noise to characterize sensors, and it is a mandatory requirement of the EMVA1288 measurement standard.

When there is a non-zero offset, it has to be taken into account before regulating the image brightness and dynamic range, especially when the offset changes. Therefore, all statistical data used to control the exposure time, gain, and dynamic range of the sensor must be calculated after offset correction.

The offset setting is adjusted based on a black reference. This black reference is provided by black pixels that are not sensitive to light. Such pixels are usually metal-covered pixels on the side of the pixel array. The pixel array always has more pixels than the sensor's resolution because dummy pixels are required to provide a uniform design on all sides of the first real pixels along all four sides of the array. This more uniform design reduces the DSNU and

PRNU on the edges. They also partly shield the pixel array from bulk injected carriers of the surrounding electronics. The black reference pixels are usually within this dummy boundary.

Such black pixels unfortunately do not provide a good black reference. The metal covering the pixel changes several capacitances within the pixel and therefore changes its behavior. The metal coverage is also not a perfect light shield. Another approach places a small array of real pixels in the corner of the device; such pixels are sensitive to light but are not exposed to light as they are outside of the lens spot or covered by the sensor's package. Unfortunately, such pixels are also not a good black reference because they are too far from the pixel array and may behave differently. Both methods are even more difficult to implement in an HDR application.

Other techniques include pixels tied to a known voltage or pixels without an exposure time, but they fail to represent the correct black level at elevated temperatures due to dark current variations. Thus, they must be combined with diode-based temperature measurement devices. Such devices directly represent the dark current and can be calibrated to represent the actual junction temperature.

Chapter 7
High Dynamic Range File Formats

In memory, HDR images are treated with the highest possible precision. When it comes to saving an image or video data, it is important to find formats that reduce the file size to a minimum without losing the dynamic range or fine details. This tradeoff is similar to lossy versus lossless compression schemes: how to compress as much as possible without losing too many details.

Image encoding involves data formatting and data compression. A few data compression and data encoding methods are described here, but first, color space is introduced, along with the issues encountered when working with changes in color space.

7.1 Color Space

7.1.1 Introduction

As described earlier, the human eye and most sensors have a limited set of different color pixels, usually three (red, green, and blue). This means that they suffer from metamerism (i.e., they cannot differentiate two emission spectra that convolve the same way across the three types of detectors). The complete spectrum does not need to be encoded; only proportions of the three primary colors are required. The primaries must be chosen such that they are seen as distinct by the imaging system and should sufficiently differ so that a large number of distinct colors can be reproduced. If the primary colors are the vertices of a triangle, then the area of the triangle is maximum when the primary colors are far away from each other. The area of the triangle represents the colors that can be reproduced by the color space based on these primaries.

In practice, a negative amount of one or more primaries is required. The CIE XYZ color space is designed such that these negative amounts are not required, but at the expense of primaries that cannot be physically realized; they are called imaginary primaries.

As not all chromaticities (positions in the color space) can be reached, the device is said to have a restrained gamut. Gamut stands for "range of colors" and is a certain complete subset of colors. An imaging device is limited to a given color space. This means that even if the irradiance information can be precisely recorded from an HDR device, some color information is lost. The details of this color space discussion are not given here.

Figure 7.1 shows how three additive primaries can be combined to generate other colors. Using other proportions of the same three primaries creates other colors that belong to the same color space.

Figure 7.2 represents all possible chromaticities (CIE-1931 chromaticity diagram) and a limited color space. The three vertices of the triangle are the chosen primaries. The outer edges are the monochromatic colors of the spectrum represented by their wavelength. Monochromatic colors cannot be represented with these primaries. The chromaticity diagram only represents two dimensions (x and y) of the gamut. The third dimension (luminance z) comes out of the page, so that a gamut is actually a volume.

7.1.2 Color space definition

A color space is a set of formulae that define a relationship between a color triplet (or vector) and some standard color space, typically CIE XYZ or sRGB. A color space is also a set of boundaries on the volume defined by this vector, usually determined by the minimum and maximum value of each boundary, a range that is called the color gamut. A gamut is not a color space, but only part of the definition of a color space.

Color spaces can be defined to a given display so that no resources are wasted encoding color information that cannot be represented. This is called

Figure 7.1 Example of the combination of three additive primary colors.

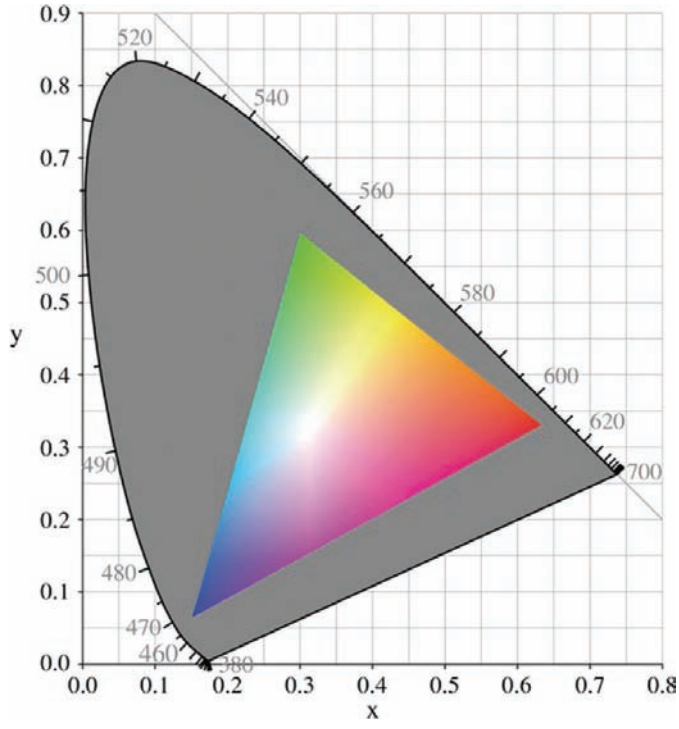

Figure 7.2 CIE-1931 representation of an arbitrary set of primaries (reproduced from CIE).

an output-referred color space. Alternately, color spaces can be referred to a scene or to an imaging device and are then referred to as scene-referred color spaces. Scene-referred color spaces require tone mapping to be presented in a display but have the advantage of not losing color information during encoding. This color information can then be retrieved if necessary, even if it cannot be displayed. Common color spaces are CIE-XYZ, sRGB, and Adobe® RGB. RGB, YUV, YCrCb, HIS, and CMYK are various color models based on the same color space. From an HDR point of view, it is clear that scene-referred color spaces should be used as much as possible. Sensor color space can be converted to a conventional color space (typically sRGB) using a color correction matrix, which can be obtained by matching a sensor image of a test chart with the expected image of the test chart.

