2008

Color Filter Array Image Analysis for Joint Denoising and Demosaicking

Keigo Hirakawa
University of Dayton, khirakawa1@udayton.edu

9

Color Filter Array Image Analysis for Joint Demosaicking and Denoising

Keigo Hirakawa

9.1 Introduction

Noise is among the worst artifacts that affect the perceptual quality of the output from a digital camera (see Chapter 1). While cost-effective and popular, single-sensor solutions to camera architectures are not adept at noise suppression. In this scheme, data are typically obtained via a spatial subsampling procedure implemented as a color filter array (CFA), a physical construction whereby each pixel location measures the intensity of the light corresponding to only a single color [1], [2], [3], [4], [5]. Aside from undersampling, observations made under noisy conditions typically deteriorate the estimates of the full-color image in the reconstruction process commonly referred to as *demosaicking* or *CFA interpolation* in the literature [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16]. A typical CFA scheme involves the canonical color triples (i.e., red, green, blue), and the most prevalent arrangement called Bayer pattern is shown in Figure 9.1b.

As the general trend of increased image resolution continues due to prevalence of multimedia, the importance of interpolation is de-emphasized while the concerns for computational efficiency, noise, and color fidelity play an increasingly prominent role in the decision making of a digital camera architect. For instance, the interpolation artifacts become less noticeable as the size of the pixel shrinks with respect to the image features, while the

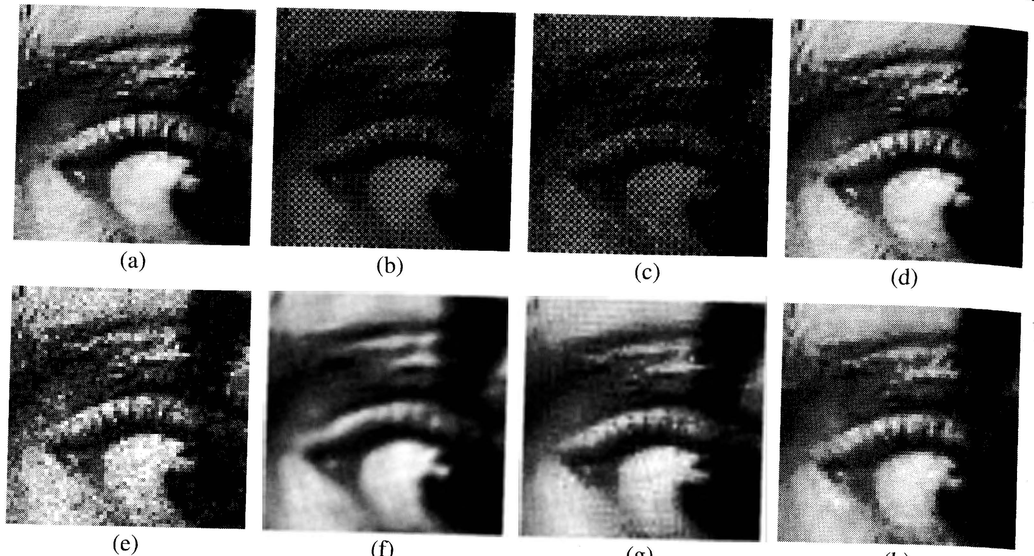

FIGURE 9.1 (See color insert.)

Zoomed portion of the *Clown* image: (a) original color image, (b) color version of ideal CFA image, (c) color version of noisy CFA image, (d) demosaicking the ideal CFA image, (e) demosaicking the noisy CFA image, (f) demosaicking the noisy CFA image followed by denoising, (g) denoising the noisy CFA image followed by demosaicking, and (h) joint denoising and demosaicking of the noisy CFA image.

decreased dimensionality of the pixel sensors on the complementary metal oxide semiconductor (CMOS) and charge coupled device (CCD) sensors make the pixels more susceptible to noise. Photon-limited influences are also evident in low-light photography, ranging from a specialty camera for precision measurement to indoor consumer photography.

Sensor data, which can be interpreted as subsampled or incomplete image data, undergo a series of image processing procedures in order to produce a digital photograph. Refer to Chapters 1 and 3 for details. However, these same steps may amplify noise introduced during image acquisition. Specifically, the demosaicking step is a major source of conflict between the image processing pipeline and image sensor noise characterization because the interpolation methods give high priority to preserving the sharpness of edges and textures. In the presence of noise, noise patterns may form false edge structures, and therefore the distortions at the output are typically correlated with the signal in a complicated manner that makes noise modelling mathematically intractable. Thus, it is natural to conceive of a rigorous tradeoff between demosaicking and image denoising.

For better illustration, Figure 9.1a shows a typical color image. Suppose we simulate the noisy sensor observation by subsampling this image according to a CFA pattern (Figure 9.1b) and corrupting with noise (Figure 9.1c). While state-of-the-art demosaicking methods such as the ones in [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16] do an impressive job in estimating the full-color image given ideal sensor data (Figure 9.1d), the interpolation may also amplify the noise in the sensor measurements, as demonstrated in Figure 9.1e. The state-of-the-art denoising methods applied to Figure 9.1f yield unsatisfactory results (Figure 9.1g), suggesting a lack of coherent strategy to address interpolation

and noise issues jointly. For comparison, the output from a joint demosaicking and denoising method [17] is shown in Figure 9.1h, clearly demonstrating the advantages.

In this chapter, the problem of estimating the complete noise-free image signal of interest given a set of incomplete observations of pixel components that are corrupted by noise is approached statistically from a point of view of Bayesian statistics, that is modelling of the various quantities involved in terms of priors and likelihood. The three design regimes that will be considered here can be thought of as simultaneous interpolation and image denoising, though this chapter has a wider scope in the sense that modelling the image signal, missing data, and the noise process explicitly yield insight into the interplay between the noise and the signal of interest. The chapter is not intended to comprise detailed step-by-step instructions of how to estimate a complete noise-free image; rather we present a theoretical basis for generalizing the image signal models to the noisy subsampled case, and propose major building blocks for manipulating such data. The author feels that leading the discussion in this manner is most effective, as it allows flexibility in the choice of models.

There are a number of advantages to the proposed estimation schemes over the obvious alternative, which is the serial concatenation of the independently designed interpolation and image denoising algorithms. For example, the inherent image signal model assumptions underlying the interpolation procedure may differ from those of the image denoising. This discrepancy is not only contradictory and thus inefficient, but also triggers mathematically intractable interactions between mismatched models. Specifically, interpolating distorted image data may impose correlation structures or bias to the noise and image signal in an unintended way. Furthermore, a typical image denoising algorithm assumes a statistical model for natural images, not that of the output of interpolated image data. While grayscale and color image denoising techniques have been suggested [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], removing noise after demosaicking, however, is impractical. Likewise, although many demosaicking algorithms developed in the recent years yield impressive results in the absence of sensor noise, the performance is less than desirable in the presence of noise.

In this chapter, we investigate the problem of estimating a complete color image from the noisy undersampled signal using spectral and wavelet analysis of the noisy sensor data. In Section 9.2, we characterize the noise corresponding to CMOS and CCD sensors and evaluate it with respect to human visual system sensitivities and current image denoising techniques. Section 9.3 identifies the structure in the loss of information due to sampling and noise by examining the sensor data in the Fourier domain, and motivates a unified approach to interpolation and denoising. To exploit the local aliasing structures, Section 9.4 refines the spectral analysis of sensor data using time-frequency analysis. Conditioned on the signal image model, we propose three frameworks for estimating the complete noise-free image via the manipulation of noisy subsampled data. In Section 9.5, we discuss the design of a spatially-adaptive linear filter whose stop-band annihilates color artifacts and whose pass-band suppresses noise. Section 9.6 demonstrates the modelling of noisy subsampled color images in the wavelets domain using a statistical missing data formulation. As outlined in Section 9.7, however, it is possible to estimate the wavelet coefficients corresponding to the desiderata from the wavelet coefficients of the sensor data. This section presents example output images obtained using the techniques presented in this chapter. Finally, concluding remarks are listed in Section 9.8.

9.1.1 A Comment About Model Assumptions

The wavelet-based statistical models for image signals play a dominant part in the image denoising literature. In this paradigm, wavelet coefficients corresponding to image signals exhibit a heavy-tailed distribution behavior, motivating the use of Laplacian distribution, Student's t-distribution, and Gaussian mixtures, to name a few. These heavy-tailed priors can be written as a continuous mixture of Gaussian with the general form,

$$x|q \sim \mathcal{N}(\mu_x, \sigma_x^2/q),$$

where μ_x and σ_x^2 are the mean and variance parameters of a random variable x, and $q \neq 0$ is an augmented random variable with its own distribution specific to the choice of heavy-tail. Thus, x is conditionally normal; conditioned on q, its posterior distribution can largely be manipulated with second-order statistics. Alternatives to wavelet-based models include image patches [32], principal components [33], and anisotropic diffusion equations [28]. Many of them make use of the sum of (sometimes spatially-adaptive) outer-products of vectorized pixel neighborhoods, which is the deterministic-counterpart to the pixel-domain second-order statistics.

