Snapshot multispectral image demosaicing and classification

Sofiane Mihoubi

UNIVERSITÉ DE LILLE

DOCTORAL THESIS
BY
SOFIANE MIHOUBI

Snapshot multispectral image demosaicing and classification

Jury:

- Pr. Pierre Chainais, École Centrale de Lille (jury president)
- Pr. Yannick Berthoumieu, IPB/ENSEIRB-Matmeca (reviewer)
- Dr. Sylvie Treuillet, Université d'Orléans (reviewer)
- Pr. Christine Fernandez-maloigne, Université de Poitiers (examinor)
- Dr. Jean-Baptiste Thomas, Université de Bourgogne (examinor)
- Pr. Ludovic Macaire, Université de Lille (supervisor)
- Dr. Olivier Losson, Université de Lille (co-supervisor)
- Dr. Benjamin Mathon, Université de Lille (co-supervisor)

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy

in the

CRIStAL laboratory
Color Image team

November 22, 2018

Acknowledgements

First, I would like to thank Pr. Ludovic Macaire, Dr. Olivier Losson and Dr. Benjamin Mathon for their investment and confidence in the supervision of my thesis, which guided me and teached me so much over the past three years.

I would also like to thank my committee members, professor Christine Fernandez, professor Yannick Berthoumieu, doctor Sylvie Treuillet, professor Pierre Chainais, and doctor Jean-Baptiste Thomas for serving as my committee members.

Further, I wish to express my gratitude to Jean-Baptiste Thomas, Jon Yngve Hardeberg, Haris Ahmad Khan and the NTNU laboratory team from Norway for welcoming me and for their contribution in our work.

I also thank my wife and my parents for encouraging me and supporting me during my PhD.

UNIVERSITÉ DE LILLE

Abstract

CRIStAL laboratory
Color Image team

Doctor of Philosophy

Snapshot multispectral image demosaicing and classification

by Sofiane MIHOUBI

Multispectral cameras sample the visible and/or the infrared spectrum according to specific spectral bands. Available technologies include snapshot multispectral cameras equipped with filter arrays that acquire raw images at video rate. Raw images require a demosaicing procedure to estimate a multispectral image with full spatio-spectral definition. In this manuscript we review multispectral demosaicing methods and propose a new one based on the pseudo-panchromatic image estimated directly from the raw image. We highlight the influence of illumination on demosaicing performances, then we propose pre- and post-processing normalization steps that make demosaicing robust to acquisition properties. Experimental results show that our method provides estimated images of better objective quality than classical ones and that normalization steps improve the quality of state-of-the-art demosaicing methods on images acquired under various illuminations.
Multispectral images can be used for texture classification. To perform texture analysis, local binary pattern operators extract texture descriptors from color texture images. We extend these operators to multispectral texture images at the expense of increased memory and computation requirements. We propose to compute texture descriptors directly from raw images, which both avoids the demosaicing step and reduces the descriptor size. For this purpose, we design a local binary pattern operator that jointly extracts the spatial and spectral texture information from a raw image. In order to assess classification on multispectral images we have proposed the first significant multispectral database of close-range textures in the visible and near infrared spectral domains. Extensive experiments on this database show that the proposed descriptor has both reduced computational cost and high discriminating power with regard to classical local binary pattern descriptors applied to demosaiced images.

UNIVERSITÉ DE LILLE

Résumé

CRIStAL laboratory
Color Image team

Docteur

Dématriçage et classification d'images multispectrales

Sofiane MIHOUBI

Les caméras multispectrales échantillonnent le spectre du visible et/ou de l'infrarouge selon des bandes spectrales spécifiques. Parmi les technologies disponibles, les caméras multispectrales snapshot équipées d'une mosaïque de filtres acquièrent des images brutes à cadence vidéo. Ces images brutes nécessitent un processus de dématriçage permettant d'estimer l'image multispectrale en pleine définition spatio-spectrale. Dans ce manuscrit nous examinons les méthodes de dématriçage multispectrale et proposons une nouvelle méthode basée sur l'utilisation d'une image panchromatique estimée directement à partir de l'image brute. De plus, nous mettons en évidence l'influence de l'illumination sur les performances de dématriçage, puis nous proposons des étapes de normalisation pré- et post-dématriçage rendant ce dernier robuste aux propriétés d'acquisition. Les résultats expérimentaux montrent que notre méthode fournit de meilleurs résultats que les méthodes classiques, et que les étapes de normalisation améliorent les performances de toutes les méthodes de l'état de l'art sur des images acquises sous différentes illuminations.
Les images multispectrales peuvent être utilisées pour la classification de texture. Afin d'effectuer une analyse de texture, nous considérons les opérateurs basés sur les motifs binaires locaux, qui extraient les descripteurs de texture à partir d'images couleur. Nous étendons ces opérateurs aux images de texture multispectrale au détriment d'exigences de mémoire et de calcul accrues. Nous proposons alors de calculer les descripteurs de texture directement à partir d'images brutes, ce qui évite l'étape de dématriçage tout en réduisant la taille du descripteur. Pour cela, nous concevons un opérateur de modèle binaire local qui extrait conjointement les informations de texture spatiale et spectrale d'une image brute. Afin d'évaluer la classification sur des images multispectrales, nous avons proposé la première base de données multispectrale de textures proches dans les domaines spectraux du visible et du proche infrarouge. Des expériences approfondies sur cette base montrent que le descripteur proposé a à la fois un coût de calcul réduit et un pouvoir de discrimination élevé en comparaison avec les descripteurs classiques appliqués aux images dématriçées.

Contents

List of Abbreviations

	Abreviations related to multispectral imaging
CFA	**C**olor **F**ilter **A**rray.
CIE	**C**ommission **I**nternationale de l'**É**clairage. (international commission on illumination).
CMOS	**C**omplementary **M**etal-**O**xide **S**emiconductor.
IMEC16	Snapshot camera from IMEC with 16 channels in the visible domain.
MSFA	**M**ulti**S**pectral **F**ilter **A**rray.
NIR	**N**ear **i**nfra**r**ed spectrum.
PSNR	**P**eak **S**ignal-to-**N**oise **R**atio.
RSPD	**R**elative **S**pectral **P**ower **D**istribution.
SPD	**S**pectral **P**ower **D**istribution.
sRGB	**s**tandard **R**ed **G**reen **B**lue.
SSF	**S**pectral **S**ensivity **F**unction.
Vis	**Vis**ible (spectral) domain.
VisNIR	**Vis**ible and **N**ear **i**nfra**r**ed (spectral) domain.
VISNIR8	Snapshot camera with 8 channels in the visible and near infrared domains.

	Abreviations related to demosaicing
BTES	**B**inary **T**ree **E**dge **S**ensing.
CSC	**C**onvolutional **S**parse **C**oding.
DCT	**D**iscrete **C**osine **T**ransform.
DWT	**D**iscrete **W**avelet **T**ransform.
DS	**D**own-**S**ampled.
GF	**G**uided **F**ilter.
GU	**G**aussian **U**p-sampling.
IC	**I**deal **C**amera.
IMEC16	Snapshot camera from IMEC with 16 channels in the visible domain.
IMEC25	Snapshot camera from IMEC with 25 channels in the near infrared domain.
ItSD	**It**erative **S**pectral **D**ifferences.
JBU	**J**oint **B**ilateral **U**p-sampling.
LMMSE	**L**inear **M**inimum **MSE**.
MLDI	**M**ultispectral **L**ocal **D**irectional **I**nterpolation.

MSE	**M**ean **S**quare **E**rror.
PP	**P**rior **P**robability.
PPBTES	**P**seudo-**P**anchromatic **B**inary **T**ree **E**dge **S**ensing.
PPDWT	**P**seudo-**P**anchromatic **D**iscrete **W**avelet **T**ransform.
PPI	**P**seudo-**P**anchromatic **I**mage.
PPID	**P**seudo-**P**anchromatic **I**mage **D**ifference.
SD	**S**pectral-**D**ifferences.
SSIM	**S**tructural **Sim**ilarity.
VIS5	Snapshot camera with 5 channels in the visible domain.
VISNIR8	Snapshot camera with 8 channels in the visible and near infrared domains.
VTV	**V**ectorial **T**otal **V**ariation.
WB	**W**eighted **B**ilinear.

	Abreviations related to classification

IOBLBP	**I**mproved **OBLBP**
LBP	**L**ocal **B**inary **P**attern
LCC	**L**ocal **C**olor **C**ontrast
maLBP	**m**ap **a**dder-based **LBP**
mdLBP	**m**ap **d**ecoder-based **LBP**
MLBP	**M**SFA-based **LBP**
OBLBP	**O**pponent **B**and **LBP**

List of Symbols

	Symbols related to multispectral image formation
λ	Wavelength.
$E(\lambda)$	Relative spectral power distributions of illumination.
Ω	Spectral domain.
p	Pixel of a multispectral image.
s	Surface element observed by a pixel.
w	Observed surface element of a white diffuser.
$R_s(\lambda)$ or $R_p(\lambda)$	Spectral reflectance at λ emitted by s or recieved by p.
$\mathcal{T}^k(\lambda)$	CIE XYZ matching function $k \in \{X, Y, Z\}$.
O	Bit depth.
$Q(\cdot)$	Quantization function on O bits.
	Symbols related to multispectral images
x, y, λ	Axes of a multispectral image.
X, Y, K	Size of a multispectral image according to x, y, λ axes.
k	Channel or spectral band index.
$T^k(\lambda)$	Spectral sensitivity function of band k.
I^k/R^k	Channel of index k of a multispectral radiance/reflectance image.
$\mathbf{I}/\mathbf{R}$	Multispectral radiance/reflectance image with K channels.
I^k_p/R^k_p	Value of channel I^k/R^k at pixel p of a radiance/reflectance image.
λ^k	Central wavelength of channel I^k.
$\mathcal{B}(*)$	Central wavelengths of bands sampled by camera *.
	Symbols related to simulation validation
$\mathbf{acq}_i$	Average value of channel acq^k_i of an acquired ColorChecker patch i.
$\mathbf{sim}_i$	Average value of channel sim^k_i of a simulated ColorChecker patch i.
$(\mathbf{a}, \mathbf{b})$	Couple of vectors that minimizes the square residuals.
	Symbols related to correlation measures
δ_x	Spatial distance according to x axis.
$C[\cdot]$	Pearson correlation coefficient.
μ_C	Correlation average.
σ_C	Correlation Standard deviation.
μ^k	Mean value of channel I^k.

	Symbols related to demosaicing problem
S	Set of all pixels of a multispectral image.
S^k	Pixel subset where the MSFA samples the band k.
I^{raw}	Raw image.
$MSFA(p)$	Function that associates a spectral band index to a pixel p.
$\tilde{I}^k$	Sparse channel of index k.
m^k	Binary mask that samples the pixels in S^k.
$\hat{\mathbf{I}}$	Estimated multispectral image.
$\hat{I}^k$	Estimated channel of index k.
$\hat{I}^k_{WB}$	Estimated channel of index k by bilinear interpolation.
	Symbols related to demosaicing methods
$\hat{G}$	Guide image.
H	Bilinear filter.
F	Unnormalized bilinear filter.
$c_F(a,b)$	Normalization coefficient of element (a,b) of filter F.
$\tilde{\Delta}^{k,l}$	Sparse difference channel (between I^k and I^l) at pixels in S^k.
$\hat{\Delta}^{k,l}$	Estimated difference channel (between I^k and I^l).
$\hat{S}^k(t)$	Estimated pixel subset at step t in I^k.
$\dot{S}^k(t)$	Available or previously estimated pixel subset at step t in I^k.
q	Neighbor of pixel p.
$\tilde{N}_p(t)$	Subset of the four closest neighbors of p that belong to $\dot{S}^k(t)$.
α_q	Weight associated to neighbor q in BTES or PPBTES methods.
β_q	Weight associated to neighbor q in MLDI method.
$\mathcal{N}^{P,d}$	Support made of the P neighbors at distance d from p.
N_p	Neighborhood associated to a support $\mathcal{N}^{P,d}$.
$\dot{I}^k_q$	Pixel value available in I^{raw} or previously estimated.
$\dot{D}^{k,l}_q$	Directional difference computed at $q \in \dot{S}^k(t)$.
$\hat{D}^{k,l}_q$	Difference value between channels of index k and l at q.
	Symbols related to PPI-based demosaicing methods
I^{PPI}	Pseudo-panchromatic image (PPI).
$\bar{I}^{PPI}$	Pre-estimated PPI.
$\hat{I}^{PPI}$	Estimated PPI.
M	Averaging filter for PPI estimation.
γ_q	Weight used for PPI estimation.
κ	Coefficient used for γ_q computation.
$\Delta_p[\cdot]$	Local difference.
H_p	Adaptive bilinear filter that depends of the neighborhood of p.
Γ_p	Filter composed of local directional weights.

	Symbols related to demosaicing assessment
$PSNR[\cdot]$	Peak signal-to-noise ratio between two images.
S_c	Subset of central pixels of an image.
$DeltaE^*_{ab}[\cdot]$	Color difference measurement between two images.
$\Delta E_{00}[\cdot]$	Extended color difference measurement.
	Symbols related to scale adjustment
I'^{raw}	Scale-adjusted raw image.
$\hat{\mathbf{I}}'$	Scale-adjusted estimated image.
ρ^k_*	Normalization factor based on acquisition properties.
	Symbols related to classification scheme
$Sim[\cdot]$	Similarity between two images.
h	Concatenated LBP histograms.
$s(\cdot)$	Unit step function.
$\epsilon(q)$	Index of each neighbor q of a pixel p.
$M_n(p)$	Raw type-I moment of order n at pixel p.
$\mu_n(p)$	Central type-I moment of order n at pixel p.
$\hat{M}_n(p)$	Raw type-II moment of order n at pixel p.
$\hat{\mu}_n(p)$	Central type-II moment of order n at pixel p.
$r_p(\lambda^k)$	Reflectance normalized by its L_1-norm.
$m_n(p)$	Raw type-I moment computed using normalized reflectance.
$\hat{m}_n(p)$	Raw type-II moment computed using normalized reflectance.
$\bar{\mathbf{I}}_p$	Average value of the neighbors of a pixel p.
$\theta^{k,l}$	Angles (represented as an image) associated with bands k and l.
B^k	Set of bands associated with pixels in N_p according to the MSFA.
$N^{k,l}_p$	MSFA-based neighborhood.
	Local binary pattern operators
$LBP^k(p)[\cdot]$	Marginal LBP operator associated to a channel I^k.
$\{maLBP^m\}[\cdot]$	m-th adder-based LBP operator.
$\{mdLBP^n\}[\cdot]$	n-th decoder-based LBP operator.
$LCC[\cdot]$	LCC operator.
$OBLBP^{k,l}[\cdot]$	Opponent band (k and l) LBP operator.
$IOBLBP^{k,l}[\cdot]$	Improved opponent band LBP operator.
$MLBP[\cdot]$	MSFA LBP operator.
	Operators
$\|\|\cdot\|\|$	Euclidean norm operator.
$\odot$	Element-wise product.
$*$	Convolution operator.
$\lfloor\cdot\rfloor$	Floor function.

$\lfloor \cdot \rceil$	Nearest integer function.
$\lceil \cdot \rceil$	Ceiling function.
$\equiv$	Congruence relation.
$\langle \cdot, \cdot \rangle$	Inner product.
$\bigoplus$	Direct sum.

Introduction

Multispectral cameras are sensitive to the spectral radiance that characterizes a combination between the spectral properties of the illumination and of the material observed at each surface element of a scene. They usually sample the radiance according to many narrow *spectral bands* in the *visible* and *near infrared* domains, and provide multispectral images that represent the scene radiance of each spectral band as a separate *channel*. Such images can be analyzed in art studies [20, 42]. For instance, by considering the spectra associated to pigments of Mona Lisa painting, Elias and Cotte [20] obtain a virtual removal of the varnish. This allows to identify the pigment used by Leonardo da Vinci and to learn more about his techniques and means. Besides art studies, multispectral images are analyzed in various application fields such as medical imaging [64, 90], precision agriculture [31], vision inspection for food quality control [61, 93] or waste sorting [100]. Specifically, Qin et al. [93] show that the analysis of channels associated with narrow bands in the visible spectral domain is beneficial for automatic safety and quality evaluation of food and agricultural products. Beyond the visible domain, multispectral images often also provide information in the near infrared domain, which is valuable for material classification [103, 111] and the identification of textile fibers [28] and minerals [40].
To evaluate properties of multispectral images, many databases are acquired using *multishot* devices, *i.e.*, systems that acquire one spectral band or one single pixel row at a time [4]. Among these systems, *linescan* devices acquire one row of pixels for all channels at a time, so that a spatial scanning is required to provide the fully-defined multispectral image. The limited number of texture databases leads us to **the acquisition of our own database of texture images (HyTexiLa) in collaboration with the Norwegian Colour and Visual Computing laboratory (Gjøvik, Norway)**. This database acquired using a linescan camera is useful to study the spectral properties of materials or for texture classification [46]. However, this technology can only observe still scenes.

Recently *snapshot* multispectral cameras have emerged to acquire all spectral bands in a single shot [24, 32]. Most of them use a *multispectral filter array* (MSFA) [10, 115] laid over the sensor that spectrally samples the incident light, like the widely-used Bayer color filter array (CFA) in color imaging. The MSFA achieves a compromise between spatial and spectral samplings. It is defined by a basic periodic pattern in which each filter is sensitive to a narrow spectral band. Each pixel of the resulting *raw image* is then characterized by one single band according to the MSFA

and the other missing values have to be estimated to recover the full spectral definition. This process known as *demosaicing* or *demosaicking* is similar in its principle to the estimation of missing RGB components in raw images captured by single-sensor color cameras fitted with a Bayer CFA. CFA demosaicing is a well-studied problem for more than forty years [57], but MSFA demosaicing is a recent subject with new issues.

Demosaicing relies on two properties of scene radiance image, namely *spatial correlation* between spatially close pixels within a channel, and *spectral correlation* between levels of different components at the same pixel [73]. In order to compare the different multispectral demosaicing solutions, we study the state of the art methods that rely on these two properties. **We then propose a demosaicing method based on the *pseudo-panchromatic image* (PPI) defined as the average image over all channels of a multispectral image [14]. For this purpose, we estimate the spatially fully-defined PPI from the raw image** [73]. To assess the performances, the different methods are then implemented and compared together by considering a 16-band MSFA (called IMEC16) incorporated in a system available at IrDIVE platform[1].

By studying IMEC16 images under various illuminations, we show that the *illumination* and *camera spectral sensitivity functions* (SSFs) strongly affect spectral correlation and demosaicing performances. Indeed, the information available in each channel of a multispectral image results from a spectral integration of the product between the scene *reflectance*, illumination, and camera SSFs. We therefore **propose normalization steps that adjust channel levels before demosaicing, which improves demosaicing robustness to illumination variations and camera SSFs**. These steps use normalization factors that either depend on the camera spectral sensitivity only, on both the sensitivity and the illumination, or on the statistics extracted from the acquired raw image [74].

To classify demosaiced raw texture images acquired by single-sensor snapshot cameras, the classical approaches extract texture *features* from each demosaiced image and compare them thanks to a *similarity measure*. In this manuscript, we focus on texture features based on *local binary patterns* (LBPs). The many variants of LBP operators have indeed proved to be very efficient for a wide variety of applications [91]. LBP-based texture classification has first been performed on gray-level images since the original operator only uses the spatial information of texture [86]. Later, Palm [88] has shown that classification based on a color analysis outperforms that based on the spatial information only. Texture feature extraction is then extended to the color domain by taking both spatial and spectral textural information into account. The recent advances in multispectral imaging lead us to extend the color texture features to the multispectral domain [46], but the computational cost significantly increases with the number of channels. Thus **we propose a new computationally-efficient LBP-based feature that is directly computed from raw images, which allows us to avoid the demosaicing step.** Extensive experiments

[1] http://www.irdive.fr

on HyTexiLa database of multispectral texture images prove the relevance of our approach [75].

This manuscript is organized as follows. In Chapter 1, we define the formation model of multispectral images and describe the different technologies that are available to acquire them. We then present the different multispectral image databases proposed in the litterature and HyTexiLa database that we have built. Multispectral databases are useful to simulate multispectral images according to the image formation model. We then simulate images that would be acquired using IMEC16 camera in order to study its properties related to demosaicing. Chapter 2 focuses on the snapshot technology that uses an MSFA. We propose a state of the art of demosaicing methods and focus on the methods suitable for IMEC16 MSFA. After highlighting their limitations, we propose our own method based on the use of an estimated PPI. Demosaicing methods are assessed in Chapter 3, which highlights the effect of illumination and camera SSFs on demosaicing performances. We therefore propose normalization steps that make demosaicing robust against acquisition conditions. Demosaiced images are used in Chapter 4 in order to study the different LBP descriptors used for texture classification. These descriptors that combine spatial and spectral information are extended from the color to the multispectral domain. Their discrimination powers are then compared with that of our proposed LBP-based descriptor in order to show the relevance of our approach.

Chapter 1

Multispectral images

Contents

1.1 Introduction

This chapter presents the formation, the acquisition, and the properties of a multispectral image. Section 1.2 explains how such image results from the observation of an illuminated object and represents a combination between the spectral properties of the illumination and of the object. A multispectral image can be acquired using various technologies described in Section 1.3, which lead to the acquisition of several image databases. In order to characterize an object regardless of the illumination, some authors propose to estimate its spectral reflectance from an acquired multispectral image. Similarly, in Section 1.4, we propose our own multispectral image database that describes spectral reflectance of different textural objects . Reflectance databases are used in Section 1.5 to simulate the acquisition, by a given camera, of multispectral images that represent the objects of the database. This simulation model is then validated with a multispectral camera that is available at IrDIVE platform. Finally, we use the simulation model to study the spatial and spectral properties of multispectral images in Section 1.6.

1.2 From illumination to multispectral image

We first introduce the notion of illumination in Section 1.2.1. Then we present the reflected radiance that comes from an illuminated object in Section 1.2.2. Finally we define the multispectral image that results from the radiance captured by a multispectral digital camera in Section 1.2.3.

1.2.1 Illuminations

An illumination is a light source that emits electromagnetic waves characterized by a linear oscillation of photons spreading at the speed of light. This oscillation can be represented by a wavelength λ in nanometers (nm), defined in free space by the ratio between the speed of light and the oscillation frequency. An illumination is characterized by its spectral power distribution (SPD) that represents the amount of photons emitted by unit surface per unit solid angle (in $W \cdot sr^{-1} \cdot m^{-2}$) with respect to the wavelength. Light sources are often characterized by their relative SPD (RSPD) $E(\lambda)$ (with no unit), obtained as their SPD in a certain spectral domain Ω normalized by that of a reference wavelength. In this manuscript, we consider the visible (Vis) and the near infrared (NIR) spectral domains that are defined respectively as $\Omega_{Vis} = [400\,nm, 700\,nm]$ and $\Omega_{NIR} = [700\,nm, 1000\,nm]$ (see Fig. 1.1).

The most commonly used illuminations have been normalized by the International Commission on Illumination (Commission Internationale de l'Éclairage, CIE)

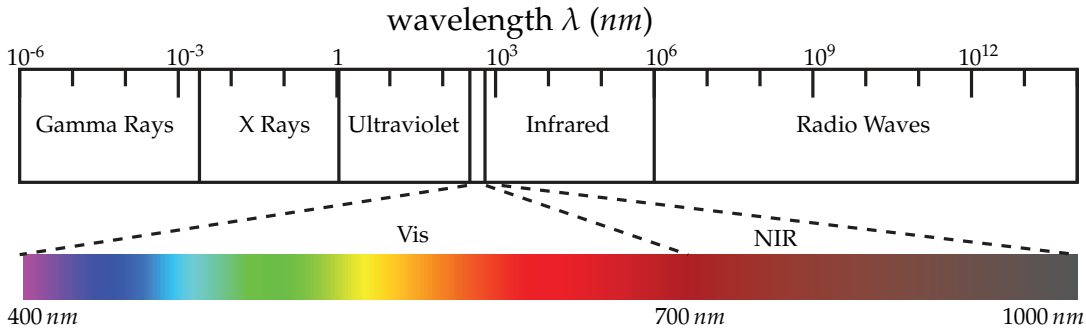


FIGURE 1.1: Visible and near infrared domains in the electromagnetic spectrum.

and are called illuminants [98]. These illuminants are arranged according to six categories prefixed by a letter from A to F. In this manuscript, we only consider the four most representative ones, namely, A, D65, E, and F12 illuminants. The A illuminant is used as a theoretical reference to represent the typical domestic tungsten-filament lighting. Its RSPD is that of a black body whose temperature is about 2856 K. The D illuminants have made B and C illuminants obsolete for daylight simulations. Among D illuminants, we focus on D65 that is considered as a reference by the International Organization for Standardization (ISO). D65 illuminant approximately corresponds to the average midday light in North/Western Europe, made of both direct sunlight and of the light diffused by a clear sky. Its RSPD is that of a black body whose temperature is close to 6500 K. The E illuminant consists in an equal-energy emission for all wavelengths. Although this illuminant is only theoretical, it can be used as a reference. The F illuminants represent different fluorescent lamps of various compositions. The RSPDs of illuminants F10, F11 and F12 present three narrow peaks in the visible domain. We focus on F12 illuminant that is originally designed for commercial lighting applications and commonly used in multispectral imaging [81].

Most illuminants are theoretical and are therefore difficult to reproduce in practice. Hence, we also consider two real illuminations, namely HA and LD. HA is composed of a set of Paulmann 2900 K halogen lamps whose behavior is similar to the A illuminant, while LD is produced by the Advanced Illumination DL097-WHIIC LED diffuse dome and has three peaks in the visible range. We have acquired the RSPDs of these two illuminations in the Vis domain using the Avantes spectrometer AvaSpec-3648[1] available at the IrDIVE platform in Tourcoing, France.

The RSPDs of the four illuminants and two illuminations are available with a step of 1 nm (F12 being linearly interpolated from 5 nm to 1 nm). In order to compare illuminants and real illuminations, they are normalized so that the maximum of each illumination in the visible domain reaches 1. The resulting RSPD $E(\lambda)$ of each of the four illuminants A, D65, E, and F12, and the two real illuminations HA and LD for all wavelengths in the visible domain are shown in Fig. 1.2.

[1] http://www.wacolab.com/avantes/spectrometers14.pdf

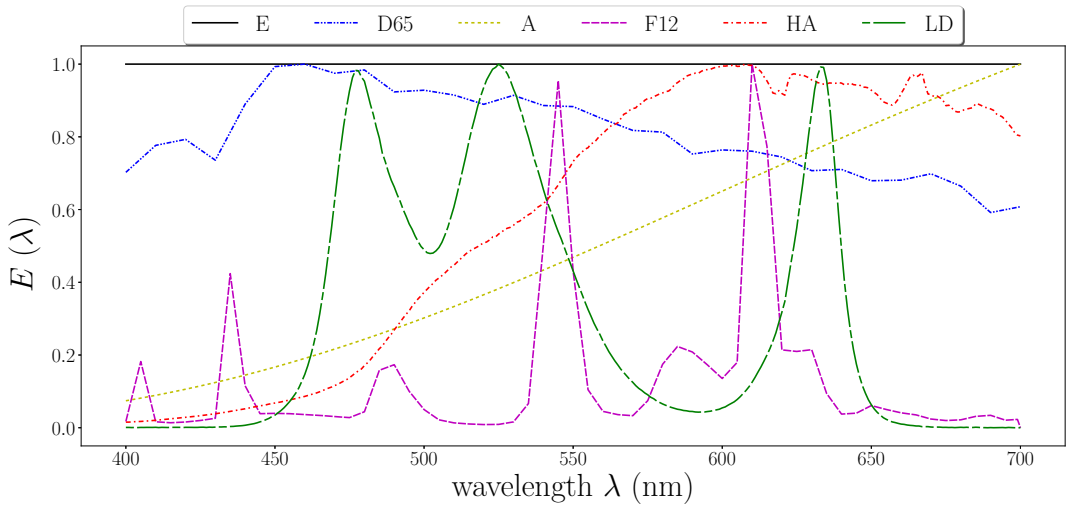


FIGURE 1.2: RSPDs of CIE E, D65, A, and F12 illuminants, and of real illuminations HA and LD in the Vis domain.

Beyond the visible domain, multispectral imaging often also considers the near infrared domain [97]. In this case, the CIE standard illuminants and our two illuminations cannot be used because the NIR part of the spectrum is not described. Hence, Thomas et al. [115] have computed or measured alternative illuminations. They provide the measures of the solar emission at the ground level, of a D65 simulator, and of a practical tungsten realization of A illuminant in the visible and near infrared (VisNIR) domain. They also extend E and A illuminants from the Vis domain to the VisNIR domain. Fig. 1.3 shows the RSPDs of these illuminations for all $\lambda \in \Omega_{VisNIR} = [400\,nm, 1000\,nm]$.

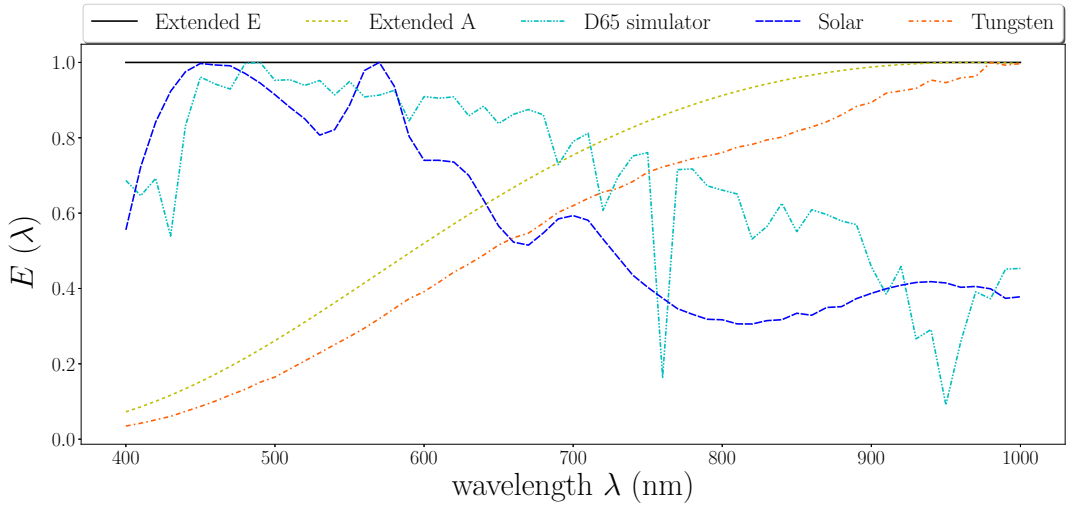


FIGURE 1.3: RSPDs of extended E and A illuminants and of measured solar, D65 simulator and tungsten illuminations in the VisNIR domain.

1.2.2 Reflected radiance

In contact with an object, the incident illumination is modified according to the spectral reflectance of material and reflected into two different ways. Specular reflection occurs when photons fall on a small (mirror-like) surface and is characterized by a reflection angle equal to the incident angle. Diffuse reflection is the scattering of photons according to many angles when it falls on a (microscopically) rough surface. In this manuscript we consider the surface of an object as Lambertian. Thus, materials that exhibit specular reflection are avoided, and we consider only diffuse reflection (and illumination), so that the reflected radiance does not depend on the angle of view [51].

The radiance function reflected by a surface element s of a material is defined by the product between its reflectance function $R_s(\lambda)$ and the illumination RSPD $E(\lambda)$ as shown in Fig. 1.4 [21]. The spectral reflectance of a material is usually normalized between 0.0 and 1.0 and depends on the pigments of which the material is made. In the VisNIR domain, a white diffuser and a black chart are used as references to characterize the reflectance. The pigments of a black chart absorb photons of all wavelengths ($R_s(\lambda) = 0$ for all $\lambda \in \Omega$) while a perfect white diffuser reflects them all ($R_s(\lambda) = 1$ for all $\lambda \in \Omega$).

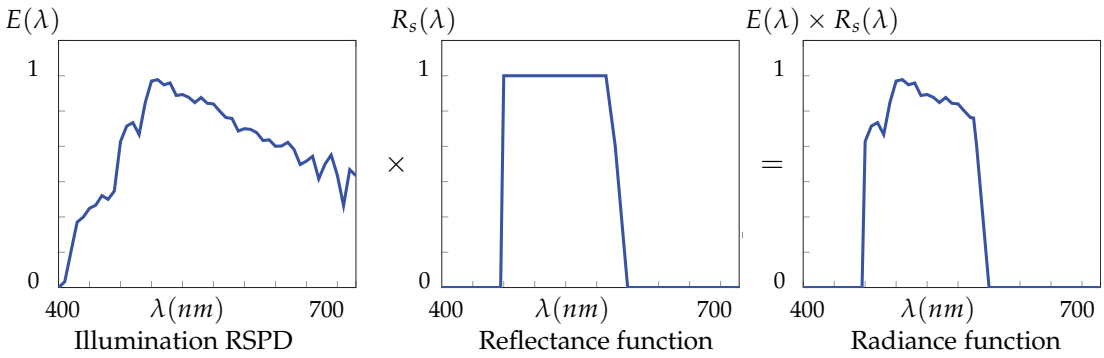


FIGURE 1.4: Computation of the radiance function from the illumination RSPD $E(\lambda)$ and the reflectance function $R_s(\lambda)$.

1.2.3 Multispectral image

The radiance that comes from a surface element in a given direction can be observed by a digital camera. The camera embeds lenses that focus the radiance and an aperture that controls the amount of photons incoming on the photosensitive surface. This surface is composed of a grid of sites which converts the amount of received photons into an electronic signal that is then digitized in binary coding by an electronic device. Thus, the resulting digital image is spatially discretized into a two-dimensional matrix of $X \times Y$ picture elements called pixels. To a pixel p is associated a value that represents the quantity of photons emitted by a surface element s of the scene in a given range of the spectrum called a spectral band. Images are

primarily acquired in the panchromatic spectral band that corresponds to the camera photosensitive surface sensitivity $T(\lambda)$. For illustration purpose, Fig. 1.5a shows the sensitivity of the SONY IMX174[2]'s complementary metal-oxide-semiconductor (CMOS) photosensitive surface [115].

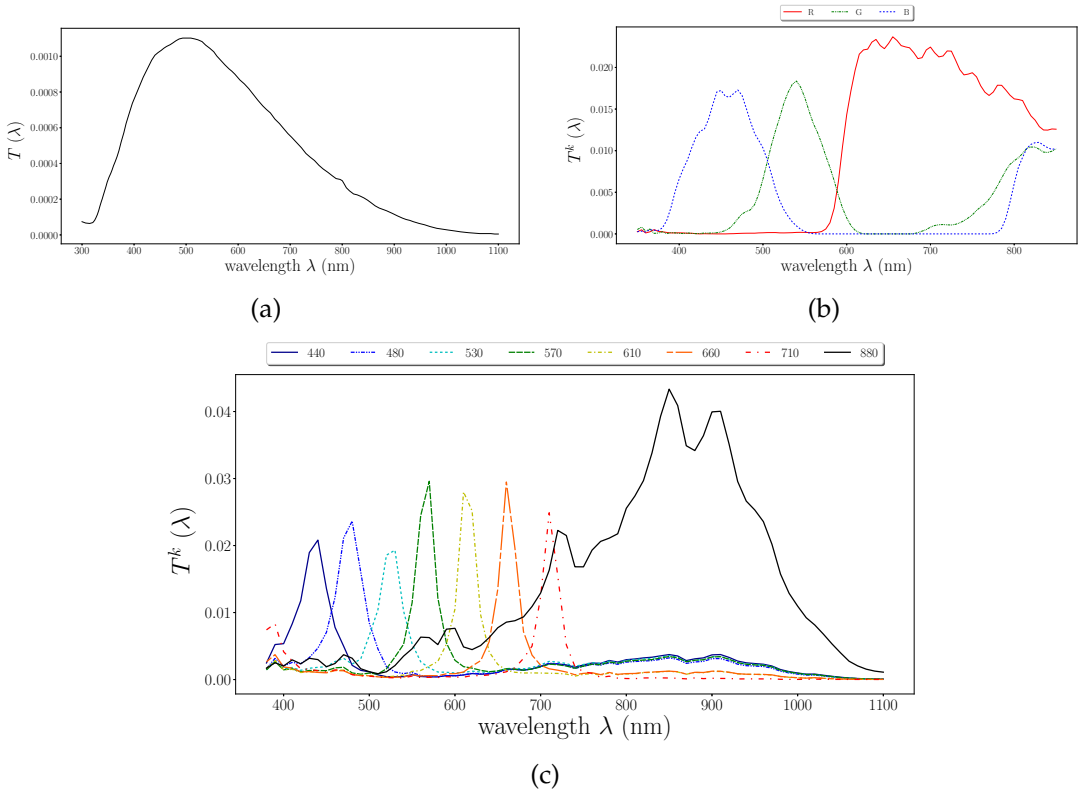


FIGURE 1.5: Normalized spectral sensitivity function $T(\lambda)$ of the SONY IMX174 CMOS photosensitive surface (a), of the Basler L301kc color camera (b), and of a multispectral camera with 8 bands in the VisNIR domain (c) [115]. Captions of (c) are band center wavelengths.

Inspired by the colorimetry, a digital color camera samples the visible domain according to three spectral bands, each being characterized by the spectral sensitivity function (SSF) $T(\lambda)$ of a band-pass filter. For illustration purpose, Fig. 1.5b shows the three SSFs of the Basler L301kc[3] color camera. Note that the SSF of a band-pass filter is not constant along its spectral range and overlaps with the SSFs of the other filters. A color camera acquires a color image composed of three channels, each one being associated to a spectral band red (R), green (G), or blue (B) according to the band-pass filters (see Fig. 1.5b). A multispectral image is more generally composed of K spectral channels, $K > 3$, whose associated filters sample the Vis, the NIR or the VisNIR domain. Each spectral channel I^k, $k \in \{1, \ldots, K\}$, is associated to the central wavelength λ^k of its SSF $T^k(\lambda)$ as illustrated in Fig. 1.5c. Note that a multispectral image with a high number of channels may be referred to as a hyperspectral image.

[2] https://eu.ptgrey.com/support/downloads/10414
[3] https://www.baslerweb.com/en/products/cameras/line-scan-cameras/l300/l301kc/

Since no consensus exists about the number of channels that makes the difference, we stick to the multispectral adjective whatever the number of channels.

1.3 Multispectral image acquisition

We first define the formation model of a multispectral radiance image in Section 1.3.1 on grounds of definitions introduced in Section 1.2. Then we briefly present the available technologies to acquire such images in Section 1.3.2. Finally we describe the various radiance databases that have been acquired using these technologies in Section 1.3.3.

1.3.1 Multispectral image formation model

Let us consider that a multispectral image is composed of K spectral channels and denote it as $\mathbf{I} = \{I^k\}_{k=1}^K$. Assuming ideal optics and homogeneous spectral sensitivity of the sensor, the value I_p^k of channel I^k at pixel p can be expressed as:

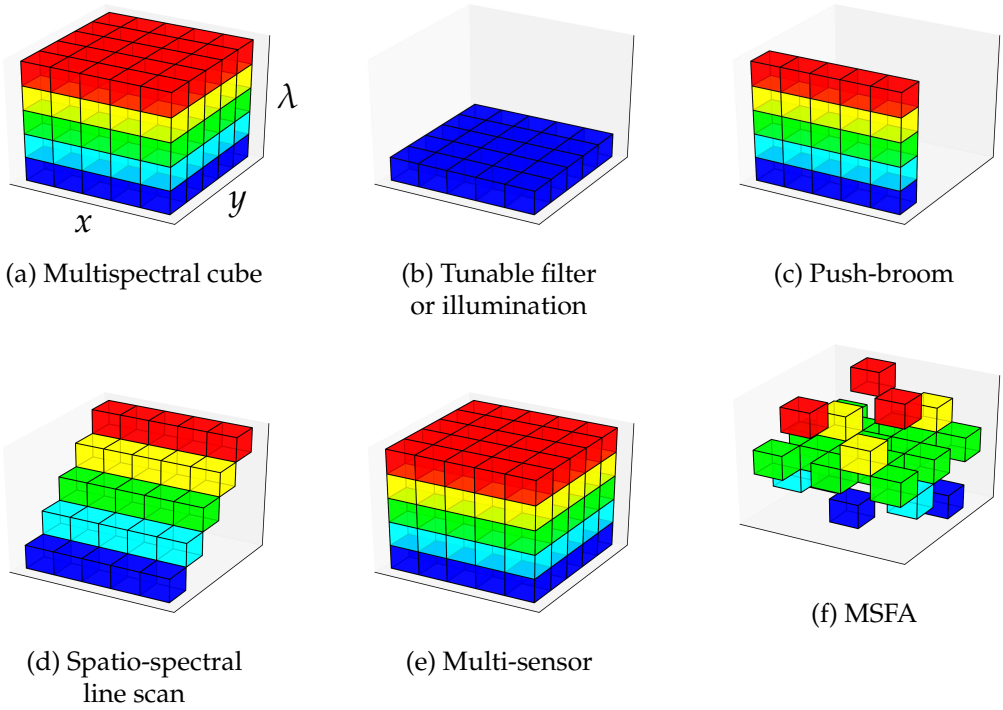
$$I_p^k = Q\left(\int_\Omega E(\lambda) \cdot R_p(\lambda) \cdot T^k(\lambda)\, d\lambda\right), \tag{1.1}$$

where Ω is the working spectral range. The term $E(\lambda)$ is the RSPD of the illumination which is assumed to homogeneously illuminate all surface elements of the scene. The surface element s observed by the pixel p reflects the illumination with the reflectance factor $R_p(\lambda)$ (supposed to be equal to $R_s(\lambda)$). The resulting radiance $E(\lambda) \cdot R_p(\lambda)$ is filtered according to the SSF $T^k(\lambda)$ of the band k centered at wavelength λ^k. The value I_p^k is finally given by the quantization of the received energy according to the function Q.

