HAL
open science

Multispectral demosaicing using pseudo-panchromatic image

Sofiane Mihoubi, Olivier Losson, Benjamin Mathon, Ludovic Macaire

To cite this version:

Sofiane Mihoubi, Olivier Losson, Benjamin Mathon, Ludovic Macaire. Multispectral demosaicing using pseudo-panchromatic image. IEEE Transactions on Computational Imaging, 2017, 3 (4), pp.982-995. 10.1109/TCI.2017.2691553 . hal-01507480

HAL Id: hal-01507480

https://hal.science/hal-01507480v1

Submitted on 25 Jan 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multispectral demosaicing using pseudo-panchromatic image

Sofiane Mihoubi, Olivier Losson, Benjamin Mathon, and Ludovic Macaire

Abstract

Single-sensor color cameras, which classically use a color filter array (CFA) to sample RGB channels, have recently been extended to the multispectral domain. To sample more than three wavelength bands, such systems use a multispectral filter array (MSFA) that provides a raw image in which a single channel value is available at each pixel. A demosaicing procedure is then needed to estimate a fully-defined multispectral image. In this paper, we review multispectral demosaicing methods and propose a new one based on the pseudo-panchromatic image (PPI). Pixel values in the PPI are computed as the average spectral values. Experimental results show that our method provides estimated images of better quality than classical ones.

Index Terms

Multispectral image demosaicing, Multispectral filter array, Spectral correlation, Spatial correlation, Illumination, Pseudo-panchromatic image.

I. Introduction

Digital color cameras are usually only sensitive to three wide and overlapping bands of the visible electromagnetic spectrum. To overcome this limitation in object reflectance representation, multispectral cameras have recently emerged to provide many channels associated with narrow (spectral) bands. Multispectral images, that are composed of up to several dozens of channels, are analyzed in various application fields, such as medical imaging [15, 24], art studies [8]. Furthermore, Qin et al. [25] 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.

Multispectral images can be captured by sequentially selecting a different optical filter for each band [4]. However, this technology is not suitable to acquire dynamic scenes because switching among filters is time-consuming. To increase the acquisition rate, one snapshot solution is to fit the camera with a single photo-sensitive sensor covered

This work was supported in part by the ANR-11-EQPX-23 IrDIVE platform.

Sofiane Mihoubi, Olivier Losson, Benjamin Mathon, and Ludovic Macaire are with Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL - Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France (email: sofiane.mihoubi@ed.univ-lille1.fr, olivier.losson@univ-lille.fr, benjamin.mathon@univ-lille.fr, ludovic.macaire@univ-lille.fr).

by a multispectral filter array (MSFA) that is a mosaic of spectrally selective filters [7, 13]. Over each photo-receptor of the sensor lies a narrow-band filter that defines the channel associated with the resulting pixel. Finally, the camera delivers a raw image in which one single channel is available at each pixel according to the MSFA and the other missing channels 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 color filter array (CFA). CFA demosaicing is a well-studied problem for more than forty years [14], but MSFA demosaicing is a recent subject with new issues. The principles of spatial and spectral correlations, that exploit the properties of radiance in CFA demosaicing, should indeed be reconsidered. First, more bands imply a lower spatial sampling rate for each of them, which weakens the assumption of spatial correlation between the values of raw image pixels 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 [18].

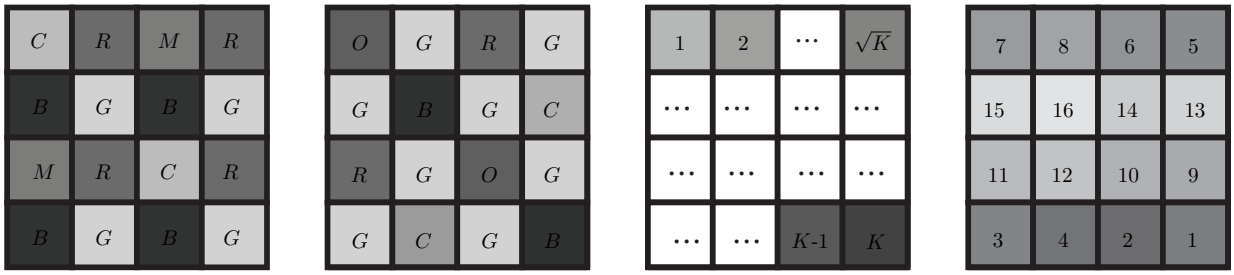
To perform multispectral demosaicing despite the weak spatial correlation, we propose to use a spatially fully-defined channel that is estimated from the raw image, namely the pseudo-panchromatic image (PPI). Chini et al. [3] define the PPI of a fully-defined multispectral image as the average image over all channels. In a preliminary work (where the PPI is called intensity image), we show that PPI-based MSFA demosaicing provides promising results [17]. We here improve this proposal and compare it with state-of-the-art MSFA demosaicing methods in extensive experiments. Moreover, as MSFA filters are sensitive to narrow bands of the visible spectrum, the non-uniformity of illumination used during image acquisition strongly affects the acquired value range (i.e., the difference between maximal and minimal acquired values) in each channel. Hence, our demosaicing procedure adjusts the raw value scale to be robust against illumination non-uniformity.

This paper is organized as follows. In Sec. II, we focus on the different MSFAs found in the literature, and on the one we consider. In Sec. III, we present the acquisition of multispectral images and compute the spatial and spectral correlations in such images. In Sec. IV, we expose the demosaicing problem, review the existing methods that can be used with our MSFA, and describe their limitations according to the multispectral image properties that they exploit. In Sec. V, we propose a new estimation of the PPI and use it for demosaicing. In Sec. VI, we experimentally assess the performances of the different methods and focus on their behavior when illumination changes.

II. MSFA-BASED MULTISPECTRAL IMAGING

A. MSFAs and associated demosaicing methods

MSFAs are composed of filters whose transmittance functions have to be carefully selected since they affect both the spectral reproduction ability and the spatial reconstruction quality [10, 20, 23]. To ensure manufacturing practicability and demosaicing feasibility, all MSFAs are defined by a basic pattern that respects a trade-off between spatial and spectral sub-samplings. Two important criteria must be considered in the MSFA design [16]: *spectral*



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

Fig. 1. Basic MSFA patterns: (a) with three dominant bands [16], (b) with one dominant band [19], (c) with K bands and no dominant one, and (d) in IMEC camera. Numbers are the band indices (see Sec. II-B).

consistency, which states that the neighborhood of all the filters associated with a band should be identical, and *spatial uniformity*, which requires that the MSFA spatially samples each band as evenly as possible. The MSFAs that respect these criteria can be generated using a binary tree as proposed by Miao et al. [16], in which each band is characterized by its probability of appearance (POA).

We classify the basic MSFA patterns and the associated demosaicing algorithms into two categories depending on the presence or not of a dominant band (i.e., with a greater POA than the other bands):

- Two MSFAs with some dominant band(s) are proposed in the literature. Miao et al. [16] design the MSFA with the basic 4×4 pattern of Fig. 1(a) that contains three dominant bands (red (R), green (G), and blue (B)) with a POA of $\frac{1}{4}$ and two other bands (cyan (C) and magenta (M)) with a POA of $\frac{1}{8}$. They propose a binary tree-based edge-sensing (BTES) demosaicing method which uses the channels associated with the dominant bands in order to estimate the values of other channels. Monno et al. [19, 21, 22] propose another 4×4 basic pattern that is inspired by the CFA pattern. It exhibits a POA of $\frac{1}{2}$ for G and of $\frac{1}{8}$ for the four other bands (R, B, C, and orange (O)) (see Fig. 1(b)). This pattern is related to several demosaicing methods: the original algorithm [19] uses an adaptive kernel up-sampling, that is improved by filtering under the guidance of the G channel [21], and further enhanced by residual interpolation [22]. Jaiswal et al. [9] also propose to exploit the redundant property of G by analyzing the spatial frequency domain.
- MSFAs with no dominant band (hereafter called non-redundant MSFAs for short), that typically have a square or rectangular basic pattern [29]. For instance, Fig. 1(c) shows a $\sqrt{K} \times \sqrt{K}$ square basic pattern composed of K filters associated with different bands whose POA is $\frac{1}{K}$. To demosaic the raw image provided by such MSFAs, Wang et al. extend the CFA demosaicing methods based on the discrete wavelet transform [31] and median filtering [30] to MSFA demosaicing. The demosaicing algorithm proposed by Brauers and Aach [1] computes the channel differences to take the spectral (inter-channel) correlation into account. Mizutani et al. [18] propose to iterate Brauers and Aach's scheme a number of times given by the distance between band centers of the considered pair of bands. Shinoda et al. [27] propose to combine the BTES method based on a spatial gradient [16] with that based on channel differences [1].

B. Our considered MSFA

As multispectral imaging privileges spectral resolution, the number of bands should be as high as possible. Although this goal conflicts with a dense spatial sub-sampling and increases the demosaicing difficulty, it leads us to retain an MSFA with no dominant band. In the following, we only consider the non-redundant MSFA formed from the square 4×4 basic pattern of Fig. 1(d) in which the $K = 16$ bands are not arranged in ascending order of the classical pixel read-out due to manufacturing constraints. This MSFA designed by IMEC [5] is embedded in the sole off-the-shelf MSFA-based systems available on the market today [24], namely XIMEA's xiSpec and IMEC's "snapshot mosaic" multispectral cameras (called IMEC camera for short in the following). The spectral narrow bands of IMEC's MSFA are centered at wavelengths $\lambda^i \in \mathcal{B} = \{469, 480, 489, 499, 513, 524, 537, 551, 552, 566, 580, 590, 602, 613, 621, 633\}$ (in nanometers), so that $\lambda^1 = 469\,nm, \ldots, \lambda^{16} = 633\,nm$.

III. Reference image simulation and properties

Let us consider the multispectral image $\mathbf{I} = \{I^i\}_{i=1}^K$ made of K fully-defined channels corresponding to the K bands. This image is called the *reference image* because it is often used as a reference to assess the demosaicing quality although it cannot be provided by single-sensor multispectral cameras. To get reference images, we simulate the acquisition process by following the simple multispectral image formation model described in section III-A. For this purpose, we consider that the radiance of various scenes from a public database (see Sec. III-B) is projected onto K (virtual) sensors, each one being associated with one of the bands sampled by IMEC camera or by an "ideal" camera with $1\,nm$-bandwidths (see Sec. III-C). This will allow us to assess the demosaicing performances (see Sec. VI) of both a real snapshot camera (IMEC camera) and of an "ideal" one with the same band centers and no inter-band overlapping. Finally, we study the statistical properties of these reference images.

A. Multispectral image formation model

Assuming ideal optics and homogeneous spectral sensitivity of the sensors, the value I_p^i of channel I^i at pixel p is expressed as:

$$I_p^i = Q\left(\int_\Omega E(\lambda) \cdot R(\lambda)_p \cdot T^i(\lambda) d\lambda\right) , \tag{1}$$

where Ω is the working spectral range. The term $E(\lambda)$ is the relative spectral power distribution of the light source which homogeneously illuminates all surface elements of the scene. The surface element observed by the pixel p reflects light with the reflectance factor $R(\lambda)_p$. The resulting radiance spectrum $(E(\lambda) \cdot R(\lambda)_p)$ is captured by the camera, filtered according to the transmission spectrum (transmittance) $T^i(\lambda)$ of the band i centered at the wavelength λ^i. The value I_p^i is finally given by quantization of the received energy according to the function Q.