The intentions of this chapter, as stipulated previously, are to provide tools for analyzing and manipulating subsampled data in a way that is relevant to the CFA image. Rather than reinvent signal models for subsampled image data, we choose to work with statistical or deterministic models for a *complete* image data. In doing so, we inherit a rich literature in image modelling that has been shown to work well for image denoising, interpolation, segmentation, compression, and restoration. Furthermore, the discussion that follows is intentionally decoupled from a *particular choice* of image signal model. Instead, *conditioned* on the complete image model, the primary focus of the discussions will be on making the necessary changes amenable to the direct manipulation of the CFA image.

Specifically, the theoretical frameworks for analyzing subsampled data below are developed in terms of second-order statistics of complete image data. By taking the expectation over the conditionals in the posterior ($E[x] = E[E[x|q]]$ in the example above, where $x|q$ in the inner expectation is normal) one can generalize the estimator derived for the multivariate normal to the heavy-tailed distribution, as in the case of Bayesian estimators. Alternatively, replacing the second-order statistics with the sum of outer-products would yield the deterministic extension of the CFA image processing. In any case, the technical frameworks presented below are nonrestrictive and compatible with a wide range of assumed models, allowing for the flexibility in selecting a model best suited for the computational and image quality requirements of the application.

9.1.2 Terminologies and Notational Conventions

Because there are several technical terms used in this chapter that sound similar but have different meanings, we would like to clarify their definitions. The term *color filter* refers to a physical device placed over photosensitive elements called pixel sensors. It yields a color coding by cutting out electromagnetic radiations of specified wavelengths. This is not to be confused with a *filter*, or convolution filtering realized by taking a linear combination of nearby pixel or sensor values. Likewise, given a two-dimensional signal, terminologies

such as frequency and spectrum are to be interpreted in the context of two-dimensional Fourier transforms and not in the sense of colorimetry.

In this chapter, all image signals are assumed to be discrete (or post-sampling). For notational simplicity, plain characters (e.g., x) represent a singleton, whereas bold-face characters (e.g., $\boldsymbol{x}$) represent a vector or a matrix. An arrow over a character symbolizes a vectorization; that is, $\vec{x}$ is a re-arrangement of $x(\cdot)$ into a vector form. Other conventions are summarized below for bookkeeping, but their formal definitions will be made explicit in the sequel:

$\boldsymbol{n} \in \mathbb{Z}^2$	pixel/sample location index
$\boldsymbol{x} : \mathbb{Z}^2 \to \mathbb{R}^3$	signal-of-interest, ideal (noise-free) color image; $\boldsymbol{x} = [x_1, x_2, x_3]^T$ are the RGB triples
$\boldsymbol{\epsilon} : \mathbb{Z}^2 \to \mathbb{R}^3$	noise for $\boldsymbol{x}$
$\boldsymbol{c} : \mathbb{Z}^2 \to \{0,1\}^3$	color filter coding indicator
$\ell : \mathbb{Z}^2 \to \mathbb{R}$	monochromatic or approximate luminance image, $\ell = \frac{1}{4}x_1 + \frac{1}{2}x_2 + \frac{1}{4}x_3$
$\alpha : \mathbb{Z}^2 \to \mathbb{R}$	color difference or approximate chrominance image, $\alpha = x_1 - x_2$
$\beta : \mathbb{Z}^2 \to \mathbb{R}$	color difference or approximate chrominance image, $\beta = x_3 - x_2$
$y : \mathbb{Z}^2 \to \mathbb{R}$	ideal (noise-free) sensor data or CFA image, $y(\boldsymbol{n}) = \boldsymbol{c}^T(\boldsymbol{n})\boldsymbol{x}(\boldsymbol{n})$
$\varepsilon : \mathbb{Z}^2 \to \mathbb{R}$	noise for y
$z : \mathbb{Z}^2 \to \mathbb{R}$	noisy sensor data, $z = y + \varepsilon$
$\boldsymbol{g} : \mathbb{Z}^2 \times \mathbb{Z}^2 \to \mathbb{R}^3$	spatially-adaptive filter coefficients
$h_0, h_1, f_0, f_1 : \mathbb{Z} \to \mathbb{R}$	one-dimensional impulse responses to convolution filters used in filterbank

In the above, the elements in the vector $\boldsymbol{x}(\boldsymbol{n}) = [x_1(\boldsymbol{n}), x_2(\boldsymbol{n}), x_3(\boldsymbol{n})]^T$ are interpreted as the red, green, blue pixel component values, respectively, though the results established in this chapter are equally applicable in other color coding schemes. The luminance-chrominance representation of a color image, $[\ell(\boldsymbol{n}), \alpha(\boldsymbol{n}), \beta(\boldsymbol{n})]$, is an invertible linear transformation of $\boldsymbol{x}(\boldsymbol{n})$. The symbols $x : \mathbb{Z} \to \mathbb{R}$ and $\varepsilon : \mathbb{Z} \to \mathbb{R}$ are also occasionally used for a generic (nondescriptive) signal and noise, respectively. Singleton functions $x(\boldsymbol{n})$ and $\varepsilon(\boldsymbol{n})$ are used interchangeably with $\boldsymbol{x}(\boldsymbol{n})$ and $\boldsymbol{\epsilon}(\boldsymbol{n})$ to generalize results to the multivariate case, respectively.

In addition, given a two-dimensional function $x : \mathbb{Z}^2 \to \mathbb{R}$, its Fourier transform is denoted by $\hat{x}(\boldsymbol{\omega})$, where, in the two-dimensional case, $\boldsymbol{\omega} = [\omega_0, \omega_1]^T \in \mathbb{R}^2$ is the modulo-2π frequency index. Similarly, let $\boldsymbol{i} \in \{0, 1, \ldots, I\}^2$ be the subband index for the $(I+1)^2$-level (separable) two-dimensional filterbank decomposition, where a smaller index value corresponds to low-frequency channel. Then $w_{\boldsymbol{i}}^x(\boldsymbol{n})$ is the filterbank (or wavelet packets) coefficient at the $\boldsymbol{i}$-th subband, $\boldsymbol{n}$-th spatial location corresponding to the signal $x(\boldsymbol{n})$.

9.2 Noise Model

In order to design an effective image denoising system, it is important to characterize the noise in an image sensor. The CMOS photodiode active pixel sensor typically uses a photodiode and three transistors, all major sources of noise [34]. The CCD sensors rely on the electron-hole pair that is generated when a photon strikes silicon [35]. While a detailed investigation of the noise source is beyond the scope of this chapter, studies suggest that $z : \mathbb{Z}^2 \to \mathbb{R}$, the number of photons encountered during an integration period (duration between resets), is a Poisson process $\mathscr{P}_y$:

$$p(z(\boldsymbol{n})|y(\boldsymbol{n})) = \frac{e^{-y(\boldsymbol{n})}y(\boldsymbol{n})^{z(\boldsymbol{n})}}{z(\boldsymbol{n})!},$$

where $\boldsymbol{n} \in \mathbb{Z}^2$ is the pixel location index, and $y(\boldsymbol{n})$ is the expected photon count per integration period at location $\boldsymbol{n}$, which is linear with respect to the intensity of the light. Note $E\left[z(\boldsymbol{n})|y(\boldsymbol{n})\right] = y(\boldsymbol{n})$ and $E\left[z^2(\boldsymbol{n}) - E\left[z(\boldsymbol{n})|y(\boldsymbol{n})\right]^2 \middle| y(\boldsymbol{n})\right] = y(\boldsymbol{n})$. Then, as the integration period increases, $p(z(\boldsymbol{n})|y(\boldsymbol{n}))$ converges weakly to $\mathscr{N}\left(y(\boldsymbol{n}), y(\boldsymbol{n})\right)$, or

$$z(\boldsymbol{n}) \approx y(\boldsymbol{n}) + \sqrt{y(\boldsymbol{n})}\varepsilon(\boldsymbol{n}), \tag{9.1}$$

where $\varepsilon \overset{\text{i.i.d.}}{\sim} \mathscr{N}(0,1)$ is independent of y. This approximation is justifiable via a straightforward application of central limit theorem to the binomial distribution. The noise term, $\sqrt{y(\boldsymbol{n})}\varepsilon(\boldsymbol{n})$ is commonly referred to as the *shot noise*.