1.3.2 Multispectral image acquisition systems

A K channel multispectral image $\mathbf{I}$ can be seen as a cube with x and y spatial axes discretized as pixels and λ spectral axis discretized as central wavelengths of the spectral bands (see Fig. 1.6a). In order to acquire such a cube of size $X \times Y$ pixels $\times$ K channels, two families of multispectral image acquisition devices can be distinguished. "Multishot" systems build the cube from multiple acquisitions, while "snapshot" systems build it from a single acquisition.

"Multishot" systems sample the cube according to spectral and/or spatial axes and require the scene to be static until the cube is fully acquired. The first emerging "multishot" technology scans a channel I^k at each time (see Fig. 1.6b), so that K acquisitions are required to provide the fully-defined multispectral image $\mathbf{I} = \{I^k\}_{k=1}^K$. According to the image formation model, such spectral scanning can be achieved by using either a specific SSF or a narrow-band illumination at each acquisition. For this purpose, the tunable filter-based technology captures one channel at a time by changing the optical filter in front of the camera mechanically (*e.g.*, filter wheel)



(a) Multispectral cube

(b) Tunable filter or illumination

(c) Push-broom

(d) Spatio-spectral line scan

(e) Multi-sensor

(f) MSFA

FIGURE 1.6: "Multishot" (b to d) and "snapshot" (e and f) multispectral acquisition technologies used to acquire the multispectral cube (a).

or electronically (*e.g.*, liquid crystal tunable filters or acousto-optical tunable filters) [26]. In order to prevent painting deterioration due to illumination radiation, tunable illumination is commonly used in cultural heritage [60]. LED-based systems successively illuminate the scene with K narrow-band LED illuminations in order to provide the K channels [89].

Another approach is the spatial sampling of one row of X pixels for all channels at a time, so that Y acquisitions are required to build the entire cube. For this purpose, the push-broom line scan technology (see Fig. 1.6c) requires a focusing mirror lens that makes the camera acquire only a narrow strip of the scene. This strip is projected onto a collimating mirror and then passed through a transmission grating that separates the incoming radiance into different wavelengths, each being focused onto a row of the detector array. In this way, all surface elements in a strip of the scene are spectrally sampled simultaneously and a spatial scanning is required to provide the fully-defined cube, either by moving the scene (using a translation stage) or the camera in the direction orthogonal to the acquired strip [30].

Spatio-spectral line scan devices consist of a camera placed behind a slit spectroscope and a dispersive element [17]. At each acquisition, such technology provides a 2-D representation of the scene of size $X \cdot K$ in which the $k-$th row provides the values of channel I^k (see Fig. 1.6d). The fully-defined multispectral image is acquired by moving the scene, the camera, or the slit alone in the direction orthogonal to the acquired rows.

"Snapshot" multispectral systems provide the whole multispectral cube after a

single-shot acquisition. As opposed to "multishot" technologies, such acquisition systems provide the cube in real time. They are therefore able to capture a moving scene, and suited to video-based applications. Multi-sensor snapshot devices straight acquire a fully-defined multispectral cube (see Fig. 1.6e) thanks to dichroic beam splitters that selectively redirect the incoming radiance by wavelength onto the K sensors [67]. Such devices are expensive and cumbersome, as well as they are often only sensitive to a limited number of bands. Several single-sensor snapshot technologies have therefore been developed [32]. Among them, we focus on those using a multispectral filter array (MSFA) laid over the sensor [52]. The MSFA spatio-spectrally samples the incoming radiance according to the sensor element's location, like the widely-used Bayer color filter array (CFA) in color imaging. Each element of the MSFA is sensitive to a specific narrow spectral band, so that each pixel of the acquired raw image is characterized by one single band according to the MSFA pattern (see Fig. 1.6f). Finally, the fully-defined multispectral image is estimated by a demosaicing process. Further details are given about snapshot MSFA technology and demosaicing in Chapter 2.

1.3.3 Databases of acquired radiance

Table 1.1 provides an overview of the existing multispectral radiance image databases acquired using the multispectral acquisition systems described in Section 1.3.2. These databases represent outdoor urban or rural landscapes and indoor objects or faces. Early databases have been acquired using tunable filters and contain multispectral images composed of about 30 channels. These channels are associated with narrow bands, most often in the Vis domain, and with a spectral resolution of 10 nm [7, 9, 12, 23, 83, 110]. Later, the push-broom or spatio-spectral line scan technologies have made possible to acquire radiance databases with an increasing number of channels (up to 519 channels with a spectral resolution of 1.25 nm [5]) in the VisNIR domain [5, 19, 84, 92][4]. Recent advances in MSFA-based snapshot technology has led to the acquisition of several databases [53, 81, 122] that are used, for instance, to assess the performances of demosaicing [53, 81] or of classification [122]. Their raw images are acquired in the Vis or the VisNIR domain and include a number of channels that ranges from 5 to 25.

1.4 Databases of estimated reflectance

A major goal of multispectral imaging is to estimate the reflectance of surface elements in order to describe the spectral properties of materials. In this section we first explain how reflectance is estimated from radiance images and we provide an overview of existing reflectance databases in Section 1.4.1. Then, we present our

[4] Singapore database [84, 92] is spectrally down-sampled to the Vis domain, but the fully-defined VisNIR images can be provided.

Database	Year	Kind of images	Technology	Images	Size $X \times Y \times K$	Spectral domain
Bristol [9]	1994	Outdoor landscapes	Tunable filters	29	$256 \times 256 \times 31$	Vis
Brainard [7]	1998	Indoor objects	Tunable filters	9	$2000 \times 2000 \times 31$	Vis
Harvard [12]	2011	Indoor and outdoor scenes	Tunable filters	50	$1392 \times 1040 \times 31$	Vis
Scien [110]	2012	Indoor objects, faces, outdoor landscapes	Tunable filters, tunable illuminations	106	Various, *e.g.*, $998 \times 693 \times 148$	VisNIR
Singapore [84, 92]	2014	Indoor and outdoor scenes	Push-broom	64	$1312 \times 1924 \times 31$	Vis
Minho [83]	2015	Outdoor landscapes	Tunable filters	30	$1344 \times 1024 \times 33$	Vis
Time-Lapse [23]	2015	Outdoor landscapes	Tunable filters	33	$1344 \times 1024 \times 33$	Vis
UGR [19]	2015	Outdoor landscapes	Spatio-spectral line scan	14	$1392 \times 1040 \times 61$	VisNIR
ICVL [5]	2016	Outdoor landscapes	Push-broom	201	$1392 \times 1300 \times 519$	VisNIR
TokyoTech 5 [81]	2017	Indoor objects	MSFA	7	$1920 \times 1080 \times 5$	Vis
Bourgogne [53]	2017	Indoor objects	MSFA	18	$319 \times 255 \times 8$	VisNIR
Koblenz 16 [122]	2017	Terrains for Ground vehicle	MSFA	N.A	$2048 \times 1088 \times 16$	Vis
Koblenz 25 [122]	2017	Terrains for Ground vehicle	MSFA	N.A	$2048 \times 1088 \times 25$	VisNIR

TABLE 1.1: Overview of existing multispectral radiance image databases acquired using technologies described in Section 1.3.2. (N.A means unavailable information).

proposed database of estimated reflectances in Section 1.4.2, and we detail its acquisition scheme in Section 1.4.3.

1.4.1 Reflectance estimation and existing databases

At each pixel p of a radiance multispectral image, one generally estimates the reflectance $\mathbf{R}_p$ of the surface element s associated to p. According to our formation model that assumes homogeneous illumination of the scene and a maximal reflectance of 1 for a white diffuser, we estimate the reflectance $\mathbf{R}_p$ as the ratio between the acquired radiance $\mathbf{I}_p$ and the acquired radiance of a white diffuser $\mathbf{I}_w$ as:

$$\mathbf{R}_p = \frac{\mathbf{I}_p}{\mathbf{I}_w}, \tag{1.2}$$

where the pixel w observes a surface element of a white diffuser in the same illumination conditions than the surface element observed by p, so that $\mathbf{R}_p$ values range between 0.0 and 1.0. Note that $\mathbf{I}_p$ may here consider a dark current compensation to

reduce sensor noise that may cause wrong reflectance estimation. Note also that according to the Lambertian model, materials that exhibit high gloss in the considered acquisition conditions should be excluded since specular radiance would provide a reflectance greater than 1.
A reflectance image $\mathbf{R} = \{R^k\}_{k=1}^K$ can therefore be deduced from a multispectral radiance image when the latter contains at least one pixel w that observes a surface element of a white diffuser. Authors thus often include a white diffuser in each observed scene so that the reflectance can be estimated using pixel values associated with the diffuser. These scenes can therefore be acquired either indoor in controlled illumination conditions (*e.g.*, [38, 124]) or outdoor (*e.g.*, [22, 82]).

Some radiance databases are provided with the acquired surface element of a white diffuser in order to perform the reflectance estimation [9, 84, 92]. Otherwise, the databases presented in Table 1.2 directly provide the reflectance images estimated by their authors. Among them, two databases consist of urban and rural outdoor scenes [22, 82], and four databases show indoor scenes acquired in controlled conditions [38, 55, 81, 124]. Some databases are composed of specific categories of close-range objects like ancient manuscripts [22], paintings [125], textiles [76], or wood pieces [37]. Our proposed database, named HyTexiLa (Hyperspectral Texture images acquired in Laboratory) [46], is explained in details in the next section.

Database	Year	Kind of images	Technology	Images	Size $X \times Y \times K$	Spectral domain
Minho 02 [82]	2002	Outdoor scenes	Tunable filters	8	$1024 \times 1024 \times 31$	Vis
Minho 06 [22]	2004	Outdoor scenes	Tunable filters	8	$1344 \times 1024 \times 33$	Vis
East Anglia [38]	2004	Indoor objects	Multi-sensor	22	N.A $\times$ N.A $\times$ 31	Vis
CAVE [124]	2008	Indoor objects	Tunable filters	32	$512 \times 512 \times 31$	Vis
Ancient manuscripts [22]	2012	Printed documents	Tunable filters	3	$1344 \times 1024 \times 33$	Vis
Nordic sawn timbers [37]	2014	Timbers	Push-broom	107	$320 \times 800 \times 1200$	VisNIR
SIDQ [55]	2015	Indoor objects	Push-broom	9	$500 \times 500 \times 160$	VisNIR
TokyoTech [81]	2017	Indoor objects	Tunable filters	35	$500 \times 500 \times 31$	Vis
Paintings [125]	2017	Paintings	Tunable filters	5	$2560 \times 2048 \times 23$	VisNIR
SpecTex [76]	2017	Textiles	Spatio-spectral line scan	60	$640 \times 640 \times 39$	Vis
HyTexiLa [46]	2018	Textured materials	Push-broom	112	$1024 \times 1024 \times 186$	VisNIR

TABLE 1.2: Overview of existing estimated reflectance image databases acquired using technologies described in Section 1.3.2. (N.A means unavailable information).

1.4.2 Our proposed database: HyTexiLa

In order to perform texture analysis, only three close-range multispectral image databases of textured materials are available, namely SpecTex [76], Nordic sawn timbers [37] and SIDQ [55] (see Table 1.2). However, these databases present some limitations. SpecTex exhibits major spatial deformations due to the acquisition process and is available only in the visible domain, thus disregarding the NIR information that is relevant for the classification of materials [111] or textile fibers [28]. The Nordic sawn timber database is composed of only nine classes, in which each image is affected by strong stripe noise. Such noise is due to temporal variations of the sensor properties during the acquisition and causes visually perceptible stripes along spatial scanning [96]. Lastly, SIDQ contains only 9 different textures, whose images are severely blurred.

Thus we create our own database of texture images (HyTexiLa) in collaboration with the Norwegian Colour and Visual Computing laboratory (Gjøvik, Norway). HyTexiLa consists of 112 images acquired in the VisNIR domain. These images represent close-range fairly flat textile, wood, vegetation, food, and stone samples. Among them, 65 textile samples of various types and colors were provided by the humanitarian association *Don De Soie* located in Templeuve, France. Moreover 4 stone samples and 18 wood samples were found in the area around Mjøsa lake, Norway. Vegetation and food images respectively consist of 15 leaf or flower samples, and of 10 tea, coffee, or spice samples. Each image of the database is provided with a size of 1024×1024 pixels $\times$ 186 channels and represents a unique texture. Each channel is associated to a narrow spectral band centered at wavelength λ^k, $k \in \{1, \cdots, 186\}$, that ranges from 405.379 nm to 995.839 nm with a step of 3.190 nm. Our database is available for public use in the form of reflectance data [46]. Fig. 1.7 shows standard red green blue (sRGB) versions of four images in each category for illustration purpose (see Appendix A for the conversion from multispectral to sRGB), and the next section details the acquisition process.

1.4.3 Database acquisition and reflectance estimation

Our database is acquired by push-broom line scans using the HySpex VNIR-1800 multispectral camera manufactured by Norsk Elektro Optikk AS[5]. This camera is coupled with a close-up lens with a working distance of 30 cm. The scene is illuminated using the 3900e DC Regulated ER Lightsource manufactured by Illumination Technologies, Inc.[6]. The camera coupled with a translation stage requires about 15 minutes to acquire a fully-defined multispectral image. Together with each textured sample, the camera observes the SG-3051 SphereOptics Diffuse Reflectance Tile[7] that has a reflectance of 99% in the VisNIR domain.

From each acquired radiance image $\mathbf{I}$ of size 1800×5000 pixels $\times$ 186 channels, we

[5] https://www.hyspex.no/products/vnir_1800.php
[6] http://bit.ly/IT3900e
[7] http://sphereoptics.de/wp-content/uploads/2014/03/Zenith_Product-Brochure.pdf

FIGURE 1.7: Color version of four samples (rows) in each of the five categories (columns) that compose the HyTexiLa database, from left to right: food, stone, textile, vegetation, wood.

crop an area of 1024×1024 pixels that contains the texture sample. In order to perform reflectance estimation, we also consider an area of 550×550 pixels that contains the white diffuser in each image (see Fig. 1.8 for an example).

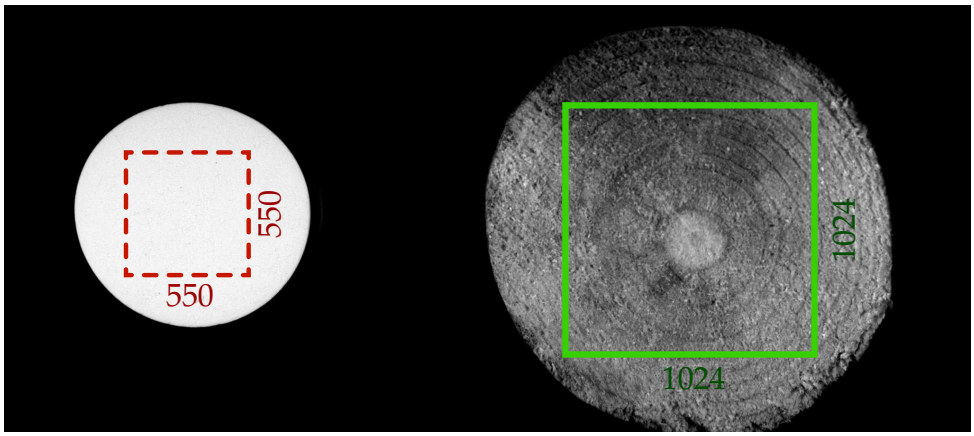


FIGURE 1.8: Acquired channel I^{93} (associated to the band centered at $\lambda^{93} = 699\,nm$) of a wood sample (right) together with the white diffuser (left). The retained texture area for all channels is displayed as a green solid square and the retained white diffuser area is displayed as a red dashed square.

The surface of the white diffuser is not perfectly flat and produces shaded areas in the acquired close-range images. To robustly estimate $\mathbf{I}_w$, we therefore consider the 5% of pixels with the greatest average values over all channels in the retained

white diffuser area. Then, for each band k, I_w^k is estimated as the median value of I^k at these pixels. Finally, we compute the reflectance image $\mathbf{R}$ according to Eq. (1.2). Note that pixels that correspond to specular reflection of illumination in the radiance image have higher values than those of the white diffuser and then than 1 in the reflectance image. We decide to keep them unchanged in the final database so that the original output radiance image can be retrieved by a multiplication with the white diffuser values.

1.5 Multispectral image simulation

By associating the values of each channel to the central wavelength of its associated band, the radiance image $\mathbf{I}$ (or the reflectance image $\mathbf{R}$) can be used to characterize the radiance (or the reflectance) of a scene at these wavelengths. Such information allows us to simulate the acquisition of any scene of the databases proposed in Tables 1.1 and 1.2 using the characteristics of a known camera. We present our proposed multispectral image simulation model based on the multispectral image formation model in Section 1.5.1. Such model is useful for the comparison of cameras properties or to simulate the fully-defined images that would be acquired using a single-sensor MSFA-based camera. Indeed, the only multispectral camera available to us is a 16-channel MSFA-based multispectral camera that is presented in Section 1.5.2. Finally we assess our simulation model with this camera in Section 1.5.3.

1.5.1 Image simulation model

We simulate the image acquisition process by discretely summing the simple multispectral image formation model described in Eq. (1.1) with $d\lambda = 1$:

$$I_p^k = Q\left(\sum_{\lambda\in\Omega} E(\lambda)\cdot R_p(\lambda)\cdot T^k(\lambda)\right), \tag{1.3}$$

where Ω denotes the minimal available common range among those of $E(\lambda)$, $R_p(\lambda)$ and $T^k(\lambda)$.

The radiance $E(\lambda)\cdot R_p(\lambda)$ of the surface element associated to a pixel p is available in one of the public radiance image databases of Table 1.1. Alternatively, radiance can be computed from estimated reflectance databases described in Table 1.2, coupled with any illumination described in Section 1.2 in either the Vis (Fig. 1.2) or the VisNIR (Fig. 1.3) domain. In both cases, the radiance can be computed for all integer $\lambda \in \Omega$ using linear interpolation of radiance or reflectance data available in the image channels associated to the band central wavelengths λ^k, $k \in \{1,\dots,K\}$. The resulting radiance is then projected onto K sensors, each one being associated with the SSF of one of the bands sampled by any considered camera. Note that $E(\lambda)$ and $R(\lambda)$ values range between 0 and 1. SSFs are normalized as $\max_k \sum_{\lambda\in\Omega} T^k(\lambda) = 1$, so that the product with the radiance provides a float value between 0 and 1. The

function Q quantifies this value on O bits as $Q(i) = \lfloor (2^O - 1) \cdot i \rceil$, where $\lfloor \cdot \rceil$ denotes the nearest integer function, so that $0 \leq I_p^k \leq 2^O - 1$. By applying such quantization, the maximal value (255 if $O = 8$) is only associated to a pixel that observes a white diffuser through a filter whose SSF area is 1. This normalization practically corresponds to setting the integration time of the camera as the limit before saturation when a white patch is observed.

1.5.2 IMEC16 multispectral filter array (MSFA) camera

The "snapshot" camera shown in Fig. 1.9 is available at the IrDIVE platform and we refer to it as IMEC16 for short in the following. It embeds a single sensor, covered by a 16-band MSFA that samples the Vis spectrum. This MSFA is manufactured by IMEC [27] and embedded in the sole off-the-shelf MSFA-based systems available on the market today, namely XIMEA's xiSpec[8] and IMEC's "snapshot mosaic"[9] multispectral cameras, with applications in medical imaging [90] or terrain classification [121].

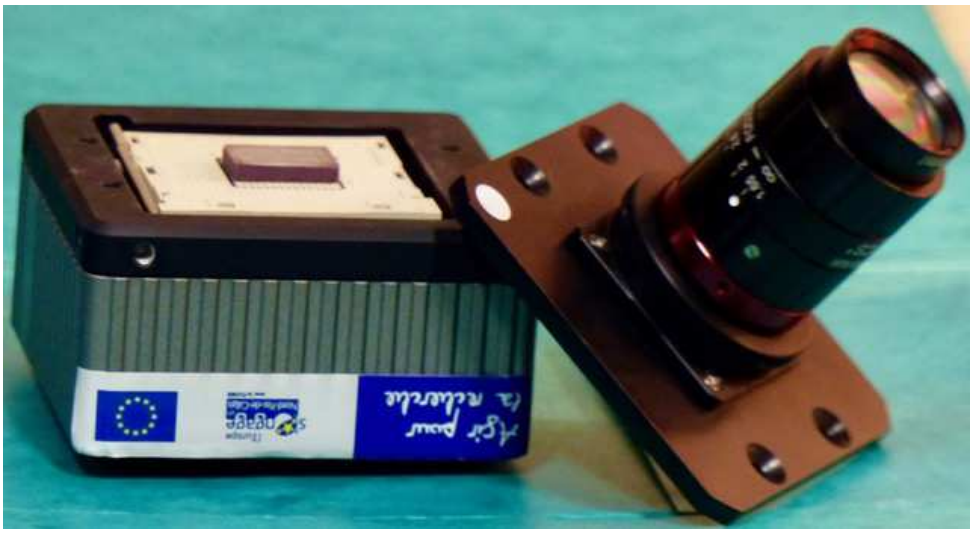


FIGURE 1.9: IMEC16 "snapshot mosaic" camera

IMEC16 camera samples 16 bands with known SSFs centered at wavelengths $\lambda^k \in \mathcal{B}(\text{IMEC16}) = \{469, 480, 489, 499, 513, 524, 537, 551, 552, 566, 580, 590, 602, 613, 621, 633\}$ (in nm), so that $\lambda^1 = 469\,nm$, $\ldots$, $\lambda^{16} = 633\,nm$. The SSFs $\{T^k(\lambda)\}_{k=1}^{16}$ (see Fig. 1.10) are provided by IMEC with 1 nm-bandwidths and normalized so that $\max_k \sum_{\lambda \in [450,650]} T^k(\lambda) = 1$. Note that in order to avoid second-order spectral artifacts, the optical device of this camera is equipped with a band-pass filter (at 450–650 nm).

[8] https://www.ximea.com/en/products/xilab-application-specific-oem-custom/hyperspectral-cameras-based-on-usb3-xispec

[9] https://www.imec-int.com/drupal/sites/default/files/inline-files/SNm4x4%20snapshot%20hyperspectral%20imaging%20sensor_0.pdf

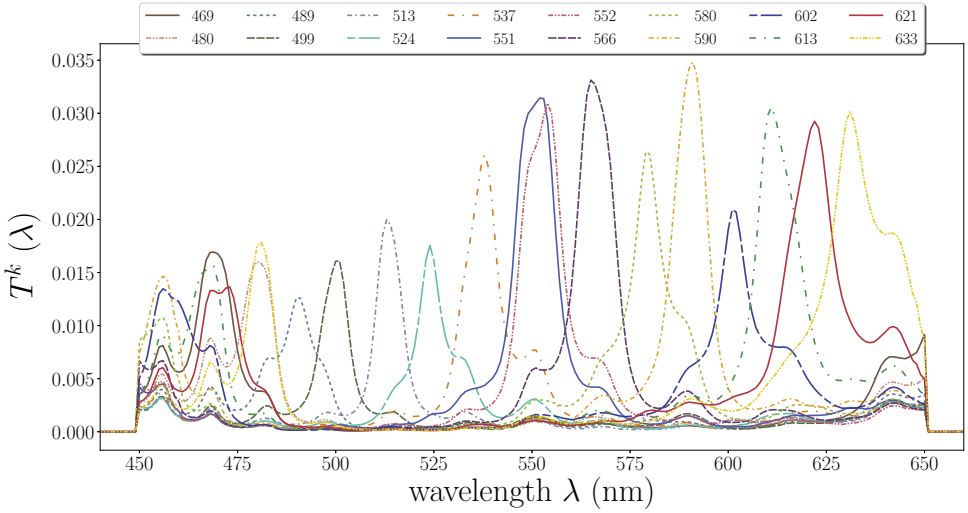


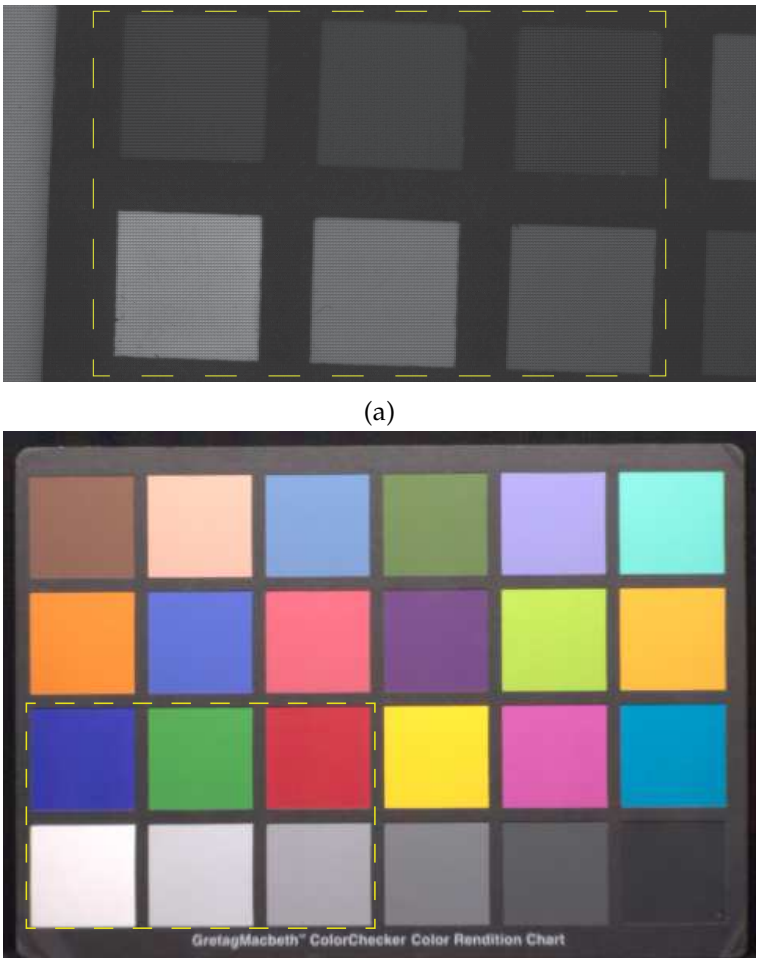
FIGURE 1.10: Normalized SSFs of IMEC16 camera. Captions: band center wavelengths $\mathcal{B}$(IMEC16) in ascending order.

1.5.3 Simulation validation with IMEC16 camera

In order to validate our simulation model, we use a Macbeth ColorChecker® Color Rendition Chart [68] that is both acquired and simulated in similar conditions. The color checker is acquired at IrDIVE platform using IMEC16 camera under HA or LD illumination (see Fig. 1.11a). Similarly, the simulation is performed using a reflectance image of the same color checker from East Anglia database [38] (see Fig. 1.11b), HA or LD RSPD (see Fig. 1.2), and SSFs of IMEC16 filters (see Fig. 1.10). The small LED dome that produces LD illumination forces us to bring the camera close to the scene, which restricts the acquired area to only six patches (see Fig. 1.11a). Assuming that this is enough to validate our simulation model, we select the red, green, and blue patches, and three gray ones (see dashed rectangle in Fig. 1.11). Finally, the pixel values of the six acquired color checker patches are compared with those of the simulated ones. Note that the normalization conditions of Section 1.5.1 require to configure the camera for each illuminant so that the acquired white patch reaches the maximum value.

Assuming that all surface elements of a patch have the same spectral response, we represent a patch as a 16-dimensional vector whose values are obtained as the averages over the available pixel values of this patch in each channel. Thus, each element of the resulting 16-dimensional vector carries the spectral response of the patch in one of the 16 bands. Then, we apply the least squares method in order to mitigate errors due to the camera optics. Specifically, we compute the vectors $\mathbf{a}$ and $\mathbf{b}$ that minimize the squared residual sum of acquired values $\mathbf{acq}_i = \{acq_i^k\}_{k=1}^{16}$ with respect to a linear function of simulated values $\mathbf{sim}_i = \{sim_i^k\}_{k=1}^{16}$ at each patch i among the 6 ones for a given illumination:

$$(\mathbf{a}, \mathbf{b}) = \underset{(\boldsymbol{\alpha}\in\mathbb{R}^{16}, \boldsymbol{\beta}\in\mathbb{R}^{16})}{\arg\min} \sum_{i=1}^{6} ||\mathbf{acq}_i - (\boldsymbol{\alpha} \cdot \mathbf{sim}_i + \boldsymbol{\beta})||^2 , \tag{1.4}$$



(a)

(b)

FIGURE 1.11: Acquired raw image of the six patches using IMEC16 camera under LD illumination (a), and sRGB representation of Macbeth Chart image from East Anglia database (b). Yellow dashed rectangles represent the area that contains the six selected patches

where $||\cdot||$ denotes the Euclidean norm. Vectors **a** and **b** are estimated using the simple linear regression proposed in [41]. The fidelity of our simulation model is measured according to the average peak signal-to-noise ratio (PSNR) between the acquired and simulated patches:

$$PSNR(\mathbf{acq}, \mathbf{sim}) = \frac{1}{16}\sum_{k=1}^{16} 10 \cdot log_{10}\left(\frac{2^O - 1}{\frac{1}{6}\sum_{i=1}^{6}\left(acq_i^k - (a^k \cdot sim_i^k + b^k)\right)^2}\right). \quad (1.5)$$

Table 1.3 shows the results according to the least squares method whose parameters are estimated under each illumination. When no least square regression is used, acquired and simulated patches have a nice PSNR. For a fixed given illumination, least square regression can be used to improve the fidelity of our simulation. For instance, the use of **a** and **b** computed from HA samples improves the PSNR by about $30\,dB$

on HA samples. However, it reduces the PSNR by about $3\,dB$ on LD samples. Therefore, when illumination changes, computing $\mathbf{a}$ and $\mathbf{b}$ using patches acquired under both illuminations represents a good compromise since it significantly improves the PSNR between all acquired and simulated samples.

$(\mathbf{a}, \mathbf{b})$	LD	HA
not used ($\forall k$, $a^k = 1$ and $b^k = 0$)	51.40	48.94
computed using LD samples	58.83	52.24
computed using HA samples	48.24	78.88
computed using HA and LD samples	54.10	66.66

TABLE 1.3: PSNR (dB) between acquired and simulated patches, under LD or HA illumination, without least square regression, or by computing $(\mathbf{a}, \mathbf{b})$ from Eq. (1.4) using patches acquired and simulated under LD, or HA, or both illuminations.

In a preliminary work on HyTexiLa database, we measure the noise power by analyzing its standard deviation on gray patches from a Macbeth ColorChecker reflectance image [46]. Results show that channels whose spectral bands are centered around $400\,nm$ are likely to be severely corrupted by noise. This can be due to the weak illumination and/or to optics and low sensor sensitivity in these spectral bands, where we are at the limit of the optical model that is being used [16]. Future works will focus on the improvement of our image formation model to take into account the noise with respect to both the SSFs and illumination in "multishot" and "snapshot" acquisition systems.

1.6 Properties of multispectral images

Multispectral raw images acquired from a "snapshot" camera must be demosaiced to provide fully-defined multispectral images. As will be detailed in Chapter 2, demosaicing generally takes advantage of spatial or/and spectral reflectance properties. We therefore study the properties of multispectral images simulated from reflectance data. We first describe the two considered multispectral image sets in Section 1.6.1. The spatial properties of these image sets are then assessed in Section 1.6.2 and their spectral properties are presented in Section 1.6.3.

1.6.1 Two simulated radiance image sets

In order to study multispectral images properties, we consider (i) CAVE scenes [124] of various objects with sharp transitions and (ii) HyTexiLa scenes [46] of smooth close-up textures (see Table 1.2). Considering these two databases allows us to highlight the influence of edge sharpness on spatial correlation.
(i) The 32 multispectral CAVE images are defined on 31 bands of width $10\,nm$ and centered at $\{400\,nm, 410\,nm, \ldots, 700\,nm\}$. By associating each surface element with

a pixel p and assuming linear continuity of reflectance, we get $R_p(\lambda)$ for all integer $\lambda \in \Omega = [400\,nm, 700\,nm]$ using linear interpolation of CAVE data. For each $\lambda \in \Omega$, the radiance is defined at each pixel p by the product between $R_p(\lambda)$ and the RSPD $E(\lambda)$ of D65 illuminant (see Section 1.2.1). Finally, we consider the SSFs of IMEC16 camera (see Fig. 1.10) in order to estimate the associated 16 channels according to Eq. (1.3). Indeed IMEC16 is the only MSFA-based camera available to us and it embeds no demosaicing method. It is therefore interesting to study the properties associated to IMEC16 SSFs for demosaicing.
(ii) We simulate the radiance of HyTexiLa scenes at Hyspex VNIR-1800 central wavelengths (see Section 1.4.3) under extended D65 illuminant (see Section 1.2.1) as:

$$I_p^k = Q\left(E(\lambda^k) \cdot R_p(\lambda^k)\right), k \in \{1, \ldots, 186\}, \tag{1.6}$$

so that no linear interpolation of the reflectance is required in that case. In order to reduce the spectral dimension of resulting images, we uniformly select 16 among the 186 channels such that their band centers range from 437 nm to 964 nm with a step of 35.07 nm.
Note that the CAVE set considers the Vis domain while HyTexiLa set considers the VisNIR domain. Considering these two sets highlights the influence of NIR information on spectral correlation.

1.6.2 Spatial properties

Most CFA demosaicing schemes assume that reflectance does not change locally across neighboring surface elements, hence that values of a color component are correlated among neighboring pixels in homogeneous areas. The sparse spatial subsampling of each channel by the MSFA may affect this spatial correlation assumption. To assess it, we use the Pearson correlation coefficient between the value I_p^k of each pixel $p(x, y)$ and that of its right neighbor $I_{p+(\delta_x,0)}^k$ at spatial distance δ_x according to the x-axis in a given channel I^k. This coefficient is defined as [29]:

$$C[I^k](\delta x) = \frac{\sum_p \left((I_p^k - \mu^k)(I_{p+(\delta_x,0)}^k - \mu^k)\right)}{\sqrt{\sum_p (I_p^k - \mu^k)^2}\sqrt{\sum_p (I_{p+(\delta_x,0)}^k - \mu^k)^2}}, \tag{1.7}$$

where μ^k is the mean value of channel I^k. For a given δ_x, we compute the average correlation $\mu_C(\delta_x)$ on the 32 scenes from CAVE set, and on the 112 scenes from HyTexiLa set. Note that the illumination has no influence on spatial correlation since we assume that it homogeneously illuminates all surface elements. The results (see Table 1.4) show that for the CAVE set, the higher the spatial distance between two pixels, the lower the correlation between them. In particular, the spatial distance between two pixels with the same available channel is $\delta_x = 2$ in the Bayer CFA and $\delta_x = 4$ in IMEC16 MSFA, which makes the correlation decreases from 0.94 to 0.88.

Regarding the 112 textures of HyTexiLa set, some images of which are mostly composed of spatial low frequencies, the spatial distance between pixels has no significant influence on spatial correlation. Note that because of the presence of non-blurry details, CAVE is the most widely used database for multispectral demosaicing that becomes a challenge for the community as spatial sampling gets sparser.

δ_x (pixels)	0	1	2	3	4
CAVE	1.00	0.98	0.94	0.91	0.88
HyTexiLa	1.00	0.96	0.95	0.94	0.96

TABLE 1.4: Spatial correlation $\mu_C(\delta_x)$ between values of two neighboring pixels for different distances δ_x (average over 16 channels of 32 images from CAVE set or 112 images from HyTexiLa set).

1.6.3 Spectral properties

Gunturk et al. [29] also experimentally show that color components are strongly correlated in natural images, such that all three channels largely share the same texture and edge locations. This strong spectral correlation can be effectively used for CFA demosaicing because SSFs of single-sensor color cameras widely overlap. On the opposite, an MSFA usually finely samples the visible spectrum according to K separated bands. We can then expect that channels associated with nearby band centers are more correlated than channels associated with distant band centers [77]. To validate this assumption, we evaluate the correlations between all pairs of channels on all scenes from CAVE set. The Pearson correlation coefficient between any pair of channels I^k and I^l is computed as [29]:

$$C\left(I^k, I^l\right) = \frac{\sum_p \left((I_p^k - \mu^k)(I_p^l - \mu^l)\right)}{\sqrt{\sum_p (I_p^k - \mu^k)^2}\sqrt{\sum_p (I_p^l - \mu^l)^2}}. \tag{1.8}$$

The results (see Fig. 1.12) confirm that channels associated with spectrally close band centers ($\lambda^k \approx \lambda^l$) are more correlated than channels associated with distant band centers ($\lambda^k \gg \lambda^l$ or $\lambda^k \ll \lambda^l$).

Fig. 1.12 shows that IMEC16 SSFs provide images with pairwise correlated channels even when the associated band centers are distant, all correlation values being higher than 0.76. It is interesting to examine the behavior of this correlation in the VisNIR domain by considering the HyTexiLa set. Fig. 1.13 shows the spectral correlation between channels on average over the 112 images. The correlation is high within each of the Vis and NIR domains: it ranges from 0.55 to 1.00 inside the Vis domain (top left), and from 0.76 to 1.00 inside the NIR domain (bottom right). But channels associated with two bands in different domains (top right and bottom left)

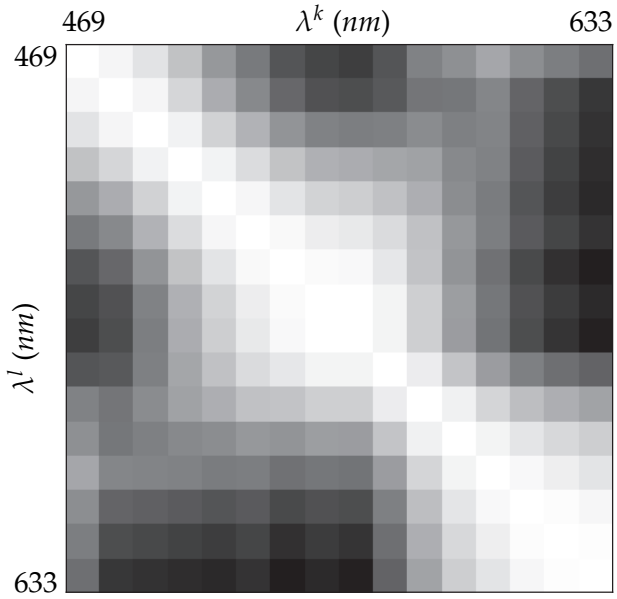


FIGURE 1.12: Correlation between channels I^k and I^l of images from the CAVE set. Values are averaged over the 32 images and range between 0.76 (black) and 1.0 (white).

are weakly correlated since values range from 0.29 to 0.64. Note that spectral correlation is higher in the NIR domain than in the Vis domain. Note also that channels from CAVE set are more correlated than channels from HyTexiLa set in the Vis domain since the 16 channels of CAVE set range from 469 nm to 633 nm while the 8 channels of HyTexiLa set in the Vis domain more widely range from 437 nm to 682 nm.

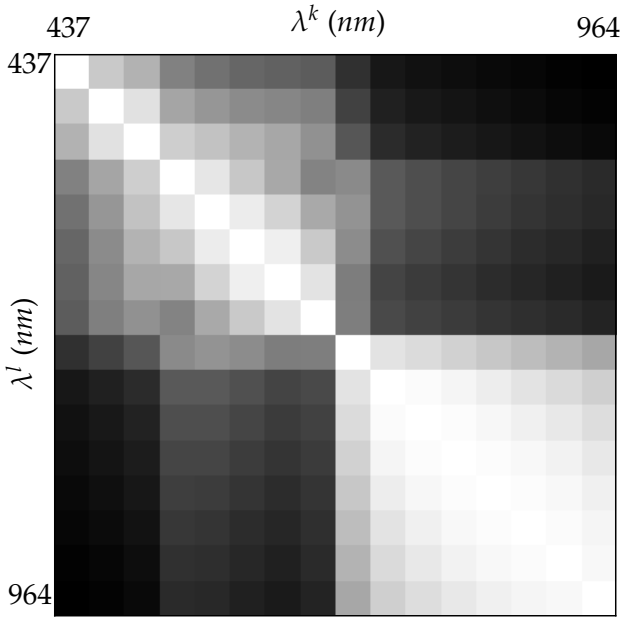


FIGURE 1.13: Correlation between channels I^k and I^l of images from the HyTexiLa set. Values are averaged over the 112 images and range between 0.29 (black) and 1.0 (white).