7.2 Storing Image Data of Extended Dynamic Range Cameras

XDR cameras only provide compressed response output. There is no access to the raw linear data (except in the case of multiple sampling, scientific CMOS, or time to saturation, if the reconstruction is not performed inside the camera). In this case, image output is similar to the output of a linear camera

with 8, 10, 12, 14, or more bits, and standard image encoding techniques can be used (typically TIFF or 16-bit RAW files).

7.3 Storing Data of Radiance Maps and High Dynamic Range Software: Direct Pixel Encoding Methods

Direct pixel encoding means that each pixel is encoded independently based on its own value, or on its own value and on the value of its neighbors (such as in JPEG encoding). Direct encoding is discussed in this section, and gradient-based and flow-based encoding that do not use direct encoding are discussed in the next section. Frequently, when special encoding is used, the data are embedded in a TIFF container. However, it is shown that some formats have their specific file name.

7.3.1 IEEE single precision floating point

This is a standard floating point representation. It can easily represent the absolute radiance value, the relative radiance map, or the sensor dynamic range mapped between two limits. This is a raw type of format.

7.3.2 Pixar™ log encoding

Instead of the 8-bit gamma-based representation used in BMP, Lucasfilms™ (later Pixar™) digitalized films with a logarithmic 11-bits-per-channel format. Assuming no glare, a dynamic range of 3600:1 is reached. This is slightly more than 70 dB, and matches well with films that have a dynamic range of about 75 to 80 dB with logarithmic response. Because the dynamic range was so low, this historically first HDR format has never been used outside of Pixar.

7.3.3 Radiance RGBE

Three 8-bit mantissas for R, G, and B share a common exponent E. This format was developed by Greg Ward in 1985. The scaling factor is $2^{(E-128)}$. This format can reach up to 76 orders of magnitude with good precision (this is about 60 useless orders of magnitude). It has the disadvantage that the ratios between the colors must be multiples of 1/256. A derived format, called XYZE and using the XYZ color space, allows for complete coverage of the visible gamut.

Because the three components share the same exponent, the level of detail is the same in all three components. Therefore, a color consisting of one strong component and two weak components will make the weak component invisible. This is a general problem of any floating point representation with a common exponent. The finest level of detail that can be represented will correspond to one step of the mantissa multiplied by the value of the exponent.

7.3.4 SGI™ LogLuv TIFF

This format separates chrominance and luminance channels, such as in YUV or YCrCb, and treats them separately. It was developed in 1997 by Greg Ward to improve the RGBE format. Luminance is log encoded, and chrominance remains linear and is encoded following the CIE look-up table. It is used in TIFF libraries (three different versions exist).

7.3.5 Industrial Light and Magic™ OpenEXR

OpenEXR is based on a 16-bit floating point representation called 16-bit half data type. There is a sign bit, a 5-bit exponent, and a 10-bit mantissa. OpenEXR can represent positive and negative values of the primary colors, allowing it to cover the complete visible color gamut. It covers 10.7 orders of magnitude of dynamic range with a relative precision better than 0.1%. It also offers additional channels that can be used for alpha, depth, and spectral information.

This format is similar to the NVIDIA and ATI high-end formats and is thus directly supported by these graphic cards. This format is popular not only because of its performance, but also because Industrial Light and Magic (a division of Lucasfilms, Limited™) has offered a royalty-free C++ library to read and write the format.

7.3.6 Unified Color™ BEF

This format is specific to Unified Color Technologies (acquired by Pinnacle Imaging Systems in 2015). It covers a very wide dynamic range and the complete visible gamut.

7.3.7 Microsoft/HP™ scRGB

This format is based on the sRGB format, which allows for negative coefficients. It unfortunately only covers little more than 3.5 orders of magnitude of dynamic range with linear encoding, and 3.2 orders of magnitude with gamma encoding. The data can be encoded directly as RGB or as YCrCb.

7.3.8 JPEG XR

JPEG XR is also a Microsoft-driven standard, formerly named Windows Media Photo or HD Photo. JPEG XR is an improvement over standard JPEG, as it supports lossy and lossless compression and improved color accuracy (16 bits per pixel and per channel instead of 8). The perceived quality of the compressed image is similar to that of JPEG-2000, although its computation time is significantly less. The compression rate is higher than for standard JPEG for similar image quality perception. As with most special image formats, JPEG XR is embedded in a TIFF container. Strangely

Table 7.1 File format characteristics (from Ref. 52).

Format	Covers full visible gamut	Bits per pixel	Dynamic range (orders of magnitude)	Quantization step
sRGB	No	24	1.6	Variable
Pixar Log	No	33	3.8	0.4%
RGBE	No	32	76	1%
XYZE	Yes	32	76	1%
LogLuv 24	Yes	24	4.8	1.1%
LogLuv 32	Yes	32	38	0.3%
EXR	Yes	48	10.7	0.1%
scRGB	Yes	48	3.5	Variable
scRGB-nl	Yes	36	3.2	Variable
scYCC-nl	Yes	36	3.2	Variable

enough, Microsoft has released this format with an open license and allows its use in open-source software.