In practice, the photodiode charge (e.g., photodetector readout signal) is assumed proportional to $z(\boldsymbol{n})$, thus we interpret $y(\boldsymbol{n})$ and $z(\boldsymbol{n})$ as the ideal and noisy sensor data at pixel location $\boldsymbol{n}$, respectively. For a typical consumer-grade digital camera, the approximation in Equation 9.1 is reasonable. The significance of Equation 9.1 is that the signal-to-noise ratio improves for a large value of $y(\boldsymbol{n})$ (e.g., outdoor photography), while for a small value of $y(\boldsymbol{n})$ (e.g., indoor photography) the noise is severe. To make matters worse, human visual response to the light $y(\boldsymbol{n})$ is often modeled as $\sqrt[3]{y(\boldsymbol{n})}$, suggesting a heightened sensitivity to the deviation in the dark regions of the image. To see this, the perceived noise magnitude is proportional to:

$$\sqrt[3]{z(\boldsymbol{n})} - \sqrt[3]{y(\boldsymbol{n})} = \sqrt[3]{y(\boldsymbol{n}) + \sqrt{y(\boldsymbol{n})}\varepsilon(\boldsymbol{n})} - \sqrt[3]{y(\boldsymbol{n})},$$

which is a monotonically decreasing function with respect to $y(\boldsymbol{n})$ for a fixed value of $\varepsilon(\boldsymbol{n})$.

There have been some hardware solutions to the sensor noise problems. For example, the cyan-magenta-yellow (CMY) CFA pattern performs better in a noisy environment, as the quantum efficiency is more favorable for CMY as compared to RGB. That is, a CMY-based CFA allows more photons to penetrate through to the photosensitive element because the pigments used in it are considerably thinner than those of the RGB-based CFA. The disadvantage is that the photo-sensitivity wavelengths of the cyan, magenta, and yellow overlap considerably, and therefore the color space conversion from CMY to the RGB color space is an unstable operation. Today, the CMY-based CFAs are more readily used in video cameras, since the frame-rate restricts the length of the integration period. Other circuit-based noise-reduction techniques include correlated double sampling. In this scheme, the pixel sensors are each sampled twice, first measuring the reset/amplifier noise alone, and second measuring the photon counts and the reset/amplifier noise combined. The difference of the two is presumed noise-free.

In reality, efforts to address signal-dependent noise in Equation 9.1 lag behind those of image interpolation and image denoising for additive white Gaussian noise (AWGN). A standard technique for working with signal-dependent noise is to apply an invertible nonlinear operator $\gamma(\cdot)$ on z such that signal and noise are (approximately) decoupled:

$$\gamma(z)|\gamma(y) \sim \mathscr{N}\left(\gamma(y), \sigma^2\right)$$

for some constant σ^2. Homomorphic filtering is one such operator designed with monotonically-increasing nonlinear pointwise function $\gamma : \mathbb{R} \to \mathbb{R}$, [36], [37]. The Haar-Fisz transform $\gamma : \mathbb{Z}^2 \times \mathbb{R} \to \mathbb{Z}^2 \times \mathbb{R}$ is a multiscale method that asymptotically decorrelates signal and noise [38], [39]. In any case, a signal estimation technique (assuming AWGN) is used to estimate $\gamma(y)$ given $\gamma(z)$, and the inverse transform $\gamma^{-1}(\cdot)$ yields an estimate of y. The advantage of this approach is the modularity of the design of $\gamma(\cdot)$ and the estimator. The disadvantage is that the signal model assumed for y may not hold for $\gamma(y)$ and the optimality of the estimator (e.g., minimum mean squared error estimator) in the new domain does not translate to optimality in the rangespace of y, especially when $\gamma(\cdot)$ significantly deviates from linearity.

An alternative to decorrelation is to approximate the noise standard deviation, $\sqrt{y(\boldsymbol{n})}$. The AWGN noise model is effectively a zero-th order Taylor expansion of the Poisson process; an affine noise model is the first order Taylor expansion of Equation 9.1 used in References [32] and [40]. In practice, these approximations yield acceptable performance because the CMOS sensors operate on a relatively limited dynamic range, giving validity to the Taylor assumption (when the expansion is centered about the midpoint of the operating range). The human visual system can also tolerate a greater degree of error in the brighter regions of the image, allowing for more accurate noise characterization for small values of y (at the cost of poorer characterization for higher rangespace of y). Alternatively, empirical methods that address signal-dependent noise take a two-step approach [21]. First, a crude estimate of the noise variance at each pixel location $\boldsymbol{n}$ is found; second, conditioned on this noise variance estimate, we assume that the signal is corrupted by signal-independent noise. A *piecewise* AWGN model achieves a similar approximation.

Methods that work with the posterior distribution of the coefficients of interests, such as Markov chain Monte Carlo and importance sampling, either have a slow convergence rate or require a large number of observations [41]. Emerging frameworks in Bayesian analysis for Poisson noise yield an asymptotic representation of the Poisson process in the wavelets domain, but the manipulation of data in this class of representation is extremely complicated [42].

For all the reasons above, it is clear that the estimation of the mean y given the Poisson process z is not a well-understood problem; and existing methods use variations of AWGN models to address the Poisson noise. Hence, while acknowledging inadequacies, we restrict our attention to the AWGN problem,

$$z(\boldsymbol{n}) = y(\boldsymbol{n}) + \varepsilon(\boldsymbol{n}), \tag{9.2}$$

where $\varepsilon \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$.

9.3 Spectral Analysis of CFA Image

In this section, we take a closer look at the sampling scheme and the structure of aliasing induced by the Bayer color filter array illustrated in Figure 9.1b, [11], [17]. The estimation of missing pixel components given observed pixel components is generally an ill-posed problem. By assuming that the image signals are highly structured, however, we effectively assume that the signal-of-interest lives in a lower-dimensional subspace that can be represented by the subspace spanned by the color filter array. Thus, although the *loss of data* at the hardware interface is inevitable, the *loss of information* due to sampling may be limited. We will show that the Fourier analysis and aliasing serve as a measure of loss of information, and that they motivate joint modelling and manipulation of subsampled data and noise (which will subsequently be fine-tuned using locally adaptive schemes in Sections 9.5 to 9.7).

In a color image, such as one shown in Figure 9.1a, the image pixel $\boldsymbol{x}(\boldsymbol{n}) = [x_1(\boldsymbol{n}), x_2(\boldsymbol{n}), x_3(\boldsymbol{n})]^T$ at the position $\boldsymbol{n} \in \mathbb{Z}^2$ denotes a vectorial value, typically expressed in terms of RGB coordinates. Figure 9.2a is a grayscale version of Figure 9.1a. Visual inspection of the original color image and its corresponding red, green, and blue channels depicted in Figure 9.2b to Figure 9.2d, respectively, reveals that the decomposed color channels may contain redundant information with respect to edge and textural formation, reflecting the fact that the changes in color at the object boundary are secondary to the changes in intensity. It follows from the (de-)correlation of color content at high frequencies and is well accepted among the color image scientists that the difference images (e.g., red-green, blue-green) exhibit rapid spectral decay relative to monochromatic image signals (e.g., gray, red, green), and are therefore slowly-varying over spatial domain. See Figure 9.2e and Figure 9.2f. Such heuristic intuitions are further backed by human physiology — the contrast sensitivity function for the luminance channel in human vision is typically modelled with a much higher pass-band than that of the chrominance channels.

An alternative to spectral modelling strategy based on color-ratio has been studied [43], [44], [45], [46]. Assuming that objects are piecewise constant color, then the ratios between color components within an object are constant, even though the intensities of pixels may vary over space. In practice, however, the numerical stability of ratios is difficult to achieve, and the spatial variation of the intensity levels is not captured explicitly by this model. For these reasons, while acknowledging the merits of the color-ratio modelling strategy, the discussions in this chapter will be confined to the difference image modelling.