B. Radiance data

To simulate the radiance data we need (i) reflectance and (ii) illumination data. (i) The CAVE database [32] contains the reflectance of surface elements in 32 scenes. The reflectance is defined on 31 bands of width $10\,nm$

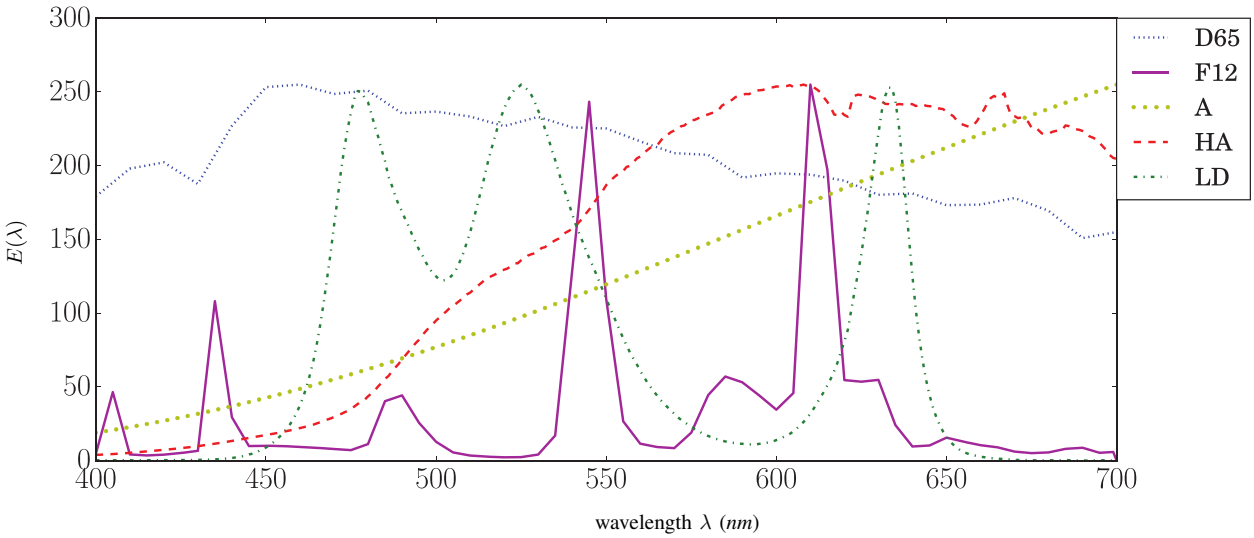


Fig. 2. Relative spectral power distributions of CIE D65, A, and F12 illuminants, and of real illuminations (acquired with the Avantes spectrometer AvaSpec-3648) provided by halogen lamps (HA) and a LED dome (LD).

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 $R(\lambda)_p$ for all integer $\lambda \in [400\,nm, 700\,nm]$ using linear interpolation of CAVE data. (ii) We consider three CIE standard illuminants (D65, A, and F12) and two real illuminations: a set of six Paulmann 2900K halogen lamps (HA) and an Advanced Illumination DL097-WHIIC LED diffuse dome (LD). Their relative power spectral distributions $E(\lambda)$ are defined for all $\lambda \in [400\,nm, 700\,nm]$ and normalized so that their maximum reaches 255 (see Fig. 2).

C. Multispectral image simulation

To simulate the reference channels that the IMEC and "ideal" cameras would have provided according to Eq. (1), we also need the transmittances $T^i(\lambda)$. These are specific to each camera:

- IMEC camera samples 16 bands with known transmittances $T^i_{IMEC}(\lambda)$ centered at $\lambda^i \in \mathcal{B}$ (see Fig. 3). By discretely summing with $d\lambda = 1$, Eq. (1) becomes:

$$I^i_p = Q\left(\sum_{\lambda=400}^{700} E(\lambda) \cdot R(\lambda)_p \cdot T^i_{IMEC}(\lambda)\right). \tag{2}$$

- To simulate the 16 channels acquired by the "ideal" camera (IC), we assume monochromatic transmittances at $\lambda^i \in \mathcal{B}$: $T^i_{IC}(\lambda) = \delta(\lambda - \lambda^i)$, where δ is the Dirac delta function. Eq. (1) then becomes:

$$I^i_p = Q\left(E(\lambda^i) \cdot R(\lambda^i)_p\right). \tag{3}$$

Note that all transmittances are normalized: for all $i = 1..K$, $\sum_{\lambda=400}^{700} T^i_{IMEC}(\lambda) = \sum_{\lambda=400}^{700} T^i_{IC}(\lambda) = 1$.

The functions $E(\lambda)$, $R(\lambda)$, and $T(\lambda)$ are coded as float variables and the Q function quantizes the energy on 8

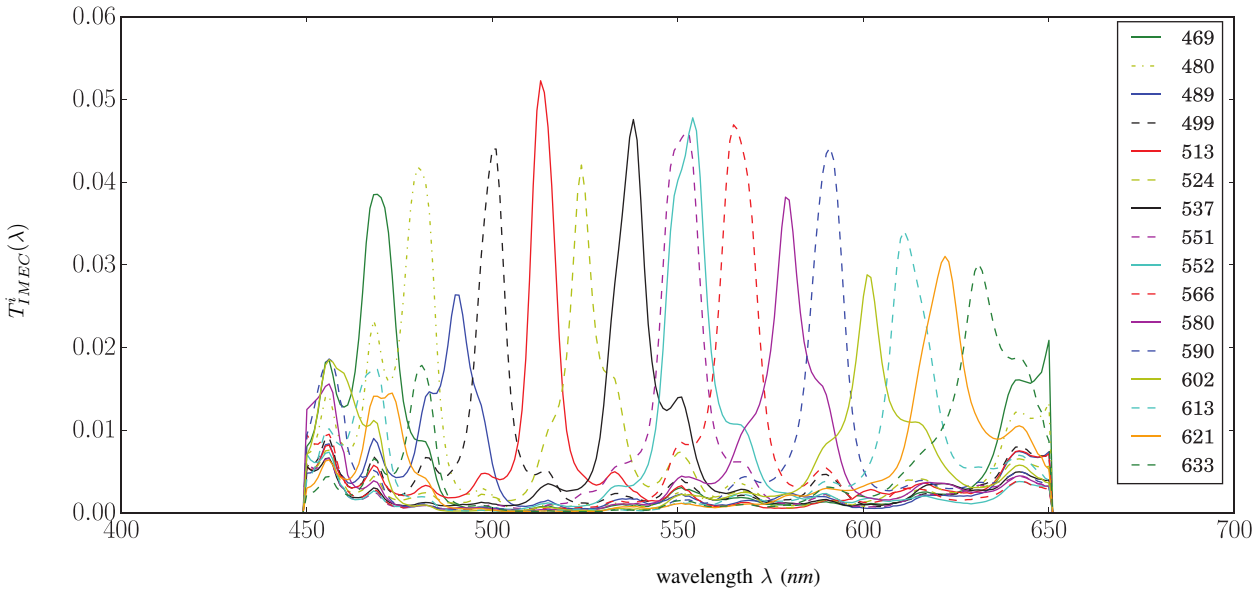


Fig. 3. Normalized transmittances of IMEC camera[1]. Caption: center wavelengths $\lambda^i \in \mathcal{B}$ in ascending order.

bits, so that $0 \leq I^i_p \leq 2^8 - 1$. Overall, this simulation provides 32 multispectral images with 16 channels of size 512×512 pixels, for each of the considered two cameras and five illuminations.

D. Properties of multispectral images

1) Spatial Properties: Like for CFA demosaicing, it would be interesting to take advantage of the reflectance properties for MSFA demosaicing. Most CFA demosaicing schemes assume that the reflectance does not change locally across object surfaces, hence that values of a color channel (R, G or B) are correlated among neighboring pixels in homogeneous areas. The sparse spatial sub-sampling 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^i_p of each pixel $p(x, y)$ and that $I^i_{p+(\delta_x,0)}$ of its right neighbor at spatial distance δ_x in a given channel I^i as [6]:

$$C[I^i](\delta x) = \frac{\sum_p \left((I^i_p - \mu^i)(I^i_{p+(\delta_x,0)} - \mu^i) \right)}{\sqrt{\sum_p (I^i_p - \mu^i)^2} \sqrt{\sum_p (I^i_{p+(\delta_x,0)} - \mu^i)^2}}, \tag{4}$$

where μ^i is the mean value of channel I^i. For a given δ_x, we compute the average correlation $\bar{C}(\delta_x)$ on the 32 IMEC and IC images simulated with the CIE D65 illuminant whose power distribution can be considered as uniform in the visible domain (see Fig. 2). The results (see Tab. I) show that 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 our considered MSFA, which makes the correlation decrease from 0.94 to 0.89.

[1]The optical device is equipped with a 450–650 *nm* band-pass filter to avoid spectral artifacts.

TABLE I

SPATIAL CORRELATION $\bar{C}(\delta_x)$ BETWEEN VALUES OF TWO NEIGHBORING PIXELS FOR DIFFERENT DISTANCES δ_x (AVERAGE OVER 16 CHANNELS OF 32 IMEC OR IC IMAGES UNDER CIE D65 ILLUMINANT).

δ_x (pixels)	0	1	2	3	4
IMEC images	1.00	0.98	0.94	0.91	0.88
IC images	1.00	0.98	0.94	0.91	0.89

2) Spectral Properties: Gunturk et al. [6] also experimentally show that color channels 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 (R, G, B) transmittances of color cameras widely overlap. On the opposite, an MSFA finely samples the visible spectrum according to K rather separated bands. We can then expect that channels associated with nearby band centers are more correlated than channels associated with distant band centers [18]. To validate this assumption, we evaluate the correlations between all pairs of channels on IMEC and IC multispectral images simulated with the CIE D65 illuminant. The Pearson correlation coefficient between any pair of channels I^i and I^j is computed as [6]:

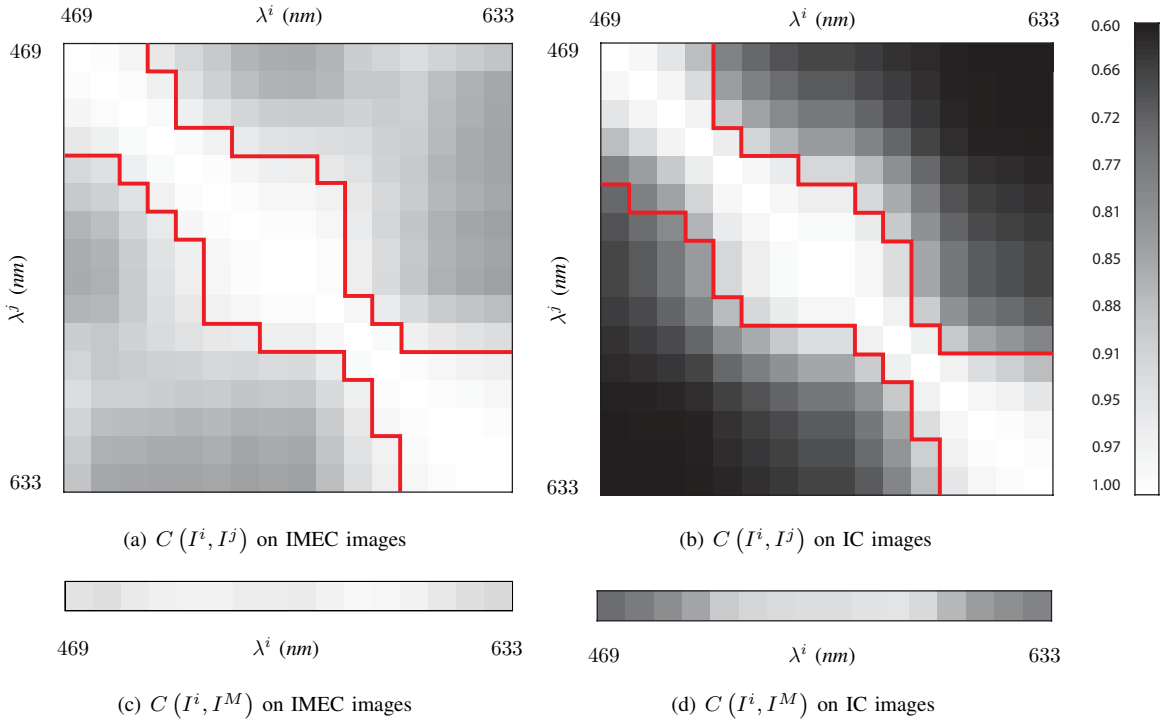
$$C\left(I^i, I^j\right) = \frac{\sum\limits_p \left((I_p^i - \mu^i)(I_p^j - \mu^j)\right)}{\sqrt{\sum\limits_p (I_p^i - \mu^i)^2}\sqrt{\sum\limits_p (I_p^j - \mu^j)^2}}. \tag{5}$$

The results (see Figs. 4(a) and 4(b)) confirm that channels associated with spectrally close band centers ($\lambda^i \approx \lambda^j$) are more correlated than channels associated with distant band centers ($\lambda^i \gg \lambda^j$ or $\lambda^i \ll \lambda^j$) for IMEC and IC images. They also show that IMEC channels are more correlated than IC ones because IMEC transmittances somewhat overlap.