7.3.9 Summary of file formats

Table 7.1 summarizes the characteristics of the presented file formats.

7.3.10 Gen<I>cam

The Gen<I>cam standard of the European Machine Vision Association (EMVA) also defines several pixel formats for HDR industrial cameras, e.g., a 32-bit linear representation and a 32-bit floating point representation. Gen<I>cam is the most-used standard communication protocol between industrial cameras and computers.

7.4 Storing Data of Radiance Maps and High Dynamic Range Software: Gradient-Based and Flow-Based Methods

Gradient-based encoding uses the fact that the variation of brightness from one pixel to one of its neighbors is usually only a portion of the image's dynamic range. Relative variations are encoded when they are small; absolute pixel values are encoded when the relative variation is large. A pixel encoded with absolute value is called a key pixel. Key pixels are introduced every n pixel, even if the relative variation is small, to avoid a drift from the actual value by accumulating small errors.

Flow-based encoding is used for videos and is based on the fact that a given pixel does not vary greatly from one image to the next. Key frames (absolute pixel values) are introduced when there are major changes in the image or every n images. The other images are relative to the previous image or to the previous key frame and usually require fewer bits per pixel. It is possible to combine flow-based encoding with other encoding methods.

Chapter 8
Testing High Dynamic Range Sensors, Cameras, and Systems

It is important for manufacturers to check the performance of their sensors and cameras over their complete dynamic range to guarantee that the devices are in accordance with specifications. This check should be made in a characterization phase and in a production phase. Performance testing methods are discussed for non-HDR systems and possible extensions to HDR systems. It is important to note that testing of HDR systems is a very difficult task, as the dynamic range and accuracy of the testing equipment should be larger than those of the device being tested. As will be shown, there is still progress required in this field.

8.1 Testing of Software-Based Systems

Software-based systems use an HDR or LDR camera or sensor. The camera or sensor is characterized separately from the algorithm using standard camera performance measurement standards. The algorithm is tested by using known image sets or synthetic images.

8.2 Testing of Non-High Dynamic Range (Linear) Sensors and Cameras

There are two groups of standards related to performance measurement of cameras and image sensors. ISO standards come from the photography world. They are usually based on illuminated testing charts and are related to standard illumination sources (like D55) or standard observers. ISO standards use photometric units and measure the camera together with a lens. The EMVA1288 standard comes from the industrial world and is based on radiometric units; the camera is measured without a lens.

8.2.1 The ISO approach

ISO standards have several documents that describe how cameras can be characterized:

ISO12231: terminology,
ISO12232: determination of ISO speed,
ISO12233: resolution measurements,
ISO14524: methods for measuring optoelectronic conversion functions (OECFs),
ISO15739: noise measurements (Fig. 8.1), and
ISO9358: veiling glare of image forming systems, definitions, and methods of measurements.

Dynamic range is the contrast between the lightest and darkest point given in contrast, f-stop, or densities. ISO15739 defines dynamic range on the basis of SNR = 1. This is difficult to measure (as mentioned earlier). Therefore, a practical value of SNR = 3 is used. Dynamic range is determined from the OECF (using the OECF corrects for nonlinearities). To measure dynamic range, the lightest point is chosen at the illumination level where the camera reaches its maximum output value. The darkest point is the illumination level, where the signal-to-noise level passes the value of 3.

8.2.2 The EMVA1288 approach

The EMVA1288 standard is dedicated for the measurement of performance parameters of any camera or sensor, independent of its target application, and

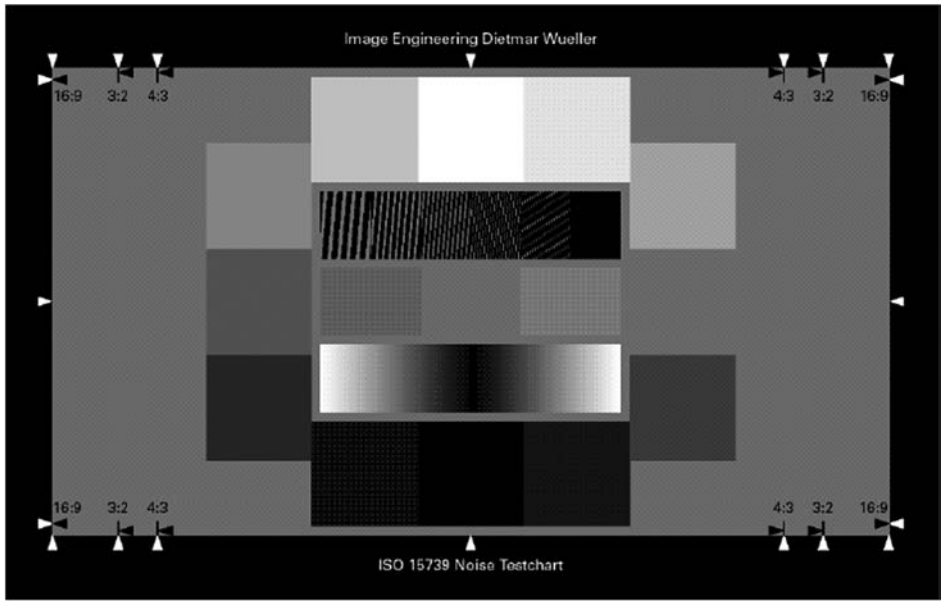

Figure 8.1 ISO15739 test chart modified for a contrast of 1000:1 (courtesy of Image Engineering).

Figure 8.2 EMVA1288 equipment (courtesy of Aphesa).

is well suited for machine vision equipment (Fig. 8.2). The current version of the standard is applicable to line scan cameras and area scan cameras (color or monochrome), but only for sensors with a linear response.

The measurement method is based on monochromatic and uniform illumination at *f*/8. Several images are acquired with increasing exposure (i.e., increasing exposure time or increasing light intensity) and are stored together with the corresponding dark (i.e., unexposed) images. Performance parameters such as responsivity, quantum efficiency, full well size, temporal dark noise, shot noise, dark FPN, PRNU, defect pixels, and many others are extracted from statistical calculations on this image set.