Let $\boldsymbol{c}(\boldsymbol{n}) = [c_1(\boldsymbol{n}), c_2(\boldsymbol{n}), c_3(\boldsymbol{n})]^T \in \{[1,0,0]^T, [0,1,0]^T, [0,0,1]^T\}$ be a CFA coding such that the noise-free sensor data can be written as an inner product, $y(\boldsymbol{n}) = \boldsymbol{c}^T(\boldsymbol{n})\boldsymbol{x}(\boldsymbol{n})$. Given

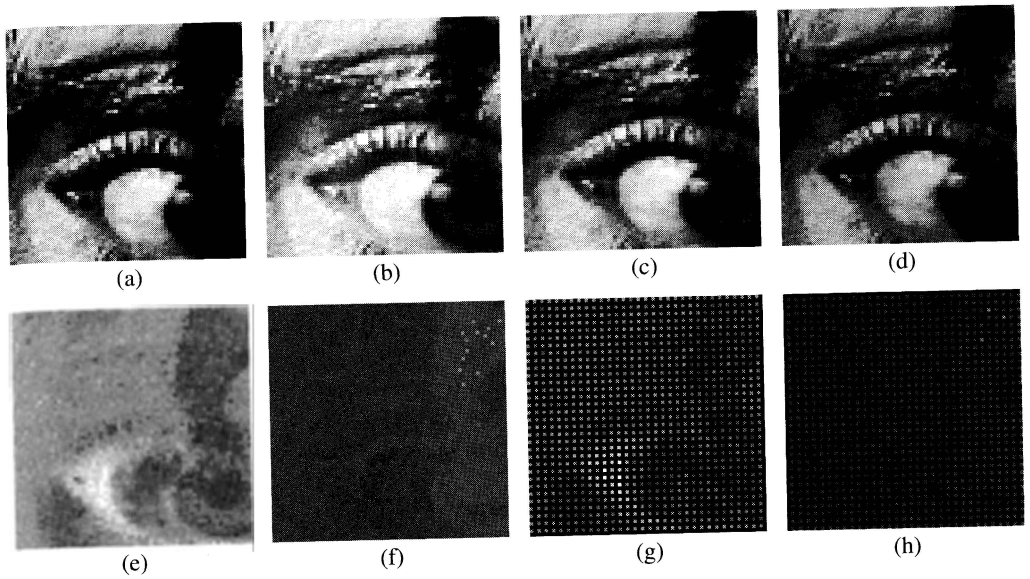

FIGURE 9.2

Zoomed portion of the *Clown* image: (a) gray-scale version of original color image, (b) decomposed red channel, (c) decomposed green channel, (d) decomposed blue channel, (e) difference image $x_1 - x_2$, (f) difference image $x_3 - x_2$, (g) subsampled version of $x_1 - x_2$, and (h) subsampled version of $x_3 - x_2$.

that it is a convex combination, we may then decompose $y(\boldsymbol{n})$ in the following manner:

$$\begin{aligned} y(\boldsymbol{n}) &= c_1(\boldsymbol{n})x_1(\boldsymbol{n}) + c_2(\boldsymbol{n})x_2(\boldsymbol{n}) + c_3(\boldsymbol{n})x_3(\boldsymbol{n}) \\ &= c_1(\boldsymbol{n})x_1(\boldsymbol{n}) + (1 - c_1(\boldsymbol{n}) - c_3(\boldsymbol{n}))x_2(\boldsymbol{n}) + c_3(\boldsymbol{n})x_3(\boldsymbol{n}) \\ &= c_1(\boldsymbol{n})(x_1(\boldsymbol{n}) - x_2(\boldsymbol{n})) + c_3(\boldsymbol{n})(x_3(\boldsymbol{n}) - x_2(\boldsymbol{n})) + x_2(\boldsymbol{n}) \\ &= c_1(\boldsymbol{n})\alpha(\boldsymbol{n}) + c_3(\boldsymbol{n})\beta(\boldsymbol{n}) + x_2(\boldsymbol{n}), \end{aligned} \tag{9.3}$$

where the difference images $\alpha(\boldsymbol{n}) = x_1(\boldsymbol{n}) - x_2(\boldsymbol{n})$ and $\beta(\boldsymbol{n}) = x_3(\boldsymbol{n}) - x_2(\boldsymbol{n})$ are crude approximations for the chrominance channels. In other words, the convex combination above can be thought of as the summation of $x_2(\boldsymbol{n})$ with the subsampled difference images, $c_1(\boldsymbol{n})\alpha(\boldsymbol{n})$ and $c_3(\boldsymbol{n})\beta(\boldsymbol{n})$; it is shown pictorially in Figure 9.2c, Figure 9.2g and Figure 9.2h, as their sum is equal to the sensor data in Figure 9.1b. It follows from the composition of the dyadic decimation and interpolation operators induced by the Bayer sampling pattern that $\hat{y}(\boldsymbol{\omega})$, the Fourier transform of sensor data $y(\boldsymbol{n})$, is a sum of $\hat{x}_2(\boldsymbol{\omega})$ and the spectral copies of $\hat{\alpha}(\boldsymbol{\omega})$ and $\hat{\beta}(\boldsymbol{\omega})$:

$$\begin{aligned} \hat{y}(\boldsymbol{\omega}) &= \hat{x}_2(\boldsymbol{\omega}) + \frac{1}{4}\Big((\hat{\alpha}+\hat{\beta})(\boldsymbol{\omega}) + (\hat{\alpha}-\hat{\beta})(\boldsymbol{\omega}-[\pi,0]^T) \\ &\quad + (\hat{\alpha}-\hat{\beta})(\boldsymbol{\omega}-[0,\pi]^T) + (\hat{\alpha}+\hat{\beta})(\boldsymbol{\omega}-[\pi,\pi]^T)\Big) \\ &= \hat{\ell}(\boldsymbol{\omega}) + \frac{1}{4}\Big((\hat{\alpha}-\hat{\beta})(\boldsymbol{\omega}-[\pi,0]^T) \\ &\quad + (\hat{\alpha}-\hat{\beta})(\boldsymbol{\omega}-[0,\pi]^T) + (\hat{\alpha}+\hat{\beta})(\boldsymbol{\omega}-[\pi,\pi]^T)\Big), \end{aligned} \tag{9.4}$$

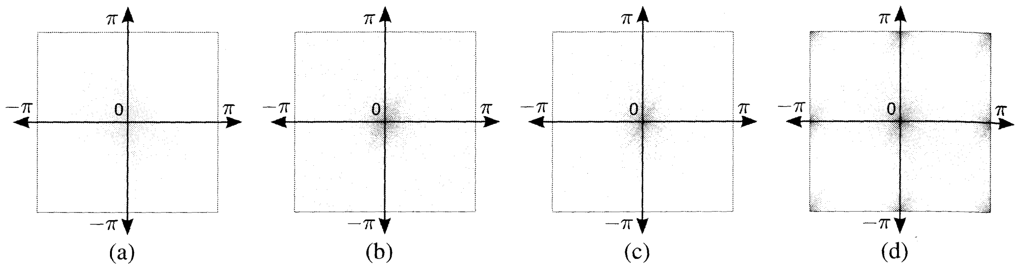

FIGURE 9.3

Log-magnitude two-dimensional spectra of: (a) $\hat{\ell}$, (b) $\hat{\alpha}$, (c) $\hat{\beta}$, and (d) $\hat{y}$. The spectra were obtained using the *Clown* image. The figure is color-coded to show contribution from each channel in figure (d): green for $\hat{\ell}$, red for $\hat{\alpha}$, blue for $\hat{\beta}$.

where, without loss of generality, the origin is fixed as $\boldsymbol{c}(0,0) = [1,0,0]^T$, and

$$\hat{\ell} = \hat{x}_2(\omega) + \frac{1}{4}\hat{\alpha}(\omega) + \frac{1}{4}\hat{\beta}(\omega) = \frac{1}{4}\hat{x}_1(\omega) + \frac{1}{2}\hat{x}_2(\omega) + \frac{1}{4}\hat{x}_3(\omega), \tag{9.5}$$

is a crude approximation to the luminance channel.

The representation of sensor data (Equation 9.4) in terms of luminance ℓ and difference images α and β is convenient because α and β are typically sparse in the Fourier domain. To see this, consider Figure 9.3, in which the log-magnitude spectra of a typical color image is shown. The high-frequency components, a well-accepted indicator for edges, object boundaries, and textures, are easily found in Figure 9.3a. In contrast, the spectra in Figure 9.3b and Figure 9.3b reveal that α and β are low-pass, which supports our earlier claim about the slowly-varying nature of the signals in Figure 9.2e and Figure 9.2f. It is typically easier to estimate a lower bandwidth signal from its sparsely subsampled versions (see Figure 9.2g and Figure 9.2h), since it is less subject to aliasing. The key observation that can be made in Equation 9.4, therefore, is that we expect a Fourier domain representation of sensor data similar to what is illustrated in Figure 9.3d — the spectral copies of $\hat{\alpha} - \hat{\beta}$ centered around $[\pi,0]^T$ and $[0,\pi]^T$ overlap with the baseband $\hat{\ell}$, while $\hat{\alpha}+\hat{\beta}$ centered around $[\pi,\pi]^T$ remain aliasing-free.