1.7 Conclusion

In this chapter, we have first provided an overview of the different illuminations that are used through the manuscript. Beams coming from illumination are modified according to reflectance properties of the object material and reach the sensor of the camera. To form the channels of a multispectral image, a multispectral camera samples the resulting radiance spectrum according to different spectral bands. A multispectral reflectance image can also be estimated by placing a white diffuser in a scene whose radiance is to be acquired. Thus, many reflectance image databases are proposed in the literature and are useful to characterize the reflectance of surface elements in different scenes.
Because no existing database is relevant for texture analysis, we have proposed our own database of estimated reflectances. This database is used especially to perform texture classification, as detailed in Chapter 4. The acquisition of an image of this database requires 15 minutes and is not appropriate for moving scenes.
In order to reduce the acquisition time, multispectral cameras based on MSFA technology can be used. IrDIVE platform provides us the IMEC16 MSFA-based camera that samples 16 bands in the Vis domain. However, such camera provides only raw images in which the value of a single channel is available at each pixel. Thus, we have proposed a model to simulate the fully-defined images that would be acquired using the SSFs of this camera. This model has been successfully assessed by comparing simulated and acquired images. However it can be criticized since it does not take into account the noise associated with camera optics or SSFs, that have an influence on multispectral images properties. A statistical study of the properties of IMEC16 multispectral images has yield to three main properties that could be exploited or at least should be kept in mind for MSFA demosaicing:

- Spatial correlation within each channel decreases as the spatial distance between pixels increases.
- Spectral correlation between channels decreases as the distance between centers of their associated bands increases.
- The correlation between NIR and Vis channels is low.

The next chapter focuses on multispectral demosaicing methods that are based on these properties.

Chapter 2

MSFA raw image demosaicing

Contents

2.1 Introduction

A multispectral filter array (MSFA) is defined by a basic repetitive pattern composed of filter elements, each of which is sensitive to a specific narrow spectral band. A camera fitted with such a device provides a raw image in which the value of a single channel is available at each pixel according to the MSFA pattern. The missing channel values are thereafter estimated by a demosaicing process that is similar in its principle to the estimation of missing values in Bayer color filter array (CFA) raw images. CFA demosaicing is a well-studied problem for more than forty years [57], while MSFA demosaicing is a recent subject with new issues. Indeed, the principles of spatial and spectral correlations, that exploit the properties of radiance in CFA demosaicing, should be reconsidered. First, more spectral bands imply a lower spatial sampling rate for each of them, which weakens the assumption of spatial correlation between the raw values that sample the same band. Second, since multispectral imaging uses narrow bands whose centers are distributed over the spectral domain, the correlation between channels associated with nearby band centers is stronger than between channels associated with distant ones. Third, the property of spectral correlation is weakened in the VisNIR domain since Vis and NIR channels are weakly correlated.

We present the "Snapshot" MSFA technology, and the specifications and issues associated to the different MSFAs of the literature in Section 2.2. In order to assess demosaicing methods, we focus on the IMEC16 MSFA. Indeed, such MSFA that privileges spectral resolution is incorporated in a camera available at the IrDIVE platform, and that embeds no demosaicing method. However, since only a few multispectral demosaicing methods exist, we first present the methods that are not dedicated to our considered MSFA. Then, Section 2.3 presents the four methods developed specifically to demosaic raw images that are acquired thanks to an MSFA that exhibits a dominant green band (like the Bayer CFA). This section also briefly presents data-dependent demosaicing methods, that are based on a learning database or a sparsity assumption of the raw images. Indeed, such methods often require fully-defined multispectral images that are not available in practice, which makes them unreliable for our considered MSFA. Section 2.4 further details the different state of the art methods that can be used to demosaic images acquired thanks to our considered IMEC16 MSFA raw images. To perform multispectral demosaicing despite the weak spatial correlation, we propose to use a spatially fully-defined channel that is estimated from the raw image, namely the pseudo-panchromatic image (PPI). Section 2.5 presents the relevance of PPI for demosaicing and its estimation from the raw image. The PPI is then used to improve two state of the art methods and in an original PPI difference demosaicing scheme in Section 2.6.

2.2 Multispectral filter array technology

This section focuses on MSFA technology whose acquisition pipeline is presented in Section 2.2.1. The MSFA design is described in Section 2.2.2 and the main MSFA patterns are detailed in Section 2.2.3.

2.2.1 MSFA-based acquisition pipeline

To acquire a color image in a single shot, the technology based on CFA is the most used in machine vision. Indeed, in addition to be cheap, such technology is light and robust enough to be embedded in every consumer electronics device. Similarly, cameras equipped with MSFAs are able to acquire images with more than three channels in a single shot. For this purpose, the single sensor of an MSFA-based camera captures the radiance spectrum through an MSFA. Each of the K spectral sensitivity functions (SSFs) of the different filters that compose it is sensitive to a specific narrow spectral band. Thus, at each pixel of the acquired raw image, only the value of the associated single channel is available according to the MSFA. The $K-1$ missing channel values at each pixel are thereafter estimated by a demosaicing process that estimates a fully-defined multispectral image. The IMEC16 MSFA acquisition pipeline is shown in Fig. 2.1.

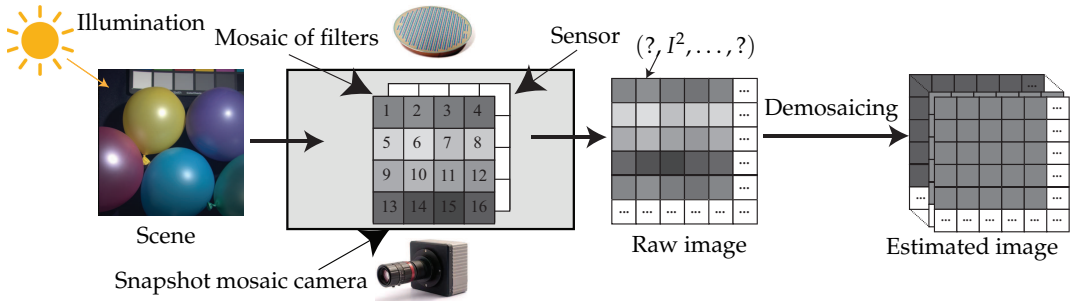


FIGURE 2.1: Acquisition pipeline in IMEC16 MSFA-based camera.

2.2.2 MSFA design

The demosaicing quality is directly related to filter array design. The Bayer CFA for instance samples the green band at half of the sites, which makes it a prominent candidate to begin the demosaicing process. Spectral correlation is then generally assumed in order to estimate red and blue channels using the well-estimated green channel. Unlike CFA design which mainly consists in Bayer CFA, the number of spectral bands in an MSFA and the shape of associated SSFs may vary with respect to the application [52].

Early MSFA-based devices aim to improve CFA-based ones. For instance Ohsawa et al. [85] combine two color cameras in order to provide a multispectral image with six channels. In order to extend CFA to the VisNIR domain, some cameras integrate both Vis and NIR photosensitive elements in a single filter array [35, 48]. Stating that

the panchromatic band (that is sensitive to the whole VisNIR domain) is less sensitive to noise than color channels, some so-called RGBW filter arrays are proposed to also sample a panchromatic band [94].
In order to improve demosaicing performances some authors design MSFAs that provide optimal demosaicing performances in term of PSNR on a given multispectral image set. For instance, Shinoda et al. [104] and Yanagi et al. [123] evaluate the filter arrangement of an MSFA by using a metric related to the PSNR between simulated and demosaiced images. By considering the VisNIR MSFA design as a spatial optimization problem, some authors propose an iterative procedure that leads to the co-design of an optimized MSFA and its demosaicing algorithm [65, 97]. Another approach favors a faithful reconstruction of the incoming radiance. For this purpose, Jia et al. [44] designs a "Fourier" MSFA that improves spectrum reconstruction using the Fourier transform spectroscopy.
When no fully-defined image set is available to assess demosaicing performances, some models provide an optimized MSFA without using training images. For this purpose, Shinoda et al. [107] measure the distances between sampling filters in a spatio–spectral domain, and assume that the demosaicing performances depend on the dispersion degree of the sampling points in this domain. Recently, Li et al. [58] present an optimization model that considers various errors associated with spectral reconstruction, namely, errors due to spectrum estimation, noise, and demosaicing. Regardless of the acquired scene, these errors only depend on tunable parameters, such as the SSFs, the MSFA pattern, the demosaicing algorithm, or the variance of the sensor noise.
To conclude, MSFA design deals with a trade-off between spatial and spectral resolutions which leads to some issues, like the lack of bands for spectral reconstruction or the performance of demosaicing.

2.2.3 MSFA basic patterns

To ensure manufacturing practicability and demosaicing feasibility, all MSFAs are defined by a basic repetitive pattern that respects a trade-off between spatial and spectral sub-samplings. The spatial arrangement of the filter elements in this basic pattern plays an important role in MSFA design. Indeed, Shrestha et al. [109] show that the influence of the pattern tends to be more prominent when the number of bands increases, *i.e.*, when the spatial distance between sites associated with the same band increases. Moreover, SSFs have to be carefully designed since they both affect the spectral reproduction ability and the spatial reconstruction quality [44].
Two important criteria must be considered in the MSFA basic pattern design [69]: *spectral consistency* and *spatial uniformity*. An MSFA is spectrally consistent if, in the neighborhood of all filters associated to any given band, the same bands are sampled the same number of times. Spatial uniformity requires that an MSFA spatially samples each band as evenly as possible. Both requirements are related to the demosaicing process that is applied to the raw image. Indeed, demosaicing independently

scans all the pixels associated to a given band and considers pixels in their neighborhoods. The neighborhood layout should then be same whatever the pixel considered in the raw image. We present here the main MSFAs that respect these criteria.
Brauers and Aach [8] propose a 6-band MSFA arranged in 3×2 basic pattern. Wang et al. [118] propose an RGBW MSFA where bands are arranged in diagonal stripes, and in which half of the sites sample the panchromatic channel. Aggarwal and Majumdar [2] also propose a 5-band "uniform" MSFA where bands are arranged in diagonal stripes.
Miao and Qi [69] propose an algorithm that generically builds MSFAs in which each band is characterized by its prior probability (PP). This algorithm associates each band to a leaf of a binary tree and defines its PP as the inverse of a power of two with the leaf depth as exponent. Fig. 2.2 shows the formation of an MSFA using such a binary tree. The resulting 4×4 basic pattern shown in Fig. 2.3a contains three dominant bands (R, G, and B) with a PP of $\frac{1}{4}$ and two under-represented bands (cyan (C) and magenta (M)) with a PP of $\frac{1}{8}$.

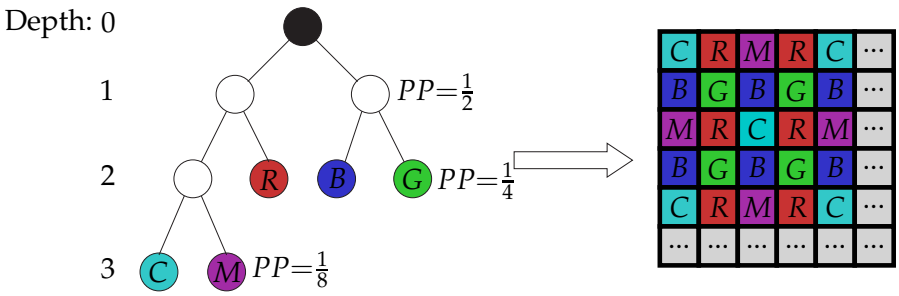
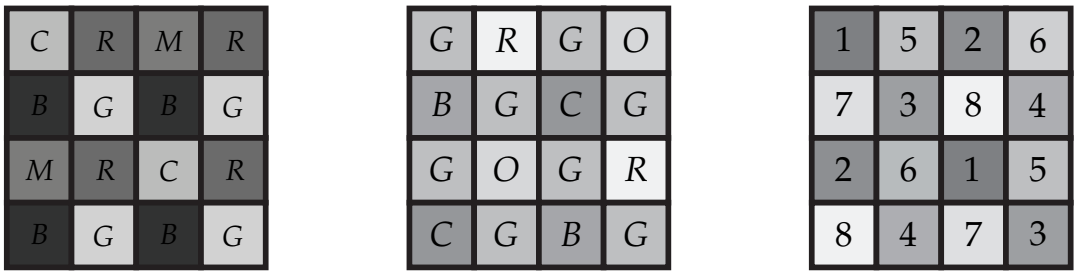


FIGURE 2.2: MSFA generation using a binary tree [69].

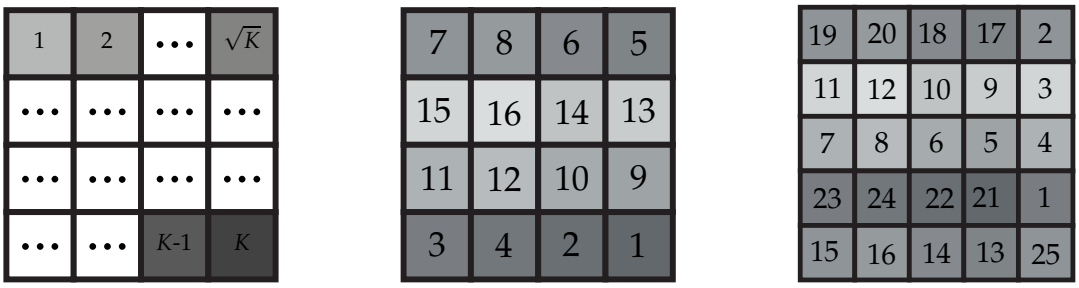
Monno et al. [78] propose the 4×4 basic pattern that is inspired by the Bayer CFA pattern. This pattern, called here VIS5, exhibits a PP of $\frac{1}{2}$ for G and of $\frac{1}{8}$ for the four other bands (R, B, C, and orange (O)) (see Fig. 2.3b). Thomas et al. [115] propose the VISNIR8 4×4 basic pattern shown in Fig. 2.3c that samples 7 bands in the Vis domain and 1 band in the NIR domain with equal PPs of $\frac{1}{8}$. The spectral sensitivity functions (SSFs) of VIS5 and VISNIR8 MSFAs can be found in the papers [81] and [115], respectively, and are represented in Appendix B.



(a) (b) (c)

FIGURE 2.3: Basic patterns of three MSFAs generated using a binary tree: That of Fig. 2.2 (a) [69], VIS5 (b) [78], and VISNIR8 (c) [115]. Band labels in (a), (b) are those of [69, 78] but could be replaced by indexes.

Increasing the number of bands to enhance spectral resolution is a goal of multispectral imaging. Some MSFAs are then defined by a basic pattern without any repeated band, although this conflicts with a dense spatial sampling. Such MSFAs have typically a square or rectangular basic pattern [115]. For instance, Fig. 2.4a shows a $\sqrt{K} \times \sqrt{K}$ square basic pattern composed of K non-redundant bands. The two MSFAs whose square basic patterns are shown in Figs. 2.4b and 2.4c are manufactured by IMEC [27]. The 4×4 basic pattern samples 16 bands in the Vis domain and the 5×5 one samples 25 bands in the NIR domain. Their band centers are not ascending in the classical pixel readout order, presumably due to manufacturing constraints. The MSFAs defined by these two patterns (or the corresponding cameras) are shortly called IMEC16 and IMEC25 in the following and their SSFs are available in Appendix B.



(a) (b) (c)

FIGURE 2.4: Square basic patterns of three MSFAs with no redundant band: $\sqrt{K} \times \sqrt{K}$ (a), IMEC16 (b), and IMEC25 (c) [27]. Numbers are band indexes.

2.3 MSFA demosaicing

In this section we first introduce a formulation of the MSFA demosaicing problem in Section 2.3.1. Then we present demosaicing methods that use the dominant band of VIS5 MSFA in Section 2.3.2 and data-dependent demosaicing methods in Section 2.3.3.

2.3.1 MSFA demosaicing problem

A single-sensor multispectral camera fitted with an MSFA provides a raw image I^{raw} of size $X \times Y$ pixels, in which a single band $k \in \{1, \ldots, K\}$ is associated with each pixel p according to the MSFA. Let S be the set of all pixels (whose cardinal is $|S| = X \times Y$) and S^k be the pixel subset where the MSFA samples the band k, such that $S = \bigcup_{k=1}^{K} S^k$. An MSFA can be defined as a function $MSFA : S \to \{1, \ldots, K\}$ that associates each pixel p with the index of its associated spectral band. Therefore the pixel subset where the MSFA samples the band k can be defined as $S^k = \{p \in S, MSFA(p) = k\}$. The raw image I^{raw} can then be seen as a spectrally-sampled

version of the reference fully-defined image $\mathbf{I} = \{I^k\}_{k=1}^K$ (that is unavailable in practice) according to the MSFA:

$$\forall p \in S,\ I_p^{raw} = I_p^{MSFA(p)}. \tag{2.1}$$

The raw image can also be seen as the direct sum of K sparse (raw valued) channel $\{\tilde{I}^k\}_{k=1}^K$, each of which contains the available values at pixels in S^k and zero elsewhere. This can be formulated as:

$$\tilde{I}^k = I^{raw} \odot m^k\ , \tag{2.2}$$

where $\odot$ denotes the element-wise product and m^k is a binary mask defined at each pixel p as:

$$m_p^k = \begin{cases} 1 & \text{if } MSFA(p) = k, \textit{i.e.}, p \in S^k, \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

Demosaicing is then performed on each sparse channel $\tilde{I}^k$ to obtain an estimated image $\hat{\mathbf{I}}$ with K fully-defined channels, among which $K-1$ are estimated at each pixel p: for all $p \in S^k$, $\hat{\mathbf{I}}_p = \left(\hat{I}_p^1, \ldots, \hat{I}_p^{k-1}, I_p^k, \hat{I}_p^{k+1}, \ldots, \hat{I}_p^K\right)$, where $\hat{I}_p^l$, $l \neq k$, is the estimated value of channel I^l at p. For illustration purpose, Fig. 2.5 shows the demosaicing problem formulation for VIS5 MSFA.

All demosaicing methods estimate missing values using spatial (i) and/or spectral (ii) correlations. (i) The spatial correlation assumes that if a pixel p and its neighborhood belong to the same homogeneous area, the value of p is strongly correlated to the values in its neighborhood. Thus, assuming that a channel is composed of homogeneous areas separated by edges, the value of a pixel can be estimated by using its neighbors within the same homogeneous area. Spatial "gradients" are often used as weights to determine whether two pixels belong to the same homogeneous area. Indeed, gradient-based methods consider the difference between values of two spatially close pixels of a subset S^k. We can therefore assume that these pixels belong to the same homogeneous area if the gradient is low, and that they belong to different homogeneous areas otherwise. (ii) Spectral correlation assumes that the areas with high frequencies (textures or edges) of the different channels are strongly correlated. If the MSFA contains a dominant band, demosaicing generally estimates the associated channel whose high frequencies can be faithfully reconstructed, then uses it as a guide to estimate other channels. Indeed, the faithfully reconstructed image can be used in order to guide the high-frequency contents estimation within the different channels [43].

2.3.2 VIS5 MSFA demosaicing

Several Bayer CFA demosaicing schemes exploit the green channel properties (either implicitly or explicitly as in [49]) for demosaicing because G is over-represented with

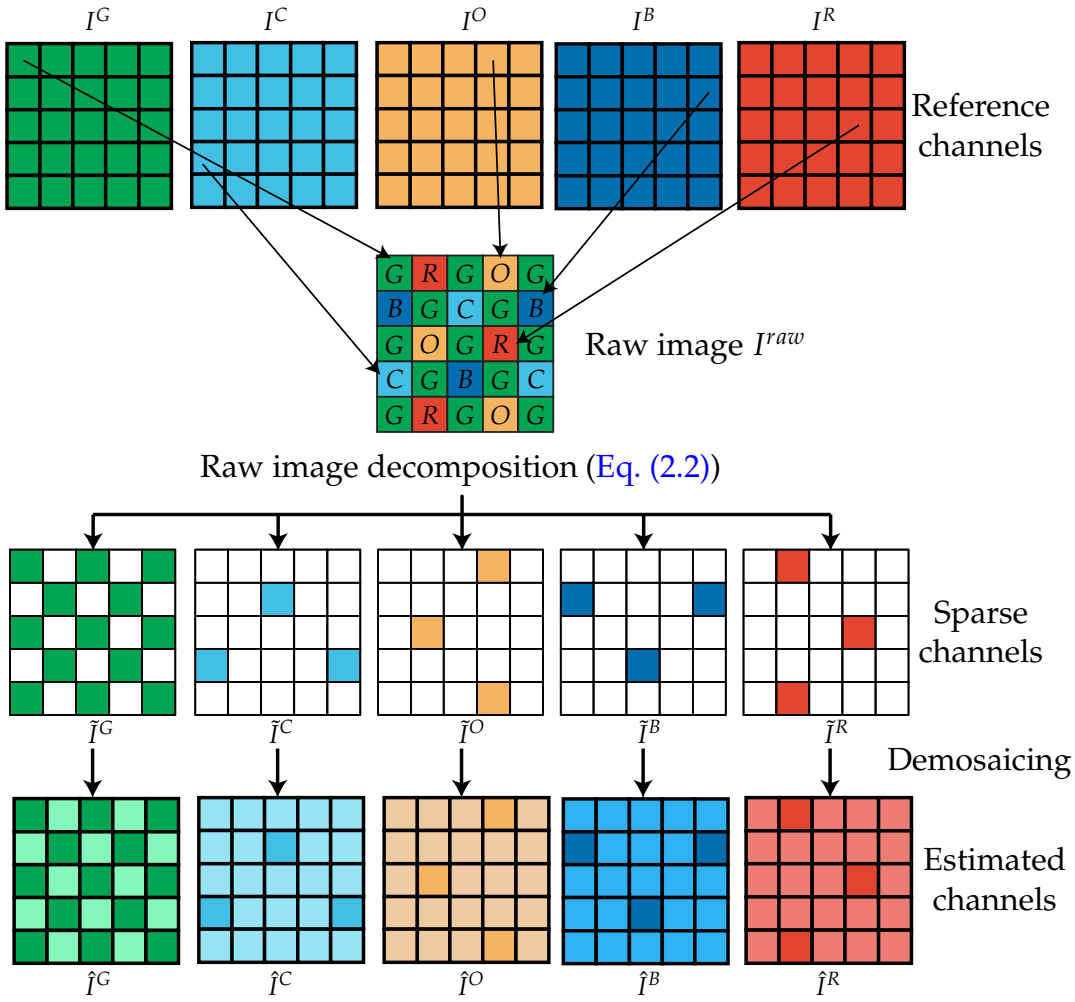


FIGURE 2.5: Demosaicing outline for VIS5 MSFA.

respect to R and B in a raw Bayer image ($|S^G| = 2|S^R| = 2|S^B|$). Similarly, multispectral demosaicing schemes applied to MSFAs with a dominant band first estimate the associated channel and use it to estimate other channels [43, 78–80]. Here we present three methods specially designed for the VIS5 MSFA that exhibits the dominant G band (see Fig. 2.3b).

Demosaicing using adaptive kernel up-sampling

Monno et al. [78] adapt Gaussian up-sampling (GU) and joint bilateral up-sampling (JBU) proposed by Kopf et al. [50] to VIS5 MSFA demosaicing. GU estimates a missing value of a sparse raw image by using a weighted value of spatially neighboring pixels, while JBU also considers the weights of a guide image. Both use a spatially-invariant Gaussian function for weight computation. Monno et al. [78] instead use an adaptive kernel for kernel regression as proposed in [114]. Such adaptive kernel considers a covariance matrix based on the diagonal gradients (computed among pixels that are associated to the same pixel subset) in a 3×3 window around the

pixel to be estimated. Adaptive GU and JBU are used in an algorithm that proceeds in 3 successive steps (see Fig. 2.6):

1. First, it estimates the adaptive kernels from the raw image.
2. Second, it generates the guide image $\hat{G}$ by applying the adaptive GU on sparse channel $\tilde{I}^G$.
3. Third, it applies the adaptive JBU using the guide image in order to estimate all channels (including the green channel).

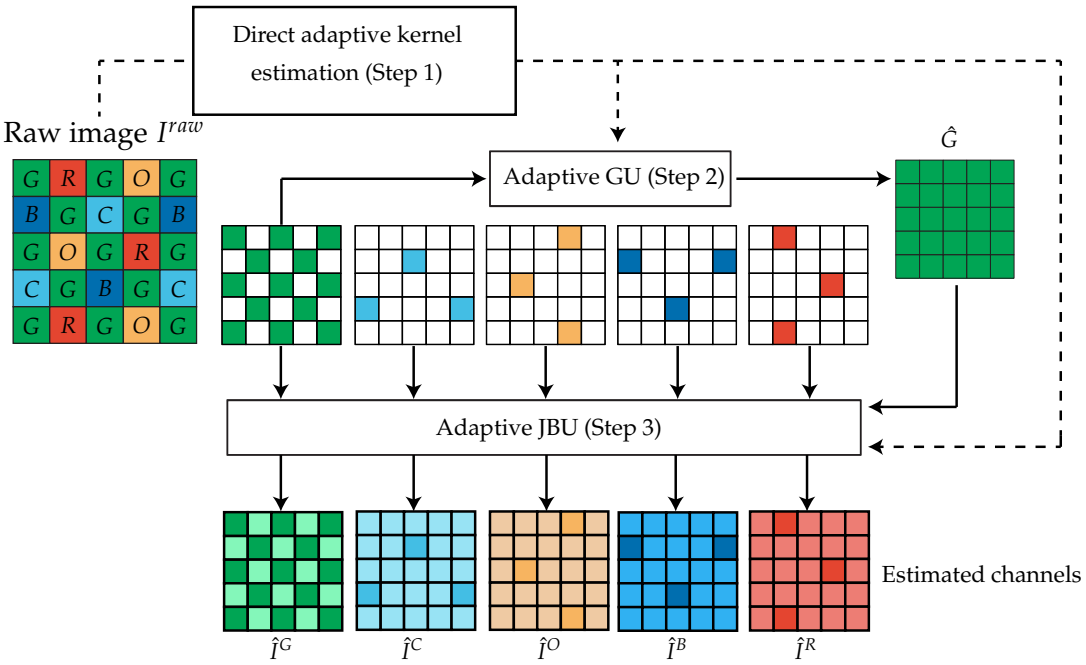


FIGURE 2.6: Demosaicing by adaptive kernel upsampling.

Note that $\tilde{I}^G$ is also interpolated by adaptive JBU so that the high-frequency properties of all spectral channels are consistent. This algorithm is further improved in [79] by considering a guided filter (GF) instead of the adaptive JBU to estimate each channel. Such GF performs a linear transform of the guide image in order to faithfully preserve its structure in the estimated image [34].

Demosaicing using residual interpolation

Monno et al. [80] adapt their CFA demosaicing method that uses residual interpolation [47] to VIS5 MSFA. Such method considers the residuals defined by a difference between an acquired and a tentatively estimated pixel value. Their algorithm first estimates the guide image $\hat{G}$ at each pixel subset S^k, $k \in \{G, R, B, C, O\}$. At pixels in S^G, $\hat{G}$ has raw image values, while at pixels in other subsets, for instance S^R, $\hat{G}$ is estimated as follows:

1. The green channel is linearly interpolated in the horizontal direction at rows that contain S^R (every two rows).

2. The red channel is estimated by residual interpolation in the same rows (see Fig. 2.7). For this purpose, the red channel is pre-estimated at these rows by using a GF with the estimated green channel as a guide. Note that such pre-estimation modifies the raw values of pixels in S^R. The red channel residuals at S^R positions are then computed by subtracting the pre-estimated red channel and $\tilde{I}^R$. Finally, the residuals are linearly interpolated in the horizontal direction, and added to the pre-estimated red channel in order to provide an horizontally estimated red channel at rows that contain S^R.

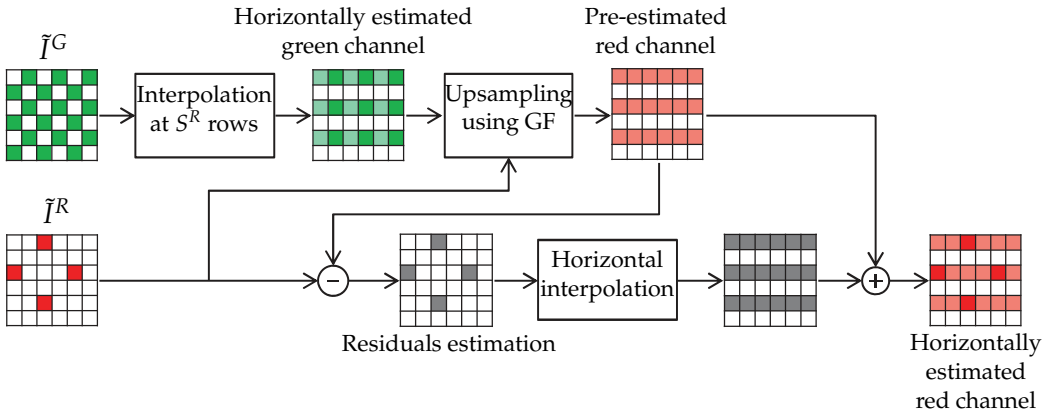


FIGURE 2.7: Horizontal residual interpolation of channel I^R.

3. A horizontal difference channel is computed by subtracting estimated green (by linear interpolation) and red (by residual interpolation) values in these rows.

4. Steps 1−3 are performed in the vertical direction.

5. Both horizontal and vertical difference channels are combined at pixels in S^R using Gaussian weighted averaging filters and weights that depend on the directional gradients (see [80] for details). Finally, the resulting difference values in S^R are added to $\tilde{I}^R$ to provide the estimation of $\hat{G}$ at these positions.

The same steps are performed for pixels in S^B, S^C, and S^O to provide the fully-defined guide image $\hat{G}$. Once the guide image has been estimated, the residual interpolation of all channels I^k, $k \in \{B, C, G, O, R\}$ is performed as in step 2, but using the fully-defined guide image and bilinear interpolations instead of simple linear interpolations.

Demosaicing using adaptive spectral correlation

Like the two algorithms above, demosaicing based on spectral differences estimate a spectral difference channel (difference between two correlated channels) in order to guide the demosaicing process. Jaiswal et al. [43] analyze the conditions that validate the assumptions of spectral difference-based schemes. They show that spectral

correlation highly differs from a database to another, concluding that spectral correlation is image-dependent. Therefore, they propose an adaptive spectral-correlation-based demosaicing scheme that privileges a bilinear interpolation in the case of weak spectral correlation and a spectral difference method in case of high spectral correlation. Their algorithm involves the following steps:

1. The green channel is estimated in the frequency (Fourier) domain by using a circular low-pass filter.

2. The raw image is divided into blocks of size 6×6 pixels. For each block, the missing values of a channel k are both pre-estimated by using a bilinear interpolation of pixels in S^k and a spectral difference scheme with the estimated green channel as a guide.

3. The final estimated values in each block are given by a weighted combination of the values provided by both methods. The weights are determined thanks to a linear minimum mean square error (LMMSE) scheme, *i.e.*, by minimizing the residual values of each block. For this purpose, a fully-defined version of each block has to be known. These blocks are therefore previously interpolated using the GF-based method of Monno et al. [79].

The algorithms presented above use a dominant green band to estimate missing values. They are designed to demosaic VIS5 raw images, but unsuitable to images from IMEC16, IMEC25, and VIS5 MSFAs that do not exhibit any dominant band.

2.3.3 Data-driven demosaicing

Here, we present demosaicing methods that require fully-defined images or that assume sparsity of the raw data.

Demosaicing using learned weights

Aggarwal and Majumdar [1] propose an algorithm based on the prior learning of weights for a given acquisition system. To assess their algorithm, they propose the "uniform" MSFA composed of diagonal stripes, each one sampling a single among $K = 5$ bands. For a given band in this MSFA, the neighborhood within a 3×3 window centered at each pixel is always composed of the same bands at the same positions and includes at least one instance of each band. A missing channel value at a pixel $p \in S^k$ is estimated according to a weighted linear combination of its 3×3 neighbors in the raw image. Thus, for each pixel subset S^k, $K-1$ vectors carrying the 9 weights associated to the 3×3 neighbors of $p \in S^k$ in I^{raw} are used to estimate the $K-1$ missing channel values. These weights that minimize a convex optimization problem are determined using a set of fully-defined multispectral images. They consider both the spectral correlation between different channels in the neighborhood and the spatial correlation of neighboring pixels in I^{raw}. In order to determine

the $K \times (K-1)$ weight vectors efficiently, the training images must be as various as possible in terms of high- and low-frequency characteristics.

Demosaicing using linear minimum mean square error

Amba et al. [3] use linear minimum mean square error (LMMSE) for multispectral demosaicing. LMMSE is a linear estimation method that minimizes the mean square error (MSE), which is a common measure to estimate the reconstruction quality of down-sampled data. The authors consider a spatio-spectral neighborhood in the MSFA raw image for demosaicing by LMMSE optimization [117]. They first express illumination, SSFs of the filters, and raw image values as column vectors. The resulting matrix is coupled with a cross-correlation matrix learned from a given database of reflectance images acquired with the same illumination and SSFs. The fully-defined estimated values are finally given by matrix multiplication of this matrix with the vectorial representation of raw values.

Demosaicing using compressed sensing

Compressed sensing consists in the recovery of a sparse signal from its Gaussian noised under-sampled measurement, by solving a L_1-norm minimization problem. An MSFA raw image can be seen as a sparse signal in the discrete cosine transform (DCT) basis, or the Fourier transform basis. The reconstruction quality of the fully-defined multispectral image thus depends on the sparsity of I^{raw} in the sparsifying basis, and on the incoherence between I^{raw} values and the sparsifying basis that is often satisfied by using a random MSFA pattern. Aggarwal and Majumdar [2] propose two approaches based on compressed sensing. The first one consists in an L_1-norm minimization problem using the DCT as sparsifying basis. The second one considers a Kronecker compressed sensing formulation that uses the representation of the raw image in the Fourier domain and the Kronecker product in the L_1-norm minimization problem.
Shinoda et al. [106] propose to recover a sparse signal using a vectorial total variation (VTV) norm instead of a simple L_1-norm minimization. Total variation norms are essentially L_1-norms of gradients, which makes them more appropriate for demosaicing since gradients are used to preserve edges in the estimated images. More precisely, for a given pixel, VTV is defined as a normalized summation of the gradients at neighboring pixels in all channels. Shinoda et al. [106] extend Ono and Yamada [87]'s VTV-based color demosaicing scheme to the multispectral domain. Their algorithm estimates the fully-defined multispectral image by minimizing the VTV. It is shown to be robust to the incoherence requirement between the MSFA raw image and the sparsifying basis.

Demosaicing using consensus convolutional sparse coding

Sparse coding attempts to parsimoniously represent a group of input vectors by means of a given dictionary. As such, sparse coding is closely related to compressed sensing, but is more general in the sense that it does not necessarily deal with an under-determined set of equations. Given a set of input vectors, it consists in finding another set of vectors (known as dictionary) such that each input vector can be represented as a linear combination of these vectors. The goal is to learn a dictionary that is as small as possible to represent the input vectors. Zeiler et al. [126] propose a convolutional implementation of sparse coding that sparsely encodes a whole image, taking thus spatial arrangement of levels in image into account. Indeed, instead of decomposing a vector as a linear combination of the dictionary elements, convolutional sparse coding (CSC) represents an image as a summation of convolution outputs. However CSC is limited by memory requirements. Thus, the consensus CSC approach splits a single large-scale problem into a set of smaller sub-problems that fit with available memory resources. The author show that their new features lead to significant improvements in a variety of image reconstruction tasks, among which is demosaicing.
The methods presented in the above subsection highly depend on the data: fully-defined images are required by learning-based methods, while sparsity of the raw data is well adapted to compressed sensing-based methods. We thus avoid them in the demosaicing assessment of our considered IMEC16 MSFA since it fits none of these requirements.

2.4 Demosaicing methods for IMEC16 MSFA

In this section, we review methods that are proposed in the literature for MSFAs with no redundant band, and we adapt them to IMEC16 MSFA (see Fig. 2.4b). Note that methods that present low demosaicing performances like [33, 43, 119] are not developed here.

2.4.1 Generic demosaicing methods

Weighted bilinear (WB) interpolation

One of the most simple demosaicing scheme estimates the missing values at each pixel thanks to a bilinear interpolation of the neighboring values. The WB estimates each channel by interpolation of neighboring pixels as [8]:

$$\hat{I}^k_{WB} = \tilde{I}^k * H \,, \tag{2.4}$$

where $*$ is the convolution operator and H is a low-pass filter. For IMEC16 MSFA, H is defined from the following 7×7 unnormalized filter:

$$F = \begin{bmatrix} 1 & 2 & 3 & 4 & 3 & 2 & 1 \\ 2 & 4 & 6 & 8 & 6 & 4 & 2 \\ 3 & 6 & 9 & 12 & 9 & 6 & 3 \\ 4 & 8 & 12 & 16 & 12 & 8 & 4 \\ 3 & 6 & 9 & 12 & 9 & 6 & 3 \\ 2 & 4 & 6 & 8 & 6 & 4 & 2 \\ 1 & 2 & 3 & 4 & 3 & 2 & 1 \end{bmatrix}, \tag{2.5}$$

such that the weight of each neighbor decreases as its spatial distance to the central pixel increases. Note that the filter size is set to the maximum size ensuring that when F is centered at a pixel p, its support window does not include any other pixel of $S^{MSFA(p)}$ (see black pixels in Fig. 2.8c). The normalization of F to get H must take care of the sparse nature of $\tilde{I}^k$ and proceed channel-wise, hence element-wise. The element of H at the a-th row and b-th column, $(a, b) \in \{1, \ldots, 7\}^2$, is then given by:

$$H(a, b) = \frac{F(a, b)}{c_F(a, b)}, \tag{2.6}$$

where the normalization factor c_F is defined at the a-th row and b-th column by:

$$c_F(a, b) = \sum_{\substack{i=1 \\ i \equiv a \,(\mathrm{mod}\, 4)}}^{7} \sum_{\substack{j=1 \\ j \equiv b \,(\mathrm{mod}\, 4)}}^{7} F(i, j). \tag{2.7}$$

The conditions here use the congruence relation $\equiv$ to consider all the pixels that underlie H and belong to the same channel subset as the pixel under $H(a, b)$, which ensures that H is normalized channel-wise according to the 4×4 basic MSFA pattern. Fig. 2.8 shows three (out of sixteen) cases of F (and of H) center locations for the convolution of a sparse channel $\tilde{I}^k$. The elements of F that affect the convolution result correspond to non-zero pixels of $\tilde{I}^k$ (displayed in black), and are normalized by the sum of all such elements of F that overlie the pixels of S^k. Note that for the particular filter F of Eq. (2.5), the normalization factor $c_F(a, b)$ is equal to 16 for all $(a, b) \in \{1, \ldots, 7\}^2$, and elements of H range from $\frac{1}{16}$ (corner element to 1 (central element).

Such interpolation is considered as the most intuitive method for MSFA demosaicing. However, as the estimation of missing values for a channel only uses available values in the same channel, WB interpolation only exploits spatial correlation.

Discrete wavelet transform (DWT) demosaicing

Wang et al. [120] extend the DWT-based CFA demosaicing to MSFA demosaicing. This approach assumes that the low-frequency contents is well estimated by WB

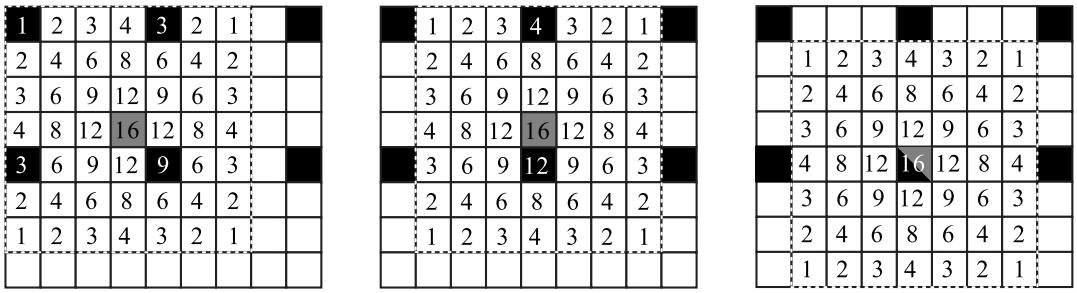


FIGURE 2.8: Normalization of F as H for the convolution of a sparse channel $\tilde{I}^k$ (with non-zero pixels in black) on three cases of filter center locations (in gray). The support window of F (dotted bound) overlies four (a), two (b) or one (the center itself) (c) non-zero pixels according to its center location. Numbers are the elements of F.

interpolation and that the high-frequency contents have to be determined more accurately. The algorithm first estimates a fully-defined multispectral image $\hat{\mathbf{I}}_{WB}$ by WB interpolation, then applies five successive steps to each channel $\hat{I}^k_{WB}$:

1. It decomposes $\hat{I}^k_{WB}$ into K down-sampled (DS) images as shown in Fig. 2.9, so that the l-th DS image of $\hat{I}^k_{WB}$ is made of the pixels in S^l. Note that only the k-th DS image of $\hat{I}^k_{WB}$ contains MSFA (available) values.