For more details on the EMVA1288 standard, read the free standard at www.standard1288.org.

8.3 Testing of High Dynamic Range Sensors and High Dynamic Range Sensor-Based Cameras

There is currently no standard describing how an HDR system should be measured, although the EMVA1288 working group is currently working on an extension of the standard. Therefore, only a few possible ideas are introduced. In the first two proposed solutions, the tested dynamic range is the

global intrascene dynamic range or the maximum achievable dynamic range, but not the local dynamic range. The next two proposed solutions introduce ideas about how the local dynamic range can be tested.

8.3.1 The ISO approach

The ISO standard can be extended to cover a larger dynamic range. This requires expensive test charts and light sources to reach the required dynamic range. Test charts (Fig. 8.3) with a contrast of up to 1000000:1 (120 dB) exist. They are sufficient to measure dynamic ranges up to 100 dB.

It is possible to combine two or more illuminated test charts side by side (some front illuminated and some back illuminated), or use some high-quality LCD displays to produce a wider dynamic range for testing cameras at a higher dynamic range.

8.3.2 The EMVA1288 approach

The EMVA1288 standard currently does not provide a solution for testing HDR sensors, but the technical working group is working internally on a solution for future releases of this standard. The philosophy of the measurement remains the same as for linear sensors, as it has already been proven that the math and models are valid.

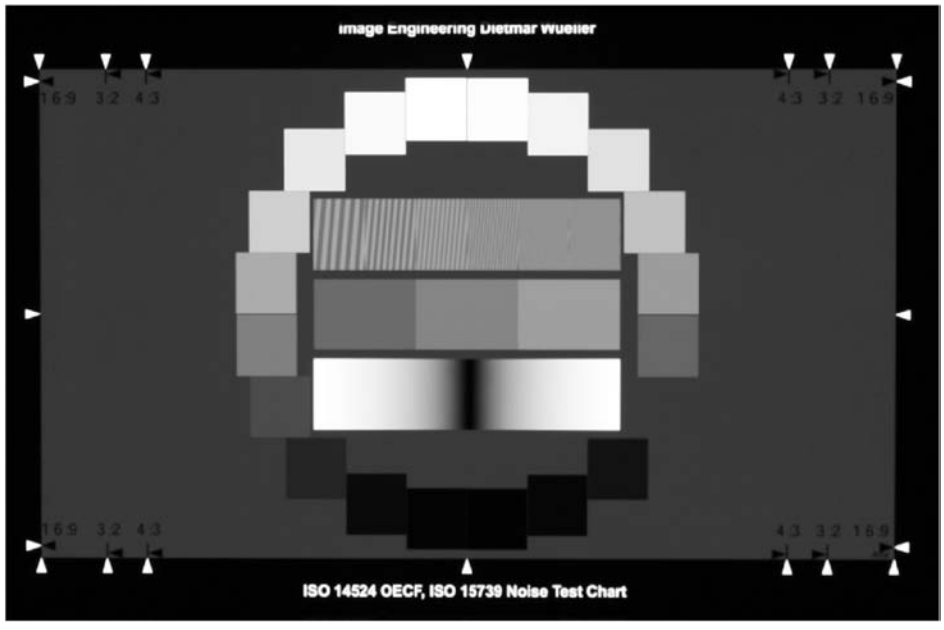

Figure 8.3 Back-illuminated HDR test chart at a contrast of 10000:1 (courtesy of Image Engineering).

8.3.3 Two-projector approach

Two projectors are used to project an image on a screen (Fig. 8.4). One of the projectors provides a high intensity, and the other projector provides a low intensity. The high-intensity projector is used to produce a large dynamic range, whereas the low-intensity projector is used to produce small intensity variations. A neutral density filter can be used to dim the low-intensity projector. A partially black cache or curtain can block part of the light from the high-intensity projector to provide better black levels. The two projectors produce their image on the same screen so that the images are superposed, and the result is an image that simultaneously has a large dynamic range and small variations of signal in both the dark and the bright areas.

Such a system can test for small variations of signal in a large overall dynamic range, but it can also provide a large local dynamic range. It also has the advantage of being able to vary the patterns quickly. This approach is limited by veiling glare.

8.3.4 Projector-and-display approach

A LCD or LED display can replace one of the projectors of the previous approach. It can be covered with a white cache (onto which the projector can generate its high-intensity pattern) and a black cache (where real black is available). If the computer display and the projector screen are side-by-side, then the capability to test the local high dynamic range is limited.

8.4 Contrast Detection Probability

The new IEEE P2020 automotive image-quality standard proposes to use a new method called contrast detection probability (CDP). Instead of separately

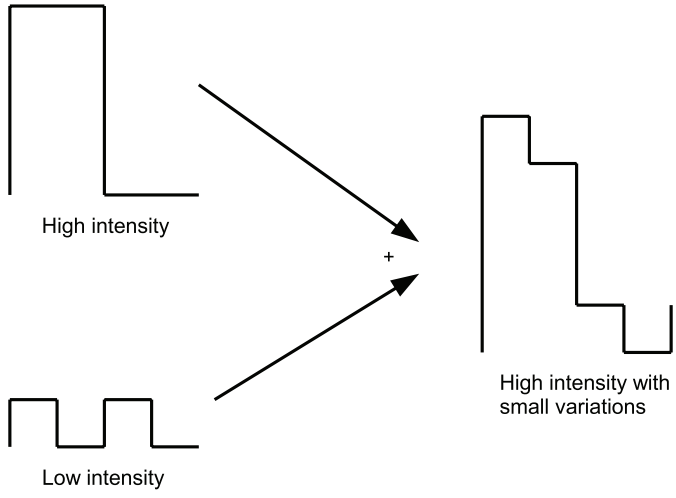

Figure 8.4 Two projectors used to test HDR cameras.

addressing the problems of sensors, cameras, lenses, and software, CDP considers that contrast loss through any element of the system under any circumstance, including HDR scenes or high-temperature operation, is what has to be measured. Contrast is measured in several areas of the scene with several types of patterns. This method has recently been proposed, but it has not been published as a standard at the time of this book's publication, and it still needs to be clearly defined.

