ugr | Universidad de Granada

Multispectral High Dynamic Range Polarimetric Imaging applied to scene segmentation and object classification

by

MIGUEL ÁNGEL MARTÍNEZ DOMINGO

B.Sc., Telecommunications Engineering, University of Málaga

M.Sc., Erasmus Mundus Master CIMET

A dissertation

submitted in partial fulfillment of the requirements for the

DOCTOR OF PHILOSOPHY

at the

UNIVERSIDAD DE GRANADA

(University of Granada)

Doctoral Programme: Física y Ciencias del Espacio

Editor: Universidad de Granada. Tesis Doctorales
Autor: Miguel Ángel Martínez Domingo
ISBN: 978-84-9163-359-4
URI: http://hdl.handle.net/10481/47628

El doctorando / The doctoral candidate ***Miguel Ángel Martínez Domingo*** y los directores de la tesis / and the thesis supervisors: ***Javier Hernández Andrés y Eva María Valero Benito,*** garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

I guarantee, by signing this doctoral thesis, that the work has been done by the doctoral candidate under the direction of the thesis supervisors and, as far as our knowledge reaches, in the performance of the work, the rights of other authors to be cited (when their results or publications have been used) have been respected.

Lugar y fecha / Place and date: Granada, 14 de febrero de 2017.

Directores de la Tesis / Thesis supervisors:

Doctorando/ Doctoral candidate:

Javier Hernández-Andrés

Miguel Ángel Martínez Domingo

Eva María Valero Benito

Multispectral High Dynamic Range Polarimetric Imaging applied to scene segmentation and object classification

Miguel Ángel Martínez Domingo

Abstract

Different advanced techniques of digital imaging such as multispectral imaging, high dynamic range (HDR) imaging, polarimetric imaging or Near-Infra-Red imaging, have been developed and applied separately for years. Researchers are trying to merge some of these techniques together into a single integrated system. However this integration is rather challenging, specially if we are dealing with general purpose applications, such as capturing outdoor urban or natural scenes.

This dissertation proposes capturing system designs, as well as algorithms and processing techniques for improving and simplifying the systems currently present in the state of the art of these different imaging techniques. This way, high dynamic range multispectral polarimetric images in the visible and near infrared can be captured and processed for many applications such as image segmentation, objects or materials classification, vegetation monitoring, food inspection, remote sensing, surveillance, etc.

A new multispectral image capturing system is proposed, based on a novel generation of sensors which are still under development. Based on simulations, this work takes advantage of the spectral tunability of these sensors, and combines it with color filter arrays, to propose an imaging system with 36 spectral channels, achieving very good colorimetric and spectral performance for spectral reflectance estimation.

Besides, a new algorithm for the automatic capture of HDR images is proposed, called Adaptive Exposure Estimation (AEE). It can be implemented in any digital imaging system, and it works online, as the capturing is ongoing. It is adaptive to scene content without the need of any prior knowledge about the scene being captured. The proposed method allows the user to tune the

performance of the algorithm, keeping the balance between exposure time and signal-to-noise ratio, by just adjusting two free parameters. It can also capture the full dynamic range of the scene (or region of interest), or just a part of it.

The proposed AEE algorithm is also adapted to multispectral polarimetric image capture. Based on a previous work which uses a Liquid Crystal Tunable Filter, a new full framework for capturing and processing 31-channels MultiSpectral HDR Polarimetric (MSHDRPol) images is proposed. New techniques for segmentation and classification of objects present in indoors scenes are proposed and tested. The results show that the algorithm outperforms other methods proposed in previous studies.

As an additional contribution, the whole capturing workflow is adapted to an 8-channels filter-wheel-based imaging system covering the visible and NIR ranges up to 1000 nm. Therefore a system and a framework able to automatically capture MultiSpectral HDR Polarimetric Visible and Near Infra-Red (MSHDRPolVISNIR) images of outdoor scenes are proposed.

A set of 8 outdoors scenes have been captured using the proposed system and methods and they will be made publicly available after the defense of this doctoral thesis.

Nuevas técnicas de adquisición y procesado de imágenes HDR multiespectrales y polarimétricas. Aplicación a segmentación y clasificación de objetos relevantes.

Miguel Ángel Martínez Domingo

Resumen

Durante años, muchas técnicas avanzadas de captura de imágenes digitales como imágenes multiespectrales, de alto rango dinámico (HDR), polarimétricas o en el infrarrojo cercano, han sido desarrolladas y utilizadas por separado. Muchos investigadores intentan unir estas técnicas en un solo sistema integrado. Sin embargo esta integración es bastante complicada, especialmente si estamos apuntando hacia una aplicación de carácter general, como captura de imágenes urbanas o naturales al aire libre.

Esta tesis propone diseños de sistemas de captura, así como algoritmos y técnicas de procesado para mejorar y simplificar los sistemas que actualmente podemos encontrar en el estado del arte de las diferentes técnicas avanzadas de imagen mencionadas. De este modo, se pueden capturar y procesar imágenes multiespectrales, de alto rango dinámico, polarimétricas en el visible e infrarrojo para usarlas en diferentes aplicaciones como segmentación de imágenes, clasificación de objetos o materiales, monitorización de vegetación, inspección de alimentos, teledetección, vigilancia, etc.

Se propone un nuevo sistema de captura de imágenes multiespectrales, basado en una nueva generación de sensores en desarrollo. Mediante simulaciones, este trabajo se beneficia de las propiedades sintonizables de dichos sensores, que combinados con arrays de filtros de color, nos permiten alcanzar un sistema de 36 canales espectrales con un rendimiento colorimétrico y espectral destacable.

Además, se propone un nuevo algoritmo llamado Adaptive Exposure Estimation (AEE, Estimación de Exposición Adaptativa), para la captura automática de imágenes HDR. Este algoritmo puede ser implementado en cualquier sistema de captura de imágenes digitales. Funciona online, durante la captura, y se adapta al contenido de la escena sin necesidad de obtener ninguna información previa acerca de ésta. El método propuesto permite al usuario sintonizar su

rendimiento, eligiendo el equilibrio entre tiempo de captura y relación señal a ruido, simplemente ajustando el valor de dos parámetros. También puede capturar el rango dinámico completo de la escena (o solo de una región de interés), total o parcialmente.

El algoritmo AEE propuesto se ha adaptado a la captura de imágenes multiespectrales y polarimétricas. Basado en un trabajo previo que usaba un filtro sintonizable de cristal líquido, se propone un nuevo marco de trabajo completo para la captura y el procesado de imágenes multiespectrales HDR polarimétricas de 31 canales. También se proponen nuevas técnicas de segmentación y clasificación de objetos presentes en escenas interiores. Se ha comprobado que dichas técnicas funcionan de manera más robusta que las anteriormente propuestas por otros autores.

Como siguiente paso, todo el marco de trabajo propuesto para la captura se adapta a un sistema de captura de imagen basado en una rueda de filtros de 8 canales. Este sistema es capaz de capturar información de imagen de hasta 1000 nm de longitud de onda. Por tanto el sistema y métodos propuestos permiten la captura de imágenes multiespectrales, HDR polarimétricas en el visible e infra-rojo cercano de escenas en exteriores.

Se ha capturado un conjunto de 8 escenas en exteriores usando el sistema y los métodos propuestos. Dichas escenas serán publicadas para el libre acceso de la comunidad científica tras la defensa de esta tesis doctoral.

Contents

The people who are crazy enough to think they can change the world, are the ones who do.

Steve Jobs.

1

Introduction

Content:

Chapter 1

Introduction

1.1 Motivation

Since the emergence of the first imaging cameras in the nineteenth century, the main aim of most imaging systems has been the perfect mimicking of the human visual system (HVS), making the captured images as realistic as possible. Scientists and engineers have tried ever since to capture and display images which are the closest possible to what the human observers would see when capturing the scene. This aim has always come together with an artistic purpose. Not only making the images captured faithful, but also to make them pleasantly looking and engaging.

On the other hand, more recently, in parallel with this aim, and thanks to the evolution of imaging systems, science has tried to take advantage of such systems, by using them to analyze different physical phenomena by means of captured image data. The need to see more than what the eye can see has gained increasing appeal. This purpose has risen specially after the appearance of digital imaging systems in the twentieth century. Images became numbers, the language of science and digital machines. Hence, scientists could use computers to perform image data analysis with a potential that humans could have never imagined before.

As curiosity is one of the main engines of science, the developers of imaging systems have not resigned themselves to just mimicking what the HVS could do. They always wanted to overcome the limitations of our own vision, as well as the limitation of new incoming imaging systems. This fact triggered the appearance of other imaging technologies which could be considered advanced compared with the traditional monochrome or color (RGB,

from Red, Green and Blue) imaging systems. It was here when techniques such as High Dynamic Range imaging (*HDR*), MultiSpectral imaging (*MS*), Polarimetric imaging (*Pol*) or Infrared imaging (*NIR*) were developed.

Each of these techniques alone poses a challenge in itself. The limitations of classical imaging systems force the scientist to change the architecture of image capturing systems. All of these techniques have some potential for helping to detect or analyze some objects of interest. Combining them to build a more complete image capture system is more challenging than dealing with each of them individually. However, when we make any change in order to achieve the goals of one of these techniques, we could accidentally lead further from the goals of the rest of them. Besides, compared with classical imaging systems (namely mainly photographic cameras), special imaging systems are not used by so many people in the world. Some of the solutions proposed can be rather costly and complex. They might be not easy to manage or involving a lot of pre and post capture image processing, thus making these kind of advanced systems rather unique.

It is clear that consumer electronics are rather useful for millions of people in the world, specially imaging consumer electronics. With the appearance of digital cameras and lately with smartphones featuring photographic cameras with a very high image quality for a relatively low price, every subject has an imaging device in its pocket, and storing, sharing and processing the image data captured is so fast and easy that the potential gets infinite.

For this reason, science is always interested in using these consumer-oriented systems for its purposes. Nevertheless, the hardware of these imaging systems is usually a monochrome sensor with an *RGB* Bayern-pattern filter on top of it. Designing new cameras at sensor level, is a complex and costly process which could lead to the loss of generality of the system. It means that each single imaging application could need for the development of one new sensor architecture which could be useful only for this application in particular and not for any other. Thus, it is always interesting to find a different approach using the same sensors and cameras, adding or removing outer hardware elements (like color filters, polarization filters or *NIR* cut-off filters) which are cheaper and easier to attach or detach. This would make the camera still valid for a different application. Besides, the fast and powerful *CPUs* and *GPUs* inside any computer or mobile device, make it possible to compensate the lack of capability of the hardware, with computationally demanding image processing methods. Hence, there is always a balance between hardware and software. A very complex hardware

would probably reduce the need of image processing, at the cost of narrowing down the potential applications of a designed system. On the other hand, a simple hardware would need more complex image processing techniques, though the system could still be used for different applications just by changing the software and/or the attached hardware.

The research team at the Color Imaging Laboratory of the University of Granada [148], has carried out research into both classical Colorimetry (e.g. color differences) and Color Vision (e.g. chromatic discrimination) since the beginning of the 1970s. In the 1990s it became interested in both human and computational color constancy ([160, 186]). From 1990 to 2000 the group decided to direct its research attention towards color images [76], while maintaining open the research lines about human vision related to color constancy and perception of chromatic gratings [40]. The research group became interested in the mathematical characterization of objects and illuminants in scenes by obtaining linear models for the representation of spectral reflectances of various different types of artificial objects, particularly acrylic paints and the Spectral Power Distribution (*SPDs*) of natural and artificial illuminants ([41, 77, 76]). The group used several algorithms based upon the use of conventional *CCD* (Charge-Coupled Device) cameras, with or without the addition of color filters in front of the lens [147]. After 2000, work has been done in computational color constancy, applying the fundamental concepts relating to the phenomenon of color constancy in the development of color-descriptors invariant to changes in illumination applicable to color imaging [146]. In addition, the group is studying the statistical properties of natural color images and the influence of daylight illumination on the statistical descriptors such as color contrast, gamut, etc [187].

At present, the group is looking into the design of optimum sensors to be used in the spectral acquisition of *HDR* color images. Based in color and/or spectral information of images, we also work in the development of algorithms for recognition of objects in natural scenes, dehazing methods and design and optimization of multispectral color scanners. In the immediate future we intend to continue along these lines of research by improving our methods of acquiring multispectral and *HDR* images both outside and in the laboratory and by applying color and multispectral techniques to several fields, like industrial inspection systems and characterization of goniochromatic materials. We are also exploring other imaging techniques such as polarimetric or infrared imaging. We still continue with experimental work into the Applied Colorimetry and Atmospheric Optics.

1.2 Structure of the work

It is in this framework where the inception of this thesis takes place. We wanted to study the present imaging technologies such as multispectral, *HDR* and *NIR* imaging, and during the study we also incorporated polarimetric imaging. We aimed to find cutting edge limitations in these fields and try to propose new alternative systems or methods to overcome them. Consequently, we performed an extensive state of the art research in these different technologies, which is written in chapter 2. Afterwards we dug into a new sensor technology under development (*Transverse Field Detectors or TFDs*) to propose a new multispectral imaging system. These sensors offer a great potential for multispectral imaging thanks to their tuning capabilities. The experimental work based on simulations as well as the results and conclusions are presented in chapter 3.

After proposing a new multispectral image capturing system, we moved on to *HDR* imaging techniques and proposed a software method to solve the problem of estimating exposure times when capturing *HDR* images via multiple exposures. We made our method universal so that it can be implemented in any imaging system as long as it can control the exposure time and it is able to perform easy computations like the cumulative histogram. The proposed method, the experiments performed and the results obtained are presented in chapter 4.

The next step was trying to merge both technologies into a single system. Since *TFDs* were in a very early stage of research and we could not use still a photographic camera featuring these sensors to capture real images yet, we moved on to a different multispectral imaging system which could allow us to change spectral responsivities easily, and therefore make us able to experiment with different possibilities.

In chapter 4, we used different standard cameras for implementing our proposed *HDR* capturing technique. Afterwards, in chapter 5, we extended this technique to multispectral systems.

We chose a Liquid Crystal Tunable Filter (*LCTF*) attached to a monochrome camera, which works in the visible range and polarizes light through it, and a filter wheel camera that allowed us to capture images in both the visible and *NIR* ranges of the spectrum, adding an extra polarizing filter in front of it. The first system has spectrally narrow bands, and was used to capture 31-channels indoors scenes with multiple common objects. These

images were used to perform image segmentation and object classification with a completely new proposed work-flow (compared with the one proposed in [179]). The second system has broader spectral bands and was used to capture 8-channels outdoors urban scenes. The description of the systems as well as the experiments made and results obtained are written in chapter 5. These last systems are the examples of merging different advanced imaging technologies in one imaging system, by applying hardware and software techniques, allowing us to retrieve MultiSpectral High Dynamic Range Polarimetric image data pixel-wise in the visible and *NIR* range of the spectrum (*MSHDRPolVISNIR*. See the scheme in figure 1-1).

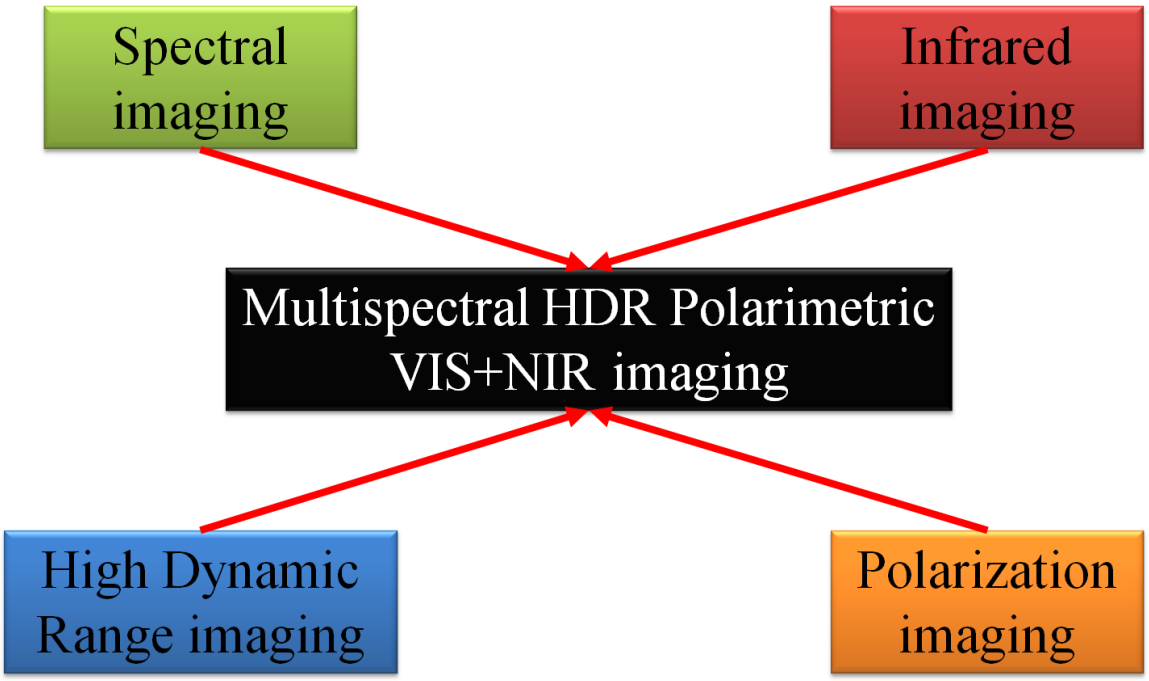

Figure 1-1: Scheme of different technologies merged in this thesis into a single imaging system.

Finally, in chapter 6, we summarize the conclusions reached in every step of this thesis, as well as the lines of future work to explore after it.

We summarize here the main contributions of this thesis:

- We have studied the state of the art in the different advanced imaging technologies such as spectral imaging, *HDR* imaging, polarimetric imaging, and near infrared (*NIR*) imaging.
- We have designed through simulations a new multispectral image capturing system, based on a new generation of *CMOS* silicon sensors that are still under development.

Combining them with custom designed color filter arrays (*CFA*), we have proposed a system able to capture up to 36 spectral channels in a fraction of a second.

- We have proposed a new algorithm to automatize the capture of *HDR* images through multiple exposures. Our method is the only one to our knowledge that adapts to any scene content with no a priori known information about it, and automatically captures the whole dynamic range of the scene on-line. Besides, our method offers the possibility to be tuned in order to balance the ratio: capturing time/ *SNR*.

- We have adapted our proposed *HDR* image capture technique to multispectral and polarimetric imaging systems, making the capture of *MSHDRPol* images fully automatic.

- We have improved a previously proposed framework for segmentation and classification of objects in indoors scenes by using *MSHDRPol* images.

- We have captured a database of *MSHDRPolVISNIR* images of outdoors scenes using the same capturing strategy proposed for indoors scenes.

The science of today is the technology of tomorrow.

Edward Teller.

2

State of the art

Content:

Chapter 2

State of the art

In this thesis, the state of the art is organized in different sections. Each section corresponds to one of the main technologies and techniques that have been studied. We analyze first how each of these technologies are developed apart. Later on we also study the attempts that some authors have made to join two or more of these technologies together into a single image capturing and processing unified framework.

This chapter is organized as follows: in section 2.1 we review the main architectures used for spectral image capture. In section 2.2 we explain the systems and techniques used to capture and process *HDR* images. In section 2.3 we describe those image capturing systems which capture image information about the polarization state of light. Later on, in section 2.4 we talk about the working principles of those systems which already combine some of these techniques together. Finally, in the last section of this chapter (section 2.5), we review the state of the art in near infrared imaging systems.

2.1 Spectral imaging systems

This section intends to be a summary of the different kinds of techniques and systems that make up the state of the art in the field of spectral imaging, rather than a exhaustive literature review with all existing spectral imaging systems. We tackle the different philosophies and approaches to get to retrieve spectral information pixel-wise in a digital image captured from a scene.

We call monochrome image to a digital image consisting in only one band of information, which actually is a two-dimensional matrix of gray values. A color image or *RGB* image

is a digital image formed by three of these monochrome images taken in different spectral bands (usually spectrally broad bands roughly corresponding to the Red, Green and Blue areas of the spectrum respectively). A multispectral image is a digital image formed by more than 3 bands, but less than 10. These bands could be broader or narrower spectrally speaking depending on the nature of the system itself. If the system has more than 9 spectral bands we can call it a hyperspectral system. Of course the border between multispectral and hyperspectral is not so well defined, and depends on the literature and authors we are studying. It is not only about the number of bands but also about how broad or narrow the spectral channels are, or the processing required to get the final spectral information pixel-wise, etc ([66]). Normally hyperspectral information refers to data from spectra that are sufficiently well sampled to be considered like a continuous spectral signal. This definition is somehow fuzzy since it depends on the nature of the spectral signal itself. If the signal is totally flat or it changes smoothly with wavelength (see figure 2-1 left), few bands are needed to represent it all along the spectrum. However, if the signal has very narrow spikes, like the ones which are typical from artificial light sources (see figure 2-1 right), then a high number of narrow bands would be required to represent the full spectrum keeping faithful information for every single wavelength. Therefore, the limits between multispectral and hyperspectral images, data or systems are not so sharp.

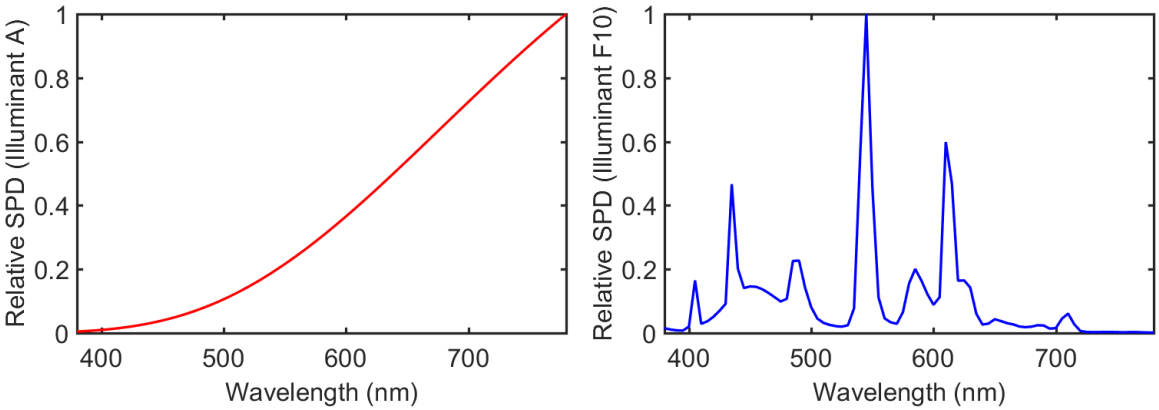

Figure 2-1: Spectral Power Distribution for two *CIE* standard illuminants. Left: Illuminant *A* (incandescent). Right: Illuminant *F10* (fluorescent). These two spectral signals are good examples of a smooth and a spiky spectral functions.

There are many different types of spectral imaging systems. We could do a classification regarding many aspects of their functioning ([42]). However, as a first level distinction we divide them into two different classes: direct measuring systems (subsection 2.1.2) and indir-

ect measuring systems (subsection 2.1.2). We develop these two categories in the following subsections.

In figure 2-2, we see a scheme with the different spectral imaging systems explained later.

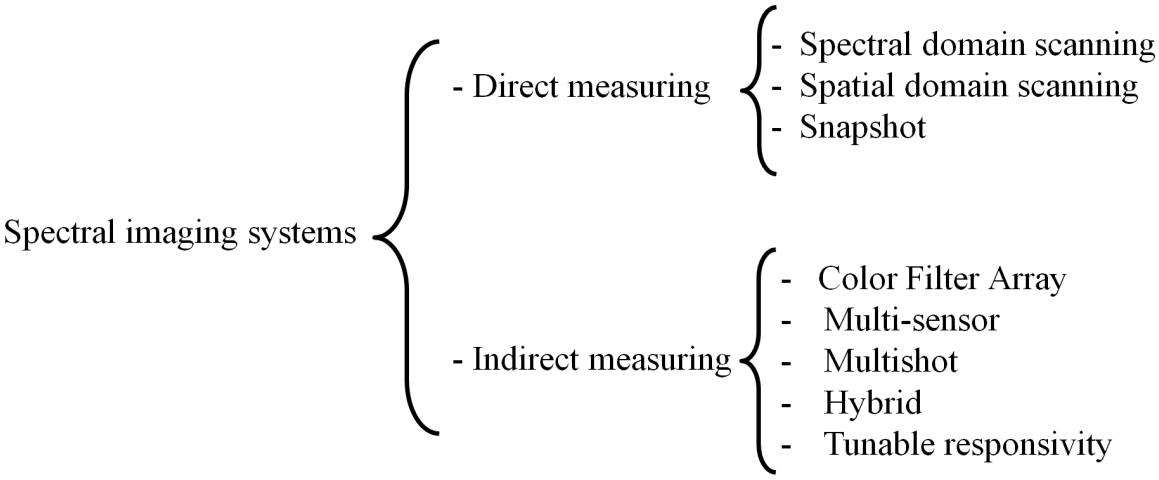

Figure 2-2: Scheme of the spectral imaging systems classification explained in this section.

2.1.1 Direct measuring systems

This kind of systems are usually hyperspectral systems. The name of direct measuring systems does not imply that the information they capture does not need a post-processing before it is useful, but rather that after the capture, we already have spectral information in enough wavelengths so that we can know the signal in its whole spectrum. There is no need of interpolation or estimation of spectral information in any wavelength because we already have information in every wavelength after the capture. These systems usually imply the use of bulky imaging devices and/or time consuming processes including data scanning along either the spectral or spatial axes. For this reason, we subdivide this section into three sub-categories: spectral domain scanning systems, spatial domain scanning systems, and snapshot systems ([42]).

Spectral domain scanning systems

To scan an image in the spectral domain, we need a system which is able to separate the light coming from each point of the scene into its spectral bands. For this purpose, these systems consist usually in a digital monochrome camera to which we attach some color filters. These filters are narrow band-pass filters that only allow a very narrow spectral component of the

light to pass through, rejecting the rest of the spectrum. Therefore, the way of scanning a whole hyperspectral cube, is to take several shots of the same scene changing the wavelengths that are allowed to pass through the filter. This fact implies that during the capturing time (the sum of all exposure times for every wavelength plus the needed change of filter), the objects in the scene must not move at all. Otherwise very complex image registration techniques would be needed to be applied in order to align properly the spectral information for each single point in the scene, which would be spread over many different pixel positions in the different band images. This is in practice not feasible since the registration depends on the movement of the objects for each scene among other things. Thus we assume that these kind of systems are meant to be used for static scenes. In practice it would also not be very practical to have a bunch of narrow band-pass color filters (like interference filters) in a big filter wheel to mechanically change them to capture each single band. This would increase a lot the time required for the whole capturing process, as well as introduce some possible sources of error due to the mechanical moving parts such as a filter wheel or any other kind of filter holder (vibrations, displacements, misalignments, etc). The very first hyperspectral devices were based on filter wheels, and they were used for different applications such as: designing a multispectral scanner for print or artwork inspection, ([81, 103, 21]), but then the technology evolved to tunable filters.

To overcome these issues, there are filters which have the ability to be electronically tuned, so their peak band-pass wavelength can be switched all along the spectrum. These filters are called tunable filters, and there are mainly two models depending on the functioning principles: Acousto-Optic Tunable Filters ($AOTF$), used for spectroscopy, or remote sensing imaging ([15, 55]), and Liquid Crystal Tunable Filters, used in applications such as: fluorescence Raman microscopy, spectral reconstruction of different color object surfaces for object recognition, or biomedical and industrial applications ($LCTF$[136, 65, 181, 2]). Based on physical properties of some special materials (like quartz for $AOTF$ or liquid crystals for $LCTF$), the filters can be electronically controlled to switch their pass wavelength applying some signal to these materials. In the case of the $AOTF$, the signal applied is an acoustic signal (sound waves) that depending on its frequency (usually always around radio-frequency) can change the diffraction properties of the quartz. Therefore changing the acoustic frequency applied to the sound signal, we can change the wavelength to which the filter is tuned. In the case of the $LCTF$, the process is more complex involving polarization

as well as induced birefringence [195, 2].

There are also some other technologies used for the spectral imaging in which the scanning is done spectrally. Recent developments of spectral imaging technology use Bragg gratings to disperse a complete image over a sensor. An example of such systems is shown in figure 2-3.

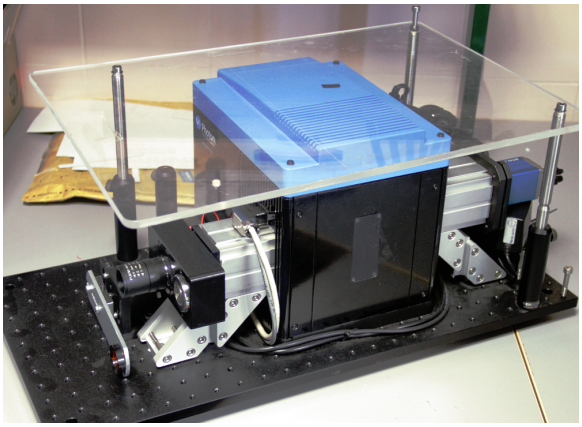

Figure 2-3: Photon Etc Hyperspectral imager in the visible and near infrared based on Bragg gratings.

Changing the angle of reflection in the diffraction grating, the peak wavelength of light impinging in the sensor changes as well. As some applications for these systems we can find material clustering or outdoors spectral reflectance image capture [159, 28]. Besides the time consuming problem of the previous systems commented in this section, it also has the problem that there is a wavelength gradient in each image taken. So, what ideally is supposed to be a monochrome image of each spectral band that only contains information from one wavelength, actually has information from different wavelengths in a gradient as we move away from the central vertical line of the image. Therefore, image registration and gradient correction is required after the whole spectral cube is captured in order to put together the image information from each wavelength in a single plane of the final spectral cube.

To finish with in this subsection, it is also worthy to mention the imaging systems that, instead of varying the spectral responsivities of the sensors by applying different transmittance filters in front of them, they vary the wavelength of the light which is illuminating the scene. Some possible application of these system are: material segmentation, spectral re-lightning, color image rendering, spectral reflectance recovery or tristimulus imager

[149, 178, 177].

These systems are called multiplexed illumination systems. For this purpose it is common to use *LED* or laser based lightning. *LED* and laser spectra are quite narrow spectrally and we can find them almost in any wavelength. Thus, with a proper illuminating system composed by many different *LEDs* or a tunable laser, and a monochrome camera we can spectrally scan a scene to get full spectral information out of it. Nonetheless this kind of systems only work for controlled conditions and cannot be used to acquire outdoors scenes where we can not control the illumination. If the number of *LEDs* used is enough to get full spectral information we can consider this system a direct measuring system. Otherwise, if it only has few *LEDs* and there is a need of estimating the spectral data to get full spectrum information, then we could include this system into the indirect measuring systems.

Spatial domain scanning systems

The spatial scanning consists of getting full spectral information of a partial area of the image (not the whole image) and then do the same for a different region of the scene, until we have the full spectral information from the whole scene. Usually, rather than by areas or pixels, this spatial scanning is done line by line. Therefore, these systems are line scanners which are able to disperse the polychromatic light impinging on them and record each wavelength component of each point of the line separately. In order to scan spatially, there is a need of moving either the object being scanned or the imaging system itself. In this category, one of the most popular systems is the *Specim's Imspector* (see figure 2-4).

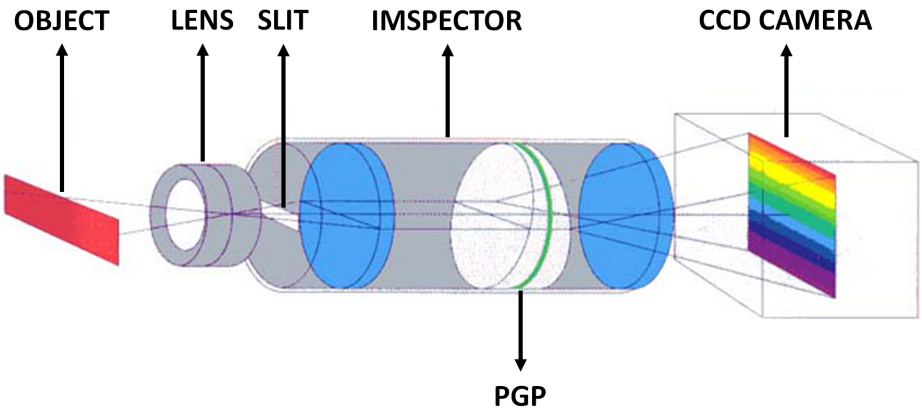

Figure 2-4: Scheme showing the working principle of the *Specim Imspector*, based on a Prism-Grating-Prism configuration (PGP).

This device is used for applications such as: industrial quality control or food inspection [79, 150]. This system images a line of the scene for instance in horizontal direction, and disperses the light of each point of the line in vertical direction by an optical element called PGP (prism-grating-prism), in such a way that this dispersed light impinges on a bi-dimensional sensor. Therefore in the horizontal direction of the sensor we have spatial dimension and in the vertical direction we have spectral dimension. If we do this operation line by line we get the full spectral information of a complete scene. This system requires a very precise set up, in which the spatial scanning is perfectly controlled and synchronized with the capture. Therefore it is more recommendable for controlled conditions.

In this category we can also find indirect measuring systems. One example of them is the 12 channels scanning system *TruePixa* ® developed by the German company *Chromasens GmbH*, and used for applications like in-line color measurement for print inspection, or spectral reflectance estimation [47, 30, 29] (see figure 2-5).

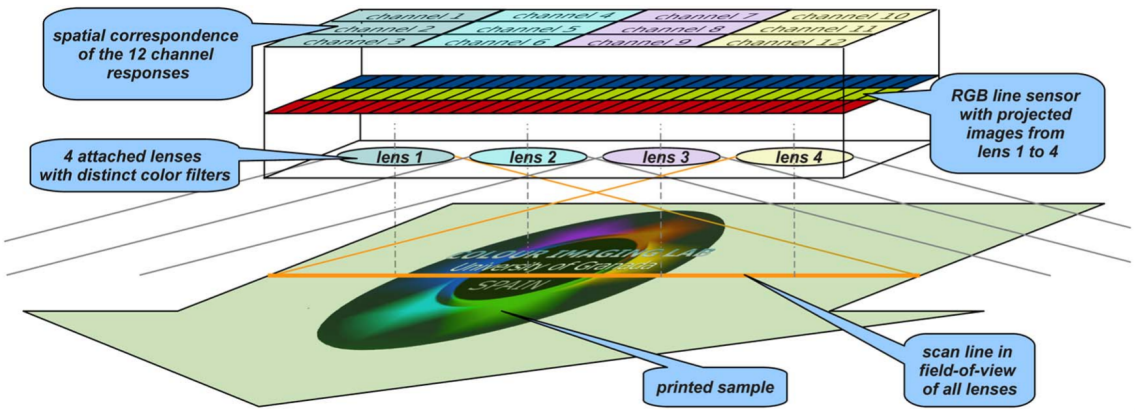

Figure 2-5: Scheme of multispectral line scanner system used in [47, 30, 29]. Image retrieved from [32].

In this system, instead of dispersive elements, they use an *RGB* line scanning sensor divided in 4 parts (each part with independent optics). These parts cover slightly different areas of the object being scanned. However there is a common area which can be registered to align the different images captured after the spatial scanning. Placing 4 different multispectral filters, one on top of each area of the sensor, and given that the line sensor is 3 color channels sensor, they get an image with 12 spectral channels. After applying spectral estimation methods, they can recover the spectral image information of the scene [31, 32]. This system is included in this category because it scans the scene spatially. However, since

the final spectral image data is obtained after spectral estimation methods, it could also fit in section 2.1.2.

Snapshot systems

Spectral imaging techniques have been developed which directly acquire high spectral resolution images in a single exposure for applications like hyperspectral microscopy or industrial applications ([39, 45, 44]). However this type of spectral imaging system suffers from inherent shortage of light power because the power of the incident light is shared both spatially and spectrally. These systems have optical components that spectrally decompose the light coming from very small areas of the scene. Therefore, in the bi-dimensional sensor we get both bi-dimensional spatial information and spectral information. They use pupil arrays and lenses arrays to re-image the dispersed areas of the scene in the sensor. Due to this, the images acquired with these sensors are not very large (up to 285x285 pixels so far) and need powerful controlled illumination. This is the reason why these systems are mainly used for spectral microscopy. In figure 2-6, we see the optical layout proposed in [39], and the image mapping principle to retrieve the spectral information from the two spatial dimensions.

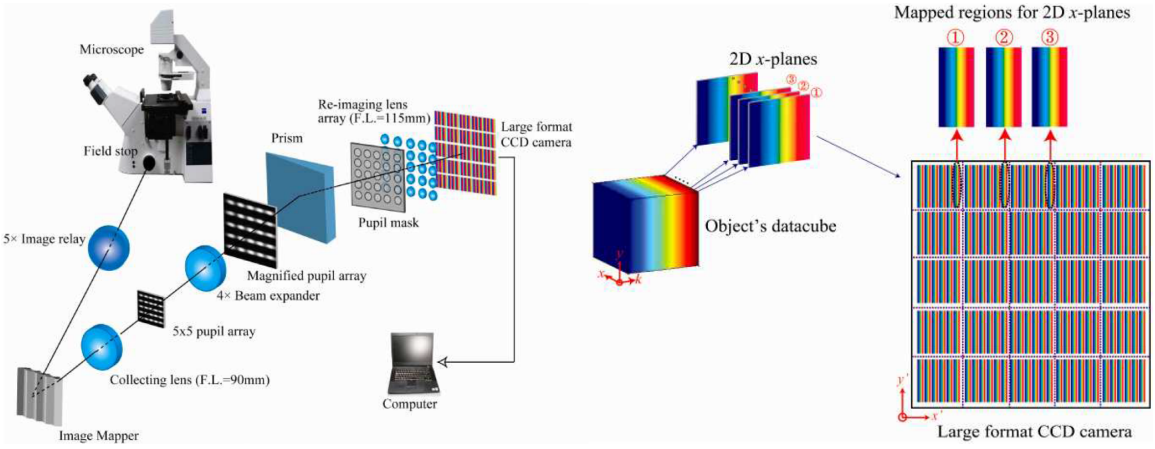

Figure 2-6: Left: optical layout of snapshot system proposed for microscopy. Right: image mapping principle to retrieve spectral image information. Images retrieves from [39].

2.1.2 Indirect measuring systems

If in the previous subsection we have seen systems that acquire full spectral data images, in this section we are going to talk about systems that acquire digital images with lower number of bands and then process the data to obtain, by estimation, the full spectral data

pixel-wise. This section, in turn, could be divided into the systems themselves, the devices, which can be very varied in philosophy and architecture, and the estimation techniques, that are basically algorithms and mathematical tools for mapping the vectorial space responses of the system into hyperspectral space (full spectral data). In this chapter however we intend to focus only in the first part.

To start with, we have to say that we could use any digital camera as a spectral imaging system. We can take any digital *RGB* camera, shoot pictures, and try to map using some techniques to spectral data. Of course, the quality of the resulting spectral data will not be desirable. Some authors studied these limitations for specific environmental applications like sugar cane monitoring in trial plots [112]. Thus, researchers try to improve the acquisition elements of the spectral imaging chain in order to make easier to retrieve full spectral data of good quality with not very complex estimation algorithms. Usually only 3 channels are not enough to estimate properly the full spectral data from digital images. Therefore, the developers look for ways to increase the number of channels. But the physical limitations of imaging systems make this a challenging and non-trivial task. There are many different approaches. In the end, a common imaging sensor is a bi-dimensional matrix of pixels where the imaging systems has to concentrate light forming an image that the sensor electronics will record. However, before reaching the sensor, the light can be pre-filtered, modulated, deviated, dispersed, etc., so that the pattern shone in the sensor is actually a complex pattern of information that, with the proper post-processing can be converted into useful multispectral data to feed the spectral estimation algorithms. Finally, the output of these algorithms would be the full spectral information pixelwise.

This subsection is in turn divided into five different categories: Color filter array systems, Multi-sensor systems, Multi-shot systems, Hybrid systems and the recently developed Tunable responsivity systems. We briefly explain them in the following subsections.

Color filter array (CFA) systems

These are probably the most common type of multispectral system. Most digital color cameras belong in this group. The basic idea is that they trade off spatial resolution to gain number of spectral bands. All pixels in the sensor have the same spectral responsivity covering the whole spectral range we are interested in. But each pixel of the sensor has a certain color filter over it, which can have broader or narrower transmittances depending on

the application we are interested in. This filter pattern is the so called Color Filter Array (*CFA*). Since the system does not acquire the information from every band in every pixel, in order to recover it, there is a need to apply interpolation algorithms called demosaicking [53].

Each camera has different patterns and each developer uses its own algorithms for this task, which most of the times are protected by industrial property regulations. In common color cameras, this filter pattern is called Bayer pattern filter ([152]) and it is composed by three different transmittances, roughly corresponding to Red, Green and Blue regions of the visible spectrum. They usually have some overlapping. However *CFAs* can have more than 3 channels. The higher the number of channels, the lower the spatial resolution is. In [142], the authors proposed a 16 channels *CFA* system (see figure 2-7) for multispectral imaging of natural scenes. This system uses 3 sensors, each of them with a color filter on top of it (red, green or blue). Splitting the incoming light in 3 beams, and adding the *CFA*, they get up to 48 channels in one shot. They present the drawback of demosaicking and low spatial resolution if the number of spectral channels is high. But they are relatively cheap and can be useful alternatives for some applications.

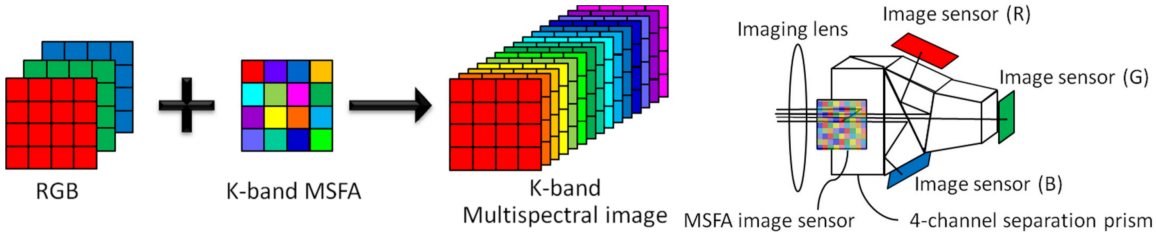

Figure 2-7: *CFA*-based system with 3 sensors. Image retrieved from [142].

Multi-sensor systems

Systems in this group try to avoid the demosaicking process and the loss of spatial resolution typical of *CFA*-based capture systems. They do so by adding more sensors and including optical elements to split the input light in order to form images from a single band in each sensor. The advantage is clear, there is no need for demosaicking to get the full information from each band. The disadvantages are that they are less robust since the light splitting has to be very well set up, in order to avoid misalignments between the images captured by the different sensors, and also they are more expensive. Besides, if we want to include

an independent sensor for each spectral band, we are limited by the size, price and amount of light impinging on each sensor (the more times we divide the beam of light, the less light each beam has). In [43] and [167], the authors use a 3-*CCD* camera by *Fluxdata*, that combined with *CFA* can yield up to 7 channels in one shot. Also the company *Quest Innovations* provides a 5-*CCD* camera composed by 3 channels in the visible part of the spectrum (*RGB*) and 2 channels in the near infra red (*NIR*, [25]).

Another approach for this category is that proposed in [166] for general spectral imaging applications. Instead of using a beam splitter, they use a stereo camera meant for 3D imaging. This camera has two independent objectives and sensors (the same as using 2 different cameras). In front of each objective, there is a different color filter which spectral transmittance function is comb shaped, so that the *RGB* spectral responsivities of each sensor are split in two and they get 6 channels covering the whole spectrum in a single shot. In the design of this system, the main challenge is to select a combination of comb-shaped filters to get a set of 6 responsivities out of the original 3 ones that performs best for the spectral reconstruction. Of course such combination will depend on the estimation method used to get the spectral data, as well as on the error metrics used to asses it and of course on the application.

This idea of stereo cameras is pushed beyond by other authors. In ([183]), they built a system with 9 cameras attached together used for digital archiving for cultural heritage preservation. Each of these cameras is not an *RGB* camera but monochrome, and has a different color filter in front of it to get 9 different channels out of the same monochrome responsivity.

Using stereo systems with two or more cameras together has the drawback of image alignment. Even though, when the cameras are properly set up, the images taken by the different cameras are very similar, their perspectives are slightly different. Thus they still need to be aligned with respect to a reference image. Besides, there could be occlusion effects happening when a 3D scene is being imaged. On the other hand, these effects are useful in order to retrieve depth information from the image (which actually is the original purpose these stereo cameras where meant for. [184]).

Multi-shot systems

Systems in this group reach a trade-off between the previous two groups. In the one hand they only have one sensor, but in the other hand they do not need demosaicking since what they do, is to change the light entering the system by adding some other components like color filters from a filter wheel. These systems are very similar (basically the same working principle) as the ones explained in subsection 2.1.1. But here the filter transmittances are usually spectrally broad and the number of bands captured is reduced (less than 10 normally). Of course, the main disadvantage is the need to mechanically move the color filters by means of a filter wheel or holder, and the time consuming of the capturing process. As some applications for these systems we can find: hyperspectral reflectance reconstruction or multispectral fluorescence imaging ([90, 75, 101]).

Hybrid systems

So far we have seen very simple and more complex systems for the acquisition of spectral images. Hybrid systems try to merge two different kinds of technologies to help each other and improve the quality of the spectral capture. The most common hybrid systems that have been developed so far are those which mix the information from a high spatial resolution *RGB* color camera and a low spatial resolution hyperspectral system that samples the scene in few points to add some spectral information as input to the estimation algorithms. In this field the most productive team so far to our knowledge has been the one led by Professor Yuri Murakami in the *Imaging Science & Engineering Laboratory* at *Tokyo Institute of Technology* in Japan. For several years, they proposed theoretical methods to recover full spectral reflectance information pixel-wise combining high resolution *RGB* images with what they call multi-point spectral measurements ([138, 137, 140, 141], see figure 2-8).

In these studies, they were performing simulations from previously acquired and estimated spectral images from multi-band cameras. They apply unmixing methods to separate the spectral components of different pixels belonging to the same spectral measurement. These authors have recently proposed and implemented a real system doing such acquisitions [139]. This system combines the *PGP* spectral line scanner explained in subsection 2.1.1, with a smart optical fiber arrangement which acquires the light in a bi-dimensional shape and converts it into a one-dimensional line to input the line scanner. Then after

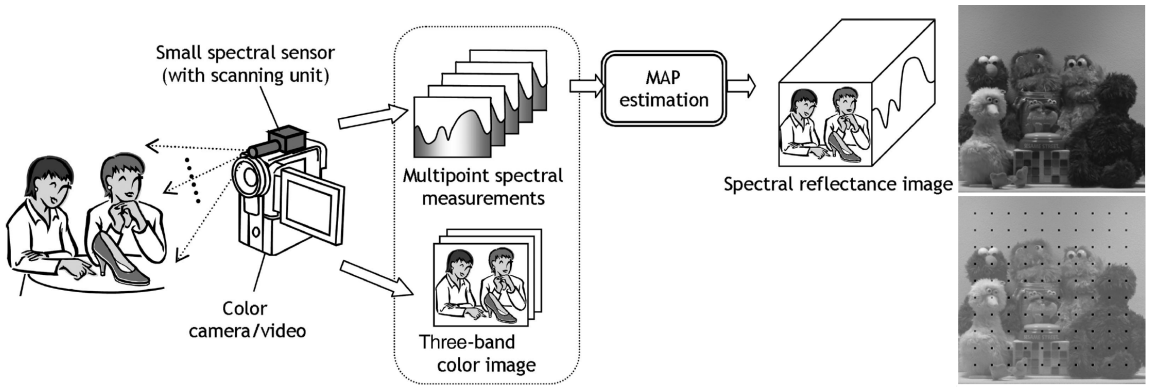

Figure 2-8: Hybrid system combining *RGB* images with multi-point spectral measurements. Left: scheme of hybrid capture. Right-top: grayscale image of one color channel image. Right-bottom: example of spot areas used for spectral measurements. Images retrieved from [141].

reading the signal from the scanner and rearranging it, they get the low spatial resolution spectral image. Applying the unmixing method presented in [141] to this image, and combining it with the high resolution color image taken by an *RGB* camera, side by side with the low resolution spectral system (*LRSS*), this system is able to get real-time spectral images. The computations needed to unmix the data, the misalignments between the *RGB* and spectral data, the optical accessories needed to convert 2-dimensional data into a line, and the need of using a spectral line scanner, make however this system complex and rather expensive.

Tunable responsivity systems

This is a new sensor technology which aims to overcome some of the limitations of the previously described capture systems. It takes advantage of the wavelength dependence of the penetration depth of photons in a silicon layer. Basically, when photons fall in a silicon sensor, they penetrate deeper the longer their wavelength is. In silicon sensors, light is absorbed following an exponential intensity decrease described by the Beer-Lambert law. Electron-hole pairs are photo-generated with different depth profiles according to the material and the radiation energy (or wavelength). In a typical pn-junction based *CMOS* pixel, photo-generated electrons are collected by the same well, regardless of the generation depth [35]. Thus, for each pixel, only one value is read in a single exposure.

Therefore, if the collection of generated electrons is done at different depths within the

silicon layer, we can retrieve different spectral bands within a pixel in only one exposure time. The first sensors using this technology were *Sigma Foveon RGB X3* sensors [89, 88, 59], mainly used for color digital photography. In only one shot and without the need of any demosaicking algorithm, we can get full resolution color information. But the responsivities of these sensors are fixed. Recently, a new generation of sensors which is still under prototype development, adds the property of tunable spectral responsivities. These new generation sensors are called Transverse Field Detectors (*TFD*), so far used mainly for color as well as infrared imaging applications ([111, 109, 108, 107]). Their working principle is similar to the *Foveon RGB X3* sensors, but in this case, a transverse electric field is applied through the silicon, in such a way that depending on the intensity of this field, the depth collection of electrons changes. Therefore, we have the ability of tuning the spectral responsivities of the channels by changing the biasing voltage. We can see a comparison scheme in figure 2-9, between normal Bayer-pattern filtered sensors, *Sigma Foveon X3* sensors, and *TFD* sensors. Also the typical spectral responsivity shapes are included in the figure.

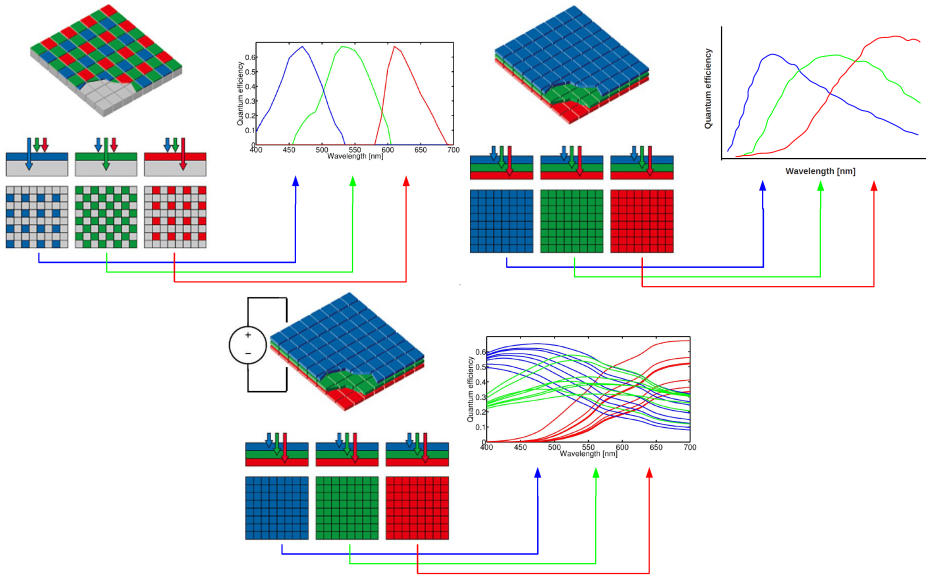

Figure 2-9: Comparison scheme between normal *RGB* Bayer-pattern filtered sensors (top-left), *Sigma Foveon X3* sensors (top-right) and *TFD* sensors (bottom). Note that *RGB* Bayer-pattern filtered sensors have reduced spatial resolution ($\frac{1}{3}$).

This system, in theory, should be included in subsection 2.1.2, because the way to get several spectral bands is taking different shots with different biasing voltages applied. However, *TFDs* offer the possibility to retrieve up to 5 channels in a single shot (3 in the visible range and 2 in the infra red). Besides, the switching of the biasing voltage is so fast and the

exposure times needed are so short, that 2 shots with different biasing voltages can happen in less than 50 ms. Therefore, we can get up to 10 channels in a very short time, and even more (15 to 20) without increasing much the capture time. Thus in practice we can say that these systems are multi-shot systems that overcome the drawback of very long capturing times. Furthermore, these sensors also offer the possibility of biasing half of their pixels in one way and other half in a different way (bipolarization). This means that the capability of retrieving up to 10 channels in only one exposure time exists. In this case we would sacrifice some spatial resolution and we would need to apply some demosaicking method. Spectral imaging systems have been proposed using *TFDs* [110, 121]. However these sensors present the disadvantage of having very broad spectral responsivities. This fact has been proven not to be very appropriate for spectral estimation [94]. In order to overcome it, and as one of the novel contributions of this thesis, a new system has been proposed based on simulations, using the *TFDs* and adding bi-polarized configurations together with color filter arrays [120, 123]. This new system could be included as a hybrid system combining different approaches to improve the spectral estimation accuracy. The proposed system will be explained in chapter 3.

2.1.3 Future trends of spectral imaging systems

We have seen many different approaches attempting to solve the problem of spectral imaging. It is not a trivial problem and there is no simple solution for it. There exist many systems that work pretty well for particular cases and under particular conditions. Capturing spectral images is easier when we deal with controlled conditions (illumination, geometry, static scenes, etc.) like it happens in a laboratory. For outdoor scenes the problem gets harder to solve. It is the future of this branch of science to design a system which has the ability to get real-time spectral images, being portable and easy to use, without the need of expensive optical components or multiple systems arranged in a complex way. It seems that the trending is to use some smart hybrid designed multispectral device combined with advanced mathematical tools to estimate the spectra pixel-wise. Yet some authors still try new complex architecture philosophies like incorporating Fabry-Perot filters directly on the sensor [189].

2.2 High dynamic range imaging systems

This section is a summary of the different approaches that engineers and researchers develop for overcoming the limited dynamic range of common imaging systems (as explained in chapter 1). Even though the variety might not be as extensive as that of spectral imaging systems (see section 2.1), there are still different approaches that are worth considering when facing the design of an *HDR* imaging system.

As done by other authors ([155, 24, 125]), we divide this section into two subsections or categories. Each of them corresponds to the two main philosophies adopted for capturing *HDR* images: the hardware approaches and the software approaches.

2.2.1 Hardware approaches for HDR imaging

In this section we explain different prototypes that some authors propose for capturing *HDR* images for general photography application, usually in one shot, but with the need of special cameras or sensors. We also made a distinction in this category. Those systems which are composed of common imaging sensors with some smart set up, and those which are composed of sensors with special architectures or features.

HDR hardware systems with common sensors

In this category we have systems using extra filters to control the exposure, like the one proposed in [145, 82] for general color photography. They use a normal *CCD* sensor coupled with a neutral density filter pattern with 4 different densities (see figure 2-10). Therefore, they can get 4 different exposures in one shot, by sacrificing spatial resolution, which is divided by 4, compared with a normal *RGB* imaging system.

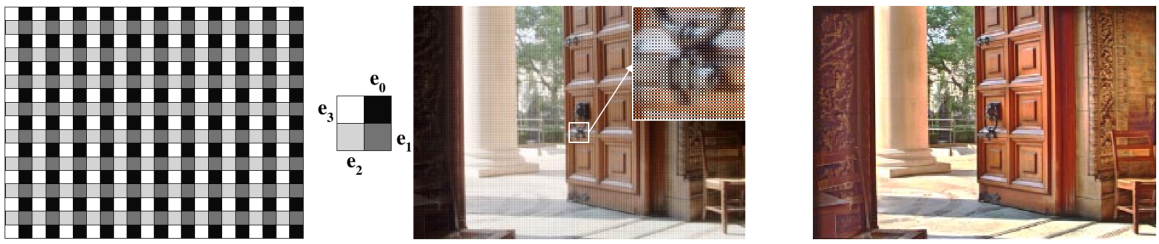

Figure 2-10: *HDR* imaging system proposed in [145]. Left: neutral density filter pattern. Center: *RGB* image with neutral density filter pattern on top. Right: final *HDR* image after tone mapping. Images retrieved from the paper.

Another approach is to use an *HDR* video system which adapts the exposure in each pixel by using a spatial light modulator (*SLM*), [144]. The *SLM* adapts the light through each pixel forming a grayscale image which is proportional to the negative of the current image recorded (see figure 2-11). Thus, those pixels which were very bright or saturated, would be dark in the *SLM*, letting less light through. And those pixels which were dark would be transparent to let the whole light through.

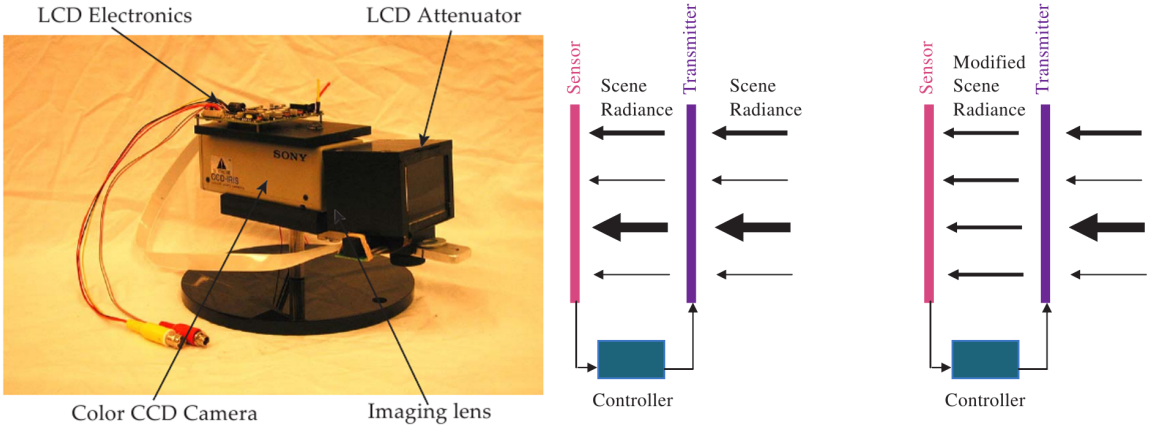

Figure 2-11: Adaptive *HDR* imaging system proposed in [144], based on a spatial light modulator and a video camera. Left: video camera attached to a spatial light modulator. Right: scheme of working principle. Images retrieved from the paper.