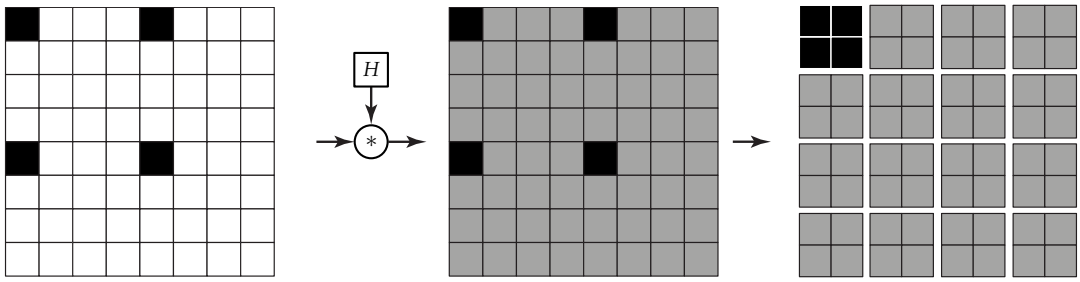


FIGURE 2.9: DS image formation. From left to right: sparse channel $\tilde{I}^k$ (with non-zero pixels in black), fully-defined channel $\hat{I}^k_{WB}$ estimated by WB interpolation, DS images of $\hat{I}^k_{WB}$.

2. It decomposes each DS image into spatial frequency sub-bands by DWT using Haar wavelet (D2).

3. It replaces the spatial high-frequency sub-bands of all (but the k-th) DS images by those of the corresponding DS images of the mid-spectrum channel assuming this is the sharpest one. The latter is associated with the band centered at $\lambda^8 = 551\,nm$ in our considered IMEC16 MSFA.

4. It computes K transformed DS images by inverse DWT.

5. It recomposes the full-resolution channel $\hat{I}^k$ from the K transformed DS images.

2.4.2 Spectral difference-based methods

Spectral difference (SD)

Brauers and Aach [8] propose a method that both uses WB interpolation and takes spectral correlation into account. It was originally designed for a 3×2 MSFA but we adapt it here to our considered MSFA. From an initial estimation $\hat{\mathbf{I}}_{WB}$ (see Eq. (2.4)), it performs the following steps:

1. First, for each ordered pair (k, l) of channel index, it computes the sparse channel difference $\tilde{\Delta}^{k,l}$ given by:

$$\tilde{\Delta}^{k,l} = \tilde{I}^k - \hat{I}^l_{WB} \odot m^k \,, \tag{2.8}$$

that is only non-zero at the pixels in S^k, and a fully-defined channel difference $\hat{\Delta}^{k,l} = \tilde{\Delta}^{k,l} * H$ by WB interpolation (see Eq. (2.4)).

2. Each channel I^k, $k \in \{1, \dots, K\}$ is estimated at each pixel p using channel $I^{MSFA(p)}$ available at p as:

$$\hat{I}^k_p = I^{MSFA(p)}_p + \hat{\Delta}^{k,MSFA(p)}_p. \tag{2.9}$$

Iterative spectral difference (ItSD)

Mizutani et al. [77] improve the SD method by iteratively updating the channel differences. The number of iterations takes the correlation between two channels I^k and I^l into account, that is strong when their associated band centers λ^k and λ^l are close (see Section 1.6.3). The number of iterations $N^{k,l}$ is given by:

$$N^{k,l} = \left\lceil \exp\left(-\frac{|\lambda^l - \lambda^k| - 100}{20\sigma} \right) \right\rceil \,. \tag{2.10}$$

where $\lceil \cdot \rceil$ denotes the ceiling function. $N^{k,l}$ decreases as the distance between λ^k and λ^l increases. For instance, setting $\sigma = 1.74$ as proposed by the authors provides $N^{k,l} = 10$ when $|\lambda^l - \lambda^k| = 20\,nm$ and $N^{k,l} = 1$ when $|\lambda^l - \lambda^k| \geq 100\,nm$. Note that for IMEC16 MSFA, the number of iterations ranges from 1 to 18.
The algorithm initially estimates all sparse channel differences $\tilde{\Delta}^{k,l}(0)$ (see Eq. (2.8)) and all channels $\hat{I}^k(0)$ (see Eq. (2.9)). At each iteration $t > 0$, it first updates the sparse channel difference:

$$\tilde{\Delta}^{k,l}(t) = \begin{cases} \tilde{I}^k - \hat{I}^l(t-1) \odot m^k & \text{if } t \leq N^{k,l}, \\ \tilde{\Delta}^{k,l}(t-1) & \text{otherwise.} \end{cases} \tag{2.11}$$

Then it estimates a fully-defined channel difference as $\hat{\Delta}^{k,l}(t) = \tilde{\Delta}^{k,l}(t) * H$ and each channel as $\hat{I}^k_p(t) = I^{MSFA(p)}_p + \hat{\Delta}^{k,MSFA(p)}_p(t)$ (see Eqs. (2.4) and (2.9)).

2.4.3 Binary tree-based methods

Binary tree-based edge-sensing (BTES)

For each channel, the methods presented previously estimate the missing values simultaneously. To determine the missing values of a channel, Miao et al. [70] propose a scheme divided into four steps for our considered MSFA. At each step t, 2^t values are known in each periodic pattern, either because these are available raw data or they have been previously estimated (see Fig. 2.10). Let us consider the k-th channel ($k \in \{1,\dots,16\}$) and denote as $\hat{S}^k(t)$ (displayed in gray in Fig. 2.10) the subset of pixels whose value of channel I^k is estimated at step t, and $\dot{S}^k(t)$ (displayed in black) the subset of pixels whose value of channel I^k is available in I^{raw} or has been previously estimated: $\dot{S}^k(0) = S^k$ and $\dot{S}^k(t) = \dot{S}^k(t-1) \cup \hat{S}^k(t-1)$ for $t > 0$. At step t, for each $k \in \{1,\dots,16\}$, the values of channel I^k at $p \in \hat{S}^k(t)$ are estimated as:

$$\hat{I}^k_p = \frac{\sum_{q \in \tilde{N}_p(t)} \alpha_q \cdot \dot{I}^k_q}{\sum_{q \in \tilde{N}_p(t)} \alpha_q}, \tag{2.12}$$

where $\dot{I}^k_q$ is available in I^{raw} or has been previously estimated, and $\tilde{N}_p(t)$ is the subset of the four closest neighbors of p that belong to $\dot{S}^k(t)$. These are vertical and horizontal neighbors for $t \in \{1,3\}$ and diagonal ones for $t \in \{0,2\}$ that are located at uniform distance $\Delta = 2 - \lfloor t/2 \rfloor$ from p (see Fig. 2.10), where $\lfloor \cdot \rfloor$ denotes the floor function.

The weights α_q, that embed the edge-sensing part of the algorithm, also depend

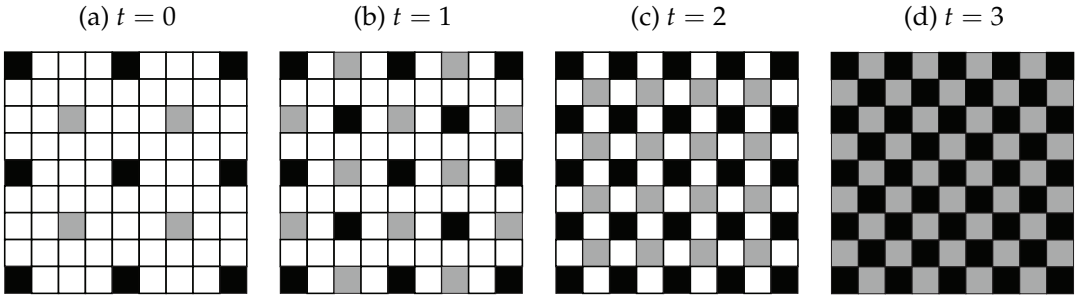


FIGURE 2.10: Estimation of $\hat{I}^k$ in four steps by BTES method. Pixels of $\hat{S}^k(t)$ whose values are estimated at t are displayed in gray, and those of $\dot{S}^k(t)$ whose values are known or previously estimated are displayed in black.

on t and on the direction (horizontal, vertical, or diagonal) given by p and q. Their computation according to the direction is presented in Appendix C.1. In the case of our considered MSFA, many values are missing at $t < 3$ to compute these weights. Miao et al. [70] propose to set missing values to 1, which leads to an unweighted bilinear interpolation at $t = 0$ and $t = 1$.

Multispectral local directional interpolation (MLDI)

Shinoda et al. [105] combine BTES and SD approaches into the MLDI method that

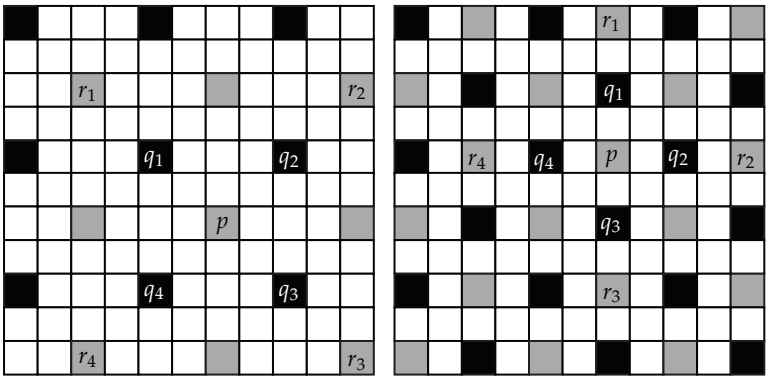
uses four steps like BTES (see Fig. 2.10). Instead of marginally estimating each channel as in Eq. (2.12), the authors compute the difference between the k-th channel being estimated and the available one at each pixel in I^{raw}. The difference value at $p \in \hat{S}^k(t)$ is computed following Eq. (2.12) as:

$$\hat{D}_p^{k,MSFA(p)} = \frac{\sum_{q \in \tilde{N}_p(t)} \beta_q \cdot \dot{D}_q^{k,MSFA(p)}}{\sum_{q \in \tilde{N}_p(t)} \beta_q}, \tag{2.13}$$

where $\dot{D}_q^{k,MSFA(p)} = \dot{I}_q^k - \frac{1}{2}\left(I_p^{MSFA(p)} + \dot{I}_r^{MSFA(p)}\right)$ is a directional difference computed at one neighbor q of p among its four closest ones that belong to $\dot{S}^k(t)$. The pixel r is the symmetric of p with respect to q, so that r belongs to $\dot{S}^{MSFA(p)}(t)$ (see Fig. 2.11). The value of the k-th channel at p is finally estimated as:

$$\hat{I}_p^k = \hat{D}_p^{k,MSFA(p)} + I_p^{MSFA(p)}. \tag{2.14}$$

Note that each weight β_q in Eq. (2.13) both depends on t and on the direction given by p and q (see Appendix C.2).



(a) $t = 0$ (b) $t = 1$

FIGURE 2.11: Estimation of $\hat{I}^k$ by MLDI method (first two steps only) at $p \in S^l$ using the neighbors $q \in \dot{S}^k(t)$ and $r \in \dot{S}^{MSFA(p)}(t)$. $\dot{S}^k(t)$ is displayed in black and $\hat{S}^k(t)$ in gray.

Shinoda et al. [105] also propose a post-processing of an initial estimation $\hat{\mathbf{I}}$ that updates each estimated channel $\hat{I}^k$ at each pixel p using Eq. (2.14) but now by considering a neighborhood N_p associated with the support $\mathcal{N}^{8,1}$ made of the eight closest neighbors of p, and:

$$\hat{D}_p^{k,MSFA(p)} = \frac{\sum_{q \in N_p} \beta_q \cdot \left(\hat{I}_q^k - \hat{I}_q^{MSFA(p)}\right)}{\sum_{q \in N_p} \beta_q}. \tag{2.15}$$

2.5 From raw to pseudo-panchromatic image (PPI)

When channels are spectrally distant, *i.e.*, when the centers of bands that are associated with channels are distant, they are more correlated with the pseudo-panchromatic

image (PPI) than between each other [73]. This interesting property allows us to expect enhanced fidelity of PPI-based demosaicing methods. We first show the limitations of existing demosaicing methods for our considered MSFA in Section 2.5.1. Then we define the PPI and study its properties in Section 2.5.2. We finally introduce how to estimate the fully-defined PPI from a raw image in Section 2.5.3.

2.5.1 Limitations of existing methods

The previous methods can be described according to the properties of channels (spatial or/and spectral correlation) that they exploit (see Table 2.1). By using bilinear interpolation in at least one initial step, all methods assume a strong spatial correlation among values within each channel. But Section 1.6.2 experimentally shows that spatial correlation decreases as the distance between neighboring pixels (or the basic MSFA pattern size) increases. Hence, the 4×4 basic pattern of our considered MSFA weakens the spatial correlation assumption. Besides, this assumption does not hold at object boundaries [13]. An edge-sensitive mechanism is then required to avoid interpolating values across boundaries. Two methods embed edge-sensitive weights in bilinear interpolation, either on each channel (BTES) or channel difference (MLDI). SD, ItSD and MLDI methods are based on channel differences assuming that channels are correlated at each pixel. But Section 1.6.3 shows that spectral correlation between channels decreases as the spectral distance between their associated band centers increases. Only ItSD relies on the property stating that channels associated with nearby band centers are more correlated than channels associated with distant ones.

	Spatial correlation		Spectral correlation		
	Bilinear interpolation	Edge-sensing	Channel difference	Nearby band centers	Frequency domain
WB	✓				
DWT	✓				✓
SD	✓		✓		
ItSD	✓		✓	✓	
BTES	✓	✓			
MLDI	✓	✓	✓		

TABLE 2.1: Properties used by existing demosaicing methods. WB: weighted bilinear [8], DWT: discrete wavelet transform [120], SD: spectral difference [8], ItSD: iterative spectral difference [77], BTES: binary tree-based edge-sensing [70], MLDI: multispectral local directional interpolation [105].

Several CFA and MSFA demosaicing schemes exploit the dominant channel (either implicitly or explicitly as in [49]) for demosaicing because it carries most of image structures [43, 78–80]. Because our considered MSFA exhibits no dominant band, we propose to compute a PPI using all raw image information, and to use it for MSFA demosaicing.

2.5.2 PPI definition and properties

The PPI is defined at each pixel p as the mean value over all channels of a fully-defined multispectral image [14]:

$$I_p^{PPI} = \frac{1}{K} \sum_{k=1}^{K} I_p^k. \tag{2.16}$$

The following demosaicing proposals assume that the PPI is strongly correlated with all channels. To assess this assumption, we propose to compute the average correlation coefficient $C\left(I^i, I^{PPI}\right)$ (see Eq. (1.8)) between each channel and the PPI CAVE image set presented in Section 1.6.1. The results (see Fig. 2.12b) show that the channels are strongly correlated with the PPI.

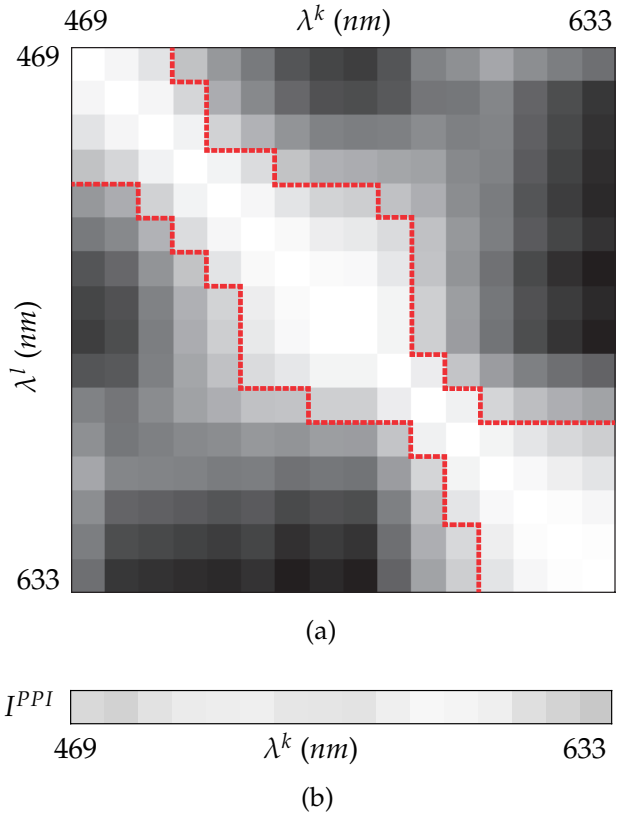


FIGURE 2.12: Correlation between channels I^k and I^l (a) and between I^k and I^{PPI} (b). Values are averaged over CAVE image set of Section 1.6.1, range between 0.76 (black) and 1.0 (white). Values of (b) are reported column-wise as dashed red lines on (a).

To compare this correlation with inter-channel correlation (see Section 1.6.3), the dashed red lines in Fig. 2.12a show the bounds $\left\{\lambda^l : C\left(I^k, I^l\right) \geq C\left(I^k, I^{PPI}\right)\right\}_{l=1}^{K}$ for each $k = \{1, \ldots, K\}$ (column-wise). When band centers are distant ($\lambda^k \gg \lambda^l$ or $\lambda^k \ll \lambda^l$), channel I^k is more correlated with I^{PPI} than with I^l. This interesting property allows us to expect enhanced fidelity of PPI-based demosaicing methods that would exploit inter-channel differences. We now introduce how to estimate the PPI from a raw image.

2.5.3 PPI estimation

Since the value of a single channel is available at each pixel in I^{raw}, we rely on the spatial correlation assumption of the fully-defined PPI (i.e., we assume that PPI values of neighboring pixels are strongly correlated). That leads us to estimate the PPI from I^{raw} by applying an averaging filter M [71]. This filter has to take all channels into account while being as small as possible to avoid estimation errors. Its size is hence that of the smallest odd-size neighborhood window including at least one pixel in all MSFA subsets $\{S^k\}_{k=1}^K$. Each element of M is set to $\frac{1}{n}$, where n is the number of times when the MSFA band associated with the underlying neighbor occurs in the support window of M. This filter is normalized afterwards so that all its elements sum up to 1.
For our considered IMEC16 MSFA (see Fig. 2.13), the size of M is 5×5 and centering M at pixel $(2,2)$ (that samples channel I^{10}) yields four available levels for channel 7, two levels for channels 3, 5, 6, 8, 11 and 15, and a single level for the other channels.

7	8	6	5	7	···
15	16	14	13	15	···
11	12	10	9	11	···
3	4	2	1	3	···
7	8	6	5	7	···
···	···	···	···	···	···

FIGURE 2.13: Raw image from IMEC16 MSFA. Numbers are band indexes.

Considering any other central pixel in I^{raw} provides the same filter M for such 4×4 non-reduncant MSFA, namely:

$$M = \frac{1}{16} \cdot \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 1 & 1 & 1 & \frac{1}{2} \\ \frac{1}{2} & 1 & 1 & 1 & \frac{1}{2} \\ \frac{1}{2} & 1 & 1 & 1 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} \end{bmatrix} = \frac{1}{64} \cdot \begin{bmatrix} 1 & 2 & 2 & 2 & 1 \\ 2 & 4 & 4 & 4 & 2 \\ 2 & 4 & 4 & 4 & 2 \\ 2 & 4 & 4 & 4 & 2 \\ 1 & 2 & 2 & 2 & 1 \end{bmatrix}. \tag{2.17}$$

A first estimation of the PPI is then computed as [71]:

$$\bar{I}^{PPI} = I^{raw} * M. \tag{2.18}$$

M is an averaging filter that may provide a smooth image. We instead propose in [73] to use local directional information to obtain another estimation $\hat{I}^{PPI}$ of the PPI that is sharper than $\bar{I}^{PPI}$. For this purpose, we consider the MSFA-specific neighborhood $\tilde{N}_p$ of each pixel p made of the eight closest pixels of p that also belong to

$S^{MSFA(p)}$ (see Fig. 2.14).

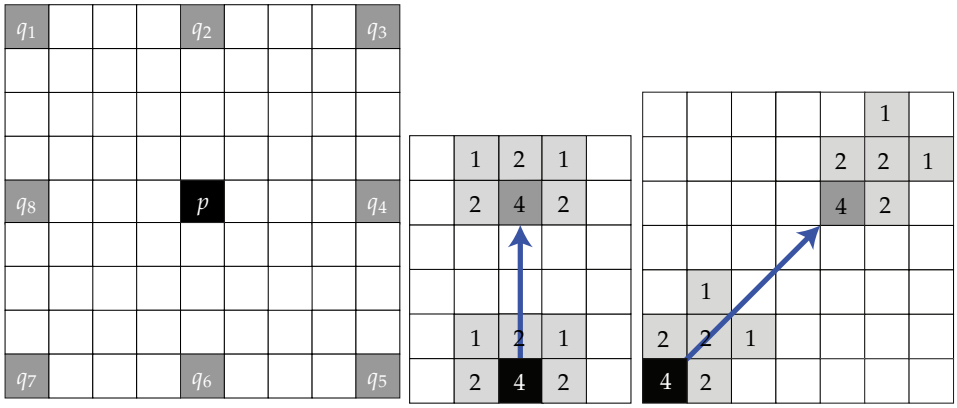


FIGURE 2.14: Proposed PPI estimation: neighborhood $\tilde{N}_p$ (in gray) of p (in black) (a), weight γ_q computation (see Eq. (2.19)) for $q = p + (0, -4)$ (b) and $q = p + (4, -4)$ (c). Numbers are coefficients $\kappa(u, v)$.

For each pixel $q \in \tilde{N}_p$, we compute a weight γ_q using the raw image I^{raw} as:

$$\gamma_q = \left(1 + \sum_{v=-1}^{1} \sum_{u=0}^{1} \kappa(u,v) \cdot |I^{raw}_{p+\rho(u,v)} - I^{raw}_{q+\rho(u,v)}|\right)^{-1}. \tag{2.19}$$

Here, $\kappa(u, v) = (2 - u) \cdot (2 - |v|) \in \{1, 2, 4\}$ is the coefficient associated with the absolute difference between the values of pixels $p + \rho(u, v)$ and $q + \rho(u, v)$ given by the following relative coordinates:

$$\rho(u,v) = \begin{cases} \left(\frac{u\cdot\delta_x + v\cdot\delta_y}{4}, \frac{u\cdot\delta_y + v\cdot\delta_x}{4}\right) & \text{if } \delta_x \cdot \delta_y = 0, \\ \left(\frac{\left(u + |v| \cdot \frac{1-v}{2}\right)\cdot\delta_x}{4}, \frac{\left(u + |v| \cdot \frac{1+v}{2}\right)\cdot\delta_y}{4}\right) & \text{otherwise,} \end{cases} \tag{2.20}$$

where $(\delta_x, \delta_y) \in \{-4, 0, 4\}^2$ are the coordinates of q relative to p. Figs. 2.14b and 2.14c show two examples of weight computation according to one of the eight cardinal directions given by the central pixel p and its neighbor q. Note that to carefully consider the direction from p to q, we only use some of the neighboring pixels of p and q, namely the five ones given by that direction and defined by $\rho(u, v)$. A weight γ_q ranges from 0 to 1 and is close to 0 when the directional variation of available values between p and q is high (see Eq. (2.19)).

We then propose to compute the local difference $\Delta_p[I^{raw}]$ between the value of any pixel p in I^{raw} and the weighted average value of its eight closest neighbors associated with the same available channel:

$$\Delta_p[I^{raw}] = I^{raw}_p - \frac{\sum_{q\in\tilde{N}_p} \gamma_q \cdot I^{raw}_q}{\sum_{q\in\tilde{N}_p} \gamma_q}. \tag{2.21}$$

Since the PPI is the average value over all channels at each pixel, we can assume that Δ_p is invariant against the PPI. Then $\Delta_p[I^{raw}] = \Delta_p[\bar{I}^{PPI}]$, which provides a new

estimation of the PPI at each pixel:

$$\hat{I}_p^{PPI} = I_p^{raw} + \frac{\sum_{q \in \tilde{N}_p} \gamma_q \cdot \left(\bar{I}_q^{PPI} - I_q^{raw} \right)}{\sum_{q \in \tilde{N}_p} \gamma_q}. \tag{2.22}$$

Correlation with estimated PPI

To validate the assumption about strong correlation between the values of each channel I^k that are available in I^{raw} and the estimated PPI $\hat{I}^{PPI}$, we consider the following Pearson correlation coefficient:

$$C_{S^k}\left(I^{raw}, \hat{I}^{PPI}\right) = \frac{\sum\limits_{p \in S^k} \left(I_p^{raw} - \mu_{S^k}^{raw} \right) \left(\hat{I}_p^{PPI} - \mu_{S^k}^{PPI} \right)}{\sqrt{\sum\limits_{p \in S^k} \left(I_p^{raw} - \mu_{S^k}^{raw} \right)^2} \sqrt{\sum\limits_{p \in S^k} \left(\hat{I}_p^{PPI} - \mu_{S^k}^{PPI} \right)^2}}, \tag{2.23}$$

where $\mu_{S^k}^{raw}$ and $\mu_{S^k}^{PPI}$ are the average values of I^{raw} and $\hat{I}^{PPI}$ at the pixels in S^k.
We compute the average values of $C_{S^k}(I^{raw}, \hat{I}^{PPI})$ and of $C\left(I^k, I^{PPI}\right)$ between each fully-defined channel I^k and the PPI (see Section 2.5.2) on the CAVE set (see Section 1.6.1). The results (not displayed here) show that $C_{S^k}(I^{raw}, \hat{I}^{PPI}) = 0.979$ and $C\left(I^k, I^{PPI}\right) = 0.980$ on average over all channels. These correlation coefficients differ by less than $7 \cdot 10^{-3}$ channel-wise for all images. We can conclude that each channel either in the fully-defined image or in the raw image is strongly correlated with the estimated PPI. This leads us to exploit the estimated PPI for demosaicing.

2.6 PPI-based demosaicing

The faithful estimation of the PPI should make it effective for demosaicing since its high frequencies can be used to guide the estimation of channels. Below we propose both an adaptation of two existing demosaicing methods (DWT and BTES) to the PPI in Section 2.6.1 and Section 2.6.2, and a new demosaicing method based on PPI difference in Section 2.6.3.

2.6.1 Using PPI in DWT (PPDWT)

For the considered IMEC16 MSFA, DWT uses the high-frequency contents of the mid-spectrum channel estimated by bilinear interpolation to estimate the other channels (see Section 2.4.1). Since the PPI has similar information than the mid-spectrum channel and is (hopefully) better estimated, we propose to replace the spatial high-frequency sub-bands by those of the PPI instead of the mid-spectrum channel (see step 3 of Section 2.4.1) [73]. The adapted method is referred to as PPDWT and assessed in Chapter 3.

2.6.2 Using PPI in BTES (PPBTES)

When a dominant band is present in the MSFA (*e.g.*, green band in VIS5), Miao et al. [70] take advantage of associated channel by estimating it first to compute the weights α_q (see Eq. (2.12)). We follow the same strategy and use the PPI as a dominant channel. For this purpose we propose weights (see Appendix C.3) that consider all the possible cases occurring with IMEC16 MSFA. Fig. 2.15 shows the pixels used to compute these weights as dotted crosses on two examples: the first diagonal direction (for $t \in \{0,2\}$) and the horizontal direction (for $t \in \{1,3\}$). Among them, the crosses that do not overlie black (known) pixels at $t < 3$ correspond to unknown values. We then replace them by the values of the estimated PPI at the same position (see Appendix C.3) [72]. This PPI-adapted method is referred to as PPBTES and assessed in Chapter 3.

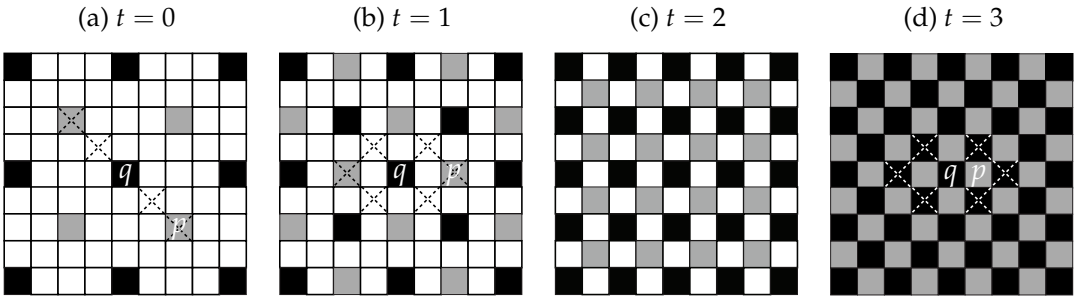


FIGURE 2.15: Estimation of $\hat{I}^k$ in four steps by BTES method. At each step t, the subset of pixels whose values are known or previously estimated are displayed in black. The subset of pixels whose values are estimated at t are displayed in gray. Considering a pixel p to be estimated, the pixels used to compute the weight α_q at neighbor q are q itself and those marked with a dotted cross.

2.6.3 Proposed PPI difference (PPID)

Instead of using the difference between channels as in SD (see Section 2.4.2), we propose to compute the difference between each channel and the PPI. The algorithm is divided into four successive steps:

1. First, it estimates the PPI image $\hat{I}^{PPI}$ (see Eq. (2.22)).

2. Second, it computes the sparse difference $\tilde{\Delta}^{k,\text{PPI}}$ between each available value in I^{raw} and the PPI at pixels in S^k, $k = \{1,\dots,K\}$:

$$\tilde{\Delta}^{k,\text{PPI}} = \tilde{I}^k - \hat{I}^{PPI} \odot m^k, \tag{2.24}$$

where $\tilde{I}^k = I^{raw} \odot m^k$.

3. Third, it uses the local directional weights computed according to Eq. (2.19) (whereas [71] directly uses H of Eq. (2.6)) to estimate the fully-defined difference $\hat{\Delta}^{k,\text{PPI}}$ by adaptive WB interpolation as:

$$\hat{\Delta}_p^{k,\text{PPI}} = \tilde{\Delta}^{k,\text{PPI}} * H_p. \tag{2.25}$$

Each element $(a,b) \in \{1,\dots,7\}^2$ of the new 7×7 adaptive convolution filter H_p is given by:

$$H_p(a,b) = \frac{F(a,b) \cdot \Gamma_p(a,b)}{\sum_{\substack{i=1 \\ i \equiv a \,(\text{mod } 4)}}^{7} \sum_{\substack{j=1 \\ j \equiv b \,(\text{mod } 4)}}^{7} F(i,j) \cdot \Gamma_p(i,j)}, \tag{2.26}$$

where $F(a,b)$ is defined by Eq. (2.5) and the denominator is a channel-wise normalization factor like in non-adaptive WB interpolation (see Eq. (2.7)). The 7×7 filter Γ_p contains the local directional weights according to each cardinal direction given by the central pixel p and its neighbor q underlying the filter elements:

$$\Gamma_p = \left[\begin{array}{ccc|c|ccc} & & & \gamma_{q_2} & & & \\ & \gamma_{q_1} \cdot J_3 & & \gamma_{q_2} & & \gamma_{q_3} \cdot J_3 & \\ & & & \gamma_{q_2} & & & \\ \hline \gamma_{q_8} & \gamma_{q_8} & \gamma_{q_8} & 1 & \gamma_{q_4} & \gamma_{q_4} & \gamma_{q_4} \\ \hline & & & \gamma_{q_6} & & & \\ & \gamma_{q_7} \cdot J_3 & & \gamma_{q_6} & & \gamma_{q_5} \cdot J_3 & \\ & & & \gamma_{q_6} & & & \end{array}\right], \tag{2.27}$$

where $q_1 = p + (-4,-4), \dots, q_8 = p + (-4,0)$ (see Fig. 2.14) and J_3 denotes the 3×3 all-ones matrix. By design, Γ_p splits H_p into eight areas matching with the directions given by p and its eight neighbors that belong to $\tilde{N}_p$. Note that H_p depends on p because γ_q also does.

4. Finally, it estimates each channel by adding the PPI and the difference:

$$\hat{I}^k = \hat{I}^{PPI} + \hat{\Delta}^{k,\text{PPI}}. \tag{2.28}$$

The proposed demosaicing method based on PPI difference (PPID) is outlined in Fig. 2.16.

2.7 Conclusion

The single sensor of an MSFA-based camera captures the radiance spectrum through an MSFA whose filter elements are sensitive to specific narrow spectral bands. Thus, only one value is available at each pixel of the acquired raw image according to the

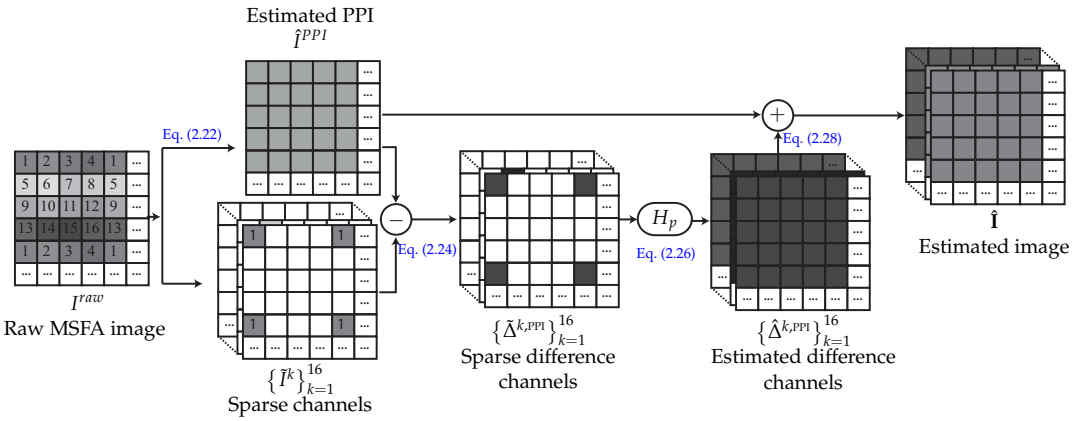


FIGURE 2.16: Outline of the proposed PPID demosaicing method.

MSFA pattern. In order to provide a fully-defined multispectral image, a demosaicing step is performed. Demosaicing is strongly related to the MSFA design which has been shown to result from a trade-off between spatial and spectral resolutions. Among MSFAs proposed in the literature, some contain a dominant band (*e.g.*, VIS5) or not (*e.g.*, IMEC16). To demosaic VIS5 raw images, authors first estimate the dominant channel then use it as a guide for demosaicing. However, no dominant band is available in IMEC16 raw images. We have detailed all demosaicing methods that can be applied to our considered IMEC16 raw images, excluding methods that highly depend on the data since they require data sparsity or fully-defined images.

All state of the art methods use properties of channels (spatial or/and spectral correlation). By using bilinear interpolation (WB), all methods use spatial correlation among values within each channel. Few methods (BTES and MLDI) apply an edge-sensitive mechanism in order to use spectral correlation more faithfully. Assuming that channels are correlated at each pixel, three methods (SD, ItSD and MLDI) use spectral correlation as channel differences and DWT uses the frequency domain to homogenize the high-frequency information among channels. ItSD iterates the estimation according to the property stating that channels associated with nearby band centers are more correlated than channels associated with distant ones.

Like several MSFA demosaicing schemes that exploit the properties of a dominant channel for demosaicing because it carries most of image structures, we propose to compute a PPI from the raw image and to use it for demosaicing. To estimate the PPI from the raw MSFA image, a simple averaging filter can be used in case of low inter-channel correlation but it may fail to restore the high-frequency contents of the reference image. We therefore propose to use local directional variations of raw values to estimate the edge information more accurately in the PPI. We then incorporate the PPI into existing DWT-based and BTES-based methods and propose a new demosaicing method based on PPI difference. As recently shown by Jaiswal et al. [43], spectral correlation is image-dependent. Thus spectral difference-based schemes have to be locally adapted with respect to the considered raw image. Future works could focus on the study of local correlation in raw image in order to

decide whether to use the PPI or not for demosaicing.
The next chapter focuses on the assessment of methods related to IMEC16 MSFA and on the effect of acquisition properties on demosaicing performances.

Chapter 3

Demosaicing assessment and robustness to acquisition properties

Contents

3.1 Introduction

In order to assess multispectral demosaicing algorithms, we only consider the IMEC16 MSFA formed from the square 4×4 basic pattern of Fig. 2.4b that privileges spectral resolution. Indeed, this MSFA is incorporated in a snapshot camera that is available at IrDIVE platform and that embeds no demosaicing method. Moreover we consider the CAVE database that is composed of scenes in which objects have sharp transitions and various properties [124]. Note that CAVE is the most widely used database for multispectral demosaicing assessment because of the presence of sharp details whereas images in other databases are often blurred (see Section 1.6.2). After a presentation of the experimental procedure, in Section 3.2 we assess the different demosaicing methods of Sections 2.4 and 2.6 on the CAVE image set of Section 1.6.1 (images of CAVE scenes that are simulated using IMEC16 SSFs and D65 illumination). The different methods are assessed in Section 3.3 under various illuminations to highlight the effects of acquisition properties on demosaicing performances. We then propose pre- and post-normalization steps that make demosaicing robust to acquisition properties in Section 3.4. Finally, we discuss about demosaicing performances on different MSFA-based cameras in Section 3.5.

3.2 Demosaicing quality assessment on CAVE image set

We objectively compare the demosaicing performances reached by PPI-based methods (see Section 2.6) with those provided by the existing methods described in Section 2.4. The experimental procedure is presented in Section 3.2.1 and the results in terms of PSNR, color difference, and computation time, are shown in Section 3.2.2. Images obtained by either selecting one spectral channel or color simulation by projection in sRGB color space are visually compared in Section 3.2.3.

3.2.1 Experimental procedure

Let us consider the multispectral image $\mathbf{I} = \{I^k\}_{k=1}^K$ made of K fully-defined channels associated to the K bands of a single-sensor MSFA camera. Although this image cannot be provided by such cameras, it is called the reference image because it is often used as a reference to assess the demosaicing quality. To get reference images, we simulate the 32 multispectral images of size 512×512 pixels from the CAVE database [124] with the CIE D65 illuminant and IMEC16 SSFs as described in Section 1.6.1. To obtain raw images, we spectrally sample reference images according to the IMEC16 MSFA whose pattern is shown in Fig. 2.4b. The fully-defined image is then estimated from the raw image using any demosaicing method and compared

with the reference image in order to assess demosaicing performances. This experimental procedure for demosaicing assessment is shown in Fig. 3.1.

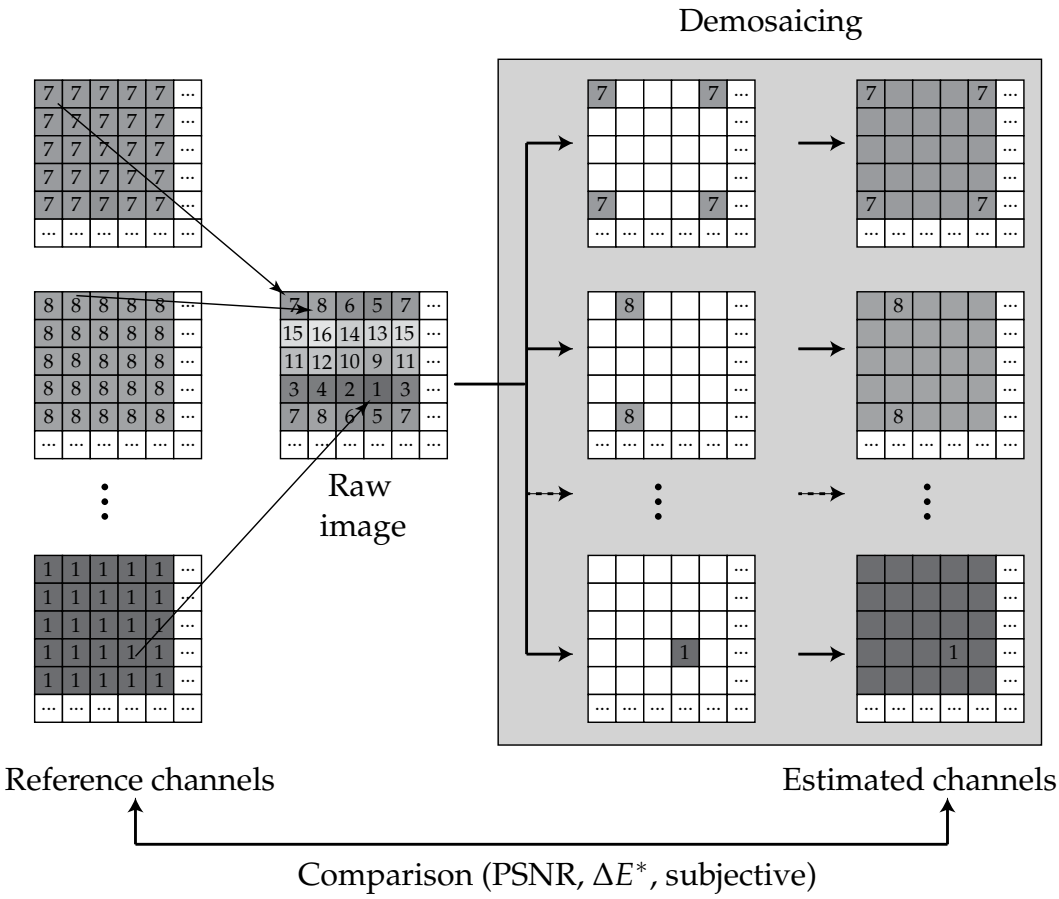


FIGURE 3.1: Experimental procedure for demosaicing assessment.