3) Illumination Properties: As seen in Sect. III-A, the acquired multispectral images depend on the spectral power of illumination $E(\lambda)$. From Eq. (1) we can expect that when $E(\lambda)$ is low over wide spectral ranges (e.g., F12 in Fig. 2) and $T^i(\lambda)$ is low at other wavelengths, the value I_p^i will also be low. To highlight this dependency upon illumination, we compute the mean value μ^i of each channel for IMEC and IC images simulated with the CIE F12 illuminant. Figure 5 confirms that μ^i is low when $E(\lambda)$ is also low, showing that the illumination strongly affects the mean value of each channel. We can therefore assume that the non-uniformity of illumination spectral distribution impacts the value range in each channel.

These statistical studies of multispectral images finally yield three main properties that could be exploited 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 non-uniformity of illumination spectral distribution impacts the acquired value range in each channel.



(a) $C\left(I^i, I^j\right)$ on IMEC images

(b) $C\left(I^i, I^j\right)$ on IC images

(c) $C\left(I^i, I^M\right)$ on IMEC images

(d) $C\left(I^i, I^M\right)$ on IC images

Fig. 4. (a), (b): Correlation between channels I^i and I^j of IMEC (a) and IC (b) images under CIE D65 illuminant. (c), (d): Correlation between channel I^i and PPI I^M of IMEC (c) and IC (d) images under CIE D65 illuminant (see Sec. V-A). Values are averaged over 32 images, range between 0.63 (black) and 1.0 (white), and are displayed with cubic scale. Values of (c) and (d) are reported column-wise as red lines on (a) and (b) respectively.

IV. Demosaicing methods for non-redundant MSFAs

In this section, we first formulate the MSFA demosaicing problem and present a simple scheme of bilinear interpolation. Then we review methods that are proposed in the literature for MSFAs with no dominant band. Finally we point out the limitations of these approaches.

A. MSFA demosaicing problem formulation

A single-sensor multispectral camera fitted with a non-redundant MSFA provides a raw image I^{MSFA} with $X \times Y$ pixels. 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 (with cardinality $|S| = X \times Y$) and S^k be the pixel subset where the MSFA samples the band k, such that $|S^k| = \frac{XY}{K}$ and $S = \bigcup_{k=1}^{K} S^k$. Then, I^{MSFA} can be seen as a spectrally-sampled version of the reference fully-defined image $\mathbf{I} = \{I^i\}_{i=1}^{K}$ according to the location of p: for all $p \in S^k$, $I_p^{MSFA} = I_p^k$. At each pixel of the raw image I^{MSFA}, only one out of the K channels is available and the values of the $K-1$ others are missing. Demosaicing is then performed on I^{MSFA} 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 $\hat{I}_p^i$, $i \neq k$, is the estimated value of channel I^i at p.

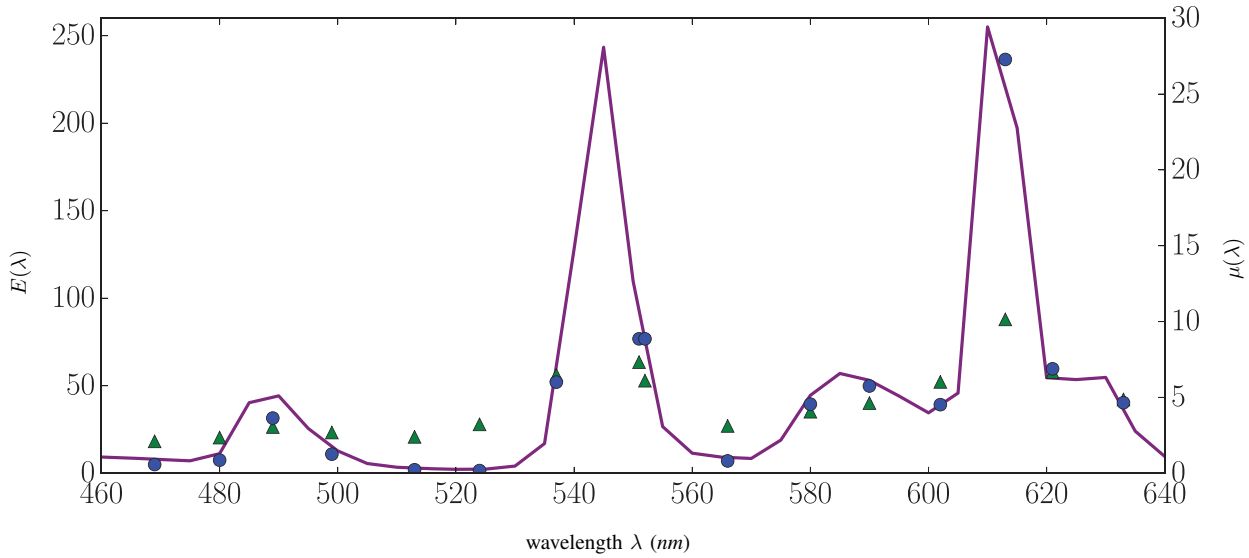


Fig. 5. Relative power spectral distribution $E(\lambda)$ of CIE F12 illuminant (solid line) and average value μ^i over 32 simulated IMEC (triangle dot) and IC (circle dot) images for each center wavelength $\lambda^i \in \mathcal{B}$.

B. 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 interpolation can be described in two steps [1]:

- First, for each channel index $i = 1..K$, it builds a sparse raw image $\tilde{I}^i$ that contains the available values at pixels in S^i and zero elsewhere (see Fig. 7(a)). This can be formulated as:

$$\tilde{I}^i = I^{MSFA} \odot m^i \ , \tag{6}$$

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

$$m_p^i = \begin{cases} 1 & \text{if } p \in S^i, \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

- Second, it estimates each channel by interpolation (see Fig. 7(b)) as:

$$\hat{I}_{WB}^i = \tilde{I}^i * H \ , \tag{8}$$

  where $*$ is the convolution operator and H is a low-pass filter. For a 4×4 non-redundant MSFA, H is defined

from the following 7×7 unnormalized filter:

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

such that the weight of each neighbor decreases as its spatial distance to the central pixel increases. The normalization of F to get H must take care of the sparse nature of $\tilde{I}^i$ 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:

$$H(a, b) = \frac{F(a, b)}{c_F(a, b)}, \tag{10}$$

where the normalization factor c_F is defined by:

$$c_F(a, b) = \sum_{\substack{k=1 \\ k \equiv a \,(\mathrm{mod}\ 4)}}^{7} \sum_{\substack{l=1 \\ l \equiv b \,(\mathrm{mod}\ 4)}}^{7} F(k, l). \tag{11}$$

The conditions here using the congruence relation $\equiv$ consider all the pixels that underlie H and belong to the same MSFA subset as the pixel under $H(a, b)$, which ensures that H is normalized channel-wise according to the 4×4 basic MSFA pattern. Figure 6 shows three (out of sixteen) cases of F (and of H) center locations for the convolution of a sparse channel $\tilde{I}^i$. The elements of F that affect the convolution result correspond to non-zero pixels of $\tilde{I}^i$ (displayed in black), and are normalized by the sum of all such elements of F that overlie the pixels of S^i. Note that for the particular filter F of Eq. (9), the normalization factor is constant: $c_F(a, b) = c = 16$.

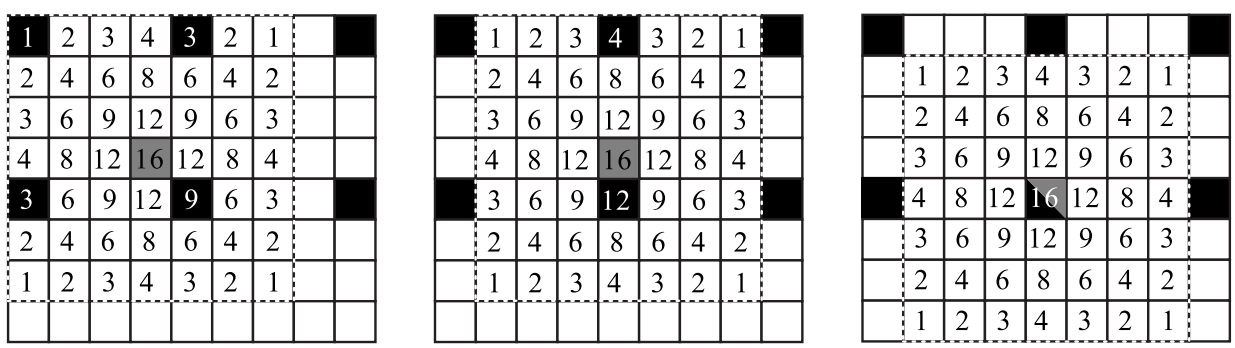
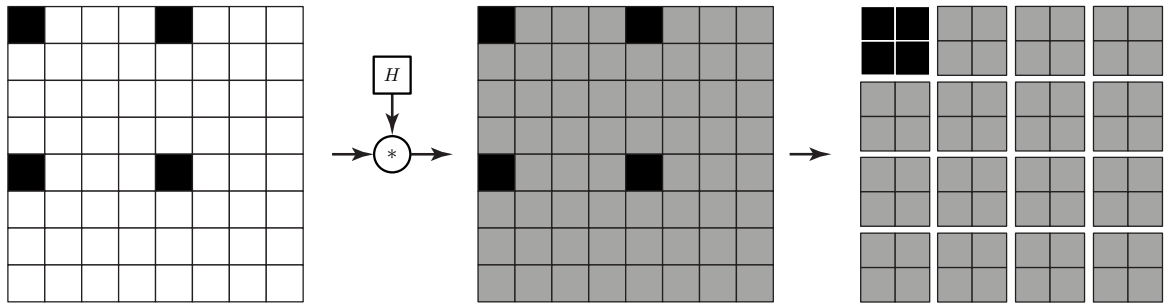


Fig. 6. Normalization of F as H for the convolution of a sparse channel (with non-zero pixels in black) on three cases of filter center locations (in gray). The support window of F (dotted bound) overlies a variable number of non-zero pixels according to its center location: four for (a), two for (b) or one (the center itself) for (c).



(a) (b) (c)

Fig. 7. DS image formation: (a) sparse channel $\tilde{I}^i$ (with non-zero pixels in black), (b) fully-defined channel $\hat{I}^i_{WB}$ estimated by WB interpolation, (c) DS images of $\hat{I}^i_{WB}$.

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.

C. Discrete wavelet transform (DWT)

Wang et al. extend the DWT-based CFA demosaicing to MSFA demosaicing [31]. This approach assumes that the low-frequency contents is well estimated by WB 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 (see Sec. IV-B), then applies five successive steps to each channel $\hat{I}^i_{WB}$:

1) It decomposes $\hat{I}^i_{WB}$ into K down-sampled (DS) images as shown in Fig. 7, so that the j-th DS image of $\hat{I}^i_{WB}$ is made of the pixels in S^j. Note that only the i-th DS image of $\hat{I}^i_{WB}$ contains MSFA (available) values.
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 i-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 4×4 MSFA.
4) It computes K transformed DS images by inverse DWT.
5) It recomposes the full-resolution channel $\hat{I}^i$ from the K transformed DS images.