As a last example in this category we have the system proposed in [3]. They designed a system featuring 6 *CCD* sensors. Each sensor has a different neutral density filter on top of it. Therefore the image recorded in each sensor has a different exposure, but synchronizing all of them, we capture 6 different exposures in one shot. Of course they need to split the incoming light beam into 6 beams. For this purpose they use a 6-sided pyramid made of mirrors. Dividing the light into 6 beams makes each of them have one sixth less intensity. This brings the need of increasing exposure times or *ISO* settings.

HDR hardware systems with special sensors

In this category we find some *HDR* imaging systems that feature a special pixel architecture. In the previous systems, we assumed that pixels in all sensors respond linearly to light. However the idea in the systems of this category, is that the final sensor response does not. For this purpose, authors design systems that include pixels with linear response, but operating them in a different way. *CMOS* sensors allow the independent driving of

single pixels since the electronics controlling them are in-pixel. In the system proposed in [80], they achieve a piecewise linear sensor response by independently resetting each pixel partially while the capture is ongoing. Another approach with linear response pixels is that called time-to-saturation (see [171]). These systems, instead of measuring the amount of charge accumulated in the well, measure the time that each pixel needs to reach saturation. In bright regions this time would be short, and in darker regions it would be longer. The problem is that they need a large amount of in-pixel circuitry in order to control this charge monitoring.

On the other hand, pixels with non-linear response are interesting for *HDR* imaging as well. Logarithmic response pixels like those proposed in [156] are interesting because they yield high responsivity to low light, and low responsivity to high lights. We can say that logarithmic sensor's responsivity is inversely proportional to the incoming light intensity. Thus extending the dynamic range able to be captured in one shot. These sensors however tend to suffer from fixed pattern noise and response delay [24].

Also frequency modulators can be used in pixel for *HDR* imaging. They provide an output signal of constant amplitude but with a variable frequency directly proportional to the photocurrent. This approach is called light-to-frequency ([191]). The problem is that when the light signal is low, the readout time needs to be long in order to see at least one period of the output signal. In highlights however, the bandwidth needs to be very high.

2.2.2 Software approaches for HDR imaging

Software approaches for *HDR* imaging systems are of special interest, because they can be applied in any common imaging system. They can yield better looking *RGB* pictures than hardware methods, however they need the scene to be as static as possible. This happens because they are based in merging multiple captures of the same scene. Thus, no moving objects or changes in illumination should occur during the capturing time. Yet there are techniques to remove the effect of ghosting due to subtle movements [97, 100, 50, 37].

The usual architecture of a software method for *HDR* imaging is to capture multiple exposures of the scene and create an *HDR* radiance map. Later on, if we want to visualize the image, we perform tone mapping operators (*TMO*). There are many different *TMO* ([198, 113]). They are intended to make the final *HDR* image look as natural as possible after compressing its dynamic range to be able to be displayed in a normal *LDR* display. It

is not within the scope of this thesis to study the different *TMO*, since we are not aiming to produce natural looking *HDR* images. Nonetheless, we will use them in order to visualize the content of the scenes we capture with our imaging systems.

The ideal case for a software method would be that sensor response is linear and therefore reciprocity law holds directly [34]. Unfortunately, in many common digital imaging systems this is not the case. Their response to light is rather sigmoid-shaped. This curve is under-responsive close to dark and close to saturation, and over-responsive in the middle range, where it behaves almost linear (see figure 2-12). This fact makes it necessary to implement a previous step before stitching the different exposures of a scene together.

Firstly Mann and Picard in [119], and later on Debevec and Malik in ([26]), demonstrated that capturing different exposures of a static scene, is effectively sampling the camera response function (*CRF*) at each pixel. This *CRF* is a function that relates the input radiance outgoing the scene, or the irradiance impinging in the sensor, with the output sensor response. If we workout what this function is like, we are able to obtain the *CRF* of a capture system, we can use its inverse function (given that this function is by definition a monotonically increasing function), to estimate the radiance (L_e)/irradiance (E_e) impinging on each pixel of the sensor. This is how the concept of radiance map emerges. In figure 2-12, we see an example of *CRF* function that we estimated for a *Canon EOS 7D* camera working in *jpeg* mode.

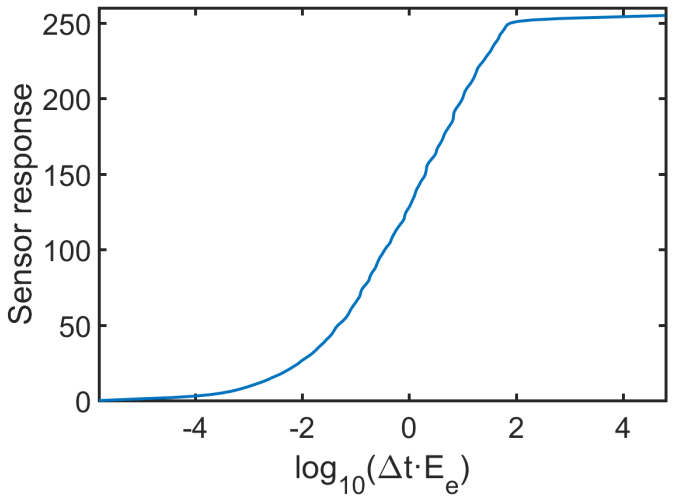

Figure 2-12: Example *CRF* of a *Canon EOS 7D* camera working in *jpeg* mode.

We see the typical S-shaped (sigmoid) function mentioned before. This is the key to recovering *HDR* image data from multiple *LDR* images captured using different exposure

settings (either changing the exposure time, or the aperture, or the ISO, etc.). We will discuss later on how to merge these differently exposed *LDR* images.

The main difference between the main software methods proposed, is the way this *CRF* is derived from image data. Mann & Picard ([119]), estimate it in an iterative way from only two differently exposed images. A set of corresponding pixels with their pixel values is generated, and from them, a power function *CRF* is estimated using regression.

Debevec & Malik ([26]) though, do not impose a power shape to compute the *CRF*. They solve it by minimizing an objective function by singular value decomposition. The resulting *CRF* can feature any shape.

Mitsunaga & Nayar ([134]), model the *CRF* by a polynomial of order k, where the order and the the $k+1$ coefficients are unknown. They do so in an iterative process which is very computationally heavy.

Robertson et al. ([157]), choose a cubic spline function to model the *CRF*. Besides, they also include constraints related with the exposure times used to capture the images when building the *HDR* image from many *LDR* images.

These are the main software methods used to obtain *HDR* images. Once the *CRF* is estimated and the *LDR* images are captured, the way we build the *HDR* image is the same (see section 4.3.2).

2.3 Polarimetric imaging and material classification

Since in this thesis the polarization information is used to perform material classification of the objects present in the captured scenes, we will tackle these two topics together. We focus more closely on material classification techniques rather than describe all the existing polarimetric imaging techniques.

Regarding material classification, initially some authors studied it by using the dichromatic reflection model ([175]). Using a spectroradiometer, they studied the spectral color signals coming from different objects and also its chromaticities. In general *RGB* or monochrome images were used to perform material classification ([70]). Limitations of those methods are that the color reflected by objects is strongly dependent on the observation geometry. In [188], they based the classification in the texture of the objects, assuming that objects with similar textures are made by the same materials. Later on some authors

approached the material classification problem via polarization-based imaging. In [194, 17], they already tried to classify objects in two main classes: metals and dielectrics. They based their method in a thresholding of the Degree of Linear Polarization ($DoLP$). They used for the first time polarization features, which do not depend on object color and illumination. However their method was strongly dependent on the observation geometry, and only available for flat object surfaces. Besides, using polarimetric images also helps getting extra information about the content of the scene that is not visible with common imaging systems [85, 104]. In [180], they found that inspecting the curvature of $DoLP$ map around specular highlights, they could classify the objects into metal or dielectric, disregarding the observation geometry. This allowed them to perform classification including curved object surfaces as well ([179]), but sometimes the $DoLP$ calculation around the highlights is unstable. Specially if the image captured is somewhat noisy or the highlight region is rather extensive. This system could also be included in next section 2.4, since the imaging system used is also a multispectral system. Besides, the images captured are HDR as well. They used a liquid crystal tunable filter ($LCTF$), and took advantage of its light polarization properties. Therefore by rotating the $LCTF$ in front of the camera (see figure 2-13), they could capture images at different polarization angles.

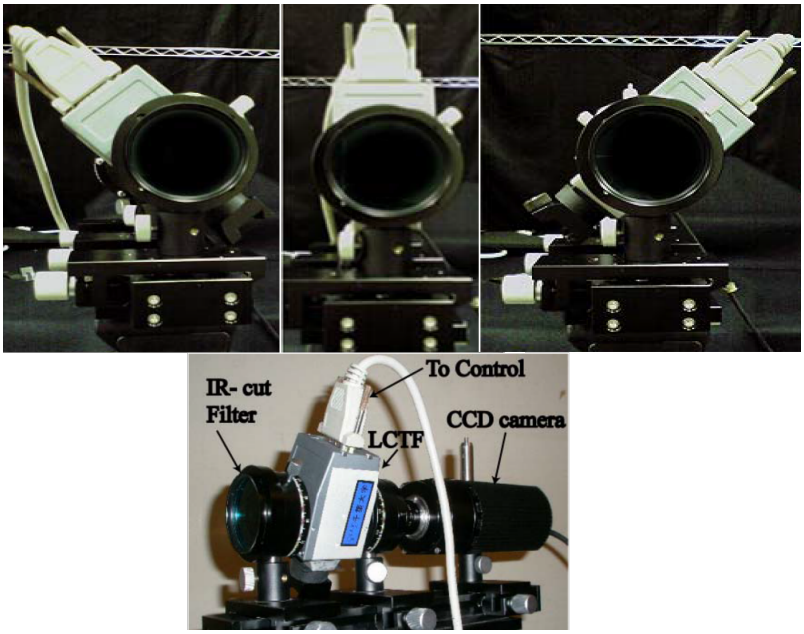

Figure 2-13: $LCTF$-based imaging system used in [179]. The LCTF is rotated in different angles to capture different polarization angles of light. Image provided by Horiuchi & Hirai laboratory, in Chiba University (Japan).

2.4 Multispectral, HDR, Polarimetric imaging crossovers

Few authors combined some of these techniques (multispectral, *HDR* and polarization imaging) before. In [61], they propose a system composed of a beam splitter dividing the incoming beam of light into two beams. Each of these is impinging in a different sensor. Both are *RGB* sensors, but each of them has a different comb-shaped filter on top. Therefore each of the 3 color channels is spectrally divided into two halves. Besides, one of the sensors has also a neutral density filter on top of it, making the whole response of this sensor less sensitive. This way they get a 6 channels *HDR* multispectral image in one shot.

In [162], they propose a system for *HDR* polarimetric imaging in remote sensing. They rigidly attach to the camera a fixed, spatially varying polarization and attenuation filter. This way, the camera motion induced multiple measurements per scene point, taken under different optical settings. This is in contrast to the redundant measurements of traditional Bayern pattern mosaics. Computational algorithms then analyze the data to extract polarization imaging with high dynamic range for the entire field of view.

Even some authors tried combining all three of them together in a system able to capture all-sky *HDR* multispectral and polarimetric image data in less than two minutes, using two filter wheels and a fish-eye lens ([151]). They use a fish-eye lens to capture all-sky images, together with two filter wheels. One filter wheel contains multispectral filters at different wavelengths and bandwidths, and the other wheel contains linear polarization filters at different polarization angles. This way they sequentially capture all-sky multispectral and polarimetric images in *HDR*.

We based a chapter of this thesis on a system proposed in [179]. They used a Liquid Crystal Tunable Filter (*LCTF*) attached to a monochrome scientific camera to capture multispectral *HDR* polarimetric images by using exposure times chosen by the user by trial and error. Their capturing process was not automatic or adaptive, and they were not accounting for the misalignment errors we found when repeating their experiments with the same work-flow (see chapter 5).

2.5 Near Infrared imaging

When it comes to Near Infrared (*NIR*) imaging, most systems are specifically application-oriented. We can find many works using *NIR* imaging. Even though *NIR* portion of the

spectrum is considered to be up to 2500 nm approximately, many studies use only up to 1000 nm, where common silicon imaging sensors are still responsive.

In [91], they study wavelength-specific features in the spectrum of particulate minerals. They use images from a remote sensing database covering the spectral range from 325 nm, to 2500 nm.

Regarding the medical imaging field, in [96], they used an array of microbolometer elements, to generate thermal maps of different regions of the human body, trying to relate features of these maps with possible diseases. Also in [87], they proposed a system based on gold nanorods with suitable aspect ratios that absorb and scatter strongly in the range from 650 nm, to 900 nm. They found a way to distinguish cancer malignant cells from benignant ones, in order to find how to thermally destroy only the malignant ones, keeping the benignant ones alive.

Concerning food industry, in [114] the authors used the line scanning hyperspectral imaging system proposed in [102], to capture 460x1200 pixels hyperspectral images with 112 spectral bands in the range from 447 nm, to 951 nm. They wanted to detect fecal contaminants present in apple surfaces treated with different concentrations of fecal treatments. Also tenderness of fresh beef, as well as its pH were studied by different authors. For this purpose, in [143], the authors used a line scanning system composed of a CCD silicon sensor and a spectrograph. They covered the range from 400 nm, to 1000 nm. In [33], they used the same line-scanning approach with *Specim Imspector N17E* spectrograph covering the spectral range from 910 nm, to 1700 nm.

As instances of some other miscellaneous applications, in [200], the authors proposed a method for enhancing the texture details of color images by using the contrast and texture information from a *NIR* image. They capture the dual images by using two synchronized commercial cameras installed in an optical set up which divides the incoming beam of light by two using a hot mirror (see figure 2-14). Therefore one of the cameras receives the visible light, and the other the *NIR* light. The latest is modified by removing the IR cut-off filter on top of the CCD sensor (which is a built-in feature of this camera model: *Sony F828*).

In [5], the authors proposed a method to assess environmental changes, by deriving baseline conditions for the radiation intercepted by vegetation. They used remote sensing imagery in the visible and *NIR* portions of the spectrum, and deduced a vegetation index as a ratio between spectral bands in both regions.

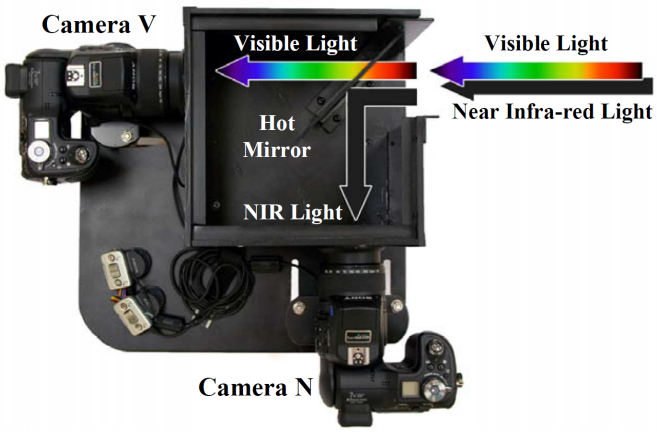

Figure 2-14: Dual camera system used in [200]. Image retrieved from the paper.

If we are going to capture outdoors images for vegetation monitoring, we should include the possibility of capturing *HDR* image data. Therefore in section 5.6, we propose a system and a capturing procedure to retrieve multispectral *HDR* and polarimetric images of outdoors scenes in the visible and near infra-red portions of the spectrum.

Light brings us the news of the Universe.

William Bragg.

3

Spectral imaging using Transverse Field Detectors

Content:

Chapter 3

Spectral imaging using Transverse Field Detectors

We have seen in the previous chapter 2 that there exist many different approaches for spectral imaging. All of them have advantages and disadvantages, and depending on the application we are focusing in, we have a wide variety of architectures to choose from. When facing a specific application ([150, 114, 27]), one of the advantages is that usually, the conditions in which the capturing is done (i.e. illumination, motion in the scene, geometry, background, etc), are rather well known or controlled. However, if we aim for a more general imaging application, specially if the capture is made outdoors ([84, 28]), the imaging conditions usually will not be controlled any more.

As mentioned in chapter 1, one of the contribution of this thesis relates to the improvement of multispectral imaging systems. Later on, in section 2.1.2, we introduced a new generation of filter-less tunable sensors that are being developed in the Polytechnic University of Milan (Italy). These sensors are called Transverse Field Detectors (*TFDs*). We studied the working principles of this new technology and tried to exploit the promising capabilities that these sensors could bring to multispectral imaging. Electronic tunability is a rather interesting feature which is worth testing for capturing this kind of images.

Thus we tried to achieve a higher color and spectral accuracy in the estimation of spectral reflectances from sensor responses. Such an improvement was done by combining these recently developed silicon-based sensors with color filter arrays (*CFAs*). As mentioned in section 2.1.2, *CFAs*are usually composed of 3 different filters (red, green and blue, also

named as *RGB*). But they can be composed of any number of filters. If the *CFA* has more than 3 different transmittances or colors, it can also be called multispectral filters array (*MSFA*). In this work, for simplicity, we will always call them *CFAs*.

As we can observe in figure 3-1, *TFD* responsivities are rather spectrally broad. Hence, we decided to include these *CFAs* in order to be able to narrow them down. Using the *CFAs*, we sacrificed the filter-less full spatial resolution property of *TFDs*, but we gained narrower spectral responsivities.

We designed and performed several experiments based on simulations, to test the influence of different design features such as: type of sensor, tunability, interleaved polarization, use of *CFAs*, type of *CFAs*, and number of shots, on the estimation quality. Some of these features are exclusive to *TFDs*. We compared systems that use a *TFD*, with systems that use normal monochrome sensors, both combined with multispectral *CFAs* as well as common *RGB* filter arrays present in commercial digital color cameras.

We propose *CFA*+*TFD*-based systems with one or two shots depending on the possibility of using longer capturing times or not. Improved *TFD* systems thus emerge as an interesting possibility for multispectral acquisition which overcomes the limited accuracy found in previous studies [121].

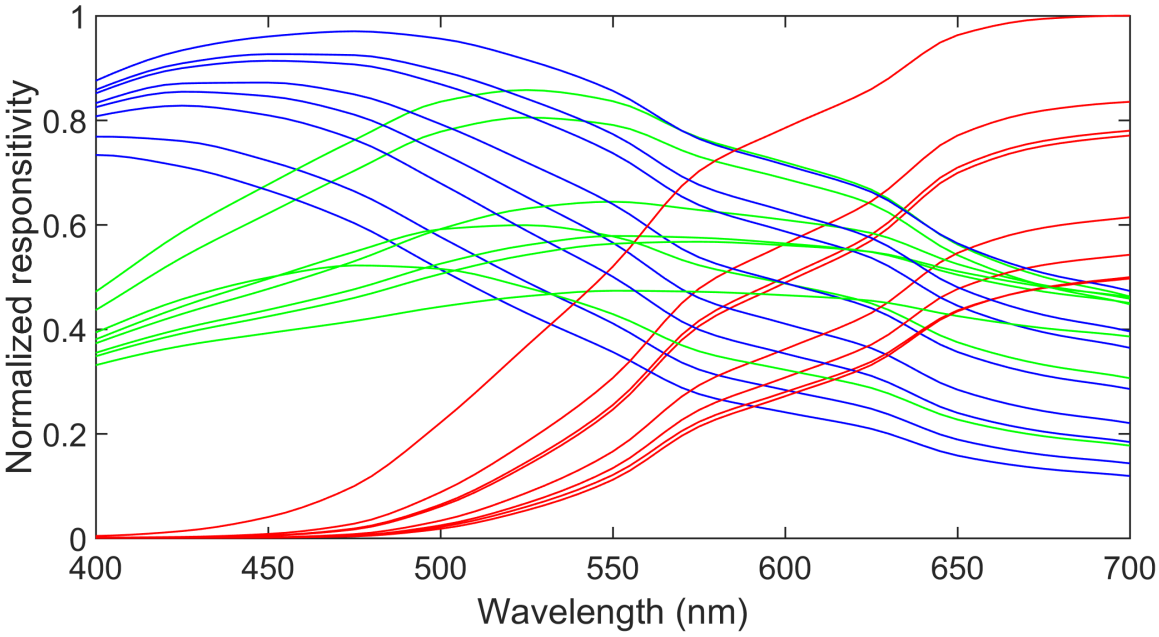

Figure 3-1: *TFD* responsivities corresponding to 8 different biasing conditions. Red, green and blue lines correspond to red, green and blue channels in each tuning condition. These responsivities were measures in laboratory by *TFD* developers in Milan (Italy).

As explained in section 2.1.2, these sensors take advantage of the wavelength-dependent penetration depth of photons in silicon. The longer the wavelength is, the deeper the penetration depth of the photon within the silicon layer. Therefore, if we collect electrons generated upon incoming photons at different depths, we are selectively separating them according to their wavelength (see figure 3-2).

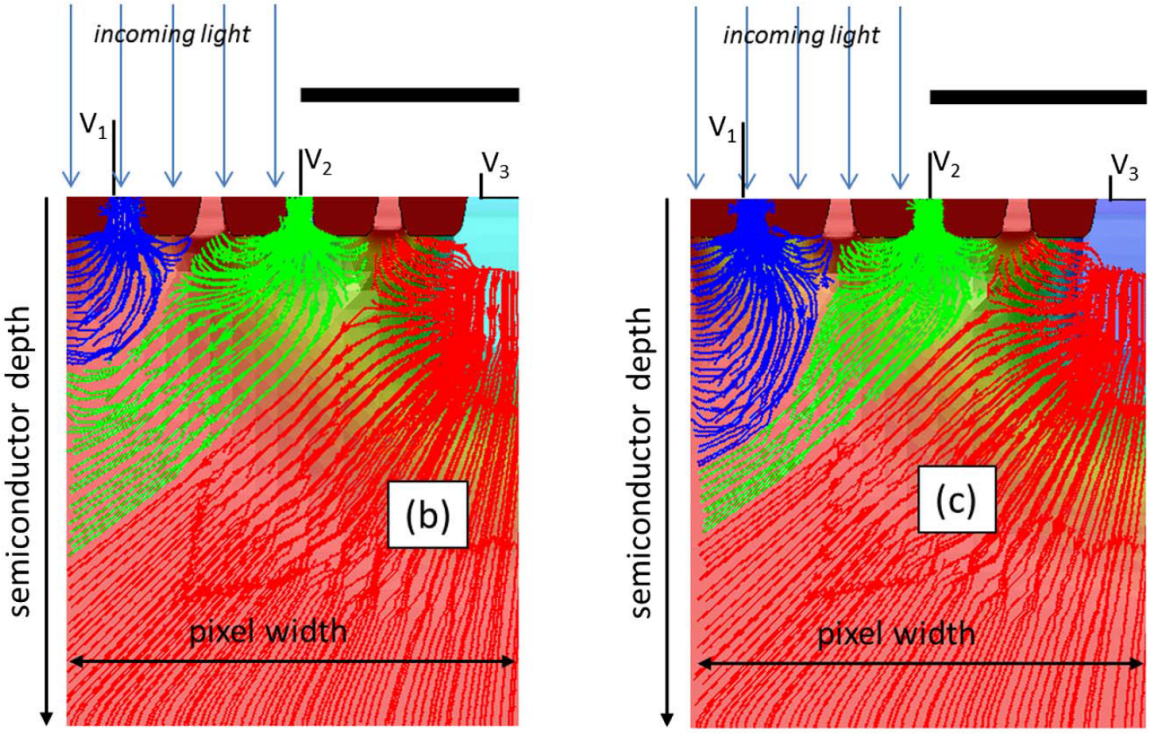

Figure 3-2: *TFD* pixel scheme of electrons collection trajectories at different depths. Left and right represent different biasing voltages. Red, green and blue lines represent the trajectories of generated electrons from photons in roughly the red, green and blue regions of the spectrum. Images retrieved from [107].

This property is also characteristic of filter-less color imaging sensors like *Sigma Foveon X3* ([133, 89, 88, 59]). As explained in section 2.1.2 these sensors are tunable, by means of applying different biasing voltages [115, 199]. Moreover, this tunability property goes beyond with the bi-polarization property (half pixels of the sensor with one configuration, and the other half with a different configuration). This bi-polarization is done in an interleaved way (like a chess pattern mosaic where black patches are tuned with one biasing condition, and white patches with a different one). When all pixels of the sensor are equally biased, we call it full-polarized. If half of the pixels are differently polarized from the other half, we call it bi-polarized. In figure 3-3 we see an example scheme of the bi-polarization pattern.

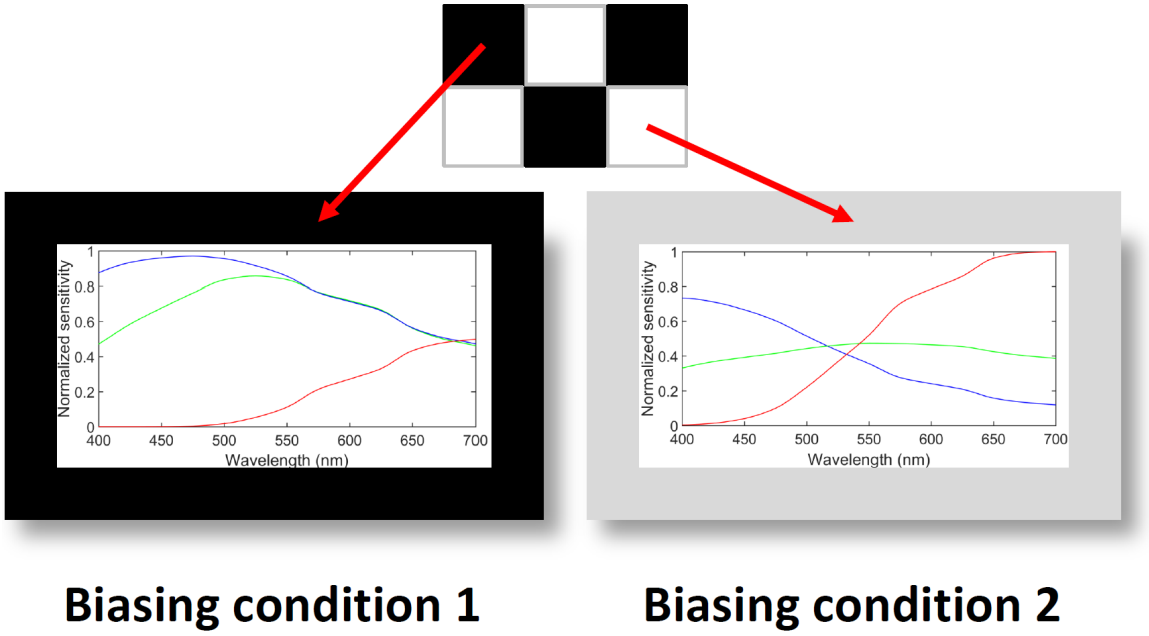

Figure 3-3: Example scheme of a 6 pixels area of a bi-polarized *TFD* sensor.

These sensors however, are still in an early stage of development. By the moment this work was done, the biggest prototype was an $8x8$ sensor, attached to a prototype board. Later on, the developers built a $64x64$ pixels sensor and included it in a camera mount for prototype sensors, operating it from an external micro processor ([12]).

Our idea was to study the potential of these sensors for spectral imaging, taking advantage of their quick tunability. Therefore, we simulated sensor responses using different biasing voltage values, using real objects' reflectances illuminated with real outdoors light sources' *SPDs* (for additional details see section 3.1). Using these multispectral sensor responses, we applied a spectral estimation algorithm to recover the spectral reflectances. We also compared the spectral estimation results with those of other typical system architectures as well.

In the following section 3.1, we will explain the physical model used to simulate *TFD* sensor responses. Section 3.2 is dedicated to explain the *TFD*-based system we propose, as well as the different architectures simulated to compare their performance with our proposed system (using both *TFD* sensors in different configurations as well as normal silicon sensors). In section 3.3, we explain the simulation experiments performed, together with the spectral estimation algorithm and the spectral data used as well as the methods for assessing the performance and comparing it with the other systems simulated. In section 3.4 we explain

the results we reached in these experiments. Finally in section 3.5, we summarize the main conclusions related to this contribution of the PhD dissertation.

3.1 Physical model to simulate TFD pixel responses

TFD developers at Polytechnic University of Milan, provided us with a physical model to simulate pixel analog to digital conversion and noisy sensor response generation. They also gave us the spectral quantum efficiencies of 8 different biasing conditions of *TFD* (3 channels per biasing condition), ranging from 400 nm to 700 nm. Therefore we could generate sensor responses from spectral radiance signals. And these radiance signals could be generated from spectral reflectances of real objects and spectral power distributions (*SPDs*) of real light sources.

In the following we describe the physical model used to simulate camera responses. $R(\lambda)$ is the object spectral reflectance illuminated with illuminant $L(\lambda)$.

$$C(\lambda) = R(\lambda) \cdot L(\lambda) \tag{3.1}$$

$$I_{ch} = q \cdot A_{light} \cdot \int_{400nm}^{700nm} C(\lambda) \cdot Qe_{ch}(\lambda) d\lambda \tag{3.2}$$

$$V_{ch} = \frac{I_{ch} \cdot t_{exp}}{C_f} \tag{3.3}$$

We firstly calculate the spectral color signal $C(\lambda)$ as the product of the two spectral signals $R(\lambda)$ and $L(\lambda)$. Later on we calculate the photo-current (I_{ch}), as the product of the elementary charge ($q = 1.6 \times 10^{-19} C$), the pixel illuminated area ($A_{light} = A_{total} \cdot FF$, where $A_{total} = 1 \mu m^2$ is the total pixel area, and $FF = 0.2$ is the pixel fill factor for the *TFD CMOS* technology), and the integral over the visible range of the product of the calculated $C(\lambda)$ and the spectral quantum efficiency of each channel of the sensor ($Qe_{ch}(\lambda)$). Later on, for the calculation of pixel output voltage (V_{ch}), we use the exposure time used to capture the image, and the pixel feedback capacitance $C_f = 3 \times 10^{-14} F$.

Since we are dealing with digital signals, all spectral signals are sampled at a sample step of $\Delta\lambda = 10nm$ from $400nm$ to $700nm$. Therefore, equations 3.1 and 3.2 transform into

the following two equations 3.4 and 3.5 respectively:

$$C(\lambda) = R(\lambda) \cdot L(\lambda) \tag{3.4}$$

$$I_{ch} = q \cdot A_{light} \cdot \sum_{400nm}^{700nm} C(\lambda) \cdot Qe_{ch}(\lambda) \cdot \Delta\lambda \tag{3.5}$$

After calculating the pixel output voltage, we check for saturation. If this voltage is higher than the maximum pixel output voltage ($V_{dd} = 1.3V$), it is clipped to this value.

This output voltage represents the noiseless analog output of the *TFD* pixel. So far we have not included any source of noise. Thus, before performing quantization in order to transform the analog signal in a digital signal (process which will also add noise to the final sensor response), we include two sources of analog noise: photo-current shot noise, and thermal noise. The former depends on the signal level, the exposure time, and the gain of the electronics. The latter is independent on signal level and exposure time. It only depends on the temperature and the value of the feedback capacitance.

The noise variances are calculated as:

$$\sigma_{ch_{shot}} = \sqrt{\frac{2 \cdot q \cdot t_{exp}}{C_f} \cdot (I_{ch} + J_d \cdot A_{total} \cdot \alpha_{ch})} \tag{3.6}$$

$$\sigma_{kT/C} = \sqrt{\frac{k \cdot T}{C_f}} \tag{3.7}$$

Where $J_d = 7 \times 10^{-5} \frac{A}{m^2}$ is the technology dark current density, and a scaling factor α_{ch} is 0.5 for red channel, 0.3 for green channel and 0.2 for the blue channel. $T = 300K$ is the absolute temperature and $k = 1.38 \times 10^{-23} \frac{J}{K}$ is the Boltzman constant.

There are other typical sources of noise, like dark current shot noise. It is independent on signal level, and dependent on exposure time. However the variance of this noise can be reduce to negligible values by capturing and subtracting dark images using the same exposure time. Therefore we did not add this source of noise.

We add now the noise to the output analog voltage. We do it sequentially because different noise sources have different probability distributions. First we add the thermal

noise as indicated in equation 3.8, where function *rndmnorm* (*Matlab R2012b*) picks a random value from a normal distribution of V_{ch} mean and $\sigma_{kT/C}$ variance.

$$V_{ch\eta_{kT/C}} = rndmnorm(V_{ch}, \sigma_{kT/C}) \tag{3.8}$$

However, the probability density function of shot noise behaves statistically as a Poisson's distribution. Therefore, after the thermal noise is added, we add shot noise as indicated in equation 3.9, where function *rndmpoiss* (*Matlab R2012b*) adds a random value to $V_{ch\eta_{kT/C}}$, according to the mass density function $f(k, \lambda)$ in equation 3.10

$$V_{ch\eta} = rndmpoiss(V_{ch\eta_{kT/C}}, \sigma_{ch_{shot}}) \tag{3.9}$$

$$f(k, \lambda) = \frac{e^{-\lambda}\lambda^{k}}{k!} \tag{3.10}$$

It might happen that, a very high unsaturated noiseless sensor response becomes saturated after adding noise. Therefore at this point we check for saturation again. This time we compare the value of $V_{ch\eta}$ with the maximum pixel output voltage ($V_{dd} = 1.3V$). If $V_{ch\eta} > V_{dd} = 1.3V$ then $V_{ch\eta}$ is clipped to V_{dd}.

Now that we already have the final noisy analog output signal for channel ch, we quantize it with $BIT = 12$ bits as indicated in equation 3.11.

$$\rho_{ch_{\eta}} = round(\frac{V_{ch\eta}}{V_{dd}} \times (2^{BIT} - 1)) \tag{3.11}$$

Where $round(x)$ is a function that rounds the value of x to the closest integer, and ρ_{ch} is the final 12 bits noisy digital sensor response for channel ch, in digital counts (DC).

We are still missing the noise introduced by the quantization in the analog to digital conversion (quantization noise). The variance of quantization noise (σ_q) is calculated as shown in equation 3.12).

$$\sigma_q = \frac{V_{dd}}{\sqrt{BIT} \times 2^{BIT}} \tag{3.12}$$

Now we can calculate the total noise variance including all sources of noise: shot noise, thermal noise, and quantization noise. Thus, the total noise variance σ_{tot} is calculated as

the squared root sum of all noise variances, as shown in equation 3.13.

$$\sigma_{ch_{tot}} = \sqrt{\sigma^2_{ch_{shot}} + \sigma^2_{kT/C} + \sigma^2_q} \tag{3.13}$$

And the *SNR* can be calculated as shown in equation 3.14, using the noiseless output voltage V_{ch} calculated in equation 3.3, and the total noise variance ($\sigma_{ch_{tot}}$) calculated in equation 3.13.

$$SNR_{ch} = 20 \times log(\frac{V_{ch}}{\sigma_{ch_{tot}}}) \tag{3.14}$$

Now that we can calculate the noisy simulated sensor responses as well as the total *SNR*, we can already compare the colorimetric and spectral estimation performances of each system, taking into account the *SNR* featured by each of them.

3.2 Systems compared

In previous sections, we have already explained the working principles of *TFD* sensors. Thanks to their easily tunable capabilities, we believe they are good candidates for spectral imaging. Spectrally narrow responsivities can be beneficial for spectral imaging if we are to estimate spectral signals pixel-wise (i.e. spectral reflectances, radiances, etc) [94]. However, as we can see in figure 3-1, the spectral responsivities of *TFDs* are far from narrow. Therefore, we studied how good the spectral estimation was using only raw *TFD* sensors, but we also studied how *TFD* sensors combined with *CFA* performed. We actually proposed a system exploiting all benefits from *TFDs* as well as *CFAs* together.

What follows is a series of simulation experiments that aim to prove that the proposed system outperforms other reasonable capture system designs. We will describe all the systems simulated and tested, and then show the results of the respective comparisons to achieve the proposed aim.

All systems studied in this work are silicon-based systems. Despite the architecture of each system in particular, the light acquisition process of them all can be realistically described with the CMOS capture model used for simulating camera responses (see section 3.1). Therefore, sensor responses for every system studied in this work have been calculated using the same model for signal and noise generation (explained in section 3.1). We can

thus be sure that a higher or lower noise level is not the reason for a better or a worse performance. The average *SNR* was calculated for each system. As it was explained in section 3.1, the noise model accounts for exposure time and quantum efficiency. Therefore, even though the values of exposure time were different for different systems, this difference was taken into account when computing the noise. As we will see in section 3.4, all systems resulted in similar *SNR* values. Actually the system which resulted in the lowest *SNR* (worst noise condition) is the proposed system number 1 (see table 3.1, where the average *SNR* is shown for each simulated system). To demonstrate the performance of the designed system, we have created 4 different experiments (explained in section 3.3) in which its features are tested across other imaging systems.

In Figure 3-4, we can see a representation of the eleven different systems simulated in this study and used to complete the four different experiments.

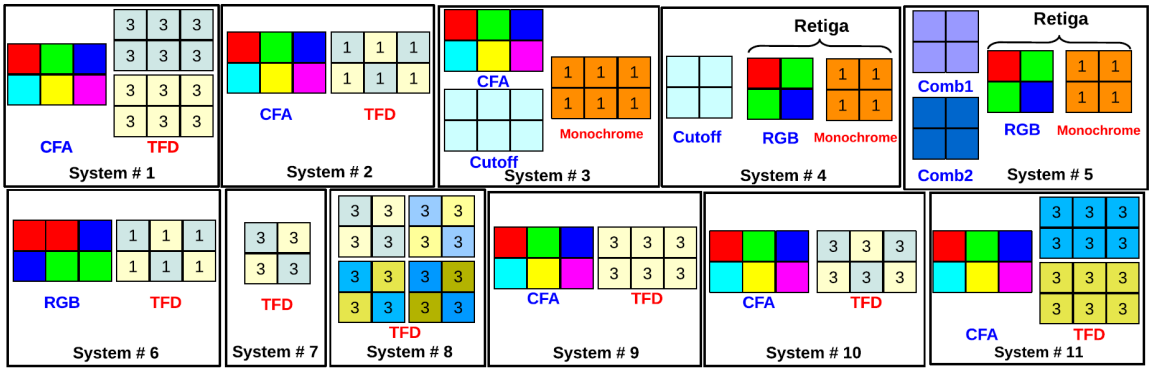

Figure 3-4: Scheme of the 11 systems simulated and compared along the 4 different experiments explained in section 3.3. The filter layer has blue caption. The sensing layer has red caption. The number in each pixel determines the number of channels retrieved in one shot out of it.

In each system there is a filter layer and a sensing layer. The number of different transmittances in the filter layer determines the spatial resolution of the captured image. A *CFA* with 3 filters provides $\frac{1}{3}$ of spatial resolution and with 6 filters $\frac{1}{6}$. If the *TFDs* are bi-polarized and use no filters, the spatial resolution reduces to $\frac{1}{2}$. The characteristics of each system are explained in the following and summarized in table 3.1.

1. This is the proposed system. The filter layer is a *CFA* made up of 6 band-pass optical filters selected from a real database (see section 3.3). The sensing layer is a *TFD* sensor full-polarized where two different biasing conditions were used in two shots. 3

Table 3.1: Systems characteristics

System #	System name	Filter layer	Sensing layer	# shots	# channels	Spatial resolution	Average SNR
1	CFA+2TFD	CFA	TFD full	2	36	1/6	40.5 dB
2	CFA+1TFDbi1	CFA	TFD bi	1	6	1/6	42.1 dB
3	CFA+Monochrome	CFA	Silicon	1	6	1/6	42.5 dB
4	Cut+Re	RGB\RGB+C_{off}	Silicon	2	6	1/3	42.3 dB
5	Comb+Re	RGB+$Comb_1$\RGB+$Comb_2$	Silicon	2	6	1/3	44.6 dB
6	RGB+TFDbi	RGB	TFD bi	1	6	1/6	44.4 dB
7	TFDbi	Empty	TFD bi	1	6	1/2	42.8 dB
8	TFDbi4	Empty	TFD bi	4	24	1/2	47.7 dB
9	CFA+1TFDfull	CFA	TFD full	1	18	1/6	47.4 dB
10	CFA+1TFDbi3	CFA	TFD bi	1	18	1/6	40.9 dB
11	CFA+2TFDalt	CFA	TFD full	2	36	1/6	45.1 dB

channels are retrieved per pixel. Thus, without the need of any moving mechanical optical component, we get information from 36 channels in two shots.

2. Same *CFA* used as in system number 1. The sensing layer however was a bi-polarized *TFD* from which we only take the information from one channel per pixel. The channel which was operative for each pixel was selected as explained in section 3.3. This system is similar to the proposed one, but it has been designed to compare fairly its performance with other systems under the same conditions, in terms of number of channels.

3. The sensing layer is a normal monochrome silicon sensor like the one from a commercial scientific camera model *Retiga SRV (Qimaging Corp., Canada)*, equipped with a *NIR* cut-off filter. In the filter layer we set a *CFA* which has been optimized for the monochrome sensor, using the same technique as in our proposed system (see section 3.3). This system has been designed to test if including a *TFD* sensor with its tunability property in a *CFA*-based system helps to improve its performance.

4. In the sensing layer we used the spectral responsivities of the *RGB* scientific camera *Retiga 1300C (QImaging Corp., Canada)*. In the first shot we set no filter in the filter layer, and in the second shot we added a *IR-UV*-cutoff filter (C_{off}) in front of the lens, which sharpened down the responsivities in the extreme of the spectrum. This system has been designed to test the performance of *TFDs* plus *CFAs* against other types of systems which were proven to work well in a previous work [121].

5. In the sensing layer we used the *RGB* camera again. This is also a 2-shots system. In each shot a different ideally custom-made comb-shaped optical filter ($Comb_1$ and $Comb_2$) was placed in the filter layer. These filters divided by half the 3 responsivities

into 6. This idea of splitting the spectral responsivities of single channels into two by using a comb-shaped filter has been used by some authors in the literature as well [68, 166, 185].

6. The sensing layer used here was the same as in system number 2, but the filter layer used was composed of an *RGB* Bayer filter like the one from the scientific camera *Retiga 1300C (QImaging Corp., Canada)* used in this study as representative of a standard *RGB CCD* camera. This system has been designed to test the effect of the selected *CFA* versus other kinds of narrow band-pass filters present in common color imaging systems.

7. The filter layer was left empty in this system. The sensing layer used was a bi-polarized *TFD* sensor. Only one shot was taken. This system and the next one were designed to test if adding a *CFA* to a *TFD* sensor improves its performance.

8. The same as the previous configuration but we took 4 shots with 4 different bi-polarization conditions. We wanted to push the number of channels to the extreme using all the biasing conditions that the developers provided to us.

9. The filter layer was the same *CFA* used in system number 1, but the sensing layer was a single shot full-polarized *TFD* sensor. This system has been designed to test the effect of bi-polarization property of the *TFD* sensor versus full-polarization state.

10. The filter layer was the same *CFA* used in system number 1. The sensing layer in this case was a one-shot bi-polarized *TFD*. This system is also proposed as an alternative to system number 1 if the application requires shorter capturing times (since it only uses one shot instead of two). In addition, this system has been compared to system number 9 to check the effect of using the bi-polarization property of *TFD* sensors.

11. This system has been designed exactly the same as the proposed system. However its channels were randomly selected from the available ones instead of using voting principal feature analysis (*VPFA*) [16, 117] as explained in section 3.3. Ten reasonable combinations were simulated in which the channels were selected to cover the whole visible range with some overlap. We selected the one with best results. This system was simulated to check whether selecting the channels using *VPFA* effectively increases the performance of our system or not.

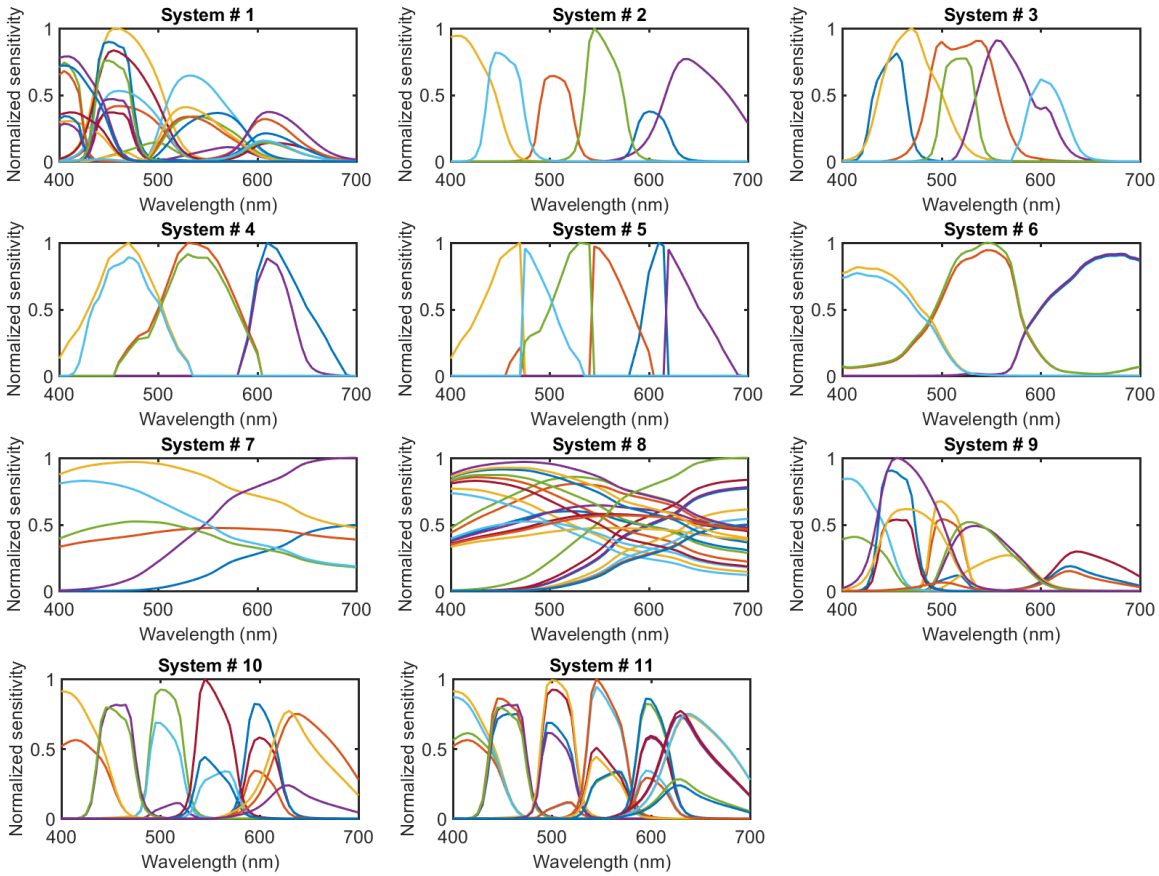

Figure 3-5: Normalized Spectral Sensitivities of all systems versus wavelength in nanometers. Normalization was performed just for displaying.

We have specifically selected these configurations to better isolate the effect of each feature that we wanted to test in our proposed system. To sum up: systems number 2, 3, 4 and 5 were designed to prove, under the same conditions in terms of number of channels, that using a *TFD* sensor improves the accuracy of the results over common monochrome sensor systems. Systems number 2, 6, 7 and 8 were designed and compared to test whether combining the *CFA* with the *TFD* also helps to improve its performance. Systems number 9 and 10 were designed to test if using the bi-polarization property of *TFDs* is better than not using it. System number 1 evolved from system number 10, pushing up the number of channels retrieved with only two biasing conditions adding one shot to the capturing process. System number 11 was designed to test the advantages of using *VPFA* to select the spectral channels.

In figure 3-5 we can see the spectral responsivities of all simulated systems.

Note how systems including a *CFA* (numbers $1, 2, 3, 9, 10$ and 11), have narrower re-

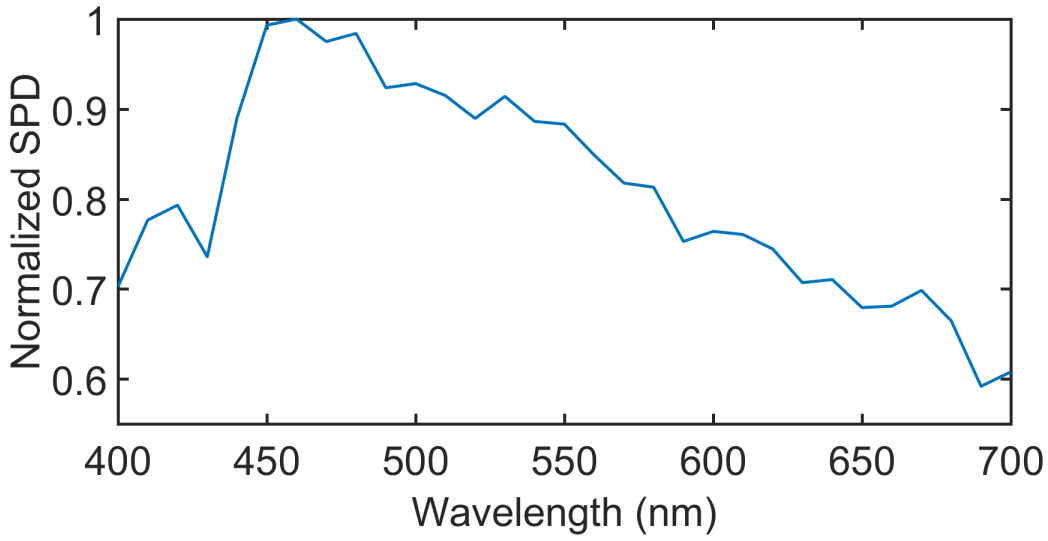

Figure 3-6: Normalized Spectral Power Distribution of *CIE* standard illuminant *D65*.

sponsivities compared with systems using *RGB* Bayer pattern filters (numbers $4, 5$ and 6), or filter-less *TFD*-based systems (7 and 8).

3.3 Methods

This section is divided into 4 subsections. In 3.3.1, we explain the filter selection as well as the spectral data used in the experiments. In 3.3.2 we explain the color and spectral metrics used to assess the spectral estimation from sensor responses and compare the performance between different systems. In 3.3.3, we explain how the spectral estimation was performed and evaluated. Finally, in 3.3.4 we explain the 4 simulation experiments conducted to test how well our proposed system performs compared with the others.

3.3.1 Spectral data and filter selection

As mentioned in section 3.1, the illuminant used for the simulations was the *CIE* standard *D65* illuminant [66] (see figure 3-6), and the spectral reflectances were 1700 samples from the Natural Color System (*NCS*) [62, 63, 64]. We selected these samples because they represent natural colors found in common scenes for a general purpose color correction application very well.

We were provided by the developers of TFD [111], with a set of 8 different *RGB* responsivities corresponding to 8 different biasing voltages (24 responsivities in total. See

figure 3-1). As candidates for filters included in the CFA, we used a set of 13 real filter transmittances from the *Andover corp.* web site [23] (see figure 3-7).

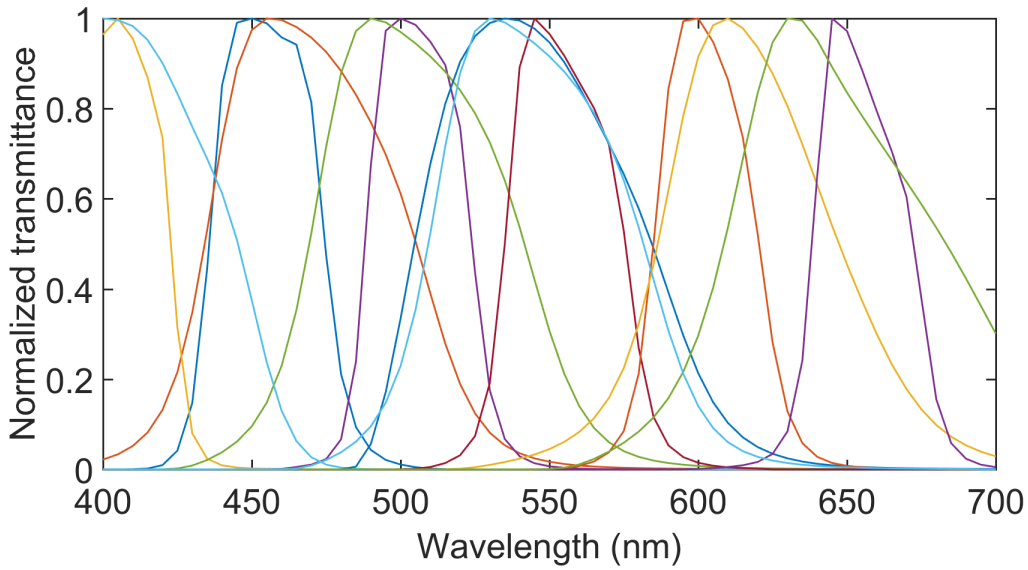

Figure 3-7: Normalized Spectral Transmittances of selected band-pass filters.

These spectral transmittances were selected so that they all covered the whole visible range of the spectrum with certain overlap between them. We carefully chose different bandwidths and maxima positions. The commercial references of these 13 selected filters are: *400FS40, 400FS70, 450FS40, 450FS80, 500FS40, 500FS80, 550FS20, 550FS40, 550FS80, 600FS40, 600FS80, 650FS40* and *650FS80*.

Combining all responsivities with all filters, we got a total of 312 channels. Out of them we wanted to select six, corresponding to only two different biasing conditions (which are the limits of bi-polarization as mentioned in the beginning of this chapter), which could best recover the spectral information of the samples imaged in the visible range.

For the channel selection, we have implemented the Voting Principal Feature Analysis method (*VPFA*). This method takes as input the sensor responses from all 312 channels of the 1700 *NCS* [62] color samples used, and performs principal component analysis (*PCA*, [193]) on them. Then we select a reduced number of projections of the data onto the principal component vectors (half of them) and cluster them into 6 clusters (which is the number of channels we aim to find) using k-means clustering method [67]. We calculate the mean point (center) of each cluster, and then select the filters corresponding to the vectors closest to each of the cluster centers. Due to the random initialization of k-means clustering method,

the clustering step is repeated for several iterations (50) and the selected vectors are voted. Finally the most voted ones which correspond to only two biasing conditions are selected. Each pixel of the *TFD* yields 3 channels responses. With *VPFA* we only selected the 6 best performing ones. This would correspond to system number 2. However once the capture is done, we already get the 3 channels per pixel. Therefore we also studied the possibility of using the other two channels per pixel. This situation corresponds to system number 10. Results showed that there is a slight improvement when considering all possible channels.

Furthermore, a two-shot system was also designed using the two full polarizations of the *TFD* resulting from *VPFA* and retrieving the information from all resulting channels. This corresponds to our proposed system number 1. After this process, the spectral responsivities of the channels selected were the ones shown in figure 3-5. As this figure shows, they span the whole visible range with some overlap. Overlap is important not to leave a gap in the spectrum without retrieved information [66]. However too much overlap could cause bad performance as well due to the consequent stronger effect of cross-talk between channels. We also simulated other systems with the same characteristics, but selecting randomly the combination of filters and sensors to get a set of responsivities which visually looks good (covering the spectrum, relatively narrow and with some spectral overlap). Out of the combinations we built, the one with the best color and spectral metrics was chosen. This corresponds to system number 11. We also compared it with the one resulting from the *VPFA*.

3.3.2 Error metrics

We used both spectral and colorimetric indexes for the evaluation of the performance.

- Goodness of fit coefficient *GFC* [161, 78, 121, 116], also known as the complement of Pearson distance [74]. It is a spectral performance metric calculated as shown in equation 3.15.

$$GFC = \frac{|\sum_{j=1}^{n} E(\lambda_j) \cdot E_R(\lambda_j)|}{|\sum_{j=1}^{n} [E(\lambda_j)]^2|^{\frac{1}{2}} \cdot |\sum_{j=1}^{n} [E_R(\lambda_j)]^2|^{\frac{1}{2}}} \tag{3.15}$$

  Where n is the number of spectral bands, $E(\lambda_j)$ is the estimated spectral signal to be compared with the reference spectral signal $E_R(\lambda_j)$. The *GFC* quality index is the

cosine of the angle formed by the two samples in the high-dimensional vector space of spectral signals. The closer the GFC is to unity, the better the estimation quality of the sample. The GFC index is independent of scale factors. In consequence, two samples differing only in scale but not in shape would result in a GFC of 1 (perfect match). If $GFC > 0.999$ the estimation was considered quite good and if $GFC > 0.9999$ the estimation was almost a exact fit [147].

- Root mean square error, $RMSE$ [95, 121]. This spectral metric focuses on absolute differences between the original and the estimated spectral signals. Thus, it is not independent of scale factors. For a perfect match we get $RMSE = 0$. It is calculated as show in equation 3.16.

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{j=1}^{n} (E(\lambda_j) - E_R(\lambda_j))^2} \tag{3.16}$$

- ΔE^*_{00} color difference. Also called *CIEDE2000* ([164]). When $\Delta E^*_{00} < 1$ units, the colorimetric performance was considered acceptable. We included this color metric because assessing how good the estimation of the spectral signal is in terms of color perception is also important. It could happen that two very similar spectral signals (with good spectral metrics), have their highest differences in spectral bands where our perception is most sensitive. In this case we would still perceive these two signals as different colors. On the other hand, two spectral signals with worse spectral metrics, but which main differences are in regions of the spectrum where we are less sensitive, could yield a very similar color, and so be acceptable for a colorimetric application.

3.3.3 Spectral reflectance estimation and evaluation

When we face the problem of spectral reflectance estimation from sensor responses, there are many algorithms proposed in the literature. In a previous work ([121]), we compared the performance of several methods for this task. Thise methods were: Pseudoinverse ([147]), Matrix R ([201]), Projection Onto Convex Sets (*POCS*, [170]), Radial Basis Functions Neural Networks (*RBFNN*, [10]), and Kernels ([73]). We will not explain the details of each of these methods in this thesis. More details about the working principles of each of these methods are presented in [121]. We summarize in table 3.2 the average color and spectral metrics

found when comparing those methods over the spectral estimation from simulated sensor responses of 1600 reflectances.