We have implemented all demosaicing methods of Sections 2.4 and 2.6 in Java under the open-source image analysis software ImageJ [95, 99]. The Java bytecodes and an ImageJ macro are available as supplementary material in [73]. These codes allow the user to create reference radiance images (CAVE set) from CAVE database and to demosaic them using the methods presented in Sections 2.4 and 2.6.

3.2.2 Objective assessment

PSNR assessment

The peak signal-to-noise ratio (PSNR), initially designed to estimate the distortion level of a compressed image with respect to its reference, is the most widely used metric to assess demosaicing performances. Another full-reference metric proposed in the literature is the Structural SIMilarity (SSIM) index between two images, with the aim of being closer to the perceived quality of an image than the PSNR. Although these metrics are not sensitive to degradations in the same way, PSNR values can be predicted from SSIM ones and vice versa [39]. In this manuscript, we only consider

the PSNR for assessment of demosaicing methods.
The quality of a multispectral image $\hat{\mathbf{I}}$ estimated by demosaicing is evaluated as the average PSNR over all reference and estimated channels as follows:

$$PSNR\left[\mathbf{I},\hat{\mathbf{I}}\right] = \frac{1}{K}\cdot\sum_{k=1}^{K} 10\cdot\log_{10}\left(\frac{(max_{p\in S_c} I_p^k)^2}{\frac{1}{|S_c|}\sum_{p\in S_c}\left(\hat{I}_p^k - I_p^k\right)^2}\right). \tag{3.1}$$

To avoid border effects, we only consider the subset S_c of pixels that represent the 500×500 central pixels of each image (that represents 95% of pixels in original image) in the PSNR computation. Because $max_{p\in S_c} I_p^k$ can be low in case of low energy in the spectral band associated to channel I^k, Eq. (3.1) takes into account this actual maximal value rather than the theoretical one (255) to avoid misleading PSNR values.

	WB	DWT	PPDWT	SD	ItSD	BTES	PPBTES	MLDI	PPID
PSNR	31.88	31.09	33.73	34.02	35.13	31.99	34.10	**37.17**	37.06

TABLE 3.1: Average PSNR (dB) over the 32 CAVE set images estimated by each demosaicing method. The best result is displayed as bold. WB: weighted bilinear [8], DWT: discrete wavelet transform [120], PPDWT: PPI-adapted DWT, SD: spectral difference [8], ItSD: iterative spectral difference [77], BTES: binary tree-based edge-sensing [70], PPBTES: PPI-adapted BTES [72], MLDI: multispectral local directional interpolation [105], PPID: pseudo-panchromatic image difference [73].

Table 3.1 displays the PSNR values provided by the demosaicing methods on average over the 32 images. WB and BTES methods give poor (similar) results because they only use spatial correlation. Using $\hat{I}^{PPI}$ as a guide for weight estimation in PPBTES improves the PSNR by 2.11 dB. DWT does not perform well because it uses the information of the bilinearly interpolated channel $\hat{I}^8_{WB}$ for the estimation of all channels. The PPI more efficiently represents the high-frequency contents of the multispectral image since the PSNR provided by PPDWT is improved by 2.64 dB with respect to DWT. Similarly, taking spectral correlation into account through spectral differences (SD) among all channels improves the PSNR of WB interpolation by 2.14 dB. Using channels associated with nearby band centers to iteratively update the results is an interesting approach that improves SD by 1.11 dB in ItSD. Using local directional information is especially efficient: this improves SD by 3.15 dB in MLDI that exploits inter-channel correlation, and by 3.04 dB in PPID that exploits the correlation between each channel and the PPI.

ΔE^* assessment

According to the MSFA pattern, the sampling is shifted from a sparse channel to another. This shift often causes misalignment between reconstructed channels. The

PSNR metric compares each channel separately, thus disregarding misalignment between channels. We then propose to project the 32 images in the L*a*b* space (see Appendix A) in order to compute the CIE color difference. Indeed, misalignment of channels generates color artifacts that can be assessed using the CIE color difference [63]. The first metric for color difference measurement, proposed in 1976, is named ΔE^*_{ab} and computed in L*a*b* color space as follows [11]:

$$\Delta E^*_{ab} = \sum_{p \in S_c} \frac{\sqrt{(I^{L^*}_p - \hat{I}^{L^*}_p)^2 + (I^{a^*}_p - \hat{I}^{a^*}_p)^2 + (I^{b^*}_p - \hat{I}^{b^*}_p)^2}}{|S_c|}. \tag{3.2}$$

ΔE^*_{ab} has further been extended as ΔE^*_{00} in 2000 to consider perceptual non-uniformities [102]. Table 3.2 displays the color difference values provided by the two metrics on average over the 32 images, both providing the same ranking of methods. Results show that using the frequency domain in DWT and PPDWT methods provides poor results. Except for PPBTES that provides nice results, other methods have similar ranking than with the PSNR, PPID having the lowest ΔE^* values.

	WB	DWT	PPDWT	SD	ItSD	BTES	PPBTES	MLDI	PPID
ΔE^*_{ab}	2.83	3.33	2.97	2.71	2.65	2.76	2.05	1.86	**1.81**
ΔE^*_{00}	2.58	2.93	2.62	2.50	2.44	2.57	1.94	1.79	**1.71**

TABLE 3.2: Average ΔE^*_{ab} and ΔE^*_{00} over the 32 images of CAVE set estimated by each demosaicing method. The best result is displayed as bold.

Computation time

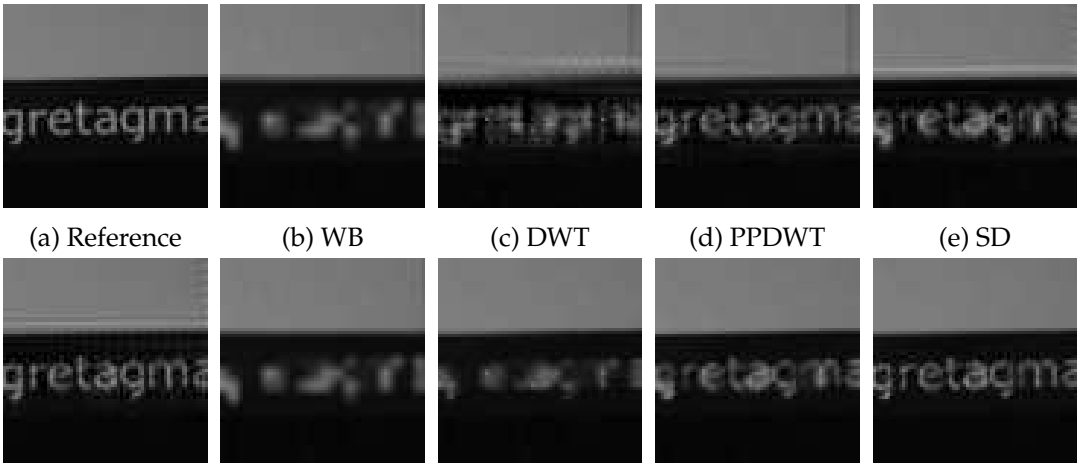
We propose to roughly consider the computation time required by the various methods. To be independent of computer performances, programming language, and image size, we propose to compare each method with WB interpolation that is the simplest method. We therefore run each demosaicing method a hundred times on a single constant image whose values are all set to 127, and we normalize the resulting computation time with that of WB interpolation. Table 3.3 shows the corresponding ratios, which gives some insights about the algorithmic complexity of the tested methods. By putting them in relation with PSNR performances, we conclude that our proposed PPID method gives nice results and requires less computation time than MLDI that is the best state of the art method in terms of PSNR.

WB	DWT	PPDWT	SD	ItSD	BTES	PPBTES	MLDI	PPID
1.00	6.16	6.26	17.20	97.49	2.97	3.07	12.76	7.87

TABLE 3.3: Computation time ratios of the methods with respect to WB.

3.2.3 Subjective assessment

With a single spectral channel



(a) Reference (b) WB (c) DWT (d) PPDWT (e) SD

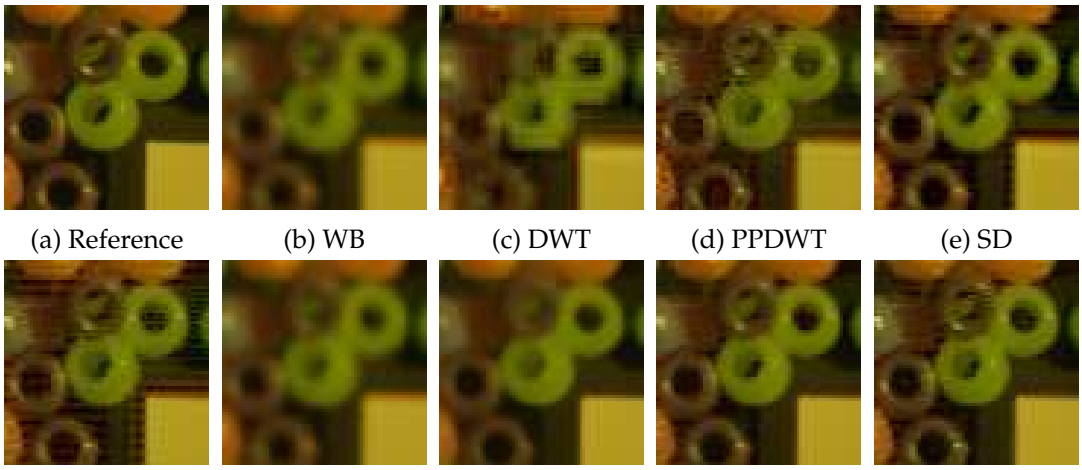
(f) ItSD (g) BTES (h) PPBTES (i) MLDI (j) PPID

FIGURE 3.2: Central extract of reference I^9 (a) and estimated $\hat{I}^9$ (b to j) channel of "Fake and Real Lemons" image from the CAVE set.

We select an area of 70×70 pixels from the "Fake and Real Lemons" image from the CAVE set, and we visually compare the results of each method in the channel associated with the band centered at $\lambda^9 = 552\,nm$ (that corresponds to the central wavelength band). Fig. 3.2 shows that the images estimated by WB (Fig. 3.2b), BTES (Fig. 3.2g) and PPBTES (Fig. 3.2h) methods are strongly blurred. Due to the wavelet transform, DWT (Fig. 3.2c) introduces some artifacts and blurred edges that are reduced in PPDWT (Fig. 3.2d). The SD (Fig. 3.2e) and ItSD (Fig. 3.2f) methods suffer from severe zipper effect (alternating colors along the edges). All demosaicing artifacts are fairly reduced with MLDI (Fig. 3.2i) and PPID (Fig. 3.2j), that notably produce far less blurry images than WB. PPID provides sharper edges than MLDI but is slightly more sensitive to zipper effect.

With color images

As previously mentioned, misaligned spectral channels generate color artifacts when they are converted into a color space. Thus we convert the reference and estimated area of 70×70 pixels from "Beads" image to the sRGB color space (see Appendix A). The results displayed in Fig. 3.3 show that WB and BTES provide highly blurred images with color artifacts, and that PPBTES both reduces this blur effect and color artifacts. PPDWT successfully reduces the image blur generated by DWT. SD is prone to false colors that are somewhat reduced by ItSD at the expense of severe zipper artifacts. These artifacts are fairly reduced with MLDI and PPID, that provide sharp edges.

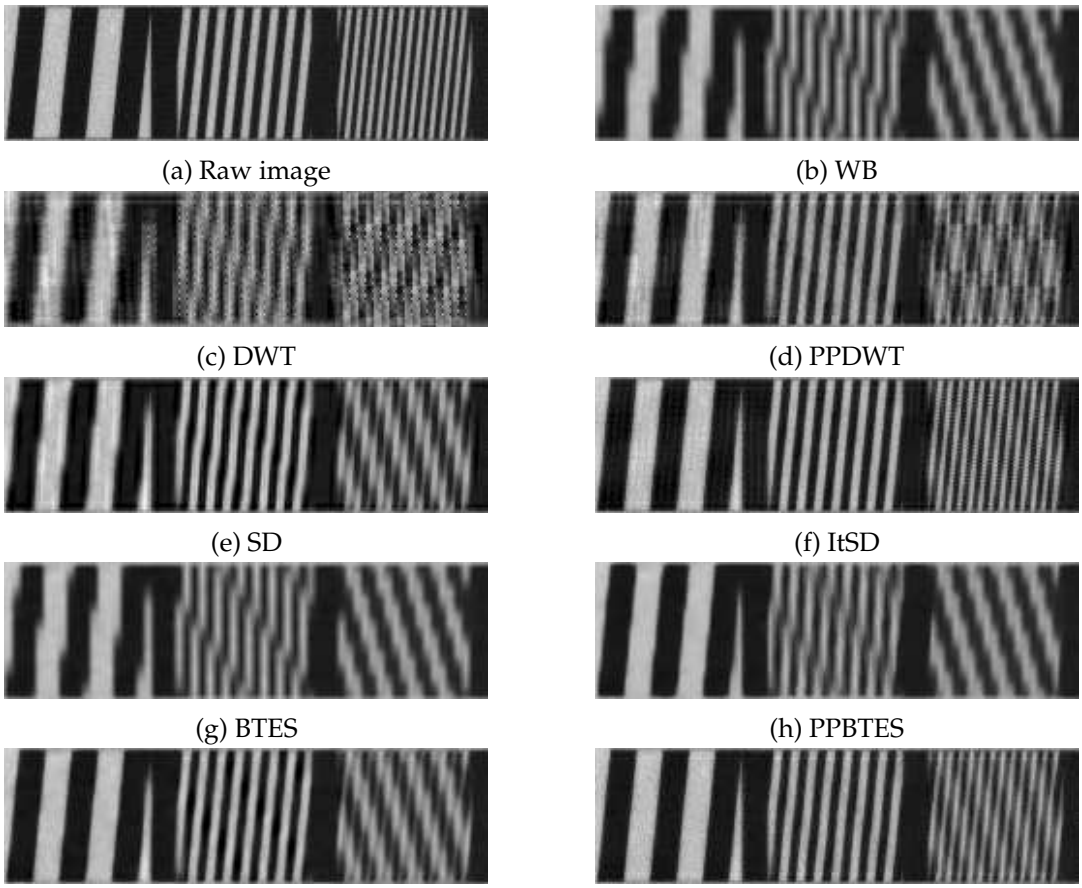


(a) Reference (b) WB (c) DWT (d) PPDWT (e) SD

(f) ItSD (g) BTES (h) PPBTES (i) MLDI (j) PPID

FIGURE 3.3: sRGB renderings of a reference (a) and demosaiced (b to j) central extract of "Beads" image from CAVE set..

With acquired IEEE Target images



(a) Raw image (b) WB

(c) DWT (d) PPDWT

(e) SD (f) ItSD

(g) BTES (h) PPBTES

(i) MLDI (j) PPID

FIGURE 3.4: Raw image I^{raw} (a) and estimated channel $\hat{I}^9$ (b to j) central extract of IEEE Target estimated reflectance from IMEC16 device.

For assessment purposes, we have also acquired an IEEE reflection Target[1] using the IMEC16 camera at IrDIVE platform. In order to avoid illumination influence, we compute the reflectance of the acquired scene using a white patch (see Section 1.4.1). An area of size 250×70 pixels of the acquired raw image and the mid-spectrum channel $\hat{I}^9$ of the corresponding images estimated by the different demosaicing methods are represented in Fig. 3.4. Note that since the acquired IEEE Target is a reflectance of a black-and-white scene, the raw image faithfully represents the edges of the Target. According to Shannon's theorem [101], such high-frequency lines cannot be reconstructed without aliasing. Hence, bilinear interpolation that only uses spatial information generates spatial artifacts (see Figs. 3.4b, 3.4c, 3.4g and 3.4h). By using spectral correlation, PPDWT, SD, and MLDI methods (Figs. 3.4d, 3.4e, 3.4i and 3.4j) fairly reduce artifacts in the middle area of the considered extract. ItSD and PPID methods also reduce artifacts in the right area.

3.3 Acquisition properties and demosaicing performances

We here study the impact of the illumination and camera properties on demosaicing performances. In Section 3.3.1 we show that demosaicing performances are fairly affected by illumination changes. According to the image formation model (see Section 1.3.1), illumination as well as camera SSFs have an influence on the values of pixels. By studying the demosaicing performances with respect to each channel, we show in Section 3.3.2 that the channels that receive little energy, *i.e.*, that have low values on average with respect to other channels, are difficult to demosaic. Finally, we show in Section 3.3.3 that such variation of values between channels has a great influence on spectral correlation.

3.3.1 PSNR assessment with respect to illumination

In order to study the influence of illumination changes on the demosaicing performances, we compute the PSNR provided by each method on the 32 IMEC16 images simulated with the various illuminations of Fig. 1.2. The results displayed in Table 3.4 show that the performances of all methods are affected by illumination changes. Images simulated under E and D65 are fairly well demosaiced since these illuminants uniformly illuminate the whole spectrum. Using A and HA illuminations reduces the performances because the spectral power distribution (SPD) of these illuminations increases with respect to the wavelength. At last, using F12 and LD illuminations whose SPDs present 3 peaks in the visible domain severely reduces demosaicing performances. Table 3.4 also shows that WB, BTES, and PPBTES methods are fairly robust to illumination variations because they are mainly based on spatial correlation. Other methods use spectral correlation assumption that is weakened by illumination changes.

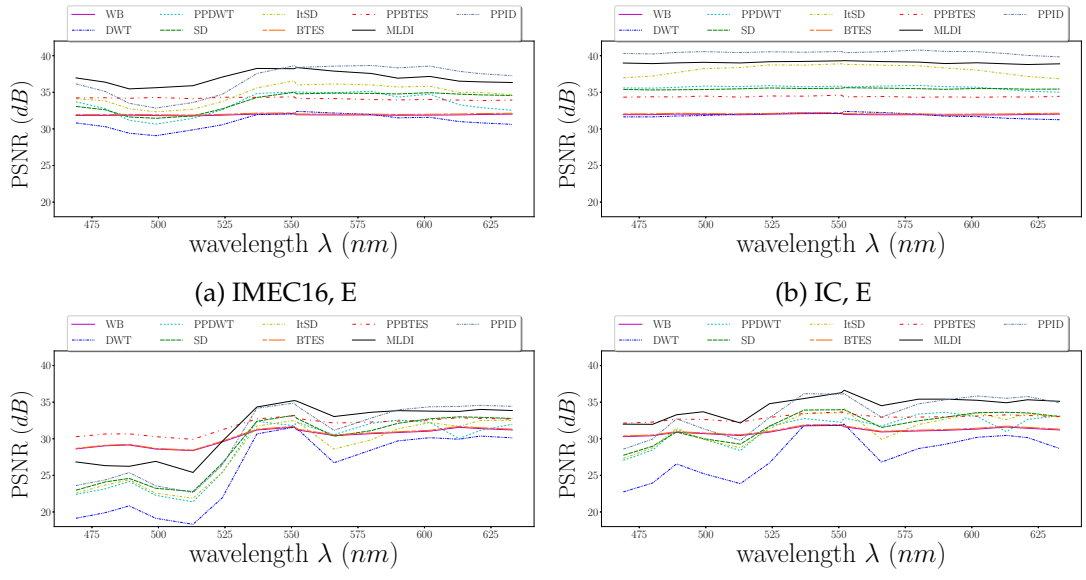
[1] https://www.edmundoptics.com/test-targets/resolution-test-targets/ieee-target/

	E	D65	F12	A	HA	LD
WB	31.91	31.88	30.28	31.69	31.75	31.48
DWT	31.01	31.09	26.15	30.25	29.67	30.41
PPDWT	33.45	33.73	28.48	31.70	31.23	32.45
SD	33.80	34.02	29.23	32.26	32.02	32.68
ItSD	34.75	35.13	28.49	32.51	32.14	32.50
BTES	32.02	31.99	30.38	31.80	31.87	31.59
PPBTES	34.13	34.10	31.84	33.72	33.82	33.42
MLDI	**36.95**	**37.17**	**31.37**	**34.99**	**34.70**	**35.50**
PPID	36.71	37.06	30.32	34.36	34.08	34.71

TABLE 3.4: Average PSNR (dB) over the 32 CAVE images estimated by each demosaicing method according to illumination (average over all channels of the 32 images). The best result for each illumination is displayed as bold.

3.3.2 PSNR with respect to spectral sensitivity function (SSF)

According to the image formation model (see Section 1.3.1), illumination and camera SSFs have an influence on the values of pixels in each channel. We thus propose to study the demosaicing performances with respect to each channel under different illuminations. In order to study the influence of the camera SSFs on demosaicing performances, we also consider an Ideal Camera (IC) whose SSFs are the same as IMEC16 but are normalized so that all of them have the same area of 1 over Ω.



(a) IMEC16, E

(b) IC, E

(c) IMEC16, F12

(d) IC, F12

FIGURE 3.5: Average PSNR (dB) over the 32 CAVE images estimated by each demosaicing method according to band centers. The considered cameras are IMEC16 (a, c) and IC (b, d), and illuminations are E (a, b) and F12 (c, d).

Fig. 3.5 shows the PSNR with respect to each band center for IMEC16 and IC images simulated under E and F12 illuminants. Fig. 3.5b shows that, under a uniform illumination (E) and with SSFs that have the same area, demosaicing performances are similar for all channels. By analyzing Fig. 3.5a, we see that methods based on spectral correlation (DWT, PPDWT, SD, ItSD, MLDI, and PPID) provide poor demosaicing performances in channels whose band centers are around 500 nm. In opposition with IC SSFs that have all the same area ($\sum_{\lambda\in\Omega} T^k(\lambda) = 1$ for all k), IMEC16 SSFs have different areas (they only satisfy $\max_k \sum_{\lambda\in\Omega} T^k(\lambda) = 1$), and SSFs of bands centered at around 500 nm have the smallest areas. According to Eq. (1.3), small SSF areas implies low pixel values. We can therefore deduce that a channel with low pixel values has low demosaicing performances.
In opposition with E illuminant that homogeneously illuminates Ω, F12 illuminant only lights some bands. Channels from bands that receive almost no energy (*e.g.*, bands centered at 469, 480, 524, and 566 nm) have very low values. As shown in Figs. 3.5c and 3.5d, methods based on spectral correlation exhibit low PSNR values at wavelengths where F12 SPD is low with respect to E illuminant (see Figs. 3.5a and 3.5b).
Thus, low pixel values due to spectrally non-uniform illumination or to SSF areas significantly impact demosaicing performances for methods based on spectral correlation.

3.3.3 Effect of illumination and SSFs on spectral correlation

To highlight the effect of spectrally non-uniform illumination or SSFs areas on spectral correlation, we compute the correlation coefficient between the high-frequency information of each channel pair [29, 59]. For this purpose, we apply a circular high-pass filter with a cut-off spatial frequency of 0.25 cycle/pixel on the 2D Fourier transform of each channel. For each illumination and camera, we compute the average Pearson correlation coefficient μ_C (see Eq. (1.8)) over all possible high frequency channel pairs and the standard deviation σ_C of the correlation coefficient.

Table 3.5 shows the correlation and its dispersion on average over all 32 IMEC16 and IC images simulated with each illumination. These results show that the illuminations whose SPD is uniform (E) over Ω or can be considered as such (D65) provide channels with the highest and less scattered spectral correlations. The illuminations A and HA, for which $E(\lambda)$ increases with respect to λ over Ω, provide channels with lower and more scattered spectral correlations. The illuminations F12 and LD, for which $E(\lambda) \approx 0$ except for three marked peaks, provide channels with the lowest and most scattered spectral correlations. By comparing Table 3.5a with Table 3.5b we see that SSF areas over Ω strongly affect spectral correlation, and that channels are more correlated when SSFs have similar areas.
To conclude, the illumination SPD and camera SSF areas strongly affect values of pixels in different channels. Such variation of values from a channel to another weakens

	E	D65	F12	A	HA	LD
μ_C	0.894	0.884	0.514	0.785	0.814	0.724
σ_C	0.040	0.043	0.166	0.086	0.081	0.092

(A) IMEC16

	E	D65	F12	A	HA	LD
μ_C	0.940	0.934	0.645	0.889	0.901	0.821
σ_C	0.028	0.028	0.113	0.040	0.040	0.073

(B) IC

TABLE 3.5: Correlation average and standard deviation with respect to illumination (averages over all 32 IMEC16 (a) and IC (b) images).

spectral correlation, which affects the performance of demosaicing procedures that rely on this property. In the next section we propose three ways to overcome this issue.

3.4 Robust demosaicing for various acquisition properties

We propose pre- and post-normalization steps for demosaicing in Section 3.4.1 that adjust the values of channels before demosaicing and restore them afterwards. These steps make demosaicing robust to acquisition properties by using the normalization factors presented in Section 3.4.2. Such normalization factors depend on acquisition properties or on raw image statistics. In Section 3.4.3, we finally assess the demosaicing methods presented in Sections 2.4 and 2.6 when the proposed normalization steps are performed.

3.4.1 Raw value scale adjustment

We first proposed pre- and post-normalization steps to adjust raw values for demosaicing in **Mihoubi** et al. [73]. These procedures, illustrated in Fig. 3.6, improve the estimation of the PPI under various illuminations and are extended to any demosaicing method in [74].
Before demosaicing, the value scale of each channel is adjusted by computing a new raw value $I_p'^{raw}$ at each pixel p. For this purpose, the pre-normalization step normalizes the raw image at each pixel subset $\{S^k\}_{k=1}^K$ by a specific factor ρ_*^k:

$$I_p'^{raw} = \rho_*^k \cdot I_p^{raw} \text{ for all } p \in S^k, \tag{3.3}$$

where $*$ refers to a normalization approach among the three presented below. Demosaicing is then performed on the scale-adjusted raw image I'^{raw} to provide the

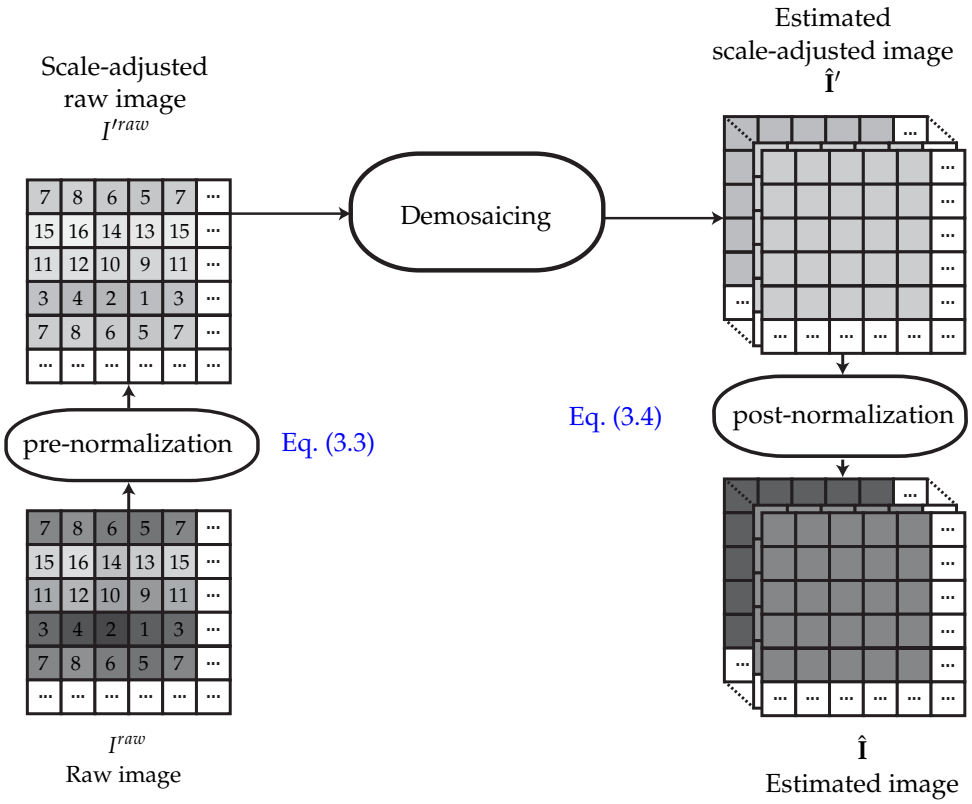


FIGURE 3.6: Normalization steps for demosaicing.

estimated image $\hat{\mathbf{I}}'$. After demosaicing, the post-normalization step restores the original value scale of all pixels of each estimated channel $\hat{I}'^k$:

$$\hat{I}_p^k = \frac{1}{\rho_*^k} \cdot \hat{I}_p'^k \text{ for all } p \in S. \tag{3.4}$$

In the following we propose three ways to compute the normalization factor ρ_*^k of each channel.

3.4.2 Normalization factors

Eq. (1.3) shows that image formation results from the product between the reflectance $R_p(\lambda)$, the illumination $E(\lambda)$, and the SSFs $T^k(\lambda)$ associated to the spectral bands. Depending on the information that are available about the camera SSFs and the illumination, three normalization approaches may then be applied:

- **Camera-based normalization:** When prior knowledge about the camera sensitivity is available, Lapray et al. [54] balance all SSFs $\{T^k(\lambda)\}_{k=1}^K$ so that the area of each of them over Ω is equal to 1. We then propose the following normalization factor ρ_{cam}^k based on camera properties:

$$\rho_{cam}^k = \frac{\max_{l=1}^K \sum_{\lambda\in\Omega} T^l(\lambda)}{\sum_{\lambda\in\Omega} T^k(\lambda)}. \tag{3.5}$$

Such normalization enhances the values of channels that receive low energy due to camera SSFs.

- **Camera- and illumination-based normalization:** When both the SSFs of the camera and the illumination $E(\lambda)$ of the scene are known, Lapray et al. [54] apply a scheme similar to a white balance on each channel. For this purpose the maximal energy that would be obtained from a perfect diffuser ($R_p(\lambda) = 1$ for all $\lambda \in \Omega$ at each pixel p) is divided by the energy of each channel. We then propose the following normalization factor ρ^k_{ci} based on camera and illumination properties:

$$\rho^k_{ci} = \frac{\max_{l=1}^{K} \sum_{\lambda\in\Omega} T^l(\lambda)E(\lambda)}{\sum_{\lambda\in\Omega} T^k(\lambda)E(\lambda)} \,. \tag{3.6}$$

  Such normalization enhances values of channels that receive low energy due to both camera SSFs and illumination.

- **Raw image-based normalization:** In contrast with the two previous approaches, raw image-based normalization does not use any prior knowledge about camera or illumination. We instead propose to balance the value ranges of all channels by only using the raw image values [73]. For this purpose, we consider the ratio between the maximum value over all channels and the maximum value that is available for each channel in the raw image I^{raw}. The normalization factor ρ^k_{raw} is then given by:

$$\rho^k_{raw} = \frac{\max_{p\in S} I^{raw}_p}{\max_{p\in S^k} I^{raw}_p} \,. \tag{3.7}$$

  Note that this is similar to the max-spectral approach proposed by Khan et al. [45] for illumination estimation.

3.4.3 Normalization assessment

To study the benefit of normalization on the demosaicing performances when illumination changes, each method is assessed without and with the normalization approaches proposed in Section 3.4.2. Table 3.6 shows the average PSNR over the 32 CAVE images simulated using IMEC16 SSFs. Results of Table 3.4 are recalled in Table 3.6 to provide an easy comparison.
Normalization has no effect on images estimated by WB and BTES since these methods only use spatial correlation. Using camera-based normalization (ρ^k_{cam}) fairly improves the performances with illuminants E and D65 whose SPD is uniform. However, performances can be reduced in the case of LD illumination whose SPD mainly lies in three dominant narrow bands. Using camera- and illumination-based normalization (ρ^k_{ci}) provides the best performances for most of illuminations and methods. However, the illumination is unknown and has to be estimated when the camera is used in uncontrolled conditions. The same performances are practically reached by raw image-based normalization (ρ^k_{raw}) that does not require any prior knowledge

about the camera or illumination. This simple approach which uses statistics of the raw image therefore gives satisfactory results whatever the demosaicing method and scene illumination conditions.

The best improvement provided by normalization is reached using the PPID demosaicing method under HA illumination. For illustration purposes, we select an extract of size 125×125 pixels from the "Chart and stuffed toy" CAVE image simulated using IMEC16 SSFs under the HA illumination. Reference and estimated images (using PPID) are converted to the sRGB color space. The results displayed in Fig. 3.7 show that the estimated image without normalization presents severe zipper artifacts and false colors. Applying camera-based normalization reduces those artifacts and the other two normalization approaches slightly further improve the visual results.

(a) Reference (b) None (c) ρ^k_{cam} (d) ρ^k_{ci} (e) ρ^k_{raw}

FIGURE 3.7: sRGB renderings of a central extract from "Chart and stuffed toy" CAVE image simulated using IMEC16 SSFs under HA illumination (a). Images (b to e) are estimated by PPID demosaicing method with different normalization approaches.

3.5 Demosaicing HyTexiLa images with various cameras

We propose to study the demosaicing performances on multispectral images acquired by different cameras in the Vis, the NIR, or the VisNIR domain. For this purpose we use HyTexiLa database that contain 112 reflectance images in the VisNIR domain (see Section 1.4.2). We consider four cameras and two demosaicing methods that are presented in Section 3.5.1. The demosaicing methods are extended to the four MSFAs in Section 3.5.2 and assessed in Section 3.5.3.

3.5.1 Considered cameras and demosaicing methods

Among MSFA-based cameras, we select VIS5 and IMEC16 that sample the Vis domain, IMEC25 that samples the NIR domain, and VISNIR8 that samples the VisNIR domain. The SSFs associated to each of the four cameras are available in Appendix B. We use Eq. (1.3) to simulate the HyTexiLa images that would be acquired using these cameras under the extended D65 illuminant. For each camera, the resulting 8-bit radiance images are sub-sampled according to the MSFA associated to each camera (see Section 2.2.3). In order to demosaic the raw images, we select only WB and

Method	Norm.	E	D65	F12	A	HA	LD
WB	Any	31.91	31.88	30.28	31.69	31.75	31.48
DWT	None	31.01	31.09	26.15	30.25	29.67	30.41
	ρ^k_{cam}	31.78	31.75	27.72	31.43	31.23	30.37
	ρ^k_{ci}	31.78	31.75	30.00	31.54	31.56	31.31
	ρ^k_{raw}	31.76	31.74	29.99	31.52	31.55	31.30
PPDWT	None	33.45	33.73	28.48	31.70	31.23	32.45
	ρ^k_{cam}	35.48	35.42	30.69	34.50	34.15	32.18
	ρ^k_{ci}	35.48	35.42	32.31	34.95	35.06	34.39
	ρ^k_{raw}	35.42	35.36	32.25	34.89	35.01	34.34
SD	None	33.80	34.02	29.23	32.26	32.02	32.68
	ρ^k_{cam}	35.30	35.23	31.19	34.51	34.36	32.70
	ρ^k_{ci}	35.30	35.24	32.29	34.80	34.94	34.30
	ρ^k_{raw}	35.25	35.19	32.20	34.75	34.89	34.26
ItSD	None	34.75	35.13	28.49	32.51	32.14	32.50
	ρ^k_{cam}	37.75	37.65	30.85	36.26	35.83	32.53
	ρ^k_{ci}	37.75	37.66	33.26	36.89	37.08	36.03
	ρ^k_{raw}	37.59	37.49	33.13	36.75	36.94	35.91
BTES	Any	32.02	31.99	30.38	31.80	31.87	31.59
PPBTES	None	34.13	34.10	31.84	33.72	33.82	33.42
	ρ^k_{cam}	34.29	34.23	31.99	33.94	34.04	33.43
	ρ^k_{ci}	34.29	34.23	32.12	33.99	34.12	33.58
	ρ^k_{raw}	34.29	34.23	32.12	33.99	34.12	33.58
MLDI	None	36.95	37.17	31.37	34.99	34.70	35.50
	ρ^k_{cam}	38.71	38.60	33.31	37.54	37.50	35.24
	ρ^k_{ci}	38.71	38.59	**34.32**	37.86	38.14	37.02
	ρ^k_{raw}	38.68	38.56	34.28	37.84	38.12	36.99
PPID	None	36.71	37.06	30.32	34.36	34.08	34.71
	ρ^k_{cam}	**39.84**	39.65	32.50	37.89	37.60	34.73
	ρ^k_{ci}	**39.84**	**39.69**	34.18	**38.49**	**38.81**	**37.46**
	ρ^k_{raw}	39.74	39.59	34.10	38.40	38.73	37.40

TABLE 3.6: Average PSNR (dB) over the 32 CAVE images (simulated using IMEC16 SSFs) estimated by each demosaicing method according to illumination. Normalizations: camera-based ρ^k_{cam}, camera- and illumination-based ρ^k_{ci}, raw image-based ρ^k_{raw} (see Section 3.4.2). The best result for each illumination is displayed as bold.

PPID demosaicing methods. Indeed, BTES, PPBTES, and MLDI are not applicable with the IMEC25 MSFA in which each band has a prior probability (PP) of $\frac{1}{25}$ that is not the inverse of a power of two. WB is the simplest and the most generic method that is applicable with all MSFAs, and PPID always provides better results than DWT, PPDWT, SD and ItSD.

3.5.2 Extension of WB and PPID methods to the four MSFAs

WB is extended to the different MSFAs by adapting the bilinear filter H according to its definition. The used bilinear filter depend on the sampling of pixels in each subset S^k. The filter of Fig. 3.8a is applied when PP is $\frac{1}{2}$ (G band in VIS5), that of Fig. 3.8b is applied when PP is $\frac{1}{8}$ (R, O, B, C bands in VIS5, and VISNIR8 channels), that of Fig. 3.8c is applied when PP is $\frac{1}{16}$ (IMEC16 bands), and that of Fig. 3.8d is applied when PP is $\frac{1}{25}$ (IMEC25 bands).

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

(a) $PP=\frac{1}{2}$

$$\begin{bmatrix} 1 & 2 & 3 & 2 & 1 \\ 2 & 4 & 6 & 4 & 2 \\ 3 & 6 & 9 & 6 & 3 \\ 2 & 4 & 6 & 4 & 2 \\ 1 & 2 & 3 & 2 & 1 \end{bmatrix}$$

(b) $PP=\frac{1}{8}$

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 3 & 2 & 1 \\ 2 & 4 & 6 & 8 & 6 & 4 & 2 \\ 3 & 6 & 9 & 12 & 9 & 6 & 3 \\ 4 & 8 & 12 & 16 & 12 & 8 & 4 \\ 3 & 6 & 9 & 12 & 9 & 6 & 3 \\ 2 & 4 & 6 & 8 & 6 & 4 & 2 \\ 1 & 2 & 3 & 4 & 3 & 2 & 1 \end{bmatrix}$$

(c) $PP=\frac{1}{16}$

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 4 & 3 & 2 & 1 \\ 2 & 4 & 6 & 8 & 10 & 8 & 6 & 4 & 2 \\ 3 & 6 & 9 & 12 & 15 & 12 & 9 & 6 & 3 \\ 4 & 8 & 12 & 16 & 20 & 16 & 12 & 8 & 4 \\ 5 & 10 & 15 & 20 & 25 & 20 & 15 & 10 & 5 \\ 4 & 8 & 12 & 16 & 20 & 16 & 12 & 8 & 4 \\ 3 & 6 & 9 & 12 & 15 & 12 & 9 & 6 & 3 \\ 2 & 4 & 6 & 8 & 10 & 8 & 6 & 4 & 2 \\ 1 & 2 & 3 & 4 & 5 & 4 & 3 & 2 & 1 \end{bmatrix}$$

(d) $PP=\frac{1}{25}$

FIGURE 3.8: Unnormalized bilinear filter F with respect to the probability of appearance.

PPID is extended to the different MSFAs by adapting the bilinear filter according to Fig. 3.8 and the average filter M according to its definition as shown in Fig. 3.9.

The weights used in Eq. (2.22) still consider the eight neighbors that sample the same channel as the central pixel (see Fig. 2.14). These neighbors are located at a spatial distance that varies from 2 to 5 pixels with respect to the considered MSFA. In the particular case of VIS5 MSFA, we only consider the PPI estimated using the averaging filters of Figs. 3.9a and 3.9b (see Eq. (2.22)) since the weights cannot be computed for the green channel. Note that PPID is applied using the raw image-based normalization that does not require any acquisition information (see Section 3.4.1).

$$\frac{1}{210}\begin{bmatrix}14 & 6 & 14 & 6 & 14\\ 6 & 42 & 6 & 42 & 6\\ 14 & 6 & 14 & 6 & 14\end{bmatrix}$$

(a) VIS5 at pixels in S^G

$$\frac{1}{40}\begin{bmatrix}1 & 4 & 1 & 4 & 1\\ 4 & 1 & 8 & 1 & 4\\ 1 & 4 & 1 & 4 & 1\end{bmatrix}$$

(b) VIS5 at pixels in S^k, $k \in \{R, O, B, C\}$

$$\frac{1}{48}\begin{bmatrix}2 & 3 & 2 & 3 & 2\\ 3 & 6 & 6 & 6 & 3\\ 2 & 3 & 2 & 3 & 2\end{bmatrix}$$

(c) VISNIR8

$$\frac{1}{64}\begin{bmatrix}1 & 2 & 2 & 2 & 1\\ 2 & 4 & 4 & 4 & 2\\ 2 & 4 & 4 & 4 & 2\\ 2 & 4 & 4 & 4 & 2\\ 1 & 2 & 2 & 2 & 1\end{bmatrix}$$

(d) IMEC16

$$\frac{1}{25}\begin{bmatrix}1 & 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1 & 1\end{bmatrix}$$

(e) IMEC25

FIGURE 3.9: Filter M used for first PPI estimation (see Eq. (2.18)) with respect to the MSFA.