Chapter 9
Dynamic Range in Non-Visible and 3D Imaging Devices

The book has thus far addressed dynamic range issues, concepts, and solutions for 2D visible-light image sensors. Dynamic range is of course a more general concept that also applies to human vision, audio, amplifier design, and other fields of engineering. This chapter considers non-visible and 3D imaging devices; these technologies will not be covered in detail, but rather it will be shown that dynamic range, HDR, and XDR exist in imaging outside the scope of this book.

9.1 Infrared Imaging

Although most of this text involves visible light and is also applicable to near-UV (NUV) and near-IR (NIR) wavelengths, longer wavelengths such as MWIR and LWIR (thermal imaging) use specific detectors and therefore specific techniques. (A significantly greater scene dynamic range is naturally available in longer wavelengths due to the much greater flux available at these wavelengths, i.e., orders of magnitude more photons/cm^2/s in the MWIR and LWIR. Higher dynamic range techniques are widely applied to infrared imaging for this reason.)

Infrared devices exist in two categories: temperature sensors and quantum devices. Both types of devices are very sensitive to their own temperature and the temperature of their package if the package is partly in the field of view. Therefore, the devices will only have good dynamic range and good performance if they are designed so that they are not affected by their own temperature. One such design technique, cooling, also improves the intrinsic performance of the detector and the noise.

Infrared devices are hybrid modules made of a detector wafer (a dedicated material such as InGaAs, HgCdTe, or PbS; micromachined structures) and a readout wafer. Therefore, the sensing node design and the readout circuit design are separate and can be optimized individually. Infrared detectors

make good use of the wafer stacking design and can achieve very good circuit performance in terms of noise, linearity, and dynamic range.

9.2 3D Imaging

3D imaging (or more correctly, 2.5D imaging, because only the outer surface of an object is seen) is divided into active 3D and passive 3D. The former involves the use of a support light source, either a pattern to measure displacements (structured light, triangulation) or a pulse of light to measure the travel time (time of flight, lidar).

Active 3D relies on the detectability of the useful signal within all of the other light sources. It is particularly difficult to use in the presence of the sun and over long distances. The sun is very powerful, and the light intensity of the emitted useful pulse of light decreases with the square of the distance. The measurement of the small variations of the useful signal within the large parasitic light signal is a dynamic range problem.

Time-of-flight systems are 3D range cameras that provide a depth image of a scene based on the round-trip time of flight of the light from the camera to the scene and back.

- The first step to solve the problem filters the parasitic light. Because the useful light is monochromatic, a narrow-bandpass filter is added in front of the detector to suppress any parasitic light outside the band of the emitted light source.
- For the second step, several pixel-design and signal-phase-detection techniques have been invented to select the useful signal out of the total signal. It is important to ensure that the remaining parasitic signal does not saturate the detector before the phase of the useful signal has been detected.

Chapter 10
Conclusions

For more than 30 years, various solutions have been found to the problem of HDR imaging. It is only more recently that significant progress has been made in the field of rolling-shutter and global-shutter CMOS image sensors, which provide a direct HDR image output without software processing and with good image quality.

Several XDR sensor types have been discussed that can be useful in many HDR applications, usually where a low-processing application or a fast solution is necessary, or when the scene is moving relative to the camera. Each sensor type has advantages and disadvantages. It has been shown how the performance of XDR sensors can be measured or calculated using SNR, dynamic range factor, and information, and how sensors can be compared to select a sensor for an application. The most useful types of sensors today are multiple-segment response sensors and those using advanced logarithmic pixels. For higher performance, such sensors usually require larger pixels; this is not compatible with the ongoing race for higher sensor resolution. 3D silicon (wafer stacking) will create new opportunities for complex XDR pixel designs.

Software methods were then discussed (which are useful in photography of still scenes), and one special example of a software-based high-speed machine vision application was given. Some sensors provide a simplified version of this software method for real-time operation and a small increase in dynamic range. Dedicated hardware platforms provide limited real-time solutions. It was also seen that optical path should be designed carefully, especially regarding optical veiling glare and MTF.

Now that sensors are capable of 150 dB or more using various design techniques, and that software methods exist to go beyond 180 dB, it is time to work on the improvement of other components of these systems such as control algorithms, optical path, and testing and characterization processes.

Finally, file formats compatible with HDR images were briefly discussed, along with testing methods for HDR cameras and systems. HDR issues in non-visible and non-2D imaging were also introduced.

10.1 Important Figures of Merit of a High Dynamic Range Sensor

HDR sensors and cameras cannot be specified and characterized in the same way as linear CMOS and CCD sensors; specifying only the type of response curve and the dynamic range for some operating points is not enough. It is highly recommended that suppliers provide the SNR and DR mathematical models developed in this book so that users can anticipate a device's performance for various settings, especially the potential presence of dynamic range gaps (SNR holes).

If an automatic exposure control with HDR control is provided, the tradeoffs in parameter adjustments should be explained in the device's datasheet. It is recommended that the EMVA1288 characterization document and mathematical models of signal, noise, SNR, and dynamic range be requested from every sensor or camera manufacturer, HDR or not.

10.2 Questions

- If a sensor has a full-well size of 10,000 electrons, which type of ADC should be selected?
- Is an XDR pixel larger than an LDR pixel?
- Is an XDR sensor larger than an LDR sensor?
- Can a sensor with a 12-bit ADC have a dynamic range of 180 dB?
- Can software methods generate a higher-dynamic-range image than hardware methods?
- If a sensor has a maximum reachable dynamic range of 120 dB, can a 120-dB scene be properly acquired within one image? Always? With only one sensor setting? Never?
- If the power supply of the board on which a sensor is assembled is noisy because of some switching regulators, will this affect the SNR of the imaging system? Will it affect the dynamic range of the imaging system?
- If, for cost reasons, lenses are made of injected plastic in an M12 plastic barrel, will this affect the SNR of the imaging system? Will it affect the DR of the imaging system?
- When combining n images using the software method, is the relation $\mathrm{DR}_{total} = \sum_{i=1}^{n} \mathrm{DR}_i$ correct?