Table 3.2: Mean (and std) of spectral and color error metrics for the different spectral estimation algorithms compared [121]

Algorithm	Pseudoinverse	MatrixR	RBFNN	POCS	Kernel
GFC	0.9941	0.9962	0.9941	0.9959	0.9965
	(0.0064)	(0.0041)	(0.0337)	(0.0049)	(0.0083)
$RMSE$	0.0497	0.0487	0.0473	0.0416	0.0385
	(0.0216)	(0.0233)	(0.0238)	(0.0192)	(0.0217)
ΔE_{ab}	9.37	4.57	8.2807	5.7098	4.3313
	(7.26)	(4.0769)	(9.5163)	(4.1218)	(3.25)

We see in table 3.2, that Kernel method outperforms the rest of the methods in all color and spectral error metrics. Hence we decided to use this algorithm for our simulation experiments. In this previous study ([121]) for comparing the different algorithms, the color error metric used was ΔE_{ab} ([196]) instead of $CIEDE00$. We did so for direct comparison with other authors' publications. However we used $CIEDE00$ as a better recommendation from *CIE* for color difference calculations.

The estimation method used was regularized in-homogeneous polynomial kernel regression [72, 73, 71]. This is a non-linear mathematical regression method. A kernel function projects the input data into a new vector space, and afterwards a regularized linear regression is performed over the data in this new vector space. As explained in [71], this model needs to set two free parameters: the degree of the polynomial (d) and the regularization term (λ_r). The values of these parameters need to be optimized.

For the optimization and evaluation, a double 10-fold cross validation method was used, one nested inside the other. In the outer loop, we randomly separated the 1700 *NCS* samples into optimization and evaluation sets. The former was used to find the optimal values of the two parameters of the mathematical model, and the latter to evaluate the estimation of the spectral reflectances. In each iteration of the outer 10-fold, the optimization set was used in the inner 10-fold loop. A grid of parameters was built and every possible combination of both parameters was tested (exhaustive search) with the testing set corresponding to each loop of the inner 10-fold. For the parameter optimization, we used the *CIEDE2000* color difference formula as a cost function (see ΔE^*_{00} in subsection 3.3.2). Using a color difference

Table 3.3: Mean (and std) of polynomial degree (d) and regularization term (λ_r) kernel parameters values found in the optimization process for every system. The regularization term values are shown in logarithmic scale. Values of d are not integers because they are the computed mean.

System #	1	2	3	4	5	6	7	8	9	10	11
d	2.68	5.70	5.9	7.08	3.47	7.00	5.41	4.13	4.46	4.30	4.60
	(0.58)	(0.48)	(1.37)	(0.80)	(1.14)	(0.00)	(1.31)	(0.83)	(0.59)	(0.48)	(0.52)
$log_{10}(\lambda_r)$	-2.85	-6.06	-4.29	-5.81	-4.15	-4.93	-8.19	-4.39	-3.17	-4.23	-4.43
	(0.43)	(0.13)	(1.06)	(0.87)	(1.94)	(0.22)	(1.19)	(0.64)	(0.68)	(0.42)	(0.50)

formula for parameter optimization gives to colorimetric performance of the results a slight advantage over spectral performance compared with using a spectral metric. The opposite happens using a spectral metric as a cost function. In average, using the color metric yields slightly better results. Therefore the metric we chose for our application domain was *CIEDE2000*.

After the 10 iterations of the inner loop, the average of the parameters found was calculated, and those were used for evaluation in the outer loop. The standard deviation was also calculated to check if the distributions of the best parameters found were stable. Table 3.3 shows the means and standard deviations (in parenthesis) of both kernel parameters found in every system out of each of the 100 folds in total (10 times 10 folds). Since the regularization parameter was optimized looking for the value that performed best in different orders of magnitude, the mean and standard deviation calculated for it are shown in logarithmic scale. The parameter grid studied by exhaustive search consisted in all possible combinations between $d \in [1, 2, 3, 4, 5, 6, 7, 8, 9]$ and $\lambda \in [10^2, 10^1, 10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}, 10^{-9}]$.

After the 10 iterations of the outer loop, the average error metrics were calculated and these were the values with which we compared the different systems. By doing things this way, we ensured that samples used for training would never be used for testing, which could lead to over-fitting and thus to overestimating the quality of the recovered samples obtained from the camera responses of our system.

3.3.4 Experiments performed

Since we wanted to test separately features like the tunability of *TFDs*, or the presence of different types of *CFAs*, we performed 4 experiments to test the following hypotheses:

- Hypothesis 1: the proposed system performs better than systems based on other types of sensors such as monochrome sensors or multi-shot systems based on *RGB* scientific cameras plus some filters (see subsection 3.4.1).

- Hypothesis 2: the proposed system performs better than systems based on *TFDs* that do not use the designed *CFA*. Therefore, this *CFA* helps to improve the performance of *TFDs* (see subsection 3.4.2).

- Hypothesis 3: using the bi-polarization property of a *TFD* sensor, the performance can be improved without the need to increase the number of shots taken (see subsection 3.4.3).

- Hypothesis 4: the proposed system performs better than a similar system in which the channels selection was not done by *VPFA* (see subsection 3.4.4).

Hypothesis 1 demonstrates the capabilities of the proposed design. Hypotheses 2 and 3 try to go deeper into the reasons justifying the superiority of our *TFD*-based proposed system. Hypothesis 4 demonstrates that using *VPFA* for selecting candidate filters and channels is worthy.

3.4 Results

Table 3.4 shows the average error metrics obtained for the 11 systems simulated. In the following subsections these results are commented on.

Table 3.4: Results for all systems studied. Mean (and standard deviation).

System #	1	2	3	4	5	6	7	8	9	10	11
ΔE^*_{00}	**0.23**	0.51	1.26	0.66	0.27	0.96	3.15	1.89	0.41	0.38	0.35
	(0.16)	(0.35)	(1.3)	(0.47)	(0.18)	(0.73)	(2.27)	(1.32)	(0.34)	(0.26)	(0.24)
GFC	**0.9997**	0.9987	0.9992	0.9982	0.9992	0.9986	0.9976	0.9982	0.9991	0.9996	0.9994
	(0.0028)	(0.0043)	(0.0022)	(0.0030)	(0.0030)	(0.0079)	(0.0038)	(0.0033)	(0.0010)	(0.0012)	(0.0021)
RMSE	**0.0064**	0.0101	0.0094	0.0103	0.0083	0.0113	0.0169	0.0143	0.0082	0.0064	0.0067
	(0.0063)	(0.0087)	(0.0082)	(0.062)	(0.0099)	(0.034)	(0.0092)	(0.008)	(0.004)	(0.0043)	(0.0053)

The best performance for the three error metrics evaluated corresponds to system number 1 (written in bold in table 3.4).

We also present the quality indexes for all systems compared in this study in figure 3-8.

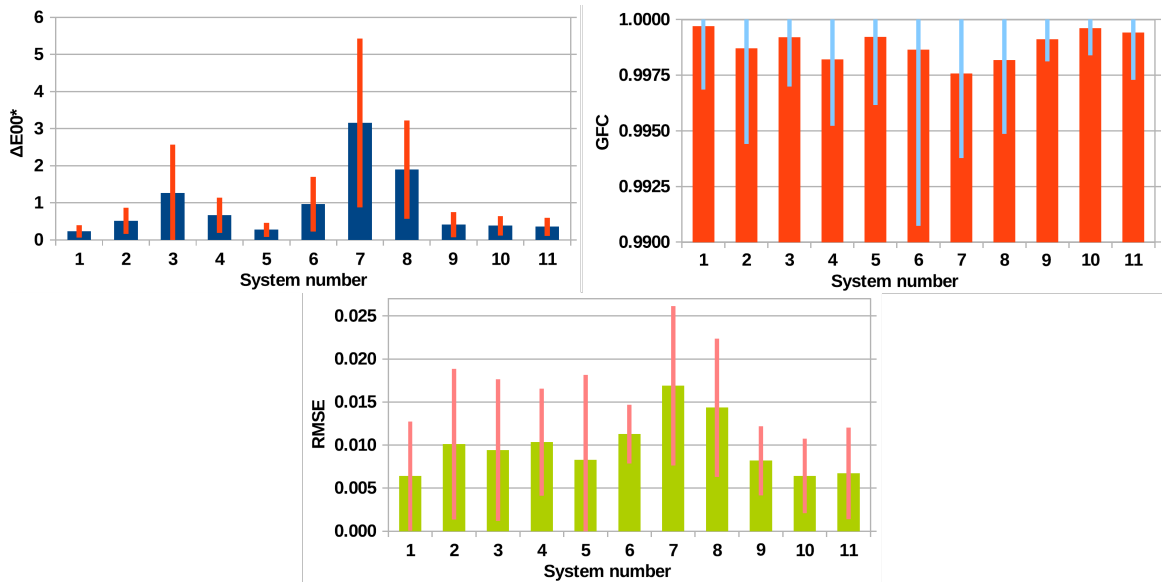

Figure 3-8: Color and spectral error metrics for all systems. The bars show the mean value and the lines the standard deviation centered on the mean.

3.4.1 Simulation experiment 1: Superiority of TFDs

This experiment was performed to check whether our proposed system yields better accuracy in spectral reflectance estimation than systems based on other types of sensors using the same kind of *CFAs* and other different ones. We included the 2-shot system made up of a *RGB Retiga* camera and a cut-off filter, studied in previous work [120], and we also designed a new system in which we divided each of the *R, G* and *B* responsivities of *Retiga* camera into two halves (see Figure 3-5 system 5) with theoretical custom-designed comb-shaped optical filters. Results are shown in Table 3.4 for systems number 2, 3, 4 and 5.

We see that system number 2 performed better both for spectral and color metrics than systems number 3 and 4. The system formed by monochrome silicon sensor plus the *CFA* performed worst. This means that the good results obtained by system 2 were not only due to the *CFA*, but also that the tunability and bi-polarization property of *TFD* sensor helped to improve the performance. The systems formed by *RGB Retiga* camera plus filters yielded good results, especially the one with the two comb-shaped filters which worked best. This is in agreement with previous studies [121, 120]. This comparison was made in the same conditions as far as number of channels is concerned. However, when we simulated system number 2, we were neglecting the information from 2 channels from each pixel. This meant that we are capturing 18 channels, but only using 6. If we add this available information to

the recovery (we get system number 10) the results improved. They were closer to system number 5 colorimetrically and better spectrally. Besides we would only need one shot instead of two, and there would be no need to mechanically move in and out any filter.

Moreover, we designed a two shots fully-polarized system, and we retrieved the information from all channels under each pixel, increasing the number of channels to up to 36 in just two shots. This was the strategy of system number 1. In this case we used two shots which were sequentially captured just by switching the biasing voltage of the *TFD*. These results outperformed the ones from system 5. In practice systems 1 and 5 have the same number of shots (2 shots). For system number 5 the capturing process cannot be real time since the filters need to be switching all the time. In the case of the proposed system, even also being a 2-shot system, the switching of the biasing voltage occurs so fast and electronically controlled that it can be considered as a snapshot system in practice, as long as the lighting conditions in the scene being imaged allow short exposure times. We thus propose system number 10 as a good and convenient option to be used in applications which require fast capturing times. And if we could afford longer capturing times, then system number 1 is our proposal to enhance performance by increasing the number of shots.

3.4.2 Simulation experiment 2: Superiority of a CFA-based system

In this experiment, we wanted to assess whether including the *CFA* in the *TFD*-based system is the key factor leading to the superior performance of the system shown in first experiment. For this we simulated different systems using *TFD* sensors. One of them used no filters in front (system number 7). The other used the filters present in a common *RGB* scientific color camera (system number 6). This was done as we had found out that the *RGB* *Retiga* camera was working very well and that these filters drastically narrowed down raw responsivities. The results of this experiment are shown in Table 3.4 for systems number 2, 6 and 7. We see that the system we propose performs best for both spectral and color metrics. Including other kinds of filters such as *RGB* color filters (system number 6) yielded reasonable results, but still far from the ones reached by system number 2. A system using only two *TFD* polarizations such as the proposed one, but with no *CFA* (system number 7), yielded quite bad results in comparison. Even pushing this strategy to the limit of taking many shots with different polarizations (up to 4 shots with bi-polarized *TFD* as in system number 8), the performance was far from our proposed system. Therefore it did not

outperform a system with only one shot, two *TFD* polarizations and a *CFA*.

If the difference is clear when the experiment is performed for systems with the same number of channels, the difference is still more evident if we use the systems number 10 or 1. We pointed out their advantages in subsection 3.4.1. Therefore we can conclude that adding in a *TFD*-based system, a *CFA* which has been calculated to be the best choice out of a set of available color filters, helps to improve the performance for spectral reflectance estimation, over systems without *CFA* or with other filters that are not specially selected.

3.4.3 Simulation experiment 3: Bi-polarization vs full-polarization

In this experiment we wanted to assess whether using the capability of bi-polarizing the *TFD* sensor helps with its performance. So the easiest way was to test it by comparing systems number 9 and 10. The results are shown in table 3.4.

We see how using two polarization states of the *TFD* improves the system performance slightly for both color and spectral metrics. Therefore it is worth adding a second biasing voltage, since in practice, the acquisition process is the same. Apparently, using responsivities from different biasing voltages perform better than using them all from the same biasing condition.

3.4.4 Simulation experiment 4: VPFA performance

This last experiment was carried out to check whether it makes sense to invest time in selecting the filters using *VPFA*, or simply any combination of them, randomly chosen, covering well the visible range of the spectrum and having some little overlap between bands, could yield better or similar results. Out of ten random combinations, we selected the one giving the best results (system number 11) and compared it with our proposed system (number 1). Results show that the *VPFA* procedure improves the filter selection process.

3.5 Conclusions

Two different approaches for a multispectral imaging system have been proposed which are based on a novel technology still under development. These are *TFD*-based systems which exploit the tunability and bi-polarization properties of this new type of sensor. They are

coupled with a *CFA* whose filter transmittances have been selected via *VPFA* method out of a set of real available color filters from a commercial database. Four simulation experiments have been conducted to prove if the proposed systems outperform others with similar characteristics but using other design strategies. We aimed to assess if the combination of *TFD* plus *CFA* works better than any of them separately. We also studied whether we could use different architectures depending on the requirements of a particular application. We found out that the proposed systems perform both colorimetrically and spectrally better or as well as other systems, while offering an easier and more elegant solution to the problem of spectral imaging. Some systems yield close spectral results, and only one of them comes significantly close in the color metrics. However this system is made up of a scientific *RGB* color camera plus two ideal filters that are placed in front of it alternatively (system number 5), so each capture would need two shots and the mechanical switch between filters from the first shot to the second. This would make the real time capture unsuitable.

One of our proposed systems (system number 1) is also a two-shot system, but the tuning of the responsivities is so fast and easy that we could still use it for real time spectral imaging if the conditions of the amount of light in the scene being imaged allowed it. Otherwise we offer a single shot alternative system (system number 10) which would reduce drastically the time needed for capturing and would still give good colorimetric and spectral results. This study demonstrates the potential of *TFD* sensors as candidates to be part of spectral imaging systems that are portable, real-time, versatile and low cost, as soon as they can be implemented as part of a real capturing system.

The work and results presented in this chapter were published in [123] which is the publication number 1 in the Appendix A.

3.6 Future work

After performing these theoretical experiments, simulations showed the potential of *TFD* sensors combined with *CFAs* for spectral imaging. The next step would be to implement a real camera featuring a *TFD* sensor and the *CFA* we have designed for it. Unfortunately, by the time this study was done, *TFD* sensors were still in a prototype phase of development. Therefore it was still too early for this step.

As the next step in this thesis was to move on to *HDR* imaging and combine it with

Figure 3-9: *Pixelteq Spectrocam VIS-NIR*. Left: 8-slots filter wheel. Right: camera with filter wheel attached, and 16 band-pass filters covering the visible and near infrared range of the spectrum $[400nm, 1000nm]$.

spectral imaging, we decided to switch to different spectral imaging systems which. These systems do not offer so many advantages as the *TFD*-based system proposed, but they could still be easily tuned in order to perform different experiments combining all imaging techniques studied in this work. For this reason we continued our spectral imaging experiments with a system based on a monochrome camera and a Liquid Crystal Tunable Filter ($LCTF$), and another system based on a scientific monochrome camera synchronized with a filter wheel. The latest was *Pixelteq SpectroCam VIS-NIR* (see figure 3-9). We will talk about the experiments done with these imaging systems in chapter 5, but before that, we will move on to the techniques developed regarding *HDR* imaging in next chapter 4.

In order for the light to shine so brightly, the darkness must be present.

Francis Bacon.

4

Adaptive Exposure Estimation for HDR imaging

Content:

Chapter 4

Adaptive Exposure Estimation for HDR imaging

As explained in section 2.2, there are many different approaches for tackling the *HDR* imaging problem. Many authors are focusing in developing new camera or sensor hardware architectures ([145, 144, 3, 80, 171, 156, 191, 99, 98]). However we wanted to focus rather in the software methods. As our final goal is to combine *HDR* imaging techniques with spectral imaging techniques, including special sensors or camera set ups could affect the capability of our system for spectral imaging. Therefore, it is more convenient to exploit some of the already existing hardware architectures for spectral imaging, and try to incorporate to them an *HDR* software method.

The most common software method is multiple exposures. It is popular because it is simple and can be carried out by any imaging system. As explained in section 2.2.2, this methods consists on capturing different images of the same scene. The only requirement is to change the exposure settings of the camera between shots. This can be done either by changing the exposure time (which is the most popular technique), or by changing other parameters like the *ISO* gain, the aperture, the presence of neutral density filters (*ND*-filters), etc.

Out of all these options, we focused in the variable exposure time option. Even though the other options are plausible, we discarded them because of the following reasons:

- *ISO* setting in digital sensors is a electronic gain applied at pixel level. Increasing it results in a higher voltage signal. However it also causes a higher amount of noise due

to electronic processes of amplification.

- Aperture setting is easily adaptive to scene content. Using an automated aperture system, which nowadays is embedded in many commercial camera systems, we can implement an algorithm within camera chip to open or close the aperture as the capturing process requires it. The problem is that when we increase the aperture of an optical system the depth of focus is reduced. This might bring defocussing problems, specially if we are working at close distances. The effect can be even worse if we are dealing with spectral imaging systems with narrow spectral responsivities (using interference or liquid crystal tunable filters). In these cases, due to chromatic aberration combined with reduced depth of focus, the system might produce blurry images in some spectral bands, or even in objects at different distances within the same spectral band image.

- Regarding the presence of *ND*-filters, this approach is indeed a hardware approach. Since we would in all likelihood be using some filters already as part of the spectral imaging capture device, we decided to avoid including extra filters as much as possible.

We decided then to take the multi-shot approach changing the exposure time between shots. We wanted to develop an algorithm to make the *HDR* image capture as automatic as possible. Despite the content of the scene, we aimed for an algorithm which could search for the different exposure times with no need of user intervention.

When we capture an *HDR* scene, either with a monochrome, *RGB* or spectral imaging system, we always need to decide which exposure time we are going to use. Auto-exposure settings available in most commercial cameras and smartphones work fine for a single *LDR* capture. This means that they analyze the content of the scene (usually in live view, which means capturing several shots, or they resort to cues like the reading from a built-in photometer in high end cameras). They focus in a region of interest (*ROI*, e.g. the whole image, the center region, etc). If the pixel values in this *ROI* are too dark, the exposure time setting increases its value. If on the other hand it is too bright, the exposure time setting decreases its value. However, in most outdoors conditions, we could find regions of the image where the incoming light is too bright, and regions of the image where the shadowy areas are too dark. Thus, with a single shot, whatever its exposure time is, we can not correctly expose the whole information in our *ROI*. We show an example of this in figure 4-1. Both

images are taken from the same scene, using the exposure time values 20 ms (calculated by the auto-exposure setting of the camera) for the image in the left and 500 ms for the image in the right. The aperture setting was $f5.6$. We can observe underexposed areas in the left image and saturated areas in the right image. Whatever exposure time value we choose in between these two values used, would yield underexposed and/or saturated pixels as well.

Figure 4-1: Example of outdoors scene captured using *Canon EOS 7D* commercial *RGB* camera. Left: image captured using 20 ms of exposure time. Right: image captured using 500 ms of exposure time. The red ellipse highlights an underexposed region, and the green ellipse highlights a saturated region.

Usually, the dynamic range of a given sensor (i.e. the ratio between the maximum and minimum irradiance impinging on the sensor that produces an effective response) is much lower than the dynamic range found in natural open air scenes. The dynamic range of these scenes (ratio between the maximum and minimum radiances emitted by the objects in the scene) can vary from 2 to 8 orders of magnitude depending on the season and scene content [125] (see table 4.1 from [190, 155], showing the typical luminance levels present in common scenes). The human visual system can simultaneously adjust to a difference of up to 3.73 orders of magnitude (or logarithmic units) [106], when adaptation is accomplished. However, most imaging and display devices can only account for barely 2 orders of magnitude in a single image (either for capturing or for displaying) [197, 163].

Consequently, when we use our camera to capture a digital image of a scene with *HDR* content, we can not know in advance which exposure times would be useful for composing the *HDR* image afterwards. We could just take a large number of images with different exposure times (or even all available ones in the camera), and then use all of these *LDR* images to compose the final *HDR* image. But this option could take too much time and could be very computationally demanding, so it is not always feasible.

Table 4.1: Average luminances present in common scenes. Table data retrieved from [155]

Condition	**Illumination:** $(\frac{cd}{m^2})$
Starlight	10^{-3}
Moonlight	10^{-1}
Indoor lighting	10^{2}
Sunlight	10^{5}
Max. Intensity *CRT* monitors	10^{2}

Our aim was to present a method for the selection of a set of exposure times (bracketing set), which allowed us to retrieve useful information from all pixels within our *ROI* (or at least from as many of them as possible). As we will explain in section 4.1, other proposed methods also aim to find bracketing sets for HDR capture.

A correct HDR capture is very important for capturing useful image information from scenes with *HDR*content, either indoors ([176, 179]) or outdoors [84]. Thus, a digital camera can be a useful tool for composing a *HDR* radiance map of these scenes. Of course, if the scene captured has very dark regions that need long exposure times to be correctly exposed, then it is important that there is no relative movement between camera and scene content during the capturing time. If small movements happen, there are ghosting-compensation techniques to correct for artifacts [37]. In this thesis we did not focus in these issues, and assumed that the content in all scenes captured was static.

We wanted to design a method with the following characteristics:

- Blind: no information from the scene content is needed to be known a-priori. Thus the method would work whatever the scene content is as long as it is within the physical limits of the camera.

- Adaptive: the method adapts to scene content by dynamically adjusting the required exposure times.

- Full range: the complete dynamic range of the scene can be correctly exposed if it is within the physical limits of the camera and the veiling glare conditions allow it.

- Universal: works for any camera provided that it allows us to control exposure settings and trigger.

- On-line: the exposure times are calculated as the capturing process is ongoing and every single shot acquired is used in the *HDR* radiance map generation. Therefore there is no need to capture some shots in advance to make any estimations or assumptions about the content of the scene, and the method will not waste time capturing shots that are not useful for the composition of the *HDR* image.

- Minimal: the method gives as default output the minimal bracketing set (i.e. the bracketing set that has minimum number of shots, yet recovering the full dynamic range of the scene).

- Tunable: the method can be tuned to yield longer exposure times with a higher *SNR*. This tunability is introduced to control the amount of overlapping between consecutive exposures, increasing *SNR* in the resulting *HDR* image at the cost of increasing the number of shots taken, and hence the capturing time.

- Flexible: in some applications, it could happen that we are not really interested in the very brightest or darkest areas of a scene, even if they are inside our *ROI*. Then can introduce the possibility to set the percentage of pixel population that we can accept to be underexposed or saturated in all shots.

In the following section, we will study the previous approaches proposed by other authors to solve the problem of finding an appropriate bracketing set for *HDR* imaging (see section 4.1). Afterwards we will explain the details of our proposed method (see section 4.2). Later on, we will show the experiments performed to test our method against the the most similar one developed so far ([7]). Finally we will draw our conclusions (section 4.4) and the future lines of research (section 4.5).

4.1 Previous approaches

Several approaches have been proposed in the literature for solving the problem of finding the exposure time values for *HDR* image capture via multiple exposures. In [18], the authors proposed an scheduling for capture times. They were assuming a known illuminant in the scene, which in practice is a rather non-realistic assumption. In [51], they proposed a method to simulate the response of any camera (linear, logarithmic, gamma function, etc) using a single camera with known Camera Response Function (*CRF*), by just selecting a set of

exposure times. So their aim was not to find a minimum bracketing set for radiance map generation. In [172], they proposed a method for capturing *HDR* images of the sun and sky. They thresholded the images and checked if there were saturated and/or underexposed pixels. If so, they added new shots by increasing or decreasing the exposure a fixed amount of 3-stops. This approach is not adaptive to the scene content and could lead to situations where the exposure times are not well fitted to cover the full dynamic range of the scene. In [9], the authors proposed a method for overcoming the limitations of mobile devices for *HDR* imaging. Their proposal was also done by iteratively trying every available value of exposure time and afterwards decide which is the correct one. Therefore many images need to be taken before a bracketing set is selected, which makes it not an on-line or adaptive approach.

In [7], they proposed a method for finding minimal bracketing sets for *HDR* capture. Firstly they studied how the camera responds to radiance using every available exposure time. Then they selected only those exposure times that completely covered the full dynamic range of the camera, with certain overlap. This is the so-called Minimal System Bracketing Set (*MSBS*), and whatever the content of the scene being imaged is, using all these exposures will always cover the full radiance range that the camera can effectively acquire. To adapt this *MSBS* to scene content, they proposed to select a subset of it called Minimum Image Bracketing Set (*MIBS*) by capturing a first shot with an intermediate exposure time (that belongs to the *MSBS*), and checking if there are saturated or underexposed pixels. If so, they add the next exposure time included in the *MSBS* until the full dynamic range is covered. Though the underlying idea in this method is similar to the one proposed in this thesis, it is still not totally adaptive to scene content since they limit the exposure times selected to those belonging to the *MSBS* (usually just 4 or 5). We believe that the same scenes could be captured with less shots and shorter exposure times, yet covering their full dynamic range, and we have demonstrated this by comparing the results of our method with Barakat et al's clairvoyant algorithm, using their *MIBS* approach. More details about this method are given in section 4.3.1.

In [49], they proposed a method which used as stating point the mean *HDR* irradiance histogram of the scene being captured. This method worked only for linear cameras, and they used a greedy algorithm, iteratively capturing the same scene many times until they got the optimal *SNR* solution. Thus, this method is not on-line. In [83], the authors used

a mathematical method based on training for *HDR* exposure time selection. They assumed linear sensor response and known noise sources in the capture, which is not always a realistic scenario. They did not really aim for finding minimum bracketing sets, but for optimal *SNR* in the *HDR* radiance map.

In [46], they adapted *HDR* imaging to mobile devices as well. They merged *LDR* images iteratively two by two. If there are still saturated or underexposed pixels, they keep adding the next available exposure time that the camera offers. In [69], the authors proposed a method to calculate a bracketing set that is optimal in terms of *SNR* (see figure 4-2) but not minimal. They just try to fit it within a given time budget, by varying both the exposure time and the *ISO* settings of the camera. Besides, their method assumes linear raw sensor responses only, and known information about scene radiance content.

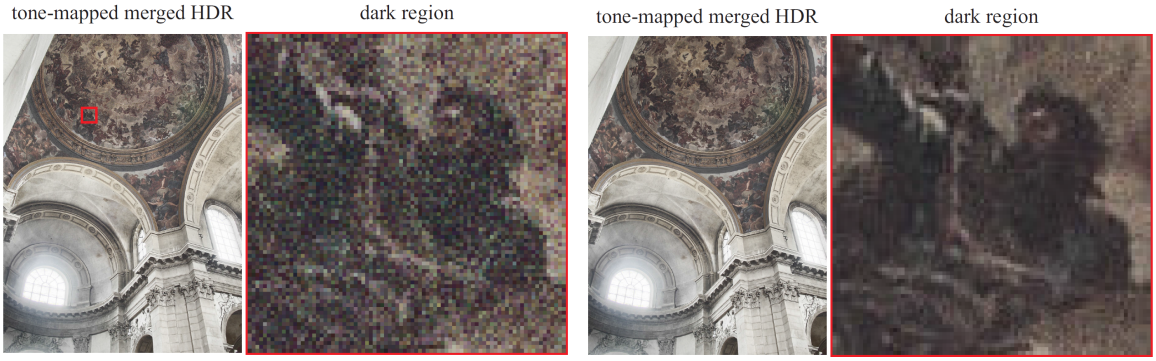

Figure 4-2: Left: Reference 3-shot exposure bracketing sequence, with images 2 stops apart. Right: *SNR*-optimal 3-shots capture sequence maximizing worst-case *SNR* in the same time budget. Images taken from [69].

Later on, in [38], the authors proposed a method which takes advantage of mobile phone camera *APIs* (Application Programming Interface). They programmed the automatic histogram calculation in mobile phones to build a reduced *HDR* histogram of the scene, which will be the target to be captured. This is however not possible if the camera used for the captures does not feature this automatic process. Besides, if the scene imaged has very dark regions, the long exposures needed to create this histogram make the process slow. Moreover, the method is not on-line. After this histogram is calculated, they capture several exposures of the scene and then study many possible combinations of them until the optimal one is selected.

Afterwards, in [56], they followed the lines previously proposed by the authors in [97], who implemented a method for sequentially adjusting the exposure for real-time *HDR* video.

Both methods in [56] and [97], are iterative and limited to only two shots for building the *HDR* image. Later on, in [57, 58], they developed their algorithm further for *HDR* video capture. They propose a weighting function which punishes low signal values, and iteratively try to fit the histogram of the incoming captured frames to this weighting function by adjusting the exposure time. This approach is smart for a video system. However it needs some iterations until convergence is reached, and also assumes that histograms of the captured images match somehow the shape of the weighting function (unimodal histogram).

In [54], they proposed a Fibonacci-series-based bracketing set determination algorithm in which each exposure time is the sum of the previous two. However, their purpose was not full dynamic range recovery, but image registration for *HDR* video. Afterwards, in [86], the authors proposed a method to be implemented in-camera chip. They get information from the camera metering system (i.e. analysis of certain *ROI* to check if it is correctly exposed).

First they capture an image exposing the whole image as good as possible. Afterwards they capture two more images. One of them has a longer exposure time and the other one a shorter exposure time (over and underexposed respectively). The way they calculate the new exposure times for the two new images is by finding the areas in the image with most underexposed and saturated pixels, and extracting the camera measurement data in those regions. This method however does not ensure that the whole scene is correctly exposed. It is limited to three shots and can not be implemented for cameras with no metering system. Therefore it is neither universal, nor full range or adaptive.

After this extensive literature research, we conclude that only the method proposed in [7] has similar features to the one we intend to develop. However it still has disadvantages that we will comment later on in section 4.3.1.

4.2 Proposed method

Since we wanted to propose a universal method applicable to any imaging system, we implemented it in different cameras. The first series of experiments (explained in this chapter), were performed with a scientific monochrome camera (*Retiga SRV 1300*), a scientific *RGB* camera (*Retiga 1300C*), and a commercial *RGB* camera (*Canon EOS 7D*). The three cameras are shown in figure 4-3. The technical details of each of these cameras can be found in the devices and instrumentation section of the web site of our research group ([148]). The

method was proven to work satisfactorily for all of them. In this chapter we will show the results obtained for the commercial *RGB* camera. In the following chapter 5, we will show the results of applying the proposed method to multispectral imaging systems, including some modifications.

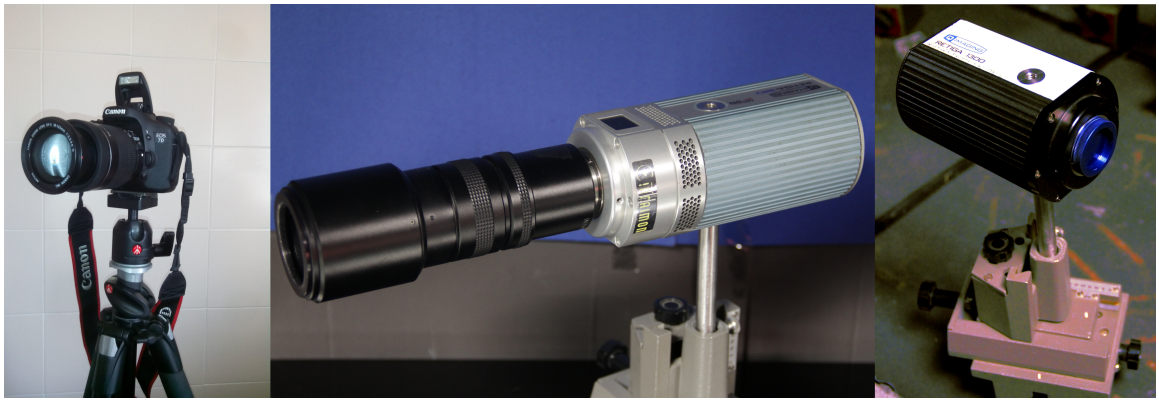

Figure 4-3: Left: *Canon EOS 7D* commercial *RGB* camera. Center: *Retiga SRV 1300* monochrome scientific camera. Right: *Retiga 1300 C, RGB* scientific camera.

We drove our *Canon EOS 7D* camera from our laptop via *USB* port, using the open-source libraries called *GPhoto2* ([1]), which were accessed through our algorithm implemented in *Matlab R2014a* and working on-line. Thus, the images captured were processed in the computer on the spot as the capture was ongoing.

The method proposed is full range, because it finds a bracketing set which covers the full dynamic range radiance map of the scene. The *HDR* radiance map is unknown a priori, therefore we can not know in advance how many shots are we going to need to expose it completely.

In figure 4-4, we show an example of *HDR* relative radiance histogram. This is the target histogram we have to expose completely by capturing multiple shots with different exposure times.

We can see in figure 4-5, the *CRF* of the camera we are using to capture the scene.

When we fix an exposure time and we capture the first shot, we block the *CRF* of the camera over a window of the *HDR* relative radiance histogram. We see an example of this in figure 4-6 (top).

We see how, if we increase the exposure time, the areas of the scene which would be well exposed are those darker than in the previous shot. Therefore the exposure window in the second shot is shifted to the left in the histogram (yellow *CRF*). On the other hand,

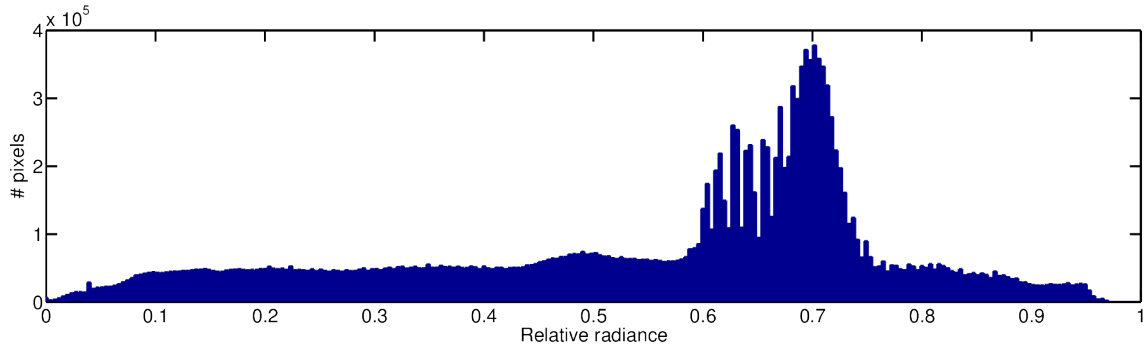

Figure 4-4: *HDR* relative radiance histogram of the scene to be captured. This information is unknown a priori. We plot it for the explanation of the *AEE* method.

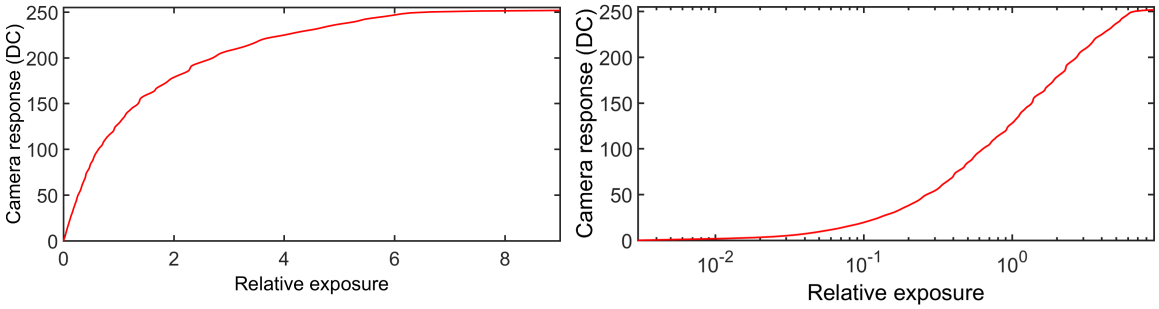

Figure 4-5: *CRF* of *Canon EOS 7D* camera in 8-bits *jpeg* mode. Left: linear scale. Right: logarithmic scale.

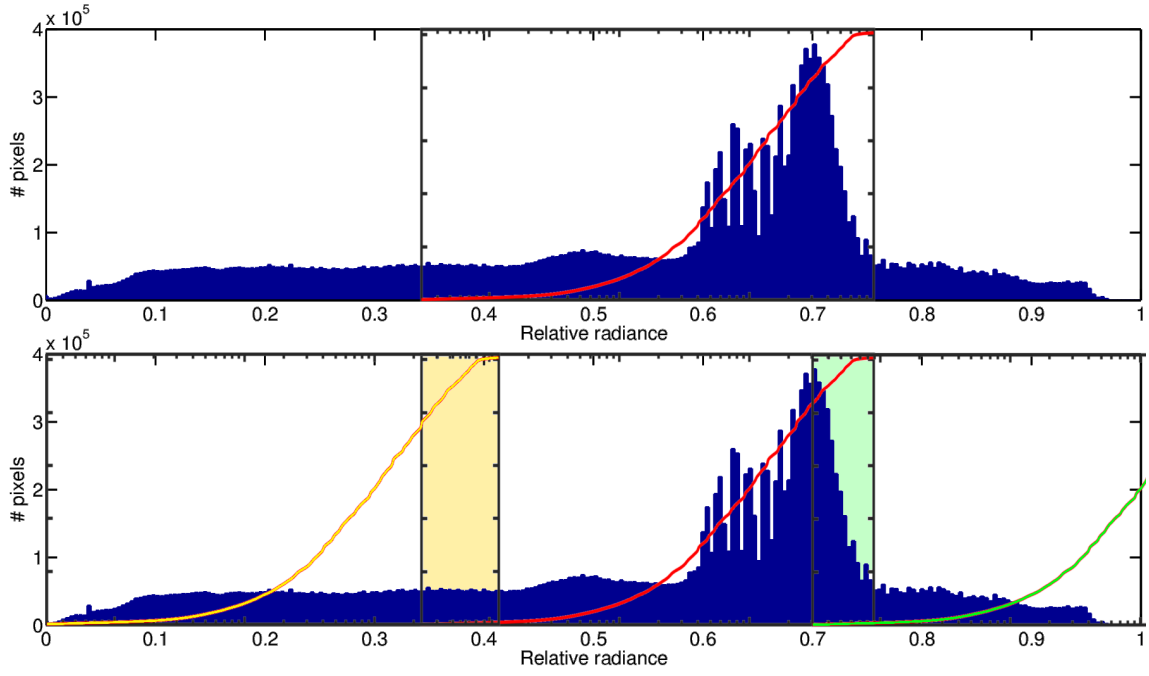

Figure 4-6: *HDR* relative radiance histogram superimposed with *CRF*. Top: first shot. Bottom: second and third shots. Note the overlapping radiance areas covered in two shots colored in yellowish and greenish color.

if we decrease the exposure time, then the exposure window would be shifted to the right, exposing well the light-most regions of the scene (green CRF). We then see how we can slide the exposure window along the histogram by just changing the exposure time. In the example plotted in figure 4-6, we see how there are two areas (shaded in yellowish and greenish colors), that are covered by two exposure windows at the same time. This overlap works fine for *HDR* image reconstruction. But if we would choose shorter or longer exposure times (thus sliding the exposure windows further towards right or left sides respectively), we could risk leaving some gap in the radiance histogram, not covered by any shot. Therefore the information in this range of radiances would be lost. A *MIBS* for this scene could be any combination of three shots with certain overlap between them (since any combination of only two shots would not be enough to cover the whole radiance range without overlap). However, we can say so, because in this example we know the *HDR* relative radiance histogram in advance. As this information is unknown in real scenarios, after the first shot (red CRF), the way to maximize the possibilities to cover the whole dynamic range with the least number of shots would be to push the CRF of the next shots as far as possible yet keeping certain overlap. This way, with the minimum number of shots, we could cover the widest possible dynamic range. But how could we choose these new exposure time values?

As explained in section 2.2.2, the CRF is a function which relates the response of the camera, in digital counts (DC), with the exposure that the sensor receives. This function depends on each camera, and even could be different for different settings of the same camera (e.g. camera working in *raw* mode or in *jpeg* mode, etc). Knowing the CRF of our camera is a key factor to build the radiance map.

The computation of the CRF of a camera is a well known procedure ([26]). The CRF of our *Canon* camera working in *jpeg* mode is shown in figure 4-5. Both plots represent the same function. However the one on the left is in linear scale, and the one on the right in logarithmic scale. Having a look at the one in the left, we see how the CRF is clearly not linear.

The exposure axis is in relative units and it is normalized so that the center of the DC values ($128DC$ for 8-bits images) corresponds to a relative exposure value of 1. The function is the same for the 3 color channels R, G and B of the camera, since it is a property of the sensor. Therefore we process the 3 color channels together like it is done in [38]. For each LDR image we capture, we know the exposure time (Δt) used as well as the digital counts

(DC) for each pixel and channel. Therefore, using the CRF we can easily work out the relative irradiance (E) by computing a simple ratio shown in equation 4.1.

$$E_{i,k,e} = \frac{H(i,k,e)}{\Delta t(e)} \tag{4.1}$$

Where $H(i,k,e)$ is the relative exposure for pixel i, in color channel k, and for exposure index e (or number of shot). Once we have captured an initial image with a known exposure time Δt_0, the CRF relates H_0 with DC_0 as shown in equation 4.2.

$$DC_0 = \mathrm{CRF}(H_0) = \mathrm{CRF}(E_0 \cdot \Delta t_0) \tag{4.2}$$

Therefore we can work out the relative irradiance value of a point of the image (E_0, calculated as shown in equation 4.3), knowing the CRF, the Δt and its DC value in the first shot (Δt_0 and DC_0 respectively).

$$E_0 = \frac{\mathrm{CRF}^{-1}(DC_0)}{\Delta t_0} \tag{4.3}$$

Where CRF^{-1} refers to the inverse CRF function which always exists since CRF is a monotonically increasing function (a higher radiance/irradiance would never result in a lower sensor response). Then, to shift the sensor responses DC_0 to this same irradiance value E_0, into a new value DC_1, we just have to workout which new exposure time Δt_1, is needed for a new shot, like equation 4.4 shows.

$$\Delta t_1 = \frac{\mathrm{CRF}^{-1}(DC_1)}{E_0} \tag{4.4}$$

If our camera has only a limited set of available exposure time values (like our *Canon* camera), we can select the available value which is closest to the calculated one.

We already have a tool to control the values of sensor responses, by tuning the exposure time used to acquire the images. In other words, we know now how to slide the exposure window as shown in figure 4-6. Now we explain how to use it for our purpose of optimizing the HDR capture. For this aim, we propose a method based on cumulative histograms of the scene and inspired on the work by Grossberg and Nayar in [52], originally applied to pixel selection for CRF computation. If the scene content does not change, then the same value of percentile of population in the cumulative histogram will correspond to the same

areas in the image. In figure 4-7, we show several cumulative histograms of the same scene captured using different exposure times.

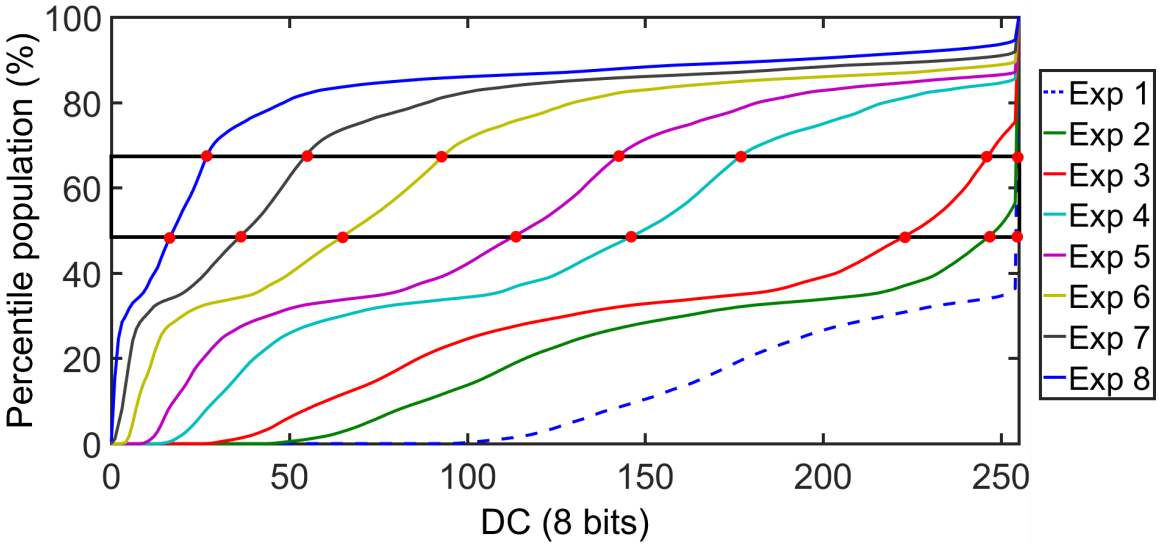

Figure 4-7: Cumulative histograms of the same scene captured using different exposure times and plotted in increasing value of exposure time from *Exp 1* to *Exp 8*.

The horizontal lines with red circles represent the same areas of the image. Depending on the exposure (8 shots in this example named in increasing value of exposure time from *Exp 1* to *Exp 8*), the sensor response values for the same areas of the scene (red circles) increase with exposure time. If a given percentile is below some exposure value for a given exposure time, then, for a different exposure time, the same percentile will correspond to a different exposure value but they will still keep its location within the scene. Therefore, the points where the horizontal lines intersect the histograms, report information corresponding to the same areas of the scene. Our idea is to shift sensor responses by calculating exposure times, to control the sensor responses to pixel populations of key percentile values. As a starting point, we calculate the cumulative histogram of the image captured with the automated exposure of the camera. But in principle any image can be used as starting point as long as it has some pixels that are neither saturated nor underexposed.

We sample the scene's radiance using the CRF of the camera between two DC levels. Unless the scene has a very reduced dynamic range, there will be pixel values below and above these DC values. Since in the default version of the algorithm we aim to find the $MIBS$, we have set the low level (LO) to 3 DC and the high level (HI) to 252 DC for considering a pixel underexposed or saturated respectively out of these bounds.

Thus, any pixel with camera responses out of the [LO,HI] range will need to be re-captured with different exposure settings.

We introduce now two novel features of our method. One is the possibility to set a tolerance for the percentage of useless pixels. If we choose 0% tolerance, the algorithm will look for longer or shorter exposure times if at least one pixel is underexposed or saturated, like Barakat et al. proposed [7]. However, for some scenes we can set a different tolerance threshold accept a certain amount of incorrectly exposed pixels (e.g. when we directly image the sun and our region of interest (ROI) is in a different area). This feature makes the method flexible.

The other feature is the possibility of setting the LO and HI values. Setting values very close to the extremes (0 and 255 DC for 8 bits), will result in less number of shots, at the cost of lower SNR. In contrast if we set values further from these extremes, we will sample the scene's radiance with more overlap between contiguous shots and therefore the SNR will increase, at the cost of higher number of shots. This shows how our algorithm is tunable, allowing us to adapt to different requirements regarding SNR of the captured HDR.

After commenting these functionalities, we describe now in detail how the exposure time search is done. With the information present in the cumulative histogram of the first shot captured, we check the percentile of pixel population which is below the LO level. If it is higher than the maximum value set, then a longer exposure time is calculated. The same is done if the difference between 100 and the percentile of pixel population above the HI level is higher than the tolerance threshold. In this case a shorter exposure time will be calculated.

To find a longer exposure time, we will use equation 4.4, to shift the camera response value from the LO level to HI level. Therefore we use the HI level as DC_1 , and E_0 is substituted by equation 4.3, where we use the LO level as DC_0. Δt_0 is the exposure time used to acquire the current image (see eq. 4.5).

$$\Delta t_{longer} = \frac{\mathrm{CRF}^{-1}(\mathrm{HI})}{\mathrm{CRF}^{-1}(\mathrm{LO})} \cdot \Delta t_0 \tag{4.5}$$

On the other hand, to find a shorter exposure time, we will use equation 4.4, to shift the camera response value from the HI level to LO level. Therefore we use the LO level as

DC_1, and E_0 is substituted by equation 4.3, where we use the *HI* level as DC_0 (see eq. 4.6).

$$\Delta t_{shorter} = \frac{\mathrm{CRF}^{-1}(\mathrm{LO})}{\mathrm{CRF}^{-1}(\mathrm{HI})} \cdot \Delta t_0 \tag{4.6}$$

In this way, if the population that has a sensor response at *HI* level in one shot, shifts to *LO* level in the next shot, we can cover the full dynamic range of the scene with certain overlap between contiguous shots.

The process described here goes on checking the cumulative histograms of the longer and shorter exposure times until the tolerance requested is met or the system reaches its maximum or minimum available exposure time values.

4.3 Experiments and results

We will first explain in section 4.3.1 the only method found following the same philosophy as ours: the *MIBS* method proposed by Barakat et al. (*BAR*) [7]. Later on, in section 4.3.2, we describe the two experiments performed:

- Experiment 1: we tested the default version of our Adaptive Exposure Estimation (*AEE*) method (see section 4.3.2).
- Experiment 2: we explore the tunability and evaluate the *SNR* performance (see section 4.3.2).

4.3.1 BAR method and MSBS

Regarding *BAR* method, the *MSBS* found for our *Canon* camera, as explained in [7], using 5.6 aperture setting, was composed by 4 exposure times which values were: 30 s , 300 ms, 1 ms and 0.0125 ms. By definition, a *MSBS* contains the longest and shortest exposure times that the imaging system allows us to use.

In figure 4-8, we can see an example of overlaid cumulative histograms for a scene captured using the *MSBS*. As we can see, contiguous shots have certain overlap. Choosing exposure times further apart would end up in leaving some gaps in the exposure axis not covered by any shot, and therefore some areas of the scene would not be correctly captured.

Depending on the scene captured, it can happen that we do not need all four shots to record the full dynamic range. In these cases, a sub-set of the *MSBS* is used. This subset

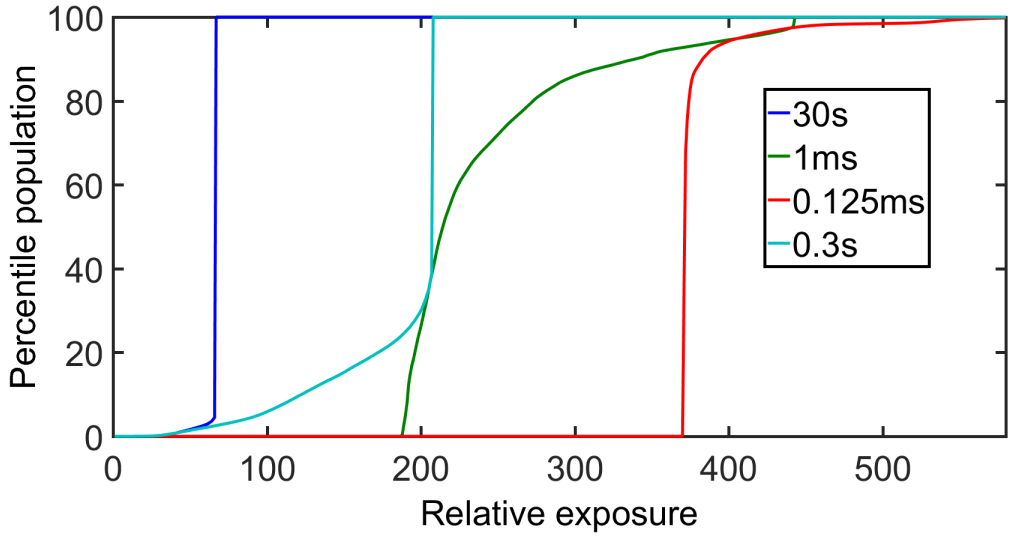

Figure 4-8: Minimal System Bracketing Set found for *Canon EOS 7D* camera. Composed of 4 shots with increasing exposure times: 0.0125 ms, 1 ms, 300 ms, and 3000 ms.

is called Minimal Image Bracketing Set (*MIBS*). We could use as first shot's exposure time the automatic exposure selected by the camera, as done for the *AEE* method. However we found that doing so, we only got the same number of shots or even one more (see figure 4-9).

The exposure time chosen by the auto-exposure mode of the camera using the whole image for metering was 66.7 ms (top-center). It was used as first shot for *AEE* method, since our method adapts to any exposure value chosen as first. However this value was in between 300 ms and 1 ms (both belong to *MSBS*). Therefore if we used it also as first shot for *BAR* method, it would mean that the capture of this scene would end up with 5 shots instead of 4.

Also if we change the aperture setting of the camera, as the *CRF* is not changing, *AEE* method would work the same, adapting to the new exposure levels in the sensor. However for *BAR* method we would need to calculate a new *MSBS*, since the same exposure times for a different aperture could not be valid any more. Thus we fixed our aperture setting to 5.6 for both methods.

4.3.2 Experiments performed

For the first experiment (section 4.3.2), we captured 30 scenes (25 outdoors and 5 indoors) using both *AEE* and *BAR* methods and we studied the number of shots taken, the total exposure time used, and the percentage of pixel population which was not properly exposed.

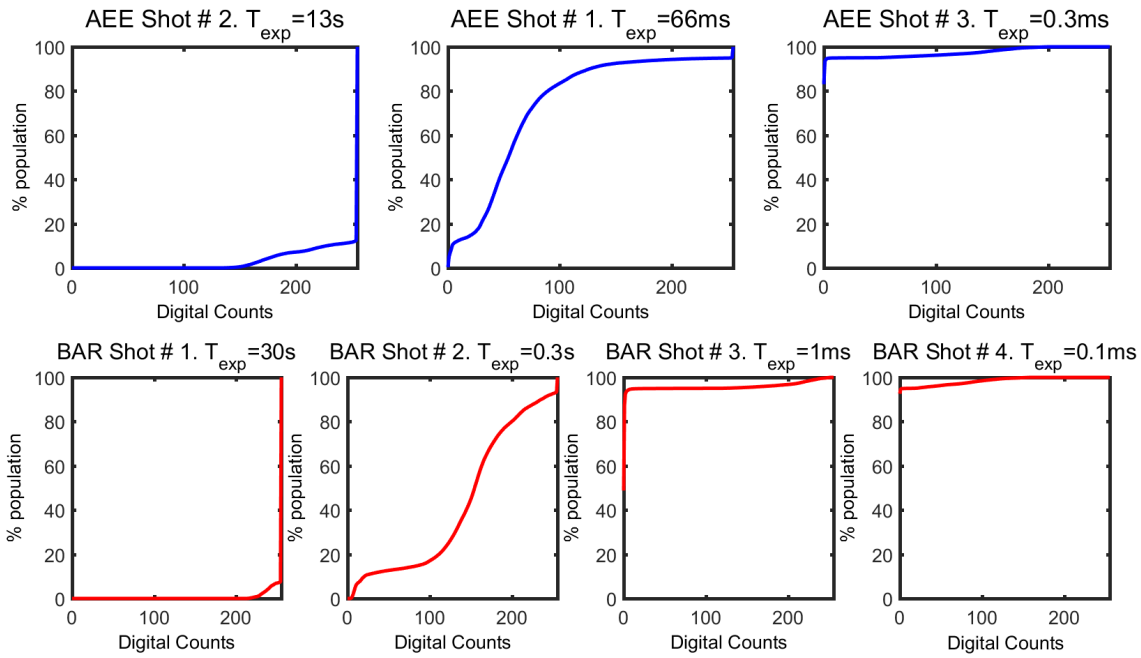

Figure 4-9: Cumulative histograms of the same scene using *AEE* method (top row) and *BAR* method (bottom row). The histograms are ordered by decreasing exposure time to observe their continuity.

This way we assessed how efficiently did both methods recover the full dynamic range of the scene, comparing their resulting bracketing sets.

In the second experiment (section 4.3.2), we built an indoors *HDR* scene with controlled illumination conditions. We captured 10 *HDR* images of it using *AEE* and *BAR* methods. Besides, for the *AEE* method we repeated the capture 4 times using different values for the *LO* and *HI* levels (see section 4.2). Finally we captured 10 ground-truth (*GT*) *HDR* images using all available exposure times in the camera. These *GT* represent the highest *SNR* that our camera can achieve to record the scene.

HDR capturing efficiency

As mentioned before, we acquired 30 *HDR* scenes. 23 scenes were captured outdoors with natural illumination and 7 indoors with artificial illumination. Outdoors, daylight casts *HDR* illumination over objects, including clouds. Indoors we used a light booth and a fluorescent lamp oriented directly to the camera in a dark room to generate *HDR* content.

The different *LDR* images captured were stitched together using the weighted average

equation 4.7 ([155]).

$$E'_{\lambda}(x,y) = \frac{\sum_{n=1}^{N} \omega(\rho_{n\lambda}(x,y)) \cdot \frac{CRF^{-1}(\rho_{n\lambda}(x,y))}{\Delta t_n}}{\sum_{n=1}^{N} \omega(\rho_{n\lambda}(x,y))} \tag{4.7}$$

Where E' is the value of *HDR* radiance map generated, λ is the spectral band index, (x, y) are pixel coordinates, N is the number of shots captured for each spectral band, ρ is the sensor response in each *LDR* image captured, CRF^{-1} the inverse camera response function, Δt_n is the exposure time used for shot number n, and ω the weighting function used to build the *HDR* image (see figure 4-10).