Note that there exists no straightforward global strategy such that we recover unaliased $\hat{\ell}$ because both spectral copies centered around $[\pi,0]^T$ and $[0,\pi]^T$ are aliased with the baseband $\hat{\ell}$. Dubois et al., however, emphasized that the local image features of the baseband, $\hat{\ell}$, exhibit a strong directional bias, and therefore either $(\hat{\alpha}-\hat{\beta})(\omega-[\pi,0]^T)$ or $(\hat{\alpha}-\hat{\beta})(\omega-[0,\pi]^T)$ is locally recoverable from the sensor data [47]. This observation motivates nonlinear processing that is locally adaptive — in fact, most existing demosaicking methods can be reexamined from this perspective. Specifically, Figure 9.4 illustrates the presumed local aliasing pattern. The locally horizontal images suffer from aliasing between $\hat{\ell}$ and $(\hat{\alpha}-\hat{\beta})(\omega-[\pi,0]^T)$ while we expect that $(\hat{\alpha}-\hat{\beta})(\omega-[0,\pi]^T)$ remains relatively intact. Conversely, locally vertical images suffer from aliasing between $\hat{\ell}$ and $(\hat{\alpha}-\hat{\beta})(\omega-[0,\pi]^T)$ while $(\hat{\alpha}-\hat{\beta})(\omega-[\pi,0]^T)$ is clean. On a sidenote, locally diagonal image features, which are often ignored by the demosaicking algorithm designs, do not interfere with $(\hat{\alpha}-\hat{\beta})(\omega-[\pi,0]^T)$ and $(\hat{\alpha}-\hat{\beta})(\omega-[0,\pi]^T)$, making the reconstruction of diagonal features a trivial task.

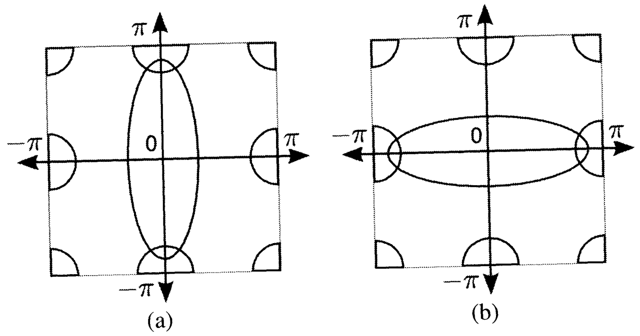

FIGURE 9.4

Presumed aliasing structure in local spectra, conditioned local image features of the surrounding. Images correspond to: (a) $\hat{y}$ given horizontal features, and (b) $\hat{y}$ given vertical features. Compare with Figure 9.3d.

Finally, let $z(\boldsymbol{n})$ be the noisy sensor data,

$$z(\boldsymbol{n}) = y(\boldsymbol{n}) + \varepsilon(\boldsymbol{n}) = c_1(\boldsymbol{n})\alpha(\boldsymbol{n}) + c_3(\boldsymbol{n})\beta(\boldsymbol{n}) + x_2(\boldsymbol{n}) + \varepsilon(\boldsymbol{n}), \tag{9.6}$$

where $\varepsilon \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma_\varepsilon^2)$. Recall that Fourier transform is a unitary transformation — a spatially white noise in space domain remains uncorrelated in the frequency representation. It follows that the Fourier transform of a noisy observation is

$$\begin{aligned}\hat{z}(\omega) = \hat{\ell}(\omega) + \frac{1}{4}\Big((\hat{\alpha}-\hat{\beta})(\omega-[\pi,0]^T)\\ +(\hat{\alpha}-\hat{\beta})(\omega-(0,\pi)^T) + (\hat{\alpha}+\hat{\beta})(\omega-[\pi,\pi]^T)\Big) + \hat{\varepsilon}(\omega).\end{aligned}$$

In other words, the sensor data is the baseband luminance image $\hat{\ell}$ distorted by the noise $\hat{\varepsilon}$ and aliasing due to spectral copies of $\hat{\alpha}$ and $\hat{\beta}$, where $\hat{\varepsilon}$, $\hat{\alpha}$, and $\hat{\beta}$ are conditionally normal. A unified strategy to demosaicking and denoising, therefore, is to design an estimator that suppresses noise and attenuates aliased components simultaneously. We will see how this can be accomplished via a spatially-adaptive linear filter whose stop-band contains the spectral copies of the difference images and pass-band suppresses noise (Section 9.5).

9.4 Wavelet Analysis of CFA Image

In the previous section, we established the inadequacy of taking the global approach to CFA image processing. In this section, we develop a time-frequency analysis framework to exploit the local aliasing structures [17]. Specifically, image signals are highly nonstationary/inhomogeneous and thus an orthogonal filterbank (or wavelet packet) expansion for sparsely sampled signal would prove useful.

For simplicity, consider first a one-dimensional signal $x : \mathbb{Z} \to \mathbb{R}$. A one-level filterbank structure defined by filters $\{h_0, h_1, f_0, f_1\}$ is shown in Figure 9.5. It is a linear transformation composed of convolution filters and decimators. The channel containing the

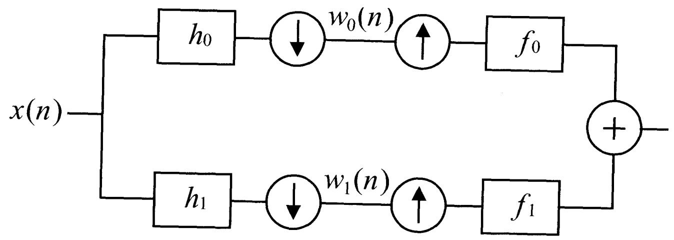

FIGURE 9.5
One-level filterbank structure.

low-frequency components is often called *approximation* (denoted $w_0^x(n)$), and the other containing the high-frequency components is referred to as the *detail* (denoted $w_1^x(n)$). The decomposition can be nested recursively to gain more precision in frequency. The approximation and detail coefficients from one-level decomposition can be analyzed in the Fourier domain as:

$$\hat{w}_i^x(\omega) = \frac{1}{2}\left(\hat{h}_i\left(\tfrac{\omega}{2}\right)\hat{x}\left(\tfrac{\omega}{2}\right) + \hat{h}_i\left(\tfrac{\omega}{2}-\pi\right)\hat{x}\left(\tfrac{\omega}{2}-\pi\right)\right),$$

where $i \in \{0,1\}$. With a careful choice of filters $\{h_0, h_1, f_0, f_1\}$, the original signal, $x(n)$ can be recovered exactly from the filterbank coefficients $w_0^x(n)$ and $w_1^x(n)$. To see this, consider the reconstruction of one-level filterbank, as in Figure 9.5. The transfer function of the system (or the reconstructed signal $x^{\mathrm{rec}}(n)$) has the following form in the frequency domain:

$$\begin{aligned}\hat{x}^{\mathrm{rec}}(\omega) &= \hat{f}_0(\omega)\hat{w}_0^x(2\omega) + \hat{f}_1(\omega)\hat{w}_1^x(2\omega)\\ &= \frac{1}{2}\left(\hat{f}_0(\omega)\hat{h}_0(\omega) + \hat{f}_1(\omega)\hat{h}_1(\omega)\right)\hat{x}(\omega)\\ &\quad + \frac{1}{2}\left(\hat{f}_0(\omega)\hat{h}_0(\omega-\pi) + \hat{f}_1(\omega)\hat{h}_1(\omega-\pi)\right)\hat{x}(\omega-\pi).\end{aligned}$$

In other words, the output is a linear combination of the filtered versions of the signal $\hat{x}(\omega)$ and a frequency-modulated signal $\hat{x}(\omega-\pi)$. The structure in Figure 9.5 is called a perfect reconstruction filterbank if

$$\begin{aligned}\hat{f}_0(\omega)\hat{h}_0(\omega) + \hat{f}_1(\omega)\hat{h}_1(\omega) &= 2\\ \hat{f}_0(\omega)\hat{h}_0(\omega-\pi) + \hat{f}_1(\omega)\hat{h}_1(\omega-\pi) &= 0.\end{aligned}$$

The filters corresponding to $\hat{x}(\omega)$ constitute a constant, whereas the filters corresponding to the aliased version are effectively a zero.