Bibliography

Pixel designs and HDR pixels

1. M. Waeny, S. Lauxtermann, N. Blanc, M. Willemin, M. Rechsteiner, E. Doering, J. Grupp, P. Seitz, F. Pellandini, and M. Ansorge, "High-sensitivity, high-dynamic, digital CMOS imager: Paper B," *Proc. SPIE* **4306**, 78 (2001) [doi:10.1117/12.426992].
2. B. Fowler, "High dynamic range image sensor architectures," Stanford Univ. WDR Workshop (Sep. 2009).
3. T. Lulé, B. Schneider, M. Bohm, et al, "Design and fabrication of a high dynamic range image sensor in TFA technology," *IEEE J. Solid State Cir.* **34**(5), 704–711 (1999).
4. W. Yang, "A wide dynamic range low power photosensor array," *Proc. IEEE ISSCC* **37**, 230–231 (1994).
5. O. Vietze, "Active pixel image sensors with application specific performance based on standard silicon CMOS processes," PhD Thesis, No. 12038, ETH Zurich, Switzerland, (1997).
6. D. X. D. Yang, A. El Gamal, B. Fowler, and H. Tian, "A 640×512 CMOS image sensor with ultrawide dynamic range floating-point pixel-level ADC," *IEEE J. Solid State Circ.* **34**(12), 1821–1834 (1999).
7. T. F. Knight Jr., "Design of an integrated optical sensor with on-chip preprocessing," PhD Thesis, MIT, Boston, MA (June 1983).
8. M. Barbaro, P.-Y. Burgi, A. Mortara, P. Nussbaum, and F. Heitger, "A 100×100 pixel silicon retina for gradient extraction with steering filter capabilities and temporal output coding," *IEEE J. Solid State Circ.* **37**(2), 160–172 (2002).
9. O. Yadid-Pecht, "Wide-dynamic-range sensors," *Opt. Eng.* **38**(10), 1650–1660 (1999).
10. B. Dierickx, N. Witvrouwen, and B. Dupont, "DC and AC high dynamic range pixels," Image Sensor Conf., London (March 2008).
11. J. Huppertz, R. Hauschild, B.J. Hosticka, T. Kneip, S. Miler, and M. Schwarz, "Fast CMOS imaging with high dynamic range," Proc. Workshop CCD and Advanced Image Sensors, R7-1–R7-4, Bruges, Belgium (June 1997).

12. M. Schanz, C. Nitta, A. Bumann, B. J. Hosticka, and R. K. Wertheimer, "A high dynamic range CMOS image sensor for automotive applications," *IEEE J. Solid State Circ.* **35**(7), 932–938 (2000).
13. M. Schanz, W. Brickhende, R. Hauschild, B. J. Hoticka, and M. Schwarz, "Smart CMOS image sensor arrays," *IEEE Trans. Electron Dev.* **44**, 1699–1705 (1997).
14. O. Yadid-Pecht and A. Belenky, "In-pixel autoexposure CMOS APS," *IEEE J. Solid State Circ.* **38**(8), 1425–1428 (2003).
15. O. Yadid-Pecht and E. Fossum, "Wide intrascene dynamic range CMOS APS using digital sampling," *IEEE Trans. Electron Dev.* **44**, 1721–1723 (1997).
16. B. Fowler, D. Yang, A. El Gamal, and H. Tian, "A CMOS image sensor with ultrawide dynamic range floating-point pixel-level ADC," *IEEE J. Solid State Circ.* **34**, 1821–1834 (1999).
17. K. Brehmer, S. J. Decker, R. D. McGrath, and C. G. Sodini, "A CMOS imaging array with wide dynmaic range pixels and column-parallel digital output," *IEEE J Solid State Circ*. **33**, 2081–2091 (1998).
18. R. Hauschild, M. Hillebrand, B. J. Hosticka, J. Huppertz, T. Kneip, and M. Schwarz, "A CMOS image sensor with local brightness adaptation and high intrascene dynamic range," *Proc. ESSCIRC*, 308–311 (1998).
19. E. Funatsu, Y. Nitta, Y. Miyake, T. Toyoda, K. Hara, H. Yagi, J. Ohta, and K. Kyuma, "Artificial retina chip with a 256×256 array of n-mos variable sensitivity photodetector cells," *Proc. SPIE* **2597**, 283–291 (1995) [doi:10.1117/12.223990].
20. J. Trumblin, A. Agrawal, and R. Raskar, "Why I want a gradient camera," *Proc. IEEE CVPR* (2005), see http://www.umiacs.umd.edu/~aagrawal/1323GradCamFinal.pdf (last accessed July 2012).
21. V. Branzoi, S. K. Nayar, and T. Boult, "Programmable imaging using a digital micromirror array," *Proc. IEEE CVPR*, 436–443 (2004).
22. S. K. Nayar and T. Mitsunaga, "High dynamic range imaging: spatially varying pixel exposures," *Proc. IEEE CVPR*, 472–479 (2000).
23. E. Fossum, "CMOS acive pixel image sensor," *Nucl. Instrum. Methods A* **395**, 291–297 (1995).
24. E. Fossum, "Image capture circuits in CMOS," *VLSI Tech. Syst. Appl.*, 52–57 (1997).
25. S. Kavusi and A. El Gamal, "Quantitative study of high dynamic range image sensor architectures," *Proc. SPIE* **5301**, 264 (2004) [doi:10.1117/12.544517].
26. A. Uehara, K. Kagawa, T. Tokuda, J. Ohta, and M. Nunoshita, "A high-sensitive digital photosensor using MOS interface-trap charge pumping," *IECE* **1**(18), 556–561 (2004).
27. L. G. McIlrath, "A low-power low-noise ultrawide-dynamic-range CMOS imager with pixel-parallel A/D conversion," *IEEE J. Solid State Circ.* **36**(5), 15 (2001).