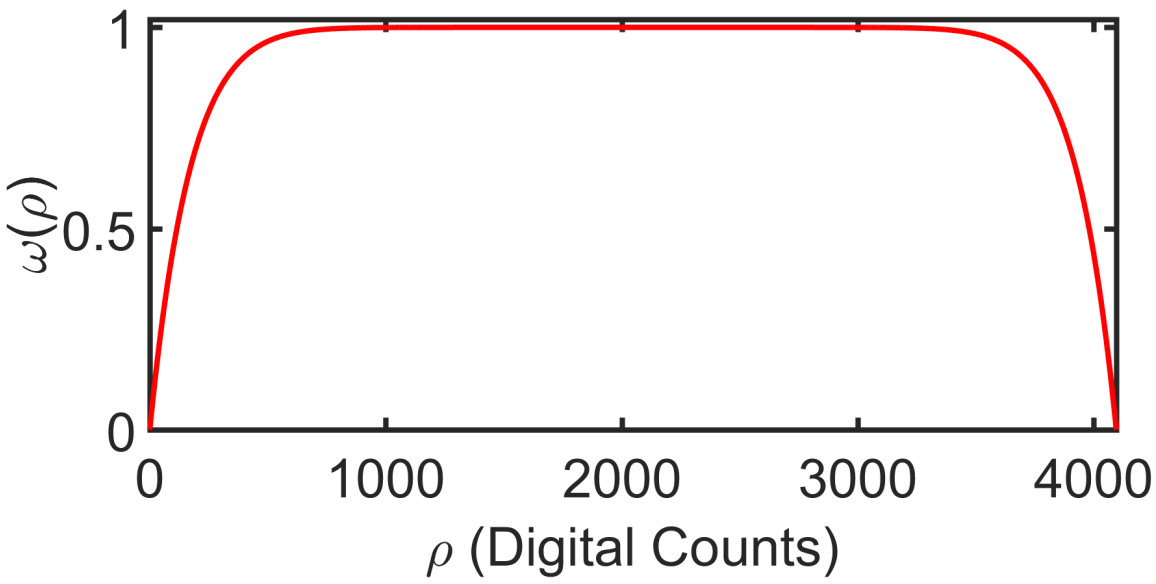

Figure 4-10: Weighting function used to build *HDR* images from multiple 12-bits *LDR* images.

In the literature we find some other shapes for this weighting function like the triangular or hat shape ([26]), derivatives of the *CRF* ([119, 158, 134]), or more complex shaped functions ([49]). We found the smooth (also called broad hat function) function proposed in [155, 4] and shown in figure 4-10 was less prone to introduce artifacts in the final *HDR* image than the other alternatives.

To check the performance of both methods in terms of full range recovery, we plotted the cumulative histograms of all shots taken for each scene, checking that no irradiance gaps were left uncovered between consecutive shots. We set the maximum percentile allowed to be lost to 0%. In table 4.2 we can see the results for 7 out of the 30 scenes captured.

Table 4.2: *MIBS* found for *BAR* and *AEE* methods in 7 of the 30 scenes captured.

Scene #	Method	# shots	% lost	$\sum_{i=1}^{n} T_i(s)$
1	BAR	3	1.61	0.301
	AEE	2	1.61	0.025
2	BAR	3	0	0.301
	AEE	2	0	0.101
3	BAR	3	0	0.301
	AEE	2	0	0.040
4	BAR	3	0	0.301
	AEE	2	0	0.050
5	BAR	3	0	0.301
	AEE	2	0	0.034
6	BAR	3	0	0.301
	AEE	2	0	0.025
7	BAR	4	0	30.301
	AEE	3	0	13.067

We observe how the number of shots is always equal or lower for *AEE* method. The percentage of useless pixels is always the same for both methods. Many scenes had a percentage of lost pixels equal to 0, since both methods managed to retrieve the full dynamic range of the scene. For the rest of the scenes, the useless pixels were due to direct sunlight (like in scene 1). The total exposure time is always lower for the *AEE* method. The same commented trends are found for the remaining 23 scenes captured. In total, for the 30 scenes captured, *BAR* method took a total of 96 shots, using 218.734 seconds, and *AEE* method 81 shots and 139.869 seconds. This means that the number of shots was 15.63% less, an the total exposure time 36.06% less for *AEE* method.

In figure 4-11 we show the *LDR* pictures and the tone-mapped *HDR* radiance maps generated for some of the scenes. The tone-mapping algorithm used was contrast-limited adaptive histogram equalization, introduced by Ward [155] and implemented in *Matlab R2014a*.

In figure 4-9 we plot an example of the cumulative histograms corresponding to both methods of scene number 7. Although at a first glance the central histograms for both methods may seem to reach percentiles 0 and 100 respectively, the percentage of underexposed pixels in the central shot of *AEE* was 8.53%, and the percentage of saturated pixels in this same shot was 5.08%. In the case of *BAR* method, the shot taken with $T_{exp} = 0.3s$ had a percentage of underexposed pixel of 0.065%, while the percentage of saturated pixels in the

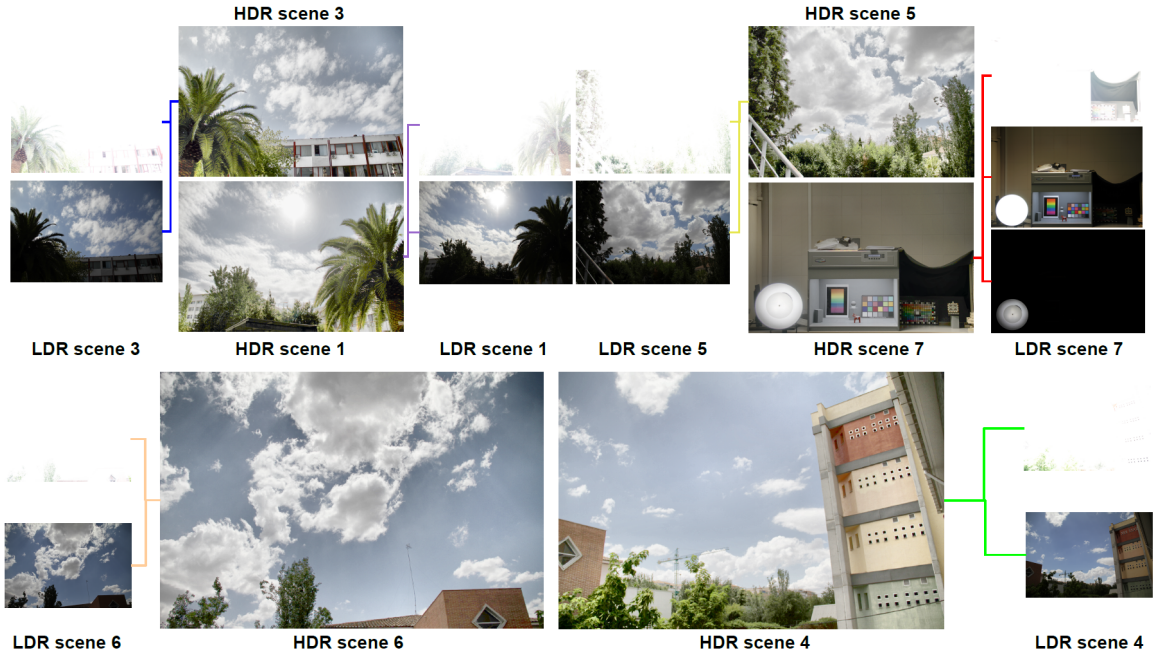

Figure 4-11: *LDR* images and *HDR* radiance maps for some of the indoors and outdoors captured scenes using *AEE* method. Figure taken from [122]

image taken with $T_{exp} = 1ms$ was 0.04%.

We observe how both methods succeeded in recovering the full dynamic range of the scene. However, the *AEE* method managed to do it in 3 shots, while the *BAR* method needed 4 shots. Therefore we can conclude with this experiment that *AEE* method recovers the dynamic range of the scene as satisfactorily as *BAR* method does, but is more effective in terms of time and number of shots.

Signal to Noise Ratio

We did a second experiment to study the *SNR* behavior of our method. As explained in [6], there is a trade off between *SNR* and total amount of shots in *HDR* imaging. For the *AEE* method, we tested 4 different conditions:

A $LO = 3$ and $HI = 253$.

B $LO = 16$ and $HI = 240$.

C $LO = 56$ and $HI = 200$.

D $LO = 106$ and $HI = 150$.

We compared all these *AEE* conditions with *BAR* method and the ground-truth (*GT*).

We set up an indoor scene and measured its dynamic range using a spectroradiometer (*Photo Research: PR745*). We measured the integrated radiances of both the brightest and the darkest points of the scene. The resulting dynamic range measured was 4.1 logarithmic units. We captured the scene 10 times using each method.

For each pixel and each color channel of these *HDR* radiance maps, we calculated its average *HDR* value and its standard deviation across the ten images corresponding to each method. The average *HDR* value provides information about the signal level in the pixel, and the standard deviation provides information about the level of noise generated by all noise processes present in the *HDR* capture process. Thus, computing the ratio of average *HDR* value (E_{xy}) over the standard deviation (σ_{xy}) as equation 4.8 shows, we have a *SNR* estimate [11].

$$SNR_{xy} = 20 \times \log_{10}(\frac{\bar{\mathrm{E}}_{xy}}{\sigma_{xy}}) \tag{4.8}$$

The subindex xy stands for pixel position within the *HDR* radiance map.

We show the number of shots, the total exposure times and the average *SNR* for each method in table 4.3.

Table 4.3: *SNR* performance for 4 *AEE*, *GT*, and *BAR* methods.

Method	# shots	Δ t (s)	$\overline{SNR}$ (dB)
AEE A	3	14.79	27.19
AEE B	4	15.02	30.47
AEE C	5	31.32	32.58
AEE D	16	61.22	33.57
GT	55	151.43	35.32
BAR	4	30.30	29.97

As expected, setting the *LO* and *HI* values further from the extremes of the range in *AEE* method, yields a higher number of shots and also higher total exposure time, but the *SNR* increases as well. For condition *D*, we reached an average *SNR* only less than 2 dB below the ideal case (*GT*), yet using only about 40% of the total exposure time. The minimum bracketing set found was *AEE A*, with only 3 shots and 14.79 s of total exposure time, but also had the lowest *SNR*. *BAR* method needed the full *MSBS* to recover this

scene, using 4 shots. It got better SNR than our minimum bracketing set, but we also got a second option ($AEE\ B$), using 4 shots, with shorter total exposure time (less than half), and higher SNR than BAR method.

We can also observe in figure 4-12 the SNR for each pixel of the radiance maps generated, plotted as a function of the signal level. We can observe how $AEE\ D$ has the most similar distribution compared with GT. Also BAR method has a very similar distribution compared with cases $AEE\ A$ and $AEE\ B$, as expected.

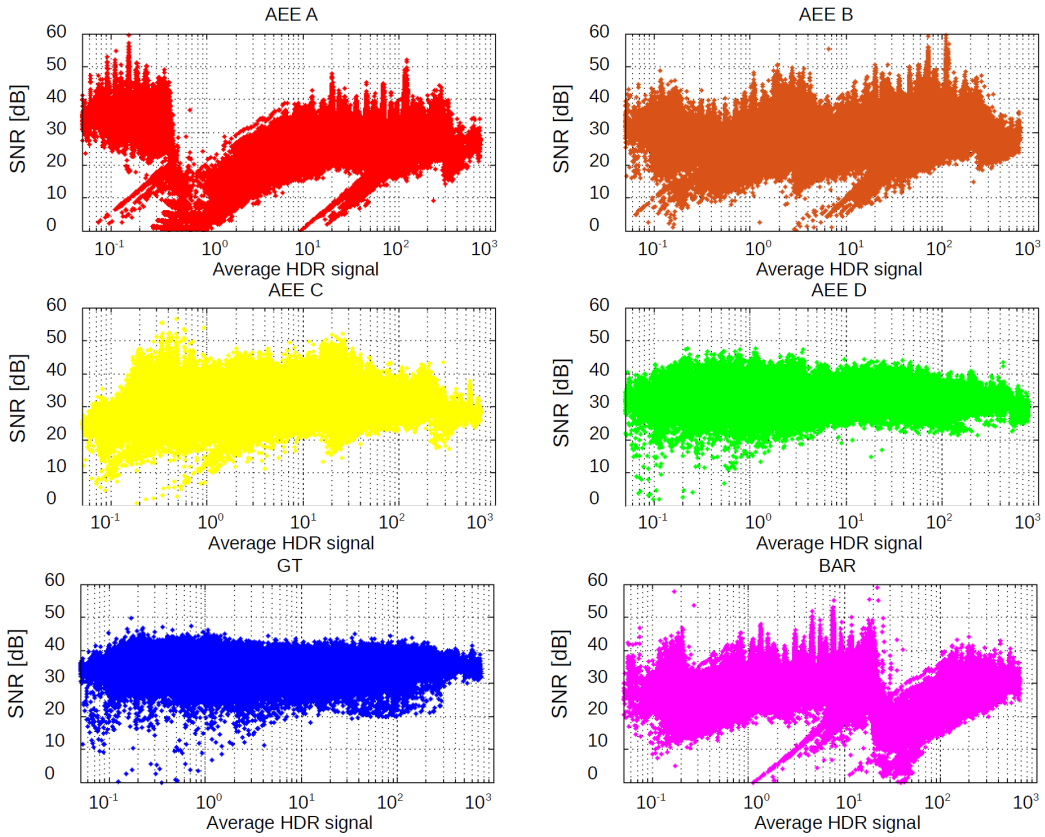

Figure 4-12: SNR vs Average HDR signal present in the radiance maps.

In figure 4-13 we plot the SNR histograms for all conditions and methods, including the ground truth.

We observe how for AEE method, the main lobe gets narrower and shifts towards higher mean SNR (rightwards) as we tune the LO and HI levels further from the extremes of the range (thus closer to each other). $AEE\ D$ is quite close to GT in position and shape. But the number of shots captured is higher, and therefore the total exposure time used for the HDR capture increases.

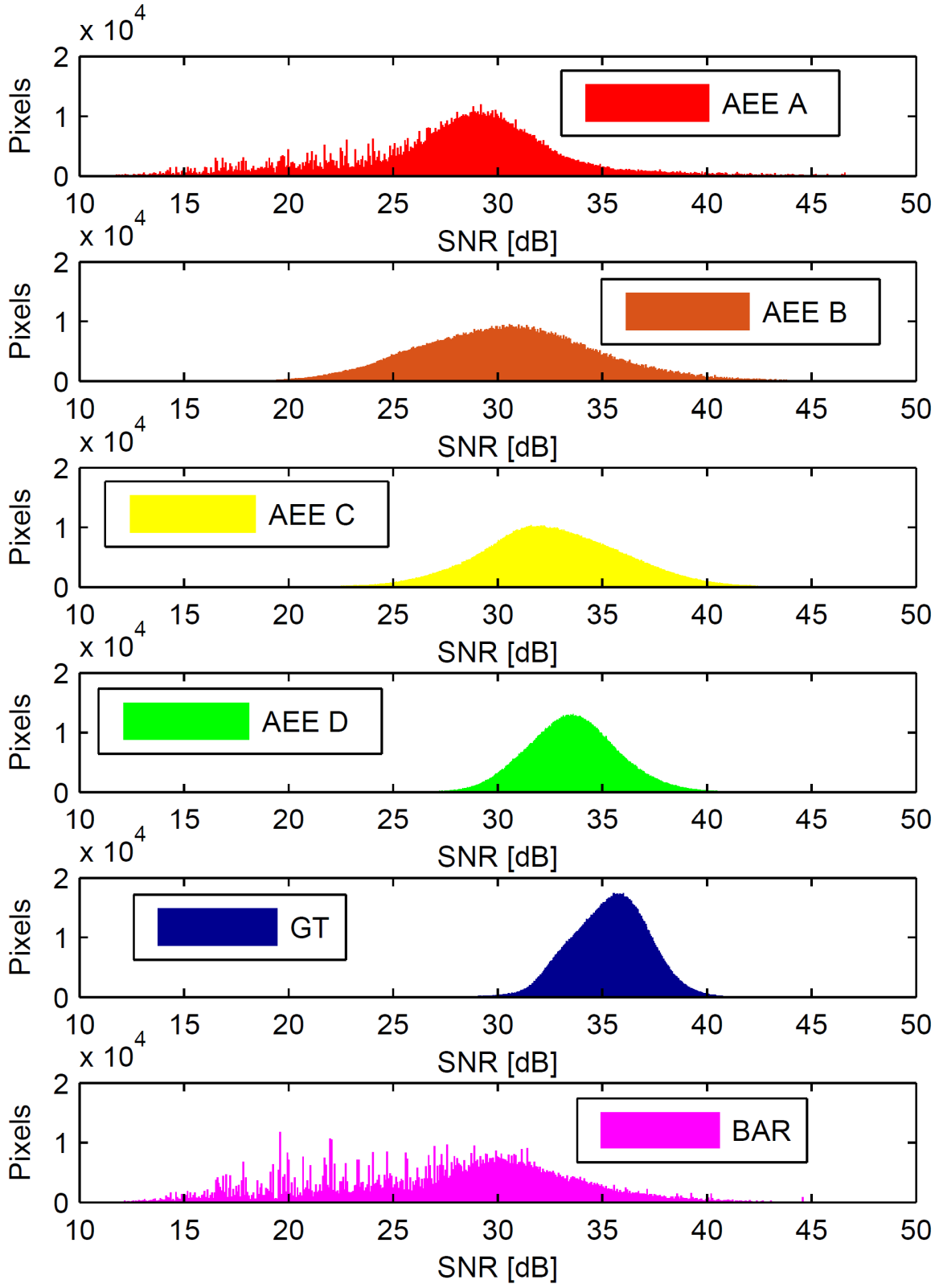

Figure 4-13: Histograms of *SNR* values for all conditions tested.

On the other hand, *BAR* method yields a histogram which is the most spread over a wide range of *SNR* values, as well as the second one with the lowest mean *SNR* value, yet using a relatively high total exposure time.

4.4 Conclusions

We have presented a new method for estimating the exposure times needed to recover the *HDR* radiance map from a scene via multiple exposures. We compared the performance of our method with that of the only method found in the literature with similar design and purpose (finding minimum image bracketing sets).

Our proposed method is adaptive because it finds a bracketing set adapted to any *HDR* scene content, and universal because it works for any camera. We only need to calculate the *CRF* (which is anyway needed to build the radiance map).

Besides, the method is tunable, since we can decide if we prefer to find a minimum bracketing set, at the cost of lower *SNR*, or increase the *SNR*, by sampling the radiance of the scene with more overlapping between consecutive shots (increasing the number of shots and capturing time as well). For the minimum bracketing set case, the bracketing sets found were minimal in the 30 scenes tested.

Moreover our method is blind, which means that no information about the content of the scene needs to be known a-priori. The multiple *LDR* images are captured on-line as the process is ongoing, and every single shot taken is used to compose the *HDR* radiance map.

We can also control what is the percentage of total pixel population we can accept as useless (underexposed or saturated). This way we can find the minimum bracketing set only for our region of interest.

We have applied the method for natural scenes where partially cloudy skies were present in order to increase the dynamic range of the capture. We have successfully covered the full dynamic range of the 30 scenes imaged. We have shown how our method can find bracketing sets which are shorter than those found by Barakat method (*BAR*), yet keeping higher *SNR* levels in the *HDR* radiance map reconstructed from the multiple exposures.

We studied the *SNR* performance of our method comparing it not only with *BAR* method, but also with an ideal-case ground-truth *HDR* image built using all available exposure times in the camera. We have demonstrated how we can tune our method to suit

different requirements of *SNR* at the cost of increasing the number of shots.

The proposed method offers a solution for the blind acquisition of *HDR* images using multiple exposures, which can be used in any *HDR* imaging context: machine vision, outdoors urban or natural scenes, sky imaging, *HDR* photography, etc. And in particular, for studying optical phenomena present in open air scenes, where the illumination conditions are extreme.

4.5 Future work

As future lines of research for the *HDR* method proposed, we would like to include *ISO* and aperture settings to the capturing. We have to take care of both noise and depth of focus in order to be able to play with these two parameters as well. If we manage to use them and include them in a new version of the *AEE*, maybe we could accelerate even more the *HDR* automatic capture where the imaging conditions are rather dim. Accelerating enough the adaptive capture of *HDR* images, could lead us to propose a new method for the capture of *HDR* video.

We would also like to consider the implementation of some ghost removal techniques ([100, 93]) for those cases where slight movements are present in the captured *LDR* images.

Also we plan to consider variational techniques [36] to try to reduce the impact of veiling glare during the post-processing of the images, rather than doing so during the acquisition process ([173]).

When you open your mind to the impossible, sometimes you find the truth.

Walter Bishop.

5

Multispectral HDR Polarimetric imaging for object segmentation and classification

Content:

Chapter 5

Multispectral HDR Polarimetric imaging for object segmentation and classification

After presenting the advances proposed for spectral imaging in chapter 3, and for *HDR* imaging in chapter 4, we will now explain how we combined both technologies. In chapter 2, we described many different system architectures for capturing multispectral images. One of the aims of this thesis was object classification. Polarimetric imaging shows potential, since different object materials interact differently with light from the polarization point of view. Liquid Crystal Tunable Filters (*LCTFs*, [136, 65, 181]) receive incoming unpolarized light, and produce as output linearly polarized light. Hence we propose a method for the capture of *HDR*, Multispectral (*MS*), Polarimetric (*Pol*) images of indoor scenes using a *LCTF*.

In a 100 days stay in University of Chiba (Japan), we collaborated with the members of the *Horiuchi & Hirai's laboratory* in the Graduate School of Advanced Integration Science. We based our work there on improving the capture of *MSHDRPol* images for object segmentation and classification ([180, 179]).

We have included our Adaptive Exposure Estimation method (*AEE*, explained in chapter 4) to fully automatize the capturing process. We also propose a pre-processing method which can be applied for the registration of *HDR* images after they are already built as the result of combining different *LDR* images. This method is applied to ensure a correct alignment of the different polarization *HDR* images for each spectral band. We have focused our efforts

in two main applications: object segmentation and classification into metal and dielectric classes.

We have simplified the segmentation using mean shift combined with cluster averaging and region merging techniques. We compared the performance of our segmentation algorithm with that of *Ncut* ([165]) and Watershed (*WS*, [8]) methods. For the classification task, we propose to use information not only in the highlight regions but also in their surrounding area, extracted from the degree of linear polarization (*DoLP*) maps. We present experimental results which proof that the proposed image processing pipeline outperforms previous techniques developed specifically for *MSHDRPol* image cubes.

As explained in chapter 2.3, in [180], the authors found that inspecting the curvature of *DoLP* map around specular highlights, they could classify the objects into metal or dielectric, disregarding the observation geometry. This allowed them to perform classification including curved object surfaces as well ([179]), but sometimes the *DoLP* calculation around the highlights is unstable. Specially if the image captured is somewhat noisy or the highlight region is rather large.

We proposed a new method which is also based on the study of the spatial distribution of the *DoLP* map around the highlights, and which overcomes these limitations. We also found and showed that when we rotate the *LCTF* for changing the polarization angle, we introduce some slight unwanted arbitrary translation, which was unaccounted for before. This translation is specially problematic if we want to get high resolution images and retrieve pixel-wise *MSHDRPol* data (for instance for studying the metallic sparkling elements in special effects coating car paints [131]).

As explained in section 5.3, methods for registering the differently exposed *LDR* images captured to compose a *HDR* image [192, 174, 93] do not yield satisfactory results with already built *HDR* images. We therefore proposed a pre-processing step before the registration, based in the compression of the dynamic range of the *HDR* scenes. This way, we also saved computation time, since for each spectral band we only need to register one *HDR* image, and not many *LDR* images.

We also wanted to move the capture of *MSHDRPol* images out of the laboratory. However when we work outdoors, the imaging conditions become uncontrolled. Outdoor illumination might suffer relatively rapid changes, which make unfeasible the use of an *LCTF* to capture the full spectrum of each pixel in an outdoor scene. Thus we tried to implement the

same capturing method to the system commented at the end of section 3.6. This filter-wheel camera (*Pixelteq SpectroCam VIS-NIR*), provides us a way to capture multispectral *HDR* images with the advantage of changing the spectral responsivities of the spectral channels by switching to different filters. If we also include a rotating polarizing filter in front of it, we add the polarization information to our image data. As an extra feature, this camera is sensitive in both visible and *NIR* ranges of the spectrum. Therefore our outdoors images were *MSHDRPolVISNIR* images.

The remainder of this chapter is structured as follows: in section 5.1, we explain the absolute radiometric calibration performed to the camera, as well as the spectral calibration of the imaging system. In section 5.2 we briefly explain the capturing system and also the capture work-flow. In section 5.3, we explain all the image processing applied in order to get correctly registered *HDR* multispectral polarimetric images. In sections 5.4 and 5.5, we explain our proposed work-flows for the segmentation and classification applications respectively, showing the results obtained and comparing them with those of the methods proposed in [179]. Later on, in section 5.6, we explain how we captured a new set of *MSHDRPol* images outdoors, using a very similar method with *Pixelteq SpectroCam VIS-NIR*. Finally, in section 5.6.4, we draw the final conclusions of this chapter and in section 5.6.5 the future lines of research proposed.

5.1 Radiometric and spectral calibration

When capturing *HDR* images via combination of multiple *LDR* exposures, the resulting image data is in a different scale from the original *LDR* images [119]. The former scale is exposure-time-independent, and the latter is exposure-time-dependent. Usually for the *HDR* images, we aim to get pixel values in the final image that are proportional to the amount of light coming from each region of the scene. For this purpose we perform a radiometric calibration of the camera.

5.1.1 Radiometric Calibration

Before starting with the capture of the scenes, we performed an absolute radiometric calibration of the imaging system [129, 130] by measuring the *CRF* of our camera. We also did it for the imaging system used for outdoors scenes, but this will be explained in section

5.6.2.

For the *LCTF*-based system, we first detached the *LCTF* from the camera, because the *CRF* is unaffected by the *LCTF*, and detaching it allows us to work with shorter exposure times, while the signal level is still high enough to get useful image data. We prepared a scene containing ten homogeneous gray patches of different gray levels (see figure 5-1).

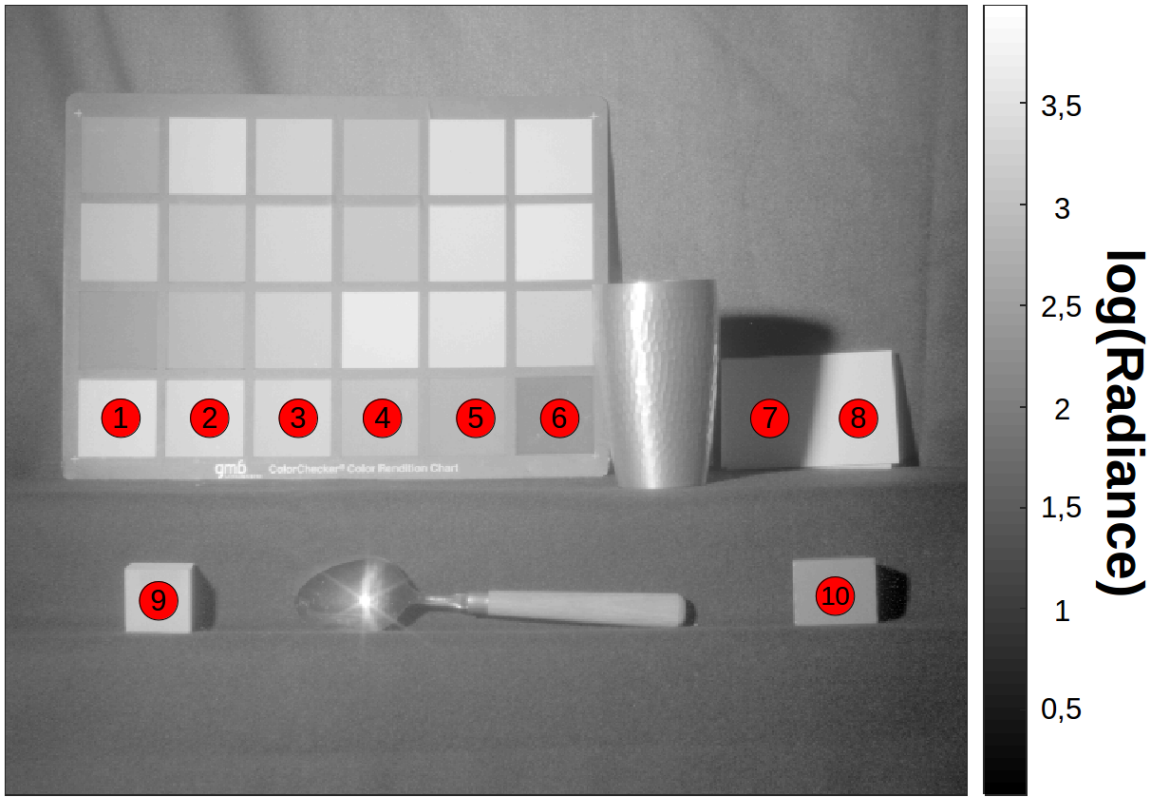

Figure 5-1: Tone-mapped version of monochrome *HDR* image of the scene used for radiometric calibration. The scale shows scene radiance (W/m^2sr) in a logarithmic scale. The spectral radiances of highlighted areas are shown in figure 5-2.

The patches were non-homogeneously illuminated (different levels of irradiance from direct light to shadow). This way we generated radiance signals covering a wide range of values in the scene (up to 4 orders of magnitude). We captured several images of this static scene using different exposure times and aperture values. After that, we substituted the camera by a spectroradiometer model *Photo Research PR-645* (this model is different from the spectroradiometer used in the Color Imaging Laboratory at University of Granada, explained in section 4.3.2), and measured the radiance coming from the ten patches in the scene. Since the scene was static, the radiance outgoing from each patch was constant in time. Therefore we did ten radiance measurements for each patch and averaged them in

order to reduce noise. The spectral radiances measured for the ten patches can be seen in figure 5-2.

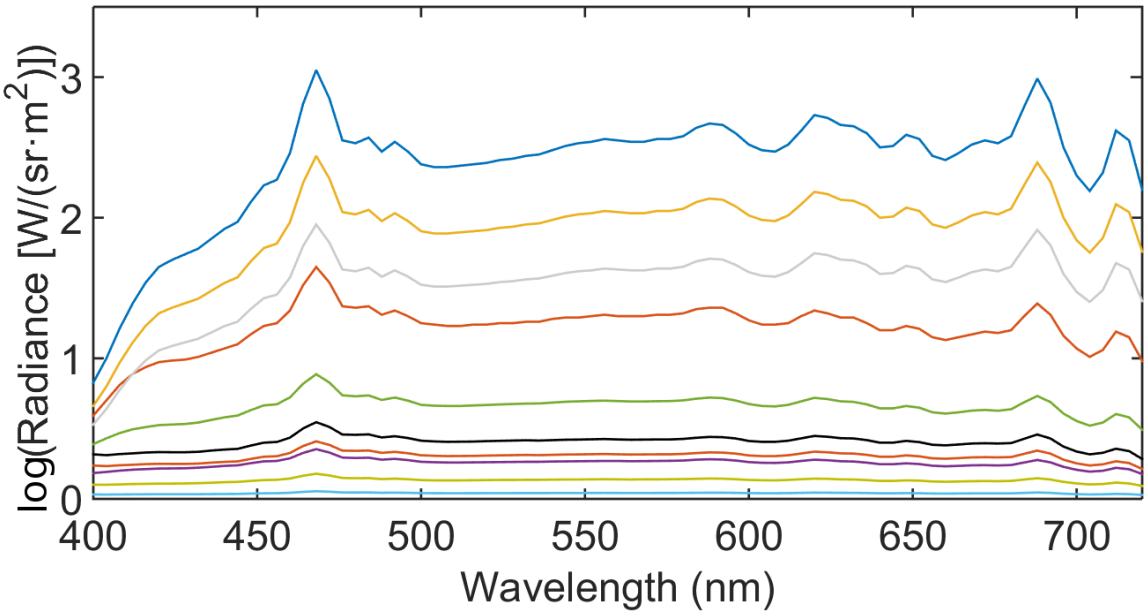

Figure 5-2: Spectral radiances from the ten gray patches in the scene used for radiometric calibration (shown in figure 5-1). The vertical axis is in logarithmic scale.

Multiplying all the integrated radiance values in the spectral range from $400nm$ to $700nm$, by all exposure times used to capture the images for each aperture, we got a matrix of pseudo-exposure values. We call it pseudo-exposure because the exposure is the product of exposure time times irradiance ($T_{exp} \cdot E_e$), but our spectroradiometer measured radiance instead (L_e) of irradiance. Radiance and irradiance are different magnitudes, but both of them give us information about the amount of light in each region of the scene.

According to the reciprocity law [34], halving the exposure time and doubling the irradiance should result in the same sensor response. Since we can also average a small area in each patch to retrieve sensor responses for each exposure time and aperture value, then we could estimate the *CRF* for each aperture setting. This function relates exposure/pseudo-exposure with sensor responses, and afterwards it is used to estimate a radiance map of the scene captured via multiple exposures technique [26].

Several factors influence the shape of the *CRF*. As *Ferrero et al.* pointed out in [34], reciprocity law does not hold when we study the response of the whole camera to light (rather than just the response of the sensor). This happens even working in raw mode. This is due to the fact that other phenomena also determine the way the camera responds to light.

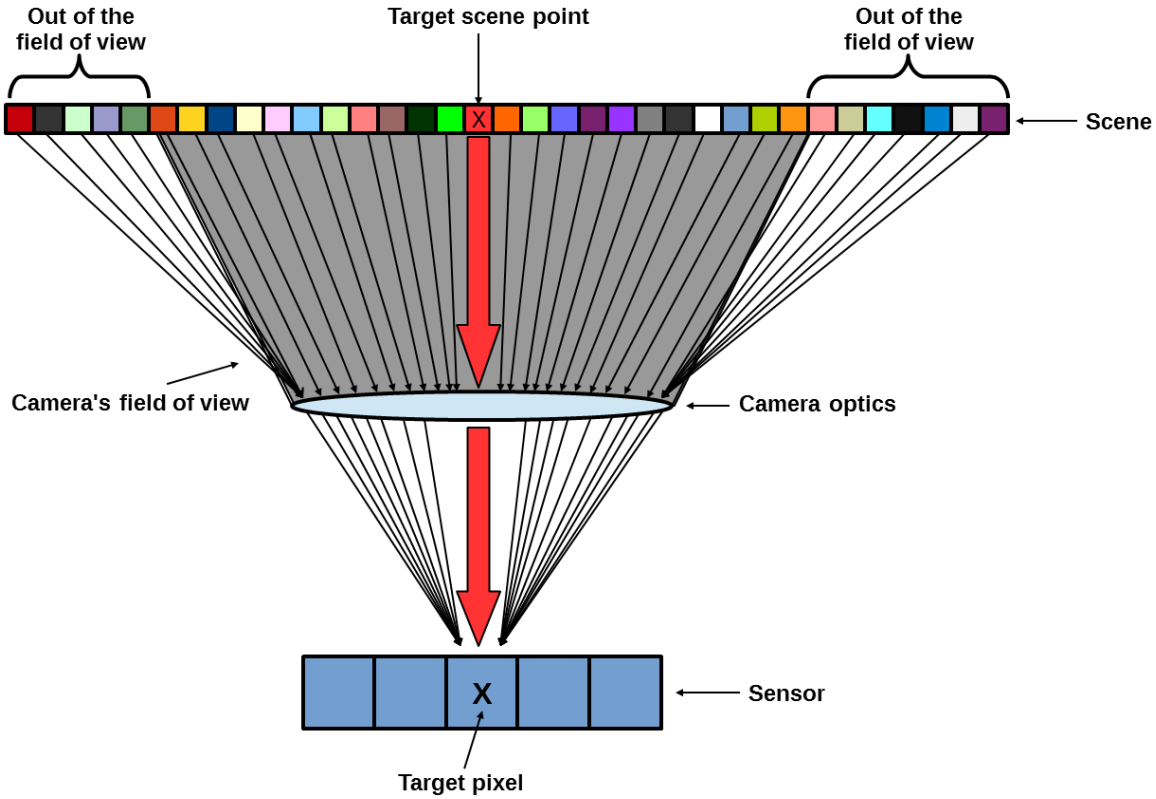

Figure 5-3: Scheme of how veiling glare affects the responses of pixels. The arrows represent contributions from each scene point to the target pixel's response, not light rays.

The strongest impact is that of optical veiling glare. Camera optics are not perfect, and as light goes through them, unwanted light from some regions of the scene, ends up impinging in pixels from other regions. Even light coming from regions of the scene out of the field of view of the camera can influence the camera responses. This effect is specially severe when there are very brilliant regions present in the scene together with very dark areas. *McCann and Rizzi* have thoroughly characterized this phenomenon [126, 127, 129, 128, 130, 125].

In figure 5-3, we a diagram to better illustrate the effect of veiling glare. The thin black arrows represent the contribution of each point of the scene to the sensor response in the target pixel. They do not represent light rays. The target pixel should only receive the contribution (thick red arrow) from the target scene point (marked with an X sign). However we can see how all points in the camera's field of view, as well as those out of the field of view, contribute to the target pixel's response. If the target scene point is dark, and other scene points are very light, their contributions are very significant compared with the contribution of the target scene point. Therefore the target pixel's response is wrongly

increased.

Consequently, the way the camera responds to light (i.e. the shape of its *CRF*), is different depending on the content of the scene. Thus, we can conclude that there is not a single *CRF* for a single camera. In other words, the *CRF* of a camera is scene-dependent. To demonstrate this, we prepared 3 scenes with different veiling glare conditions (see figure 5-4).

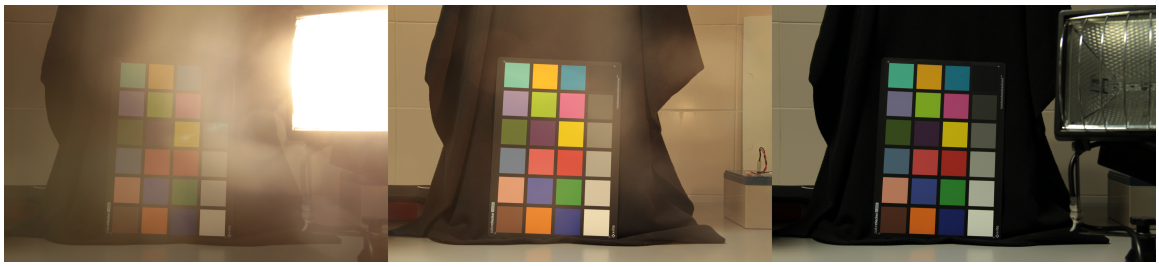

Figure 5-4: Scenes captured with different veiling glare conditions to test the influence of this phenomenon in the calculation of *CRF*. From left to right, scenes 1 to 3.

Scene number 1 had a very unfavorable veiling glare condition, with a powerful light source included in the scene, directly pointing towards the camera. Scene Number 2 was also unfavorable regarding veiling glare, but not as bad as scene number 1, since the light source was right out of the camera's field of view. Scene number 3 was rather favorable in terms of veiling glare, since we switched of the light source. We included a color checker in the scenes with a 6 patches grayscale, and used the camera responses in these patches to estimate the *CRF* in the three conditions. Figure 5-5 shows the resulting *CRFs*. Both axes are in logarithmic scale.

We can see the different shapes of the *CRFs* obtained for the different veiling glare conditions. This fact is a problem for *HDR* imaging. No matter how good the quality of our optics is, we can not completely get rid of the veiling glare effect. In this work, we have not considered correcting the effects of veiling glare (like they propose in [173] using some complex and not really practical system set up). We assumed that the imaging conditions are not so extreme so that the *CRF* curves measured for the different apertures hold for all the images captured. The *CRF* curves fitted from the measured data for the different aperture settings are shown in figure 5-6.

In figure 5-6 we can appreciate a clear linear behavior in the *CRF* curves. We used the *CRF* corresponding to the aperture number in the capture to estimate the *HDR* radiance

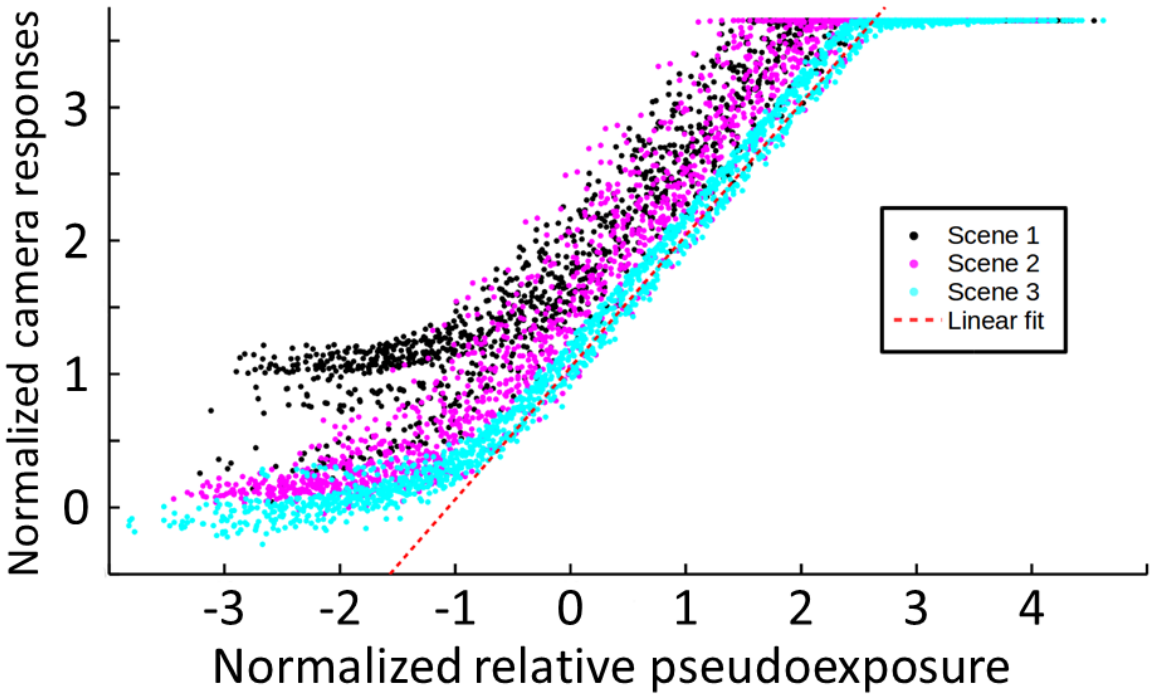

Figure 5-5: *CRFs* estimated for our camera using three different scenes with distinct veiling glare conditions, from very unfavorable (scene 1) to favorable (scene 3).

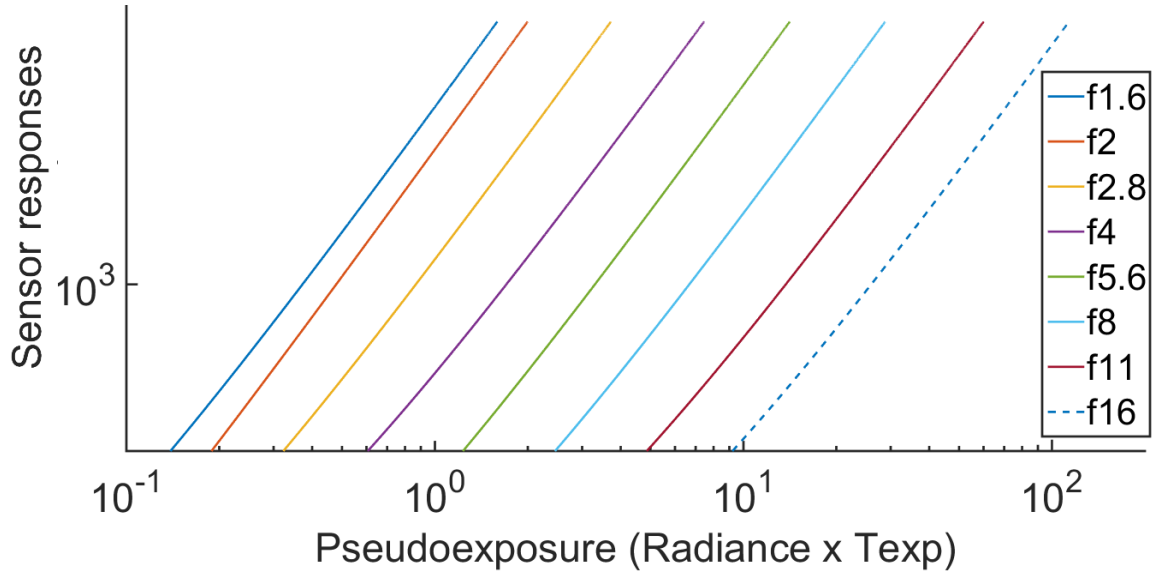

Figure 5-6: 12 bits Camera Response Functions (CRF) for different apertures settings. Both axes are displayed in logarithmic scale. Only the sensor response range above noise floor and below saturation is shown. X axis in $(W \cdot s/sr \cdot m^2)$ and Y axis in digital counts.

map. We checked this absolute radiometric calibration including some gray patches in different scenes and measuring its integrated radiance with a spectroradiometer. The results are shown in figure 5-7.

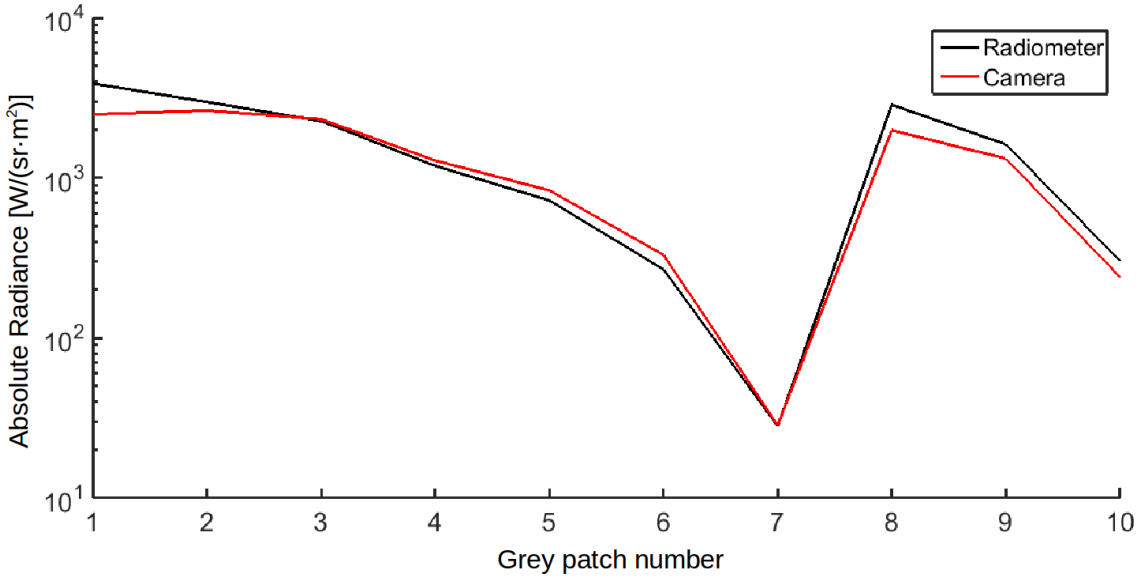

Figure 5-7: Comparison between radiances measured with the spectroradiometer and estimated from the *HDR* radiance map. The vertical axis is in logarithmic scale.

We can observe how the estimation of *HDR* radiances is reasonably close to the measurements done with the spectroradiometer. Therefore we conclude that, for scenes where the veiling glare condition is not very unfavorable, the *HDR* radiance map built is acceptably similar to the real radiances we can measure using a spectroradiometer.

The last part of the radiometric calibration consisted in identifying which pixels of the sensor were considered as hot pixels. These pixels produce a very high response even in darkness condition. This high response depends on the exposure time and temperature among other factors. We considered hot pixels those which average sensor response exceeded 5% of the dynamic range of our camera, or 205 digital counts, when we averaged 10 images with the aperture closed and 50 ms or higher exposure time. We stored the pixel coordinates of such pixels in order to make their impact negligible in the *AEE* method during the capture.

Since the response of these pixels is not relevant, we located them in every captured image and interpolated their pixel values by the mean value of their closest 8-neighbors. We found 35 hot pixels in total (less than 0.003% of the pixels contained in a captured image). However, since depending on the settings, the *AEE* algorithm might keep on looking for

new exposure times until every pixel is correctly exposed in at least one shot, it is important to account for them in order to mitigate the impact of these pixels and achieve convergence.

5.1.2 Spectral Calibration

The camera spectral responsivity $R(\lambda)$, as well as the spectral transmittances of the different tune modes of the *LCTF* (weighted by the *IR* cut-off filter spectral transmittance) $T_i(\lambda)$, and the Spectral Power Distribution (*SPD*) of the lamp $L(\lambda)$ were measured using the spectroradiometer previously mentioned, and a monochromator. The members of the *Horiuchi & Hirai's laboratory* performed these measurements and provided us with all the data. All these spectral measurements are used during the capturing process explained in section 5.2.

5.2 Image capturing

Our imaging device includes a 12-bits *CCD monochrome camera* (*QImaging Retiga SRV* 1200) with hot mirror attached, and a *LCTF* model *Cambridge Research Instruments VIS* $10-20$ (see figure 2-13).

We illuminated the scenes using a 500 watts incandescent lamp. Thanks to the polarizing properties of the *LCTF* technology [195], we can capture images at different linear polarization angles just by rotating the *LCTF* along its optical axis. For a more detailed explanation about the capturing system see [179]. A work-flow diagram for the capture process is shown in figure 5-8.

As explained before, we selected the exposure times needed to capture each spectral band using the *AEE* method described in chapter 4. This method had been tested so far for color *RGB* camera (*Canon EOS 7D*) and also for a monochrome camera (*Qimaging Retiga SRV* 1200). Both cameras were capturing all their spectral channels (3 and 1 respectively) in a single shot. Thus, we needed to adapt the method for the first time to a multispectral imaging system, which takes one shot to capture each spectral channel. We therefore run the *AEE* method for each individual spectral band.

As explained in chapter 4, the AEE method only needs as input one initial exposure time value. This value only has to accomplish one condition: at least some pixels of the resulting image must be properly exposed (*i.e.* neither below noise level, nor above saturation level).

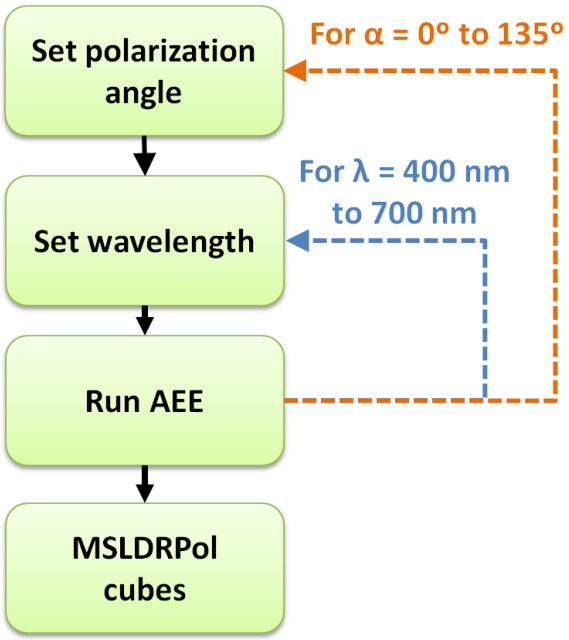

Figure 5-8: Work-flow diagram of the capture process.

For this first shot of each band, due to the wide differences for different spectral bands in the spectral responsivity of the sensor, the spectral transmittances of the $LCTF$ and the SPD of the light, the exposure time value that accomplishes the mentioned condition for a spectral band (e.g. 550 nm), does not accomplish it for a different spectral band (e.g. 400 nm or 700 nm). This fact makes it necessary to determine the initial exposure time value to run the AEE algorithm in each spectral band.

Our idea was to select a valid initial exposure time in the spectral band we usually use to point and focus the imaging system (550 nm since is the central band in the range from 400 nm to 700 nm). Once we chose this exposure time value, the AEE could run for this band. Afterwards, for each of the remaining spectral bands, we calculated a new initial exposure time value by using a weighting function. This function was derived from the product of the relative weights of spectral responsivity, SPD of light and the integral of the spectral transmittance of each mode of the $LCTF$ multiplied by the IR cut-off filter used. The resulting relative weighting function is plotted in figure 5-9.

In order to capture polarimetric information, we rotated the $LCTF$ in front of the camera. As we assumed unknown the directions of maximum and minimum transmission through the polarizer, we make the capture in four relative rotation angles of the $LCTF$, which results in the 4 relative polarization images:

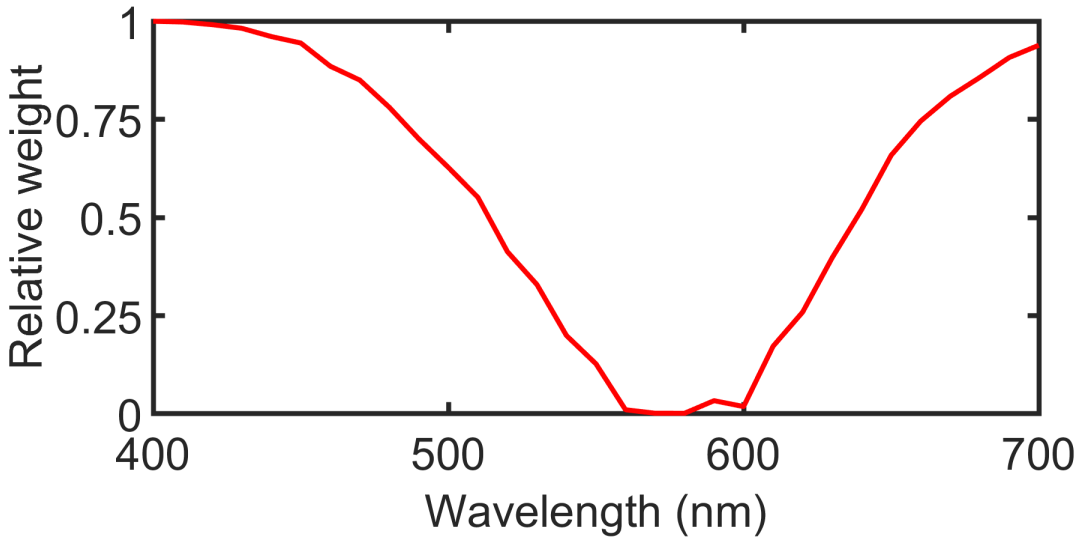

Figure 5-9: Weighting function used to scale the initial exposure time values to run the *AEE* algorithm in each spectral band. Values around [560 nm, 600 nm] are non-zero values.

- 0° $I(\lambda, \theta)$
- 45° $I(\lambda, \theta) + 45°$
- 90° $I(\lambda, \theta) + 90°$
- 135° $I(\lambda, \theta) + 135°$

Each of these images is a *HDR* image built from several *LDR* images captured with different exposure times. In section 5.5 we explain how to calculate the *DoLP* map out of these 4 polarization images.

5.3 Image pre-processing

There are three pre-processing steps in our proposed pipeline, which are applied to the *MSLDRPol* cubes resulting from the capture process: dark image subtraction, *HDR* image building and polarimetric registration (see figure 5-10). We explain them in the following subsection.

Once the capture is finished, for each spectral band and polarization angle, we get a set of images captured with different exposure times. The number of images for each spectral band and polarization angle would depend on the dynamic range of the scene we

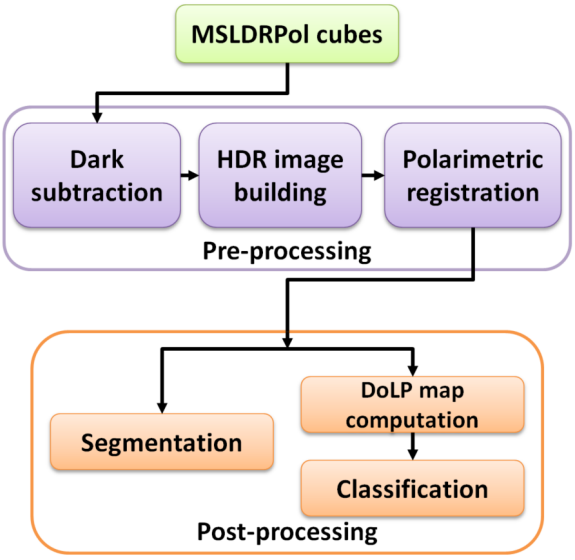

Figure 5-10: Work-flow diagram of the whole image processing pipeline from the captured images to the final classification step.

are capturing, as well as the parameters we set for the *AEE* method, and the capturing device responsivity. Since our *AEE* parameterization was configured to get *MIBS* (i.e. the set of different exposure times needed to capture correctly every point of the scene with the lowest number of shots), we obtained from two to three images per band and polarization angle.

5.3.1 Dark subtraction

As it usually happens in most digital imaging systems, there are some sources of noise which make the sensor yield a non-zero response even in the absence of light a signal [182, 24]. The impact of this dark noise component can be reduced by subtracting black images from the different images captured from the scene. These black images are captured using the same exposure times as the scene images, but the aperture of the camera is completely closed. For each exposure time value used during the capture of all *LDR* images, ten black images were captured and averaged (in order to reduce noise). Afterwards, and before the next step of building the *HDR* images, these averaged black images were subtracted from every *LDR* image captured.

5.3.2 HDR image building

Once the values of the hot pixels are corrected (as explained in section 5.1.1) and the dark noise is subtracted from every *LDR* image, we are ready to build the *HDR* image of each spectral band and polarization angle. We did so using equation 5.1, previously used in [26, 155].

$$E'_{\lambda\theta}(x,y) = \frac{\sum_{n=1}^{N} \omega(\rho_{n\lambda\theta}(x,y)) \cdot \frac{CRF^{-1}(\rho_{n\lambda\theta}(x,y))}{\Delta t_n}}{\sum_{n=1}^{N} \omega(\rho_{n\lambda\theta}(x,y))} \tag{5.1}$$

This equation is similar to equation 4.7 shown in chapter 4. However now we have added the subindex θ, which is the polarization angle index. The weighting function (ω) used is the one shown in figure 4-10.