3.5.3 PSNR comparison

Table 3.7 shows the average PSNR over the 112 images simulated using each of the four cameras with respect to WB and PPID methods.

	IMEC16	IMEC25	VIS5	VISNIR8
WB	31.82	32.78	36.93	**31.59**
PPID	**36.48**	**38.86**	**38.16**	31.17

TABLE 3.7: Average PSNR (dB) reached by WB and PPID demosaicing methods over the 112 HyTexiLa images simulated using each of the four cameras under extended D65 illuminant. The best result for each camera is displayed as bold.

These results show that for IMEC16, IMEC25, and VIS5 images, using the correlation between each channel and the PPI in PPID provides better demosaicing performances than using only spatial correlation in WB. As seen in Section 1.6.3 indeed, channels that sample the Vis domain are strongly correlated as are channels that sample the NIR domain. However, channels that sample the Vis domain are not correlated with channels that sample the NIR domain. Thus applying PPID on VISNIR8 images provides poor demosaicing performances. In order to use PPID in case of low spectral correlation, a solution is to avoid Eq. (2.22) in PPI estimation since it assumes a high spectral correlation [73].

3.6 Conclusion

Our extensive experiments show that PPI-based demosaicing methods provide nice performances with respect to the existing demosaicing schemes that are suited to our considered MSFA (IMEC16) both in terms of PSNR and in a visual assessment. Indeed, the proposed method based on PPI difference (PPID) provides high-quality

estimated images with sharp edges and reduced color and zipper artifacts at a moderate computational cost.

By studying the impact of the illumination and camera properties on demosaicing performances, we notice that demosaicing performances decrease when channel values highly differ on average among channels. This is due to a reduction of spectral correlation when illumination is non-homogeneous over the spectrum or when the camera SSFs differ in areas. This severely affects the performance of demosaicing schemes that mainly rely on assumptions about spectral correlation. We then propose a normalization scheme that adjusts channel values before demosaicing, which improves demosaicing robustness to acquisition properties. The associated normalization factors either depend on the camera spectral sensitivity only, on both the sensitivity and the illumination, or on the statistics extracted from the acquired raw image. Experimental results show that normalization based on the sole SSFs provides good but illumination-sensitive results. Normalization based on SSFs and a known of illumination provides the best results but illumination information is not always available in practice. At last raw image-based normalization provides promising results without any a priori knowledge about the camera or illumination, and thus constitutes a good compromise for demosaicing.

This raw image-based normalization is then applied to PPID demosaicing method in order to compare the performances of demosaicing on four different MSFA-based cameras proposed on the marked or in the literature, namely IMEC16, IMEC25, VIS5, and VISNIR8. In comparison with the simple WB demosaicing method, PPID provides good demosaicing performances on cameras whose bands belong to either the Vis or the NIR domain. However, when the bands belong to Vis and to the NIR domains, performance are fairly reduced.

The next chapter focuses on the classification of images acquired by MSFA-based cameras.

Chapter 4

MSFA raw image classification

Contents

4.1 Introduction

A texture image is a characterization of the spatial and spectral properties of the physical structure of a material or an object. Texture analysis classically relies on a set of features that provide information about the spatial arrangement of the spectral responses of an object. In a preliminary work, Khan et al. [46] have shown that taking spectral information, by the way of multispectral images, improves classification accuracy for textures compared to color or gray-scale images. To classify multispectral texture images acquired by single-sensor snapshot cameras, the classical supervised approach is to demosaic the raw images, extract texture features from the estimated images, then compare features with those computed from different known images thanks to a similarity measure. Our classification scheme applied on HyTexiLa database is presented in Section 4.2. Such classification scheme requires a texture descriptor in order to extract texture features. Among texture descriptors, we select the histogram of local binary patterns (LBPs) that is one of the most robust descriptors in the literature. In this chapter the feature extracted from the descriptor is the descriptor itself, *i.e.* the histogram of LBPs, so that the two terms are equivalent. The existing color LBP-based descriptors are extended to any K-channel image in Section 4.3. In addition to spatial information, these descriptors also consider the spectral information available among the K channels of a multispectral image. However, the computational cost significantly increases with respect to the number of channels due to demosaicing and feature extraction. Thus, in Section 4.4 we propose a new computationally-efficient LBP-based descriptor that is directly computed from raw images, which allows us to avoid the demosaicing step [75]. Extensive experiments on HyTexiLa database prove the relevance of our approach in Section 4.5.

4.2 Classification scheme

In order to perform texture classification on MSFA raw images, we consider the four MSFAs that are either related to research works with detailed publications (VIS5 and VISNIR8) or available in consumer cameras (IMEC16 and IMEC25) (see Section 2.2.3). The classification scheme is presented in Section 4.2.1. It uses the LBP descriptor whose marginal approach is presented in Section 4.2.2 and that is then combined with the similarity measure and the decision algorithm presented in Section 4.2.3.

4.2.1 Classification of MSFA raw texture images

The goal of texture classification is to assign a sample texture image to one among several known texture classes. For this purpose discriminant texture features are

extracted from test images and compared to those extracted from training images whose classes are known, as represented in Fig. 4.1.

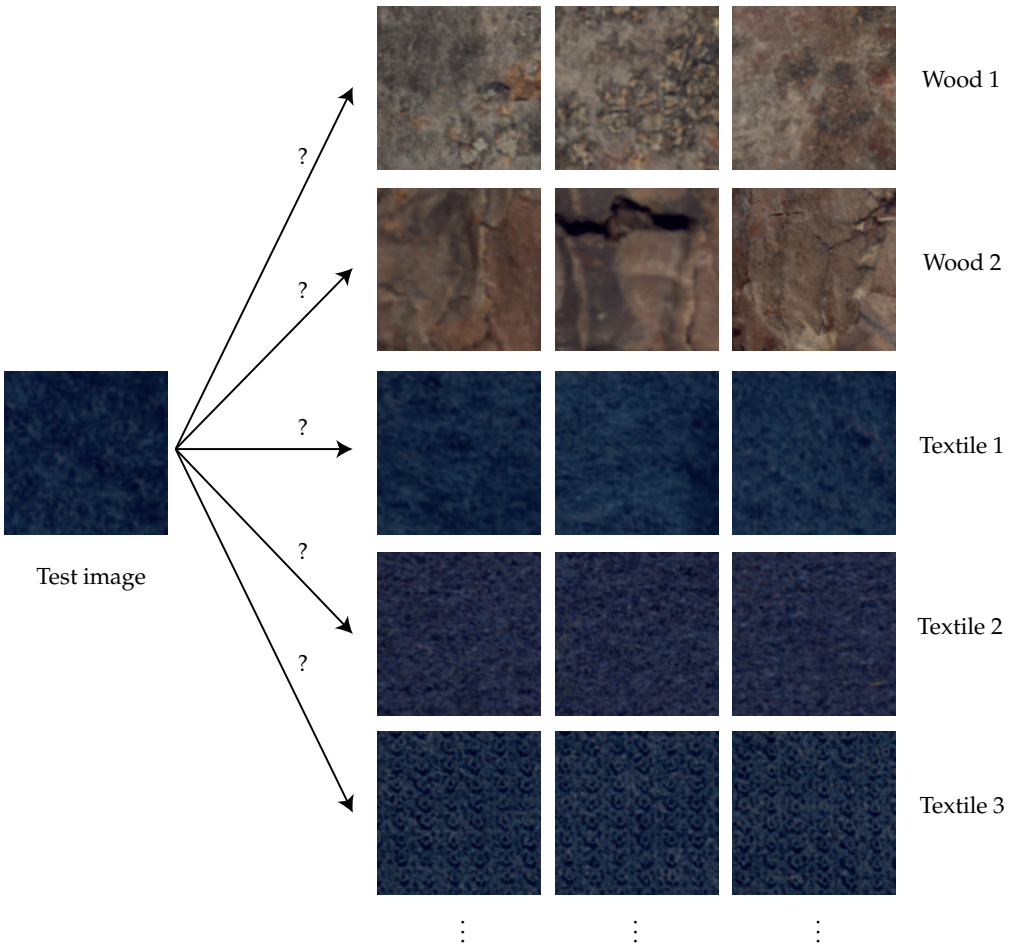
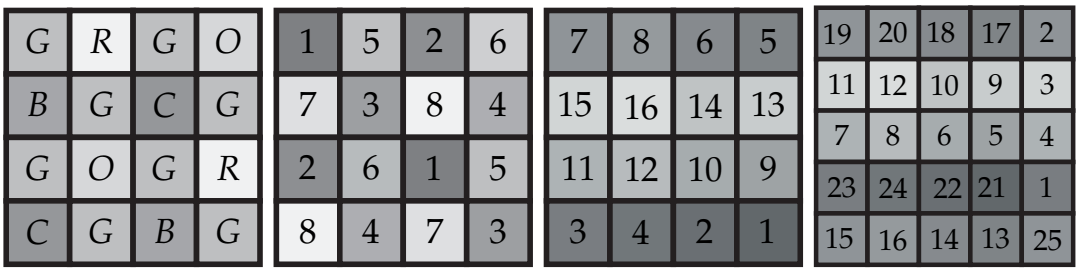


FIGURE 4.1: Texture classification scheme. For illustration purpose, images from HyTexiLa database are rendered in sRGB space.

In order to perform and assess texture classification a database of different textures is needed. As seen in Section 1.4.2, our proposed HyTexiLa database [46] is currently the only suitable database of multispectral texture images for texture classification. Texture feature assessment can be performed on HyTexiLa database by considering each of the 112 texture images as a class.

Texture images are simulated in order to provide raw images that would be acquired using a single-sensor camera. For this purpose we consider three illuminations (extended E, extended A, and D65 simulator) and four MSFA-based cameras (VIS5, VISNIR8, IMEC16, and IMEC25). We first simulate fully-defined radiance images from the reflectance of textures, illuminations, and camera SSFs using Eq. (1.3). Then, we sample these radiance images according to an MSFA among those of Fig. 4.2 to simulate the raw images that would be acquired by the associated snapshot multispectral camera. For a considered camera and illumination, these simulations provide 112 8-bit raw images of size 1024×1024 pixels. Finally,

we estimate radiance images $\hat{\mathbf{I}}$ of size 1024 × 1024 pixels × K channels from raw images by demosaicing. Then, we split each of them into 25 images of size 204 × 204 pixels among which 12 are randomly picked for training and the 13 others for testing. Note that K depends on the considered camera and that the last four columns and rows are left out when splitting $\hat{\mathbf{I}}$.



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

FIGURE 4.2: Basic patterns of four square periodic MSFAs: VIS5 (a) [78], VISNIR8 (b) [115], IMEC16 (c) and IMEC25 (d) [27]. Numbers are band indexes (see Appendix B) and labels in (a) are those of [78] but could be replaced by indexes.

In the learning phase, LBP histogram features are extracted from each training estimated image. Then, to assign a test image to one of the classes, the same features are extracted from it and compared to those of each training image. This comparison is performed using a similarity/dissimilarity measure between test and train features. Finally, each test image is assigned to the class of the training image with the best match by using a decision algorithm. The performances of a classification algorithm are determined by the rate of well-classified test images, and depend on three main parts of classification, namely the choice of discriminative textural features, the feature similarity measure, and the decision algorithm. The chosen three parts of classification are presented in next subsections.

4.2.2 Local binary patterns (LBPs)

To extract a texture feature from an image, LBP is a prominent operator. By characterizing the local level variation in a neighborhood of each pixel, this operator is robust to grayscale variations. Due to its discrimination power and computational efficiency, LBP has also proved to be a very efficient approach in a wide variety of applications, among which texture classification and face recognition [91, 116].
LBP-based texture classification has first been performed on gray-level images since the original operator only uses the spatial information of texture [86]. This LBP operator can be applied marginally on a multispectral radiance image $\mathbf{I} = \{I^k\}_{k=1}^K$ by considering the K channels separately. In this section and in the next section, we formulate several LBP-based texture features for any fully-defined K-channel image, that we generically denote as $\mathbf{I}$ for simplicity whatever the value of K and even it has been estimated by demosaicing. For a given pixel p of a channel I^k, the LBP operator considers the neighborhood N_p defined by its support $\mathcal{N}^{P,d}$ made of P pixels

at spatial distance d from p:

$$LBP^k[\mathbf{I}](p) = \sum_{q \in N_p} s\left(I_q^k, I_p^k\right) \cdot 2^{\epsilon(q)}, \tag{4.1}$$

where I_p^k is the value of channel I^k at p, $\epsilon(q) \in \{0, ..., P-1\}$ is the index of each neighboring pixel q in N_p, and $s(\cdot)$ is the unit step function:

$$s(\alpha, \beta) = \begin{cases} 1 & \text{if } \alpha \geq \beta, \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

An example of LBP computation at a pixel p of a channel I^k is shown in Fig. 4.3.

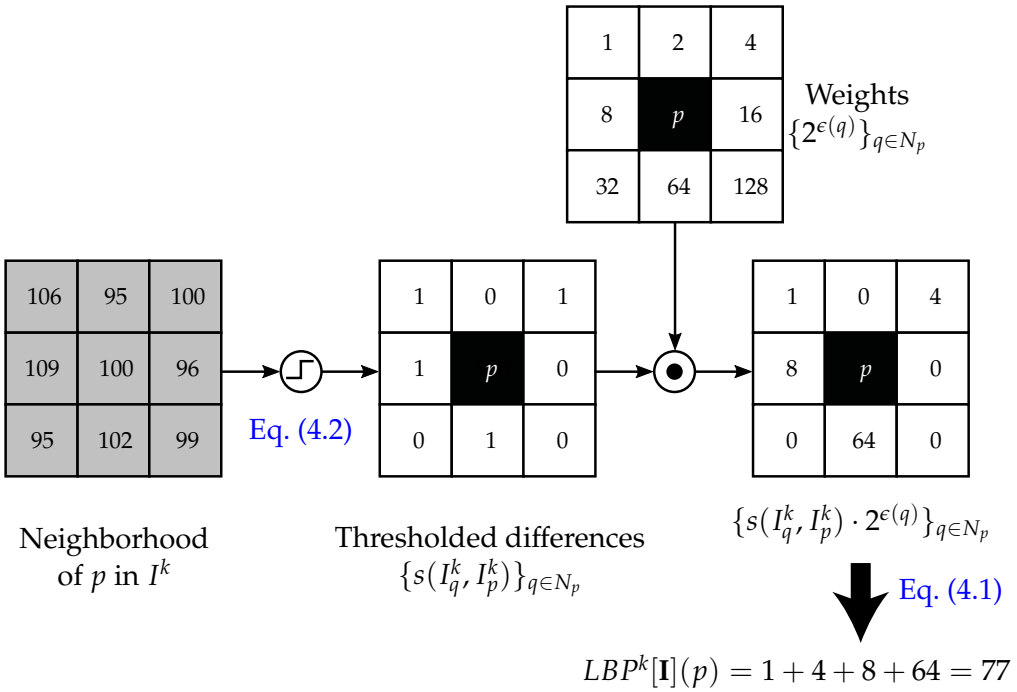


FIGURE 4.3: Marginal LBP operator applied to a pixel p of channel I^k.

Each channel I^k, $k \in \{1, ..., K\}$, is characterized by the 2^P-bin un-normalized histogram of its LBP values. The multispectral texture image $\mathbf{I}$ is then described by the concatenation of the K histograms of $\{LBP^k[\mathbf{I}]\}_{k=1}^K$. This feature, whose size is $K \cdot 2^P$, represents the spatial interaction between neighboring pixels within each channel independently. The next section reviews some extensions of the original LBP operator from gray-scale to color images ($K = 3$) and generalizes them to the multispectral domain by considering $K \geq 4$ channels.

Note that in this chapter we only consider the few variants of the basic LBP operator that can straightforwardly be applied to a multispectral image even though many LBP variants have been described in the literature [91]. Also note that the definition of Eq. (4.1) ignores border effects for readability sake and that only those pixels at

which N_p is fully enclosed in the image are actually taken into account to compute the LBP histogram.

4.2.3 Decision algorithm and similarity measure

In order to determine the most discriminant LBP-based texture feature, we propose to retain the similarity measure based on intersection between histograms [113] coupled with the 1-Nearest Neighbor decision algorithm since this classification scheme requires no additional parameter.
The similarity measure between two images $\mathbf{I}$ and $\mathbf{I'}$ is defined by the normalized intersection between their concatenated LBP histograms h and h' as

$$Sim[\mathbf{I},\mathbf{I'}] = \frac{\sum_{i=0}^{|h|} min[h(i),h'(i)]}{\sum_{i=0}^{|h|} h(i)}, \tag{4.3}$$

where $|h|$ is the size (number of bins) of LBP histograms, and $\sum_{i=0}^{|h|} h(i) = \sum_{i=0}^{|h|} h'(i)$ represents the number of pixels from which the histogram is computed (possibly not all pixels). $Sim[\mathbf{I},\mathbf{I'}]$ ranges from 0 to 1 and equals 1 when the two images are identical.
In order to highlights the intrinsic properties of the descriptor we choose the 1-nearest neighbor that simply considers the class associated to the training image with the highest similarity to the tested image. This non-parametric decision algorithm outputs the closest training samples in the feature space according to the similarity measure.

4.3 LBP-based Spectral texture features

Palm [88] has shown that classification based on a color analysis outperforms that based on the spatial information only. Texture feature extraction has been extended to the color domain by taking both spatial and spectral textural information into account. Below we formulate several color LBP-based texture features for any fully-defined K-channel image.

4.3.1 Moment LBPs

Mirhashemi [76] proposes an LBP-based spectral feature using mathematical moments to characterize the reflectance spectrum shape. The LBP operator of Eq. (4.1) is no longer applied to pixel values but to moment values of the pixel spectral signatures.

Different moments can be extracted from the reflectance $\{R_p(\lambda^k)\}_{k=1}^{K}$ sampled over K bands at each pixel p. Raw and central type-I moments of order $n \in \mathbb{N}$ are

defined as

$$\begin{aligned} M_n(p) &= \sum_{k=1}^{K} \left(\lambda^k\right)^n R_p(\lambda^k) \\ \text{and} \quad \mu_n(p) &= \sum_{k=1}^{K} \left(\lambda^k - \frac{M_1(p)}{M_0(p)}\right)^n R_p(\lambda^k). \end{aligned} \tag{4.4}$$

Type-II moments are estimated moments of the probability density function from which reflectance values are sampled. Raw and central type-II moments are expressed as

$$\begin{aligned} \hat{M}_n(p) &= \frac{1}{K} \sum_{k=1}^{K} \left(R_p(\lambda^k)\right)^n \\ \text{and} \quad \hat{\mu}_n(p) &= \frac{1}{K} \sum_{k=1}^{K} \left(R_p(\lambda^k) - \frac{\hat{M}_1(p)}{\hat{M}_0(p)}\right)^n. \end{aligned} \tag{4.5}$$

Alternatively, these moments can be computed from the reflectance normalized by its L_1-norm at each pixel p: $r_p(\lambda^k) = \frac{R_p(\lambda^k)}{\sum_{i=1}^{K} R_p(\lambda^i)}$. We then denote type-I and type-II raw moments as $m_n(p)$ and $\hat{m}_n(p)$.

Mirhashemi [76] assesses the texture classification performance of all the possible moment-based features (namely the 38 moment LBP histograms obtained for $n = \{1, \ldots, 6\}$), either considered alone or concatenated in 2- or 3-feature combinations. The most powerful combinations use three features based on the following moments: $m_1(p)$ or $\hat{m}_1(p)$, $M_1(p)$ or $\hat{M}_1(p)$, and $\mu_3(p)$, $\mu_5(p)$, $\hat{\mu}_3(p)$, or $\hat{\mu}_5(p)$. The texture feature is then a concatenated histogram with $3 \cdot 2^P$ bins.

4.3.2 Map-based LBPs

Dubey et al. [18] propose two kinds of LBP operators that can theoretically be applied to any K-channel image. These operators use the spectral information in the encoding scheme by testing the sum of the marginal comparison patterns between each pixel p and its neighbors over all channels.

- The adder-based LBPs $\{maLBP^m\}_{m=0}^{K}$ are defined as

$$maLBP^m[\mathbf{I}](p) = \sum_{q \in N_p} \begin{cases} 2^{\epsilon(q)} & \text{if } \sum_{k=1}^{K} s(I_q^k, I_p^k) = m, \\ 0 & \text{otherwise.} \end{cases} \tag{4.6}$$

- The decoder-based LBPs $\{mdLBP^n\}_{n=0}^{2^K-1}$ are defined as

$$mdLBP^n[\mathbf{I}](p) = \sum_{q \in N_p} \begin{cases} 2^{\epsilon(q)} & \text{if } \sum_{k=1}^{K} s(I_q^k, I_p^k) \cdot 2^{(K-k)} = n, \\ 0 & \text{otherwise.} \end{cases} \tag{4.7}$$

The concatenation of the histograms of maLBP or mdLBP operator outputs provides the final feature of size $(K+1) \cdot 2^P$ or $2^K \cdot 2^P$, respectively.

4.3.3 Luminance–spectral LBPs

By analogy with the luminance–chrominance model for a color image, a multispectral image can be represented as both a panchromatic channel and the joint information computed from two or more channels. The PPI that carries the spatial information of the luminance is computed as the average value over all channels at each pixel p. We recall it here from Eq. (2.16):

$$I_p^{PPI} = \frac{1}{K} \sum_{k=1}^{K} I_p^k.$$

To form the final feature, the histogram of the output of the LBP operator applied to I^{PPI} is concatenated with a histogram based on the spectral content according to one of the following propositions that we extend here to the multispectral domain:

- Cusano et al. [15] define the local color contrast (LCC) operator that depends on the angle between the value of a pixel p and the average value $\bar{\mathbf{I}}_p = \frac{1}{P} \sum_{q \in N_p} \mathbf{I}_q$ of its neighbors in the spectral domain:

$$LCC[\mathbf{I}](p) = \arccos\left(\frac{\langle \mathbf{I}_p, \bar{\mathbf{I}}_p \rangle}{||\mathbf{I}_p|| \cdot ||\bar{\mathbf{I}}_p||} \right), \tag{4.8}$$

  where $\langle \cdot, \cdot \rangle$ and $|| \cdot ||$ denote the inner product and the Euclidean norm. The histogram of $LBP[I^{PPI}]$ is concatenated to that of $LCC[\mathbf{I}]$ quantized on 2^P bins to provide the final feature of size $2 \cdot 2^P$.

- Lee et al. [56] consider $\mathbf{I}$ in a K-dimensional space and compute spectral angular patterns between bands at each pixel. Specifically, for each pair of bands $(k, l) \in \{1, ..., K\}^2, k \neq l$, the authors apply the LBP operator to the image $\theta^{k,l}$ defined at each pixel p as the angle between the axis of the band k and the projection of $\mathbf{I}_p$ onto the plane associated with bands k and l:

$$\theta_p^{k,l} = \arctan\left(\frac{I_p^k}{I_p^l + \eta} \right), \tag{4.9}$$

  where η is a small-valued constant to avoid division by zero. The histogram of $LBP[I^{PPI}]$ is concatenated to the $K(K-1)$ histograms of $\{LBP[\theta^{k,l}]\}_{k,l=1,k\neq l}^{K}$ to provide the final feature of size $(1 + K(K-1)) \cdot 2^P$.

4.3.4 Opponent band LBPs

To fully take spectral correlation into account, Mäenpää et al. [66] apply the opponent color LBP operator to each pair of channels of a color image. This operator can be directly generalized as the opponent band LBP (OBLBP) applied to each pair of

channels (I^k, I^l), $(k,l) \in \{1,...,K\}^2$, of a multispectral image:

$$OBLBP^{k,l}[\mathbf{I}](p) = \sum_{q \in N_p} s\left(I_q^l, I_p^k\right) \cdot 2^{\epsilon(q)}. \tag{4.10}$$

Bianconi et al. [6] similarly considers both intra- and inter-channel information but with a different thresholding scheme. Their improved OBLBP (IOBLBP) operator uses a local average value rather than the sole central pixel value as threshold:

$$IOBLBP^{k,l}[\mathbf{I}](p) = \sum_{q \in \{p\} \cup N_p} s\left(I_q^l, \bar{I}_p^k\right) \cdot 2^{\epsilon(q)}, \tag{4.11}$$

where $\bar{I}_p^k = \frac{1}{P+1} \sum_{r \in \{p\} \cup N_p} I_r^k$ and $\epsilon(p) = P$.

In both cases, the texture feature is the concatenation of the K^2 2^P-bin (OBLBP) or the 2^{P+1}-bin (IOBLBP) histograms of $\{(I)OBLBP^{k,l}[\mathbf{I}]\}_{k,l=1}^K$.

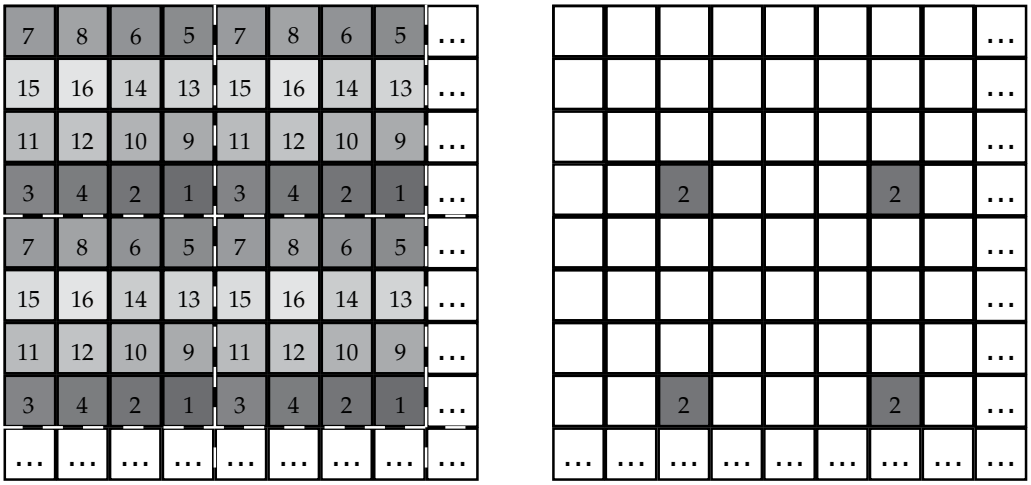
4.4 LBP-based MSFA texture feature

We intend to design an LBP-like operator to characterize multispectral texture images directly from the images acquired by MSFA-based snapshot cameras. A similar approach was proposed by Losson and Macaire [62] for color texture representation from raw CFA images. Rather than a straightforward extension that would neglect spectral correlation, we here propose a new operator dedicated to raw MSFA images and inspired by OBLBPs. We first present the raw image neighborhoods in details in Section 4.4.1, including the specific neighborhoods defined by the MSFAs of Figs. 4.2b and 4.2c. Then we describe our MSFA-based LBP operator in Section 4.4.2, and we explain how it is related to OBLBPs in Section 4.4.3. The neighborhoods considered in our proposed operator in association to each of the four cameras is studied in Section 4.4.4.

4.4.1 MSFA neighborhoods

As defined in Section 2.3.1, an MSFA associates a single spectral band with each pixel. It can be defined as a function $MSFA : S \to \{1,\dots,K\}$ over the set S of all pixels. Let $S^k = \{p \in S, MSFA(p) = k\}$ be the pixel subset where the MSFA samples the band k, such that $S = \bigcup_{k=1}^K S^k$. Fig. 4.4 shows the example of the IMEC16 MSFA and one among its $K = 16$ pixel subsets.

For a given pixel $p \in S^k$, $k \in \{1,\dots,K\}$, let $B^k = \{l \in \{1,\dots,K\}, MSFA(q) = l\}_{q \in N_p}$ be the set of bands that are associated with the neighboring pixels in N_p according to the MSFA. Note that N_p is always composed of pixels with the same associated bands whatever the location of p in S^k. Moreover, we assume that any neighbor $q \in N_p$ is always associated with the same band for a given relative position of q with respect to p in the MSFA pattern. A necessary but not sufficient condition for this assumption to be fulfilled is spectral consistency (see Section 2.2.3). Then, the



(a) (b)

FIGURE 4.4: IMEC16 MSFA (a) and its S^2 pixel subset (b). Dashes on (a) bound the basic pattern of Fig. 2.4b.

neighborhood of $p \in S^k$ can be decomposed into

$$N_p = \bigcup_{l \in B^k} N_p^{k,l}, \tag{4.12}$$

where $N_p^{k,l} = N_p \cap S^l$ is the MSFA-based neighborhood made of the neighboring pixels of p that belong to S^l. Let us notice that $N_p^{k,l} \neq \emptyset \iff l \in B^k$ and stress out that B^k and $N_p^{k,l}$ both depend on $\mathcal{N}^{P,d}$ and on the basic MSFA pattern.

For illustration purposes, let us consider the IMEC16 and VISNIR8 MSFAs and focus on the 3×3 neighborhood defined by the support $\mathcal{N}^{8,1}$ as shown in Fig. 4.5. In the IMEC16 MSFA of Fig. 4.5a, the neighbors of any pixel $p \in S^2$ are associated with the bands $B^2 = \{12,\ 10,\ 9,\ 4,\ 1,\ 8,\ 6,\ 5\}$ and $\left|N_p^{2,l}\right| = 1$ for all $l \in B^2$, where $|\cdot|$ is the cardinal operator. In the VISNIR8 MSFA of Fig. 4.5b, we have $B^2 = \{4,\ 7,\ 3,\ 5,\ 6,\ 8\}$ and $\left|N_p^{2,l}\right| = 1$ for $l \in \{5, 6, 7, 8\}$, but $\left|N_p^{2,3}\right| = \left|N_p^{2,4}\right| = 2$.

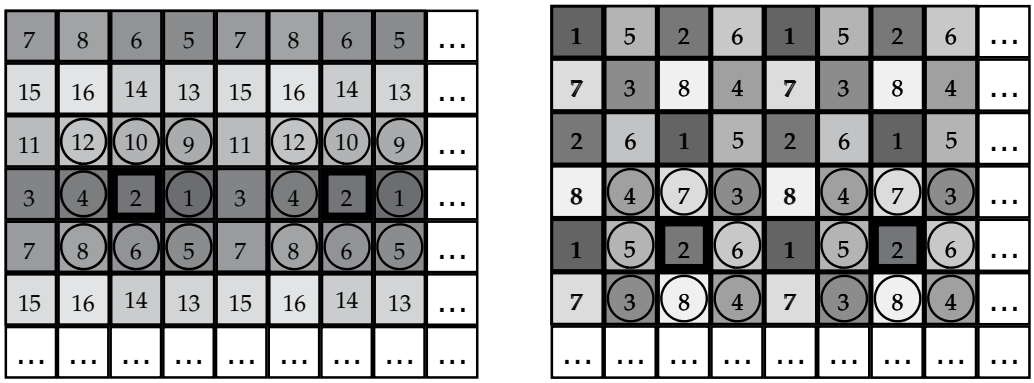
4.4.2 MSFA-based LBPs

A snapshot multispectral camera provides a raw image I^{raw} in which a single band is associated with each pixel according to the MSFA. Then, I^{raw} can be seen as a spectrally-sampled version of the reference fully-defined image $\mathbf{I} = \{I^k\}_{k=1}^K$ (that is unavailable in practice) according to the MSFA (see Section 2.3.1):

$$\forall p \in S,\ I_p^{raw} = I_p^{MSFA(p)}. \tag{4.13}$$

To design a texture feature dedicated to the raw image, let us first consider applying the basic LBP operator of Eq. (4.1) directly to I^{raw} considered as a gray-level image:

$$MLBP\left[I^{raw}\right](p) = \sum_{q \in N_p} s(I_q^{raw}, I_p^{raw}) \cdot 2^{\epsilon(q)}. \tag{4.14}$$



(a) (b)

FIGURE 4.5: Neighborhood N_p defined by the support $\mathcal{N}^{8,1}$ for two pixels $p \in S^2$ (bold squares) in IMEC16 (a) and VISNIR8 MSFAs (b), with associated bands B^2 shown in solid circles.

The LBP operator is here renamed as MSFA-based LBP (MLBP) to make clear the key difference introduced by its application to I^{raw} and its dependency upon the considered MSFA. Unlike Eq. (4.1), Eq. (4.14) combines the spectral information of $B^{MSFA(p)}$, *i.e.*, the different bands that are associated with the neighbors of p.

Because this set of bands depends on the band $MSFA(p)$ associated with p, we separately consider each pixel subset S^k to compute the LBP histogram. Specifically, we compute the histogram of $MLBP[I^{raw}]$ for each band $k \in \{1, ..., K\}$ [75]:

$$\begin{aligned} h^k\left[MLBP\left[I^{raw}\right]\right]\colon [0, 2^P - 1] &\to \{0, \ldots, |S^k|\} \\ j \quad &\mapsto \left|\left\{p \in S^k, MLBP\left[I^{raw}\right](p) = j\right\}\right|. \end{aligned} \tag{4.15}$$

Let us point out that only pixels in S^k are considered to compute the k-th histogram. The concatenation of all the K histograms provides the final feature of size $K \cdot 2^P$.

4.4.3 Relation between MSFA-based and opponent band LBPs

To show that the MSFA-based LBP defined by Eq. (4.14) bears an analogy to OBLBP (see Eq. (4.10)), let us consider its output as the direct sum of the sparse outputs of the same operator restrictively applied to each pixel subset S^k:

$$\operatorname{im}\left\{MLBP\left[I^{raw}\right]\right\} = \bigoplus_{k=1}^{K} \operatorname{im}\left\{MLBP\big|_{S^k}\left[I^{raw}\right]\right\}, \tag{4.16}$$

where $\operatorname{im}\{\cdot\}$ is a function output. According to the definition of S^k, we have $I^{raw}_p = I^k_p$ for each pixel $p \in S^k$. From Eq. (4.14) and the decomposition of the neighborhood N_p according to Eq. (4.12), we can then express $MLBP\big|_{S^k}$ from $\{I^k\}_{k=1}^K$ as

$$MLBP\big|_{S^k}\left[I^{raw}\right](p) = \sum_{l \in B^k} \sum_{q \in N_p^{k,l}} s(I^l_q, I^k_p) \cdot 2^{\epsilon(q)}. \tag{4.17}$$

Therefore, MLBP is related to OBLBP since both operators take opponent bands into account. But unlike OBLBP that considers any band l at all the neighbors of p, each MLBP code combines opponent band information from the $|B^k|$ bands that are available at the neighbors of $p \in S^k$.

4.4.4 Neighborhoods in MSFA-based LBPs

As explained in Section 4.4.1, the neighbors of any pixel p are associated with different bands according to the MSFA. It is thus impossible to consider interpolated values in a circular neighborhood of p as is usually done for LBP-like operators. To avoid interpolation, we therefore consider the uniform spatial distance (hence square neighborhoods) rather than the Euclidean one. Moreover, LBP operators classically use neighborhoods with $P = 8$, 16, or 24 pixels. But $P = 16$ with $d = 3$ does not match the image lattice and requires interpolation, and $P = 24$ would yield extremely large features. We therefore set $P = 8$ and consider the three supports $\mathcal{N}^{8,d}$ with uniform distance $d \in \{1, 2, 3\}$ as shown in Fig. 4.6.

Fig. 4.6 also shows that the number of bands available in the neighborhood of a pixel p generally depends on the distance d for a given MSFA. This number is formalized by $|B^k|$, where $k = MSFA(p) \in \{1, \dots K\}$ is the band associated with p (i.e., $p \in S^k$), and its dependency upon d is summarized in Table 4.1. In VISNIR8 for instance (see Fig. 4.6b), the neighborhood of $p \in S^3$ contains eight different bands for $d = 1$ and $d = 3$, but only the bands 3 and 4 for $d = 2$. $|B^k|$ is also lower for $d = 2$ with VIS5 and IMEC16 MSFAs but is constant to eight whatever $d \in \{1, 2, 3\}$ with IMEC25 due to the large 5×5 basic pattern of this MSFA. Note that $|B^k|$ reflects the degree to which spectral correlation is taken into account by an MSFA neighborhood.

MSFA	$d = 1$	$d = 2$	$d = 3$
VIS5	3 or 5	1 or 2	3 or 5
VISNIR8	6	2	6
IMEC16	8	3	8
IMEC25	8	8	8

TABLE 4.1: Number of available bands $|B^k|$, $k \in \{1, \dots, K\}$, in the neighborhood of any pixel according to each MSFA and each distance.

The basic pattern of the VIS5 MSFA (see Fig. 2.3b) is particular because of its single dominant G band. Unlike in the other considered MSFAs, $|B^k|$ in VIS5 depends on d but also on p (i.e., on k) (see Table 4.1). Considering $d = 1$ for instance, the neighbors of a pixel $p_1 \in S^G$ belong to all the bands (hence $|B^G| = 5$) while those of a pixel $p_2 \in S^C$ belong to $B^C = \{R, G, O\}$. Moreover, VIS5 contradicts our assumption that a neighbor of $p \in S^k$ is always associated with the same band for a given relative position whatever the location of p. Indeed, for two pixels associated with the G band, vertical neighbors may either be associated with R and O or

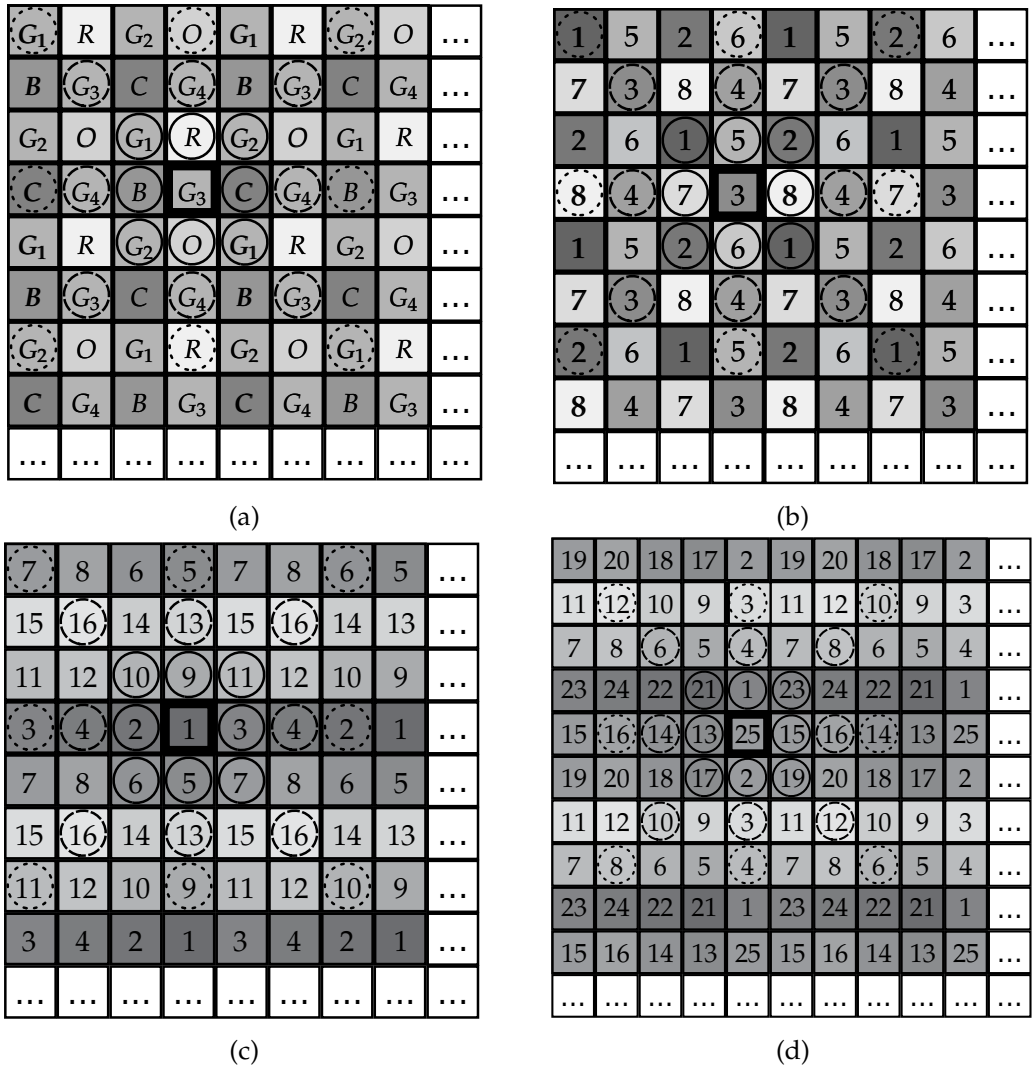


FIGURE 4.6: Neighborhood N_p of a pixel p (bold square) in VIS5 (a), VISNIR8 (b), IMEC16 (c), and IMEC25 (d) MSFAs, considering the supports $\mathcal{N}^{8,1}$ (solid circles), $\mathcal{N}^{8,2}$ (dashed), and $\mathcal{N}^{8,3}$ (dotted).

with B and C. To fulfill our assumption and compute MLBP with VIS5, we therefore split S^G into four pixel subsets $\{S^{G_i}\}_{i=1}^4$ as shown in Fig. 4.6a. The information of $\{MLBP|_{S^{G_i}}[I^{raw}]\}_{i=1}^4$ is then merged into a single histogram $h^G[MLBP[I^{raw}]]$ for the G band.