A large body of literature exists on designing a set of filters $\{h_0, h_1, f_0, f_1\}$ that comprise a perfect reconstruction filterbank [48]. For example, wavelet packets belong to a class of filterbanks arising from the factorizing filters satisfying the Nyquist condition (Smith-Barnwell [48]). In this case, the following are met by construction:

$$\begin{aligned}\hat{h}_1(\omega) &= -e^{-j\omega m}\hat{h}_0(-\omega-\pi)\\ \hat{f}_0(\omega) &= \hat{h}_1(\omega-\pi)\\ \hat{f}_1(\omega) &= -\hat{h}_0(\omega-\pi).\end{aligned} \tag{9.7}$$

In other words, h_1 is a *time-shifted, time-reversed,* and *frequency-modulated* version of h_0; and f_0 and f_1 are *time-reversed* versions of h_0 and h_1, respectively. Derivation of these filters is beyond of the scope of this chapter, and interested readers are referred to Reference [48] for details.

Define modulated signal and subsampled signal of $x(n)$, respectively, as

$$\begin{aligned}x_m(n) &= (-1)^n x(n)\\ x_s(n) &= \frac{1}{2}\left(x(n) + x_m(n)\right) = \begin{cases} x(n) & \text{for even } n\\ 0 & \text{for odd } n.\end{cases}\end{aligned}$$

To derive an explicit filterbank representation of $x_s(n)$, we are interested in characterizing the relationship between filterbank coefficients of $x(n)$ and $x_m(n)$. Let $w_0^{x_m}(n)$ and $w_1^{x_m}(n)$ be the approximation and detail coefficients of the one-level filterbank decomposition of $(-1)^n x(n)$. Then substituting into Equation 9.7 we obtain

$$\begin{aligned}\hat{w}_0^{x_m}(\omega) &= \frac{1}{2}\left(\hat{h}_0\left(\tfrac{\omega}{2}\right)\hat{x}\left(\tfrac{\omega}{2}-\pi\right) + \hat{h}_0\left(\tfrac{\omega}{2}-\pi\right)\hat{x}\left(\tfrac{\omega}{2}\right)\right)\\ &= \frac{1}{2}\left(e^{-jm\frac{\omega}{2}}\hat{h}_1\left(-\tfrac{\omega}{2}-\pi\right)\hat{x}\left(\tfrac{\omega}{2}-\pi\right) + e^{-jm\left(\frac{\omega}{2}-\pi\right)}\hat{h}_1\left(-\tfrac{\omega}{2}\right)\hat{x}\left(\tfrac{\omega}{2}\right)\right)\\ &= \frac{e^{-jm\frac{\omega}{2}}}{2}\left(\hat{h}_1^*\left(\tfrac{\omega}{2}-\pi\right)\hat{x}\left(\tfrac{\omega}{2}-\pi\right) - \hat{h}_1^*\left(\tfrac{\omega}{2}\right)\hat{x}\left(\tfrac{\omega}{2}\right)\right)\end{aligned}$$

$$\begin{aligned}\hat{w}_1^{x_m}(\omega) &= \frac{1}{2}\left(\hat{h}_1\left(\tfrac{\omega}{2}\right)\hat{x}\left(\tfrac{\omega}{2}-\pi\right) + \hat{h}_1\left(\tfrac{\omega}{2}-\pi\right)\hat{x}\left(\tfrac{\omega}{2}\right)\right)\\ &= \frac{1}{2}\left(-e^{-jm\frac{\omega}{2}}\hat{h}_0\left(-\tfrac{\omega}{2}-\pi\right)\hat{x}\left(\tfrac{\omega}{2}-\pi\right) - e^{-jm\left(\frac{\omega}{2}-\pi\right)}\hat{h}_0\left(-\tfrac{\omega}{2}\right)\hat{x}\left(\tfrac{\omega}{2}\right)\right)\\ &= \frac{e^{-jm\frac{\omega}{2}}}{2}\left(-\hat{h}_0^*\left(\tfrac{\omega}{2}-\pi\right)\hat{x}\left(\tfrac{\omega}{2}-\pi\right) + \hat{h}_0^*\left(\tfrac{\omega}{2}\right)\hat{x}\left(\tfrac{\omega}{2}\right)\right),\end{aligned}$$

where m is an odd integer, and $*$ denotes the complex conjugation. A subtle but important detail of the equations above is that if the approximation and detail coefficients of $x(n)$ were computed using $h_0(-n-m)$ and $h_1(-n-m)$ instead of $h_0(n)$ and $h_1(n)$, these coefficients behave exactly like the *detail* ($w_1^{x_m}(n)$) and *approximation* ($w_0^{x_m}(n)$) coefficients for $(-1)^n x(n)$, respectively (note the *reversed* ordering of detail and approximation). It is straightforward to verify that if $\{h_0(n), h_1(n)\}$ comprise perfect reconstruction filterbank, then $\{h_0(-n-m), h_1(-n-m)\}$ constitute a legitimate perfect reconstruction filterbank as well (we will refer to the latter as the time-reversed filterbank). Reversal of coefficients is illustrated in Figure 9.6 — the systems in Figure 9.6a and Figure 9.6b are equivalent.

Restricting our attention to the Haar decomposition for the rest of discussion and fixing $m = 1$, we have that $h_0(n) = h_0(-n-1)$ and $h_1(n) = -h_1(-n-1)$ and the approximation coefficient of $(-1)^n x(n)$ is exactly equal to the detail coefficient of $x(n)$ by construction, and vice-versa — i.e., $w_0^{x_m}(n) = w_1^x(n)$ and $w_1^{x_m}(n) = w_0^x(n)$. It follows that the multi-level filterbank decomposition of $(-1)^n x(n)$ is equivalent to the time-reversed filterbank decomposition of $x(n)$, but with the reversed ordering of low-to-high frequency coefficients. This reversed-order filterbank can be used to derive the filterbank representation of $x_s(n)$.

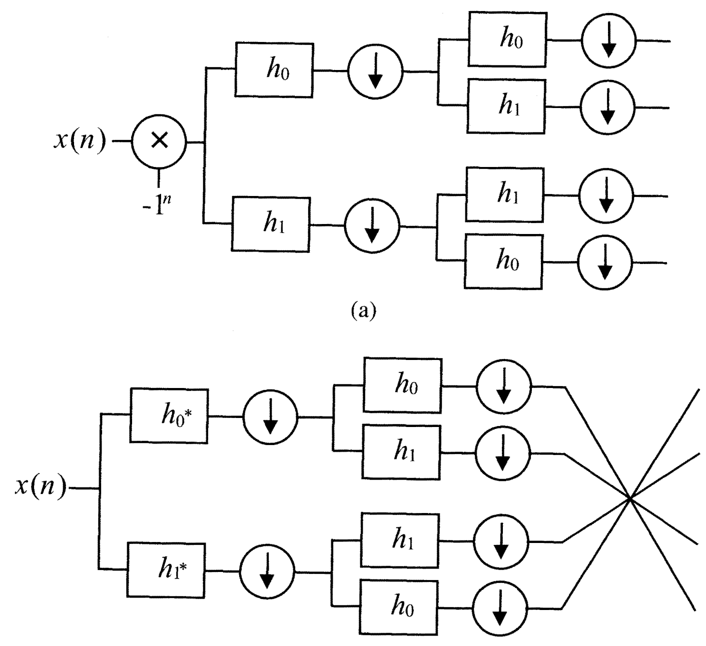

FIGURE 9.6

Two equivalent filterbanks for $x_m(n) = (-1)^n x(n)$: (a) filterbank transform of x_m, (b) reversed-order filterbank transform of x. Here, * indicates time-reversed filter coefficients.

Specifically, let $w_0^{x_s}(n)$ and $w_1^{x_s}(n)$ be the approximation and detail coefficients of the one-level filterbank decomposition of $x_s(n)$. Then

$$w_0^{x_s}(n) = w_0^{1/2(x+x_m)}(n) = \frac{1}{2}\left(w_0^{x}(n) + w_0^{x_m}(n)\right) = \frac{1}{2}\left(w_0^{x}(n) + w_1^{x}(n)\right)$$
$$w_1^{x_s}(n) = w_1^{1/2(x+x_m)}(n) = \frac{1}{2}\left(w_1^{x}(n) + w_1^{x_m}(n)\right) = \frac{1}{2}\left(w_1^{x}(n) + w_0^{x}(n)\right) = w_0^{x_s}(n).$$

Now, update the definition of w_i^x to mean the i-th subband of $(I+1)$-level filterbank decomposition. Then by recursion, we have a general form

$$w_i^{x_s}(n) = \frac{1}{2}\left(w_i^{x}(n) + w_{I-i}^{x}(n)\right). \tag{9.8}$$

Also see Figure 9.7. Equation 9.8 should not come as a surprise, as it is analogous to the Fourier domain aliasing where the high frequency component is summed to the low. Similar analysis for x_s can be performed for nonHaar decompositions, but omitted here for simplicity.