5.3.3 Polarimetric registration

Once the *HDR* images were built for each spectral band, and each polarization angle, we checked for misalignments between different bands and polarization angles within each band. We found no noticeable misalignment between spectral bands. This is in all likelihood happening because the depth of field was larger than the longitudinal chromatic aberration in our capturing conditions. Thus, there was no need for registration among different bands captured with the same polarization angle.

However we did find misalignment between the images corresponding to different polarization angles for a given spectral band. Apparently, the manual rotation of the *LCTF* introduced small translations of the images. Thus, we had to perform image registration. Since all spectral bands were correctly aligned for each polarization angle individually, as long as we could find the transformation needed to align the four polarization images for one spectral band, we could apply the same transformation to every spectral band. These misalignments could cause future inaccurate results in the metal-dielectric classification based on polarimetric images. The difficulty in this step lies in the fact that the images we are going to align are already *HDR* images.

In figure 5-11, we show some of these misalignments. We captured some test scenes including objects covered with a chessboard pattern. Each image is an overlay of two polarization angle images. If the images were perfectly registered (aligned), we would not

see any color shades, only gray shades. However, the green and purple color shades indicate that the two overlaid images are not perfectly aligned.

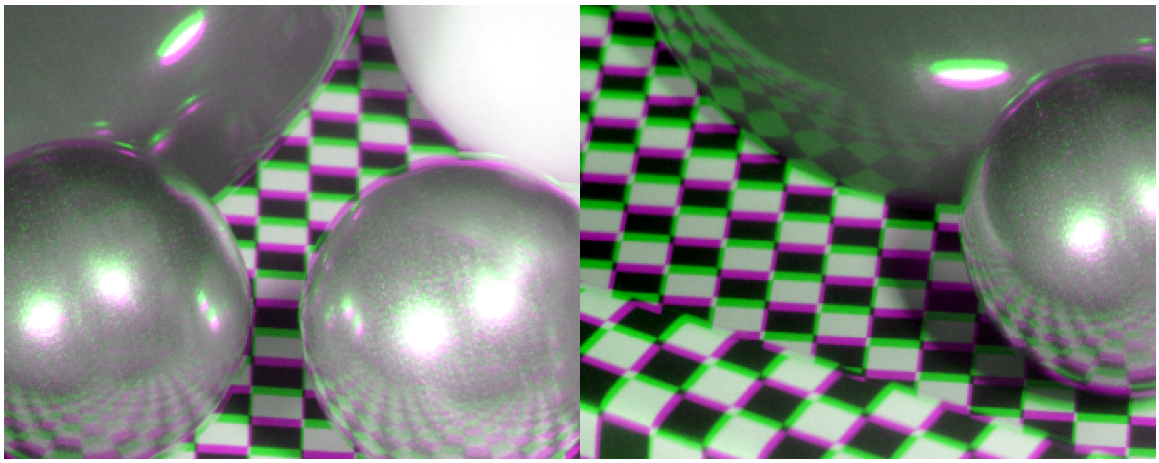

Figure 5-11: Polarization images overlaid of test scene. Image overlay 0° over 135° for 550 nm. Same results were found for any two of the four polarization angles captured.

Several algorithms have been proposed for the registration of the differently exposed *LDR* images before building the final *HDR* image. Some of them are based in percentile threshold bitmaps like [192]. Some others are based on *SIFT* key points extraction like [174], or novel proposed methods such as optimum target frame estimation [93]. However all these methods are not applicable for the registration of different *HDR* images already built. This is due to the fact that when the *HDR* images are already built, their pixel values are too different from the darkest regions to the brightest ones. Key features such as edges or corners are based in the high contrast between different objects. However, the magnitude of this contrast is very different for *HDR* images depending whether we are in a very dark or a very bright region.

We tried standard feature-based registration methods for estimating the transformation needed in order to align the *HDR* images of the 4 different polarization angles. We considered different types of affine transformations (translation, rotation, scale and shear), as well as combinations of them, but none succeeded. Therefore we decided to modify the registration work-flow by including a three-stage pre-processing step before applying the standard key-point based registration procedure [48, 202].

Our method has three stages:

- Stage 1: dynamic range logarithmic compression. We compress the dynamic range

in logarithmic scale to reduce the large difference between very bright and very dark regions.

- Stage 2: normalization. We scale the image values in the range $[0, 1]$, by subtracting minimum value and dividing by maximum value.

- Stage 3: contrast enhancement. We stretch the contrast of the image so that we allow 5% underexposed pixels and 5% saturated pixels. This way, even if up to 5% of pixel population is either very bright or very dark, we will still highlight the details in the middle exposure region.

Figure 5-12 shows a portion of overlay between the polarization images corresponding to 0° and 135° for the spectral band of 550 nm, before and after pre-processing and registration are applied. Same results were found for any two of the four polarization angles captured.

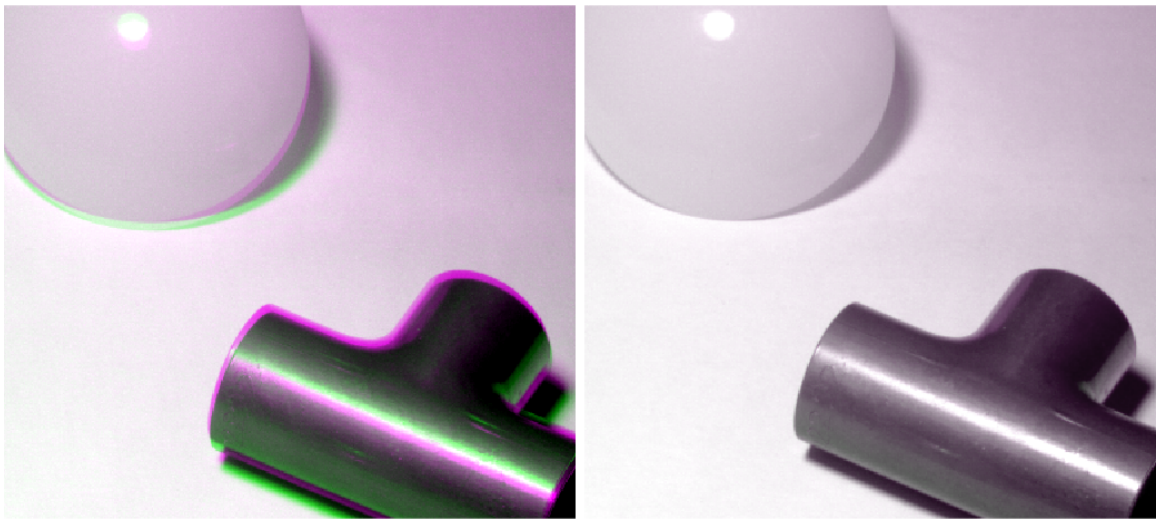

Figure 5-12: Image overlay 0° over 135° for 550 nm. Left: before registration. Right: after registration.

We see that before registration (left), there are colored bands close to the edges of the objects and their highlights. This means there is misalignment between the two polarization images. A similar situation is found for all pairs of images corresponding to different polarization angles. After the registration procedure, we can see in figure 5-12 (right) how the two images are correctly aligned. So we can conclude that the registration was performed satisfactorily.

The final transformation found for all images was a translation. The magnitude of this

translation depended on the scene, ranging from 5 to 15 pixels both in horizontal and vertical directions.

After all the described pre-processing steps were finished, our *MSHDRPol* image cubes were ready to use. These images will be made publicly available in the website of the *Color Imaging Laboratory* ([148]), after the defense of this thesis.

We focused our work in two main applications: segmentation and material classification. For each of those, we concentrated our efforts in trying to overcome the limitations found in the methods proposed in [179], which is to our knowledge the only reference available so far which deals with *MSHDRPol* image cubes for segmentation and material classification. We will explain both applications in the next sections. The first one is dedicated to the segmentation method description and validation experiments, and the next one to the classification method and its validation experiments.

5.4 Segmentation method and evaluation

The images we are going to capture and process contain common objects present in indoors scenes, made out of different metal and dielectric materials. The segmentation procedure we propose is composed of several steps. We used *Matlab R2016a* for processing the captured image data. In the pseudo-code of algorithm 1 we describe the different steps of the segmentation. We explain each of them in the following subsections from 5.4.1 to 5.4.5. The input to the algorithm is a *MSHDRPol* cube obtained after applying the pre-processing method explained in the previous section 5.3.

1: **Segment**$(cube.original)$
2: $RGB1 \leftarrow mshdr2rgb(cube.original)$
3: $RGB2 \leftarrow hlremoval(RGB1)$
4: $labels1 \leftarrow meanshift(RGB2)$
5: $cube2 \leftarrow labels2clusters(cube.original, labels1)$
6: $RGB3 \leftarrow mshdr2rgb(cube2)$
7: $labels2 \leftarrow meanshift(RGB3)$
8: $RGB4 \leftarrow labels2clusters(RGB2, labels2)$
9: $labels3 \leftarrow meanshift(RGB4)$
10: $labelsfinal \leftarrow mergeregions(labels3)$
11: **return** $labelsfinal$

Algorithm 1: Segmentation algorithm

Our proposed segmentation method involves sequential application of the mean-shift

algorithm [19, 22], together with the clustering of all pixels belonging to the same superpixel after a mean shift iteration is applied. Each time we apply a new iteration of mean-shift, we reduce the number of superpixels in the pre-segmented image.

We first produce a usable *RGB* image from the original cube with highlights removed (see subsections 5.4.1 and 5.4.2). Afterwards we apply mean-shift to get an initial crude estimation of labels for the different regions (see subsection .5.4.3). Later on, we use the original spectral cube and these labels to perform an initial clustering of superpixels data. From here on, we use only the *RGB* image of this initial clustering to further apply mean-shift and clustering to obtain a second set of more refined labels (see subsections 5.4.3 and 5.4.4). These labels are finally segmented and merged (see subsection 5.4.5) to obtain the output of the segmentation algorithm.

5.4.1 RGB image from spectral HDR image cube (mshdr2rgb)

This function receives as input a *HDR* multispectral image cube and uses the *CIE* 1931 Color Matching Functions ([118]) to get a *HDR XYZ* image, which afterwards is transformed into *RGB* color space. Later on we perform a tone mapping to get finally a *LDR RGB* image. The tone mapping includes logarithmic compression, normalization, contrast enhancement, histogram equalization and saturation boost. In figure 5-13 we can see an instance of the output *RGB1* image for one of the scenes.

Figure 5-13: RGB1 (see algorithm 1) image rendered after a *HDR* multispectral cube using function *mshdr2rgb* in algorithm 1.

5.4.2 Highlights removal (hlremoval)

This function receives an RGB image containing objects with highlights, and returns another RGB image where the highlights are reduced or eliminated. This is done by computing the negative of the image (for an $8-bits$ image: $Image_{negative} = 255 - Image$). Then we perform a morphological operator called region filling ([168]) to the negative image, and finally calculate again the negative of the filled image. Figure 5-14 shows the $RGB1$ image for one scene and the resulting $RGB2$ image after removing highlights. It is important to note that we remove the highlights to help in the segmentation task. As we will explain in subsection 5.5, we will use the information in the highlight regions for this task, getting it again from the original pre-processed images ($cube.original$).

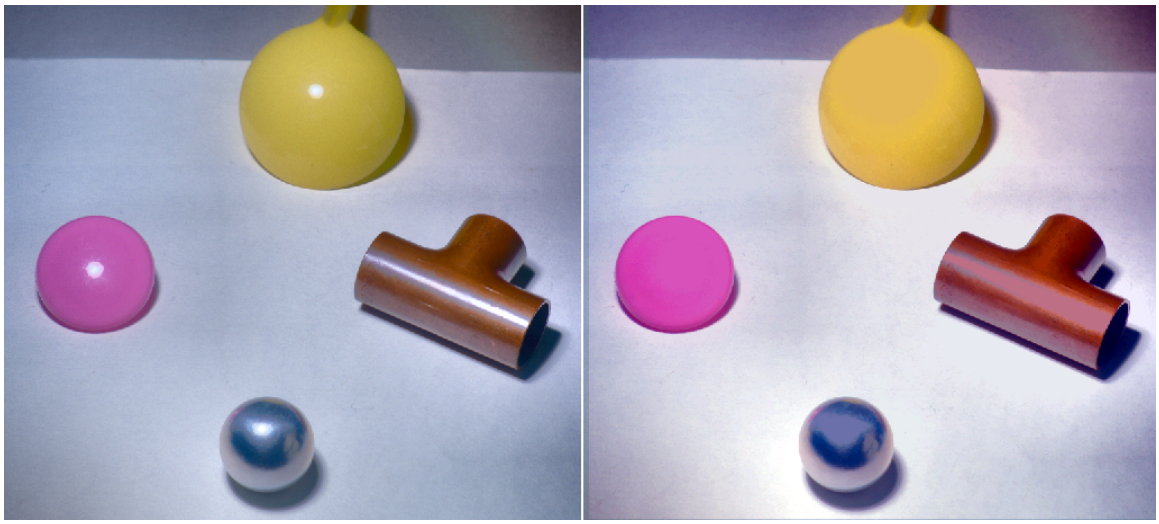

Figure 5-14: Left: $RGB1$ image rendered. Right: $RGB2$ image after highlights removal.

5.4.3 Mean shift

This function receives as input and RGB image and returns a label image. A label image is an image where each pixel value corresponds to a label. Mean shift algorithm ([19, 22]) is a pre-segmentation performed to find areas of the image where the pixels have similar values, and group them together under the same label. This cluster is often called superpixel.

5.4.4 Labels to clusters (labels2clusters)

Once the label images are generated, we can use a multispectral image cube or a RGB image to average the pixel values in each spectral or color channel in order to generate a new cube

or image which is much more simple than the original one. This is what this function does. It receives a labels image and an original image (color or multispectral), and returns an image of the same kind of the original, where the pixel values for those pixels under the same label (belonging to the same superpixel) are the same. This value is the average of all pixels in the same superpixel.

This new image can in turn be an input for a new mean shift iteration. In figure 5-15, we can see an *RGB* renderization of one of the highlights-removed images (*RGB2*), and the resulting *RGB* image (*RGB3*), after applying the first mean shift iteration.

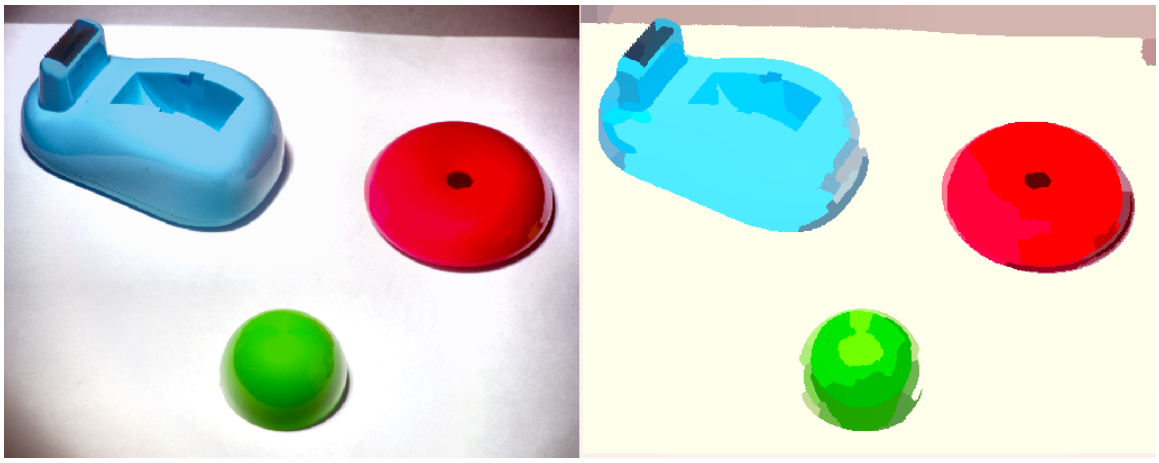

Figure 5-15: Left: *RGB2* image after highlights removal. Right: *RGB3* image after the first iteration of mean-shift and *labels2clusters* processing.

The performance was found to be better if we apply *labels2clusters* method using the label images and the spectral cubes in the first iteration, and then the label images and the *RGB* highlight-removed rendered image for the second iteration. The reason why we use first the spectral cube (*cube.original*) and second the *RGB* image (*RGB2*) to perform the *labels2clusters* method is that when the image is still rather complex, averaging spectral information produces a better input for the next step. However once the second mean-shift iteration has been performed, the *RGB* images are already simpler, yielding satisfactory results. Thus we do not need to use the spectral cubes anymore. This way we can save a step of *mshdr2rgb* method.

5.4.5 Region merging (mergeregions)

After mean shift is applied and we have used the label image to cluster regions in our image, we tried to reduce the number of clusters as much as possible, yet keeping the different objects

apart both from each other and from the background. We then apply an iterative region merging process which consists in the following steps. First of all we find the background, assuming it is the region with the biggest pixel area. We know that this might not occur in some cases where there is a very large and homogeneous object in the scene. However for the general case this condition worked fine. Afterwards we compute the adjacency matrix of the label image ignoring the background. This means that any region adjacent with the background would account as no adjacent, since we will never want to merge any object with the background. Once we have the adjacency matrix, we compute a dissimilarity matrix. We treat the image as a graph, in which the weight ($\omega_{i,j}$) between nodes i and j is ∞ for non adjacent regions, and calculated as shown in equation 5.2 for adjacent regions.

$$\omega_{i,j} = \sqrt{\sum_{k=1}^{K}(\rho_{k,i} - \rho_{k,j})^2} \tag{5.2}$$

where K is the number of color or spectral channels, and $\rho_{k,a}$ is the average sensor response for image channel k and region a. The fourth step is thresholding the dissimilarity matrix. Those regions which dissimilarity lies below a threshold value found by trial and error to be convenient for all our scenes were merged. Merging two regions means computing the average sensor response values for each channel and giving the same label to both of them. Therefore after each merging, we recomputed the adjacency and dissimilarity matrices again. This iterative process goes on until there are no adjacent regions for which dissimilarity metric falls below the threshold. In figure 5-16, we can see a false color labels image before and after applying the region merging. We can see how the number of regions in the image was reduced from 36 to 8.

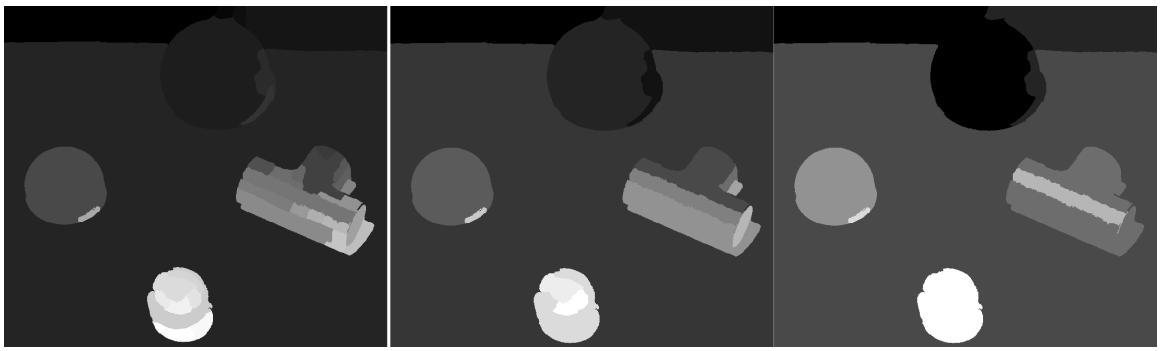

Figure 5-16: Left: Grayscale labels image before region merging (36 regions). Center: labels after thresholding with $th = 60$ (15 regions remaining). Right: labels after thresholding with $th = 120$ (8 regions remaining).

In this scene, the optimal number of regions would be 6 or 7. They would correspond with the 4 objects, the ground, and the wall (divided in two sides by the top-most object). We are reasonably close to this number using a threshold value of 120. Increasing this value would make the walls merge with the top-most object.The threshold value used performs reasonably well even with the residual highlights or shadows remaining in the scene after the previous steps.

In general, a threshold value of 120 performed rather well for all scenes tested. However a fine adjustment of this value would increase the performance of the segmentation for certain scenes.

5.4.6 Evaluation of segmentation procedure

In order to evaluate our segmentation results, we need to compare them with a reference segmentation or ground truth. For this purpose we segmented manually the images captured (see figure 5-17), with the help of a graphical user interface ([169]). We considered this segmentation as a perfect segmentation, and then compared how close our segmentation was from this ground truth (see figure 5-17). We considered to be background whatever was not an object in the scene.

Figure 5-17: Top row: *RGB* renderization of original spectral cubes. Bottom row: benchmark manually segmented. From left to right, scenes from 1 to 5.

The figure of merit we used for evaluation was the Jaccard's similarity index [60, 154, 153]. This metric measures the similarity between finite sample sets. It is defined as the size of the intersection between sets divided by the union of these sets. In our case, we use it to compare the benchmark labeling and the automatically segmented labeling for each scene. It is a measure of how much each labeled object overlaps with its ground truth labeling. A

Jaccard's index value of 100% would mean a perfect match, while the lower this value is, the more dissimilar the two segmentation results would be.

In order to compare our segmentation with that proposed in [179], we also used *Ncut* algorithm with the *RGB* highlights removed images (*RGB2*) of the scenes captured. Besides, we included a third widely used image segmentation method implemented in most image processing toolboxes, like Matlab. This method is called Watershed (*WS*), and it is based on morphological Watershed transforms, embodying edge detection, thresholding and region growing [8]. Table 5.1 shows the Jaccard index values for the five scenes captured and segmented using the three methods.

Table 5.1: Jaccard index values of the five scenes captured and processed for the three segmentation methods compared.

Scene	1	2	3	4	5
Our method	92.49%	87.62%	91.99%	95.23%	90.82%
Ncut	81.11%	84.72%	52.94%	90.01%	39.01%
Watershed	59.61%	71.85%	52.77%	54.41%	35.20%

We can see how, scene-wise, the smallest improvement of our proposed method is 2.9% better than *Ncut*, which we could consider almost negligible, and 15.77% better than Watershed. The largest improvement is 51.81% from *Ncut* and 55.62% from Watershed. If we average results, we find that our proposed method yields a mean segmentation accuracy of 91.63%, while *Ncut* method yields a mean accuracy of 69.56% and Watershed of 54.77%. That means that applying our method resulted in a relative 31.73% increase in segmentation accuracy from *Ncut*, and 36.86% from Watershed.

5.5 Classification method and evaluation

When dealing with the classification problem, we based our work in the method used in [180, 179], that uses the *DoLP* map curvature around the highlights. The *DoLP* map is computed from the four polarization angles images. Using these four images, we can calculate the Stokes parameters images ([14]), as well as the *DoLP* map for wavelength λ using the following equations:

$$S_0(\lambda,\theta) = I(\lambda,\theta) + I(\lambda,\theta+90°) \tag{5.3}$$

$$S_1(\lambda,\theta) = I(\lambda,\theta) - I(\lambda,\theta+90°) \tag{5.4}$$

$$S_2(\lambda,\theta) = I(\lambda,\theta+45°) - I(\lambda,\theta+135°) \tag{5.5}$$

$$DoLP(\lambda) = \frac{\sqrt{S_1(\lambda,\theta)^2+S_2(\lambda,\theta)^2}}{S_0(\lambda,\theta)} \tag{5.6}$$

In [180, 179], they classified the objects into metal and dielectrics by studying an area of 9 x 9 pixels centered in the brightest point of the highlights, detected by simple thresholding in the luminance image. Then the curvature along the direction of maximum variation of the *DoLP* surface was computed. According to the sign of this curvature (which indicates if the surface is concave or convex), the object can be classified into metal or dielectric material.

When we implemented this method, we found some difficulties in the calculation of this magnitude. Sometimes, the shapes of the highlights and the *DoLP* maps are not regular or rounded, but rather spiky and having elongated or non-rounded shapes. In these cases, calculating the curvature in both the direction of maximum variation of the *DoLP* map or even the direction of the highlight in the *HDR* image, was not yielding satisfactory classification results. We even tried studying the mean curvature along all directions around the brightest point, but results were still poor.

In figure 5-18, we can see two examples of *DoLP* map surfaces together with their corresponding *HDR* image areas and highlight surfaces. Both these highlights belong to two different metal objects. However, using the curvature of the *DoLP* map resulted in wrongly classifying one of these two objects as a dielectric. With this kind of *DoLP* pattern (with the brightest point in the middle of a ramp-like surface), the curvature varies quite a lot depending on the direction we choose. We found this pattern rather often, and not only in objects presenting elongated highlights.

We propose a new method of material classification based on the *DoLP* maps which consists in thresholding a ratio between the *DoLP* values within the highlight and in the surrounding area (see figure 5-19). Equation 5.7 shows how to calculate this *DoLP* ratio (R) value for each highlight. N is the total number of pixels in the highlight central area. M the total number of pixels in the diffuse surrounding area. δ_n^{hl} is the *DoLP* value in pixel n of the highlight area and so δ_m^{su} is the same for the surrounding area. As we can observe,

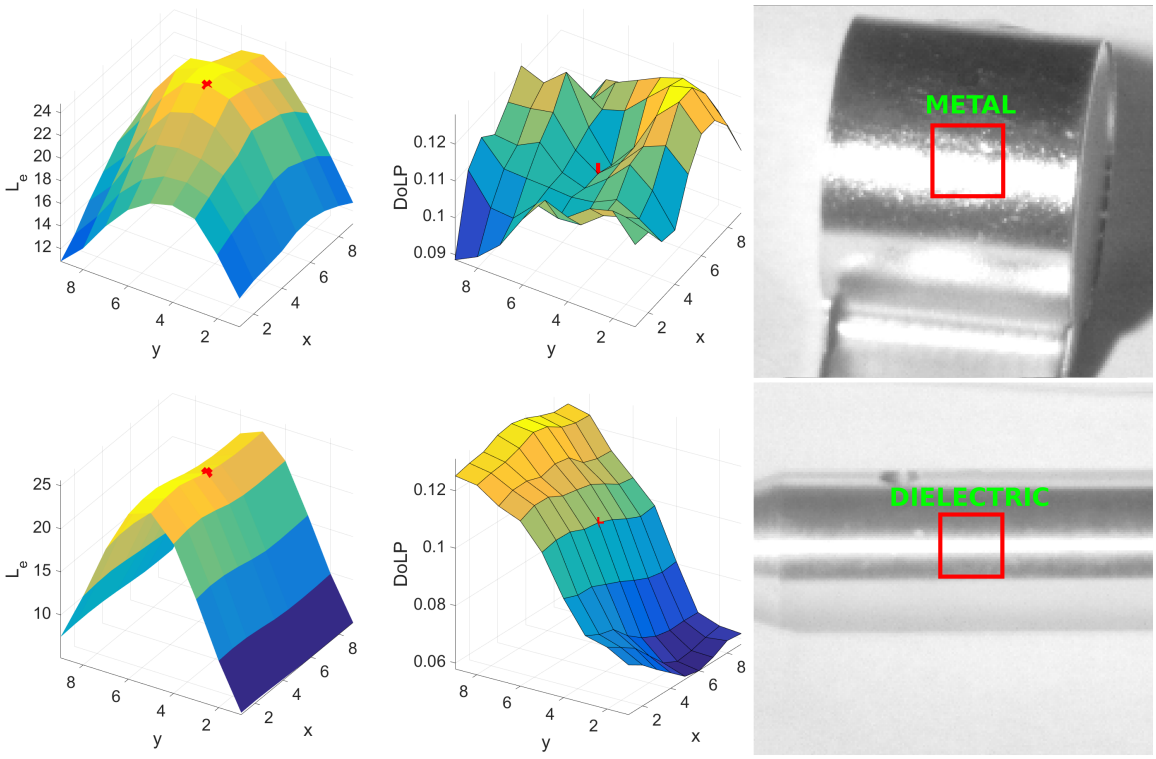

Figure 5-18: Example of two *HDR* highlight surfaces (left) of a metal object. *DoLP* surfaces (center) and 550 nm band images (right). The brightest point is highlighted with a red mark on the surfaces. Both are metallic objects. The top example is correctly classified as metal, but the bottom example is wrongly classified as a dielectric.

R just a ratio of average *DoLP* values in the two areas.

$$R = \frac{\sum_{n=1}^{N} \delta_n^{hl}}{\sum_{m=1}^{M} \delta_m^{su}} \cdot \frac{M}{N} \tag{5.7}$$

We keep a safe zone between these two areas to avoid including pixels of a highlight area in the computation of the surrounding area or viceversa. The central and surrounding areas are selected via a double region growing morphological image processing operator ([132]), around the brightest points of each detected highlight. We used a disk-shaped growing element of radius 5 two times consecutively. First we grow to get the safe area out of the central area, and then we grow once more to get the surrounding area. Afterwards we just subtract both to end up with the ring-shaped surrounding area. We can see an example of this in figure 5-19.

On the left side we have a zoomed-in area of the image, which contains a highlight. The central region of the highlight (which sensor responses are in the numerator in equation 5.7)

is shaded in green color. The surrounding area of the highlight (which sensor responses are in the denominator in equation 5.7), is shaded in blue color. Between both regions, there is a safe area shaded in white color.

On the right side se can see all the highlights detected for one of the scenes. In each of them we see both the central and surrounding regions, as well as the value of the *DoLP* ratio (R) calculated for each highlight and the class assigned automatically to it.

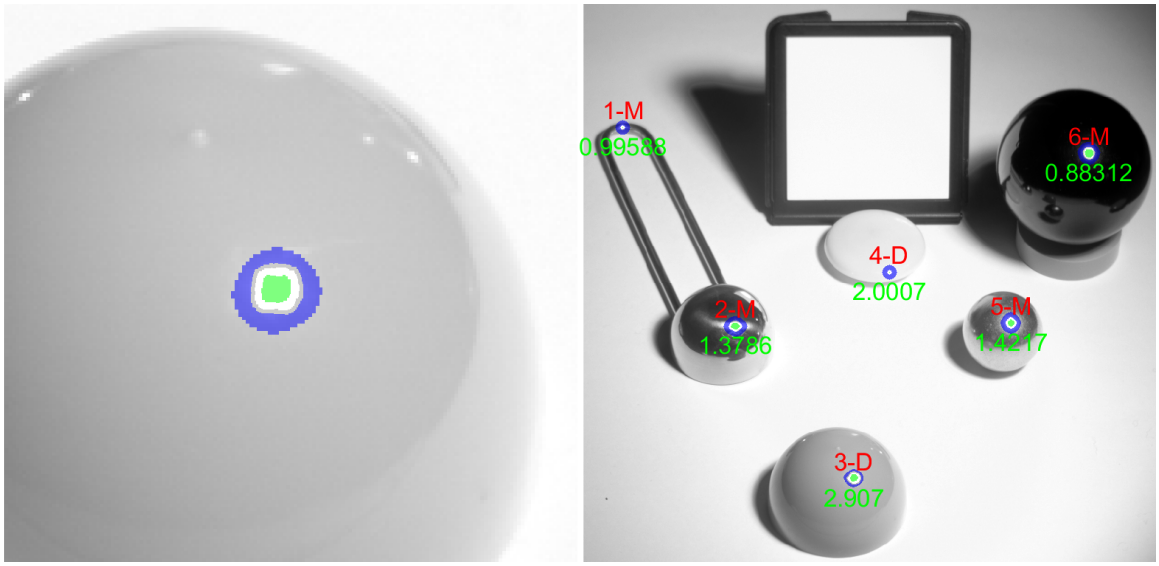

Figure 5-19: Example of highlight detected. The central area is shaded in green color, and the surrounding area in blue color. The green numbers are the R ratio of each highlight calculated as shown in equation 5.7. The reference white patch was included just for monitoring the illuminant.

Once the ratios for all the objects included in our scene set are computed as shown in equation 5.7, we noticed that their values for metal objects were lower than for dielectric objects. Thus, our idea was to make a thresholding of this ratio value to classify each highlight. In order to find an optimal threshold value, we proposed to use a set of highlights with known labels (training set). For this set, the optimal value is found by brute force approach in a fast and easy way. Once we find the threshold value yielding the highest classification accuracy for the training set, we compute its percentile value for the whole distribution of ratio values in this set. Afterwards, when we get a set of test highlights, we can compute the optimal threshold value using the same percentile with the new distribution of ratio values and check the classification accuracy for validating our proposed method.

In figure 5-20, we see another example of scene with the highlights automatically detected

and classified. On the left side we see the results of our proposed method using a threshold value of 1.2 (which was optimal for this particular scene). However, as we explain in section 5.5.1, the optimal threshold value found for all scenes tested is 1.5. Therefore, using this global optimal value, highlight number 4 would be misclassified as a metal. On the right side we see the result of applying the method used in [179], where highlight number 4 was also misclassified.

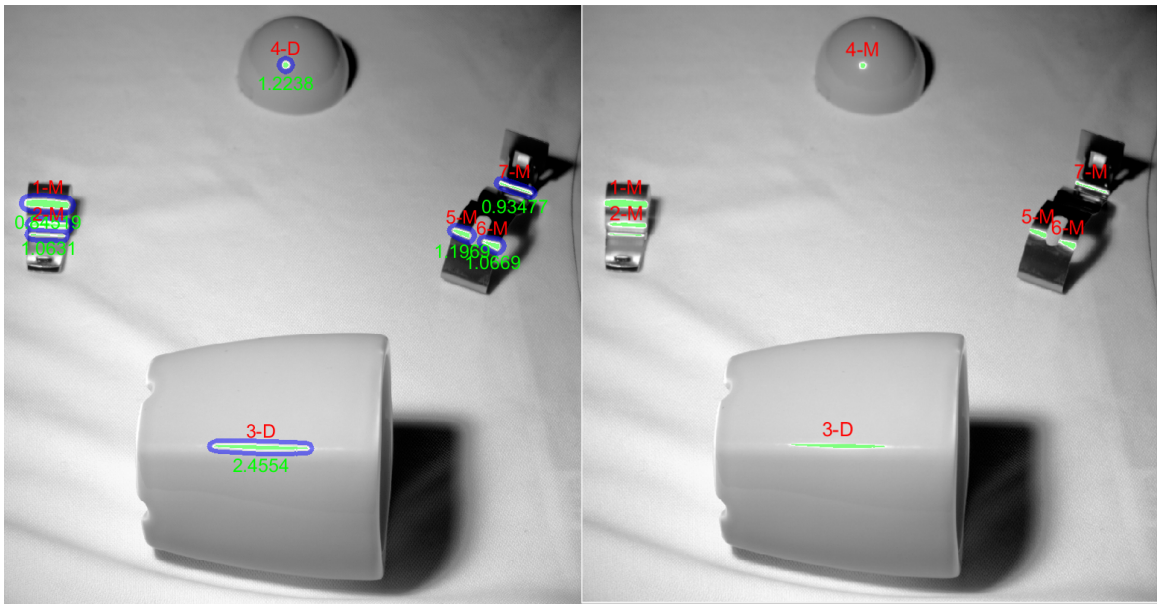

Figure 5-20: Example of highlights detected (green) and their surrounding areas (blue) in scene 2. Each highlight is automatically classified as metal (M) or dielectric (D), according to its ratio value (green text under each highlight). Left: proposed method using a threshold value of 1.2. Right: method in [179].

5.5.1 Evaluation of classification method

We have created a benchmark dataset of manually classified highlights detected in the scenes. After running the automatic classification using our proposed method (explained in subsection 5.5), and the method proposed in [179], we compared their respective classification accuracies.

We randomly divided the full highlights data set (72 highlights from 10 scenes captured only in the spectral band of 550 nm) into two halves. One of them was the training set, and the other one the test set. In the training set, we looked for the optimal threshold value by a brute force approach. We tested all threshold values from 0 to 4 by steps of 0.1. In figure

5-21, we can see the classification accuracy in the training set for each threshold value.

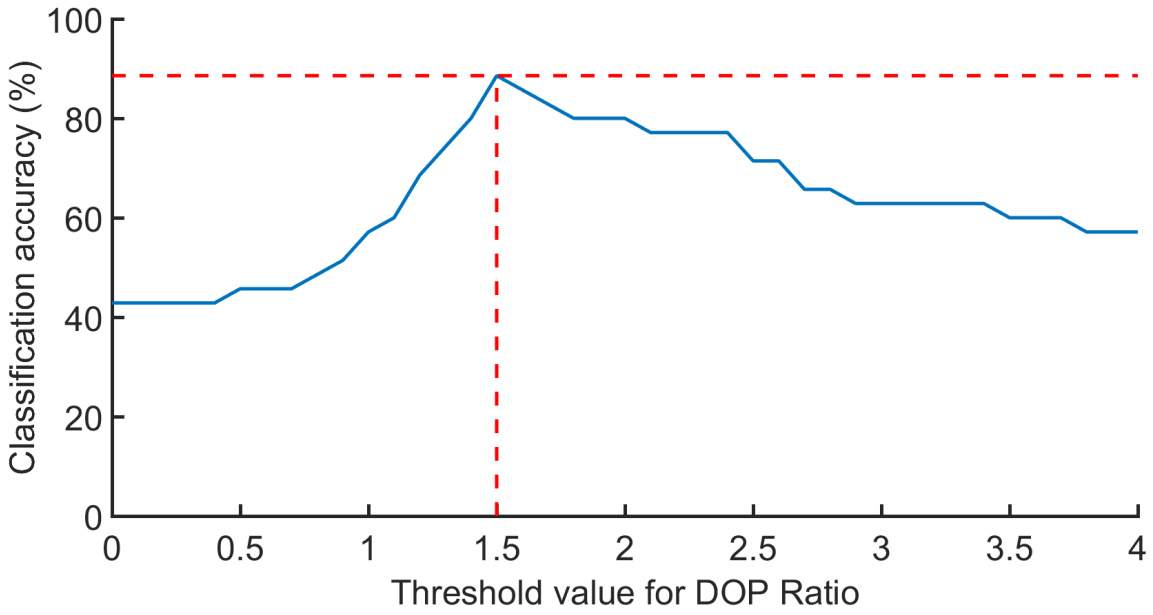

Figure 5-21: Classification accuracy vs threshold value for training set.

As we can see, the optimal threshold value calculated was 1.5 with an maximum accuracy of 88.57% for the training set. Since we divided the data set into train and test sets randomly, different instances of this process yielded different overall classification accuracies. After 20 trials, we found that the same threshold value was estimated in all of them.

Applying this threshold value for the classification of the test sets, gave us accuracy values ranging between 86.11% and 91.89%. The mean accuracy for the 20 trials was 89.19%. As an overall accuracy for the whole data set, using a threshold value of 1.5, we got a 90.28% classification accuracy. In figure 5-22, we see all the highlights in a cloud of points relating *DoLP* ratio and radiance of brightest area in each highlight. The threshold value found as optimal for the full set is drawn in green color.

On the other hand, using the classification method in [179] based on curvature of *DoLP* map, we got a total accuracy of 66.67% for the full set. We found that this method was failing to classify most objects for which the brightest point in the highlight did not correspond to a local maximum or minimum in the *DoLP* map. On the other hand, a fair comparison should mention that this method does not require for a training set.

In general, using our method we also observe that the classification accuracy for the dielectric objects is lower (77.42%) than for the metal objects (95.12%). In other words, the

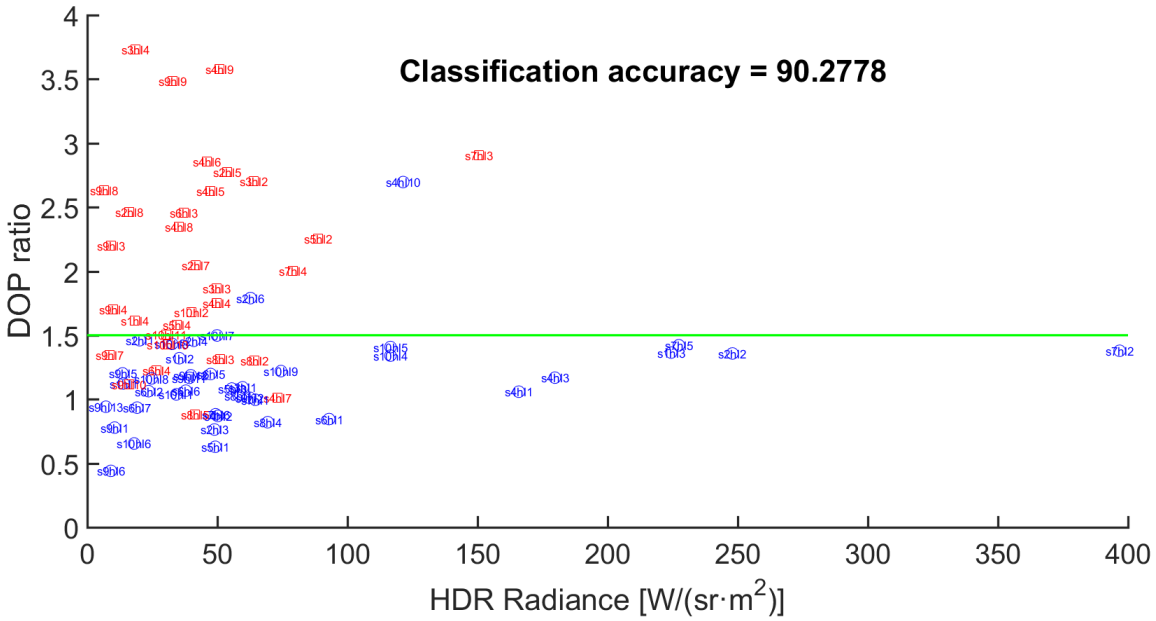

Figure 5-22: *DoLP* ratio vs *HDR* Highlight radiance for the whole data set of highlights. Red: dielectric objects. Blue: metal objects.

number of false positives in the metal class (classifying a dielectric as a metal) is 28.52%, and the number of false positives in the dielectric class (classifying a metal as a dielectric) is only 4.88%. This could be due to the fact that dielectric objects are from very different base materials (plastics, ceramics, etc, which can also be coated). Thus they behave in a more heterogeneous way in terms of *DoLP* ratio than metals.

Finally, if we only consider the results of the five scenes which were used as well in the segmentation procedure, we get a global classification accuracy of 90.91%. Therefore for the full framework, and in realistic conditions, we would get roughly more than 9 out of every 10 objects correctly classified.

We have seen how for the segmentation method, the full set of multispectral images has been used during the procedure. However in the classification method, only one spectral band was used. We found the best candidate spectral band for classification to be 550 nm. Besides, the responsivity of the capturing system is highest in this band, thus reducing the exposure times needed for the capture. Hence, if only classification of highlights is going to be performed, capturing a single spectral band could be enough to do it.

5.6 MSHDRPolVISNIR outdoors

The full method proposed in this chapter, for both capturing and processing, and also the two applications segmentation and object classification, were tested for indoors scenes. The imaging system, as explained in section 5.2, was a scientific monochrome camera attached to a *LCTF*.

We also wanted to test the capturing method in outdoors scenes. Outdoor illumination might suffer relatively rapid changes as well, which make not convenient the use of an *LCTF* to capture the full spectrum of each pixel in an outdoor scene. For this reason, instead of using the same imaging system, we decided to capture 8-channels images using a filter wheel camera model *Pixelteq Spectrocam VIS-NIR* (see figure 3-9). This camera is a 12-bits monochrome camera attached and synchronized with a 8-slots filter wheel. We chose this system because a filter wheel camera allows us certain spectral tunability by changing the interference filters. Compared with the *LCTF*-based system used in previous experiments, we lost the easy tunability (changing the interference filters was easy but not to be done on-line) and the narrow-band filtering, but we gained information in the *NIR* range of the spectrum ranging from 700 nm to 1000 nm. Besides, the exposure times were reduced with respect to any 8 channels of the *LCTF* configuration, because the filters of the *Pixelteq SpectroCam* have wider transmittances. This is another reason for choosing this system for outdoors scenes.

As this system does not polarize light, we included a linear polarizer in front of it, attached to a goniometer to control the angle in which we rotate it (see figure 5-23).

In our case we rotated the polarizer manually, but we could consider as well to use a motorized rotation stage.

5.6.1 Filter selection

The first thing we did was to select the set of 8 filter we were going to use. We were offered a catalog of 145 commercially available interference filters among which we could choose. Out of this list we discarded some of them as we will explain later and remained with a set of 121 candidate filters. In appendix B, we show the filter data we were provided by the manufacturers of the 121 candidate filters. We were provided with the central wavelength and bandwidth data of each filer, as well as with the relative spectral responsivity of the

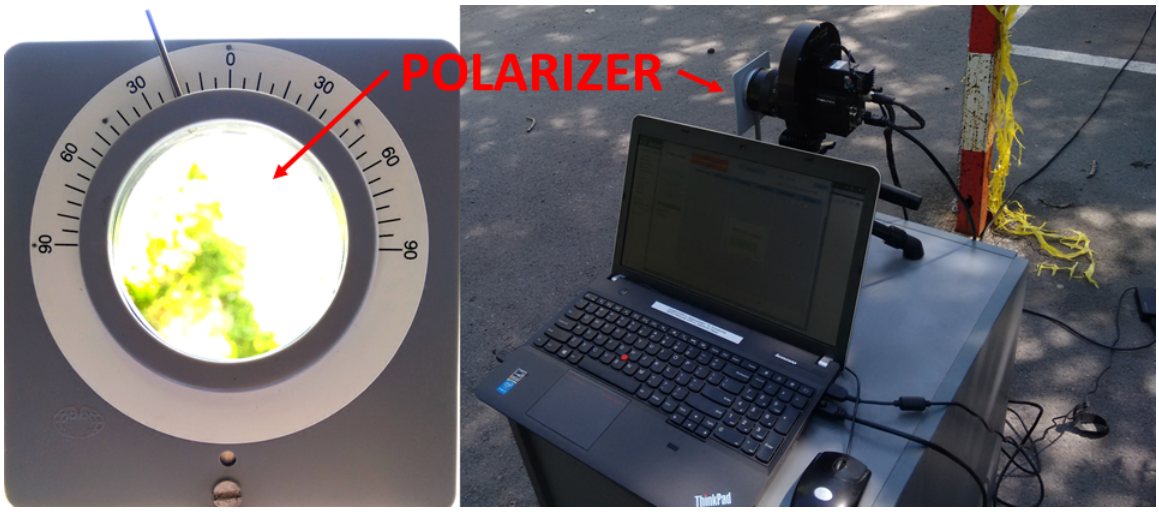

Figure 5-23: Spectrocam set up used for capturing outdoor scenes with a linear polarizer attached to a goniometer in front of the camera.

sensor. Therefore, we performed filter selection by simulations. In figure 5-24 we can see the spectral data we used for simulations.

We simulated camera responses to spectral radiance images, which were computed using spectral reflectance images of urban scenes acquired using a hyperspectral imaging device model *PhotonEtc. HImager* ([28]). Then we divided the spectrum from 400 nm to 1000 nm in 8 different regions (see table 5.2), and tried every possible combination of filters in which there was only one filter per region. This was done to reduce the huge number of possible combinations, as well as for ensuring that the whole visible and *NIR* ranges of the spectrum were covered with a single 8-channels system with certain overlap between adjacent bands. In figure 5-25 we show the *CIE Lab* coordinates for the 2526 reflectances used. These reflectances were selected by downsampling the full spectral images, in order to get spectral information from all objects present in the scenes.

Table 5.2: Regions of the spectrum in which we divided it to search a filter in each of them. Minimum and maximum wavelength of each region

Region	A	B	C	D	E	F	G	H
Minimum λ (nm)	400	451	501	551	601	651	701	851
Maximum λ (nm)	450	500	550	600	650	700	850	1000

The spectral reflectances were in the range $[0, 1]$. As the reflectances resulting from the *HImager* capture are slightly spiky, we did a pre-processing step consisting in smoothing each of them, using the Matlab function *smooth*, and a window spanning 5 data points.

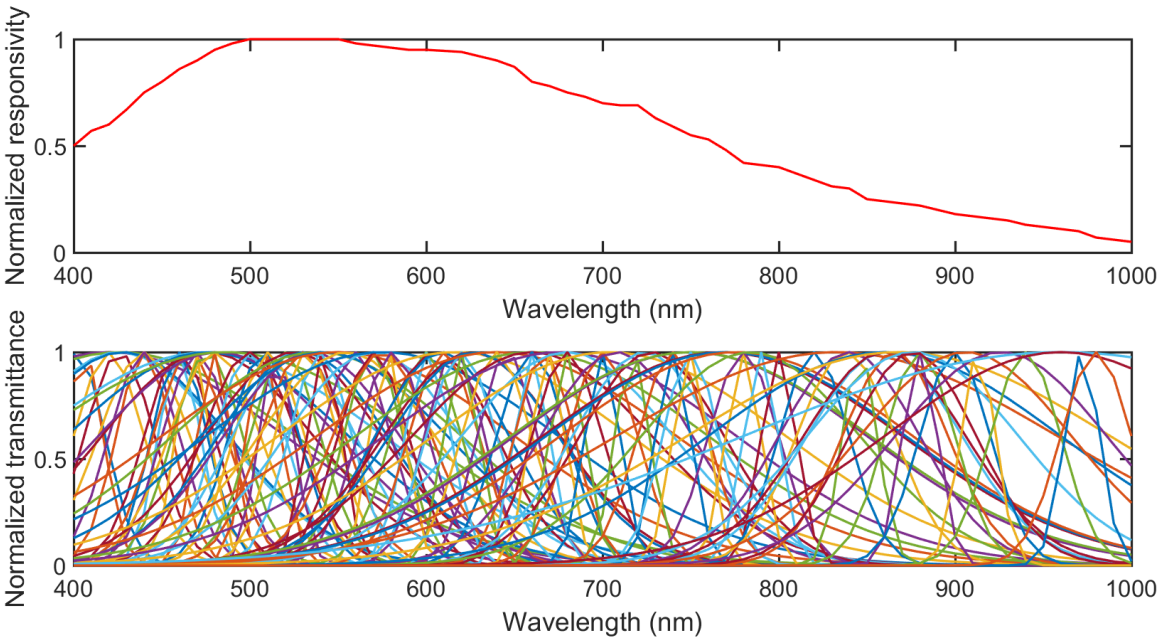

Figure 5-24: Top: spectral responsivity of *Spectrocam* sensor. Bottom: spectral transmittances of the 121 commercially available filter candidates.

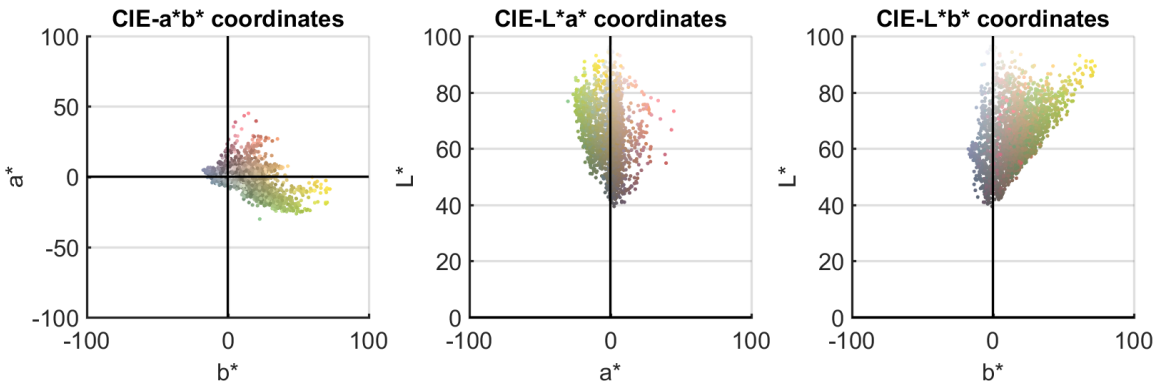

Figure 5-25: *CIE Lab* coordinates of the 2526 spectral reflectances used for the simulation of camera responses for filter selection. Left: $a^* - b^*$ plot. Center: $L^* - a^*$ plot. Right: $L^* - b^*$ plot.

As the imaging system allows us to choose a different exposure time for each spectral channel, we scaled the sensor responses for each channel so that the response for the whitest pixel (that of highest integrated radiance), was 100 of the dynamic range (meaning 4095 digital counts for 12 bits). This does not ensure the lack of saturated or underexposed pixels, but yielded a good average of less than 2 of sensor responses not properly exposed (very bright or very dark ones).

From the list of 145 candidate filters, we discarded several of them since their bandpass were so narrow that the exposure time needed to be very long as the incoming signal was too low on the sensor. Every filter with *FWHM* lower than 20 nm was discarded. We also discarded those filters with similar *FWHM*, if their maximum transmittance wavelengths were closer than 5 nm to another similar filter. So after this purge, 121 candidate filters were remaining.

We tried all possible combinations of 8 channels in which only one channel has its maximum responsivity in each of the 8 spectral regions. This way, we ensure that the whole spectral range is covered without leaving any gap in every combination tested. There were a total of 54432 combinations tested (before discarding some filters the possible combinations were several hundred billions). For each combination, the 2526 samples' reflectances were recovered using a double 10-folds cross-validation method and logarithmic kernel-ridge regression method. This way we ensured that the samples used for optimizing the regularization parameter of the logarithmic kernel regression method, were never used for evaluating its performance ([31]). The samples were previously randomly scrambled to ensure similarity in the statistical distribution of each fold.

Since some previous trials showed in the error surfaces that the optimal parameter found was pretty stable for all folds of the inner loop, the search for optimal parameters was done using only the following 4 possible values: $[10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}]$. In every outer iteration, the mode of the 10 inner 10-folds iterations was chosen as optimal parameter. We chose the mode instead of the mean because since we are dealing with different orders of magnitude, sometimes we found that one or relatively few outliers could bias the mean calculation towards a non-optimal value.

The parameter optimization was done according to minimal *RMSE* over the whole spectral range, since it was found to perform better than using any other error metric as a cost function. For each filter combination we compared *RMSE*, *GFC* and *CIEDE00* color differ-

ence with the original reflectance signals. Then the combinations that yielded best *RMSE*, *GFC* and *CIEDE00* results were chosen. We can see in table 5.3, the central wavelength and *FWHM* of the best performing filter combinations.

Table 5.3: Central wavelengths (λ) and *FWHM* of the filter combinations found to yield best color and spectral metrics. All data given in nanometers.

Best *RMSE*		Best *GFC*		Best $\Delta E00$	
λ	*FWHM*	λ	*FWHM*	λ	*FWHM*
425	50	425	50	425	50
482	56	482	56	482	56
555	50	530	24	530	24
615	100	570	50	555	50
680	50	650	100	606	34
770	102	660	59	660	59
833	125	833	125	833	125
950	100	950	100	950	100

We see in table 5.3 that the filter combinations yielding best *RMSE*, best *GFC* and best $\Delta E00$, are pretty similar. We found during the optimization that in general, using the best *RMSE* metric produced the best results. Therefore our selection was the filter set in the left column of table 5.3. In figure 5-26 we show the real filter transmittances of the selected filters, measured using spectroradiometer *Photo Research PR*-745.

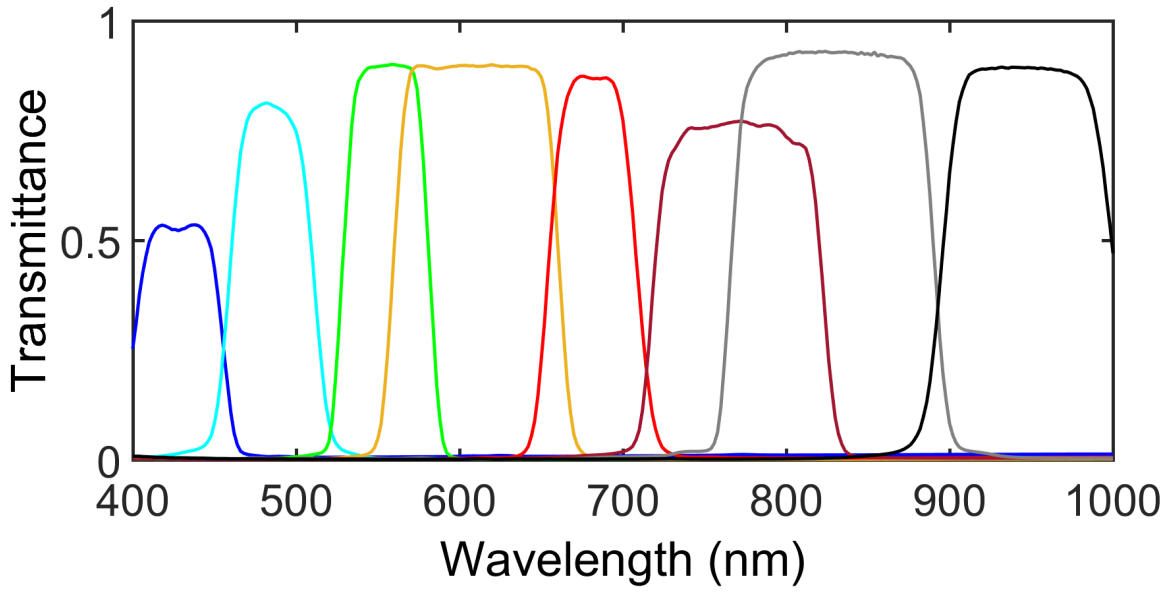

Figure 5-26: Spectral transmittances of the 8 filters chosen for the spectral system, measured using a spectroradiometer.

5.6.2 Radiometric calibration

We have set up a scene in our lab, containing an *X-rite color Checker* classic with 24 color patches. The target was illuminated with a 500W incandescent lamp with an angle of approximately 45°. We eliminated all color filters from the *Spectrocam*. Thus, the images acquired were monochrome images covering the whole responsivity range of the camera, roughly ranging from 400 nm to 1000 nm (see figure 5-24, top).

We captured 10 images with each of the following 11 values of exposure time: 0.064 ms, 0.1 ms, 0.5 ms, 1 ms, 5 ms, 10 ms, 50 ms, 100 ms, 500 ms, 1 s, and 2 s. In figure 5-27, we can see the monochrome images captured with the different exposure times. We skipped the images captured with 1 s and 2 s because both of them are completely saturated.