D. Spectral difference (SD)

Brauers and Aach [1] 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 any non-redundant MSFA with square basic pattern. From an initial estimation $\hat{\mathbf{I}}_{WB}$ (see Sec. IV-B), it performs the following steps:

1) First, for each ordered pair (i, j) of channel indices, it computes the sparse channel difference $\tilde{\Delta}^{i,j}$ given by:

$$\tilde{\Delta}^{i,j} = \tilde{I}^i - \hat{I}^j_{WB} \odot m^i \ , \tag{12}$$

that is only non-zero at the pixels in S^i, and a fully-defined channel difference $\hat{\Delta}^{i,j} = \tilde{\Delta}^{i,j} * H$ by WB interpolation (see Eq. (8)).

2) Second, for each $i = 1..K$, it estimates $\hat{I}^i$ by adding the sparse raw image (see Eq. (6)) and channel difference $\hat{\Delta}^{i,j}$:

$$\hat{I}^i = \sum_{j=1}^{K} \left(\tilde{I}^j + \hat{\Delta}^{i,j} \odot m^j \right). \tag{13}$$

Note that a missing value at each pixel $p \in S^k$ is equivalently given by: $\hat{I}^i_p = I^k_p + \hat{\Delta}^{i,k}_p$.

E. Iterative spectral difference (ItSD)

Mizutani et al. [18] improve the SD method by iteratively updating the channel differences. The number of iterations takes the correlation between two channels I^i and I^j into account, that is strong when their associated band centers λ^i and λ^j are close (see Sec. III-D). The number of iterations $N^{i,j}$ is given by:

$$N^{i,j} = \left\lceil \exp\left(-\frac{|\lambda^j - \lambda^i| - 100}{20\sigma} \right) \right\rceil . \tag{14}$$

It decreases as the distance between λ^i and λ^j increases. For instance, setting $\sigma = 1.74$ as proposed by the authors provides $N^{i,j} = 10$ when $|\lambda^j - \lambda^i| = 20\,nm$ and $N^{i,j} = 1$ when $|\lambda^j - \lambda^i| \geq 100\,nm$.

The algorithm initially estimates all sparse channel differences $\tilde{\Delta}^{i,j}(0)$ (see Eq. (12)) and all channels $\hat{I}^i(0)$ (see Eq. (13)). At each iteration $t > 0$, it first updates the sparse channel difference:

$$\tilde{\Delta}^{i,j}(t) = \begin{cases} \tilde{I}^i - \hat{I}^j(t-1) \odot m^i & \text{if } t \leq N^{i,j}, \\ \tilde{\Delta}^{i,j}(t-1) & \text{otherwise.} \end{cases} \tag{15}$$

Then it estimates a fully-defined channel difference as: $\hat{\Delta}^{i,j}(t) = \tilde{\Delta}^{i,j}(t) * H$, and each channel as: $\hat{I}^i(t) = \sum_{j=1}^{K} \tilde{I}^j + \hat{\Delta}^{i,j}(t) \odot m^j$ (see Eqs. (8) and (13)).

F. Binary tree-based edge-sensing (BTES)

For each channel, the previously presented methods estimate the missing values simultaneously. To determine the missing values of a channel, Miao et al. [16] propose a scheme divided into four steps in the case of a 4×4 non-redundant 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. 8). Let us consider the i-th channel ($i = 1..16$) and denote as $\hat{S}^i(t)$ (displayed in gray in Fig. 8) the subset of pixels whose value of channel I^i is estimated at step t, and $\dot{S}^i(t)$ (displayed in black) the subset of pixels whose value of channel I^i is available in I^{MSFA} or has been previously estimated: $\dot{S}^i(0) = S^i$ and $\dot{S}^i(t) = \dot{S}^i(t-1) \cup \hat{S}^i(t-1)$ for $t > 0$. At step t, for each $i = 1..16$, the values of channel I^i at $p \in \hat{S}^i(t)$ are estimated as:

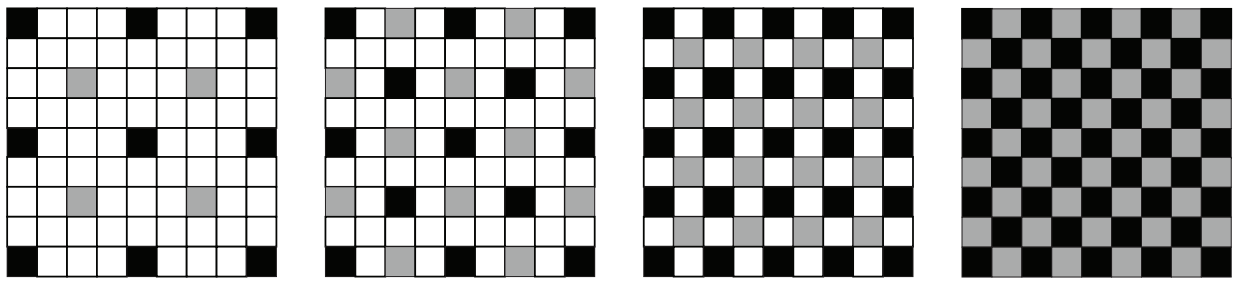
$$\hat{I}^i_p = \frac{\sum_{q \in \tilde{N}_p(t)} \alpha_q \cdot \dot{I}^i_q}{\sum_{q \in \tilde{N}_p(t)} \alpha_q} , \tag{16}$$

where $\dot{I}^i_q$ is available in I^{MSFA} or has been previously estimated, and $\tilde{N}_p(t)$ is the subset of the four closest neighbors of p that belong to $\dot{S}^i(t)$. 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. 8). The weights α_q, that embed the edge-sensing part of the algorithm, also depend on t and on the direction (horizontal, vertical, or diagonal) given by p and q. For instance, the weight α_q associated with an horizontal neighbor q is computed as:

$$\begin{aligned}\alpha_q = \Big(1 + \left|\dot{I}^i_{q+(2,0)} - \dot{I}^i_q\right| + \left|\dot{I}^i_{q-(2,0)} - \dot{I}^i_q\right| + \\ \tfrac{1}{2}\left|\dot{I}^i_{q+(-1,-1)} - \dot{I}^i_{q+(1,-1)}\right| + \tfrac{1}{2}\left|\dot{I}^i_{q+(-1,1)} - \dot{I}^i_{q+(1,1)}\right|\Big)^{-1}.\end{aligned} \tag{17}$$

In the case of a non-redundant MSFA, many values are missing at $t < 3$ to compute these weights. Miao et al. propose to set missing values to 1, which leads to an unweighted bilinear interpolation at $t = 0$ and $t = 1$.



(a) $t = 0$ (b) $t = 1$ (c) $t = 2$ (d) $t = 3$

Fig. 8. Estimation of $\hat{I}^i$ in four steps by BTES method. Pixels of $\hat{S}^i(t)$ whose values are estimated at t are displayed in gray, and those of $\dot{S}^i(t)$ whose values are known or previously estimated are displayed in black.

G. Multispectral local directional interpolation (MLDI)

Shinoda et al. [27] combine BTES and SD approaches into the MLDI method that uses four steps like BTES (see Fig. 8). Instead of marginally estimating each channel as in Eq. (16), the authors compute the difference between the i-th channel being estimated and the available one at each pixel of I^{MSFA}. Let us denote below as $k(p)$ the channel index whose value is available at p in I^{MSFA} (i.e., $I^{k(p)}_p = I^{MSFA}_p$) when its dependency upon p should be recalled for clarity sake. The difference value at $p \in \hat{S}^i(t)$ is computed following Eq. (16) as:

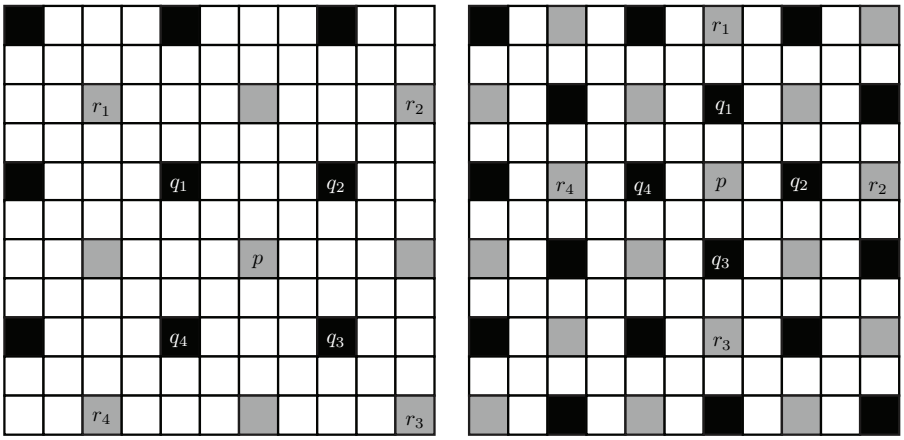
$$\hat{D}^{i,k(p)}_p = \frac{\sum_{q \in \tilde{N}_p(t)} \beta_q \cdot \dot{D}^{i,k(p)}_q}{\sum_{q \in \tilde{N}_p(t)} \beta_q}, \tag{18}$$

where $\dot{D}^{i,k(p)}_q = \dot{I}^i_q - \frac{1}{2}\left(I^{k(p)}_p + \dot{I}^{k(p)}_r\right)$ is a directional difference computed at one of the four closest neighbors q of p that belong to $\dot{S}^i(t)$. The pixel r is the symmetric of p with respect to q, so that r belongs to $\dot{S}^{k(p)}(t)$ (see Fig. 9). The value of the i-th channel at p is finally estimated as:

$$\hat{I}^i_p = \hat{D}^{i,k(p)}_p + I^{k(p)}_p. \tag{19}$$

Note that each weight β_q in Eq. (18) both depends on t and on the direction given by p and q (see Appendix A).

Shinoda et al. [27] also propose a post-processing of an initial estimation $\hat{\mathbf{I}}$, that updates each estimated channel $\hat{I}^i$ at each pixel p using Eq. (19) but now with $\hat{D}^{i,k(p)}_p = \frac{\sum_{q \in N_p} \beta_q \cdot \hat{d}^{i,k(p)}_q}{\sum_{q \in N_p} \beta_q}$ (see Eq. (18)), where N_p is made of the eight closest neighbors of p, and $\hat{d}^{i,k(p)}_q = \hat{I}^i_q - \hat{I}^{k(p)}_q$.



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

Fig. 9. Estimation of $\hat{I}^i$ by MLDI method (first two steps only) at $p \in S^k$ using the neighbors $q \in \dot{S}^i(t)$ and $r \in \dot{S}^{k(p)}(t)$. $\dot{S}^i(t)$ is displayed in black and $\hat{S}^i(t)$ in gray.

H. 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 Tab. II).

By using bilinear interpolation, all methods assume a strong spatial correlation among values within each channel. But Sec. III-D experimentally shows that the 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 [2]. 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).

Like SD, ItSD and MLDI methods are based on channel differences assuming that channels are correlated at each pixel. But Sec. III-D 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.

Several CFA demosaicing schemes exploit the properties of green channel (either implicitly or explicitly as in [12]) for demosaicing because it carries most of image structures. Indeed, the green band is dominant in Bayer CFA and the green channel is often assumed to represent the luminance information. Similarly, multispectral demosaicing schemes applied to MSFAs with a dominant band use the associated channel to estimate other channels [9, 19, 21, 22]. Because our considered MSFA exhibits no dominant band, we propose to compute a pseudo-panchromatic image (PPI) and use it for MSFA demosaicing.