Extending to two-dimensional signals, let us show the decomposition of CFA image in the separable wavelet packet domain. Let $w_{\boldsymbol{i}}^{\ell}(\boldsymbol{n}), w_{\boldsymbol{i}}^{\alpha}(\boldsymbol{n}), w_{\boldsymbol{i}}^{\beta}(\boldsymbol{n})$ be the filterbank coefficients corresponding to $\ell(\boldsymbol{n}), \alpha(\boldsymbol{n}), \beta(\boldsymbol{n})$, respectively, where $\boldsymbol{i} = [i_0, i_1]^T \in \{0, 1, \ldots, I\}^2$ indexes the horizontal and the vertical filterbank channels, respectively. As before, assume without loss of generality that $\boldsymbol{c}(0,0) = [1,0,0]^T$. In order to apply the filterbank analysis to the

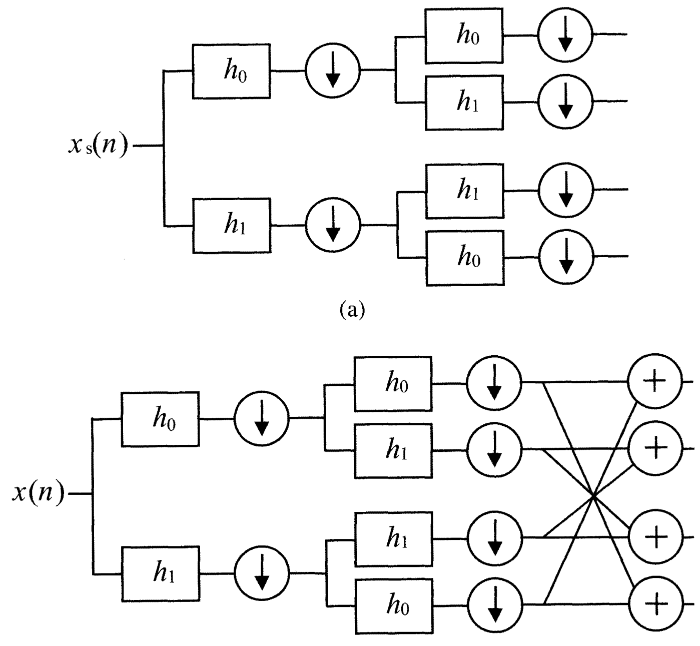

FIGURE 9.7

Two equivalent filterbanks for $x_s = \frac{1}{2}(x + x_m)$; up to multiplicative constant 2: (a) filterbank transform of x_s, and (b) ordinary and reversed-order filterbank transform of x. Here, we assume the Haar decomposition.

sensor data, we re-write $y(\boldsymbol{n})$ in the following manner:

$$\begin{aligned} y(\boldsymbol{n}) &= x_2(\boldsymbol{n}) + c_1(\boldsymbol{n})\alpha(\boldsymbol{n}) + c_3(\boldsymbol{n})\beta(\boldsymbol{n}) \\ &= x_2(\boldsymbol{n}) + \left(1 + (-1)^{n_0} + (-1)^{n_1} + (-1)^{n0+n1}\right)\frac{\alpha(\boldsymbol{n})}{4} \\ &\quad + \left(1 + (-1)^{n_0+1} + (-1)^{n_1+1} + (-1)^{n0+n1}\right)\frac{\beta(\boldsymbol{n})}{4}, \end{aligned}$$

and its corresponding filterbank representation:

$$\begin{aligned} w_{\boldsymbol{i}}^{y}(\boldsymbol{n}) &= w_{\boldsymbol{i}}^{x_2}(\boldsymbol{n}) + \frac{1}{4}\left(w_{\boldsymbol{i}}^{\alpha}(\boldsymbol{n}) + w_{(i_0,I-i_1)}^{\alpha}(\boldsymbol{n}) + w_{(I-i_0,i_1)}^{\alpha}(\boldsymbol{n}) + w_{(I-i_0,I-i_1)}^{\alpha}(\boldsymbol{n})\right) \\ &\quad + \frac{1}{4}\left(w_{\boldsymbol{i}}^{\beta}(\boldsymbol{n}) - w_{(i_0,I-i_1)}^{\beta}(\boldsymbol{n}) - w_{(I-i_0,i_1)}^{\beta}(\boldsymbol{n}) + w_{(I-i_0,I-i_1)}^{\beta}(\boldsymbol{n})\right) \\ &= w_{\boldsymbol{i}}^{\ell}(\boldsymbol{n}) + \frac{1}{4}\left(w_{(i_0,I-i_1)}^{\alpha}(\boldsymbol{n}) + w_{(I-i_0,i_1)}^{\alpha}(\boldsymbol{n}) + w_{(I-i_0,I-i_1)}^{\alpha}(\boldsymbol{n})\right) \\ &\quad \frac{1}{4}\left(-w_{(i_0,I-i_1)}^{\beta}(\boldsymbol{n}) - w_{(I-i_0,i_1)}^{\beta}(\boldsymbol{n}) + w_{(I-i_0,I-i_1)}^{\beta}(\boldsymbol{n})\right), \end{aligned}$$

where the minus signs in some w^{β} terms occur due to translation in space, and $w^{\ell}(\boldsymbol{n})$ is the filterbank coefficients of the signal in Equation 9.5. The globally bandlimitedness of difference images, as argued in the previous section, allows us to conclude that $w_{\boldsymbol{i}}^{\alpha}(\boldsymbol{n}) \approx 0$ and $w_{\boldsymbol{i}}^{\beta}(\boldsymbol{n}) \approx 0$, $\forall i_0 > \hat{I}$ or $i_1 > \hat{I}$ for some $\hat{I}$. The above simplifies to a form that reveals the energy compaction structure within CFA image:

$$w_i^y(\boldsymbol{n}) \approx \begin{cases} w_i^\ell(\boldsymbol{n}) + \left(w^\alpha_{(I-i_0,i_1)}(\boldsymbol{n}) - w^\beta_{(I-i_0,i_1)}(\boldsymbol{n})\right)/4 & \text{if } I-i_0<\hat{I},\, i_1<\hat{I} \\ w_i^\ell(\boldsymbol{n}) + \left(w^\alpha_{(i_0,I-i_1)}(\boldsymbol{n}) - w^\beta_{(i_0,I-i_1)}(\boldsymbol{n})\right)/4 & \text{if } i_0<\hat{I},\, I-i_1<\hat{I} \\ w_i^\ell(\boldsymbol{n}) + \left(w^\alpha_{(I-i_0,I-i_1)}(\boldsymbol{n}) + w^\beta_{(I-i_0,I-i_1)}(\boldsymbol{n})\right)/4 & \text{if } I-i_0<\hat{I},\, I-i_1<\hat{I} \\ w_i^\ell(\boldsymbol{n}) & \text{otherwise} \end{cases} \tag{9.9}$$

Recall Equation 9.2 and that the filterbank transforms with appropriate choices of filters constitute a unitary transform. Thus, $w_i^z(\boldsymbol{n}) = w_i^y(\boldsymbol{n}) + w_i^\varepsilon(\boldsymbol{n})$, providing

$$w_i^z(\boldsymbol{n}) \approx \begin{cases} w_i^\ell(\boldsymbol{n}) + \left(w^\alpha_{(I-i_0,i_1)}(\boldsymbol{n}) - w^\beta_{(I-i_0,i_1)}(\boldsymbol{n})\right)/4 + w_i^\varepsilon(\boldsymbol{n}) & \text{if } I-i_0<\hat{I},\, i_1<\hat{I} \\ w_i^\ell(\boldsymbol{n}) + \left(w^\alpha_{(i_0,I-i_1)}(\boldsymbol{n}) - w^\beta_{(i_0,I-i_1)}(\boldsymbol{n})\right)/4 + w_i^\varepsilon(\boldsymbol{n}) & \text{if } i_0<\hat{I},\, I-i_1<\hat{I} \\ w_i^\ell(\boldsymbol{n}) + \left(w^\alpha_{(I-i_0,I-i_1)}(\boldsymbol{n}) + w^\beta_{(I-i_0,I-i_1)}(\boldsymbol{n})\right)/4 + w_i^\varepsilon(\boldsymbol{n}) & \text{if } I-i_0<\hat{I},\, I-i_1<\hat{I} \\ w_i^\ell(\boldsymbol{n}) + w_i^\varepsilon(\boldsymbol{n}) & \text{otherwise,} \end{cases} \tag{9.10}$$

where $w_i^\varepsilon \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma_\varepsilon^2)$ is a filterbank transform of $\varepsilon(\boldsymbol{n})$. In other words, the filterbank transformation of noisy sensor data w^z is the baseband luminance coefficient w^ℓ distorted by the noise w^ε and aliasing due to reversed-order filterbank coefficients w^α and w^β, where w^ℓ, w^α, and w^β are (conditionally) normal. A unified strategy to demosaicking and denoising, therefore, is to design an estimator that estimates w^ℓ, w^α, and w^β from the mixture of w^ℓ, w^α, w^β. and w^ε. We will see how this can be accomplished in Section 9.7.