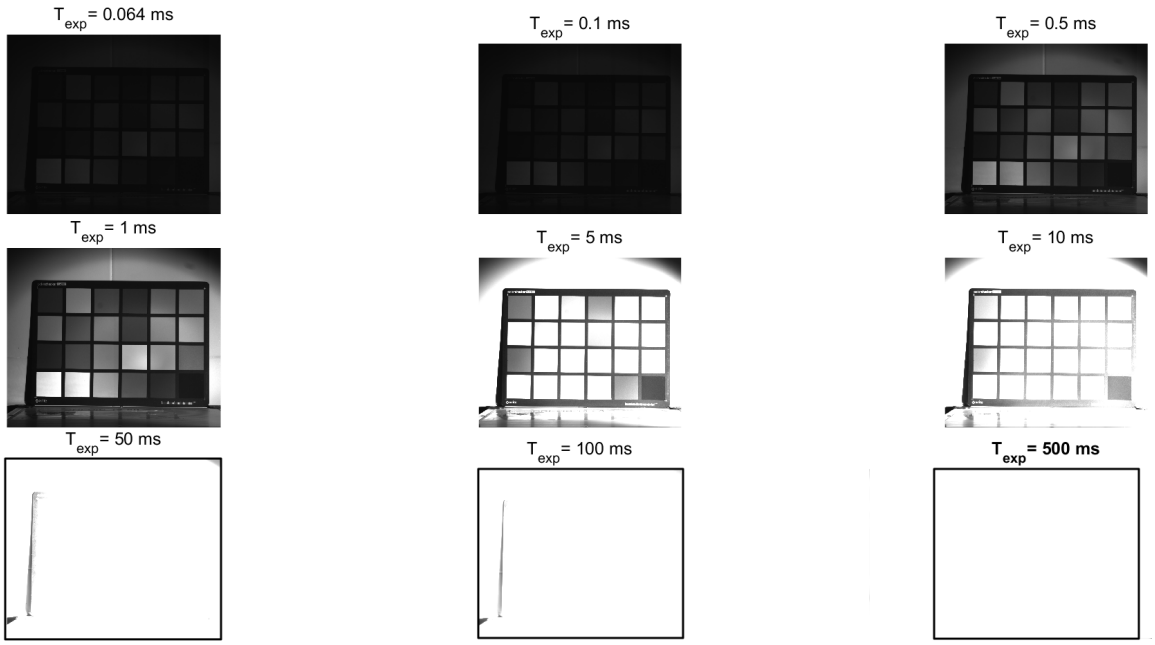

Figure 5-27: Monochrome images captured with different exposure times, used to perform radiometric calibration of the Spectrocam.

As the whole scene was completely static during the capturing process, we consider the illumination temporally (not spatially) constant over it. In other words, during the whole capture, the radiance coming from each point of the scene was not changing. We monitored it by measuring the spectral radiances of the 24 patches of the *Color Checker* twice. Once just before the image capturing and once just after it. The final spectral radiances used for the calculations were the average of these two measurements for each color patch. No significant changes were found in these radiances along time. In figure 5-28, we can see the

24 radiances measured before (green) and after (red) the image captures and their average vectors (blue). In all plots, the X-axis correspond to wavelength in nanometers, and the Y-axis to spectral radiance in $\frac{W}{sr \cdot m^2 \cdot nm}$. Axis labels and legends were excluded for the sake of clarity in the plots.

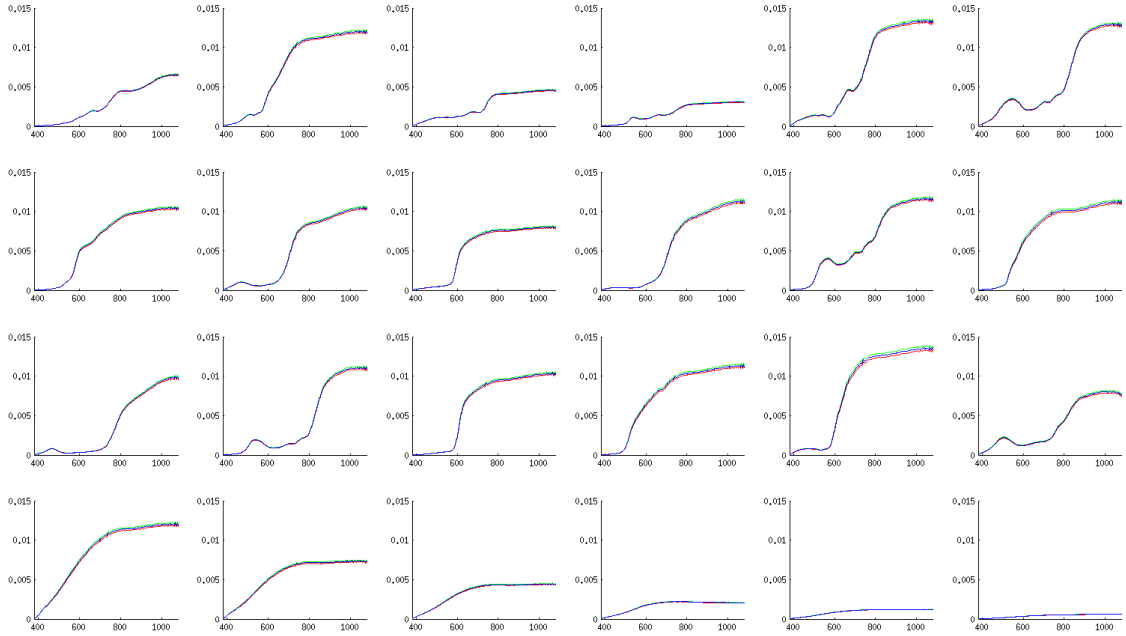

Figure 5-28: Spectral radiances measured for each patch of the *Color Checker* to check illumination stability. Green: before the scene capture. Red: after the scene capture. Blue: average of both measurements.

For each exposure time, the ten images captured were averaged. Besides, also in each exposure time, 10 dark images were captured with the lens cover on, and averaged in order to have images with dark signal information. The mean dark image calculated was subtracted to the mean image, in order to eliminate the dark current noise. These resulting images were then studied.

We calculated the mean and standard deviation values of the dark images over the whole image area. The camera captures 12 bits raw image data but we converted it to 16 bits since we anyway need to use 16 bits data container to store 12 bits image data. We found that the average dark noise was pretty constant with exposure time, with a value around 2100 DC (3.2% of the entire dynamic range). However the standard deviation did increase from 50 DC for exposure time values below 100 ms, to 250 DC for the 2 seconds exposure time.

Once the images were averaged for each exposure time and the dark images were subtracted, we retrieved the sensor responses corresponding to each of the 24 color patches. We did it averaging an area of 22×22 pixels in the center of each color patch (roughly the area seen by the spectroradiometer (PR) with the aperture of 0.2 degrees at the same distance than the camera captured the images). We assumed that the areas seen by the PR in the 24 color patches and the areas averaged to retrieve camera sensor responses in each color patch were roughly the same in size and place. We plotted these average sensor responses versus exposure time (for each color patch), and versus pseudoexposure (product of integrated radiance times exposure time).

In figure 5-29, we see the sensor responses in 16 bits for the 24 color patches with the different exposure times. We can observe how the behavior of the camera is rather linear with exposure time for each individual patch (same radiance). In the left-side axis we plot the first 12 color patches and in the right-side the last 12.

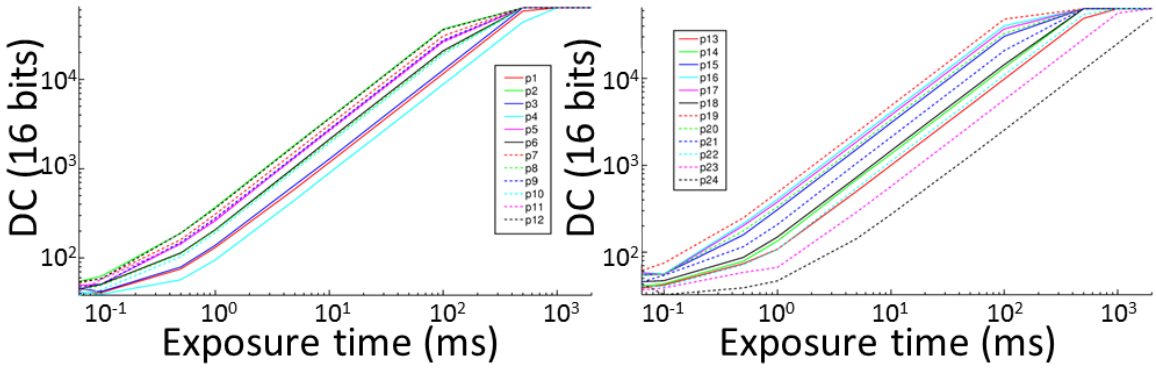

Figure 5-29: Camera responses in 16 bits digital counts versus exposure time in milliseconds Left: first 12 color patches of the color checker. Right: last 12 color patches of the color checker.

As we could expect, the sensor responses are linear with exposure time, and the slope of this function is the same (lines are parallel) for every different integrated radiance (color patch) within the scene. However, in order to say that our camera is linear, it would no be enough to check its response versus exposure time only. Therefore we also have to check how the camera responds to the amount of light impinging on the sensor. For this purpose, in figure 5-30 we plotted the product of integrated radiance times exposure time (pseudoexposure) in the X-axis. We have included the complete cloud of points, but we used different marker shapes for the data corresponding to different exposure times, to see how they distribute. The final shape of the cloud of points, is what Debevec and Malik [26]

call the *CRF*. In this case, it is an absolute *CRF*, since the X-axis is composed by absolute values of time and radiance.

We have linearly fitted the cloud of points (red line in the plot). We cropped those values below 3300 DC (approximately around 5% of the dynamic range) and above 62300 DC (approximately around 95% of the dynamic range). We consider the former underexposed (dark noise) and the latest saturated. Thus, for all our calculations, we will not consider those pixel values out of the linear range of the *CRF*. If some pixels have values below this linear range, the *AEE* algorithm will search for a longer exposure time. If on the other hand some pixels have values above the linear range, the *AEE* algorithm will search for a shorter exposure time. The fitted linear *CRF* is the one we used for the *HDR* capture using *AEE* method (see chapter 4) and for building the *HDR* radiance maps (see equation 5.1). The weighting function (ω) used for building the *HDR* images, was the 16-bits version of the one shown in 4-10. As we see, this function has a range in which its values are 1. If we make this range to match the linear range of the calculated *CRF*, then we will ensure that those values we considered underexposed or saturated will not be taken into account when building the *HDR* image.

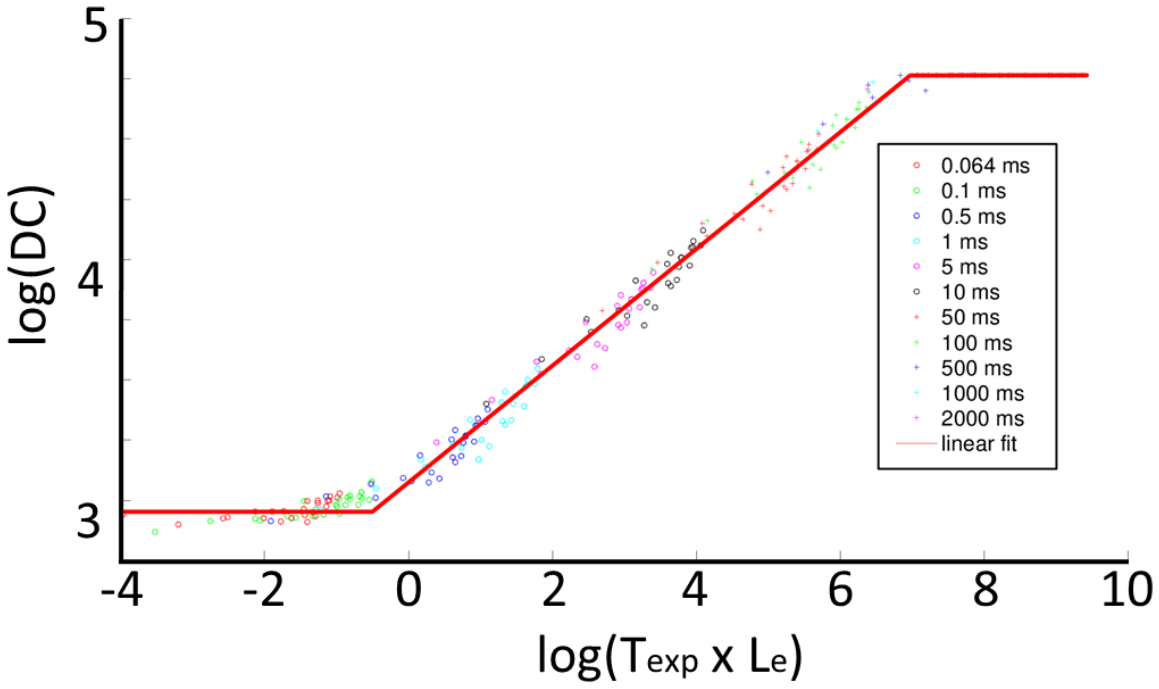

Figure 5-30: Absolute *CRF* measured for the Spectrocam, and linear fit used for *HDR* capture with *AEE* and *HDR* image building.

5.6.3 Scenes captured

Once the camera was calibrated, we captured several *MSHDRPol* images of outdoor scenes using the same method explained in section 5.2. Afterwards we also processed them using the method explained in section 5.3. For this camera, we did not need to perform any registration though, as we found no misalignments between different spectral bands or polarimetric angles. The rest of the processing done after capturing and before the *MSHDRPol* images were ready was the same.

We captured a set of 8 scenes. Since the images captured were multispectral, including the visible and *NIR* ranges of the spectrum, we can call these images *MSHDRPolVISNIR* images. The whole set of images after pre-processing will be made publicly available at our Lab's website ([148]) after the defense of this thesis.

In figure 5-31, we can see a false *RGB* version of the 8 images captured rendered just for displaying the content of the scene. As these images were *HDR*, polarimetric and multispectral images, we built for displaying a tone-mapped version of the 0° polarization angle taking channels number: 5, 3 and 1 as red, green and blue color channels respectively.

Figure 5-31: False *RGB* tone-mapped image renderizations from the *MSHDRPolVISNIR* images captured from outdoors scenes. Scenes are numbered from left to right and from top to bottom from 1 to 8.

As we can see, the scenes contain elements commonly present in urban scenes, like: buildings, vehicles, ground, sky and cars. Each of these classes is rather heterogeneous. For instance if we study the class 'vehicle', we can see how these objects include metal, glass, tires and plastics. This fact makes the application of segmentation or classification

techniques based only in spectral information potentially more complex.

To illustrate of how big the amount of information of one of these images is, compared with the amount of information of a monochrome image of the same size, we have designed figure 5-32.

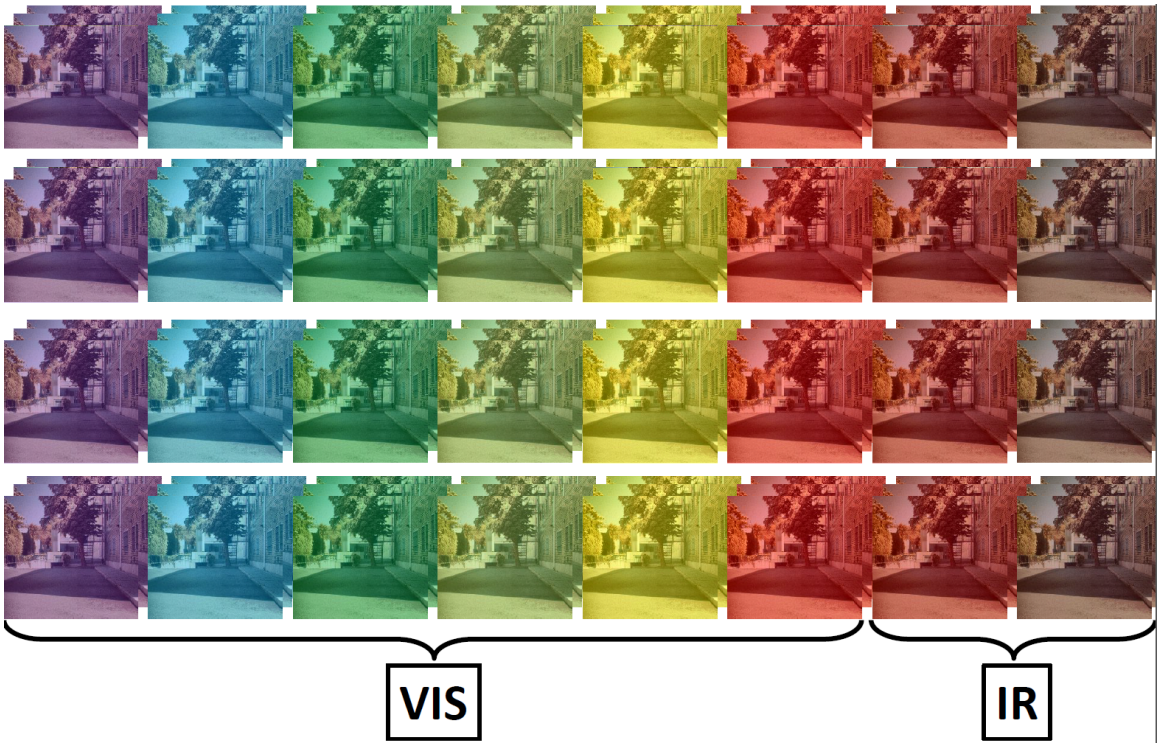

Figure 5-32: Scheme of one multispectral *HDR* polarimetric visible and *NIR* image. Each image plane is equivalent to one 12-bits *LDR* image captured.

In figure 5-32, we have made a spatial 3D composition of the *LDR* images captured for a single scene and exposure time set. In columns we have all spectral band images for the same polarimetric angle, and in different rows the different polarization angles. Note that the third dimension in this figure corresponds to the differently exposed images captured. In general, for the scenes captured, the *AEE* algorithm always needed 3 different exposures per band and angle. However this number can vary depending on the content of the scene and also on the settings used for the *AEE* algorithm. The different false colors given to the images are just for reference. Note also that the first 6 channels cover the visible range of the spectrum (roughly from 400 nm to 700 nm), and the last two channels the *NIR* portion of the spectrum approximately from 700 nm to 1000 nm. Each of these *MSHDRPolVISNIR* images was approximately 300 Mega-Bytes in size.

Some of the benefits of the captured images are shown in the following figures. First of all, in figure 5-33 we can see the comparison between a normal *LDR* spectral image (left), and its tone-mapped *HDR* version (right). They are captured in the spectral band centered in 555 nm with a bandwidth of 50 nm, with polarization angle 0°. We can see how on the *LDR* image, the details in the bright and dark areas are saturated and underexposed respectively. Conversely, in the *HDR* image, no details are lost. This trend was found for all spectral bands and polarization angles.

Figure 5-33: Grayscale images of scene 1 (see figure 5-31) captured at 555 nm with a bandwidth of 50 nm and polarization angle 0°. Left: *LDR* image. Right: tonemapped *HDR* image.

Another one of the applications we can use our captured images for, is detecting vegetation present in the scenes. We can do it by just calculating a simple ratio image as shown in equation 5.8.

$$I_{Ratio} = \frac{I_{833nm}}{I_{615nm}} \tag{5.8}$$

Where I_{Ratio} is the vegetation ratio image, I_{833nm} is the image captured at 833 nm, and I_{615nm} is the image captured at 615 nm. Once this ratio image is computed (center images in figure 5-34), we normalize it to the 8-bits range from 0 to 255 DC. Later on we only need to threshold it using the threshold value of 50 DC. This value was found by trial and error and worked for all scenes captured.

In figure 5-34, we show the false *RGB* images of scenes number 4 and 6 (left column), the

ratio image (I_{Ratio}) computed as shown in equation 5.8 (center column), and the vegetation detected after thresholding the ration image with a threshold value of 50 DC (highlighted in green color).

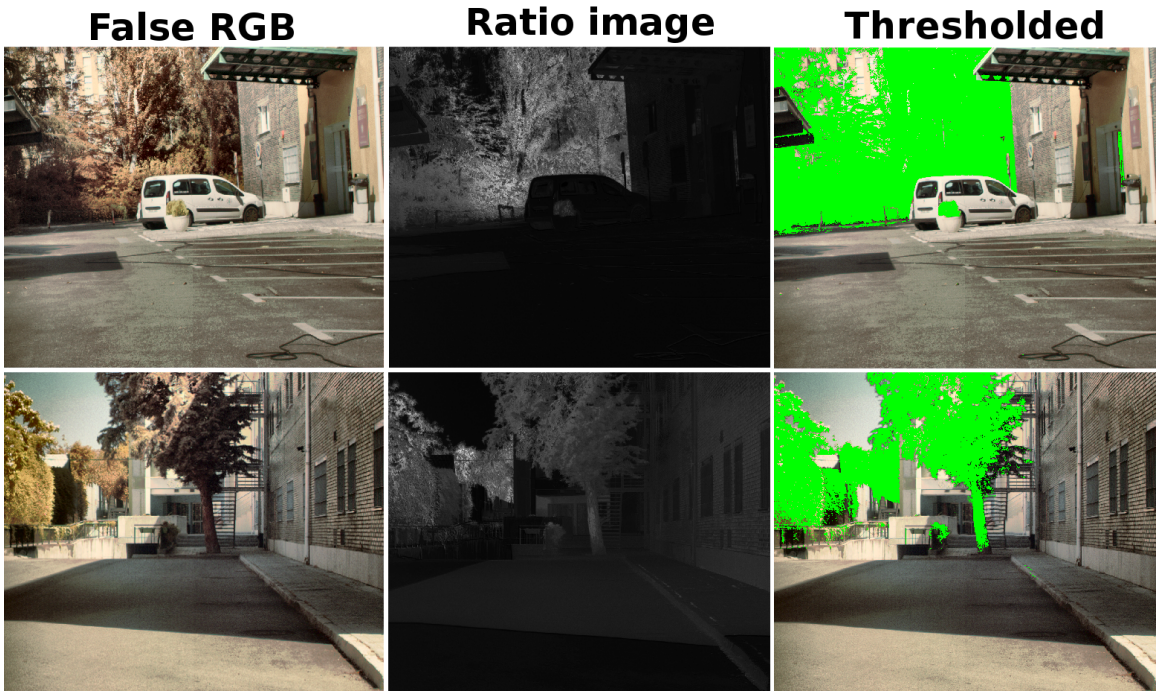

Figure 5-34: From left to right: false *RGB* image, vegetation ration image (computed as show in equation 5.8), and vegetation-detected image after thresholding the ratio image with a threshold value of 50 DC. Top row: scene number 4. Bottom row: scene number 6.

Finally, in image 5-35, we see an example of the *DoLP* maps (computed as shown in equation 5.6), for the different spectral bands of scene number 1. These *DoLP* maps are rendered in false color in order to see better the differences in polarization information from different regions of the scene. We can observe how some materials like glass have homogeneously higher *DoLP* values in almost all spectral bands (but the one centered at 950 nm). Others like concrete are more heterogeneous though still having high *DoLP* values. On the other hand, building walls have rather low *DoLP* values for all spectral bands captured.

5.6.4 Conclusions

We have introduced a complete image processing pipeline from image capture to object segmentation and classification for multispectral *HDR* polarimetric images of indoors scenes. We based our work in a previous study [179] done with an *LCTF*-based imaging system, and introduced a remarkable number of novel features.

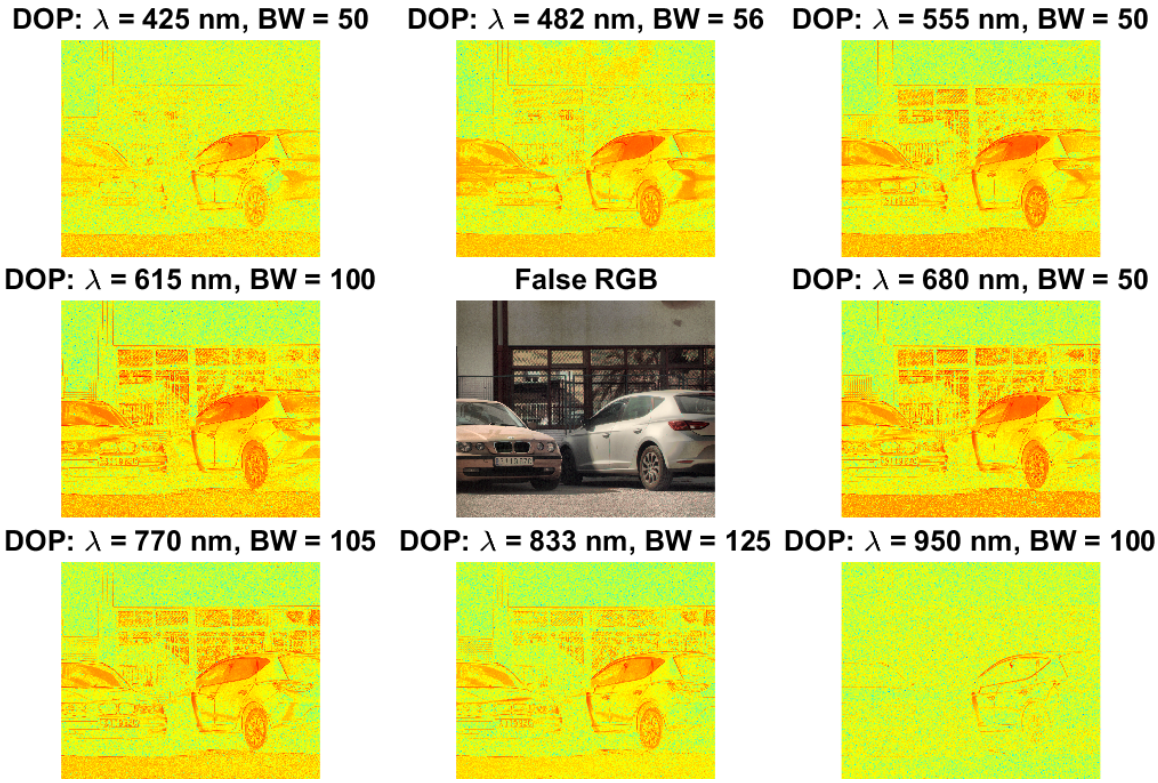

Figure 5-35: False color rendered *DoLP* maps computed (as shown in equation 5.6) for each spectral band captured for scene number 1.

Our capturing device was successfully calibrated and its *CRF* determined. We used this function to build four *HDR* images at different polarization angles for each spectral band, which were afterwards correctly registered after range compression, normalization and contrast enhancement. Our method succeeded in registering the already built *HDR* images where common methods proposed for registering *LDR* images failed.

We have also introduced new segmentation and classification approaches using as building blocks existing techniques which are combined aiming to simplify the computational load of previous approaches to this problem for *MSHDRPol* image cubes. We got a relative improvement of 31.73% in segmentation accuracy and 35.41% in classification accuracy compared with previous methods.

We have also implemented our proposed capturing method to a filter wheel camera in order to capture *MSHDRPolVISNIR* images of outdoors scenes. We calibrated this camera radiometrically and also performed filter selection from a big set of commercially available filters to purchase. We will make these *MSHDRPolVISNIR* images free for public use after the defense of this thesis.

The work and results presented in this chapter have been submitted to Pattern Recognition Journal, and we are currently waiting for the journal's response [124].

5.6.5 Future work

As future work, we will try to improve the segmentation procedure by assessing the clustering made after each mean-shift, so we can automatically perform a fine tune of the threshold for the final region merging step. How to tune this parameter automatically to find its ideal value is a matter of future study. We could use clustering quality metrics to make the full segmentation process completely automatic.

Regarding the classification, we believe that including also infrared information by extending the spectral range we are capturing, could help us including more candidate classes. However the *LCTF* model we used in this study could not be used above 720 nm. So we will use the filter wheel camera for indoor scenes as well.

In you can’t explain it simply, you don’t understand it well.

Albert Einstein.

6

Conclusions and future lines of research

Content:

Chapter 6

Conclusions and future lines of research

6.1 Conclusions

In this thesis, multiple advanced imaging technologies have been studied. Spectral, *HDR*, polarimetric and *NIR* imaging have been researched and merged together into a single imaging system.

New system architectures and methods have been proposed for the different fields of digital imaging investigated. Regarding multispectral image capture, two different systems have been proposed based on a novel technology still under development. Tunable *TFD*-based systems coupled with a *CFA* whose filter transmittances have been selected from a real available color filters commercial database. Simulation experiments have proved that the proposed systems outperform others with similar characteristics using other design strategies.

We proposed to use different architectures depending on the requirements of the application we are working in. We proposed a two shots strategy if the application allows us to afford a slightly longer capturing time. Otherwise, we also propose a single shot approach with very good colorimetric and spectral performance, for those applications that need faster captures. We found out that the proposed systems perform both colorimetrically and spectrally better or as well as other systems, while offering an easier and more elegant solution to the problem of spectral imaging. We conclude that *TFD* sensors are a promising

technology for the future of multispectral imaging, offering systems that are portable, real-time, versatile and low cost, as soon as they can be implemented as part of a real capturing system.

Concerning *HDR* imaging, we have presented a new method for estimating the exposure times needed to recover the *HDR* radiance map from a scene via multiple exposures. After looking for similar methods in the literature aiming for a completely automatic *HDR* capture, we compared the performance of our proposed method with the only method to our knowledge with most similar design strategy and features (Barakat method [7]).

Our proposed method is adaptive because it finds a bracketing set adapted to any *HDR* scene content. It is universal because it works for any imaging system as long as the exposure time can be controlled and the system is able to make simple computations like a cumulative histogram. It is full range because it aims to expose correctly at least in one *LDR* shot the whole radiance range of the scene. It is tunable, since we can decide if we prefer to find a minimum bracketing set, at the cost of lower *SNR*, or increase the *SNR* by increasing the number of shots and capturing time as well. Then, if we set it up for finding the smallest bracketing set, we can say the method is minimal. It is blind because no information about the content of the scene needs to be known a-priori. It is on-line as it works during the capture, and every single shot taken is used to compose the *HDR* radiance map. It is flexible because we can also control what is the percentage of total pixel population we can assume underexposed or saturated. Hence we can focus only in our region of interest (not only spatially but also in terms of dynamic range).

We have applied our *AEE* proposed method in monochrome and *RGB* cameras, as well as in multispectral imaging systems like a scientific camera attached to a *LCTF*, and a monochrome camera attached to a filter wheel. In all cases, the method worked well and yielded full range *HDR* images which are either monochrome, color or multispectral depending on the system used.

In the case of *RGB* images, we studied the *SNR* performance in 30 captured outdoors scenes where partially cloudy skies were present in order to increase the dynamic range of the capture. We compared it with that of Bakarat's method and found that *AEE* method can find bracketing sets which are shorter than those found by Barakat's method (*BAR*), yet keeping higher *SNR* levels in the *HDR* radiance map reconstructed from the multiple exposures.

The proposed method brings a solution for the blind acquisition of *HDR* images using multiple exposures, which can be used in any *HDR* imaging context: machine vision, sky imaging, daylight illuminated scenes, *HDR* photography, surveillance, remote sensing, medical imaging, etc.

As a combination of multispectral and *HDR* imaging system, together with polarimetric imaging, we have proposed a complete image processing pipeline from image capture to object segmentation and classification for *MSHDRPol* images of indoors scenes, with a remarkable number of novel features involved.

We calibrated our device and used it to build four *HDR* images at different polarization angles for each spectral band, which were afterwards correctly registered after range compression, normalization and contrast enhancement. Our method succeeded in registering the already built *HDR* images, contrary to common methods proposed for registering *LDR* images failed.

We have also introduced new segmentation and classification approaches using a combination of existing techniques. We compared our work with that of [179] and got improved performance with less computationally demanding methods. We obtained a relative improvement of 31.73% in segmentation accuracy and 35.41% in classification accuracy.

We have also used a filter wheel camera to capture *MSHDRPolVISNIR* images of outdoors scenes. We studied which filters to use and calibrated the camera as well. The processed *MSHDRPolVISNIR* images will be publicly available after the defense of this thesis.

6.2 Future lines of research

We still believe there is a lot of future work to be done. Many lines of research are still open to improve and advance for the development of cheaper, faster and more efficient *MSHDRPol* imaging systems in the visible and *NIR* ranges of the spectrum.

As long as big size *TFD* sensors become available in cameras, the next step would be to implement a real *MSHDRPol* imaging system featuring a *TFD* sensor and the *CFA* we have designed for it, as well as a rotating polarizing filter. We would implement the *AEE* method for the capturing process, and also research the proposed methods for processing the images and studying the segmentation and classification of their content, not only in

indoor scenes, but also in common outdoors urban scenes.

In the *HDR* field, we want to test the possibility of including other parameters in a new version of the *AEE* algorithm. Besides exposure time, we want to consider adding aperture setting as well as *ISO* setting to try to accelerate the automatic adaptive *HDR* capture in very dim lightning conditions, yet keeping satisfactory noise and focus levels using current high end *CMOS* sensors.

Also ghost removal techniques as well as the possibility to record *HDR* video using *AEE* method will be studied.

Simplifying the capture of indoors scenes by reducing the number of bands and using alternative and cheaper capture devices is also among the future lines of research we are contemplating.

Regarding both segmentation and classification, we will study a machine learning as well as a deep learning approach ([20, 13, 105]). Methods like support vector machines, kernels, or deep neural networks can work well for the big amount of image data generated in each captured scene. After capturing a big set of urban scenes, and performing a manual benchmark segmentation and classification which will be used as ground truth, we can apply local and global methods for detecting, segmenting and classifying singular objects present in common urban scenes like: buildings, plants, vehicles, people, urban furniture, traffic signs, ground, sky, etc. We believe that polarimetric as well as *NIR* information can be crucial for this task.

Some of these steps have already been implemented. We have captured a new set of 44 *MSHDRVISNIR* images of outdoors scenes. We have processed them and manually created a benchmark for segmentation and classification of objects present in them. We will use this data base for the study of deep learning methods for the classification of objects in urban scenes. For this purpose, multispectral invariant image representations ([135, 92]) will also be considered.

Nothing in life is to be feared, it is only to be understood.

Marie Curie.

Appendices

Content:

Appendix A

Publications

Combining transverse field detectors and color filter arrays to improve multispectral imaging systems

Miguel A. Martínez,[1,*] Eva M. Valero,[1] Javier Hernández-Andrés,[1] Javier Romero,[1] and Giacomo Langfelder[2]

[1]Color Imaging Laboratory, Department of Optics, Faculty of Sciences, University of Granada, Spain

[2]Politechnic University of Milano, Italy

*Corresponding author: m.martinez.domingo@gmx.com

Received 3 December 2013; revised 19 March 2014; accepted 19 March 2014; posted 19 March 2014 (Doc. ID 201775); published 16 April 2014

This work focuses on the improvement of a multispectral imaging sensor based on transverse field detectors (TFDs). We aimed to achieve a higher color and spectral accuracy in the estimation of spectral reflectances from sensor responses. Such an improvement was done by combining these recently developed silicon-based sensors with color filter arrays (CFAs). Consequently, we sacrificed the filter-less full spatial resolution property of TFDs to narrow down the spectrally broad sensitivities of these sensors. We designed and performed several experiments to test the influence of different design features on the estimation quality (type of sensor, tunability, interleaved polarization, use of CFAs, type of CFAs, number of shots), some of which are exclusive to TFDs. We compared systems that use a TFD with systems that use normal monochrome sensors, both combined with multispectral CFAs as well as common RGB filters present in commercial digital color cameras. Results showed that a system that combines TFDs and CFAs performs better than systems with the same type of multispectral CFA and other sensors, or even the same TFDs combined with different kinds of filters used in common imaging systems. We propose CFA+TFD-based systems with one or two shots, depending on the possibility of using longer capturing times or not. Improved TFD systems thus emerge as an interesting possibility for multispectral acquisition, which overcomes the limited accuracy found in previous studies. © 2014 Optical Society of America

OCIS codes: (110.4234) Multispectral and hyperspectral imaging; (110.0110) Imaging systems; (330.1690) Color; (330.1730) Colorimetry.

http://dx.doi.org/10.1364/AO.53.000C14

1. Introduction

Spectral science has been receiving gradually increased attention in the last few decades. New applications come up on a daily basis, offering an interesting range of possibilities [1–3]. Industrial, medical, military, remote sensing, and many more fields of research, focus nowadays on the use of spectral sciences for their new technological advances. One of the outstanding disciplines in the spectral science field is spectral imaging. It is mainly concerned with the problem of obtaining pixel-wise spectral information in an image. Many types of systems are found in the literature for this purpose. We could divide the spectral imaging devices into two main groups: measuring devices and estimating devices. The first group is made up mainly of hyperspectral devices, which acquire full spectral information during the capturing process. In this group, we can find systems formed by monochrome cameras attached to tunable filters [4], such as acousto-optic tunable filters (AOTFs) [5,6] or liquid-crystal tunable filters (LCTFs) [7]. We also find Bragg-grating-based systems [8], or snapshot systems [9] used in spectral microscopy [10]. The clear disadvantage of this group

1559-128X/14/130C14-11$15.00/0
© 2014 Optical Society of America

of systems is either their bulky size (usually coupled to long capture times, since they require as many shots as spectral bands), or else a reduced spatial resolution, which is clearly not enough for most applications. In the second group, we find numerous systems that can acquire three or more spectral bands (commonly termed multispectral imaging devices), which are used for estimation. These systems usually have a lower number of bands than those in the first group and their spectral sensitivities are not so spectrally narrow-banded. Such devices acquire the sensor responses and then require different techniques to map these responses onto spectra [11]. The quality of the estimated spectra depends critically on the number of channels, their spectral shape, and the algorithms used to obtain the mapping from sensor responses to full spectra. Within estimation capturing systems, we find as representative instances: monochrome cameras with a color filter wheel with few filters [12], multi-sensor cameras [13–15], color filter array (CFA)-based cameras [16,17], and hybrid systems [18]. Some of these systems suffer from relatively long capture times and need the mechanical movement of parts that might cause misalignment problems (e.g., filter wheel devices). Others have expensive hardware, including many sensors and light splitting optical components, which suffer from low-light throughput, increased exposure times, limited portability, etc. Given all these problems, it is clear that the question of how to design a capturing system that is portable, nonexpensive, easy to use, and capable of acquiring full spectral information, pixel-wise, in real-time is still far from solved. In this study, we present a proposal for a multispectral imaging system based on a recently developed technology as a potential answer to this question. The transverse field detector (TFD) technology definitely has the advantage of allowing for the capture of multiple channels at a pixel with full spatial resolution. However, the TFD by itself does not provide enough accuracy in the estimated spectra [19–21]. In [22] we pointed out the hypothesis that one of the main causes of this decrease in accuracy is the width of the spectral response functions typical of TFDs. This width is such that we deem that conventional sharpening techniques are not enough to solve the problem [23]. Instead, in this paper, we propose combining the TFD sensors with a specifically designed CFA [17] to obtain the required accuracy in the spectral estimation. The filters narrow down the spectral sensitivities of TFDs and improve their color and spectral estimation performance, trading off their full spatial resolution property. We carried out different experiments to check if the improvement of performance is due to combining TFDs and CFAs. We also checked whether our method for selecting the filters to be used was effective for selecting the combinations of filter time sensitivities (what we call channels).

Using simulated sensor responses, we compared this system with several others. We present convincing evidence in Section 5 of the advantages of such a design, even allowing for the inevitable need to sacrifice the full spatial resolution feature of conventional TFD-based systems over the ones presented in previous works [21,22]. We offer some alternative designs depending on the requisites that each individual application may require, regarding capturing times, colorimetric, and spectral accuracy.

The remainder of this paper is organized as follows: in Section 2 we present the TFD sensors. Section 3 describes all the systems simulated and compared in this study. In the methods section (4), we clarify the filter selection procedure, the computations of sensor responses, the estimation procedure and the experiments performed. In Section 5 we provide the results found in the different experiments. Finally, Section 6 draws relevant conclusions from these results.

2. TFD Sensors

In silicon sensors, light is absorbed following an exponential intensity decrease, described by the Beer–Lambert law. Electron-hole pairs are photogenerated with different depth profiles, according to the material and the radiation energy (or wavelength). In a typical pn-junction-based CMOS pixel, photogenerated electrons are collected by the same well, regardless of the generation depth [24]. Thus, for each pixel, only one value is read in a single exposure. In contrast to these normal sensors, others, such as filter-less sensors like Foveon X3 [25] or TFDs [19], account for the fact that photons are able to penetrate deeper into the silicon as their wavelength increases. Therefore, electrons generated at different depths can be separated by junctions built at different depths (as in the case of Foveon sensors), or by suitable transverse electric fields (as in the case of TFDs). Understanding this principle makes it possible to set the collection of generated electrons at different depths within the silicon, so that in a single pixel, and in a single exposure, we can retrieve information from three (Foveon) and up to five (TFDs) spectral bands or channels. Usually, in a 3-channel configuration, the responsivities have peaks in the short, middle, and long wavelength ranges, similar to standard RGB sensors; however, they are spectrally wider since there is no sharpening filter in front of the sensor. Furthermore, TFDs also offer the capability of modifying the collection depth of photons by modifying the applied transverse electric field via a tunable biasing voltage. Thus, using a given biasing voltage, the sensor presents a set of three sensitivities with certain spectral shape and maxima positions, and setting a different value for this voltage, the spectral shapes and positions of the sensitivities change. We can see this in system number 8 (Fig. 2). In this figure, we present eight different sets of RGB sensitivities corresponding to eight different values of biasing voltage. The difference in the spectral shapes and peaks is quite clear. The possibilities offered by this technology are many,

and few studies have so far been done to examine the capability of TFDs for multispectral imaging [21,22]. As pointed out before, TFD sensitivities are too spectrally broad to provide satisfactory accuracy in the estimated spectra within the context of 3-channel noisy systems. This is a drawback when trying to use them for spectral imaging [26], because even when adding more shots corresponding to different biasing voltages (up to 8, with a total of 24 channels), the results were equaled or outperformed by a simple RGB system based on CFAs [22], where spectral sensitivities were much narrower in comparison.

In previous studies, many authors have reported good estimation accuracies obtained with both simulated and real camera responses from narrow spectral responsivities [27] and, more recently, some have proposed the possibility of increasing this accuracy by using tunability of the channels' peak response positions and widths with simulated camera responses of Gaussian-shaped channels [28]. Nevertheless, narrow-band tunable responses are not so far achievable in practice without the use of external filters. Nonetheless, TFDs have the advantage of not needing any demosaicking process to get the full spatial information for every spectral band. This prompted an idea. Does adding filters embedded in the TFD sensor matrix (and so combining CFAs with TFDs) boost the estimation accuracy of the system? We present in the next sections a set of experiments comparing the color and spectral accuracy of different capturing system designs, to provide clear evidence on these points. We have assessed the color and spectral reconstruction accuracy of a set of reflectance samples with no spatial information present. Therefore, we did not study the effect of varying the spatial arrangement of the different transmittances present in the CFA pattern. This factor has been shown to influence image quality [29]. Other studies have been made using similar multispectral CFAs with up to seven channels, providing methods to design their spatial arrangement, as well as the demosaicking techniques used to retrieve multispectral information with good image quality [30,31].

TFDs offer the possibility of either being full polarized (all pixels with the same state of polarization), or bi-polarized (the sensor is biased in an interleaved configuration [19] with different bias voltage). Hence, some pixels can have an RGB set of sensitivities, and others can have different ones being biased by a different voltage. We also checked the effect of using this property. We compared a system taking advantage of this property to reduce the number of shots against a system taking more shots with different full-polarized biasing conditions. This second system had a larger number of channels, but also needed a longer capturing time.

As pointed out before, TFDs offer the capability of retrieving up to five channels out of a pixel in a single shot [32], but two of them would be in the near-infrared range (NIR) of the spectrum. Getting NIR information is an advantage in many applications. However, as a starting point for our study, we have decided to use the three channels within the visible range. This matches a general-purpose application. We have left the remaining two channels for future studies that target a more specific field of research, for which the NIR content might be critical. Including data beyond the limits of the visible range of the spectrum would make the simulations closer to the real behavior of the imaging systems in those channels close to the red and blue limits of the visible range. However, no spectral data was available for the set of spectral reflectances used (explained in Subsection 4.A) out of the visible range, so we restricted our camera response simulation to the limits of the visible range. Moreover, common imaging systems use a hot mirror to get rid of crosstalk effects between visible and NIR ranges of the spectrum [33], and most common optics components block naturally the UV components of radiation, to a certain extent.

3. Systems Simulated for Study and Comparison

TFDs are still under development and only prototype sensors exist so far. Nonetheless, their physical properties are very well characterized. The simulations of sensor response calculations are very realistic with the physical model provided by the developers (explained in Subsection 4.B). All systems studied in this work are silicon-based systems. Despite the architecture of each system in particular, the light acquisition process of them all can be realistically described with the CMOS capture model used for simulating camera responses. Therefore, sensor responses for every system studied in this work have been calculated using the same model for signal and noise generation. We can thus be sure that a higher or lower noise level is not the reason for better or worse performance. The average signal-to-noise ratio (SNR) of sensor responses is shown for each system in the last column of Table 1. As it will be explained in Subsection 4.B, the noise model accounts for exposure time and quantum efficiency. Therefore, even though the values of exposure time were different for different systems, this difference was taken into account when computing the noise. The calculation of SNR is also explained in Subsection 4.B. All systems result in similar SNR values. To demonstrate the performance of the designed system, we have created four experiments (explained in Subsection 4.D), in which its features are tested across other imaging systems. In Fig. 1, we can see a representation of the 11 different systems simulated in this study and used to complete the four different experiments.

In each system, there is a filter layer and a sensing layer. The number of different transmittances in the filter layer determines the spatial resolution. A CFA with 3 filters provides 1/3 of spatial resolution and, with 6 filters, 1/6. If the TFDs are bi-polarized (as explained in Section 2) and use no filters, then the spatial resolution reduces to 1/2. The characteristics of each system are explained and summarized in

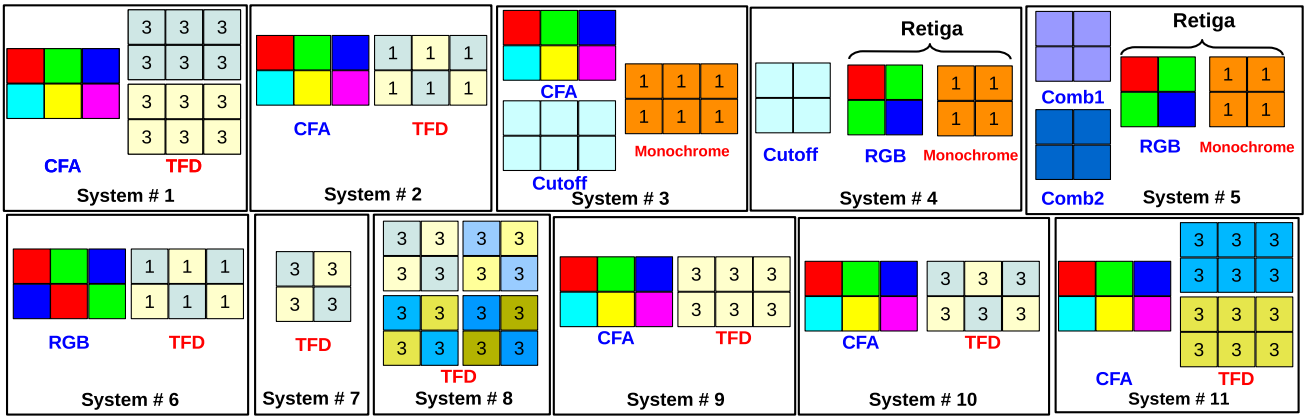

Fig. 1. Schemes for the 11 system configurations studied. Filter layer with blue caption, sensing layer with red caption. The number in each pixel determines the number of channels retrieved in one shot out of it.

Table 1. Figure 2 shows the spectral sensitivities of all systems, which are described in detail in the following paragraphs.

1. This is the proposed system. The filter layer is a CFA made up of six bandpass optical filters selected from a real database (see Subsection 4.A). The sensing layer is a TFD sensor full-polarized where two different biasing conditions were used in two shots. Three channels are retrieved per pixel; thus, without the need of any moving mechanical optical component, we get information from 36 channels in two shots.

2. Same CFA used as in system 1. The sensing layer, however, was a bi-polarized TFD from which we only take the information from one channel per pixel. The channel that was operative for each pixel was selected as explained in Subsection 4.A. This system is similar to the proposed one, but it has been designed to compare fairly its performance with other systems under the same conditions, in terms of number of channels.

3. The sensing layer is a normal monochrome silicon sensor like the one from a commercial scientific camera model Retiga SRV (QImaging Corp., Canada), equipped with a NIR cut-off filter. In the filter layer we set a CFA that has been optimized for the monochrome sensor, using the same technique as in our proposed system (see Subsection 4.A). This system has been designed to test if including a TFD sensor with its tunability property in a CFA-based system helps to improve its performance.

4. In the sensing layer, we used the spectral responsivities of the RGB scientific camera Retiga 1300C (QImaging Corp., Canada). In the first shot, we set no filter in the filter layer, and in the second shot we added an IR–UV cut-off filter (C_{off}) in front of the lens, which sharpened down the sensitivities in the extremes of the spectrum. This system has been designed to test the performance of TFDs plus CFAs against other types of systems and was proven to work well in a previous work [22].

5. In the sensing layer, we used the RGB camera again. This is also a 2-shot system. In each shot, a different ideally custom-made comb-shaped optical filter (Comb_1 and Comb_2) was placed in the filter layer. These filters divided by half the 3 sensitivities into 6. This idea of splitting the spectral sensitivities of single channels into two by using a comb-shaped filter has been used by other authors in the literature, as well [13,34,35].

6. The sensing layer used here was the same as in system 2, but the filter layer used was composed of an RGB Bayer filter like the one from the scientific camera Retiga 1300C (QImaging Corp., Canada)

Table 1. System Characteristics

System #	System Name	Filter Layer	Sensing Layer	# Shots	# Channels	Spatial Resolution	Average SNR (dB)
1	CFA+2TFD	CFA	TFD full	2	36	1/6	40.5
2	CFA+1TFDbi1	CFA	TFD bi	1	6	1/6	42.1
3	CFA+Monochrome	CFA	Silicon	1	6	1/6	42.5
4	Cut+Re	RGB RGB + C_{off}	Silicon	2	6	1/3	42.3
5	Comb+Re	RGB + Comb_1 RGB + Comb_2	Silicon	2	6	1/3	44.6
6	RGB+TFDbi	RGB	TFD bi	1	6	1/6	44.4
7	TFDbi	Empty	TFD bi	1	6	1/2	42.8
8	TFDbi4	Empty	TFD bi	4	24	1/2	47.7
9	CFA+1TFDfull	CFA	TFD full	1	18	1/6	47.4
10	CFA+1TFDbi3	CFA	TFD bi	1	18	1/6	40.9
11	CFA+2TFDalt	CFA	TFD full	2	36	1/6	45.1

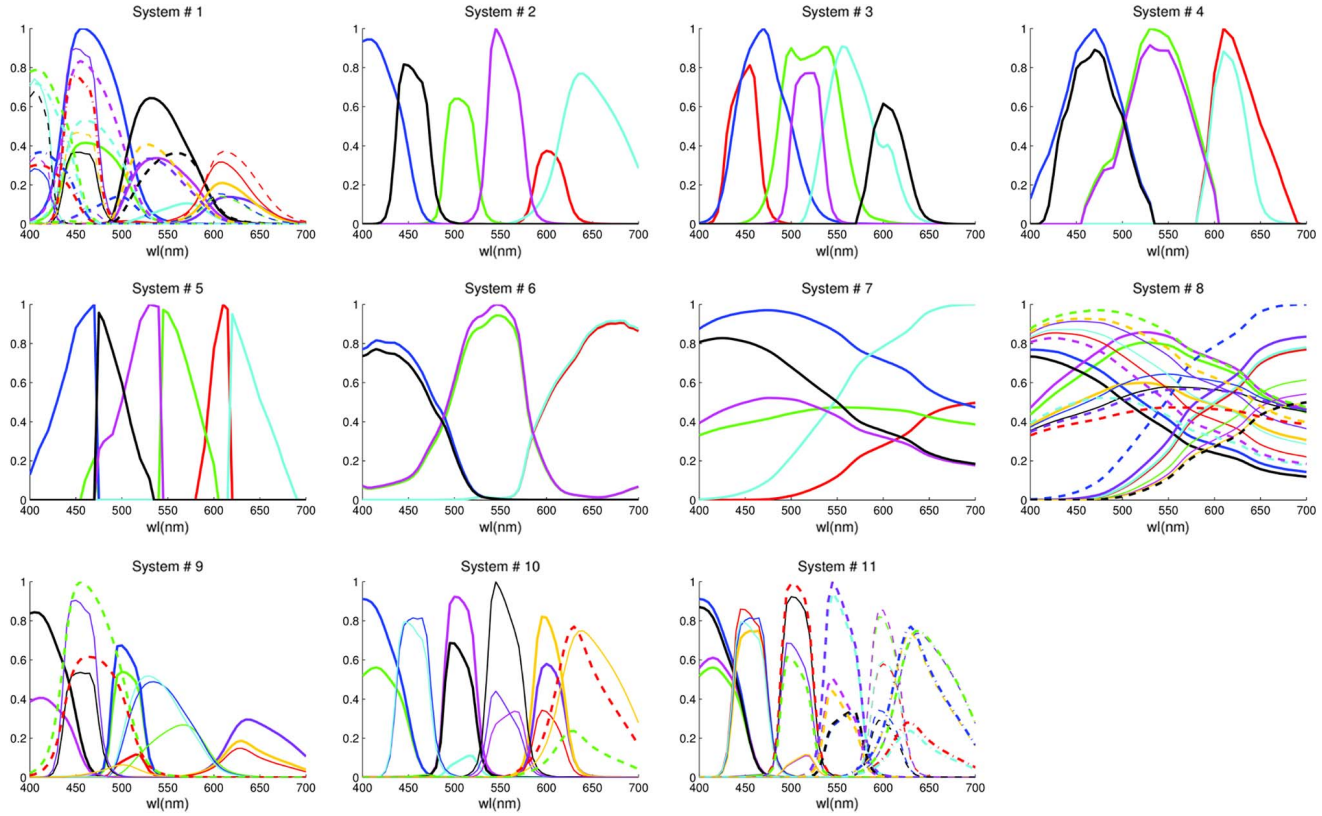

Fig. 2. Normalized spectral sensitivities of all systems versus wavelength in nanometers. Normalization was done just for displaying.

used in this study as representative of a standard RGB CCD camera. This system has been designed to test the effect of the selected CFA versus other kinds of narrow bandpass filters present in common color imaging systems.

7. The filter layer was left empty in this system. The sensing layer used was a bi-polarized TFD sensor. Only one shot was taken. This system and the next one were designed to test if adding a CFA to a TFD sensor improves its performance.

8. The same as the previous configuration but we took 4 shots with 4 different bi-polarization conditions. We wanted to push the number of channels to the extreme using all the biasing conditions that the developers provided.

9. The filter layer was the same CFA used in system number 1, but the sensing layer was a single shot full-polarized TFD sensor. This system has been designed to test the effect of bi-polarization property of the TFD sensor versus full-polarization state.

10. The filter layer was the same CFA used in system number 1. The sensing layer in this case was a one-shot bi-polarized TFD. This system is also proposed as an alternative to system number 1 if the application requires shorter capturing times (since it only uses one shot instead of two). In addition, this system has been compared with system number 9 to check the effect of using the bi-polarization property of TFD sensors.

11. This system has been designed exactly the same as the proposed system. However, its channels were randomly selected from the available ones instead of using voting principal feature analysis (VPFA) [36,37], as explained in Subsection 4.A. Ten reasonable combinations were simulated, in which the channels were selected to cover the whole visible range with some overlap. We selected the one with best results. This system was simulated to check whether selecting the channels using VPFA effectively increases the performance of our system or not.

We have specifically selected these configurations to better isolate the effect of the feature that we wanted to test in our proposed system. To sum up: system numbers 2, 3, 4, and 5 were designed to prove, under the same conditions in terms of number of channels, that using a TFD sensor improves the accuracy of the results over common monochrome sensor systems. System numbers 2, 6, 7, and 8 were designed and compared to test whether combining the CFA with the TFD also helps to improve its performance. System numbers 9 and 10 were designed to test if using the bi-polarization property of TFDs is better than not using it. System number 1 evolved from system number 10, pushing up the number of channels retrieved with only two biasing conditions adding one shot to the capturing process. System number 11 was designed to test the advantages of using VPFA to select the spectral channels.

4. Data, Methods, and Experiments

In this section, we explain the methods used to calculate the simulated sensor responses, select the set of sensors and filters, estimate the reflectances from

sensor responses, and evaluate the performance of each system.

A. Spectral Data, Sensors and Filters Selection

The illuminant used for the simulations was the CIE standard D65 illuminant, and the spectral reflectances were 1700 samples from the Natural Color System (NCS) [38,39], since they represent natural colors found in common scenes and are well-used for general-purpose color correction.

The TFD developers [40] provided a set of 8 different RGB sensitivities corresponding with 8 different biasing voltages (24 sensitivities in total). We can see them in Fig. 2 (system number 8). We used a set of 13 real filter transmittances from the Andover Corp. website [41]. These transmittances were selected so that they all covered the whole visible range of the spectrum, with certain overlap between them. We carefully chose different bandwidths and maxima positions. The commercial references of these 13 selected filters are: *400FS40, 400FS70, 450FS40, 450FS80, 500FS40, 500FS80, 550FS20, 550FS40, 550FS80, 600FS40, 600FS80, 650FS40*, and *650FS80*.

Combining all sensitivities with all filters, we obtained a total of 312 channels. Out of them we wanted to select six, corresponding to only two different biasing conditions (which are the limits of bi-polarization, as mentioned in Section 2) that could best recover the spectral information of the samples imaged in the visible range.