TABLE II

PROPERTIES USED BY EXISTING DEMOSAICING METHODS. WB: WEIGHTED BILINEAR [1], DWT: DISCRETE WAVELET TRANSFORM [31], SD: SPECTRAL DIFFERENCE [1], ItSD: ITERATIVE SPECTRAL DIFFERENCE [18], BTES: BINARY TREE-BASED EDGE-SENSING [16], MLDI: MULTISPECTRAL LOCAL DIRECTIONAL INTERPOLATION [27].

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

V. PPI DETERMINATION AND PROPOSED DEMOSAICING METHODS

In this section, we first define the PPI of a fully-defined multispectral image and show that it is strongly correlated with all channels. Then we introduce how to estimate the PPI from a scale-adjusted raw image. We finally propose two multispectral demosaicing methods based on the PPI.

A. PPI of a multispectral image

The PPI I^M is defined at each pixel p as the mean value over all channels of a multispectral image [3]:

$$I_p^M = \frac{1}{K}\sum_{i=1}^{K} I_p^i. \tag{20}$$

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 $C\left(I^i, I^M\right)$ (see Eq. (5)) between each channel and the PPI on IMEC and IC images simulated with the CIE D65 illuminant. The results (see Figs. 4(c) and 4(d)) show that the channels are strongly correlated with the PPI. To compare this correlation with inter-channel correlation given by Eq. (5), the red lines in Figs. 4(a) and 4(b) show the bounds of the domain $\left\{\lambda^j : C\left(I^i, I^j\right) \geq C\left(I^i, I^M\right)\right\}_{j=1}^{K}$ for each $i = 1..K$ (column-wise). When band centers are distant ($\lambda^i \gg \lambda^j$ or $\lambda^i \ll \lambda^j$), channel I^i is more correlated with the PPI I^M than with I^j. This interesting property allows us to expect enhanced fidelity of PPI-based demosaicing methods that would exploit inter-channel differences.

B. Raw value scale adjustment for illumination robustness

As seen in Sec. III-D3, the non-uniformity of illumination spectral distribution impacts the acquired value range in channels. As a consequence, some channels have very low values that may bias the PPI estimation. To robustly estimate the PPI in varying illumination conditions, we propose to adjust the value scale of each subset S^i in the raw image I^{MSFA} before PPI estimation.

We compute a new raw value I'^{MSFA}_p at each pixel p as:

$$I'^{MSFA}_p = I^{MSFA}_p \cdot \frac{\max_{s\in S}(I^{MSFA}_s)}{\max_{s\in S^{k(p)}}(I^{MSFA}_s)}\ , \tag{21}$$

so that all channels have the same maximal available value in I'^{MSFA}. This pre-processing ensures that spectral correlation is efficiently handled by channel differences. Because it modifies the values of the raw image and of the estimated image $\hat{\mathbf{I}}'$ consequently, we need to restore the original value scale after demosaicing by post-processing each estimated channel $\hat{I}'^i$ as:

$$\hat{I}^i_p = \hat{I}'^i_p \cdot \frac{I^{MSFA}_p}{I'^{MSFA}_p}. \tag{22}$$

Below we explain how to estimate the PPI from the scale-adjusted raw image I'^{MSFA} and how to use this PPI to estimate $\hat{\mathbf{I}}'$ according to two demosaicing methods.

C. PPI estimation

Since the value of a single channel is available at each pixel in I'^{MSFA}, 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'^{MSFA} by applying an averaging filter M [17].

This filter has to take all channels into account while being as small as possible. Its size is hence that of the smallest odd-size neighborhood window including at least one pixel in all MSFA subsets $\{S^i\}_{i=1}^K$. 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 any 4×4 non-redundant MSFA, M is given by:

$$M = \frac{1}{64}\cdot\begin{bmatrix} 1 & 2 & 2 & 2 & 1\\ 2 & 4 & 4 & 4 & 2\\ 2 & 4 & 4 & 4 & 2\\ 2 & 4 & 4 & 4 & 2\\ 1 & 2 & 2 & 2 & 1\end{bmatrix}. \tag{23}$$

A first estimation of the PPI is then computed as [17]:

$$\bar{I}^M = I'^{MSFA} * M. \tag{24}$$

M is an averaging filter that may provide a smooth image. We instead here propose to use local directional information to obtain another estimation $\hat{I}^M$ of the PPI that is sharper than $\bar{I}^M$. 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 $S^{k(p)}$ (see Fig. 10(a)).

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

$$\gamma_q = \left(1+\sum_{v=-1}^{1}\sum_{u=0}^{1}\kappa(u,v)\cdot|I'^{MSFA}_{p+\rho(u,v)} - I'^{MSFA}_{q+\rho(u,v)}|\right)^{-1}. \tag{25}$$

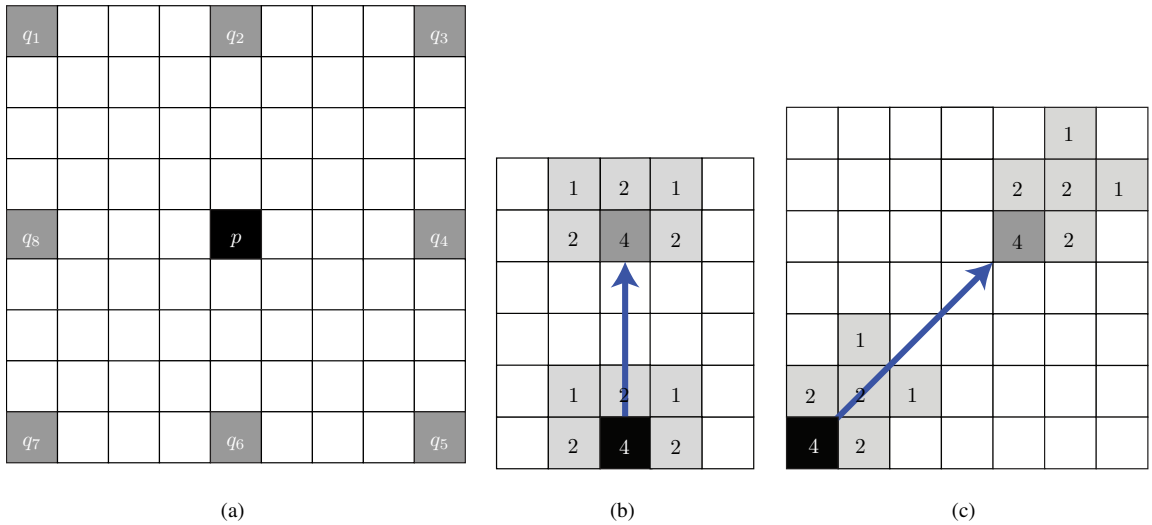


Fig. 10. Proposed PPI estimation: (a) neighborhood $\tilde{N}_p$ (in gray) of p (in black), (b)–(c) weight γ_q computation (see Eq. (25)) for $q = p+(0,-4)$ (b) and $q = p+(4,-4)$ (c) with coefficients $\kappa(u,v)$.

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:

$$\rho(u,v) = \begin{cases} \left(\frac{u\cdot\delta_x+v\cdot\delta_y}{4}, \frac{u\cdot\delta_y+v\cdot\delta_x}{4}\right) & \text{if } \delta_x\cdot\delta_y\text{=0,} \\ \left(\frac{\left(u+|v|\cdot\frac{1-v}{2}\right)\cdot\delta_x}{4}, \frac{\left(u+|v|\cdot\frac{1+v}{2}\right)\cdot\delta_y}{4}\right) & \text{otherwise,} \end{cases} \tag{26}$$

where $(\delta_x,\delta_y) \in \{-4,0,4\}^2$ are the coordinates of q relative to p. Figures 10(b) and 10(c) 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.

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

$$\Delta_p[I'^{MSFA}] = I'^{MSFA}_p - \frac{\sum_{q\in\tilde{N}_p}\gamma_q\cdot I'^{MSFA}_q}{\sum_{q\in\tilde{N}_p}\gamma_q}\ . \tag{27}$$

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 $\Delta_p[I'^{MSFA}] = \Delta_p[\bar{I}^M]$, which provides a new estimation of the PPI at each pixel:

$$\hat{I}^M_p = I'^{MSFA}_p + \frac{\sum_{q\in\tilde{N}_p}\gamma_q\cdot\left(\bar{I}^M_q - I'^{MSFA}_q\right)}{\sum_{q\in\tilde{N}_p}\gamma_q}. \tag{28}$$

To assess the two above PPI estimations, we compute the peak signal-to-noise ratio (PSNR) between each estimation and the reference PPI I^M. The PPI $\hat{I}^M$ estimated according to Eq. (28) provides a PSNR of $42.32\,dB$ (on average on IMEC and IC images simulated with the CIE D65 illuminant), which improves the estimation given by $\bar{I}^M$ [17] (see Eq. (24)) whose PSNR is $37.46\,dB$.

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

$$C_{S^i}\left(I'^{MSFA}, \hat{I}^M\right) = \frac{\sum\limits_{p\in S^i}\left(I'^{MSFA}_p - \mu^{MSFA}_{S^i}\right)\left(\hat{I}^M_p - \mu^M_{S^i}\right)}{\sqrt{\sum\limits_{p\in S^i}\left(I'^{MSFA}_p - \mu^{MSFA}_{S^i}\right)^2}\sqrt{\sum\limits_{p\in S^i}\left(\hat{I}^M_p - \mu^M_{S^i}\right)^2}}, \tag{29}$$

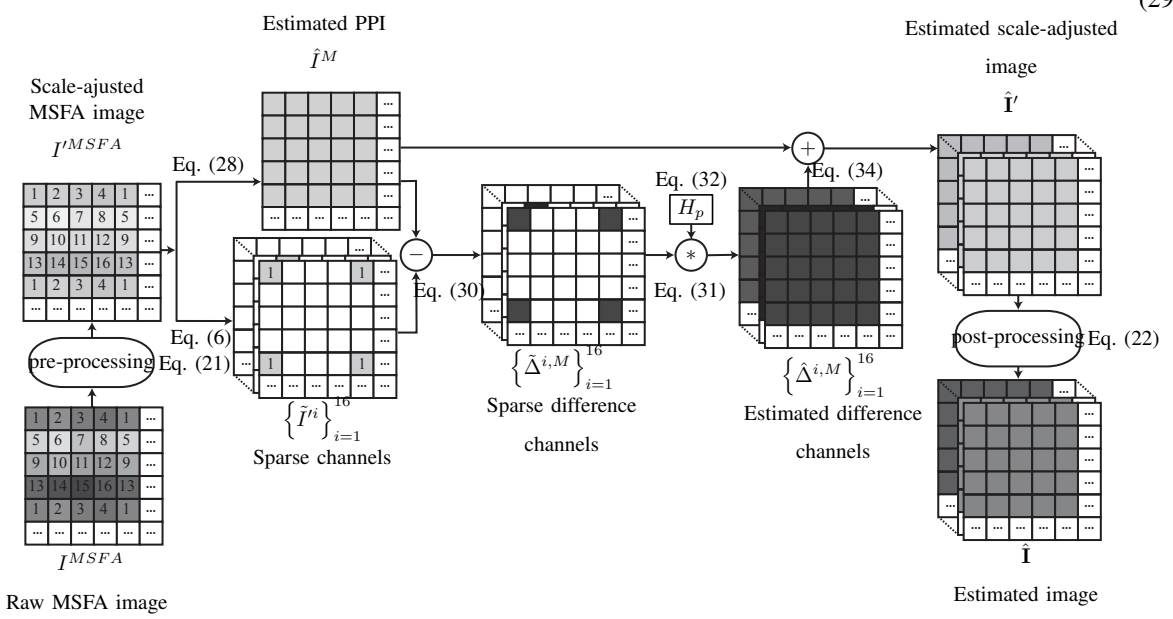


Fig. 11. Outline of the proposed PPID demosaicing method.

where $\mu^{MSFA}_{S^i}$ and $\mu^M_{S^i}$ are the average values of I'^{MSFA} and $\hat{I}^M$ at the pixels in S^i. We compute the average values of $C_{S^i}(I'^{MSFA}, \hat{I}^M)$ and of $C\left(I^i, I^M\right)$ between each reference channel I^i and the reference PPI (see Sec. V-A) on IMEC and IC multispectral images simulated with the CIE D65 illuminant. The results (not displayed here) show that $C_{S^i}(I'^{MSFA}, \hat{I}^M) = 0.979$ and $C\left(I^i, I^M\right) = 0.980$ for IMEC (0.929 and 0.927 for IC) on average over all channels. These correlations differ by less than 7.10^{-3} channel-wise for all images. We can conclude that the strong correlation between each reference channel and the PPI (see Figs. 4(c) and 4(d)) holds with the estimated PPI.