28. A. El Gamal, "High dynamic range image sensors," *Proc. IEEE ISSCC* (2002).
29. B. Dierickx, D. Scheffer, G. Meynants, W. Ogiers, and J. Vlummens, "Random addressable active pixel image sensors," *Proc. SPIE* **2950**, 1 (1996) [doi:10.1117/12.262512].
30. G. Meynants, B. Dierickx, D. Scheffer, and A. Krymsky, "Sensor for optical flow measurements based on differencing in space and time," *Proc. SPIE* **2645**, 108 (1996) [doi:10.1117/12.236090].
31. D. Scheffer, B. Dierickx, and F. Pardo, "Design and fabrication of a CMOS log-polar image sensor," *Proc. SPIE* **2784**, 2 (1996) [doi:10.1117/12.248520].
32. N. Ricquier and B. Dierickx, "Active-pixel CMOS image sensor with on-chip non-uniformity correction, Proc. IEEE Workshop on CCD and Advanced Image Sensors, 20–21 (1995).
33. N. Ricquier, I. Debusschere, B. Dierickx, A. Alaerts, J. Vlummens, and C. Claeys, "The CIVIS sensor: a flexible smart imager with programmable resolution," *Proc. SPIE* **2172**, 2 (1994) [doi:10.1117/12.172755].
34. N. Ricquier and B. Dierickx, "Smart image sensors for robotics and automation," 3rd ESA ASTRA Workshop, Estec, Noordwijk, NL (17–18 April 1994).
35. N. Ricquier and B. Dierickx, "Pixel structure with logarithmic response for intelligent and flexible imager architectures," *Microelectron. Eng.* **19**, 631 (1992).
36. D. Yang and A. El Gamal, "Comparative analysis of SNR for image sensors with enhanced dynamic range," *Proc. SPIE* **3649**, 197 (1999) [doi:10.1117/12.347075].
37. Y. Ni, Y. Zu, and B. Arian, "A 768 × 576 logarithmic image sensor with photodiode in solar cell mode," New Imaging Technologies SA, Verrières le Buisson, France, see http://www.new-imaging-technologies.com/media/doc/isw2011.pdf (last accessed July 2012).
38. Y. Ni and K. Matou, "A CMOS log image sensor with on-chip FPN compensation," see http://www.imec.be/esscirc/esscirc2001/Proceedings/data/24.pdf (last accessed July 2012).

Software HDR methods, tone-mapping algorithms, and perception

39. F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high dynamic range images," Proc. ACM SIGGRAPH, 257–265 (2002).
40. E. Reinhard, M. Stark, P. Shirley, and J. Ferwada, "Photographic tone reproduction for digital images," *Proc. ACM SIGGRAPH* **21**(3) (2002).

41. S. Mann and R. Picard, "Being 'undigital' with digital cameras: extending dynamic range by combining different exposure times," *Proc IST*, 422–428 (1995).
42. S. Winder, S. B. Kang, M. Uyttendaele, and R. Szeliski, "High dynamic range video," *Proc ACM SIGGRAPH* **22**(3), 319–325 (2003).
43. S. K. Nayar and V. Branzoi, "Adaptive dynamic range imaging: optical control of pixel exposures over space and time," *Proc. IEEE* **2**, 1168–1175 (2003).
44. M. D. Grossberg and S. K. Nayal, "Determining the camera exposure from images: what is knowable?," *IEEE Trans. Patt. Anal. Mach. Intell.* **35**(11), 1455–1467 (2003).
45. Y. Y. Schechner and S. K. Nayar, "Generalized mosaicing: high dynamic range in a wide field of view," *Int. J. Comput. Vision* **53**(3), 245–267 (2003).
46. M. Aggarwal and N. Ahuja, "High dynamic range panoramic imaging," Proc. of International Conference on Computer Vision (ICCV), 2–9 (2001).
47. R. Mantiuk and H.-P. Seidel, "Modeling a generic tone mapping operator," *Eurographics* **27**(2) (2008).
48. E. Reinhard and K. Devlin, "Dynamic range reduction inspired by photoreceptor physiology," *IEEE Trans. Visual. Comput. Graphics* **11**, 15–24 (2011).
49. R. Mantiuk, "High dynamic range imaging: towards the limits of the human visual perception," *Forschung und wissenschaftliches Rechnen* **72**, 11–27 (2006).
50. See http://www.openexr.com (last accessed July 2012).
51. See http://www.unifiedcolor.com (last accessed July 2012).
52. G. Ward, "High Dynamic Range Image Encodings," see http://www.anyhere.com/gward/hdrenc/hdr_encodings.html (last accessed July 2012).
53. E. Reinhard, E. Arif Khan, A. Oguz Akyüz, and G. M. Johnson, *Color Imaging Fundamentals and Applications*, AK Peters Ltd, Boca Raton, FL (2008).
54. A. O. Akyüz and E. Reinhard, "Noise reduction in high dynamic range imaging," *J. Vis. Commun. Image R.* **18**(5), 366–376 (2007).
55. J. J. McCann, "Art science and appearance in HDR images," *J. Soc. Info. Display* **15**(9), 709–719 (2007).
56. J. J. McCann and A. Rizzi, "Camera and visual veiling glare in HDR images," *J. Soc. Info. Display*, **15**(9), 721–730 (2007).
57. P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," *Proc. ACM SIGGRAPH*, 369–378 (1997).
58. M. Ashikhmin, "A tone mapping algorithm for high-contrast images," 13th Eurographics Workshop on Rendering, 1–11 (2002).