Basically, the VPFA method employed for this selection consisted of taking the sensor responses from all 312 channels of the 1700 NCS [39] color samples used, and perform principal component analysis on them. Then, we selected a reduced number of projections of the data onto the principal component vectors (half of them) and clustered them into 6 clusters (which was the number of channels we aimed to find) using the k-means clustering method [42]. We calculated the mean point (center) of each cluster, and then selected the filters corresponding to the vectors closest to each of the cluster centers. Due to the random initialization of k-means clustering, the clustering step is repeated over several iterations (50) and the selected vectors are voted. Finally, the most voted ones that correspond to only two biasing conditions are selected. Since each pixel of the TFD still gives us 3 channels, even though we only accounted for the 6 best performing ones (system number 2), we also studied the possibility of using the other two channels per pixel (system number 10), since results showed that there is a slight improvement when adding all of them. Furthermore, a two-shot system was also designed using the two full polarizations of the TFD resulting from VPFA and retrieving the information from all resulting channels (system number 1). After this process, the spectral sensitivities of the channels selected were the ones shown in Fig. 2 (upper left). As this figure shows, they span the whole visible range with some overlap. Overlap is important, so as not to leave a gap in the spectrum without retrieved information. However, too much overlap could cause bad performance as well, due to the consequent stronger effect of cross talk between channels. We also simulated other systems with the same characteristics, but selecting randomly the combination of filters and sensors to get a set of sensitivities that visually looks good (covering the spectrum, relatively narrow, and with some spectral overlap). The one with the best results was chosen (system number 11) and compared with the one resulting from the VPFA. The metrics used for evaluation of the performance were spectral (goodness-of-fit coefficient GFC [22,43], also known as complementary Pearson distance [44], and root mean square error, RMSE [22,45]), and colorimetric (ΔE^*_{00} also called CIEDE2000 [46]). If GFC > 0.999, then the estimation was considered quite good, and if GFC > 0.9999, then the estimation was almost an exact fit [47]. Regarding CIEDE2000 colorimetrics, differences of less than $1\Delta E^*_{00}$ units were considered acceptable.

B. Sensor Responses Simulation

As mentioned before, TFDs are still a technology under development, from which only prototype sensors exist and there is no implementation of a complete imaging system available yet. We were provided by the developers [40] with a physical model that well describes the opto-electronic behavior of such sensors. They are based on CMOS technology. The first step was to calculate the power spectral density ($\mathrm{PSD}_i(\lambda)$) of the color signal incident on a pixel, composed by the illuminant ($D65$) spectral photon flux per unit time and area ($\mathrm{SPF}_{D65}(\lambda)$), and the sample spectral reflectance of the ith color sample ($R_i(\lambda)$), as Eq. (1) shows:

$$\mathrm{PSD}_i(\lambda) = R_i(\lambda) \cdot \mathrm{SPF}_{D65}(\lambda). \tag{1}$$

Afterwards, we calculated the photocurrent I_{ik} and the output voltage V_{oki} generated in each channel f for each sample i, as Eqs. (2) and (3) show:

$$I_{ik} = q \cdot A_{\mathrm{light}} \cdot \int_\lambda \mathrm{PSD}_i(\lambda) \cdot Q_{ek}(\lambda)\mathrm{d}\lambda, \tag{2}$$

$$V_{oki} = I_{ik} \cdot \frac{T_{\mathrm{int}}}{C_f}, \tag{3}$$

where q is the elementary charge, A_{light} is the illuminated area on the TFD surface, $Q_{ek}(\lambda)$ is the spectral quantum efficiency of channel k, T_{int} is the exposure time, and C_f is the pixel feedback capacitance. After calculating the noiseless output voltage (V_{oki}) and before quantifying the signal into digital counts, the additive noise component (η_σ) is added to get the noisy output voltage ($V_{\eta oki}$). This noise was Gaussian and its total variance σ_{tot} was composed by kTC noise (σ_T), and dark current and shot

noise (σ_{DCS}). These calculations are shown in Eqs. (4)–(7):

$$\sigma_T = \sqrt{\frac{K \cdot T}{C_f}}, \tag{4}$$

$$\sigma_{\mathrm{DCS}} = \sqrt{\frac{q \cdot T_{\mathrm{int}}}{C_f^2} \cdot (I_{ik} + J_d \cdot A_{\mathrm{tot}})}, \tag{5}$$

$$\sigma_{\mathrm{tot}} = \sqrt{\sigma_T^2 + \sigma_{\mathrm{DCS}}^2}, \tag{6}$$

$$V_{\eta oki} = V_{oki} + \eta_\sigma, \tag{7}$$

where K is the Boltzmann constant, T is the absolute temperature (300°K = 26.85°C), J_d is the dark current density, and A_{tot} is the overall sensor area. As a last step, we quantify the noisy voltage signal to get the noisy sensor response in digital counts (ρ_{ki}) using B bits, as shown in Eq. (8):

$$\rho_{ki} = \mathrm{round}\left[\frac{V_{\eta oki \cdot (2^B - 1)}}{V_{dd}}\right], \tag{8}$$

where V_{dd} is the pixel maximum output range. The values of all TFD-specific parameters were extracted from Finite Elements Simulation using the Dessis software, as described in [19]. Typical biasing voltages of the TFD anodes range between 0.5 V and 8 V. The simulations of all systems compared in this study were done using the same model for sensor responses calculations. As it would happen in reality, including filters on top of the sensor reduces their quantum efficiency ($Q_{ek}(\lambda)$). Moreover, in some systems, the difference in magnitude between $Q_{ek}(\lambda)$ of different channels would cause low dynamic range problems. The exposure time selected for each system was the one exposing most of the samples within the sensor range between noise floor and saturation level. This means that in all cases, some samples were underexposed and some overexposed, but in all cases this is taken into account for computing the results, which reflect all these realistic limitations of imaging systems. Due to the fact that the noise model used accounts for different $Q_{ek}(\lambda)$ levels and exposure time values, the resulting levels of noise in all systems were similar, as mentioned in Section 3. The calculation of SNR was done by calculating the ratio in logarithmic units between the noiseless output voltage signal (V_{oki}) calculated in Eq. (3) over the total noise variance (σ_{tot}) calculated in Eq. (6).

C. Spectral Reflectance Estimation and Evaluation

The estimation method used was regularized inhomogeneous polynomial kernel regression [48,49], which was the one found to perform best in previous studies [22]. The model used for the estimation needs to set two free parameters, the degree of the polynomial (d) and the regularization term (λ_r), and these values need to be optimized. Then, a double 10-fold cross validation method was used, one nested inside the other. In the outer loop, we randomly separated the 1700 NCS samples into optimization and evaluation sets. The former was used to find the optimal values of the two parameters of the mathematical model, and the latter to evaluate the estimation of reflectances. In each iteration of the outer 10-fold, the optimization set was used in the inner 10-fold loop. A grid of parameters was built and every possible combination of both parameters was tested with the testing set corresponding to each loop of the inner 10-fold. For the parameter optimization, we used the CIEDE2000 color difference formula as a cost function (ΔE_{00}^*). Using a color difference formula for parameter optimization gives to colorimetric performance of the results a slight advantage over spectral performance, compared with using a spectral metric. The opposite happens using a spectral metric as a cost function. In this work, we aimed for a general color correction application via spectral estimation. Therefore, the metric we chose for our application domain was CIEDE2000. After the 10 iterations of the inner loop, the average of the parameters found was calculated, and those were used for evaluation in the outer loop. The standard deviation was also calculated to check if the distributions of the best parameters found were stable. Table 2 shows the means and standard deviations (in parenthesis) of both kernel parameters found in every system out of each of the 100 folds in total (10 times 10 folds). Since the regularization parameter was optimized, looking for the value that performed best in different orders of magnitude, the mean and standard deviation calculated for it are shown in logarithmic scale. After the 10 iterations of the outer loop, the average error metrics were calculated, and these were the values with which we compared the different systems. By doing things this way, we ensured that samples used for training would never be used for testing, which

Table 2. Means and Standard Deviations of Polynomial Degree (d) and Regularization Term (λ_r) Kernel Parameters Values [a]

System #	1	2	3	4	5	6	7	8	9	10	11
d	2.68	5.70	5.9	7.08	3.47	7.00	5.41	4.13	4.46	4.30	4.60
	(0.58)	(0.48)	(1.37)	(0.80)	(1.14)	(0.00)	(1.31)	(0.83)	(0.59)	(0.48)	(0.52)
$\log_{10}(\lambda_r)$	−2.85	−6.06	−4.29	−5.81	−4.15	−4.93	−8.19	−4.39	−3.17	−4.23	−4.43
	(0.43)	(0.13)	(1.06)	(0.87)	(1.94)	(0.22)	(1.19)	(0.64)	(0.68)	(0.42)	(0.50)

[a]The regularization term values are shown in logarithmic scale. Values of d are not integers because they are the computed mean.

could lead to over-fitting and, thus, to overestimating the quality of the recovered samples obtained from the camera responses of our system.

D. Experiments

In this study, four experiments were conducted to test the following four hypotheses:

1. The proposed system performs better than systems based on other types of sensors, such as monochrome sensors or multi-shot systems based on RGB scientific cameras, plus some filters.
2. The proposed system performs better than systems based on TFDs that do not use the designed CFA. Therefore, this CFA helps to improve the performance of TFDs.
3. Using the bi-polarization property of a TFD sensor, the performance can be improved without the need to increase the number of shots taken.
4. The proposed system performs better than a similar system in which the channels selection was not done by VPFA.

Hypothesis 1 demonstrates the capabilities of the proposed design. Hypotheses 2 and 3 try to go deeper into the reasons justifying the superiority of our TFD-based proposed system. Hypothesis 4 demonstrates that using VPFA for selecting candidate filters and channels is worthy. The results for these four experiments are shown in the following Section 5.

5. Results and Discussion

In this section, results from the four experiments are shown in tabular format in Table 3, and then commented on.

A. Experiment 1: Superiority of TFDs

This experiment was performed to check whether our proposed system yields better accuracy in spectral reflectance estimation than systems based on other types of sensors using the same kind of CFAs, and other different ones. We included the 2-shot system made up of a RGB Retiga camera and a cut-off filter, studied in previous work [50], and we also designed a new system, in which we divided each of the R, G, and B sensitivities of the Retiga camera into two halves (see Fig. 2, system 5) with theoretical custom-designed comb-shaped optical filters. The results are shown in Table 3 for system numbers 2, 3, 4, and 5.

We see that system number 2 performed better both for spectral and color metrics than system numbers 3 and 4. The system formed by a monochrome silicon sensor plus the CFA performed worst. This means that the good results obtained by system 2 were not only due to the CFA, but also that the tunability and bi-polarization property of TFD sensor helped to improve the performance. The systems formed by RGB Retiga camera plus filters yielded good results, especially the one with the two comb-shaped filters which worked best. This is in agreement with previous studies [22,50]. This comparison was made in the same conditions as far as the number of channels is concerned. However, when we simulated system number 2, we were neglecting the information from 2 channels from each pixel. This meant that we were not using 12 channels, which we achieved anyway in the capturing process. If we add this available information to the recovery (system number 10), then the results improved, and were closer to system number 5 colorimetrically and better spectrally, and we would not need to add a second shot. Besides, there is no need to mechanically move any filter in and out. Moreover, we went further and designed a two-shot fully-polarized system, and we retrieved the information from all channels under each pixel, increasing the number of channels to up to 36 in just two shots. This was the strategy of system number 1. In this case, we used two shots, which were sequentially captured just by switching the biasing voltage of the TFD. These results outperformed the ones from system 5. Practice systems 1 and 5 have the same number of shots. For system number 5, the capturing process cannot be real-time since the filters need to be switching all the time. In the case of the proposed system, even though it is also a 2-shot system, the switching of the biasing voltage occurs so fast (electronically controlled) that it can be considered as a snapshot system in practice, as long as the lighting conditions in the scene being imaged allow short exposure times. We thus propose system number 10 as a good and convenient option to be used in applications requiring fast capturing times. If we could afford longer capturing times, then system number 1 is our proposal to enhance performance by increasing the number of shots.

B. Experiment 2: Superiority of a CFA-Based System

In this experiment, we wanted to assess whether including the CFA in the TFD-based system is the key factor leading to the superior performance of

Table 3. Results for All Systems Studied (Mean and STD)

System #	1	2	3	4	5	6	7	8	9	10	11
ΔE^*_{00}	**0.23** (0.16)	0.51 (0.35)	1.26 (1.3)	0.66 (0.47)	0.27 (0.18)	0.96 (0.73)	3.15 (2.27)	1.89 (1.32)	0.41 (0.34)	0.38 (0.26)	0.35 (0.24)
GFC	**0.9997** (0.0028)	0.9987 (0.0043)	0.9992 (0.0022)	0.9982 (0.0030)	0.9992 (0.0030)	0.9986 (0.0079)	0.9976 (0.0038)	0.9982 (0.0033)	0.9991 (0.0010)	0.9996 (0.0012)	0.9994 (0.0021)
RMSE	**0.0064** (0.0063)	0.0101 (0.0087)	0.0094 (0.0082)	0.0103 (0.062)	0.0083 (0.0099)	0.0113 (0.034)	0.0169 (0.0092)	0.0143 (0.008)	0.0082 (0.004)	0.0064 (0.0043)	0.0067 (0.0053)

the system shown in the first experiment. For this, we simulated different systems using TFD sensors. One of them used no filters in front (system number 7). The other used the filters present in a common RGB scientific color camera (system number 6). This was done as we had found out that the RGB Retiga camera was working very well and that these filters drastically narrowed down raw sensitivities. The results of this experiment are shown in Table 3 for system numbers 2, 6, and 7. We see that the system we propose performs best for both spectral and color metrics. Including other kinds of filters, such as RGB color filters (system number 6), yielded reasonable results, but still far from the ones reached by system number 2. The system using only two TFD polarizations, such as the proposed one, but with no CFA (system number 7), yielded quite bad results, in comparison. Even pushing this strategy to the limit of taking many shots with different polarizations (up to 4 shots with bi-polarized TFD, as in system number 8), the performance was far from our proposed system. Therefore, it did not outperform a system with only one shot, two TFD polarizations, and a CFA. If the difference is clear when the experiment is performed for systems with the same number of channels, then the difference is still more evident if we use the system number 10 or 1, the advantages of which were pointed out in Subsection 5.A. Therefore, we can conclude that adding in a TFD-based system and a CFA that has been calculated to be the best choice out of a set of available color filters, helps improve the performance for spectral reflectance estimation beyond systems without CFA, or with other filters that are not specially selected.

C. Experiment 3: Bi-Polarization versus Full-Polarization

In this experiment, we wanted to assess whether using the capability of bi-polarizing the TFD sensor helps with its performance. The easiest way to test this was by comparing system numbers 9 and 10. The results are shown in Table 3. We see how using two polarization states of the TFD improved the system performance slightly for both color and spectral metrics. Therefore, it is worth adding a second biasing voltage since, in practice, the acquisition process is the same. Apparently, the use of sensitivities from different biasing voltages perform better than the use of all sensitivities from the same biasing condition.

D. Experiment 4: VPFA Performance

This last experiment was carried out to check whether it makes sense to invest time in selecting the filters using VPFA, or to simply choose any combination of them randomly to cover the visible range of the spectrum and having some little overlap between bands, to yield better results. Out of ten random combinations, we selected the one that produced the best results (system 11) and compared it with our proposed system (number 1). Results show that the VPFA procedure improves the filter selection process. As a final summary of results, we present the quality indexes for all systems compared in this study in Fig. 3.

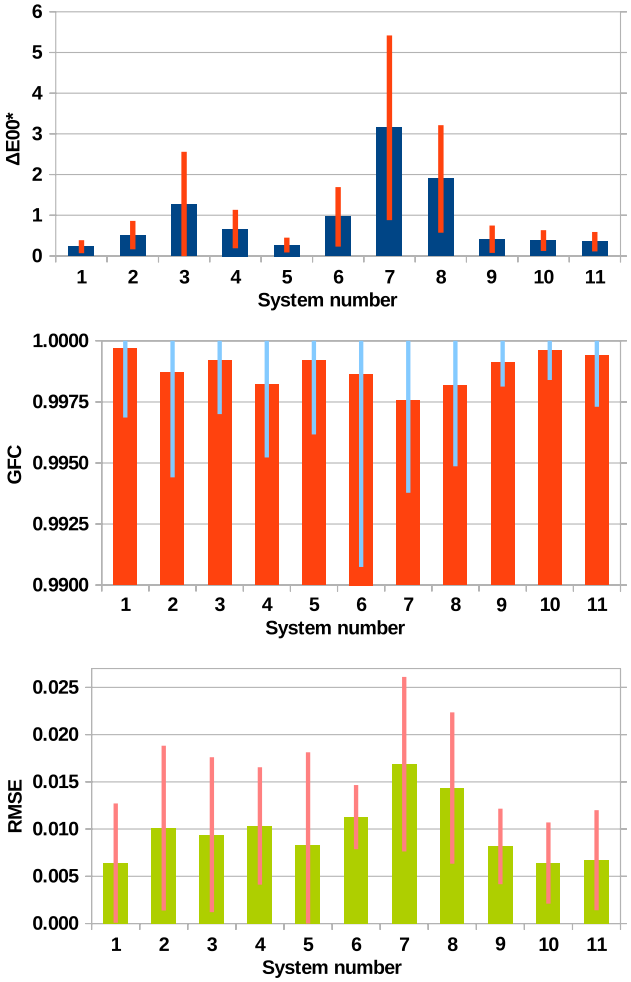

Fig. 3. Color and spectral error metrics for all systems. The bars show the mean value and the lines the standard deviation centered on the mean.

6. Conclusions

Two different approaches for a multispectral imaging system have been proposed, based on a novel technology still under development. TFD-based systems, which exploit the tunability and bi-polarization properties of this new type of sensor, are coupled with a CFA whose filter transmittances have been selected via VPFA method out of a set of real available color filters from a commercial database. Four simulation experiments have been conducted to demonstrate whether the proposed systems outperform others with similar characteristics, but using other design strategies. We aimed to assess whether the combination TFD plus CFA works better than any of them separately. We also studied whether we could use different architectures, depending on the requirements of particular applications. We found that the proposed systems perform both colorimetrically and spectrally as well as or better than the other systems while offering an easier

and more elegant solution to the problem of spectral imaging. Some systems give close spectral results, and only one of them comes significantly close in the color metrics. However, this system is made up of a scientific RGB color camera plus two ideal filters that are placed in front of it, alternatively, such that each capture would need two shots and the mechanical switch between filters from one shot to the next. This would make the real-time capture unsuitable. One of our proposed systems (system number 1) is also a two-shot system, but the tuning of the sensitivities is so fast and easy that we could still use it for real-time spectral imaging if the conditions of the amount of light in the scene being imaged allowed it. Otherwise, we offer a single-shot alternative system (system number 10), which would reduce drastically the time needed for capturing and would still give good colorimetric and spectral results. This study demonstrates the potential of TFD sensors as candidates to be part of spectral imaging systems that are portable, real-time, versatile, and low cost, as soon as they can be implemented as part of a real capturing system.

This work was funded by the Spanish Ministry of Economy and Competitiveness through the research project DPI2011-23202. We thank our colleague A. L. Tate for revising our English text.

References

1. C. I. Chang, *Hyperspectral Data Exploitation: Theory and Applications* (Wiley, 2007).
2. H. Grahn and P. Geladi, *Techniques and Applications of Hyperspectral Image Analysis* (Wiley, 2007).
3. Y. Garini, I. T. Young, and G. McNamara, "Spectral imaging: principles and applications," Cytometry **69**, 735–747 (2006).
4. H. R. Morris, C. C. Hoyt, and P. J. Treado, "Imaging spectrometers for fluorescence and Raman microscopy: acousto-optic and liquid-crystal tunable filters," Appl. Spectrosc. **48**, 857–866 (1994).
5. S. E. Harris and R. W. Wallace, "Acousto-optic tunable filter," J. Opt. Soc. Am. **59**, 744–747 (1969).
6. I. Chang, "Acousto-optic tunable filters," Opt. Eng. **20**, 206824 (1981).
7. J. Y. Hardeberg, F. Schmitt, and H. Brettel, "Multispectral color image capture using a liquid-crystal tunable filter," Opt. Eng. **41**, 2532–2548 (2002).
8. A. Rodríguez, J. L. Nieves, E. Valero, E. Garrote, J. Hernández-Andrés, and J. Romero, "Modified fuzzy c-means applied to a Bragg-grating-based spectral imager for material clustering," in *IS&T/SPIE Electronic Imaging* (International Society for Optics and Photonics, 2012), p. 83000J.
9. N. Hagen and M. W. Kudenov, "Review of snapshot spectral imaging technologies," Opt. Eng. **52**, 090901 (2013).
10. L. Gao, R. T. Kester, N. Hagen, and T. S. Tkaczyk, "Snapshot image mapping spectrometer (IMS) with high sampling density for hyperspectral microscopy," Opt. Express **18**, 14330–14344 (2010).
11. J. Y. Hardeberg, *Acquisition and Reproduction of Color Images: Colorimetric and Multispectral Approaches* (Universal-Publishers, 2001).
12. J. Brauers, N. Schulte, and T. Aach, "Multispectral filter-wheel cameras: geometric distortion model and compensation algorithms," IEEE Trans. Image Process. **17**, 2368–2380 (2008).
13. R. Shrestha, J. Y. Hardeberg, and A. Mansouri, "One-shot multispectral color imaging with a stereo camera," Proc. SPIE **7876**, 787609 (2011).
14. M. Tsuchida, T. Takahashi, K. Ito, T. Kawanishi, J. Yamato, and T. Aoki, "A stereo one-shot multi-band camera system for accurate color reproduction," in *ACM SIGGRAPH 2010 Posters* (ACM, 2010), p. 66.
15. M. Tsuchida, T. Kawanishi, K. Kashino, and J. Yamato, "A stereo nine-band camera for accurate color and spectrum reproduction," in *ACM SIGGRAPH 2012 Posters* (ACM, 2012), p. 18.
16. B. Gunturk, J. Glotzbach, Y. Altunbasak, R. Schafer, and R. Mersereau, "Demosaicking: color filter array interpolation," IEEE Signal Process. Mag. **22**, 44–54 (2005).
17. Y. Murakami, M. Yamaguchi, and N. Ohyama, "Hybrid-resolution multispectral imaging using color filter array," Opt. Express **20**, 7173–7183 (2012).
18. Y. Murakami, A. Tanji, and M. Yamaguchi, "Development of a low-resolution spectral imager and its application to hybrid-resolution spectral imaging," in *Proceedings 12th International AIC Congress* (2013), pp. 363–366.
19. G. Langfelder, "Spectrally reconfigurable pixels for dual-color-mode imaging sensors," Appl. Opt. **51**, A91–A98 (2012).
20. A. Longoni, F. Zaraga, G. Langfelder, and L. Bombelli, "The transverse field detector (TFD): a novel color-sensitive CMOS device," IEEE Electron. Device Lett. **29**, 1306–1308 (2008).
21. G. Langfelder, A. Longoni, and F. Zaraga, "Implementation of a multi-spectral color imaging device without color filter array," Proc. SPIE **7876**, 787607 (2011).
22. M. Martínez-Domingo, E. Valero, J. Hernández-Andrés, and G. Langfelder, "Spectral reflectance estimation from transverse field detectors responses," in *Conference on Color in Graphics, Imaging, and Vision* (Society for Imaging Science and Technology), 2012, pp. 378–383.
23. G. D. Finlayson, J. Vazquez-Corral, S. Süsstrunk, and M. Vanrell, "Spectral sharpening by spherical sampling," J. Opt. Soc. Am. A **29**, 1199–1210 (2012).
24. E. R. Fossum, "CMOS image sensors: electronic camera-on-a-chip," IEEE Trans. Electron Dev. **44**, 1689–1698 (1997).
25. P. M. Hubel, "Foveon technology and the changing landscape of digital cameras," in *Color and Imaging Conference* (Society for Imaging Science and Technology, 2005), pp. 314–317.
26. M. S. Drew and G. D. Finlayson, "Spectral sharpening with positivity," J. Opt. Soc. Am. A **17**, 1361–1370 (2000).
27. A. Lin and F. Imai, "Efficient spectral imaging based on imaging systems with scene adaptation using tunable color pixels," in *Color and Imaging Conference* (Society for Imaging Science and Technology, 2011), pp. 332–338.
28. F. H. Imai, "Computational spectral imaging based on adaptive spectral imaging," in *Computational Color Imaging* (Springer, 2013), pp. 35–52.
29. K. Hirakawa and P. J. Wolfe, "Spatio-spectral color filter array design for optimal image recovery," IEEE Trans. Image Process. **17**, 1876–1890 (2008).
30. L. Miao, H. Qi, R. Ramanath, and W. Snyder, "Binary tree-based generic demosaicking algorithms for multispectral filter arrays," IEEE T. Image Process. **15**, 3550–3558 (2006).
31. J. Brauers and T. Aach, "A color filter array-based multispectral camera," in *Workshop Farbbildverarbeitung* (Ilmenau, 2006).
32. G. Langfelder, T. Malzbender, A. Longoni, and F. Zaraga, "A device and an algorithm for the separation of visible and near infrared signals in a monolithic silicon sensor," Proc. SPIE **7882**, 788207 (2011).
33. G. Langfelder, "CMOS pixels directly sensitive to both visible and near-infrared radiation," IEEE Trans. Electron Dev. **60**, 1695–1700 (2013).
34. M. Hashimoto and J. Kishimoto, "Two-shot type 6-band still image capturing system using commercial digital camera and custom color filter," in *Conference on Color in Graphics, Imaging, and Vision* (Society for Imaging Science and Technology, 2008), pp. 538–541.
35. M. Tsuchida, A. Takayanagi, Y. Sakaguchi, and R. Mukai, "Estimation of spectral reflectance from six-band images based on partial least-squares regression," in *Proceedings 12th International AIC Congress*, 2013, pp. 1785–1788.
36. I. Chatzis, V. Kappatos, and E. Dermatas, "Filter selection for multi-spectral image acquisition using the feature vector

analysis methods," in *Intelligent Production Machines and Systems: 2nd I* PROMS Virtual Conference* (Elsevier Science, 2007).
37. Y. Lu, I. Cohen, X. Zhou, and Q. Tian, "Feature selection using principal feature analysis," in *Proceedings of the 15th International Conference on Multimedia* (ACM, 2007), pp. 301–304.
38. Natural Color System (NCS). Scandinavian Colour Institute of Stockholm, Sweden, (1979).
39. A. Hård and L. Sivik, "NCS- natural color system: a Swedish standard for color notation," Color Res. Appl. **6**, 129–138 (1981).
40. G. Langfelder, F. Zaraga, and A. Longoni, "Tunable spectral responses in a color-sensitive CMOS pixel for imaging applications," IEEE Trans. Electron Dev. **56**, 2563–2569 (2009).
41. http://www.andovercorp.com/Web_store/Standard_BP/Std_BP_General.php.
42. J. A. Hartigan and M. A. Wong, "Algorithm as 136: a k-means clustering algorithm," J. Roy. Stat. Soc. C-App **28**, 100–108 (1979).
43. J. Hernández-Andrés, J. Romero, and R. L. Lee, "Colorimetric and spectroradiometric characteristics of narrow-field-of-view clear skylight in Granada, Spain," J. Opt. Soc. Am. A **18**, 412–420 (2001).
44. V. Heikkinen, A. Mirhashemi, and J. Alho, "Link functions and Matérn kernel in the estimation of reflectance spectra from RGB responses," J. Opt. Soc. Am. A **30**, 2444–2454 (2013).
45. F. H. Imai, M. R. Rosen, and R. S. Berns, "Comparative study of metrics for spectral match quality," in *Conference on Colour in Graphics, Image and Vision* (Society for Imaging Science and Technology, 2002), pp. 492–496.
46. G. Sharma, W. Wu, and E. Dalal, "The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations," Color Res. Appl. **30**, 21–30 (2005).
47. J. L. Nieves, E. M. Valero, S. M. Nascimento, J. Hernández-Andrés, and J. Romero, "Multispectral synthesis of daylight using a commercial digital CCD camera," Appl. Opt. **44**, 5696–5703 (2005).
48. V. Heikkinen, T. Jetsu, J. Parkkinen, M. Hauta-Kasari, T. Jaaskelainen, and S. Lee, "Regularized learning framework in the estimation of reflectance spectra from camera responses," J. Opt. Soc. Am. A **24**, 2673–2683 (2007).
49. V. Heikkinen, "Kernel methods for estimation and classification of data from spectral imaging," Ph.D. dissertation (Faculty of Forestry and Natural Sciences, University of Eastern Finland, 2011).
50. M. Martínez-Domingo, E. Valero, V. Heikkinen, Jr., and G. Langfelder, "Design of a multispectral system based on transverse field detectors," in *Proceedings 12th International AIC Congress*, 2013, pp. 371–374.

Adaptive exposure estimation for high dynamic range imaging applied to natural scenes and daylight skies

Miguel A. Martínez,* Eva M. Valero, and Javier Hernández-Andrés

Color Imaging Laboratory, Department of Optics, Faculty of Sciences, University of Granada, Granada, Spain

*Corresponding author: m.martinez.domingo@gmx.com

Received 2 September 2014; revised 15 December 2014; accepted 17 December 2014;
posted 17 December 2014 (Doc. ID 222306); published 26 January 2015

Digital imaging of natural scenes and optical phenomena present on them (such as shadows, twilights, and crepuscular rays) can be a very challenging task because of the range spanned by the radiances impinging on the capture system. We propose a novel method for estimating the set of exposure times (bracketing set) needed to capture the full dynamic range of a scene with high dynamic range (HDR) content. The proposed method is adaptive to scene content and to any camera response and configuration, and it works on-line since the exposure times are estimated as the capturing process is ongoing. Besides, it requires no *a priori* information about scene content or radiance values. The resulting bracketing sets are minimal in the default method settings, but the user can set a tolerance for the maximum percentage of pixel population that is underexposed or saturated, which allows for a higher number of shots if a better signal-to-noise ratio (SNR) in the HDR scene is desired. This method is based on the use of the camera response function that is needed for building the HDR radiance map by stitching together several differently exposed low dynamic range images of the scene. The use of HDR imaging techniques converts our digital camera into a tool for measuring the relative radiance outgoing from each point of the scene, and for each color channel. This is important for accurate characterization of optical phenomena present in the atmosphere while not suffering any loss of information due to its HDR. We have compared our method with the most similar one developed so far [IEEE Trans. Image Process. **17**, 1864 (2008)]. Results of the experiments carried out for 30 natural scenes show that our proposed method equals or outperforms the previously developed best approach, with less shots and shorter exposure times, thereby asserting the advantage of being adaptive to scene content for exposure time estimation. As we can also tune the balance between capturing time and the SNR in our method, we have compared its SNR performance against that of Barakat's method as well as against a ground-truth HDR image of maximum SNR. Results confirm the success of the proposed method in exploiting its tunability to achieve the desired balance of total Δt and SNR. © 2015 Optical Society of America

OCIS codes: (110.1758) Computational imaging; (100.2000) Digital image processing; (040.1490) Cameras; (010.7295) Visibility and imaging; (110.1085) Adaptive imaging; (110.0110) Imaging systems.
http://dx.doi.org/10.1364/AO.54.00B241

1. Introduction

Natural scenes are usually composed by a wide variety of radiance signals outgoing from the objects in the scene, which are very different in magnitude. This fact makes their correct capture with a normal digital camera a nontrivial problem. Capturing a scene with high dynamic range (HDR) content with a single low dynamic range (LDR) image would cause loss of information in those regions of the scene where the light level reaching the sensor is too low or too high to be correctly registered with a single exposure time. The HDR imaging techniques solve this problem.

Common imaging sensors suffer from limitations in the process of capturing the light. Usually, the dynamic range of the sensor (i.e., the ratio between the maximum and minimum irradiance impinging

1559-128X/15/04B241-10$15.00/0
© 2015 Optical Society of America

on the sensor that produces an effective response) is much lower than the dynamic range found in natural open air scenes. The dynamic range of these scenes (ratio between the maximum and minimum radiances emitted by the objects in the scene) can vary from 2 to 8 orders of magnitude depending on the season and scene content [1]. The human visual system can simultaneously adjust to a difference of up to 3.73 orders of magnitude (or log units) [2] when adaptation is accomplished. However, most imaging and display devices can only account for barely 2 orders of magnitude in a single image (either for capturing or for displaying) [3,4].

There have been many techniques proposed [5–7] as well as sensor architectures [8,9] to achieve this goal. The most common techniques are based on building a HDR image from the information of a number of different LDR images. The difference between these LDR images is the exposure (i.e., the product of irradiance impinging in the sensor times the exposure time used to acquire the image). It is changed by varying either exposure time or the aperture. Usually, the former is adjusted since it does not affect the depth of field between different captures. These differently exposed versions of the same scene, when combined in the correct way, can be used to build an image that contains extended dynamic range information compared to just a single exposure [10].

When we use our camera to capture a digital image of a scene, we cannot know in advance which exposure times would be useful for composing the HDR image afterwards. We could just take a large number of images with different exposure times (or even all available ones in the camera), and then use all of these LDR images to compose the HDR image. But this option is often time consuming and very computationally demanding, so it is not always feasible. Commercial cameras usually have an auto-exposure mode that estimates an exposure time value based on some cues like the reading from a built-in photometer in high-end cameras that measures the average brightness value in certain regions of the image.

All of these cases aim just for finding one exposure time that works well for correctly imaging most parts of the scene. However, whether we would be able to find it or not, there is not a single value of exposure time that could make all pixels in one shot be correctly exposed for most common natural scenes.

The aim of this paper is not to explain how to merge LDR images into a HDR radiance map. The process we used for this is very well explained in the literature [11]. It is rather to present a method for the selection of a set of exposure times (bracketing set) to use in order to retrieve useful information from all pixels (or at least from most of them). This is very important for the study of optical phenomena present in the atmosphere and open air natural scenes with shadows, twilights, clouds, crepuscular rays, and so forth [12], all of which have HDR content that cannot be captured with a single shot of a commercial digital camera. Thus, a digital camera can be a useful tool for composing a HDR radiance map of these phenomena in order to study them. Of course, if the scene captured has very dark regions that need long exposure times to be correctly exposed, then it is important that there is no relative movement between the camera and scene content during the capturing time. If small movements happen, there are ghosting-compensation techniques to correct for artifacts [13].

We aimed for a method that is blind (no information from the scene content is known *a priori*), adaptive (adapts to scene content dynamically by adjusting required exposure times), universal (works for any camera that we have tested so far), and on-line (the exposure times are calculated as the capturing process is ongoing and every single shot acquired is used in the HDR radiance map generation). It will also give as default output the minimal bracketing set (the bracketing set that has the minimum number of shots, yet recovers the full dynamic range of the scene), but it can be tuned to yield longer exposure times with a higher signal-to-noise ratio (SNR). This tunability is introduced as a method that controls the amount of overlapping between consecutive exposures to increase the SNR in the resulting HDR image at the cost of increasing the number of shots taken, and hence the capturing time. We have also introduced a method to control the percentage of pixel population that we can accept to be useless.

The remainder of this paper is organized as follows. Section 2 summarizes the state-of-the-art technology for estimating exposure times for HDR imaging. Section 3 explains the details of the method we propose. Section 4 explains the experiments made to compare our method with the most advanced method of those described in Section 2, as well as the results obtained. Finally, Section 5 draws the main conclusions of this work.

2. State of the Art

Several approaches have been proposed in the literature for solving the problem of finding the exposure time values for HDR image capture via multiple exposures. Chen and Gamal [14] proposed scheduling for capture times. They were assuming a known illuminant in the scene, which in practice is a rather non-realistic assumption, especially for optical open air phenomena. Grossberg and Nayar [15] proposed a method to simulate the response of any camera (linear, logarithmic, gamma function, etc.) using a single camera with a known camera response function (CRF) by just selecting a set of exposure times. So their aim was not to find a minimum bracketing set for radiance map generation. Stumpfel *et al.* [16] proposed a method for capturing HDR images of the sun and sky. They threshold the images and check if there are saturated and/or underexposed pixels. If any, they add new shots by increasing or decreasing the exposure by a fixed amount of three-stops. This approach is not adaptive to the scene content and could lead to situations where the exposure times

are not well fitted to cover the full dynamic range of the scene. Bilcu *et al.* [17] proposed a method for overcoming the limitations of mobile devices for HDR imaging. Their proposal was also done by iteratively trying every available value of exposure time and afterwards deciding which is the correct one. Therefore, many images need to be taken before a bracketing set is selected.

Barakat *et al.* [18] proposed a method for finding minimal bracketing sets for HDR capture. Firstly, they studied how the camera responds to radiance using every available exposure time. Then, they select only those exposure times that completely cover the full dynamic range of the camera with certain overlap. This is the so-called minimal system bracketing set (MSBS), and whatever the content of the scene being imaged is, using all these exposures will always cover the full radiance range that the camera can effectively acquire. To adapt this MSBS to scene content, they proposed to select a subset of it called the minimum image bracketing set (MIBS) by capturing a first shot with an intermediate exposure time (that belongs to the MSBS) and checking if there are saturated or underexposed pixels. If so, they add the next exposure time included in the MSBS until the full dynamic range is covered. Though the underlying idea in this method is similar to the one proposed in this work, it is still not totally adaptive to scene content since they limit the exposure times selected to those belonging to the MSBS. We believe that the same scenes could be captured with less shots and shorter exposure times, yet covering their full dynamic range, and we have demonstrated this by comparing the results of our method with Barakat *et al.*'s insightful algorithm using their MIBS approach.

Granados *et al.* [19] proposed a method assuming the known mean HDR irradiance histogram of the scene being captured. Besides, their method only works for linear cameras, and they used a greedy algorithm, iteratively capturing the same scene many times until they obtained the optimal SNR solution. Thus, their method is not on-line. Hirakawa and Wolfe [20] used a mathematical method based on training for HDR exposure time selection. They assumed linear sensor response and known noise sources in the capture, which is not always a realistic scenario. They did not really aim for defining minimum bracketing sets, but for optimal SNRs instead. Gelfand *et al.* [21] adapted HDR imaging to mobile devices as well. They merge LDR images iteratively two by two. If there are still saturated or underexposed pixels, they keep adding the next available exposure time that the camera offers. Hasinoff *et al.* [22] proposed a method to calculate a bracketing set that is optimal in terms of the SNR but not minimal (they just try to fit it within a given time budget) by varying both the exposure time and the ISO (i.e., sensitivity) settings of the camera. Besides, their method assumes linear raw sensor responses only and known information about scene radiance content.

Gallo *et al.* [23] proposed a method for taking advantage of mobile phone camera APIs (application programming interfaces). They programmed the automatic histogram calculation in mobile phones to construct a reduced HDR histogram of the scene, which will be the target to be captured. This is however not possible if the camera used for the captures does not feature this automatic process. Besides, if the scene imaged has very dark regions, the long exposures needed to create this histogram make the process slow. Moreover, the method is not on-line. After this histogram is calculated, they capture several exposures of the scene and then study many possible combinations of them until the optimal one is selected.

Guthier *et al.* [24] followed the lines previously proposed by Kang *et al.* [25], who implemented a method for sequentially adjusting the exposure for real-time HDR video. Both are iterative and limited to only two shots for building the HDR image. Finally, Gupta *et al.* [26] proposed a Fibonacci-series-based bracketing set determination algorithm in which each exposure time is the sum of the previous two. This technique does not aim for full dynamic range recovery though, but image registration for HDR video.

3. Proposed Method

We drove our Canon EOS 7D camera from our laptop via the USB (universal serial bus) port using the open-source libraries called *GPhoto2* from our algorithm implemented in Matlab R2014a and working on-line.

The method proposed in this paper is full range because it finds a bracketing set that covers the full dynamic range radiance map of the scene. This HDR radiance map would be potentially useful for studying the behavior of light in HDR open air phenomena.

Our method uses the CRF to compute the relative irradiance impinging on the sensor, which corresponds to a certain population of pixels in the image (using its cumulative histogram). Then, a new exposure time is calculated in order to shift the camera responses to this irradiance to a different value. This way, the cumulative histogram is shifted and the new shot would capture a different range of irradiances, which are contiguous to the range captured in the previous shot.

The CRF is a function that relates the response of the camera, in digital counts (DC), with the exposure that the sensor receives. This function depends on each camera, and it even can be different for different settings of the same camera (e.g., a camera working in raw mode or in jpeg mode). Knowing the CRF of our camera is a key factor to build the radiance map. A detailed explanation of how to calculate the CRF is given in [11]. The calculated CRF of the camera used for the experimental part in this work is shown in Fig. 1. It is clearly not linear.

The exposure axis is in relative units, and it is normalized so that the center of the DC values (128 DC for the 8-bits case) corresponds to a relative exposure

value of 1. The function is the same for the three color channels R, G, and B of the camera, since it is a property of the sensor. Therefore, we process the three color channels together like the technique in [23]. For each LDR image we capture, we know the Δt used as well as the DC values for each pixel and channel. Therefore, by using the CRF we can easily work out the relative irradiance (E) by computing the simple ratio shown in Eq. (1):

$$E_{i,k,e} = \frac{H(i,k,e)}{\Delta t(e)}. \tag{1}$$

The subindex i accounts for the pixel index, k accounts for the color channel, and e accounts for the exposure index (or number of shots). Thus, once we have captured an initial image with a known exposure time Δt_0, the CRF relates H_0 with DC_0 as shown in Eq. (2):

$$\mathrm{DC}_0 = \mathrm{CRF}(H_0) = \mathrm{CRF}(E_0 \cdot \Delta t_0). \tag{2}$$

Therefore, we can work out the relative irradiance value of a point of the image [E_0, calculated as shown in Eq. (3)] by knowing the CRF, the Δt, and its DC value in the first shot (Δt_0 and DC_0, specifically):

$$E_0 = \frac{\mathrm{CRF}^{-1}(DC_0)}{\Delta t_0}, \tag{3}$$

where CRF^{-1} refers to the inverse CRF function that always exists since CRF is a monotonically increasing function. Then, to shift the sensor responses DC_0 to this same irradiance value E_0 into a new value DC_1, we just have to workout which new exposure time Δt_1 is needed for a new shot, like shown in Eq. (4):

$$\Delta t_1 = \frac{\mathrm{CRF}^{-1}(\mathrm{DC}_1)}{E_0}. \tag{4}$$

If our camera has only a limited set of values from which to choose the exposure time, we can select the available value that is closest to the calculated one.

We already have a tool to control the values of sensor responses, which is done by tuning the exposure time used to acquire the images. Now, we explain how to use it for our purpose of optimizing HDR capture. For this aim, we propose a method based on cumulative histograms of the scene inspired by Grossberg and Nayar [27], who originally applied it to pixel selection for CRF computations. If the scene content does not change, then the same value for the percentile of population in the cumulative histogram will correspond to the same areas in the image. In Fig. 2, we can see a plot where several cumulative histograms of the same scene differently exposed are drawn together.

If a given percentile is below some exposure value for a given exposure time, then, for a different exposure time, the same percentile of population will correspond to a different exposure value but they will still keep its location within the scene. Therefore, the points where the horizontal lines in Fig. 2 intersect the histograms report information corresponding to the same areas of the scene. Our idea is to shift sensor responses by calculating exposure times to control the sensor responses to pixel populations of key percentile values. As a starting point, we calculate the cumulative histogram of the image captured with the automated exposure of the camera. But in principle, any image can be used as starting point as long as it has some pixels that are neither saturated nor underexposed.

We are going to sample the scene's radiance using the CRF of the camera between two DC levels. Unless the scene has a very reduced dynamic range, there will be pixel values below and above these DC values. Since in the default version of the algorithm we aim for minimum bracketing sets, we have set the low level (LO) to 3 DC and the high level (HI) to 252 DC for considering a pixel to be underexposed or saturated, respectively, when it is out of these bounds. Thus, whatever pixel population is above the HI level or below the LO level, we will sample it using a different exposure time. Here, we introduce two novel features of our method. One is the possibility of setting a tolerance for the percentage of useless pixels. If we choose 0% tolerance, the algorithm will look for longer or shorter exposure times if at least one pixel is underexposed or saturated, like Barakat *et al.* [18] proposed. However, for some scenes we can set a different tolerance threshold to renounce to a certain percentage of the population to be properly exposed [e.g., when we directly image the sun and our region of interest (ROI) is in a different area].

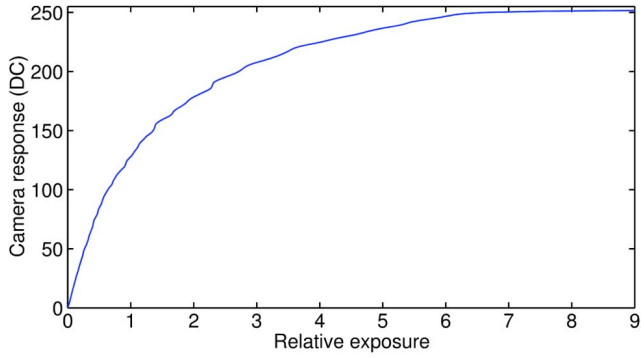

Fig. 1. CRF of the Canon EOS 7D camera in jpeg mode.

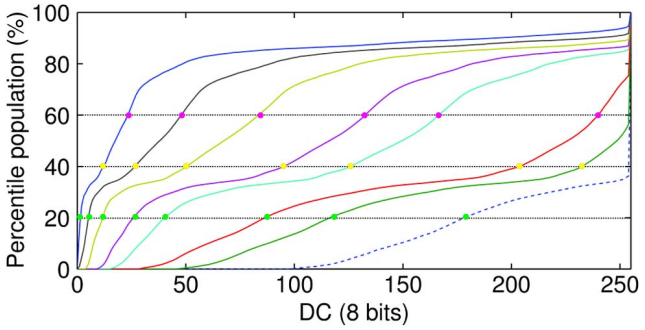

Fig. 2. Cumulative histograms of the same scene captured using different exposure times.

The other novelty involves controlling the LO and HI values of the CRF. Setting values very close to the extremes (0 and 255 DC for 8 bits) will result in a lower number of shots at the cost of a lower SNR. In contrast, if we set values further from these extremes, we will sample the scene's radiance with more overlap between contiguous shots and therefore the SNR will increase, at the cost of a higher number of shots. This shows how our algorithm can be tuned to adapt to different requirements regarding the SNR of the captured HDR.

After commenting on these functionalities, we describe now in detail how the exposure time search is done. With the information present in the cumulative histogram of the first shot captured, we check the percentile of the pixel population that is below the LO level. If it is higher than the maximum value set, then a longer exposure time is calculated. The same is done if the difference betwen 100 and the percentile of the pixel population above the HI level is higher than the tolerance threshold. In this case, a shorter exposure time will be calculated.

To find a longer exposure time, we will use Eq. (4) to shift the camera response value from the LO level to HI level. Therefore, we use the HI level as DC_1 and E_0 is substituted by Eq. (3), where we use the LO level as DC_0. Δt_0 is the exposure time used to acquire the current image [see Eq. (5)]:

$$\Delta t_{\text{longer}} = \frac{\text{CRF}^{-1}(\text{HI})}{\text{CRF}^{-1}(\text{LO})} \cdot \Delta t_0. \tag{5}$$

In contrast, to find a shorter exposure time, we will use Eq. (4) to shift the camera response value from the HI level to LO level. Therefore, we use the LO level as DC_1 and E_0 is substituted by Eq. (3), where we use the HI level as DC_0 [see Eq. (6)]:

$$\Delta t_{\text{shorter}} = \frac{\text{CRF}^{-1}(\text{LO})}{\text{CRF}^{-1}(\text{HI})} \cdot \Delta t_0. \tag{6}$$

In this way, if the population that has a sensor response at the HI level in one shot shifts to the LO level in the next shot, we can cover the full dynamic range of the scene with certain overlap between contiguous shots.

The process described here goes on checking the cumulative histograms of the longer and shorter exposure times until the tolerance requested is met or the system reaches its maximum or minimum available exposure times.

4. Experiments and Results

Our camera (Canon EOS 7D) allows the choice of only a discrete set of exposure times. We tuned the HI and LO levels (explained in Section 3) to get the minimum bracketing sets (lowest SNR). In the first experiment, we tested the default version of the adaptive exposure estimation (AEE) method (see Section 4.B.1); in the second experiment, we explored the tunability and evaluated the SNR performance (see Section 4.B.2). In all scenes tested, the method built a full dynamic range radiance map of the scenes. We implemented our proposed method as well as the only method found following the same philosophy, which is the MIBS method proposed by Barakat *et al.* [18] (hereafter termed BAR), to compare their performances.

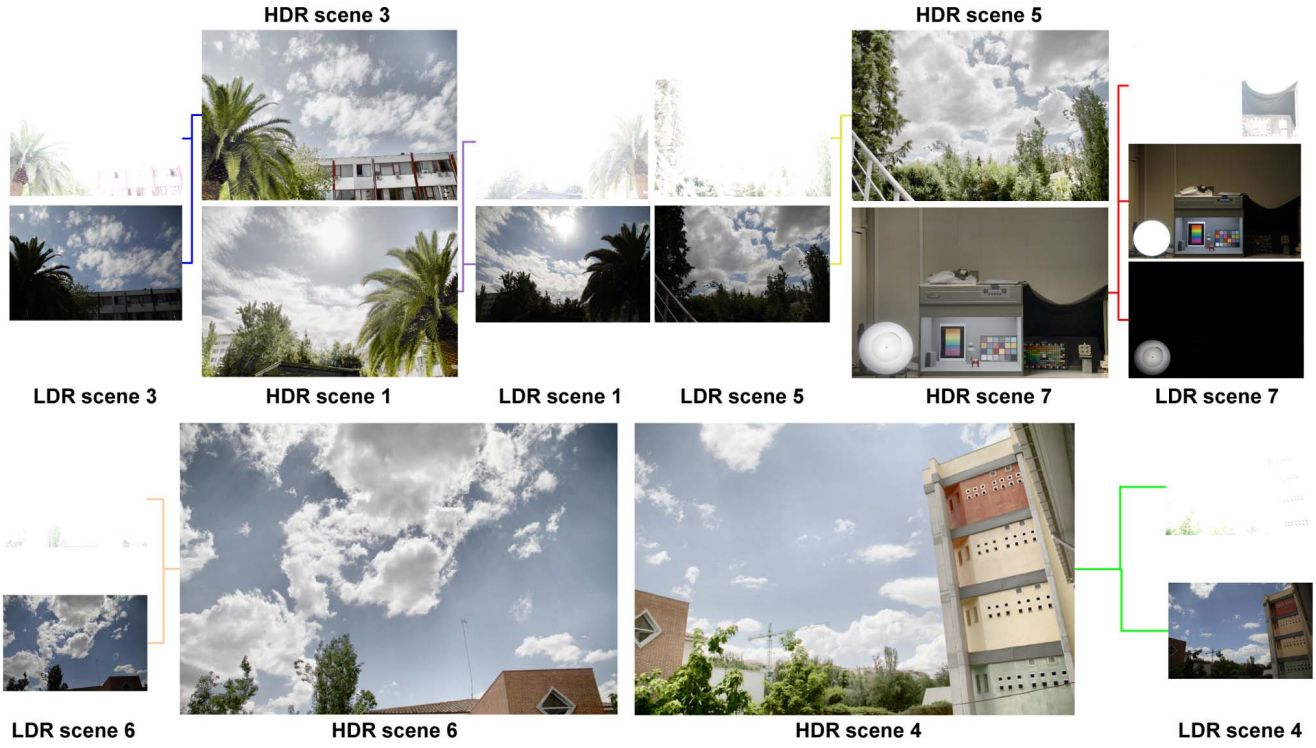

Fig. 3. LDR images and HDR radiance maps for some captured scenes.

A. Bar Method and MSBS

Regarding the BAR method, the MSBS found for our Canon camera, as explained in [18], using a 5.6 aperture setting, was composed by four exposure times for which the values were: 30 s, 300 ms, 1 ms, and 0.0125 ms. Sometimes, not all of these four shots were needed to record the full dynamic range of the scene. In these cases, a sub-set of the MSBS is used omitting some of its shots. This represents the MIBS. We could use as well as a first shot, the one chosen by the auto-exposure mode of the camera, as was done for the AEE method. However, we found that when doing so, we only got the same number of shots or even one more. So we did not use it. We see an example of this in Fig. 4. The exposure time chosen by the auto-exposure mode of the camera was 66.7 ms (top-center). It was used as first shot for the AEE method, since this method adapts to any exposure value chosen as first. However, this value was in between 300 and 1 ms (both belong to MSBS). Therefore, if we used it also as a first shot for the BAR method, it would mean that the capture of this scene would end up with five shots instead of four.

Also, if we change the aperture setting of the camera, as the CRF is not changing, the AEE method would work the same by adapting to the new exposure levels impinging in the sensor. However, for the BAR method we would need to calculate a new MSBS, since the same exposure times for a different aperture would not be valid any more. Thus, we fixed our aperture setting to 5.6 for both methods.

B. Comparison between the AEE and BAR Methods

For the first experiment (Section 4.B.1), we captured 30 scenes using both methods and we studied the number of shots taken, the total exposure time used, and the percentage of the pixel population that was not properly exposed. In this way, we assessed how efficiently did both methods recover the full dynamic range of the scene by comparing their resulting bracketing sets.

In the second experiment (Section 4.B.2), we built an indoors HDR scene with controlled illumination conditions. We captured 10 HDR images of it using the AEE and BAR methods. Besides, for the AEE method we repeated the capture four times using different values for the LO and HI levels (see Section 3). Finally, we captured 10 ground-truth (GT) HDR images using all available exposure times in the camera. These GT images represent the highest SNR that our camera can achieve to record a scene without repeating shots with the same exposure.

1. *HDR Capturing Efficiency*

As mentioned before, we acquired 30 HDR scenes; 23 scenes were captured outdoors with natural illumination and 7 were captured indoors with artificial illumination. Outdoors, daylight cast HDR illumination over objects including clouds. Indoors we used a light booth and a fluorescent lamp oriented directly to the camera in a dark room to generate HDR content. To check the performance of both methods in terms of full range recovery, we plotted the

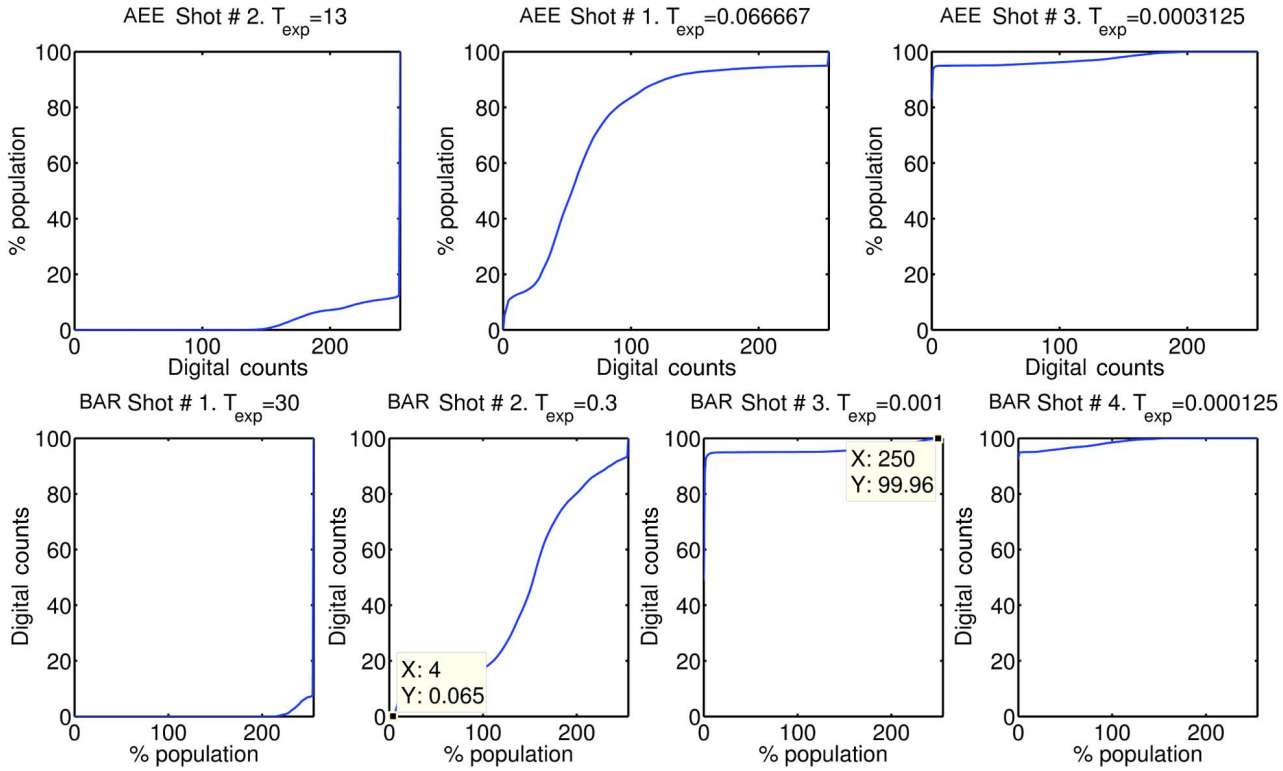

Fig. 4. Cumulative histograms of the same scene using the AEE method (top row) and BAR method (bottom row). The histograms are ordered by decreasing exposure time to observe their continuity.

cumulative histograms of all shots taken for each scene and checked that no irradiance gaps were left uncovered between consecutive shots. We set the maximum percentile allowed to be lost to 0%. In Table 1, we can see the results for 7 out of the 30 scenes captured.

We observed how the number of shots is always equal or lower for the AEE method. The percentage of useless pixels is always the same for both methods. Many scenes had a percentage of lost pixels equal to 0, since both methods managed to retrieve the full dynamic range of the scene. For the rest of the scenes, the useless pixels were due to direct sunlight (like the case of scene 1). This made some pixels impossible to recover even for the shortest exposure time available in the camera. The total exposure time is always lower for the AEE method. The same trends commented on were found for the remaining 23 scenes captured. In total, for the 30 scenes captured, the BAR method took a total of 96 shots using 218.734 seconds and the AEE method took a total of 81 shots using 139.869 seconds. This means that the number of shots was 15.63% less, and the exposure time was 36.06% less, for the AEE method.

In Fig. 3, you can see the LDR pictures and the tone-mapped HDR radiance maps generated for some of the scenes. The tone-mapping algorithm used was a contrast-limited adaptive histogram equalization, which was introduced by Ward [28] and implemented in Matlab R2014a.

In Fig. 4, we plot an example of the cumulative histograms corresponding to both methods for scene number 7. The data-tips in histograms for shots 2 and 3 for the BAR method highlight that there are still underexposed and saturated pixels in those exposures, although at first glance the histograms may seem to reach percentiles 0 and 100, respectively.

We observed how both methods succeeded in recovering the full dynamic range of the scene. However, thanks to the adaptation of the AEE method, only three shots were needed instead of the four shots the BAR method used. Therefore, we can conclude with this experiment that the AEE method recovers the dynamic range of the scene as well as the BAR method does, but by using a more reduced bracketing set.