That leads us to exploit the estimated PPI $\hat{I}^M$ for demosaicing. Below we propose both an example of application of the PPI on an existing demosaicing method (DWT) and a new demosaicing method initiated in [17].

D. Using PPI in DWT

For the considered non-redundant MSFA, DWT uses the high-frequency contents of the mid-spectrum channel estimated by bilinear interpolation to estimate the other channels (see Sec. IV-C). 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 Sec. IV-C). The adapted method is referred to as PPDWT, and assessed in Sec. VI.

E. Pseudo-panchromatic image difference (PPID)

Instead of using the difference between channels as in SD (see Sec. IV-D), 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 $\hat{I}^M$ (see Eq. (28)).
2) Second, it computes the sparse difference $\tilde{\Delta}^{i,M}$ between each available value of I'^{MSFA} and the PPI at pixels in S^i, $i = 1..K$:
$$\tilde{\Delta}^{i,M} = \tilde{I}'^i - \hat{I}^M \odot m^i, \tag{30}$$
where $\tilde{I}'^i = I'^{MSFA} \odot m^i$.
3) Third, it uses the local directional weights computed according to Eq. (25) (whereas [17] directly uses H of Eq. (10)) to estimate the fully-defined difference $\hat{\Delta}^{i,M}$ by adaptive WB interpolation as:
$$\hat{\Delta}^{i,M}_p = \tilde{\Delta}^{i,M} * H_p. \tag{31}$$
Each element $(a, b) \in \{1, \ldots, 7\}^2$ of the new 7×7 adaptive convolution filter H_p is given by:
$$H_p(a, b) = \frac{F(a, b) \cdot \Gamma_p(a, b)}{\sum_{\substack{k=1 \\ k \equiv a \,(\mathrm{mod}\ 4)}}^{7} \sum_{\substack{l=1 \\ l \equiv b \,(\mathrm{mod}\ 4)}}^{7} F(k, l) \cdot \Gamma_p(k, l)}, \tag{32}$$
where $F(a, b)$ is defined by Eq. (9) and the denominator is a channel-wise normalization factor like in non-adaptive WB interpolation (see Eq. (11)). 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:
$$\Gamma_p = \left[\begin{array}{ccc|c|ccc} & & & \gamma_{q_2} & & & \\ & \gamma_{q_1} \cdot J_3 & & \gamma_{q_2} & & \gamma_{q_3} \cdot J_3 & \\ & & & \gamma_{q_2} & & & \\ \hline \gamma_{q_8} & \gamma_{q_8} & \gamma_{q_8} & 1 & \gamma_{q_4} & \gamma_{q_4} & \gamma_{q_4} \\ \hline & & & \gamma_{q_6} & & & \\ & \gamma_{q_7} \cdot J_3 & & \gamma_{q_6} & & \gamma_{q_5} \cdot J_3 & \\ & & & \gamma_{q_6} & & & \end{array}\right], \tag{33}$$
where $q_1 = p + (-4, -4)$, $\ldots$, $q_8 = p + (-4, 0)$ (see Fig. 10(a)) 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:
$$\hat{I}'^i = \hat{I}^M + \hat{\Delta}^{i,M}. \tag{34}$$

The proposed demosaicing method based on PPI difference (PPID) is outlined in Fig. 11.

VI. Experimental demosaicing results

A. PSNR assessment of demosaicing methods

We first objectively compare the demosaicing performances reached by PPI-based methods with those provided by the existing methods described in Sec. IV. Our reference images are the 32 IMEC and 32 IC multispectral images of size 512×512 pixels simulated from the CAVE database [32] with the CIE D65 illuminant as described in Sec. III-C. To obtain raw images, we spectrally sample these images according to the MSFA pattern shown in Fig. 1(d). The quality of a multispectral image estimated by demosaicing is evaluated by the average PSNR between reference and estimated channels, where the maximum pixel value is set to 255 for each channel. To avoid border effects, we consider only the 500×500 central pixels of each image in the PSNR computation.

Figure 12 displays the PSNR values provided by the demosaicing methods on average over the 32 images for each camera. The table in Fig. 12(a) shows that WB and BTES methods give poor results because they only use spatial correlation. DWT neither performs well because it uses the information of bilinearly interpolated channels for the estimation of all channels. The PPI more efficiently represents the high-frequency contents than the channel $\hat{I}^8_{WB}$ used by DWT, since the PSNR provided by PPDWT is improved by $3.28\,dB$ (on average over IMEC and IC images) with respect to DWT. Taking spectral correlation into account through spectral differences (SD) among all channels improves the PSNR of WB interpolation by $3.12\,dB$. Using channels associated with nearby band centers to iteratively update the results is an interesting approach that improves SD by $1.95\,dB$ in ItSD. Using local directional information is especially efficient: this improves SD by $3.22\,dB$ in MLDI that exploits inter-channel correlation, and by $4.01\,dB$ in PPID that exploits the correlation between each channel and the PPI. Our PPID method provides the best result for both IMEC and IC images. Figures 12(b) and 12(c) confirm these results, and show that PPID outperforms the other methods for almost all channels.

B. Visual comparison

We select an area of 70×70 pixels from the "Fake and Real Lemons" IMEC image simulated with the D65 illuminant to visually compare the results of each method in the channel associated with the band centered at $\lambda^9 = 552\,nm$. Figure 13 shows that the images estimated by the WB (b) and BTES (c) methods are strongly blurred. Due to the wavelet transform, DWT (d) introduces some artifacts and blurred edges that are reduced in PPDWT (e). The SD (f) and ItSD (g) methods suffer from severe zipper effect. All demosaicing artifacts are fairly reduced with MLDI (h) and PPID (i), that notably generate far less blur than WB. PPID provides sharper edges than MLDI but is slightly more sensitive to zipper effect.

We also convert the reference and estimated IMEC "Feathers" images to the sRGB color space via the CIE XYZ standard observer functions [11]. The conversion from XYZ to sRGB is performed using the CIE D65 illuminant as reference white. The results displayed in Fig. 14 show that WB and BTES provide highly blurred images with colors artifacts, and that PPDWT successfully reduces the image blur and color artifacts 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, that however still produces some false colors. PPID provides almost artifact-free images with sharp edges.

C. Illumination and demosaicing performances

As the non-uniformity of illumination impacts the acquired value range in channels (see Sec. III-D), it would be interesting to study the influence of illumination changes on the demosaicing method performances. For this purpose, we compute the PSNR provided by each method on IMEC and IC images simulated with the various illuminations of Fig. 2. The results are displayed in Tabs. III(a) and, III(b) the best one for each illumination being highlighted as bold. They show that the performances of all methods are affected by illumination changes. Thanks to the pre- and post-processing steps associated with PPI estimation (see Eq. (21) and (22)), PPI-based methods are fairly robust to illumination variations and especially show very good performance with the narrow-band F12 illuminant. To highlight the usefulness of the pre- and post-processing in PPI estimation, we also consider PPID method without scale adjustment, namely PPID_0. Results show that scale adjustment improves the performance by up to $6.94\,dB$ for all illuminations except D65 illuminant whose power distribution can be considered as uniform in the visible domain. Our PPID method always provides the best results; it performs slightly better than MLDI for D65 and outperforms all existing methods by more than $1.7\,dB$ for other illuminations.

TABLE III

PSNR (*dB*) OF IMAGES ESTIMATED BY EACH DEMOSAICING METHOD ACCORDING TO ILLUMINATION (AVERAGE OVER ALL CHANNELS OF THE 32 IMEC (A) AND IC (B) IMAGES). ILLUMINATIONS: SEE FIG. 2.

(a) IMEC

	D65	F12	A	HA	LD
WB	34.68	48.78	39.32	37.60	41.86
BTES	34.79	48.89	39.43	37.72	41.97
DWT	34.56	45.78	39.06	37.06	40.70
PPDWT	38.29	51.35	42.86	41.13	45.03
SD	38.14	49.63	42.35	40.33	43.22
ItSD	40.70	49.29	44.41	41.95	43.08
MLDI	41.70	52.20	45.83	43.83	46.01
PPID_0	42.99	51.36	46.37	43.98	45.61
PPID	**42.92**	**53.91**	**47.16**	**45.52**	**48.77**

(b) IC

	D65	F12	A	HA	LD
WB	34.27	50.89	39.60	38.04	42.20
BTES	34.39	50.98	39.72	38.16	42.32
DWT	33.68	43.16	38.26	35.41	38.47
PPDWT	36.53	51.97	41.76	40.20	44.07
SD	36.95	45.84	40.46	37.72	39.48
ItSD	38.28	43.77	40.74	37.63	38.03
MLDI	39.84	48.62	43.12	40.25	41.61
PPID_0	40.13	46.63	42.44	39.41	40.60
PPID	**40.18**	**53.57**	**45.10**	**43.57**	**46.88**

D. Computation time

Here we roughly consider the computation time required by the various methods. We have implemented all of them in Java under the open-source image analysis software ImageJ [26]. The Java bytecodes and an ImageJ macro are available as supplementary material. 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 IV 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 better results and requires less computation time than MLDI that is the best state-of-the-art method in terms of PSNR.

TABLE IV

COMPUTATION TIME RATIOS OF THE METHODS WITH RESPECT TO WB.

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

E. Experiments on landscape radiance images

Figure 4 shows that IMEC and IC cameras provide images with pairwise correlated channels even when the associated band centers are distant, all correlation values being higher than 0.60. The PPI is therefore strongly correlated with all channels and our PPI-based demosaicing method provides good results. It is interesting to examine the behavior of demosaicing methods when spectral correlation is weak. For this purpose, let us consider the six landscape images from Stanford database [28] acquired over the visible and near infrared (VIS+NIR) spectral domains. These are radiance 16-bit images with 148 channels, from which we pick $K = 16$ channels associated with narrow bands ($3.7\,nm$ resolution) whose centers are equally spaced band centers from 414.72 to $903.15\,nm$.

Figure 15(a) shows the spectral correlation between channels on average over the six images. The correlation is high within each of the VIS and NIR domains: it ranges from 0.65 to 1.00 inside the VIS domain (top left), and from 0.78 to 1.00 inside the NIR domain (bottom right). But channels associated with two bands in different domains (top right and bottom left) are weakly correlated as values range from -0.32 to 0.63. Figure 15(b) shows the correlation between each channel and the PPI. Like in Fig. 4, these values are reported column-wise on Fig. 15(a) where the red lines are the bounds of the domain $\left\{\lambda^j : C\left(I^i, I^j\right) \geq C\left(I^i, I^M\right)\right\}_{j=1}^{K}$ for each $i = 1..K$ (column-wise). When band centers are in different domains (top right and bottom left), channel I^i is more correlated with the PPI I^M than with I^j. These results allow us to expect that PPI-based demosaicing methods still provide enhanced fidelity in the VIS+NIR spectral domain. Because the correlation between PPI and several NIR channels is very low, we propose to estimate the PPI by using Eq. (24) or (28) that requires strong inter-channel correlation. Table V shows the average PSNR over all channels of the six images with respect to the different methods. WB and BTES provide good results since they use only spatial correlation. DWT performs poorly and is improved by using Eq. (24) of PPI in PPDWT. The performances of SD, ItSD, MLDI, and methods based on PPID computed with Eq. (28) are severely affected because they rely on the spectral correlation among all channels. PPID provides the best results when the PPI is estimated by Eq. (24) that is based on the spatial correlation assumption.