4.5 Experimental results

We propose to study the sizes of the texture features described in Sections 4.3 and 4.4.2 and their required number of operations per pixel in Section 4.5.1. By considering the simple case with $d = 1$ and the D65 illuminant, we assess the classification accuracy provided by each feature with regard to its computation cost in Section 4.5.2. Finally we extensively assess the performances of our MLBP-based feature with respect to those of other features in various experimental conditions in Section 4.5.3.

4.5.1 Feature extraction

Table 4.2 summarizes the sizes of the texture features described in Section 4.3 as the size of each histogram (that depends on P) and the number of histograms (that depends on K). Setting $P = 8$ makes the histogram size to be 256 but the mdLBP operator provides a prohibitively large number of histograms when $K \geq 16$. All approaches except mdLBP are hence tested against our MSFA-based LBP in the experiments. Besides, among the 16 moment combinations from the format $\{m_1|\hat{m}_1\}$ $\{M_1|\hat{M}_1\}\{\mu_{3|5}|\hat{\mu}_{3|5}\}$ (see Section 4.3.1), we only retain $m_1 M_1 \mu_3$ whose LBP histogram provides the best classification result on average over all the experiments.
The number of histograms may impact the accuracy and the computational burden of classification. The approaches can then be divided into three groups, depending on whether this number is constant (Cusano and Moment LBPs), proportional to K (Marginal LBPs, maLBP, and MLBP), or to K^2 (OBLBP, IOBLBP, and Lee LBPs).
The computation cost also deserves some attention as an indication of the required processing time independently of the implementation. The last two columns of Table 4.2 show this cost as the number of elementary operations per pixel required to compute a feature. This estimation includes all arithmetic operations at the same cost of 1, and excludes array indexing and memory access.

Approach	Eq.	Feature size		Number of operations	
		Hist. size	Number of hist.	Demosaicing	Feature computation
Marginal LBPs [86]	Eq. (4.1)	2^P	K	$4K-4$	$24K$
Moment LBPs [76]	Eq. (4.4), Eq. (4.5)		3		$K^2+10K+69$
maLBP [18]	Eq. (4.6)		$K+1$		$17K+8$
mdLBP [18]	Eq. (4.7)		2^K		$2^K+24K+7$
Cusano LBPs [15]	Eq. (4.8)		2		$15K+27$
Lee LBPs [56]	Eq. (4.9)		$1+K(K-1)$		$27K^2-26K+24$
OBLBP [66]	Eq. (4.10)		K^2		$24K^2$
IOBLBP [6]	Eq. (4.11)	2^{P+1}	K^2		$27K^2+10K$
MLBP [75]	Eq. (4.14), Eq. (4.15)	2^P	K	0	24

TABLE 4.2: Feature size (histogram size and number of concatenated histograms) and required number of operations per pixel for each approach according to the number K of spectral bands.

All features but ours first require to estimate a fully-defined multispectral image by demosaicing. In our experiments, we only consider WB [8] that is both the most simple and generic method (see Section 2.4) and PPID [73] that provides the best demosaicing results in most cases (see Section 3.2). We have adapted PPID to the VIS5 MSFA and we retain it because it globally yields better classification results than the dedicated guided filtering method provided by Monno et al. [81]. Since demosaicing is not the greedier feature computation step, we minimally evaluate its

number of operations as that of the weighted average of two values required by WB to estimate each missing value at a pixel, namely $4(K-1)$.

The feature computation costs are given in the last column of Table 4.2. They result from the equation(s) of each feature recalled in the second column. As previously stated, the computation cost of mdLBP is prohibitive. Our MLBP-based feature requires 24 operations per pixel. In contrast, the cost of Marginal LBPs, maLBP, and Cusano LBPs grows with K, and that of other approaches with K^2. By considering both the feature size and computation cost, Lee LBPs, OBLBP and IOBLBP are the most greedy features. MLBP is the most efficient feature and is represented by the same number of histograms as Marginal LBPs and maLBP. This should be kept in mind while analyzing the classification results.

4.5.2 Accuracy vs. computation cost

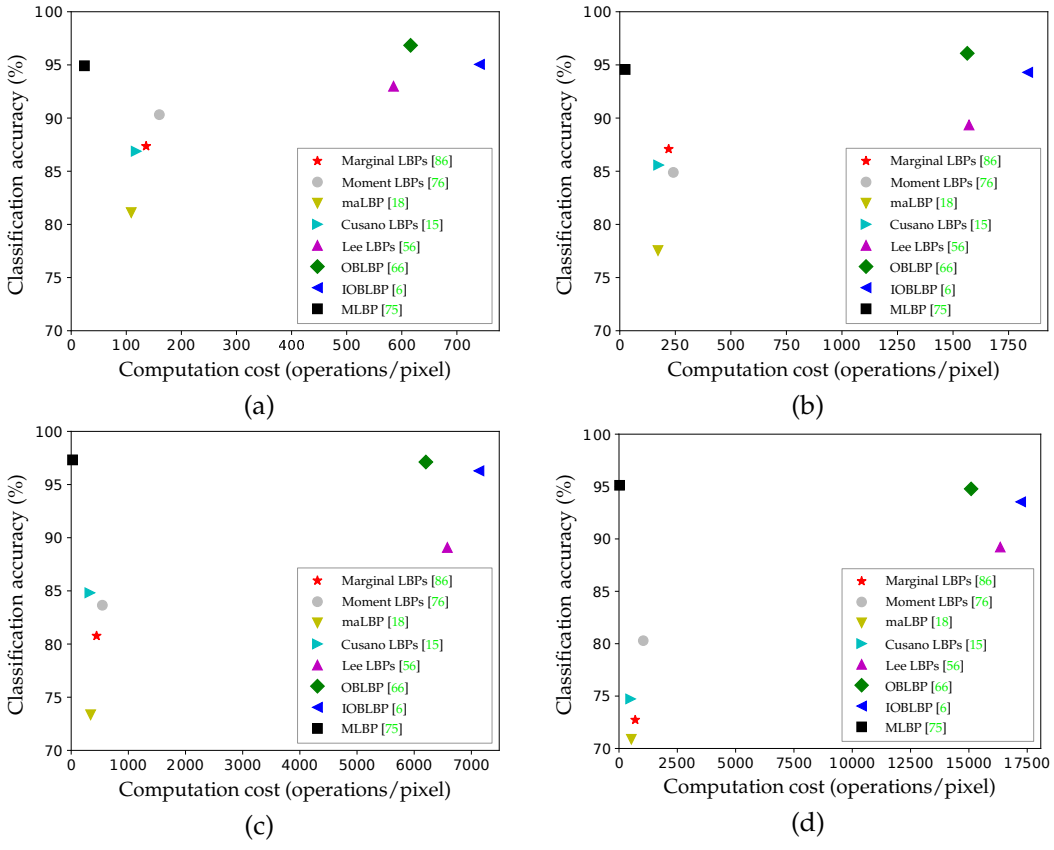


FIGURE 4.7: Classification accuracy (%) vs. computation cost (number of operations per pixel) of the different approaches (with D65 illuminant, $d = 1$, and WB demosaicing) for each MSFA: VIS5 (a, $K = 5$), VISNIR8 (b, $K = 8$), IMEC16 (c, $K = 16$), and IMEC25 (d, $K = 25$).

We first propose a study to highlight the above remark about feature computation costs. Let us consider the case $d = 1$ and the D65 illuminant, and assess the classification accuracy provided by each feature with regard to its cost. For all approaches except MLBP, WB is chosen to demosaic the raw image.

Fig. 4.7 separately shows the results for the four considered MSFAs. OBLBP globally outperforms other features but at a very high computation cost. MLBP provides only slightly lower results than OBLBP for VIS5 and VISNIR8 MSFAs and similar results for IMEC16 and IMEC25 MSFAs, at about a K^2 times smaller cost. Considering features with comparable costs, Lee LBPs and IOBLBP perform worse than OBLBP. Moment LBPs generally provide fair results with regard to the other three features with moderate costs (marginal LBPs, maLBP, and Cusano LBPs). MLBP clearly outperforms the latter four features in all cases with the benefit of reduced computation requirements.

4.5.3 Classification results and discussion

We now extensively assess the performances of our MLBP-based feature with respect to those of other features in various experimental conditions. For this purpose, all the features described in Sections 4.3 and 4.4.2 are implemented in Java under the open-source image analysis software ImageJ [95, 99]. Table 4.3 shows the classification accuracies provided by the different features for each of the four MSFAs using three illuminants (E, D65, and A), three distance values ($d \in \{1,2,3\}$), and two demosaicing methods (WB and PPID). For each combination of experimental settings (illuminant, distance, and demosaicing method), the better classification accuracies than that of our descriptor are highlighted as bold.

Let us first study the behavior of our MLBP-based feature with respect to the different settings. The neighborhood distance d slightly influences MLBP performances for IMEC25 MSFA (see Table 4.3d) whereas $d = 2$ provides clearly lower performances than $d \in \{1,3\}$ for IMEC16 (see Table 4.3c). This was expected since d determines the number $|B^k|$ of available bands in a pixel neighborhood, that does not depend on d for IMEC25 but is reduced from 8 to 3 bands when $d = 2$ for IMEC16 (see Table 4.1). Regarding VIS5 and VISNIR8, there is no systematic performance loss for $d = 2$ though $|B^k|$ is reduced. This is because the poor information about spectral (inter-channel) correlation in these cases is completed by taking spatial (intra-channel) correlation into account. When $d = 2$ indeed, $k \in B^k$ (see Figs. 4.6a and 4.6b), which means that MLBP takes both intra- and inter-channel correlation into account. No such case occurs with IMEC16 and IMEC25 (see Figs. 4.6c and 4.6d) for which MLBP only takes inter-channel correlation into account because $k \notin B^k$ whatever the value of d. Two other outstanding results are obtained with VIS5 MSFA and illuminant A when $d \in \{1,3\}$. We explain these lower accuracies by the shape of illuminant A that increases with respect to the wavelength, which makes pixel values in B and C channels significantly lower than in R and O. Besides, values of pixels in $S^B \cup S^C$ are always compared to neighboring pixels in $S^O \cup S^R$ (or vice versa) when $d \in \{1,3\}$, whereas R values are compared to O values and B values to C values when $d = 2$ (see Fig. 4.6a). Therefore, due to the particular band arrangement in VIS5 MSFA, MLBP is less discriminative and performances are notably reduced when $d \in \{1,3\}$ with A illuminant.

	E						D65						A					
	$d=1$		$d=2$		$d=3$		$d=1$		$d=2$		$d=3$		$d=1$		$d=2$		$d=3$	
	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID
Marginal LBPs [86]	87.71	86.95	88.39	87.50	88.67	88.46	87.36	87.23	88.60	88.26	88.46	88.60	**84.89**	**84.62**	87.71	87.84	88.05	88.19
Moment LBPs [76]	90.11	88.05	88.26	88.67	86.74	86.13	90.32	88.53	89.29	89.01	88.53	87.36	**87.98**	**87.16**	86.74	86.74	85.99	85.71
maLBP [18]	82.83	84.55	86.06	86.40	85.65	86.95	81.11	82.01	84.34	84.89	85.30	86.13	81.66	82.21	85.99	86.20	86.88	88.32
Cusano LBPs [15]	87.64	87.36	88.60	88.26	88.80	89.15	86.88	87.36	88.46	88.39	89.29	89.22	**88.80**	**89.01**	90.52	90.25	**90.32**	**90.87**
Lee LBPs [56]	93.41	93.75	**95.40**	**95.12**	**95.60**	94.85	92.99	93.13	**94.57**	**94.44**	95.47	95.47	**91.48**	**91.96**	**94.23**	**93.82**	**94.64**	**94.37**
OBLBP [66]	**97.39**	**97.46**	**97.39**	**97.53**	**97.12**	**97.46**	**96.84**	**96.70**	**97.18**	**97.39**	**97.12**	**97.39**	**94.64**	**94.51**	**95.67**	**95.95**	**96.29**	**96.57**
IOBLBP [6]	**96.02**	**95.95**	**95.60**	**96.29**	**96.70**	**96.09**	**95.05**	94.57	**95.81**	**96.15**	96.36	96.15	**91.55**	**92.24**	**93.06**	**93.41**	**94.57**	**94.09**
MLBP [75]	93.82		92.86		94.99		94.92		92.45		96.63		84.48		92.99		89.90	

(A)

	E						D65						A					
	$d=1$		$d=2$		$d=3$		$d=1$		$d=2$		$d=3$		$d=1$		$d=2$		$d=3$	
	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID
Marginal LBPs [86]	84.89	86.40	85.44	85.78	86.06	86.13	87.09	88.94	87.50	88.67	87.64	88.53	84.00	85.51	85.03	85.37	85.78	85.51
Moment LBPs [76]	86.26	86.40	85.85	85.37	86.06	84.75	84.89	84.62	85.71	84.55	85.51	85.23	82.76	83.04	83.93	83.79	84.20	83.79
maLBP [18]	77.75	77.88	79.46	79.33	82.55	79.95	77.54	79.33	80.22	81.11	79.95	79.95	78.23	76.99	79.12	78.78	81.46	80.29
Cusano LBPs [15]	82.49	82.55	83.93	83.72	84.34	85.03	85.58	85.51	87.16	87.09	87.16	87.57	80.77	80.43	82.42	82.01	82.49	83.17
Lee LBPs [56]	91.28	91.69	93.96	93.89	**95.05**	**95.26**	89.35	91.41	91.83	91.69	93.48	93.41	91.62	92.86	93.82	93.89	**95.26**	**94.16**
OBLBP [66]	**96.29**	**96.91**	**96.63**	**96.57**	**96.57**	**96.63**	**96.09**	**96.22**	**96.15**	**95.67**	**96.09**	**95.54**	**95.19**	**95.40**	**95.47**	**95.12**	**95.12**	**95.33**
IOBLBP [6]	**94.09**	93.61	94.64	94.30	94.44	**94.57**	94.30	94.02	93.82	93.61	94.23	93.48	92.10	**93.34**	92.65	92.31	92.86	92.51
MLBP [75]	93.75		95.67		94.57		94.57		94.71		94.64		93.06		94.02		93.20	

(B)

	E						D65						A					
	$d=1$		$d=2$		$d=3$		$d=1$		$d=2$		$d=3$		$d=1$		$d=2$		$d=3$	
	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID
Marginal LBPs [86]	81.59	85.10	83.79	86.54	83.86	86.20	81.11	84.89	83.72	86.95	85.30	87.16	80.70	84.89	83.10	85.92	85.71	87.50
Moment LBPs [76]	85.30	89.56	87.50	90.32	87.71	90.04	84.41	88.67	88.12	89.42	87.02	89.90	84.34	88.05	88.05	89.84	88.05	89.15
maLBP [18]	74.93	78.57	75.21	76.17	76.51	78.57	73.56	79.12	75.41	77.13	75.34	80.08	75.07	77.13	74.52	73.97	76.51	74.86
Cusano LBPs [15]	84.07	86.40	86.20	87.98	85.92	87.98	85.30	86.20	86.33	87.91	86.54	87.02	85.44	86.61	87.02	88.19	87.16	88.67
Lee LBPs [56]	89.77	95.40	**92.45**	**96.02**	93.68	95.95	89.22	95.26	92.58	**95.95**	93.41	95.81	87.09	94.44	**90.93**	**94.09**	92.93	94.64
OBLBP [66]	**96.91**	**96.57**	**96.22**	**96.77**	96.09	**96.50**	97.25	**97.60**	**97.05**	**97.05**	96.84	96.98	**95.67**	**96.09**	**95.47**	**95.81**	95.05	**95.95**
IOBLBP [6]	95.40	**96.36**	**95.19**	**95.88**	95.33	95.47	96.02	96.91	**95.88**	**96.57**	95.74	96.29	93.54	94.71	**94.44**	**94.78**	93.61	94.64
MLBP [75]	96.22		90.87		96.29		97.32		94.92		97.05		95.33		90.32		95.19	

(C)

	E						D65						A					
	$d=1$		$d=2$		$d=3$		$d=1$		$d=2$		$d=3$		$d=1$		$d=2$		$d=3$	
	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID	WB	PPID
Marginal LBPs [86]	73.15	78.57	74.93	78.71	75.62	78.50	73.63	78.71	75.76	80.01	77.47	79.81	72.87	78.09	74.45	78.43	75.82	77.54
Moment LBPs [76]	79.46	83.17	80.56	83.72	80.77	84.27	78.98	84.20	81.25	85.51	81.46	85.16	78.71	83.31	80.15	83.93	80.36	84.07
maLBP [18]	63.67	78.64	61.26	78.50	67.03	76.79	68.27	78.09	62.77	77.06	65.73	75.14	64.08	77.88	60.37	78.37	67.17	75.55
Cusano LBPs [15]	77.13	81.80	78.57	82.69	77.68	81.39	75.34	80.15	76.51	82.35	78.98	82.14	76.65	81.18	78.64	81.94	77.75	80.70
Lee LBPs [56]	87.77	92.58	92.03	**93.89**	**92.79**	**94.57**	89.49	94.16	91.14	94.78	92.31	**94.92**	88.12	91.96	**91.96**	**93.54**	**92.65**	**94.23**
OBLBP [66]	94.02	**95.05**	**93.41**	**94.92**	**93.48**	**94.71**	94.99	**95.88**	94.51	**95.95**	94.51	**95.60**	93.75	94.51	**93.13**	**94.71**	**93.06**	**94.16**
IOBLBP [6]	92.93	**94.37**	91.83	**93.68**	91.21	**92.58**	94.09	94.92	93.82	**95.33**	92.79	**94.85**	92.51	93.48	91.35	**93.54**	90.73	**92.24**
MLBP [75]	94.30		93.13		92.03		95.40		95.33		94.78		94.71		91.76		92.24	

(D)

TABLE 4.3: Classification accuracy (%) of the different approaches for each experimental setting (illuminant, neighborhood distance, demosaicing method) and each MSFA: VIS5 (a, $K=5$), VISNIR8 (b, $K=8$), IMEC16 (c, $K=16$), and IMEC25 (d, $K=25$).

Let us now compare the performances reached by our descriptor with those of other approaches. Table 4.3 shows that, except with VIS5 MSFA images simulated under A illuminant, our MLBP-based feature always outperforms approaches with either smaller (Cusano and Moment LBPs) or similar-size features (Marginal LBPs and maLBP). Moreover, our lightweight approach obtains close results to greedy ones (Lee LBP, OBLBP and IOBLBP), especially with IMEC16 and IMEC25 MSFAs, and even performs better than them in 95 out of the 216 tested cases. The best accuracy reached by MLBP is 97.32% (with IMEC16 MSFA under D65 illuminant using $d = 1$) while the best descriptor (OBLBP) reaches 97.60% (with the same settings and PPID demosaicing).

4.6 Conclusion

To classify multispectral texture scenes from the images that would have been acquired by single-sensor snapshot cameras, we have adopted a classification scheme based on histogram of local binary pattern (LBP) as texture descriptor, intersection between histograms as similarity measure, and 1-nearest neighbor as decision rule. We have extended some state of the art LBP operators that extract features using both spatial and color properties to any multispectral image. However, the computational cost significantly increases with the number of channels. We have then introduced a conceptually simple and highly-discriminative LBP-based feature for multispectral raw images. In addition to its algorithmic simplicity, our operator is directly applied to raw images, which avoids the demosaicing step and keeps its computational cost low. We have performed extensive experiments of texture classification on multispectral images simulated from HyTexiLa database with four well-referenced MSFAs. The results show that the proposed approach outperforms existing ones using features of similar sizes, and provides comparable results to that of features with large size and high computational cost.

Conclusion and future works

Conclusion

This manuscript can be summarized into four main contributions, each of which is detailed in a specific chapter.
First, in collaboration with the Norwegian Colour and Visual Computing laboratory, we have extended the collection of available multispectral image databases by proposing ours. This database is composed of 112 close-range images of textured surfaces observed in the visible and near infrared domains. It has been born from the need of the community and should be used in many application fields such as object recognition by multispectral texture classification or material characterization.
The second contribution is the improvement of MSFA demosaicing performances by using the strong correlation between all channels and the pseudo-panchromatic image (PPI). Indeed, the latter can be estimated directly from the MSFA raw image using a simple averaging filter. This first estimation is then improved using local directional variations of raw values to restore edge information. We then incorporate the estimated PPI into existing DWT-based and edge-sensing-based methods, and propose a new demosaicing method (PPID) based on the difference between each channel and the PPI. Extensive experiments show that PPI-based demosaicing outperforms the existing demosaicing methods at a moderate computational cost. PPID compares favorably with the state of the art both objectively in terms of $PSNR$ and ΔE^* color difference, and in a subjective visual assessment.
The third contribution is based on the study of the effect of acquisition conditions on MSFA demosaicing. Indeed, when illumination or spectral sensitivity functions (SSFs) of the camera are weak in term of energy, spectral correlation is strongly reduced. Demosaicing methods that use this property are then affected. To overcome this limitation, we propose to insert normalization steps in the imaging pipeline to adjust the channel levels before demosaicing and restore them afterwards. The channel-specific normalization factor can be deduced either from the SSFs of the camera, from the relative spectral power distribution of illumination, or directly estimated from the raw image. Experimental results show that normalization based on the sole SSFs provides good but illumination-sensitive results. Normalization based on SSFs and illumination information provides the best results despite illumination function is not always available in practice. At last raw image-based normalization provides promising results without any prior knowledge about the camera or illumination, and thus constitutes a good compromise for MSFA demosaicing.

The fourth contribution is related to multispectral texture image classification. Indeed, we propose a feature based on the local binary pattern (LBP) operator that is directly applied to the raw image. In addition to its algorithmic simplicity, our feature allows us to avoid the demosaicing step, which makes it fast to be computed with respect to classical LBP-based approaches. Extensive experiments of texture classification on simulated multispectral images with four well-referenced multispectral filter arrays show that the proposed approach outperforms existing ones using features of similar sizes, and provides comparable results to those of features with large size and high computational cost.

Future works

We can identify several future works from this thesis. First work focuses on MSFA raw image analysis, that can then be used for weed recognition. Recent new cameras equipped with polarized filter arrays (PFAs) give new challenge that are detailed in the two last parts.

MSFA raw images analysis

Although our multispectral simulation model has been validated (see Section 1.5.3), a preliminary work shows that some channels are likely to undergo noise. This can be due to the weak illumination or to low sensitivities of the filters associated to those spectral bands, where the limit of the used optical model is reached [16]. Future works will focus on the improvement of our image formation model to take noise into account with respect to both the SSFs and illumination in "multishot" and "snapshot" acquisition systems. By using such a model we could be able to adapt joint CFA denoising-demosaicing methods (*e.g.*, [36]) to MSFA demosaicing. Then we could compare the performances of these methods with those of the recent learning-based and compressed sensing-based methods (see Section 2.3.3).
Regarding multispectral texture classification, future works will study how our feature embeds spatial and spectral correlations according to the MSFA and neighborhood parameters. Since MLBP is a small-size feature, there is room for additional correlation information that could still improve its classification results. For instance, it could be made more robust to the neighborhood distance by concatenating several MLBP histograms computed with different spatial distances. Other investigations could use a demosaiced dominant channel or focus on the spectral distance between the considered neighbors.
Our proposed MLBP operator has only been tested on simulated raw images. Future work will focus on the creation of an MSFA raw image database composed of the same textures as HyTexiLa. This database will be acquired using IMEC16 snapshot camera and will be useful to validate classification accuracy results on ground truth data.
In this thesis, we have proposed texture features that are well adapted for given

MSFAs. We can also attempt to design the optimal MSFA pattern for texture classification using our MLBP operator. Indeed, as seen in Chapter 4, the camera SSFs and number of channels, illuminations, and MSFA basic pattern impact classification performances. Future works will further adjust these parameters in order to find an optimal MSFA pattern for texture classification.

Application to weed recognition

Texture classification has many applications, among which weed detection. Indeed, weed control coupled with precision agriculture limits the use of herbicides and is a major challenge for farmers and a priority of the Ecophyto plan. Future works will focus on real-time weed recognition from images acquired by snapshot multispectral cameras embedded on drones. As these cameras observe outdoor field crops, the lighting and field of view may vary. Therefore, the spatial resolution and spectral properties of images that represent the same weed species may change. Such problem will lead us to the development of new MSFAs and texture features that are invariant to illumination and spatial resolution changes.

PFA raw images analysis

Recent advances in PFAs that provide panchromatic images with four different polarization angles constitute also a new challenge for demosaicing. In our future work, we will study the correlations between polarization channels and use them to demosaic PFA raw images. Indeed polarization channels have different properties that may be used for demosaicing, while classical PFA demosaicing methods are mainly based on spatial correlation only [25]. Moreover, it would be interesting to extend/adapt classical MSFA or CFA demosaicing approaches to PFA images.

MPFA raw images analysis

Recently, Multispectral PFA (MPFA) that provide multispectral images with four different polarization angles have been developed [108]. Such MPFAs involve to study the relation between spectral channels and polarized channels that can be used for demosaicing.
In order to improve texture classification, texture descriptors have been extended from the color to the multispectral domain. It would be interesting to adapt them to multispectral polarized images that can also characterize glossed textures. Future work would extend our MLBP operator to MPFA raw images.

Appendix A

Conversions from multispectral to XYZ, sRGB and L*a*b* spaces

A multispectral image can be converted to any color space by using CIE XYZ space as a lever. XYZ space associates the Y channel to luminance, and describes visible chromaticity using X and Z channels, such that the associated color matching function have always positive values as shown in Fig. A.1.

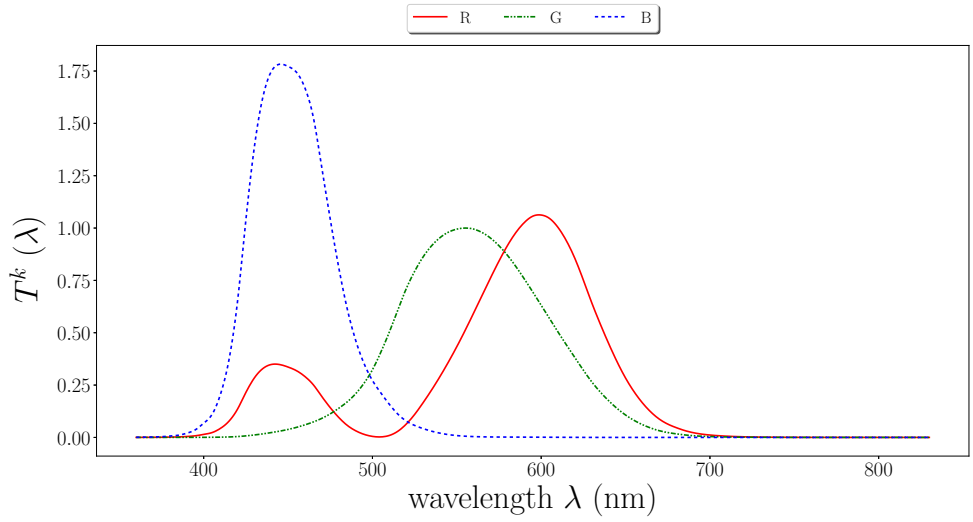


FIGURE A.1: CIE XYZ color matching functions.

The conversion from a multispectral space to XYZ space is done according to CIE XYZ 2° standard observer. Each XYZ channel I^k, $k \in \{X, Y, Z\}$, is defined at each pixel p as [11]:

$$I_p^k = \frac{100}{\sum_{\lambda \in \Omega} E(\lambda) \mathcal{T}^Y(\lambda)} \cdot \sum_{\lambda \in \Omega} E(\lambda) \cdot R_p(\lambda) \cdot \mathcal{T}^k(\lambda) \,. \tag{A.1}$$

The reflectance $R_p(\lambda)$ can be computed from estimated reflectance databases described in Table 1.2, coupled with any illumination described in Section 1.2 either in the Vis (Fig. 1.2) or in the VisNIR (Fig. 1.3) domain. Alternatively, the radiance $E(\lambda) \cdot R_p(\lambda)$ can be computed from one of the public radiance image databases of

Table 1.1. In this case illumination data must be known to compute this transformation. The three channels (I^X, I^Y, and I^Z) compose the representation of the multispectral image in XYZ space. Then, images can be converted from CIE XYZ to sRGB or CIE L*a*b* spaces among others (see Appendices A.1 and A.2).
The standard RGB (sRGB) color space is a digital color space used on monitors, printers, and the Internet [112]. We use it in this manuscript to represent a color version of multispectral images. L*a*b* color space is supposed to be perceptually uniform with respect to human color vision because the distance between two points in this space represents the difference between these colors as visually perceived. In this model, I^{L^*} is the magnitude of lightness, while I^{a^*} and I^{b^*} represent respectively the red−green and yellow−blue difference channels (with signed values). This color space is especially required in order to compute the CIE ΔE^* that is a measure of the difference between two colors (see Section 3.2.2).

A.1 From XYZ to sRGB color space

The transformation from XYZ channels $\{I^k\}_{k=X,Y,Z}$, to sRGB channels $\{I^{l'}\}_{l=R,G,B}$ proceeds at each pixel p as follows:

- Normalization of XYZ channels $\{I^k\}_{k=X,Y,Z}$ to values that range between 0 and 1:

$$I_p^k = \frac{I_p^k}{max(I_p^X, I_p^Y, I_p^Z)} \tag{A.2}$$

- Linear transformation of XYZ channels (using D65 as reference white):

$$\begin{bmatrix} I_p^{\tilde{R}} \\ I_p^{\tilde{G}} \\ I_p^{\tilde{B}} \end{bmatrix} = \begin{bmatrix} 3.2406 & -1.5372 & -0.4986 \\ -0.9689 & 1.8758 & 0.0415 \\ 0.0557 & -0.2040 & 1.0570 \end{bmatrix} \cdot \begin{bmatrix} I_p^X \\ I_p^Y \\ I_p^Z \end{bmatrix} \tag{A.3}$$

  This transformation uses values of channels I^X, I^Y, and I^Z normalized between 0 and 1 as inputs and provides channels $I^{\tilde{R}}$, $I^{\tilde{G}}$, and $I^{\tilde{B}}$ whose values are afterwards clipped to 1.

- Non-linear transformation of $\tilde{R}\tilde{G}\tilde{B}$ channels $\{I^{\tilde{l}}\}_{l=R,G,B}$ into sRGB channels $\{I^{l'}\}_{l=R,G,B}$ defined at each pixel p as [112]:

$$I_p^{l'} = \begin{cases} Q(12.92 \cdot I_p^{\tilde{l}}) & \text{if } I_p^{\tilde{l}} < 0.0031308\ , \\ Q(1.055 \cdot \sqrt[2.4]{I_p^{\tilde{l}}} - 0.055) & \text{otherwise,} \end{cases} \tag{A.4}$$

  where $Q(\cdot)$ quantizes $I_p^{l'}$ on 8 bits as $Q(I_p^{l'}) = \lfloor 2^8 - 1 \cdot I_p^{l'} \rceil$.

The three sRGB channels form the sRGB image.

A.2 From XYZ to L*a*b* color space

The conversion from XYZ channels $\{I^k\}_{k=X,Y,Z}$ to L*a*b* channels $\{I^l\}_{l=L^*,a^*,b^*}$ proceeds at each pixel p as follows:

- Normalization of channels $\{I^k\}_{k=X,Y,Z}$ with respect to the response of a perfect diffuser for a CIE XYZ 2° standard observer:

$$I_p^k = \frac{\sum\limits_{\lambda\in\Omega} \mathcal{T}^Y(\lambda)E(\lambda)}{100 \cdot \sum\limits_{\lambda\in\Omega} \mathcal{T}^k(\lambda)E(\lambda)} \cdot I_p^k \,. \tag{A.5}$$

- Computation of values $\{f_p^k\}_{k=X,Y,Z}$ of normalized XYZ channels:

$$f_p^k = \begin{cases} \sqrt[3]{I_p^k} & \text{if } I_p^k > 0.008856 \,, \\ \frac{903.3 \cdot I_p^k + 16}{116} & \text{otherwise.} \end{cases} \tag{A.6}$$

- The channels of L*a*b* color space are finally given by:

$$\begin{cases} I_p^{L^*} = 116 \cdot f_p^Y - 16 \,, \\ I_p^{a^*} = 500 \cdot (f_p^X - f_p^Y) \,, \\ I_p^{b^*} = 200 \cdot (f_p^Y - f_p^Z) \,, \end{cases} \tag{A.7}$$

  where $I_p^{L^*}$ values range between 0% and 100%, $I_p^{a^*}$ values go from the green (negative values down to -300) to the red (positive values up to 300), and $I_p^{b^*}$ values go from the blue (negative values down to -300) to the yellow (positive values up to 300).

Appendix B

Spectral sensitivity functions

IMEC16 camera samples 16 bands with known SSFs $T^k(\lambda)$ that are unevenly centered at wavelengths $\lambda^k \in \mathcal{B}(\text{IMEC16}) = \{469\,nm, \ldots, 633\,nm\}$, so that $\lambda^1 = 469\,nm, \ldots, \lambda^{16} = 633\,nm$ (see Fig. B.1).
Similarly, IMEC25 camera samples 25 bands whose SSFs are unevenly centered at wavelengths $\lambda^k \in \mathcal{B}(\text{IMEC25}) = \{678\,nm, \ldots, 960\,nm\}$ (see Fig. B.2). Note that the optical device of both cameras is equipped with a band-pass filter (at 450–650 nm for IMEC16 and 675–975 nm for IMEC25) in order to avoid second-order spectral artifacts.
The SSFs associated to VISNIR8 and VIS5 cameras are linearly interpolated to 1 nm-bandwidths and shown in Figs. B.3 and B.4. VISNIR8 SSFs are associated with their peaks at $\lambda^k \in \mathcal{B}(\text{VISNIR8}) = \{440, \ldots, 880\,nm\}$ (see Fig. B.3) while VIS5 SSFs are associated with their dominant color $\lambda^k \in \mathcal{B}(\text{VIS5}) = \{B, Cy, G, Or, R\}$ (see Fig. B.4).
Note that the SSFs of each camera (Cam) are scaled so that $\max_{k \in \mathcal{B}(Cam)} \sum_{\lambda \in \Omega} T^k(\lambda) = 1$.

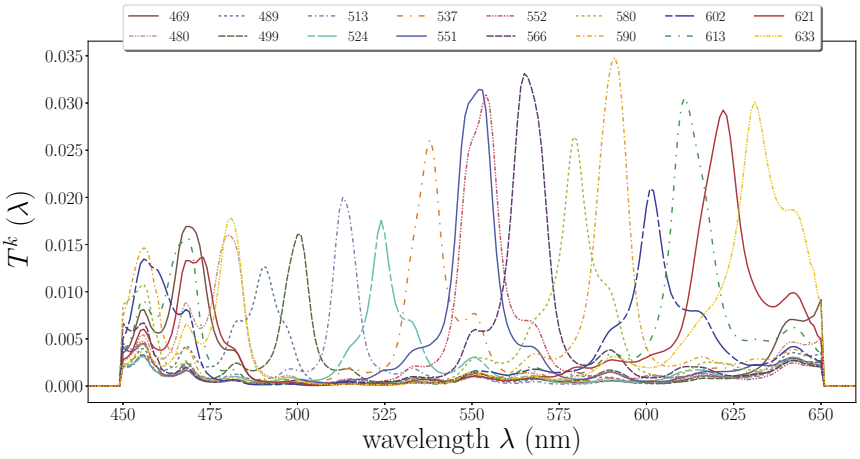


FIGURE B.1: Normalized SSFs of IMEC16 camera. Captions: band center wavelengths $\{\lambda^k\}_{k=1}^{16}$ in ascending order.

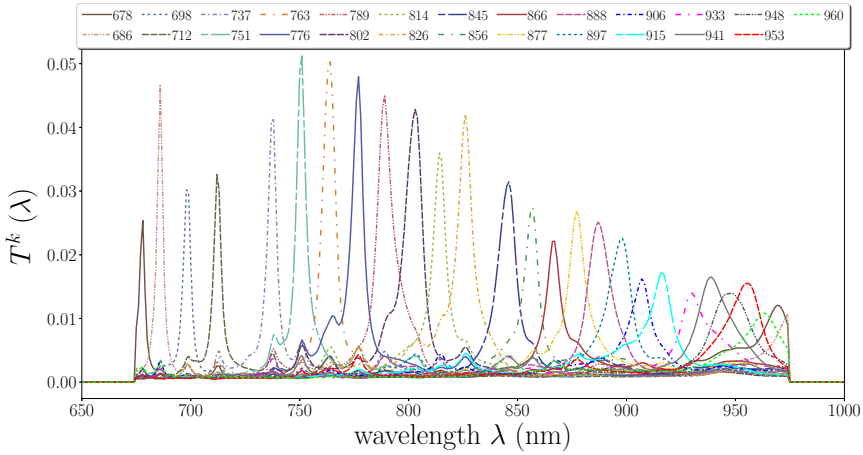


FIGURE B.2: Normalized SSFs of IMEC25 camera. Captions: band center wavelengths $\{\lambda^k\}_{k=1}^{25}$ in ascending order.

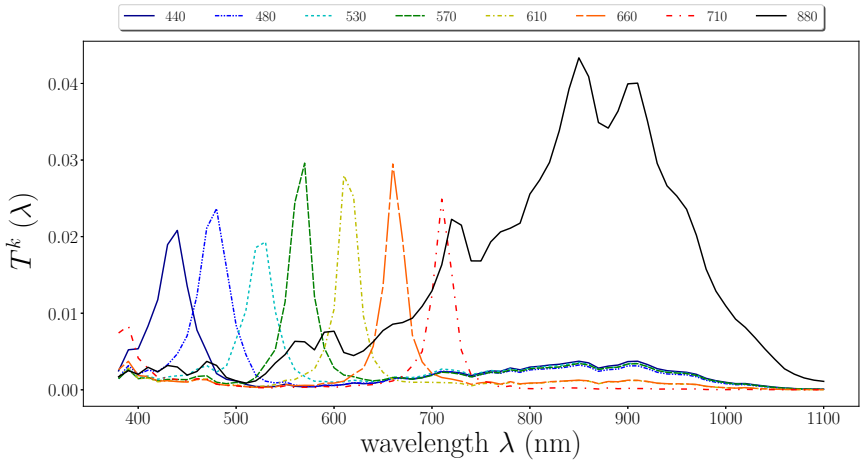


FIGURE B.3: Normalized SSFs of VISNIR8 camera. Captions: band center wavelengths $\{\lambda^k\}_{k=1}^{8}$ in ascending order.

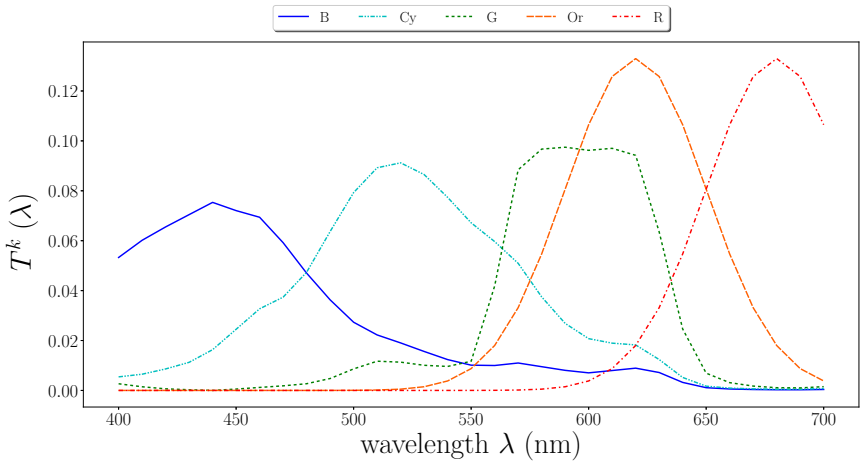


FIGURE B.4: Normalized SSFs of VIS5 camera. Captions: dominant colors.