Table 1. Minimum Bracketing Sets Found for BAR and AEE Methods in 7 of the 30 Scenes Captured

Scene #	Method	# Shots	% Lost	$\sum_{i=1}^{n} T_i(s)$
1	BAR	3	1.61	0.301
	AEE	2	1.61	0.025
2	BAR	3	0	0.301
	AEE	2	0	0.101
3	BAR	3	0	0.301
	AEE	2	0	0.040
4	BAR	3	0	0.301
	AEE	2	0	0.050
5	BAR	3	0	0.301
	AEE	2	0	0.034
6	BAR	3	0	0.301
	AEE	2	0	0.025
7	BAR	4	0	30.301
	AEE	3	0	13.067

2. *Signal-to-Noise Ratio*

We did a second experiment to study the SNR behavior of our method. For the AEE method, we tested four different conditions named A (LO = 3 and HI = 253), B (LO = 16 and HI = 240), C (LO = 56 and HI = 200), and D (LO = 106 and HI = 150). We compared all these AEE conditions with the BAR method and the ground-truth (GT) images.

The dynamic range of the scene was measured using a spectroradiometer (Photo Research, PR-745) to measure the integrated radiances of both the brightest and the darkest points of the scene. The resulting dynamic range measured was 4.1 log units.

For each pixel and each color channel of these HDR radiance maps, we calculated its average HDR value and its standard deviation across the ten images corresponding to each method. The average HDR value provides information about the signal level in the pixel, and the standard deviation provides information about the level of noise generated by all noise processes present in the HDR capture process. Thus, by computing the ratio of the average HDR value ($\bar{E}_{xy}$) over the standard deviation (σ_{xy}) as Eq. (7) shows, we obtain a SNR estimate [29]:

$$\mathrm{SNR}_{xy} = 20 \times \log_{10}\left(\frac{\bar{E}_{xy}}{\sigma_{xy}}\right). \tag{7}$$

The subindex xy stands for pixel position within the HDR radiance map.

We can see the number of shots, the total exposure times, and the average SNR for each method in Table 2.

As expected, setting the LO and HI values further from the extremes of the range in the AEE method yields a higher number of shots and also higher total exposure time, but the SNR increases as well. For condition D, we reached an average SNR only less than 2 dB below the ideal case (GT), yet using only about 40% of the total exposure time. The minimum bracketing set found was AEE A, with only three shots and 14.79 s of total exposure time, but this also had the lowest SNR. The BAR method needed the full MSBS to recover this scene using four shots. It had a better SNR than our minimum bracketing

Table 2. SNR Performance for Four AEE Conditions and the GT and BAR Methods

Method	# Shots	Δt (s)	$\overline{\mathrm{SNR}}$ (dB)
AEE A	3	14.79	27.19
AEE B	4	15.02	30.47
AEE C	5	31.32	32.58
AEE D	16	61.22	33.57
GT	55	151.43	35.32
BAR	4	30.30	29.97

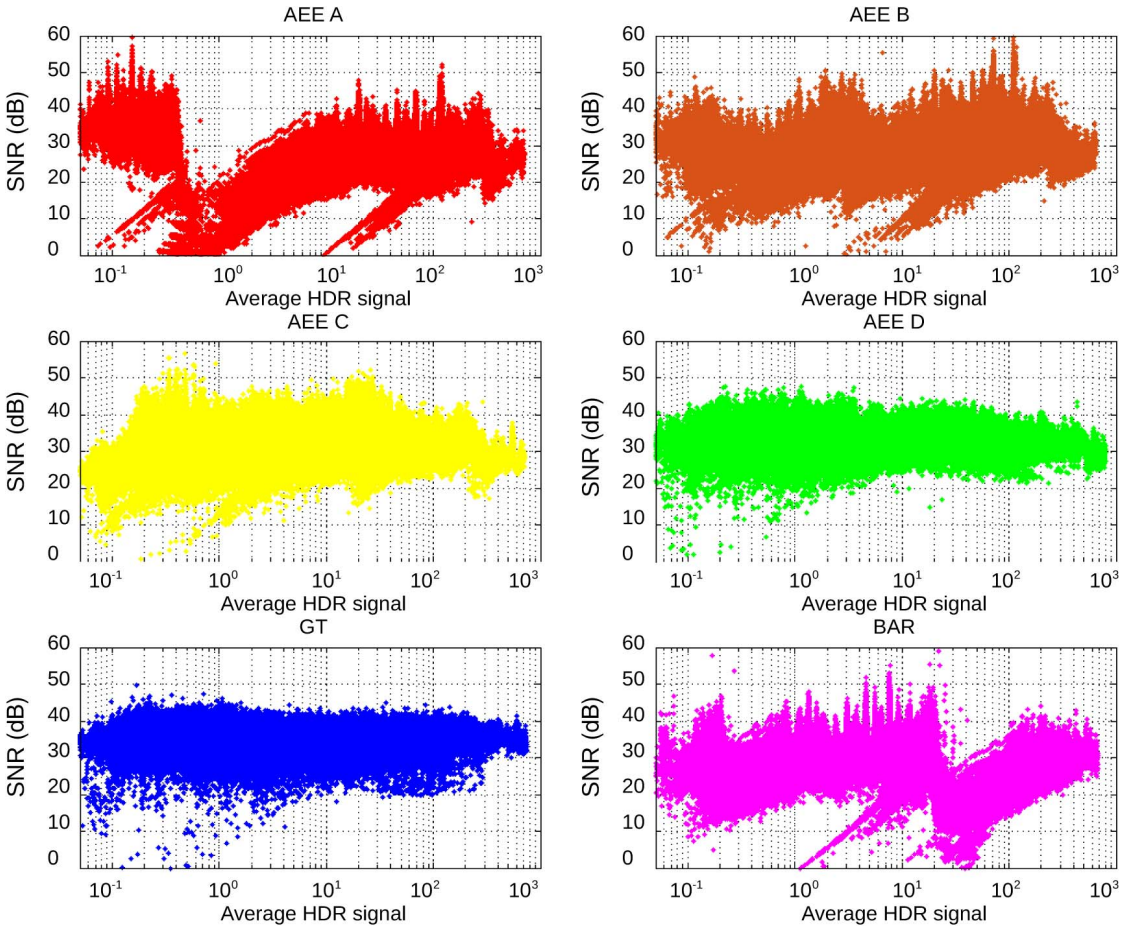

Fig. 5. SNR versus average HDR signals present in the radiance maps.

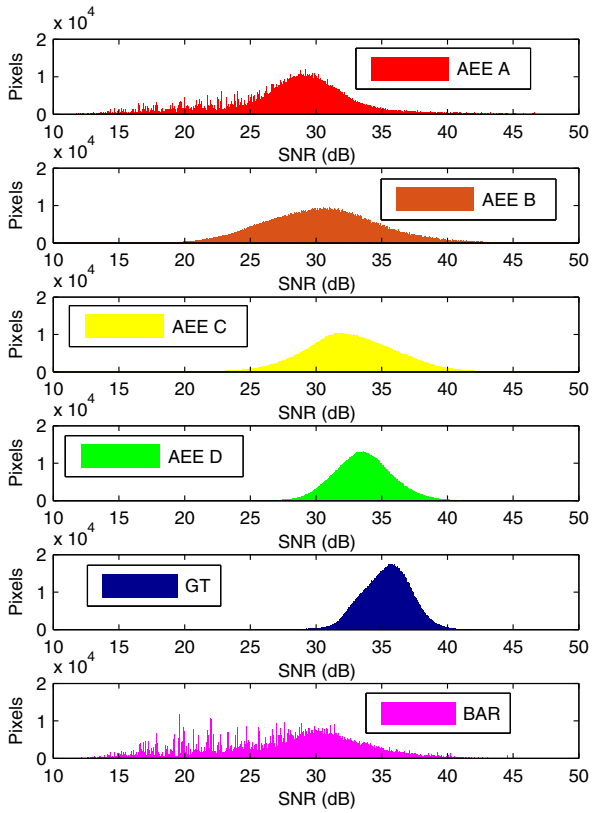

Fig. 6. Histograms of SNR values.

set, but we also had a second option (AEE B) using four shots with a shorter total exposure time (less than half) and a higher SNR than the BAR method.

We can observe in Fig. 5 the SNR for each pixel of the radiance maps generated versus the signal level. We can also see how the AEE D has the most similar distribution compared with GT. Also, the BAR method has a very similar distribution compared with cases AEE A and AEE B, as expected.

Figure 6 plots the SNR histograms for all methods. We can observe how for the AEE method, the main lobe gets narrower and shifts towards a higher mean SNR as we tune the LO and HI levels further from the extremes of the range. The AEE D is quite close to GT in position and shape. In contrast, the BAR method yields a histogram that is the mostly spread over a wide range of SNR values.

5. Conclusions

We present a new method for estimating the exposure times needed to recover the HDR radiance map from a scene via multiple exposures. We compared the performance of our method with that of the only method found in literature that aims for the same purpose (i.e., finding minimum bracketing sets) and performs under the same conditions (adapted to scene content with no *a priori* information about it and valid for any camera whether it is linear or not).

Our proposed method is adaptive because it finds a bracketing set adapted to any HDR scene content, and it is universal because it works for any camera. We only need to calculate its CRF (which is needed anyway to build the radiance map).

Moreover, the method is tunable, since we can decide if we prefer to find a minimum bracketing set at the cost of higher SNR or increase the SNR sampling the radiance of the scene with more overlapping between consecutive shots (increasing the number of shots and capturing time as well). For the minimum bracketing set case, the bracketing sets found were minimal in the 30 scenes tested.

Futhermore, our method is blind, which means that no information about the content of the scene needs to be known *a priori*. The multiple LDR images are captured on-line as the process is ongoing, and every single shot taken is used to compose the HDR radiance map.

We can also control the percentage of the total pixel population that we can assume is useless (underexposed or saturated). This way, we can find the minimum bracketing set only for our region of interest.

We have applied the method for HDR imaging of natural scenes where partially cloudy skies were present in order to increase the dynamic range of the capture. We have successfully covered the full dynamic range of the 30 scenes imaged. We have shown how our method can find bracketing sets that are shorter than those found by the BAR method, yet keeping higher SNR levels in the HDR radiance map reconstructed from the multiple exposures.

We studied the SNR performance of our method comparing it not only with the BAR method but also with an ideal-case ground-truth HDR image built using all available exposure times in the camera. We have demonstrated how we can tune our method to suit different requirements for the SNR at the cost of increasing the number of shots.

The proposed method brings a solution for the blind acquisition of HDR images using multiple exposures, which can be used in any HDR imaging context: machine vision, sky imaging, daylight illuminated scenes, HDR photography, etc. And in particular, the proposed method may be useful for studying optical phenomena present in open air scenes where the illumination conditions are extreme (i.e., direct sunlight that might be surrounded by regions of interest like halos, clouds casting shadows, and rainbows, just to give a few examples).

This work was funded by the Spanish Ministry of Economy and Competitiveness through the research project DPI2011-23202.

References

1. J. J. McCann and A. Rizzi, *The Art and Science of HDR Imaging* (Wiley, 2011), Vol. **26**.
2. T. Kunkel and E. Reinhard, "A reassessment of the simultaneous dynamic range of the human visual system," in *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization* (ACM, 2010), pp. 17–24.
3. K. Yamada, T. Nakano, S. Yamamoto, E. Akutsu, and I. Aoki, "Wide dynamic range vision sensor for vehicles," in *Proceedings of Vehicle Navigation and Information Systems Conference* (IEEE, 1994), pp. 405–408.
4. H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs, "High dynamic range display systems," in *ACM Transactions on Graphics* (ACM, 2004), Vol. **23**, pp. 760–768.
5. S. Mann and R. Picard, *Being Undigital with Digital Cameras* (MIT, 1994).
6. S. Nayar and T. Mitsunaga, "High dynamic range imaging: Spatially varying pixel exposures," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2000), pp. 472–479.
7. S. Nayar and V. Branzoi, "Adaptive dynamic range imaging: Optical control of pixel exposures over space and time," in *Proceedings of the Ninth IEEE International Conference on Computer Vision* (IEEE, 2003), pp. 1168–1175.
8. S. Kavusi, K. Ghosh, and A. El Gamal, *Architectures for High Dynamic Range, High Speed Image Sensor Readout Circuits* (Springer, 2008).
9. S. Kavusi and A. El Gamal, "Quantitative study of high-dynamic-range image sensor architectures," in *Electronic Imaging* (International Society for Optics and Photonics, 2004), pp. 264–275.
10. M. A. Robertson, S. Borman, and R. L. Stevenson, "Dynamic range improvement through multiple exposures," in *Proceedings of the International Conference on Image Processing* (IEEE, 1999), pp. 159–163.
11. P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *ACM SIGGRAPH 2008 Classes* (ACM, 2008), p. 31.
12. D. K. Lynch and W. C. Livingston, *Color and Light in Nature* (Cambridge University, 2001).
13. O. Gallo, N. Gelfand, W. Chen, M. Tico, and K. Pulli, "Artifact-free high dynamic range imaging," in *IEEE International Conference on Computational Photography (ICCP)* (IEEE, 2009), pp. 1–7.
14. T. Chen and A. El Gamal, "Optimal scheduling of capture times in a multiple-capture imaging system," in *Electronic Imaging* (SPIE, 2002), pp. 288–296.
15. M. D. Grossberg and S. K. Nayar, "High dynamic range from multiple images: which exposures to combine," in *Proceedings of the ICCV Workshop on Color and Photometric Methods in Computer Vision (CPMCV)* (IEEE, 2003).
16. J. Stumpfel, C. Tchou, A. Jones, T. Hawkins, A. Wenger, and P. Debevec, "Direct HDR capture of the sun and sky," in *Proceedings of the 3rd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa* (ACM, 2004), pp. 145–149.
17. R. C. Bilcu, A. Burian, A. Knuutila, and M. Vehvilainen, "High dynamic range imaging on mobile devices," in *the 15th IEEE International Conference on Electronics, Circuits and Systems, ICECS* (IEEE, 2008), pp. 1312–1315.
18. N. Barakat, A. N. Hone, and T. E. Darcie, "Minimal-bracketing sets for high-dynamic-range image capture," IEEE Trans. Image Process. **17**, 1864–1875 (2008).
19. M. Granados, B. Ajdin, M. Wand, and C. Theobalt, "Optimal HDR reconstruction with linear digital cameras," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR* (IEEE, 2010), pp. 215–222.
20. K. Hirakawa and P. J. Wolfe, "Optimal exposure control for high dynamic range imaging," in *17th IEEE International Conference on Image Processing (ICIP)* (IEEE, 2010), pp. 3137–3140.
21. N. Gelfand, A. Adams, S. H. Park, and K. Pulli, "Multi-exposure imaging on mobile devices," in *Proceedings of the International Conference on Multimedia* (ACM, 2010), pp. 823–826.
22. S. W. Hasinoff, F. Durand, and W. T. Freeman, "Noise-optimal capture for high dynamic range photography," in *IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2010), pp. 553–560.
23. O. Gallo, M. Tico, R. Manduchi, N. Gelfand, and K. Pulli, "Metering for exposure stacks," Computer Graphics Forum **31**, 479–488 (2012).

24. B. Guthier, S. Kopf, and W. Effelsberg, "Optimal shutter speed sequences for real-time HDR video," in *IEEE International Conference on Imaging Systems and Techniques* (IEEE, 2012), pp. 303–308.
25. S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video," in *ACM Transactions on Graphics (TOG)* (ACM, 2003), Vol. **22**, pp. 319–325.
26. M. Gupta, D. Iso, and S. K. Nayar, "Fibonacci exposure bracketing for high dynamic range imaging," in *IEEE International Conference on Computer Vision* (IEEE, 2013), pp. 1473–1480.
27. M. D. Grossberg and S. K. Nayar, "What can be known about the radiometric response from images?" in *Computer Vision ECCV* (Springer, 2002), pp. 189–205.
28. E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting* (Morgan Kaufmann, 2010).
29. J. T. Bushberg and J. M. Boone, *The Essential Physics of Medical Imaging* (Lippincott Williams and Wilkins, 2011).

Appendix B

Extra data

Table B.1: Central wavelength and bandwidth provided for the 121 candidate filters for the *SpectrCam* (see section 5.6.1). All data shown in nanometers.

Filter #	λ	BW	Filter #	λ	BW	Filter #	λ	BW
1	402	16	42	542	10	83	724	41
2	406	11	43	542	70	84	728	41
3	420	20	44	550	100	85	740	10
4	420	100	45	555	50	86	740	20
5	426	100	46	560	20	87	750	10
6	426	60	47	570	17	88	750	100
7	426	20	48	570	50	89	750	102
8	428	60	49	570	20	90	750	16
9	440	16	50	575	100	91	768	102
10	440	26	51	575	28	92	770	102
11	440	20	52	578	10	93	776	100
12	460	16	53	581	20	94	780	10
13	460	12	54	600	20	95	780	20
14	460	20	55	601	13	96	784	62
15	460	16	56	606	34	97	790	10
16	466	21	57	609	38	98	800	10
17	468	10	58	610	20	99	820	20
18	470	20	59	616	100	100	833	125
19	476	20	60	616	17	101	838	28
20	476	100	61	620	10	102	840	40
21	476	60	62	620	20	103	846	10
22	476	80	63	626	50	104	860	44
23	480	10	64	627	48	105	860	46
24	480	26	65	640	102	106	860	100
25	482	68	66	640	20	107	866	30
26	486	108	67	646	17	108	868	120
27	500	20	68	660	100	109	880	10
28	500	27	69	660	10	110	880	20
29	508	20	70	660	68	111	876	60
30	508	10	71	660	20	112	876	36
31	511	20	72	667	17	113	900	20
32	515	10	73	669	10	114	906	60
33	520	20	74	676	100	115	913	10
34	525	35	75	680	10	116	926	60
35	525	50	76	680	60	117	940	20
36	530	48	77	680	20	118	960	180
37	530	24	78	700	10	119	960	100
38	532	18	79	700	20	120	972	10
39	535	50	80	704	100	121	980	20
40	538	82	81	708	43			

Appendix C

Glossary

- AEE Adaptive Exposure Estimation.
- AOTF Acousto-Optic Tunable Filter.
- API Application Programing Interface.
- B Blue.
- BAR Barakat.
- CRT Cathode Rays Tube.
- CCD Charge Coupled Device.
- CFA Color Filter Array.
- CIE Commission Internationale de l'Éclairage.
- CMOS Complementary Metal-Oxide-Semiconductor.
- CPU Control Process Unit.
- CRF Camera Response Function.
- DC Digital Counts.
- DoLP Degree of Linear Polarization.
- FWHM Full Width Half Maximum.

- G Green.
- GFC Goodness of Fit Coefficient.
- GPU Graphics Processor Unit.
- GT Ground Truth.
- HDR High Dynamic Range.
- HI High.
- HVS Human Visual System.
- IR-UV Infra Red - Ultra Violet.
- ISO International Standardization Organization.
- LCTF Liquid Crystal Tunable Filter.
- LDR Low Dynamic Range.
- LED Light Emitting Diode.
- LO Low.
- LRSS Low Resolution Spectral System.
- MIBS Minimal Image Bracketing Set.
- MS MultiSpectral.
- MSBS Minimal System Bracketing Set.
- MSFA MultiSpectral Filter Array.
- MSHDRPol MultiSpectral High Dynamic Range Polarimetric.
- MSHDRPolVISNIR Multispectral High Dynamic Range Polarimetric in the Visible and Near Infra Red.
- NCS Natural Color System.
- Ncut Normalized Cut.

- ND Neutral Density.
- NIR Near Infra Red.
- PCA Principal Component Analysis.
- PGP Prism Grating Prism.
- PR Photo Research.
- POCS Projection Onto Convex Sets.
- Pol Polarimetric.
- R Red.
- RBFNN Radial Basis Functions Neural Networks.
- RGB Red, Green and Blue.
- RMSE Root Mean Square Error.
- ROI Region Of Interest.
- SIFT Scale Invariant Feature Transform.
- SLM Spatial Light Modulator.
- SNR Signal to Noise Ratio.
- SPD Spectral Power Distribution.
- TFD Transverse Field Detector.
- TMO Tone-Mapping Operator.
- VIS Visible range.
- VPFA Voting Principal Feature Analysis.
- WS WaterShed.

Appendix D

List of figures and tables

List of Figures

List of Tables

Bibliography

[1] http://www.gphoto.org/.

[2] M. Abuleil and I. Abdulhalim. Narrowband multispectral liquid crystal tunable filter. *Optics Letters*, 41(9):1957–1960, 2016.

[3] M. Aggarwal and N. Ahuja. Split aperture imaging for high dynamic range. *International Journal of Computer Vision*, 58(1):7–17, 2004.

[4] A. O. Akyüz and E. Reinhard. Noise reduction in high dynamic range imaging. *Journal of Visual Communication and Image Representation*, 18(5):366–376, 2007.

[5] D. Alcaraz-Segura, J. Cabello, and J. Paruelo. Baseline characterization of major iberian vegetation types based on the ndvi dynamics. *Plant Ecology*, 202(1):13–29, 2009.

[6] N. Barakat, T. E. Darcie, and A. N. Hone. The tradeoff between snr and exposure-set size in hdr imaging. In *IEEE International Conference on Image Processing*, pages 1848–1851, 2008.

[7] N. Barakat, A. N. Hone, and T. E. Darcie. Minimal-bracketing sets for high-dynamic-range image capture. *IEEE Transactions on Image Processing*, 17(10):1864–1875, 2008.

[8] S. Beucher and F. Meyer. The morphological approach to segmentation: the watershed transformation. *Optical Engineering*, 34:433–481, 1992.

[9] R. C. Bilcu, A. Burian, A. Knuutila, and M. Vehvilainen. High dynamic range imaging on mobile devices. In *IEEE International Conference on Electronics, Circuits and Systems*, pages 1312–1315. IEEE, 2008.

[10] M. D. Buhmann. *Radial basis functions: theory and implementations*, volume 12. Cambridge university press, 2003.

[11] J. T. Bushberg and J. M. Boone. *The essential physics of medical imaging*. Lippincott Williams & Wilkins, 2011.

[12] A. Caspani, G. Langfelder, A. Longoni, E. Linari, and V. Tommolini. Analysis of a 64 x 64 matrix of direct color sensors based on spectrally tunable pixels. In *IS&T SPIE Electronic Imaging*, pages 90230O–90230O. International Society for Optics and Photonics, 2014.

[13] T. Chan, K. Jia, S.a Gao, J.n Lu, Z. Zeng, and Y. Ma. Pcanet: A simple deep learning baseline for image classification? *IEEE Transactions on Image Processing*, 24(12):5017–5032, 2015.

[14] S. Chandrasekhar. *Radiative transfer*. Dover Publications, 1960.

[15] I. Chang. Acousto-optic tunable filters. *Optical Engineering*, 20(6):206824–206824, 1981.

[16] I.S. Chatzis, V.A. Kappatos, and E.S. Dermatas. Filter selection for multi-spectral image acquisition using the feature vector analysis methods. In *I* PROMS Virtual Conference on Intelligent production machines and systems*, page 283. Elsevier Science, 2007.

[17] H. Chen and L. B. Wolff. Polarization phase-based method for material classification in computer vision. *International Journal of Computer Vision*, 28(1):73–83, 1998.

[18] T. Chen and A. El Gamal. Optimal scheduling of capture times in a multiple-capture imaging system. In *Electronic Imaging*, pages 288–296. International Society for Optics and Photonics, 2002.

[19] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE transactions on pattern analysis and machine intelligence*, 17(8):790–799, 1995.

[20] D. Ciregan, U. Meier, and J. Schmidhuber. Multi-column deep neural networks for image classification. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3642–3649. IEEE, 2012.

[21] P. Colantoni, R. Pillay, C. Lahanier, and D. Pitzalis. Analysis of multispectral images of paintings. In *14th European Signal Processing Conference*, pages 1–5. IEEE, 2006.

[22] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 24(5):603–619, 2002.

[23] Andover Corporation. https://www.andovercorp.com/.

[24] A. Darmont. *High Dynamic Range Imaging: Sensors and Architectures*. SPIE press, 2012.

[25] M. De Biasio, T. Arnold, R. Leitner, G. McGunnigle, and R. Meester. Uav-based environmental monitoring using multi-spectral imaging. In *SPIE Defense, Security, and Sensing*, pages 766811–766811. International Society for Optics and Photonics, 2010.

[26] P.E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH classes*, page 31, 2008.

[27] X. Delpueyo Español, M. Vilaseca Ricart, S. Royo Royo, M. Ares Rodríguez, F. Sanàbria Ortega, J. A. Herrera Ramírez, F. J. Burgos Fernández, J. Pujol Ramo, S. Puig, and G. Pellacani. Handheld hyperspectral imaging system for the detection of skin cancer. In *AIC conference Proceedings, Tokyo*, pages 385–390, 2015.

[28] J. Eckhard, T. Eckhard, E. M. Valero, J. L. Nieves, and E. Contreras. Outdoor scene reflectance measurements using a bragg-grating-based hyperspectral imager. *Applied Optics*, 54(13):D15–D24, 2015.

[29] T. Eckhard. *Design considerations for line-scan multi-spectral imaging systems*. PhD thesis, University of Granada, 2016.

[30] T. Eckhard, M. Klammer, E. M Valero, and J. Hernández-Andrés. Improved spectral density measurement from estimated reflectance data with kernel ridge regression. In *International Conference on Image and Signal Processing*, pages 79–86. Springer, 2014.

[31] T. Eckhard, E. M Valero, J. Hernández-Andrés, and V. Heikkinen. Evaluating logarithmic kernel for spectral reflectance estimation-effects on model parametrization,

training set size, and number of sensor spectral channels. *Journal of the Optical Society of America*, 31(3):541–549, 2014.

[32] T. Eckhard, E. M. Valero, J. Hernández-Andrés, and M. Schnitzlein. Adaptive global training set selection for spectral estimation of printed inks using reflectance modeling. *Applied Optics*, 53(4):709–719, 2014.

[33] G. ElMasry, D. W. Sun, and P. Allen. Near-infrared hyperspectral imaging for predicting colour, ph and tenderness of fresh beef. *Journal of Food Engineering*, 110(1):127–140, 2012.

[34] A. Ferrero, J. Campos, and A. Pons. Apparent violation of the radiant exposure reciprocity law in interline ccds. *Applied Optics*, 45(17):3991–3997, 2006.

[35] E. R. Fossum. Cmos image sensors: Electronic camera-on-a-chip. *IEEE Transactions on Electroninc Devices*, 44(10):1689–1698, 1997.

[36] A. Galdran, J. Vazquez-Corral, D. Pardo, and M. Bertalmío. Enhanced variational image dehazing. *SIAM Journal on Imaging Sciences*, 8(3):1519–1546, 2015.

[37] O. Gallo, N. Gelfandz, W. Chen, M. Tico, and K. Pulli. Artifact-free high dynamic range imaging. In *IEEE International Conference on Computational Photography*, pages 1–7. IEEE, 2009.

[38] O. Gallo, M. Tico, R. Manduchi, N. Gelfand, and K. Pulli. Metering for exposure stacks. In *Computer Graphics Forum*, volume 31, pages 479–488. Wiley Online Library, 2012.

[39] L. Gao, Robert T. K., N. Hagen, and T. S. Tkaczyk. Snapshot image mapping spectrometer (ims) with high sampling density for hyperspectral microscopy. *Optics Express*, 18(14):14330–14344, 2010.

[40] J. A. Garcıa, J. L. Nieves, E. Valero, and J. Romero. Stochastic independence of color-vision mechanisms confirmed by a subthreshold summation paradigm. *Journal of the Optical Society of America A*, 17(8):1485–1488, 2000.

[41] A. García-Beltrán, J. L. Nieves, J. Hernández-Andrés, and J. Romero. Linear bases for spectral reflectance functions of acrylic paints. *Color Research & Application*, 23(1):39–45, 1998.

[42] Yuval Garini, Ian T Young, and George McNamara. Spectral imaging: principles and applications. *Cytometry Part A*, 69(8):735–747, 2006.

[43] A. Gebejes, J. Orava, N. Penttinen, V. Heikkinen, J. Hiltunen, and M. Hauta-Kasari. Color and image characterization of a three ccd seven band spectral camera. In *International Conference on Image and Signal Processing*, pages 96–105. Springer, 2014.

[44] B. Geelen, N. Tack, and A. Lambrechts. A snapshot multispectral imager with integrated tiled filters and optical duplication. In *SPIE MOEMS-MEMS*, pages 861314–861314. International Society for Optics and Photonics, 2013.

[45] B. Geelen, N. Tack, and A. Lambrechts. A compact snapshot multispectral imager with a monolithically integrated per-pixel filter mosaic. In *SPIE MOEMS-MEMS*, pages 89740L–89740L. International Society for Optics and Photonics, 2014.

[46] N. Gelfand, A. Adams, S. H. Park, and K. Pulli. Multi-exposure imaging on mobile devices. In *Proceedings of the international conference on Multimedia*, pages 823–826. ACM, 2010.

[47] C. Godau, M. Klammer, T. Eckhard, M. Schnitzlein, D. Nowack, B. Frei, and P. Urban. Evaluation of a multi-spectral camera system for inline color measurement. In *Annual meeting of the German Colour Group*, 2013.

[48] A. Goshtasby. *Image registration: Principles, tools and methods*. Springer Science & Business Media, 2012.

[49] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H.P. Seidel, and H.P.A. Lensch. Optimal hdr reconstruction with linear digital cameras. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 215–222. IEEE, 2010.

[50] T. Grosch. Fast and robust high dynamic range image generation with camera and object movement. *Vision, Modeling and Visualization, RWTH Aachen*, pages 277–284, 2006.

[51] M. D Grossberg and S. K. Nayar. High dynamic range from multiple images: Which exposures to combine. In *Proceedings ICCV Workshop on Color and Photometric Methods in Computer Vision*, 2003.

[52] Michael D Grossberg and Shree K Nayar. What can be known about the radiometric response from images? In *Computer Vision ECCV*, pages 189–205. Springer, 2002.

[53] B.K. Gunturk, J. Glotzbach, Y. Altunbasak, R.W. Schafer, and R.M. Mersereau. Demosaicking: color filter array interpolation. *IEEE Signal Processing Magazine*, 22(1):44–54, 2005.

[54] Mohit Gupta, Daisuke Iso, and Shree K Nayar. Fibonacci exposure bracketing for high dynamic range imaging. In *IEEE International Conference on Computer Vision*, pages 1473–1480, 2013.

[55] N. Gupta. Acousto-optic tunable filters. *Optics and Photonics News*, 8(11):23–27, 1997.

[56] B. Guthier, S. Kopf, and W. Effelsberg. Optimal shutter speed sequences for real-time hdr video. In *IEEE International Conference on Imaging Systems and Techniques*, pages 303–308, 2012.

[57] B. Guthier, S. Kopf, and W. Effelsberg. A real-time system for capturing hdr videos. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 1473–1476. ACM, 2012.

[58] B. Guthier, S. Kopf, and W. Effelsberg. Algorithms for a real-time hdr video system. *Pattern Recognition Letters*, 34(1):25–33, 2013.

[59] R. J. Guttosch. Investigation of color aliasing of high spatial frequencies and edges for bayer-pattern sensors and foveon x3® direct image sensors. Technical report, Sigma, 2002.

[60] L. Hamers, Y. Hemeryck, G. Herweyers, M. Janssen, H. Keters, R. Rousseau, and A. Vanhoutte. Similarity measures in scientometric research: the jaccard index versus salton's cosine formula. *Information Processing & Management*, 25(3):315–318, 1989.

[61] H. Haneishi, S. Miyahara, and A. Yoshida. Image acquisition technique for high dynamic range scenes using a multiband camera. *Color Research & Application*, 31(4):294–302, 2006.

[62] A. Hård and L. Sivik. Ncs-natural color system: a swedish standard for coloer notation. *Color Research & Application*, 6(3):129–138, 1981.

[63] A. Hård, L. Sivik, and G. Tonnquist. Ncs, natural color system - from concept to research and applications. part i. *Color Research & Application*, 21(3):180–205, 1996.

[64] A. Hård, L. Sivik, and G. Tonnquist. Ncs, natural color system - from concept to research and applications. part ii. *Color Research & Application*, 21(3):206–220, 1996.

[65] J. Y. Hardeberg, F. Schmitt, and H. Brettel. Multispectral color image capture using a liquid crystal tunable filter. *Optical Engineering*, 41(10):2532–2548, 2002.

[66] J.Y. Hardeberg. *Acquisition and reproduction of color images: colorimetric and multispectral approaches*. Thesis Dissertation, 2001.

[67] J. A Hartigan and M. A. Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979.

[68] M. Hashimoto and J. Kishimoto. Two-shot type 6-band still image capturing system using commercial digital camera and custom color filter. In *Conference on Colour in Graphics, Imaging, and Vision*, volume 2008, pages 538–541. Society for Imaging Science and Technology, 2008.

[69] S. W Hasinoff, F. Durand, and W. T. Freeman. Noise-optimal capture for high dynamic range photography. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 553–560, 2010.

[70] G. Healey. Using color for geometry-insensitive segmentation. *Journal of the Optical Society of America A*, 6(6):920–937, 1989.

[71] V. Heikkinen. *Kernel methods for estimation and classification of data from spectral imaging*. PhD thesis, Faculty of Forestry and natural Sciences, University of Eastern Finland, 2011.

[72] V. Heikkinen, T. Jetsu, J. Parkkinen, M. Hauta-Kasari, T. Jaaskelainen, and S.D. Lee. Regularized learning framework in the estimation of reflectance spectra from camera responses. *Journal of the Optical Society of America A*, 24(9):2673–2683, 2007.

[73] V. Heikkinen, R. Lenz, T. Jetsu, J. Parkkinen, M. Hauta-Kasari, and T. Jääskeläinen. Evaluation and unification of some methods for estimating reflectance spectra from rgb images. *Journal of the Optical Society of America A*, 25(10):2444–2458, 2008.

[74] V. Heikkinen, A. Mirashemi, and A. Juha. Link functions and matérn kernel in the estimation of reflectance spectra from rgb responses. *Journal of the Optical Society of America A*, 30(11):2444–2454, Nov 2013.

[75] S. Helling, E. Seidel, and W. Biehlig. Algorithms for spectral color stimulus reconstruction with a seven-channel multispectral camera. In *Conference on Colour in Graphics, Imaging, and Vision*, volume 2004, pages 254–258. Society for Imaging Science and Technology, 2004.

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

[77] J. Hernández-Andrés, J. Romero, A. García-Beltrán, and J. L. Nieves. Testing linear models on spectral daylight measurements. *Applied Optics*, 37(6):971–977, 1998.

[78] J. Hernández-Andrés, J. Romero, and R. L. Lee. Colorimetric and spectroradiometric characteristics of narrow-field-of-view clear skylight in granada, spain. *Journal of the Optical Society of America A*, 18(2):412–420, 2001.

[79] E. Herrala and J. Okkonen. Imaging spectrograph and camera solutions for industrial applications. *International Journal of Pattern Recognition and Artificial Intelligence*, 10(01):43–54, 1996.

[80] D. Hertel, A. Betts, R. Hicks, and M. ten Brinke. An adaptive multiple-reset cmos wide dynamic range imager for automotive vision applications. In *IEEE Intelligent vehicles symposium*, pages 614–619. IEEE, 2008.

[81] B. Hill and F.W. Vorhagen. Multispectral image pick-up system. *US Patent*, 5(319):472, 1991.

[82] K. Hirakawa and P. M. Simon. Single-shot high dynamic range imaging with conventional camera hardware. In *International Conference on Computer Vision*, pages 1339–1346. IEEE, 2011.

[83] K Hirakawa and P. J. Wolfe. Optimal exposure control for high dynamic range imaging. In *IEEE International Conference on Image Processing*, pages 3137–3140, 2010.

[84] M. Hori, N. Osawa, K. Hirai, T. Horiuchi, and S. Tominaga. Spectral gigapixel imaging system for omnidirectional ooutdoor scene measurement. In *Proceedings of AIC conference, Tokyo*, pages 379–384, 2015.

[85] G. Horváth, R. Hegedüs, A. Barta, A. Farkas, and S. Åkesson. Imaging polarimetry of the fogbow: polarization characteristics of white rainbows measured in the high arctic. *Applied Otics*, 50(28):F64–F71, 2011.

[86] K. Huang and J. Chiang. Intelligent exposure determination for high quality hdr image generation. In *IEEE International Conference on Image Processing*, pages 3201–3205. IEEE, 2013.

[87] X. Huang, I. H. El-Sayed, W. Qian, and M. A. El-Sayed. Cancer cell imaging and photothermal therapy in the near-infrared region by using gold nanorods. *Journal of the American Chemical Society*, 128(6):2115–2120, 2006.

[88] P. M. Hubel. Foveon technology and the changing landscape of digital cameras. In *Color and Imaging Conference*, volume 2005, pages 314–317. Society for Imaging Science and Technology, 2005.

[89] P. M. Hubel, J. Liu, and R. J. Guttosch. Spatial frequency response of color image sensors: Bayer color filters and foveon x3. In *Electronic Imaging*, pages 402–407. International Society for Optics and Photonics, 2004.

[90] W. C. Hung, P. Sun, and R. Jiang. Hyperspectral relfectance reconstruction using a filter-based multispectral camera. In *Proceedins of AIC conference in Tokyo.*, pages 1350–1355. AIC, May 2015.

[91] G. R. Hunt. Spectral signatures of particulate minerals in the visible and near infrared. *Geophysics*, 42(3):501–513, 1977.

[92] C. Huynh and A. Robles-Kelly. A solution of the dichromatic model for multispectral photometric invariance. *International Journal of Computer Vision*, 90(1):1–27, 2010.

[93] J. Im, S. Lee, and J. Paik. Improved elastic registration for removing ghost artifacts in high dynamic imaging. *IEEE Transactions on Consumer Electronics*, 57(2):932–935, 2011.

[94] F. H Imai, M. R. Rosen, and R. S. Berns. Comparison of spectrally narrow-band capture versus wide-band with a priori sample analysis for spectral reflectance estimation. In *Color and Imaging Conference*, volume 2000, pages 234–241. Society for Imaging Science and Technology, 2000.

[95] F. H. Imai, M. R. Rosen, and R. S. Berns. Comparative study of metrics for spectral match quality. In *Conference on Colour in Graphics, Imaging, and Vision*, volume 2002, pages 492–496. Society for Imaging Science and Technology, 2002.

[96] B. F. Jones and P. Plassmann. Digital infrared thermal imaging of human skin. *IEEE Engineering in Medicine and Biology Magazine*, 21(6):41–48, 2002.

[97] S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. In *ACM Transactions on Graphics*, number 3, pages 319–325, 2003.

[98] S. Kavusi and A. El Gamal. Quantitative study of high-dynamic-range image sensor architectures. In *Electronic Imaging*, pages 264–275. International Society for Optics and Photonics, 2004.

[99] S. Kavusi, K. Ghosh, and A. El Gamal. *Architectures for high dynamic range, high speed image sensor readout circuits*. Springer, 2008.

[100] E. A. Khan, A. O. Akyuz, and E. Reinhard. Ghost removal in high dynamic range images. In *International Conference on Image Processing*, pages 2005–2008. IEEE, 2006.

[101] M. S Kim, J. E. McMurtrey, C. L. Mulchi, C. S. T. Daughtry, E. W. Chappelle, and Y. Chen. Steady-state multispectral fluorescence imaging system for plant leaves. *Applied Optics*, 40(1):157–166, 2001.

[102] M.S. Kim, Y. R. Chen, and P. M. Mehl. Hyperspectral reflectance and fluorescence imaging system for food quality and safety. *Transactions of the ASAE*, 44(3):721, 2001.

[103] F. König and W. Praefcke. Practice of multispectral image acquisition. In *SYBEN-Broadband European Networks and Electronic Image Capture and Publishing*, pages 34–41. International Society for Optics and Photonics, 1998.

[104] G. Können. Polarization and visibility of higher-order rainbows. *Applied Optics*, 54(4):B35–B40, 2015.

[105] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[106] T. Kunkel and E. Reinhard. A reassessment of the simultaneous dynamic range of the human visual system. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, pages 17–24. ACM, 2010.

[107] G. Langfelder. Spectrally reconfigurable pixels for dual-color-mode imaging sensors. *Applied Optics*, 51(4):A91–A98, 2012.

[108] G. Langfelder, C. Buffa, AF Longoni, A. Pelamatti, and F. Zaraga. Active pixels of transverse field detector based on a charge preamplifier. In *Proceedings of SPIE*, volume 8299, page 829906, 2012.

[109] G. Langfelder, AF Longoni, and F. Zaraga. Implementation of a multi-spectral color imaging device without color filter array. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7876, page 7, 2011.

[110] G. Langfelder, T. Malzbender, AF Longoni, and F. Zaraga. A device and an algorithm for the separation of visible and near infrared signals in a monolithic silicon sensor. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7882, page 5, 2011.

[111] G. Langfelder, F. Zaraga, and A. Longoni. Tunable spectral responses in a color-sensitive cmos pixel for imaging applications. *IEEE Transactions on Electron Devices*, 56(11):2563–2569, 2009.

[112] V. Lebourgeois, A. Bégué, S. Labbé, B. Mallavan, L. Prévot, and B. Roux. Can commercial digital cameras be used as multispectral sensors? a crop monitoring test. *Sensors*, 8(11):7300–7322, 2008.

[113] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen. Evaluation of tone mapping operators using a high dynamic range display. In *ACM Transactions on Graphics*, volume 24, pages 640–648. ACM, 2005.

[114] Y. Liu, Y. Chen, M. S. Kim, D. E. Chan, and A. M. Lefcourt. Development of simple algorithms for the detection of fecal contaminants on apples from visible/near infrared hyperspectral reflectance imaging. *Journal of Food Engineering*, 81(2):412–418, 2007.

[115] A. Longoni, F. Zaraga, G. Langfelder, and L. Bombelli. The transverse field detector (tfd): a novel color-sensitive cmos device. *IEEE Electron Device Letters*, 29(12):1306–1308, 2008.

[116] M. A. López-Álvarez, J. Hernández-Andrés, E. Valero, and J. L. Nieves. Colorimetric and spectral combined metric for the optimization of multispectral systems. In *Proceedings of the 10th Congress of the International Colour Association*, pages 1685–1688. Association Internationale de la Couleur, 2005.

[117] Y. Lu, I. Cohen, X.S. Zhou, and Q. Tian. Feature selection using principal feature analysis. In *Proceedings of the 15th international conference on Multimedia*, pages 301–304. ACM, 2007.

[118] D. L. MacAdam. Color-matching functions. In *Color Measurement*, pages 178–199. Springer, 1981.

[119] S. Mann and R. Picard. *Being undigital with digital cameras.* MIT Media Lab Perceptual, 1994.

[120] M. A. Martínez, E. Valero, V. Heikkinen, J., and G. Langfelder. Design of a multispectral system based on transverse field detectors. In *Proceedings 12th International AIC Congress.*, 2013.

[121] M. A. Martínez, E. Valero, J. Hernández-Andrés, and G. Langfelder. Spectral reflectance estimation from transverse field detectors responses. In *Sixth European Conference on Colour in Graphics, Imaging, and Vision.*, 2012.

[122] M. A. Martínez, E. M. Valero, and J. Hernández-Andrés. Adaptive exposure estimation for high dynamic range imaging applied to natural scenes and daylight skies. *Applied Optics*, 54(4):B241–B250, 2015.

[123] M. A. Martínez, E. M. Valero, J. Hernández-Andrés, J. Romero, and G. Langfelder. Combining transverse field detectors and color filter arrays to improve multispectral imaging systems. *Applied Optics*, 53(13):C14–C24, 2014.

[124] M. A. Martínez, E. M. Valero, J. Hernández-Andrés, S. Tominaga, T. Horiuchi, and K. Hirai. Image processing pipeline for segmentation and material classification based on multispectral high dynamic range polarimetric images. *Submitted to Patter Recognition (not published yet)*, 2017.

[125] J. J. McCann and A. Rizzi. *The art and science of HDR imaging*, volume 26. John Wiley & Sons, 2011.

[126] JJ McCann and A. Rizzi. Optical veiling glare limitations to in-camera scene radiance measurements. *ECVP Abstracts*, 35, 2006.

[127] JJ McCann and A. Rizzi. Spatial comparisons: The antidote to veiling glare limitations in hdr images. *Proceedings ADEAC/SID&VESA, Atlanta*, pages 155–158, 2006.

[128] J.J. McCann and A. Rizzi. Spatial comparisons: The antidote to veiling glare limitations in image capture and display. *Proceedings IMQA*, 2007.

[129] JJ McCann and A. Rizzi. Veiling glare: the dynamic range limit of hdr images. In *Electronic Imaging*, pages 649213–649213. International Society for Optics and Photonics, 2007.

[130] J.J. McCann and A. Rizzi. Camera and visual veiling glare in hdr images. *Journal of the society for information display*, 15(9):721–730, 2012.

[131] J. M. Medina, J. A. Díaz, and C. Vignolo. Fractal dimension of sparkles in automotive metallic coatings by multispectral imaging measurements. *ACS applied materials & interfaces*, 6(14):11439–11447, 2014.

[132] A. Mehnert and P. Jackway. An improved seeded region growing algorithm. *Pattern Recognition Letters*, 18(10):1065–1071, 1997.

[133] R. B. Merrill. Color separation in an active pixel cell imaging array using a triple-well structure, 1999. US Patent 5,965,875.

[134] T. Mitsunaga and S. K. Nayar. Radiometric self calibration. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1. IEEE, 1999.

[135] F. Montoliu, R.land Pla and A. C. Klaren. Illumination intensity, object geometry and highlights invariance in multispectral imaging. In *Pattern Recognition and Image Analysis*, pages 36–43. Springer, 2005.

[136] H. R Morris, C. C. Hoyt, and P. J. Treado. Imaging spectrometers for fluorescence and raman microscopy: acousto-optic and liquid crystal tunable filters. *Applied Spectroscopy*, 48(7):857–866, 1994.

[137] Y. Murakami, K. Ietomi, A. Tadano, M. Yamaguchi, and N. Ohyama. Comparison of spectral image reconstruction methods using multipoint spectral measurements. In *Conference on Colour in Graphics, Imaging, and Vision*, volume 2008, pages 591–596. Society for Imaging Science and Technology, 2008.

[138] Y. Murakami, K. Ietomi, M. Yamaguchi, and N. Ohyama. Maximum a posteriori estimation of spectral reflectance from color image and multipoint spectral measurements. *Applied Optics*, 46(28):7068–7082, 2007.

[139] Y. Murakami, A. Tanji, and M. Yamaguchi. Development of a low-resolution spectral imager and its application to hybrid-resolution spectral imaging. In *Proceedings 12Th International AIC Congress*, pages 363–366, 2013.

[140] Y. Murakami, M. Yamaguchi, and N. Ohyama. Piecewise wiener estimation for reconstruction of spectral reflectance image by multipoint spectral measurements. *Applied Optics*, 48(11):2188–2202, 2009.

[141] Y. Murakami, M. Yamaguchi, and N. Ohyama. Class-based spectral reconstruction based on unmixing of low-resolution spectral information. *Journal of the Optical Society of America A*, 28(7):1470–1481, 2011.

[142] Y. Murakami, M. Yamaguchi, and N. Ohyama. Hybrid-resolution multispectral imaging using color filter array. *Optics Express*, 20(7):7173–7183, 2012.

[143] G. K. Naganathan, L. M. Grimes, J. Subbiah, C. R. Calkins, A. Samal, and G. E. Meyer. Visible/near-infrared hyperspectral imaging for beef tenderness prediction. *Computers and electronics in agriculture*, 64(2):225–233, 2008.

[144] S.K. Nayar and V. Branzoi. Adaptive dynamic range imaging: Optical control of pixel exposures over space and time. In *Proceedings of Nineth IEEE International Conference on Computer Vision*, pages 1168–1175. IEEE, 2003.

[145] S.K. Nayar and T. Mitsunaga. High dynamic range imaging: Spatially varying pixel exposures. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 472–479. IEEE, 2000.

[146] J. L. Nieves, F. Pérez-Ocón, J. Hernández-Andrés, and J. Romero. Spectral-reflectance function recovery for improved colour-constancy experiments. *Displays*, 23(5):213–222, 2002.

[147] J. L. Nieves, E. M. Valero, S. M. C. Nascimento, J. Hernández-Andrés, and J. Romero. Multispectral synthesis of daylight using a commercial digital ccd camera. *Applied Optics*, 44(27):5696–5703, 2005.

[148] Color Imaging Laboratory. Department of Optics. University of Granada. http://colorimaginglab.ugr.es/.

[149] J. Park, M. Lee, M. D. Grossberg, and S. K. Nayar. Multispectral imaging using multiplexed illumination. In *IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007.

[150] G. Polder, G. Van der Heijden, and I. T. Young. Hyperspectral image analysis for measuring ripeness of tomatoes. In *ASAE International Meeting*, number 003089, 2000.

[151] P. Porral, P. Callet, P. Fuchs, T. Muller, and E. Sandré-Chardonnal. High dynamic, spectral, and polarized natural light environment acquisition. In *SPIE/IS&T Electronic Imaging*, pages 94030B–94030B. International Society for Optics and Photonics, 2015.

[152] R. Ramanath, W.E. Snyder, G.L. Bilbro, and W.A. Sander III. Demosaicking methods for bayer color arrays. *Journal of Electronic imaging*, 11:306, 2002.

[153] R. Real. Tables of significant values of jaccard's index of similarity. *Miscellania Zoologica*, 22(1):29–40, 1999.

[154] R. Real and J. M. Vargas. The probabilistic basis of jaccard's index of similarity. *Systematic biology*, 45(3):380–385, 1996.

[155] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.

[156] N. Ricquier and B. Dierickx. Pixel structure with logarithmic response for intelligent and flexible imager achitectures. *Microelectronic Engineering*, 19(1-4):631–634, 1992.

[157] M. A. Robertson, S. Borman, and R. L. Stevenson. Dynamic range improvement through multiple exposures. In *Proceedings of International Conference on Image Processing*, pages 159–163. IEEE, 1999.

[158] M. A. Robertson, S. Borman, and R. L. Stevenson. Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *Journal of Electronic Imaging*, 12(2):219–228, 2003.

[159] A. Rodríguez, J. L. Nieves, E. Valero, E. Garrote, J. Hernández-Andrés, and J. Romero. Modified fuzzy c-means applied to a bragg grating-based spectral imager for material clustering. In *IS&T/SPIE Electronic Imaging*, pages 83000J–83000J. International Society for Optics and Photonics, 2012.

[160] J. Romero, J. A. Garcia, E. Valero, and J. L. Nieves. Measurements of the spectral modulation sensitivity function for two normal observers with crt monitors. *Journal of Optics*, 28(5):190, 1997.

[161] J. Romero, A. Garcia-Beltran, and J. Hernández-Andrés. Linear bases for representation of natural and artificial illuminants. *Journal of the Optical Society of America A*, 14(5):1007–1014, 1997.

[162] Y. Y. Schechner and S. K. Nayar. Polarization mosaicing: high dynamic range and polarization imaging in a wide field of view. In *Optical Science and Technology, SPIE's 48th Annual Meeting*, pages 93–102. International Society for Optics and Photonics, 2003.

[163] H. Seetzen, W. Heidrich, W. Stuerzlinger, G. Ward, L. Whitehead, M. Trentacoste, A. Ghosh, and A. Vorozcovs. High dynamic range display systems. In *ACM Transactions on Graphics*, number 3, pages 760–768, 2004.

[164] G. Sharma, W. Wu, and E.N. Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, 30(1):21–30, 2005.

[165] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[166] R. Shrestha, J. Y. Hardeberg, and A. Mansouri. One-shot multispectral color imaging with a stereo camera. In *Proceedings SPIE*, volume 7876, page 787609, 2011.

[167] R. Shrestha, R. Pillay, S. George, and J. Y. Hardeberg. Quality evaluation in spectral imaging–quality factors and metrics. *Journal of the International Colour Association*, 12, 2014.

[168] P. Soille. *Morphological image analysis: principles and applications*. Springer Science & Business Media, 2013.

[169] Jia Song, Eva M Valero, and Juan L Nieves. Segmentation of natural scenes: Clustering in colour space vs. spectral estimation and clustering of spectral data. *JAIC-Journal of the International Colour Association*, 12, 2014.

[170] H. Stark, Y. Yang, and Y. Yang. *Vector space projections: A numerical approach to signal and image processing, neural nets, and optics*, volume 2. Wiley, 1998.

[171] D. Stoppa, A. Simoni, L. Gonzo, M. Gottardi, and G. F. Dalla Betta. Novel cmos image sensor with a 132-db dynamic range. *IEEE Journal of Solid-State Circuits*, 37(12):1846–1852, 2002.

[172] J. Stumpfel, C. Tchou, A. Jones, T. Hawkins, A. Wenger, and P. Debevec. Direct hdr capture of the sun and sky. In *Proceedings of the 3rd international conference on Computer graphics, virtual reality, visualisation and interaction in Africa*, pages 145–149. ACM, 2004.

[173] E. Talvala, A. Adams, M. Horowitz, and M. Levoy. Veiling glare in high dynamic range imaging. In *ACM Transactions on Graphics*, volume 26, page 37. ACM, 2007.

[174] A. Tomaszewska and R. Mantiuk. Image registration for multi-exposure high dynamic range image acquisition. In *Proceedings of International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, pages 49–56, 2007.

[175] S. Tominaga. Dichromatic reflection models for a variety of materials. *Color Research & Application*, 19(4):277–285, 1994.

[176] S. Tominaga. Multichannel vision system for estimating surface and illumination functions. *Journal of the Optical Society of America A*, 13(11):2163–2173, 1996.

[177] S. Tominaga and T. Horiuchi. Spectral imaging by synchronizing capture and illumination. *Journal of the Optical Society of America A*, 29(9):1764–1775, 2012.

[178] S. Tominaga, T. Horiuchi, and A. Yoshimura. An active illumination method for tristimulus image display. In *Color and Imaging Conference*, volume 2010, pages 160–165. Society for Imaging Science and Technology, 2010.

[179] S. Tominaga, H. Kadoi, K. Hirai, and T. Horiuchi. Metal-dielectric object classification by combining polarization property and surface spectral reflectance. In *IS&T/SPIE Electronic Imaging*, pages 86520E–86520E. International Society for Optics and Photonics, 2013.

[180] S. Tominaga and A. Kimachi. Polarization imaging for material classification. *Optical Engineering*, 47(12):123201–123201, 2008.

[181] S. Tominaga and R. Okajima. Object recognition by multi-spectral imaging with a liquid crystal filter. In *Proceedings of 15th International Conference on Pattern Recognition*, volume 1, pages 708–711. IEEE, 2000.

[182] Y. Tsin, V. Ramesh, and T. Kanade. Statistical calibration of ccd imaging process. In *Proceedings of Eighth IEEE International Conference on Computer Vision*, volume 1, pages 480–487. IEEE, 2001.

[183] M. Tsuchida, T. Kawanishi, K. Kashino, and J. Yamato. A stereo nine-band camera for accurate color and spectrum reproduction. In *ACM SIGGRAPH*, page 18. ACM, 2012.

[184] M. Tsuchida, T. Takahashi, K. Ito, T. Kawanishi, J. Yamato, and T. Aoki. A stereo one-shot multi-band camera system for accurate color reproduction. In *ACM SIGGRAPH*, page 66. ACM, 2010.

[185] M. Tsuchida, A. Takayanagi, Y. Sakaguchi, and R. Mukai. Estimation of spectral reflectance from six-band images based on partial least-squares regression. In *Proceedings 12Th International AIC Congress*, 2013.

[186] E. Valero, J. A. Garcia, J. L. Nieves, and J. Romero. Measurements of sensitivity to simulated chromatic frequencies for normal and dichromatic observers. *Journal of Optics*, 29(5):339, 1998.

[187] E. M. Valero, J. L. Nieves, S. Nascimento, K. Amano, and D. H. Foster. Recovering spectral data from natural scenes with an rgb digital camera and colored filters. *Color Research & Application*, 32(5):352–360, 2007.

[188] M. Varma and A. Zisserman. Classifying images of materials: Achieving viewpoint and illumination independence. In *European Conference on Computer Vision*, pages 255–271. Springer, 2002.

[189] K. Walls, Q. Chen, D. R. S. Cumming, and T. D. Drysdale. Fabry-pérot resonator with nanostructures for multispectral visible filtering. In *12th IEEE Conference on Nanotechnology*, pages 1–5. IEEE, 2012.

[190] B. A. Wandell. *Foundations of vision*. Sinauer Associates, 1995.

[191] X. Wang, W. Wong, and R. Hornsey. A high dynamic range cmos image sensor with inpixel light-to-frequency conversion. *IEEE Transactions on Electron Devices*, 53(12):2988, 2006.

[192] G. Ward. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *Journal of Graphics Tools*, 8(2):17–30, 2003.

[193] S. Wold, K. Esbensen, and P. Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

[194] L. B. Wolff. Polarization-based material classification from specular reflection. *IEEE transactions on pattern analysis and machine intelligence*, 12(11):1059–1071, 1990.

[195] S. T. Wu. Design of a liquid crystal based tunable electrooptic filter. *Applied Optics*, 28(1):48–52, 1989.

[196] G. Wyszecki and W. S. Stiles. *Color science*, volume 8. Wiley New York, 1982.

[197] K. Yamada, T. Nakano, S. Yamamoto, E. Akutsu, and I. Aoki. Wide dynamic range vision sensor for vehicles. In *Proceedings of Vehicle Navigation and Information Systems Conference.*, pages 405–408. IEEE, 1994.

[198] A. Yoshida, V. Blanz, K. Myszkowski, and H. P. Seidel. Perceptual evaluation of tone mapping operators with real-world scenes. In *Electronic Imaging*, pages 192–203. International Society for Optics and Photonics, 2005.

[199] F. Zaraga and G. Langfelder. White balance by tunable spectral responsivities. *Journal of the Optical Society of America A*, 27(1):31–39, 2010.

[200] X. Zhang, T. Sim, and X. Miao. Enhancing photographs with near infra-red images. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.

[201] Y. Zhao and R.S. Berns. Image-based spectral reflectance reconstruction using the matrix r method. *Color Research & Application*, 32(5):343–351, 2007.

[202] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000, 2003.