TABLE V

AVERAGE PSNR (dB) OVER ALL CHANNELS OF LANDSCAPE IMAGES OF STANFORD DATABASE ESTIMATED BY EACH DEMOSAICING METHOD. IN PPI-BASED METHODS, GRAY COLORED TEXT SHOWS THE EQUATION (EQ. (24) OR EQ. (28)) USED TO COMPUTE THE PPI.

WB	BTES	DWT	PPDWT	SD	ItSD	MLDI	PPID
38.59	38.62	36.87	Eq. (24): 37.59 Eq. (28): 36.81	36.85	36.96	38.13	Eq. (24): **39.43** Eq. (28): 37.70

VII. CONCLUSION

In this paper we show that the assumption about spectral correlation between channels should be carefully used since it no longer holds when the associated band centers are distant. That leads us to propose a new scheme for demosaicing based on the correlation between each channel and the pseudo-panchromatic image (PPI).

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 restore edge information in the estimated PPI. We then incorporate the PPI into an existing DWT-based method and propose a new demosaicing method based on PPI difference.

Extensive experiments show that the proposed scheme outperforms the existing ones suited to non-redundant MSFAs, both in terms of PSNR and in a visual assessment. The PPI difference method provides high-quality estimated images with sharp edges and reduced color and zipper artifacts. It is fairly robust to illumination variations and its computational complexity is moderate. Future works will focus on how to use the PPI with redundant MSFAs or in the frequency domain to perform multispectral demosaicing.

APPENDIX A
WEIGHT COMPUTATION IN MLDI

Let $q = p + (\delta_x, \delta_y)$ denote a neighboring pixel of p, and $r = p + 2 \cdot (\delta_x, \delta_y)$ (see Fig. 9). To estimate $\hat{I}_p^i$, the weight β_q of q used in MLDI (see Eq. (18)) is computed at step t according to the direction given by p and q as:

- for an horizontal direction:

$$\begin{aligned}\beta_q^H = \Big(\epsilon &+ \left|\dot{I}_r^{k(p)} - I_p^{k(p)}\right| \\ &+ \textstyle\sum_{d=0}^{\Delta-1} \left|I^{MSFA}_{p+\left(\frac{\delta_x}{|\delta_x|}\cdot(\Delta+d)),0\right)} - I^{MSFA}_{p-\left(\frac{\delta_x}{|\delta_x|}\cdot(\Delta-d),0\right)}\right| \\ &+ \textstyle\sum_{\substack{d=-\Delta \\ d\neq 0}}^{\Delta} \omega_d \cdot \left|I^{MSFA}_{p+(2\cdot\delta_x,d)} - I^{MSFA}_{p+(0,d)}\right|\Big)^{-1},\end{aligned} \tag{35}$$

- for a vertical direction:

$$\begin{aligned}\beta_q^V = \Big(&\epsilon + |\dot{I}_r^{k(p)} - I_p^{k(p)}| \\ &+ \textstyle\sum_{d=0}^{\Delta-1} \left| I^{MSFA}_{p+\left(0,\frac{\delta_y}{|\delta_y|}\cdot(\Delta+d)\right)} - I^{MSFA}_{p-\left(0,\frac{\delta_y}{|\delta_y|}\cdot(\Delta-d)\right)} \right| \\ &+ \textstyle\sum_{\substack{d=-\Delta \\ d\neq 0}}^{\Delta} \omega_d \cdot \left| I^{MSFA}_{p+(d,2\cdot\delta_y)} - I^{MSFA}_{p+(d,0)} \right| \Big)^{-1},\end{aligned} \tag{36}$$

- for a diagonal direction:

$$\begin{aligned}\beta_q^D = \Big(&\epsilon + \left| \dot{I}^i_{p+(\delta_x,\delta_y)} - \dot{I}^i_{p-(\delta_x,\delta_y)} \right| \\ &+ \left| \dot{I}_r^{k(p)} - I_p^{k(p)} \right| + \left| \hat{I}_{WB}^{k(p)}(q) - I_p^{k(p)} \right| \Big)^{-1},\end{aligned} \tag{37}$$

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

References

[1] J. Brauers and T. Aach, "A color filter array based multispectral camera," in *Institute of Imaging and Computer Vision*, RWTH Aachen University, 52056 Aachen, Germany, Oct. 2006.

[2] 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.

[3] 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.

[4] H. Fukuda, T. Uchiyama, H. Haneishi, M. Yamaguchi, and N. Ohyama, "Development of a 16-band multispectral image archiving system," in *Procs. SPIE Electronic Imaging: Color Imaging X*, vol. 5667, San Jose, CA, USA, Jan. 2005, pp. 136–145.

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

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

[7] N. Hagen and M. W. Kudenov, "Review of snapshot spectral imaging technologies," *Optical Engineering*, vol. 52, no. 9, p. 090901, Sep. 2013.

[8] 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 *Procs. IS&T's Image Processing, Image Quality, Image Capture Systems Conference (PICS 2001)*, Montréal, Quebec, Canada, Apr. 2001, pp. 185–189.

[9] 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 Trans. Image Process.*, vol. 26, no. 2, pp. 953–968, Feb. 2017.

[10] 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.

[11] J. Jordan and E. Angelopoulou, "Gerbil–A novel software framework for visualization and analysis in the multispectral domain," in *Vision, Modeling, and Visualization*. The Eurographics Association, 2010, pp. 259–266.

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

[13] 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.

[14] X. Li, B. Gunturk, and L. Zhang, "Image demosaicing: a systematic survey," in *Proc. SPIE Visual Communications and Image Processing (VCIP)*, vol. 6822, 2008, pp. 68 221J–68 221J–15.

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

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

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

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

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

[20] Y. Monno, T. Kitao, M. Tanaka, and M. Okutomi, "Optimal spectral sensitivity functions for a single-camera one-shot multispectral imaging system," in *Procs. 19th IEEE International Conference on Image Processing (ICIP 2012)*, Orlando, FL, USA, Sep. 2012, pp. 2137–2140.

[21] Y. Monno, M. Tanaka, and M. Okutomi, "Multispectral demosaicking using guided filter," in *Procs. SPIE Electronic Imaging Annual Symposium: Digital Photography VII*, vol. 8299, Burlingame, CA, USA, Jan. 2012, pp. 82 990O–82 990O–7.

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

[23] M. Parmar and S. J. Reeves, "Selection of optimal spectral sensitivity functions for color filter arrays," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3190–3203, Dec. 2010.

[24] 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.

[25] 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.

[26] W. S. Rasband, "ImageJ," U.S. National Institutes of Health, Bethesda, MD, USA, 1997-2016. [Online]. Available: https://imagej.nih.gov/ij

[27] 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 *Procs. APSIPA Annual Summit and Conference 2015*, Hong Kong, China, Dec. 2015.

[28] T. Skauli and J. Farrell, "A collection of hyperspectral images for imaging systems research," in *Procs. SPIE Electronic Imaging Annual Symposium: Digital Photography IX*, vol. 8660, Burlingame, CA, USA, Feb. 2013, pp. 86 600C–86 600C–7. [Online]. Available: https://scien.stanford.edu/index.php/landscapes

[29] 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, 2016.

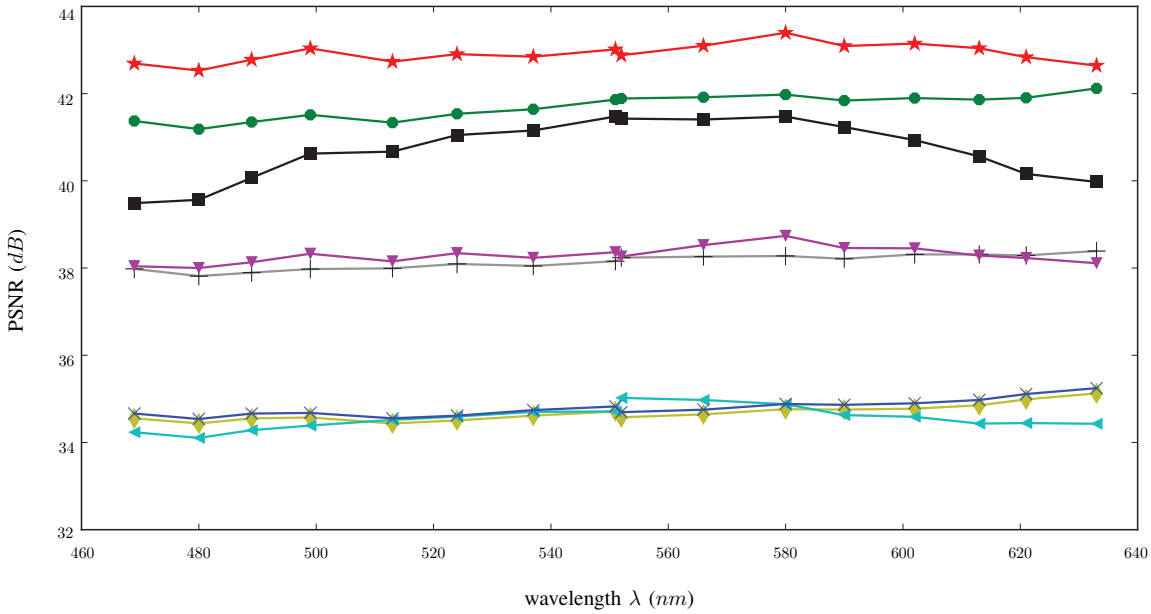
[30] X. Wang, J.-B. Thomas, J. Y. Hardeberg, and P. Gouton, "Median filtering in multispectral filter array demosaicking," in *Procs. SPIE Electronic Imaging Annual Symposium: Digital Photography IX*, vol. 8660, Burlingame, CA, USA, Feb. 2013, pp. 86 600E–86 600E–10.

[31] ——, "Discrete wavelet transform based multispectral filter array demosaicking," in *Procs. 2013 Colour and Visual Computing Symposium (CVCS 2013)*, Gjøvik, Norway, Sep. 2013, pp. 1–6.

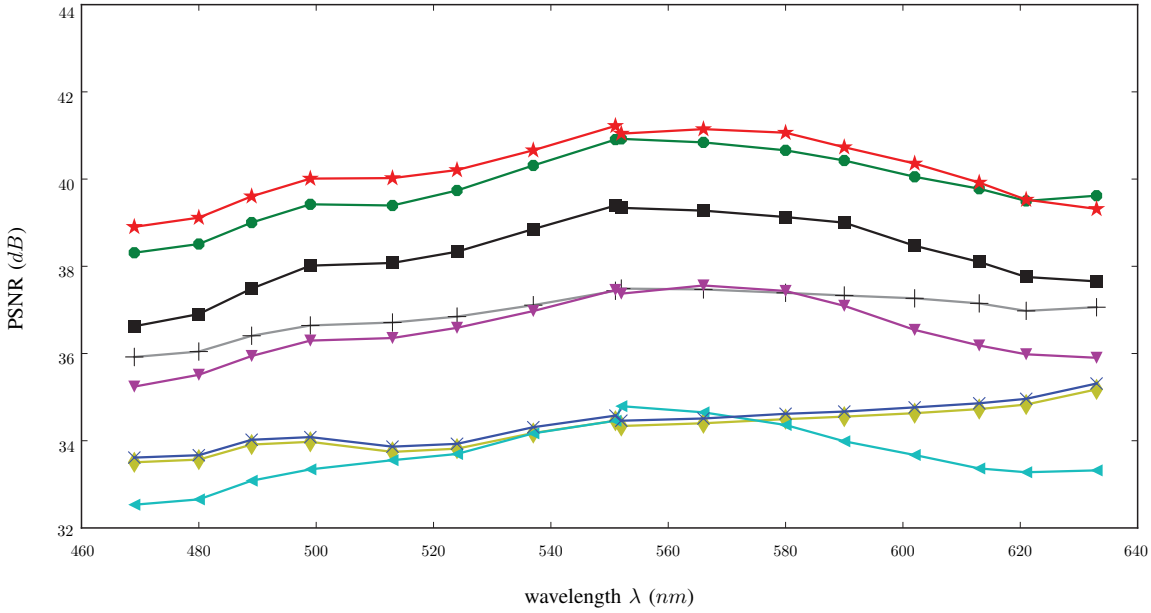
[32] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.