59. R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," *Proc. ACM SIGGRAPH*, 249–256 (2002).
60. R. Mantiuk, K. Myszkowski, and H. P. Seidel, "A perceptual framework for contrast processing of high dynamic range images," *ACM TAP* **3**(3) (2006).
61. S. Pattanaik and H. Yee, "Adaptive gain control for high dynamic range image display," *Proc. SCCG*, 24–27 (2002).

Measurement and characterization methods

62. European Machine Vision Association 1288 Standardization Group, "EMVA1288 standard 2010 release 3," see http://www.standard1288.org (last accessed July 2012).
63. F. Dierks, *Comparing Cameras using EMVA1288*, Basler AG, Ahrensburg, Germany (2006).
64. D. X. D. Yang and A. El Gamal, "Comparative analysis of SNR for image sensors with enhanced dynamic range," *Proc. SPIE* **3649**, 197 (1999) [doi:10.1117/12.347075].
65. S. O. Otim, D. Joseph, B. Choubey, and S. Collins, "Modelling of high dynamic range logarithmic CMOS image sensors," *Proc. IEEE Instrum. Measure.* **1**, 451–456 (2004).
66. J. R. Janesick, *Photon Transfer*, SPIE Press, Bellingham, WA (2007) [doi:10.1117/3.725073].
67. D. Hertel, "Extended use of incremental signal-to-noise ratio as reliability criterion for multiple-slope wide dynamic range image capture," *J. Electron. Imag.* **19**, 011007 (2010) [doi:10.1117/1.3267100].
68. D. Wueller and U. Artmann, "Test methods for digital still cameras," *Test Methods for Digital Still Cameras*, Image Engineering, Frechen, Germany, see http://www.imageengineering.com/ (last accessed July 2012).

General HDR

69. A. Darmont, "Methods to extend the dynamic range of snapshot active pixels sensors," *Proc. SPIE* **6816**, 681603 (2008) [doi:10.1117/12.761600].
70. A. Wilson, "Dynamic design," *Vision Syst. Design* **14**(10), 35–39 (2009).

General image sensors

71. A. Darmont, "Spectral response of CMOS image sensors," Aphesa, Harze, Belgium (2009).
72. G. C. Holst and T. S. Lomheim, *CMOS/CCD Sensors and Camera Systems*, Second Edition, JCD Publishing, Winter Park, FL (2011).

73. H. Tian and A. El Gamal, "Analysis of 1/f noise in CMOS APS," *Proc. SPIE* **3965**, 168 (2000) [doi:10.1117/12.385433].
74. M. T. Rahman, N. Kehtanaravaz, and Q. R. Razlighi, "Using image entropy maximum for auto exposure," *J. Electron. Imag.* **20**(1), 013007 (2011) [doi:10.1117/1.3534855].
75. J. C. Dunlap, W. C. Porter, E. Bodegom, and R. Widenhorn, "Dark current in an active pixel complementary metal-oxide-semiconductor sensor," *J. Electron. Imag.* **20**(1), 013005 (2011) [doi:10.1117/1.3533328].

Photography and optics

76. L. Johnes and H. Condit, "The brightness scale of exterior scenes and the computation of correct photographic exposures," *J. Opt. Soc. Am.* **31**, 651–678 (1941).

Miscellaneous

77. S. Maddalena, A. Darmont, and R. Diels, "Automotive CMOS image sensors," 9th International Forum on Advanced Microsystems for Automotive Applications (AMAA), 401–412 (2005).
78. A. Darmont, "CMOS imagers: from consumer to automotive," MST News (June 2004).
79. W. Karwowski, *International Encyclopedia on Ergonomics and Human Factors, Second Edition*, Vol. **1**, 1484, CRC Press, Boca Raton, FL (2006).
80. S. Masumura, "Carry out the mission of light, basic and application of lighting technology in machine vision system," Series on Lighting Technology, Eizojoho Industrial, Sangyo Kaihatsukiko Inc., Tokyo, Japan (April 2004).
81. K. Devlin, "A review of tone reproduction techniques," Tech. Report CSTR-02-005, Dept. Computer. Science, Univ. of Bristol, UK (2002).
82. L. Straniero, see http://www.flickr.com/photos/24630856@N08 (last accessed July 2012).
83. B. Dierickx, "Course on CMOS image sensors," Caeleste (2009), see http://www.caeleste.be/publicaties_bart/publicatielijst.htm (last accessed July 2012).
84. C. Lofqvist, see page http://www.flickr.com/photos/43052603@N00 (last accessed July 2012).
85. R. Xu, et al., "High dynamic range image and video data compression," *IEEE Computer Graphics and Applications*, **25**(6), 69–76 (2005).
86. B. Jähne, *Digitale Bildverarbeitung*, Springer, Berlin (2005).

Index

Arnaud Darmont (1979–2018) held a degree in Electronic Engineering from the University of Liège (Belgium, EU) oriented towards imaging science; he worked for over fifteen years in the field of imaging, HDR imaging, camera design, machine vision, and camera characterization. He authored several publications and patents in the field of HDR imaging, automotive on-board imaging, image processing applications, and camera test and characterization.

After almost 7 years developing automotive HDR CMOS image sensors at Melexis, he founded Aphesa in 2008. Aphesa specialized in image sensor consulting, HDR imaging, custom camera design, custom image processing solutions, image sensor and camera characterization and benchmarking. Aphesa was one of the main contributors to the European Machine Vision Association (EMVA)1288 standard, and Darmont was a member of the EMVA1288 working group since almost the beginning of the standard. Aphesa merged with Deltatec in 2017, and Darmont served as business development manager and technical expert.

Since December 2017, he was also the manager of standards at the EMVA, has launched new standards, and coordinated the international development of industrial imaging standards. He was a contributor to the IEEE P2020 automotive image quality standard and the IEEE P4001 hyperspectral imaging standard.

He was an SPIE instructor since 2009 and provided image sensor and industrial imaging courses to companies since 2015.