Appendix C

Weight computation for demosaicing

C.1 Weight computation in BTES

To estimate the channel $\hat{I}^k$ at pixel p, the weight α_q of neighboring pixel q used in BTES (see Eq. (2.12)) is computed according to the direction given by p and q as:

- for a horizontal direction (case $t = 3$):

$$\alpha_q^H = \Big(1 + \left|\dot{I}^k_{q+(2,0)} - \dot{I}^k_q\right| + \left|\dot{I}^k_{q-(2,0)} - \dot{I}^k_q\right| + \tfrac{1}{2}\left|\dot{I}^k_{q+(-1,-1)} - \dot{I}^k_{q+(1,-1)}\right| + \tfrac{1}{2}\left|\dot{I}^k_{q+(-1,1)} - \dot{I}^k_{q+(1,1)}\right|\Big)^{-1}, \tag{C.1}$$

- for a vertical direction (case $t = 3$):

$$\alpha_q^V = \Big(1 + \left|\dot{I}^k_{q+(0,2)} - \dot{I}^k_q\right| + \left|\dot{I}^k_{q-(0,2)} - \dot{I}^k_q\right| + \tfrac{1}{2}\left|\dot{I}^k_{q+(1,-1)} - \dot{I}^k_{q+(1,1)}\right| + \tfrac{1}{2}\left|\dot{I}^k_{q+(-1,-1)} - \dot{I}^k_{q+(-1,1)}\right|\Big)^{-1}, \tag{C.2}$$

- for the first diagonal direction (case $t = 2$):

$$\alpha_q^{D_1} = \Big(1 + \left|\dot{I}^k_{q+(2,2)} - \dot{I}^k_q\right| + \left|\dot{I}^k_{q-(2,2)} - \dot{I}^k_q\right|\Big)^{-1}, \tag{C.3}$$

- for the second diagonal direction (case $t = 2$):

$$\alpha_q^{D_2} = \Big(1 + \left|\dot{I}^k_{q+(2,-2)} - \dot{I}^k_q\right| + \left|\dot{I}^k_{q+(-2,2)} - \dot{I}^k_q\right|\Big)^{-1}, \tag{C.4}$$

where $\dot{I}^k_q$ expresses that I^k is available at q in I^{raw} or has been previously estimated at q. Note that the weights for $t = 0$ and $t = 1$ are undetermined and replaced by 1.

C.2 Weight computation in MLDI

Let $q = p + (\delta_x, \delta_y)$ be a neighboring pixel of p, and $r = p + 2 \cdot (\delta_x, \delta_y)$ (see Fig. 2.11). To estimate $\hat{I}^k_p$, the weight β_q of q used in MLDI (see Eq. (2.13)) is computed at step t according to the direction given by p and q as:

- for an horizontal direction:

$$\beta^H_q = \Big(\epsilon + \left|\dot{I}^{MSFA(p)}_r - I^{MSFA(p)}_p\right| + \sum_{d=0}^{\Delta-1} \left| I^{raw}_{p+\left(\frac{\delta_x}{|\delta_x|}\cdot(\Delta+d)),0\right)} - I^{raw}_{p-\left(\frac{\delta_x}{|\delta_x|}\cdot(\Delta-d),0\right)} \right| + \sum_{\substack{d=-\Delta \\ d\neq 0}}^{\Delta} \omega_d \cdot \left| I^{raw}_{p+(2\cdot\delta_x,d)} - I^{raw}_{p+(0,d)} \right| \Big)^{-1}, \tag{C.5}$$

- for a vertical direction:

$$\beta^V_q = \Big(\epsilon + |\dot{I}^{MSFA(p)}_r - I^{MSFA(p)}_p| + \sum_{d=0}^{\Delta-1} \left| I^{raw}_{p+\left(0,\frac{\delta_y}{|\delta_y|}\cdot(\Delta+d)\right)} - I^{raw}_{p-\left(0,\frac{\delta_y}{|\delta_y|}\cdot(\Delta-d)\right)} \right| + \sum_{\substack{d=-\Delta \\ d\neq 0}}^{\Delta} \omega_d \cdot \left| I^{raw}_{p+(d,2\cdot\delta_y)} - I^{raw}_{p+(d,0)} \right| \Big)^{-1}, \tag{C.6}$$

- for a diagonal direction:

$$\beta^D_q = \Big(\epsilon + \left|\dot{I}^k_{p+(\delta_x,\delta_y)} - \dot{I}^k_{p-(\delta_x,\delta_y)}\right| + \left|\dot{I}^{MSFA(p)}_r - I^{MSFA(p)}_p\right| + \left|\hat{I}^{MSFA(p)}_{WB}(q) - I^{MSFA(p)}_p\right| \Big)^{-1}, \tag{C.7}$$

where $MSFA(p)$ is the available channel index at p in I^{raw}, $\omega_d = \frac{\exp\left(-\frac{d^2}{2\cdot 0.5^2}\right)}{2\cdot\sum_{u=1}^{\Delta}\exp\left(-\frac{u^2}{2\cdot 0.5^2}\right)}$, $\Delta = 2 - \lfloor t/2 \rfloor$, and $\epsilon = 0.01$.

C.3 Weight computation in PPBTES

The weight α_q of neighboring pixel q used in PPBTES are computed according to the direction given by p and q, at steps $t \in \{0, \ldots, 3\}$ as:

- for a horizontal direction (case $t = 3$):

$$\alpha^H_q = \Big(1 + \left|\dot{I}^k_{q+(2,0)} - \dot{I}^k_q\right| + \left|\dot{I}^k_{q+(-2,0)} - \dot{I}^k_q\right| + \frac{1}{2}\left|\dot{I}^k_{q+(-1,-1)} - \dot{I}^k_{q+(1,-1)}\right| + \frac{1}{2}\left|\dot{I}^k_{q+(-1,1)} - \dot{I}^k_{q+(1,1)}\right| \Big)^{-1}, \tag{C.8}$$

- for a vertical direction (case $t = 3$):

$$\alpha_q^V = \Big(1 + \left|\dot{I}^k_{q+(0,2)} - \dot{I}^k_q\right| + \left|\dot{I}^k_{q+(0,-2)} - \dot{I}^k_q\right| + \frac{1}{2}\left|\dot{I}^k_{q+(-1,-1)} - \dot{I}^k_{q+(-1,1)}\right| + \frac{1}{2}\left|\dot{I}^k_{q+(1,-1)} - \dot{I}^k_{q+(1,1)}\right|\Big)^{-1}, \tag{C.9}$$

- for a first diagonal direction (case $t = 2$):

$$\alpha_q^{D_1} = \Big(1 + \left|\dot{I}^k_{q+(2,2)} - \dot{I}^k_q\right| + \left|\dot{I}^k_{q+(-2,-2)} - \dot{I}^k_q\right| + \left|\hat{I}^{PPI}_{q+(-1,-1)} - \hat{I}^{PPI}_{q+(1,1)}\right|\Big)^{-1}, \tag{C.10}$$

- for a second diagonal direction (case $t = 2$):

$$\alpha_q^{D_2} = \Big(1 + \left|\dot{I}^k_{q+(2,-2)} - \dot{I}^k_q\right| + \left|\dot{I}^k_{q+(-2,2)} - \dot{I}^k_q\right| + \left|\hat{I}^{PPI}_{q+(-1,1)} - \hat{I}^{PPI}_{q+(1,-1)}\right|\Big)^{-1}, \tag{C.11}$$

- for a horizontal direction (case $t = 1$):

$$\alpha_q^H = \Big(1 + \left|\hat{I}^{PPI}_{q+(2,0)} - \hat{I}^{PPI}_q\right| + \left|\hat{I}^{PPI}_{q+(-2,0)} - \hat{I}^{PPI}_q\right| + \frac{1}{2}\left|\hat{I}^{PPI}_{q+(-1,-1)} - \hat{I}^{PPI}_{q+(1,-1)}\right| + \frac{1}{2}\left|\hat{I}^{PPI}_{q+(-1,1)} - \hat{I}^{PPI}_{q+(1,1)}\right|\Big)^{-1}, \tag{C.12}$$

- for a vertical direction (case $t = 1$):

$$\alpha_q^V = \Big(1 + \left|\hat{I}^{PPI}_{q+(0,2)} - \hat{I}^{PPI}_q\right| + \left|\hat{I}^{PPI}_{q+(0,-2)} - \hat{I}^{PPI}_q\right| + \frac{1}{2}\left|\hat{I}^{PPI}_{q+(-1,-1)} - \hat{I}^{PPI}_{q+(-1,1)}\right| + \frac{1}{2}\left|\hat{I}^{PPI}_{q+(1,-1)} - \hat{I}^{PPI}_{q+(1,1)}\right|\Big)^{-1}, \tag{C.13}$$

- for a first diagonal direction (case $t = 0$):

$$\alpha_q^{D_1} = \Big(1 + \left|\hat{I}^{PPI}_{q+(2,2)} - \hat{I}^{PPI}_q\right| + \left|\hat{I}^{PPI}_{q+(-2,-2)} - \hat{I}^{PPI}_q\right| + \left|\hat{I}^{PPI}_{q+(-1,-1)} - \hat{I}^{PPI}_{q+(1,1)}\right|\Big)^{-1}, \tag{C.14}$$

- for a second diagonal direction (case $t = 0$):

$$\alpha_q^{D_2} = \Big(1 + \left|\hat{I}^{PPI}_{q+(2,-2)} - \hat{I}^{PPI}_q\right| + \left|\hat{I}^{PPI}_{q+(-2,2)} - \hat{I}^{PPI}_q\right| + \left|\hat{I}^{PPI}_{q+(-1,1)} - \hat{I}^{PPI}_{q+(1,-1)}\right|\Big)^{-1}. \tag{C.15}$$

Bibliography

[1] H. K. Aggarwal and A. Majumdar, "Single-sensor multi-spectral image demosaicing algorithm using learned interpolation weights," in *Proceedings of the 2014 International Geoscience and Remote Sensing Symposium (IGARSS 2014)*, Quebec City, Quebec, Canada, Jul. 2014, pp. 2011–2014.

[2] H. K. Aggarwal and A. Majumdar, "Compressive sensing multi-spectral demosaicing from single sensor architecture," in *Proceedings of the IEEE China Summit International Conference on Signal and Information Processing (ChinaSIP'2014)*, Xi'an, China, Jul. 2014, pp. 334–338.

[3] P. Amba, J.-B. Thomas, and D. Alleysson, "N-LMMSE demosaicing for spectral filter arrays," *Journal of Imaging Science and Technology*, vol. 61, no. 4, pp. 40 407–1–40 407–11, Jul. 2017.

[4] R. Arablouei, E. Goan, S. Gensemer, and B. Kusy, "Fast and robust pushbroom hyperspectral imaging via DMD-based scanning," in *Proceedings of the SPIE Electronic Imaging Annual Symposium: Novel Optical Systems Design and Optimization XIX*, vol. 9948, San Diego, California, USA, Aug. 2016, pp. 99 480A–99 480A–11.

[5] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images," in *Proceedings of the 14th European Conference on Computer Vision (ECCV'16)*, ser. Lecture Notes in Computer Science, vol. 9911. Amsterdam, The Netherlands: Springer-Verlag, Oct. 2016, pp. 19–34. [Online]. Available: http://icvl.cs.bgu.ac.il/hyperspectral

[6] F. Bianconi, R. Bello-Cerezo, and P. Napoletano, "Improved opponent color local binary patterns: an effective local image descriptor for color texture classification," *Journal of Electronic Imaging*, vol. 27, no. 1, p. 011002, Dec. 2017.

[7] D. H. Brainard, "Hyperspectral image data," 1998. [Online]. Available: http://color.psych.upenn.edu/hyperspectral

[8] J. Brauers and T. Aach, "A color filter array based multispectral camera," in *12. Workshop Farbbildverarbeitung*, Illmenau, Germany, Oct. 2006, pp. 55–64.

[9] G. J. Brelstaff, A. Parraga, T. Troscianko, and D. Carr, "Hyperspectral camera system: acquisition and analysis," in *Proc.SPIE. Geog. Inf. Sys. Photogram. and Geolog./Geophys. Remote Sensing*, vol. 2587, 1995, pp. 2587 – 2587 – 10. [Online]. Available: http://www.cvc.uab.es/color_calibration/Bristol_Hyper/

[10] X. Cao, T. Yue, X. Lin, S. Lin, X. Yuan, Q. Dai, L. Carin, and D. J. Brady, "Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world," *IEEE Signal Processing Magazine*, vol. 33, no. 5, pp. 95–108, Sep. 2016.

[11] E. Carter, Y. Ohno, M. Pointer, A. Robertson, R. Seve, J. Schanda, and K. Witt., "CIE 15: Technical report: Colorimetry, 3rd edition," International Commission on Illumination, Tech. Rep., 2004.

[12] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'11)*, Colorado Springs, USA, Jun. 2011, pp. 193–200. [Online]. Available: http://vision.seas.harvard.edu/hyperspec

[13] L. Chang and Y.-P. Tan, "Hybrid color filter array demosaicking for effective artifact suppression," *Journal of Electronic Imaging*, vol. 15, no. 1, pp. 013 003,1–17, Jan. 2006.

[14] M. Chini, A. Chiancone, and S. Stramondo, "Scale object selection (SOS) through a hierarchical segmentation by a multi-spectral per-pixel classification," *Pattern Recognition Letters*, vol. 49, pp. 214–223, 2014.

[15] C. Cusano, P. Napoletano, and R. Schettini, "Combining local binary patterns and local color contrast for texture classification under varying illumination," *Journal of the Optical Society of America A*, vol. 31, no. 7, pp. 1453–1461, Jul. 2014.

[16] F. Deger, A. Mansouri, M. Pedersen, J. Y. Hardeberg, and Y. Voisin, "A sensor-data-based denoising framework for hyperspectral images," *Journal of the Optical Society of America A*, vol. 23, no. 3, pp. 1938–1950, Feb. 2015.

[17] C. W. Dirk, M. F. Delgado, M. Olguin, and J. Druzik, "A prism–grating–prism spectral imaging approach," *Studies in Conservation*, vol. 54, no. 2, pp. 77–89, 2009.

[18] S. R. Dubey, S. K. Singh, and R. K. Singh, "Multichannel decoded local binary patterns for content-based image retrieval," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4018–4032, Sep. 2016.

[19] J. Eckhard, T. Eckhard, E. M. Valero, J. L. Nieves, and E. G. Contreras, "Outdoor scene reflectance measurements using a Bragg-grating-based hyperspectral imager," *Applied Optics*, vol. 54, no. 13, pp. D15–D24, May 2015. [Online]. Available: http://colorimaginglab.ugr.es/pages/Data#__doku_ugr_hyperspectral_image_database

[20] M. Elias and P. Cotte, "Multispectral camera and radiative transfer equation used to depict Leonardo's sfumato in Mona Lisa," *Applied Optics*, vol. 47, no. 12, pp. 2146–2154, Apr. 2008.

[21] C. Fernandez-Maloigne, L. Macaire, and F. Robert-Inacio, *Couleur numérique : Acquisition, perception, codage et rendu*, ser. Signal et Image. Lavoisier. Hermès, Jun. 2012.

[22] D. H. Foster, K. Amano, S. M. C. Nascimento, and M. J. Foster, "Frequency of metamerism in natural scenes," *Journal of the Optical Society of America A*, vol. 23, no. 10, pp. 2359–2372, Oct. 2006. [Online]. Available: http://personalpages.manchester.ac.uk/staff/d.h.foster/Hyperspectral_images_of_natural_scenes_04.html

[23] D. H. Foster, K. Amano, and S. M. C. Nascimento, "Time-lapse ratios of cone excitations in natural scenes," *Vision Research*, vol. 120, pp. 45–60, 2016. [Online]. Available: http://personalpages.manchester.ac.uk/staff/d.h.foster/Time-Lapse_HSIs/Time-Lapse_HSIs_2015.html

[24] L. Gao and L. V. Wang, "A review of snapshot multidimensional optical imaging: Measuring photon tags in parallel," *Physics Reports*, vol. 616, pp. 1–37, Feb. 2016.

[25] S. Gao and V. Gruev, "Bilinear and bicubic interpolation methods for division of focal plane polarimeters," *Optics express*, vol. 19, no. 27, pp. 26 161–26 173, Dec 2011.

[26] N. Gat, "Imaging spectroscopy using tunable filters: A review," in *Proceedings of the SPIE*, vol. 4056, San Jose, CA, USA, Jan. 2000, pp. 50–64.

[27] B. Geelen, N. Tack, and A. Lambrechts, "A compact snapshot multispectral imager with a monolithically integrated per-pixel filter mosaic," in *Proceedings of the SPIE: Advanced Fabrication Technologies for Micro/Nano Optics and Photonics VII*, vol. 8974, San Francisco, California, USA, Feb. 2014, pp. 89 740L–89 740L–8.

[28] W. Guifang, M. Hai, and P. Xin, "Identification of varieties of natural textile fiber based on vis/NIR spectroscopy technology," in *Proceedings of the 1st Conference on Advanced Information Technology, Electronic and Automation Control (IAEAC 2015)*, Chongqing, China, Dec. 2015, pp. 585–589.

[29] B. K. Gunturk, Y. Altunbasak, and R. M. Mersereau, "Color plane interpolation using alternating projections," *IEEE Transactions on Image Processing*, vol. 11, no. 9, pp. 997–1013, Sep. 2002.

[30] R. Gupta and R. I. Hartley, "Linear pushbroom cameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 9, pp. 963–975, Sep. 1997.

[31] X. Hadoux, N. Gorretta, J.-M. Roger, R. Bendoula, and G. Rabatel, "Comparison of the efficacy of spectral pre-treatments for wheat and weed discrimination in outdoor conditions," *Computers and Electronics in Agriculture*, vol. 108, pp. 242–249, Oct. 2014.

[32] N. A. Hagen and M. W. Kudenov, "Review of snapshot spectral imaging technologies," *Optical Engineering*, vol. 52, pp. 52 – 52 – 23, 2013.

[33] J. Han, G. Zhang, and X. Liu, "Taylor series-based generic demosaicking algorithm for multispectral image," in *Proceedings of the SPIE Conference on Applied Optics and Photonics China (AOPC 2017)*, vol. 10462, Beijing, China, Jun. 2017, p. 1046237.

[34] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proceedings of the 11th European Conference on Computer Vision (ECCV 2010)*, vol. 6311, Heraklion, Crete, Greece, Sep. 2010, pp. 1–14.

[35] J. Hershey and Z. Zhang, "Multispectral digital camera employing both visible light and non-visible light sensing on a single image sensor," United States Patent7, Tech. Rep., Dec. 2008.

[36] K. Hirakawa and T. W. Parks, "Joint demosaicing and denoising," *IEEE Transactions on Image Processing*, vol. 15, no. 8, pp. 2146–2157, Aug 2006.

[37] T. Hirvonen, J. Orava, N. Penttinen, K. Luostarinen, M. Hauta-Kasari, M. Sorjonen, and K.-E. Peiponen, "Spectral image database for observing the quality of nordic sawn timbers," *Wood Science and Technology*, vol. 48, no. 5, pp. 995–1003, 2014. [Online]. Available: http://www.uef.fi/web/spectral/spectral-image-database-of-nordic-sawn-timbers

[38] S. Hordley, G. Finalyson, and P. Morovic, "A multi-spectral image database and its application to image rendering across illumination," in *Proceeding of the 3rd International Conference on Image and Graphics (ICIG'04)*, Hong Kong, China, Dec. 2004, pp. 394–397. [Online]. Available: http://www2.cmp.uea.ac.uk/Research/compvis/MultiSpectralDB.htm

[39] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proceedings of the 20th IEEE International Conference on Pattern Recognition (ICPR 2010)*, Istambul, Turkey, Aug. 2010, pp. 2366–2369.

[40] B. H. Horgan, E. A. Cloutis, P. Mann, and J. F. Bell, "Near-infrared spectra of ferrous mineral mixtures and methods for their identification in planetary surface spectra," *Icarus*, vol. 234, pp. 132 – 154, 2014.

[41] E. I. Im, "A Note On Derivation of the Least Squares Estimator," University of Hawaii at Manoa, Department of Economics, Working Papers 199611, 1996.

[42] F. H. Imai, M. R. Rosen, and R. S. Berns, "Multi-spectral imaging of a van Gogh's self-portrait at the national gallery of art," in *Proceedings of the IS&T's Image Processing, Image Quality, Image Capture Systems Conference (PICS 2001)*, Montreal, Quebec, Canada, Apr. 2001, pp. 185–189.

[43] S. Jaiswal, L. Fang, V. Jakhetiya, J. Pang, K. Mueller, and O. C. Au, "Adaptive multispectral demosaicking based on frequency domain analysis of spectral correlation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 953–968, Feb. 2017.

[44] J. Jia, K. J. Barnard, and K. Hirakawa, "Fourier spectral filter array for optimal multispectral imaging," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1530–1543, Apr. 2016.

[45] H. A. Khan, J.-B. Thomas, J. Y. Hardeberg, and O. Laligant, "Illuminant estimation in multispectral imaging," *Journal of the Optical Society of America A*, vol. 34, no. 7, pp. 1085–1098, Jul. 2017.

[46] H. A. Khan, **S. Mihoubi**, B. Mathon, J.-B. Thomas, and J. Y. Hardeberg, "Hytexila: High resolution visible and near infrared hyperspectral texture images," *Sensors*, vol. 18, no. 7, p. 2045, Jun. 2018. [Online]. Available: http://color.univ-lille.fr/datasets/hytexila

[47] D. Kiku, Y. Monno, S. Kikuchi, M. Tanaka, and M. Okutomi, "Residual interpolation for color image demosaicking." in *Proceedings of the IEEE International Conference on Image Processing (ICIP'13)*, Melbourne, Australia, Sep. 2013, pp. 2304–2308.

[48] D. Kiku, Y. Monno, M. Tanaka, and M. Okutomi, "Simultaneous capturing of RGB and additional band images using hybrid color filter array," in *Proceedings of the SPIE: Advanced Fabrication Technologies for Micro/Nano Optics and Photonics VII*, vol. 9023, San Francisco, California, USA, Feb. 2014.

[49] D. Kiku, Y. Monno, M. Tanaka, and M. Okutomi, "Beyond color difference: Residual interpolation for color image demosaicking," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1288–1300, Mar. 2016.

[50] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," in *ACM SIGGRAPH 2007 Papers*, ser. SIGGRAPH '07. New York, NY, USA: ACM, 2007.

[51] S. J. Koppal, *Lambertian Reflectance*. Boston, MA: Springer US, 2014, encyclopedia of Computer Vision: A Reference guide.

[52] P.-J. Lapray, X. Wang, J.-B. Thomas, and P. Gouton, "Multispectral filter arrays: Recent advances and practical implementation," *Sensors*, vol. 14, no. 11, pp. 21 626–21 659, Nov. 2014.

[53] P.-J. Lapray, J.-B. Thomas, and P. Gouton, "A database of spectral filter array images that combine visible and NIR," in *Proceedings of the 6th Computational Color Imaging Workshop (CCIW 2017)*. Milano, Italy: Springer International Publishing, Mar. 2017, pp. 187–196. [Online]. Available: http://chic.u-bourgogne.fr

[54] P.-J. Lapray, J.-B. Thomas, P. Gouton, and Y. Ruichek, "Energy balance in spectral filter array camera design," *Journal of the European Optical Society-Rapid Publications*, vol. 13, no. 1, p. 1, Jan. 2017.

[55] S. Le Moan, S. T. George, M. Pedersen, J. Blahová, and J. Y. Hardeberg, "A database for spectral image quality," in *Proceedings of the SPIE-IS&T Electronic Imaging: Image Quality and System Performance XII*, vol. 9396, San Francisco, California, USA, Feb. 2015, p. 93960P. [Online]. Available: https://www.ntnu.edu/web/colourlab/software

[56] S. H. Lee, J. Y. Choi, Y. M. Ro, and K. N. Plataniotis, "Local color vector binary patterns from multichannel face images for face recognition," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2347–2353, 2012.

[57] X. Li, B. K. Gunturk, and L. Zhang, "Image demosaicing: A systematic survey," in *Proceedings of the SPIE Conference on Visual Communications and Image Processing (VCIP'08)*, vol. 6822, San Jose, California, USA, Jan. 2008, pp. 68 221J1–15.

[58] Y. Li, A. Majumder, H. Zhang, and M. Gopi, "Optimized multi-spectral filter array based imaging of natural scenes," *Sensors*, vol. 18, no. 4, p. 1172, 2018.

[59] N.-X. Lian, L. Chang, V. Zagorodnov, and Y.-P. Tan, "Reversing demosaicking and compression in color filter array image processing: Performance analysis and modeling," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3261–3278, Nov. 2006.

[60] H. Liang, R. Lange, H. Howard, and J. Spooner, "Non-invasive investigations of a wall painting using optical coherence tomography and hyperspectral imaging," in *Proceedings of the SPIE Electronic Imaging Annual Symposium (SPIE'11): Digital Photography VII*, vol. 8084, San Francisco Airport, California, USA, Jan. 2011, pp. 8084 – 8084 – 7.

[61] C. Liu, W. Liu, X. Lu, F. Ma, W. Chen, J. Yang, and L. Zheng, "Application of multispectral imaging to determine quality attributes and ripeness stage in strawberry fruit," *PLoS ONE*, vol. 9, no. 2, pp. 1–8, Feb. 2014.

[62] O. Losson and L. Macaire, "CFA local binary patterns for fast illuminant-invariant color texture classification," *Journal of Real-Time Image Processing*, vol. 10, no. 2, pp. 387–401, Jun. 2015.

[63] O. Losson, L. Macaire, and Y. Yang, "Comparison of color demosaicing methods," *Advances in Imaging and Electron Physics*, vol. 162, pp. 173–265, Jul. 2010.

[64] G. Lu and B. Fei, "Medical hyperspectral imaging: a review," *Journal of Biomedical Optics*, vol. 19, no. 1, p. 010901, Jan. 2014.

[65] Y. M. Lu, C. Fredembach, M. Vetterli, and S. Süsstrunk, "Designing color filter arrays for the joint capture of visible and near-infrared images," in *Proceedings of the IEEE International Conference on Image Processing (ICIP'09)*, Cairo, Egypt, Nov. 2009, pp. 3797–3800.

[66] T. Mäenpää, M. Pietikäinen, and J. Viertola, "Separating color and pattern information for color texture discrimination," in *Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02)*, vol. 1, Québec, Canada, Aug. 2002, pp. 668–671.

[67] O. D. Matchett, R. I. Billmers, E. J. Billmers, and M. E. Ludwigo, "Volume holographic beam splitter for hyperspectral imaging applications," in *Procs. 19th IST/SPIE Electronic Imaging Annual Symposium (SPIE'07)*, vol. 6668, San Jose, California, USA, Jan. 2007, pp. 6668–6668–8.

[68] C. McCamy, H. Marcus, and J. Davidson, "A color-rendition chart," *Journal of Applied Photographic Engineering*, vol. 2, Jun. 1976.

[69] L. Miao and H. Qi, "The design and evaluation of a generic method for generating mosaicked multispectral filter arrays," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2780–2791, Sept 2006.

[70] L. Miao, H. Qi, R. Ramanath, and W. Snyder, "Binary tree-based generic demosaicking algorithm for multispectral filter arrays," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3550–3558, Nov. 2006.

[71] **S**. **Mihoubi**, O. Losson, B. Mathon, and L. Macaire, "Multispectral demosaicing using intensity-based spectral correlation," in *Proceedings of the 5th International Conference on Image Processing Theory, Tools and Applications (IPTA'15)*, Orléans, France, Nov. 2015, pp. 461–466.

[72] **S**. **Mihoubi**, O. Losson, B. Mathon, and L. Macaire, "Multispectral demosaicing using intensity in edge-sensing and iterative difference-based methods," in *Proceedings of the 12th International Conference on Signal-Image Technology & Internet-based Systems (SITIS'16)*, Naples, Italy, Nov. 2016, pp. 805–810.

[73] **S**. **Mihoubi**, O. Losson, B. Mathon, and L. Macaire, "Multispectral demosaicing using pseudo-panchromatic image," *IEEE Transactions on Computational Imaging*, vol. 3, no. 4, pp. 982–995, Dec. 2017.

[74] **S**. **Mihoubi**, B. Mathon, J.-B. Thomas, O. Losson, and L. Macaire, "Illumination-robust multispectral demosaicing," in *Proceedings of the 7th International Conference on Image Processing Theory, Tools and Applications (IPTA'17)*, Montreal, Canada, Nov. 2017, pp. 1–6.

[75] **S**. **Mihoubi**, O. Losson, B. Mathon, and L. Macaire, "Spatio-spectral binary patterns based on multispectral filter arrays for texture classification," *Journal of the Optical Society of America A*, vol. 35, no. 9, pp. 1532–1542, Sep 2018.

[76] A. Mirhashemi, "Introducing spectral moment features in analyzing the SpecTex hyperspectral texture database," *Machine Vision and Applications*, vol. 29, no. 3, pp. 415–432, Apr. 2018. [Online]. Available: http://www.uef.fi/web/spectral/spectex

[77] J. Mizutani, S. Ogawa, K. Shinoda, M. Hasegawa, and S. Kato, "Multispectral demosaicking algorithm based on inter-channel correlation," in *Proceedings of the IEEE Visual Communications and Image Processing Conference (VCIP 2014)*. Valletta, Malta: IEEE, Dec. 2014, pp. 474–477.

[78] Y. Monno, M. Tanaka, and M. Okutomi, "Multispectral demosaicking using adaptive kernel upsampling," in *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP'11)*, Brussels, Belgium, Sep. 2011, pp. 3157–3160.

[79] Y. Monno, M. Tanaka, and M. Okutomi, "Multispectral demosaicking using guided filter," in *Proceedings of the SPIE Electronic Imaging Annual Symposium (SPIE'12): Digital Photography VIII*, vol. 8299, Burlingame, California, USA, Jan. 2012, pp. 82 990O–82 990O–7.

[80] Y. Monno, D. Kiku, S. Kikuchi, M. Tanaka, and M. Okutomi, "Multispectral demosaicking with novel guide image generation and residual interpolation," in *Proceedings of the 21st IEEE International Conference on Image Processing (ICIP 2014)*, Paris, France, Oct. 2014, pp. 645–649.

[81] Y. Monno, S. Kikuchi, M. Tanaka, , and M. Okutomi, "A practical one-shot multispectral imaging system using a single image sensor," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 3048–3059, Oct. 2015. [Online]. Available: http://www.ok.ctrl.titech.ac.jp/res/MSI/TIP-MSI.html

[82] S. M. C. Nascimento, F. P. Ferreira, and D. H. Foster, "Statistics of spatial cone-excitation ratios in natural scenes," *Journal of the Optical Society of America A*, vol. 19, no. 8, pp. 1484–1490, Aug. 2002. [Online]. Available: http://personalpages.manchester.ac.uk/staff/d.h.foster/Hyperspectral_images_of_natural_scenes_02.html

[83] S. M. C. Nascimento, K. Amano, and D. H. Foster, "Spatial distributions of local illumination color in natural scenes," *Vision Research*, vol. 120, pp. 39–44, Mar. 2016. [Online]. Available: http://online.uminho.pt/pessoas/smcn/hsi_spatial/HSI_illumination_2015

[84] R. M. H. Nguyen, D. K. Prasad, and M. S. Brown, "Training-based spectral reconstruction from a single RGB image," in *Proceedings of the 13th European Conference on Computer Vision (ECCV'14)*. Zürich, Switzerland: Springer-Verlag, Sep. 2014, pp. 186–201.

[85] K. Ohsawa, T. Ajito, Y. Komiya, H. Fukuda, H. Haneishi, M. Yamaguchi, and N. Ohyama, "Six band HDTV camera system for spectrum-based color reproduction," *Journal of Imaging Science and Technology*, vol. 48, pp. 85–92, Mar. 2004.

[86] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, Aug. 2002.

[87] S. Ono and I. Yamada, "Decorrelated vectorial total variation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014)*, Columbus, Ohio, USA, Jun. 2014, pp. 4090–4097.

[88] C. Palm, "Color texture classification by integrative co-occurrence matrices," *Pattern Recognition*, vol. 37, no. 5, pp. 965–976, May 2004.

[89] M. Parmar, S. Lansel, and J. Farrell, "An LED-based lighting system for acquiring multispectral scenes," in *Proceedings of the SPIE Electronic Imaging Annual Symposium (SPIE'12): Digital Photography VIII*, vol. 8299, Burlingame, California, USA, Jan. 2012, p. 82990P.

[90] J. Pichette, A. Laurence, L. Angulo, F. Lesage, A. Bouthillier, D. Nguyen, and F. Leblond, "Intraoperative video-rate hemodynamic response assessment in human cortex using snapshot hyperspectral optical imaging," *Neurophotonics*, vol. 3, no. 4, p. 045003, Oct. 2016.

[91] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, *Computer vision using local binary patterns*, ser. Computational Imaging and Vision. Springer-Verlag London, 2011, vol. 40.

[92] D. K. Prasad and L. Wenhe, "Metrics and statistics of frequency of occurrence of metamerism in consumer cameras for natural scenes," *Journal of the Optical Society of America A*, vol. 32, no. 7, pp. 1390–1402, Jul. 2015. [Online]. Available: https://sites.google.com/site/hyperspectralcolorimaging/dataset

[93] J. Qin, K. Chao, M. S. Kim, R. Lu, and T. F. Burks, "Hyperspectral and multispectral imaging for evaluating food safety and quality," *Journal of Food Engineering*, vol. 118, no. 2, pp. 157–171, Sep. 2013.

[94] M. Rafinazari and E. Dubois, "Demosaicking algorithms for RGBW color filter arrays," in *Procs. 2016 IS&T International Symposium on Electronic Imaging: Color Imaging XXI*, San Francisco, California, USA, Feb. 2016, pp. 1–6.

[95] W. S. Rasband, "ImageJ," U. S. National Institutes of Health, Bethesda, Maryland, USA, 1997-2011.

[96] C. Rogass, C. Mielke, D. Scheffler, N. K. Boesche, A. Lausch, C. Lubitz, M. Brell, D. Spengler, A. Eisele, K. Segl, and L. Guanter, "Reduction of uncorrelated striping noise—applications for hyperspectral pushbroom acquisitions," *Remote Sensing*, vol. 6, no. 11, pp. 11 082–11 106, 2014.

[97] Z. Sadeghipoor, Y. M. Lu, and S. Süsstrunk, "Correlation-based joint acquisition and demosaicing of visible and near-infrared images," in *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP'11)*, Brussels, Belgium, Sep. 2011, pp. 3165–3168.

[98] J. Schanda, *Colorimetry: Understanding the CIE System*. WILEY, Jul. 2007.

[99] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "NIH Image to ImageJ: 25 years of image analysis," *Nature Methods*, vol. 9, pp. 671–675, 2012.

[100] S. Serranti, A. Gargiulo, and G. Bonifazi, "Classification of polyolefins from building and construction waste using NIR hyperspectral imaging system," *Resources, Conservation and Recycling*, vol. 61, pp. 52–58, Apr. 2012.

[101] C. Shannon and W. Weaver, *The mathematical theory of communication"*. University of Illinois Press, 1949.

[102] G. Sharma, W. Wu, and E. Dalal, "The CIEDE2000 color-difference formula:implementation notes, supplementary test data, and mathematical observations," *Color Research & Application*, vol. 30, no. 1, pp. 21–30, Feb. 2005.

[103] H. Shin, N. H. Reyes, A. L. Barczak, and C. S. Chan, "Colour object classification using the fusion of visible and near-infrared spectra," in *PRICAI 2010: Trends in Artificial Intelligence*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 498–509.

[104] K. Shinoda, T. Hamasaki, M. Hasegawa, S. Kato, and A. Ortega, "Quality metric for filter arrangement in a multispectral filter array," in *Proceedings of the 30th Picture Coding Symposium (PCS 2013)*, San Jose, CA, USA, Dec. 2013, pp. 149–152.

[105] K. Shinoda, S. Ogawa, Y. Yanagi, M. Hasegawa, S. Kato, M. Ishikaway, H. Komagatay, and N. Kobayashi, "Multispectral filter array and demosaicking for pathological images," in *Proceedings of APSIPA Annual Summit and Conference 2015*, Hong Kong, China, Dec. 2015, pp. 697–703.

[106] K. Shinoda, T. Hamasaki, M. Kawase, M. Hasegawa, and S. Kato, "Demosaicking for multispectral images based on vectorial total variation," *Optical Review*, vol. 23, no. 4, pp. 559–570, 2016.

[107] K. Shinoda, Y. Yanagi, Y. Hayasaki, and M. Hasegawa, "Multispectral filter array design without training images," *Optical Review*, pp. 1–18, 2017.

[108] K. Shinoda, Y. Ohtera, and M. Hasegawa, "Snapshot multispectral polarization imaging using a photonic crystal filter array," *Journal of the Optical Society of America A*, vol. 26, no. 12, pp. 15 948–15 961, Jun 2018.

[109] R. Shrestha, J. Y. Hardeberg, and R. Khan, "Spatial arrangement of color filter array for multispectral image acquisition," in *Proceedings of the SPIE Electronic Imaging Annual Symposium (SPIE'11): Digital Photography VII*, vol. 7875, San Francisco Airport, California, USA, Jan. 2011.

[110] T. Skauli and J. Farrell, "A collection of hyperspectral images for imaging systems research," in *Proceedings of the SPIE Electronic Imaging Annual Symposium (SPIE'13): Digital Photography IX*, vol. 8660, Burlingame, California, USA, Feb. 2013, pp. 86 600C–86 600C–7. [Online]. Available: http://www.imageval.com/scene-database/

[111] H. Steiner, O. Schwaneberg, and N. Jung, "Advances in active near-infrared sensor systems for material classification and skin detection for safety applications," *Safety Science Monitor*, vol. 17, no. 1, 2013.

[112] M. Stokes, M. Anderson, S. Chandrasekar, and R. Motta, "Multimedia systems and equipment - colour measurement and management - part 2-1: Colour management - default RGB colour space - sRGB," *International Electrotechnical Commission*, Nov. 1996.

[113] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11–32, Nov. 1991.

[114] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Transactions on Image Processing*, vol. 16, no. 2, pp. 349–366, Feb 2007.

[115] J.-B. Thomas, P.-J. Lapray, P. Gouton, and C. Clerc, "Spectral characterization of a prototype SFA camera for joint visible and NIR acquisition," *Sensors*, vol. 16, no. 7, p. 993, Jun. 2016.

[116] A. Verma, D. Tyagi, and S. Sharma, "Recent advancement of LBP techniques: A survey," in *2016 International Conference on Computing, Communication and Automation (ICCCA)*, April 2016, pp. 1059–1064.

[117] C. Wang, X. Wang, and J. Y. Hardeberg, "A linear interpolation algorithm for spectral filter array demosaicking," in *Proceedings of the 6th EURASIP International Conference on Image and Signal Processing (ICISP'2014)*, ser. Lecture Notes in Computer Science, vol. 8509. Springer-Verlag Berlin, Jun. 2014, pp. 151–160.

[118] J. Wang, C. Zhang, and P. Hao, "New color filter arrays of high light sensitivity and high demosaicking performance," in *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP'11)*, Brussels, Belgium, Sep. 2011, pp. 3153–3156.

[119] X. Wang, J.-B. Thomas, J. Y. Hardeberg, and P. Gouton, "Median filtering in multispectral filter array demosaicking," in *Proceedings of the SPIE Electronic Imaging Annual Symposium (SPIE'13): Digital Photography IX*, vol. 8660, Burlingame, California, USA, Feb. 2013, pp. 86 600E–86 600E–10.

[120] X. Wang, J.-B. Thomas, J. Y. Hardeberg, and P. Gouton, "Discrete wavelet transform based multispectral filter array demosaicking," in *Proceedings of the 2013 Colour and Visual Computing Symposium (CVCS 2013)*. Gjøvik, Norway: IEEE, Sep. 2013, pp. 1–6.

[121] C. Winkens, V. Kobelt, and D. Paulus, "Robust features for snapshot hyperspectral terrain-classification," in *Procs. of the 17th International Conference on Computer Analysis of Images and Patterns (CAIP 2017)*, ser. Lecture Notes in Computer Science, vol. 10424. Ystad, Sweden: Springer, Aug. 2017, pp. 16–27.

[122] C. Winkens, F. Sattler, and D. Paulus, "Hyperspectral terrain classification for ground vehicles," in *Proceedings of the 12th International Conference on Computer Vision Theory and Applications (VISAPP 2017)*. Porto, Portugal: SciTePress, Feb. 2017, pp. 417–424. [Online]. Available: https://wp.uni-koblenz.de/hyko/sensor-setup/

[123] Y. Yanagi, K. Shinoda, M. Hasegawa, S. Kato, M. Ishikawa, H. Komagata, and N. Kobayashi, "Optimal transparent wavelength and arrangement for multispectral filter array," in *Proceedings of the IS&T International Symposium on Electronic Imaging, Image Processing: Algorithms and Systems XIV*, San Francisco, California, USA, Feb. 2016, pp. 1–5.

[124] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010. [Online]. Available: http://www.cs.columbia.edu/CAVE/databases/multispectral/

[125] A. Zacharopoulos, K. Hatzigiannakis, P. Karamaoynas, V. M. Papadakis, M. Andrianakis, K. Melessanaki, and X. Zabulis, "A method for the registration of spectral images of paintings and its evaluation," *Journal of Cultural Heritage*, vol. 29, pp. 10–18, Jan. 2018.

[126] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'10)*, San Francisco, CA, USA, Jun. 2010, pp. 2528–2535.