Method	WB	BTES	DWT	PPDWT	SD	ItSD	MLDI	PPID
Symbol	♦	×	◀	▼	+	■	●	★
IMEC	34.68	34.79	34.56	38.29	38.14	40.70	41.70	**42.92**
IC	34.27	34.39	33.68	36.53	36.95	38.28	39.84	**40.18**

(a) Caption of Figs. (b) and (c) and average PSNR (dB)

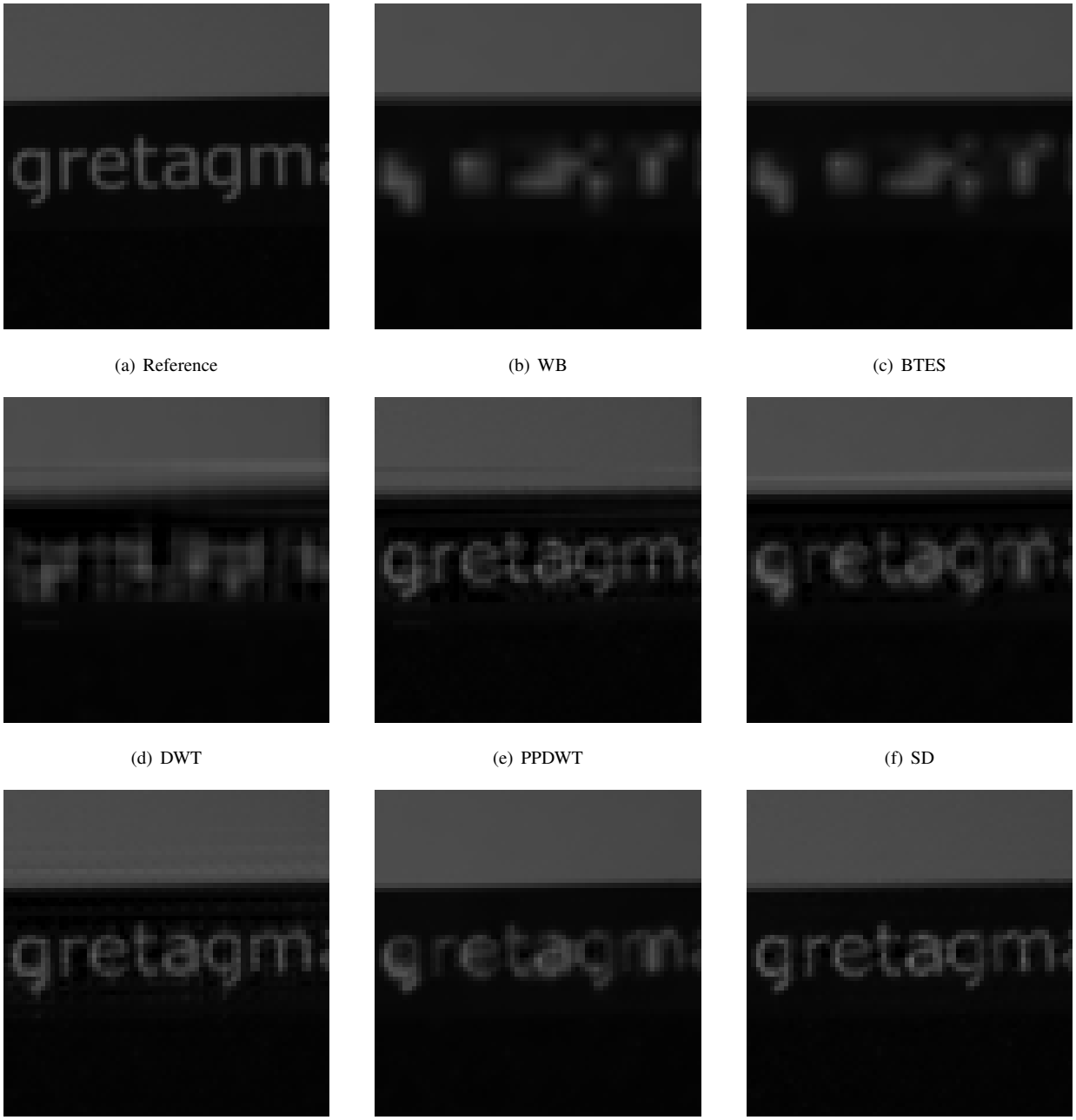


(b) IMEC



(c) IC

Fig. 12. Average PSNR (dB) of the 32 images (under CIE D65 illuminant) estimated by each demosaicing method for each camera. (a): Average PSNR (dB) over all channels for each camera, with the best result in bold. (b), (c): PSNR (dB) of each channel given by its band center. Existing methods: see Tab. II. Proposed methods: PPDWT: PPI-based DWT, PPID: PPI difference.

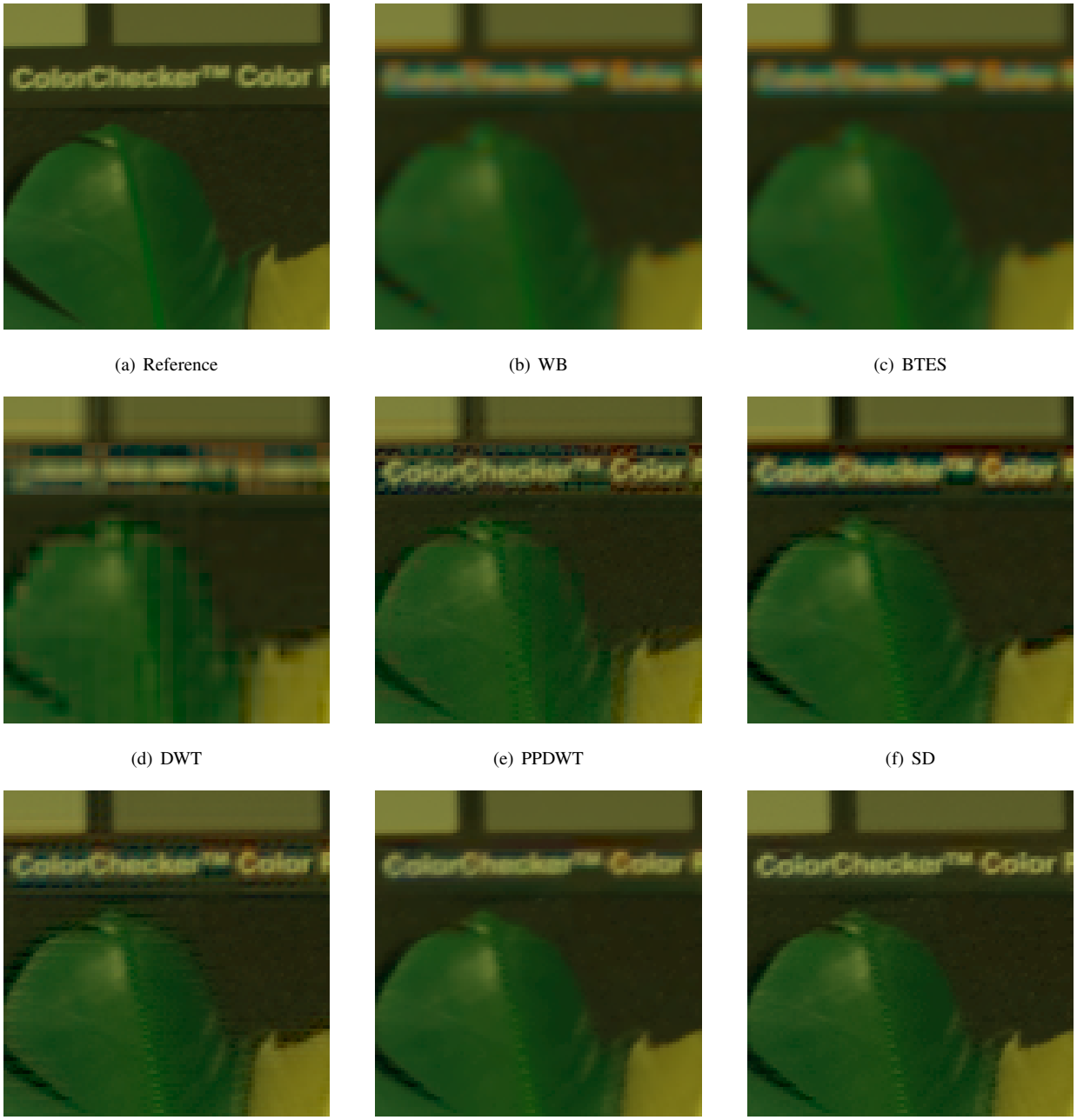


(a) Reference (b) WB (c) BTES

(d) DWT (e) PPDWT (f) SD

(g) ItSD (h) MLDI (i) PPID

Fig. 13. Central extract of reference I^9 (a) and estimated $\hat{I}^9$ (b)–(i) of IMEC "Fake and Real Lemons" image simulated under CIE D65 illuminant.

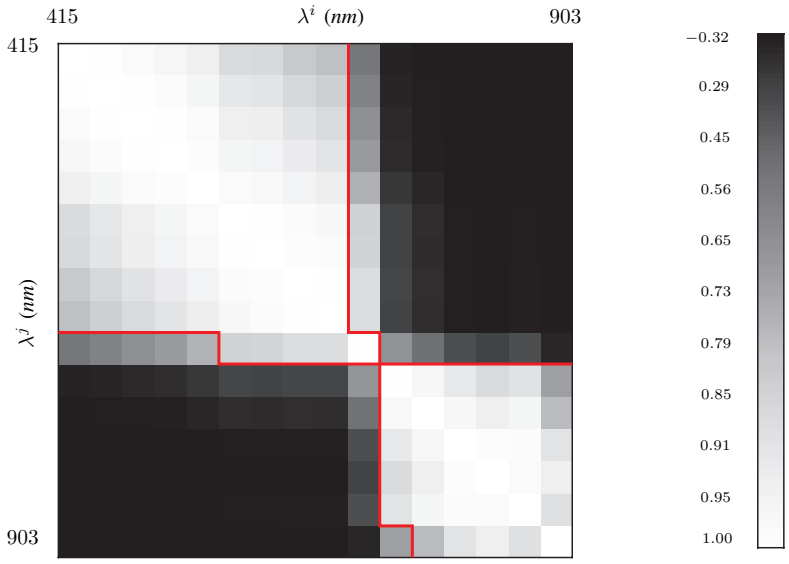


(a) Reference (b) WB (c) BTES

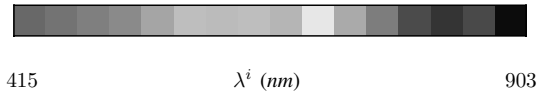
(d) DWT (e) PPDWT (f) SD

(g) ItSD (h) MLDI (i) PPID

Fig. 14. sRGB renderings of a central extract of IMEC “Feathers” image simulated under CIE D65 illuminant.



(a) $C\left(I^i, I^j\right)$ on Stanford images



(b) $C\left(I^i, I^M\right)$ on Stanford images

Fig. 15. Correlation between channels I^i and I^j (a) and between I^i and PPI I^M (b). Values are averaged over six images from Stanford database [28], range between -0.32 (black) and 1.00 (white) and are displayed with cubic scale. Values of (b) are reported column-wise as red lines on (a).