Lastly, we remind the readers that Equation 9.10 can be generalized to any filterbanks that satisfy Equation 9.7 using time-reversed filter coefficients for h_0 and h_1. However, Haar wavelets are used exclusively in this chapter to simplify the notation.

9.5 Constrained Filtering

In this section, we motivate an approach to joint demosaicking and denoising using well-understood DSP machineries [40]. Recall Equations 9.3 and 9.4. We are interested in estimating $\boldsymbol{x}(\boldsymbol{n})$ given $z(\cdot)$. It is worth noting that even if $z(\boldsymbol{n})$ for some $\boldsymbol{n}$ corresponds to an observation of a red pixel $x_1(\boldsymbol{n})$, for example, $z(\boldsymbol{n})$ does not suffice as an estimate of $x_1(\boldsymbol{n})$ (unlike the pure demosaicking problems) because it is contaminated by noise.

We begin by highlighting monochromatic image denoising methods that operate by taking a linear combination of neighboring pixels. These methods include bilateral filters [28], principal components [31], and total least squares based methods [32], where the linear weights adapt to the local image features. Transform-based shrinkage and threshold methods can also be re-interpreted as spatially-varying linear estimators, because there exists a linear combination of neighboring pixels that is equivalent to shrinkage of transform coefficients. In the Bayesian estimation framework, the linearity of estimation is (conditionally) true for (a mixture of) normally distributed transform coefficients. In any case, the

locally adaptive linear estimator, x^{est}, takes the general form:

$$x^{\text{est}}(\boldsymbol{n}) = \sum_{\boldsymbol{m}\in\eta(\boldsymbol{n})} g(\boldsymbol{n},\boldsymbol{m}) z(\boldsymbol{n}-\boldsymbol{m}),$$

where $z(\boldsymbol{n})$ is the noisy version of $x(\boldsymbol{n})$, $g(\boldsymbol{n},\boldsymbol{m})$ is the spatially-adaptive linear weights, and the summation is over $\eta(\boldsymbol{n})$, a local neighborhood of pixels centered around $\boldsymbol{n}$. Typically, we choose $g(\boldsymbol{n},\boldsymbol{m})$ such that it solves the least-squares minimization problem (though not necessarily [32]),

$$\min_g E\left\|x(\boldsymbol{n}) - x^{\text{est}}(\boldsymbol{n})\right\|^2. \tag{9.11}$$

In this section, we will show how the estimator in the above form can be modified such that the linear weights can be used to simultaneously interpolate and denoise CFA data [40]. Let $\boldsymbol{x}^{\text{est}}(\boldsymbol{n})$ be an estimate of ideal color image $\boldsymbol{x}(\boldsymbol{n})$ by taking a linear combination of noisy sensor data $z(\boldsymbol{n})$. That is,

$$\boldsymbol{x}^{\text{est}}(\boldsymbol{n}) = \sum_{\boldsymbol{m}\in\eta(\boldsymbol{n})} \boldsymbol{g}(\boldsymbol{n},\boldsymbol{m}) z(\boldsymbol{n}-\boldsymbol{m}), \tag{9.12}$$

where $\boldsymbol{g}(\boldsymbol{n},\boldsymbol{m}) \in \mathbb{R}^3$ is a spatially-adaptive linear weight.

Let $g_k(\boldsymbol{n},\boldsymbol{m})$ correspond to the linear weight for estimating x_k. In the following discussion, we focus on the estimation of $x_2(\boldsymbol{n})$ via the design of $g_2(\boldsymbol{n},\cdot)$ because Equation 9.3 already assumes $x_2(\boldsymbol{n})$ as its baseband. The results achieved here are generalized to the estimation of $x_1(\boldsymbol{n})$ and $x_3(\boldsymbol{n})$ at the end of this section. Substituting Equation 9.6 into Equation 9.12,

$$\begin{aligned} x_2^{\text{est}}(\boldsymbol{n}) &= \sum_{\boldsymbol{m}\in\eta(\boldsymbol{n})} g_2(\boldsymbol{n},\boldsymbol{m}) z(\boldsymbol{n}-\boldsymbol{m}) \\ &= \sum_{\boldsymbol{m}\in\eta(\boldsymbol{n})} \Big(g_2(\boldsymbol{n},\boldsymbol{m})\big(x_2(\boldsymbol{n}-\boldsymbol{m}) + \varepsilon(\boldsymbol{n}-\boldsymbol{m})\big) \\ &\quad + g_2(\boldsymbol{n},\boldsymbol{m})\big(c_1(\boldsymbol{n}-\boldsymbol{m})\alpha(\boldsymbol{n}-\boldsymbol{m}) + c_3(\boldsymbol{n}-\boldsymbol{m})\beta(\boldsymbol{n}-\boldsymbol{m})\big)\Big). \end{aligned} \tag{9.13}$$

The first term, $\sum_{\boldsymbol{m}} g_2(\boldsymbol{n},\boldsymbol{m})[x_2(\boldsymbol{n}-\boldsymbol{m}) + \varepsilon(\boldsymbol{n}-\boldsymbol{m})]$ represents an ordinary monochromatic image denoising. That is, $g_2(\cdot,\cdot)$ operates on the noisy version of $x_2(\cdot)$. The extra term involving $\alpha(\cdot)$ and $\beta(\cdot)$ also motivates the need for further restricting $g_2(\cdot,\cdot)$ such that the latter term is attenuated.

To accomplish this task, recall that $c_1(\boldsymbol{n})\alpha(\boldsymbol{n})$ and $c_3(\boldsymbol{n})\beta(\boldsymbol{n})$ occupy frequency regions around $\omega = \{(0,0),(0,\pi),(\pi,0),(\pi,\pi)\}$. Let us consider a class of linear filters with stop-bands near $\{(0,0),(0,\pi),(\pi,0),(\pi,\pi)\}$ (i.e., band-pass). In particular, if the filter coefficients corresponding to red and blue samples in CFA sum to zero, respectively, then $\forall\boldsymbol{n}$,

$$\begin{aligned} \sum_{\boldsymbol{m}\in\eta(\boldsymbol{n})} g_2(\boldsymbol{n},\boldsymbol{m}) c_1(\boldsymbol{n}-\boldsymbol{m}) &= 0 \\ \sum_{\boldsymbol{m}\in\eta(\boldsymbol{n})} g_2(\boldsymbol{n},\boldsymbol{m}) c_3(\boldsymbol{n}-\boldsymbol{m}) &= 0, \end{aligned} \tag{9.14}$$

and with a finite spatial support on $g_2(\cdot,\boldsymbol{m})$, we can safely assume that the frequency components in the vicinity of $\{(0,0),(0,\pi),(\pi,0),(\pi,\pi)\}$ are attenuated as well (because $\hat{g}_2$ is a linear combination of cosines in the Fourier domain).

If the restriction in Equation 9.14 holds true, then the estimator in Equation 9.13 reduces to a monochromatic image denoising problem — that is, $x_2^{\text{est}} \approx \sum_m g_2(\boldsymbol{n},\boldsymbol{m})[x_2(\boldsymbol{n}-\boldsymbol{m})+\varepsilon(\boldsymbol{n}-\boldsymbol{m})]$. Therefore, the underlying strategy for deriving a joint demosaicking and denoising operator is to solve a constrained linear estimation problem. In other words, instead of Equation 9.11, solve

$$J = \min E\left\|x_2(\boldsymbol{n}) - \sum_{\boldsymbol{m}\in\eta(\boldsymbol{n})} g_2(\boldsymbol{n},\boldsymbol{m})[x_2(\boldsymbol{n}-\boldsymbol{m})+\varepsilon(\boldsymbol{n}-\boldsymbol{m})]\right\|^2. \tag{9.15}$$
$$\text{subject to } \sum_{\boldsymbol{m}} g_2(\boldsymbol{n},\boldsymbol{m})c_1(\boldsymbol{n}-\boldsymbol{m}) = \sum_{\boldsymbol{m}} g_2(\boldsymbol{n},\boldsymbol{m})c_3(\boldsymbol{n}-\boldsymbol{m}) = 0$$

Conveniently, this optimization problem allows us to *pretend* as though we are designing a monochromatic image denoising method. However, the constraints on the filter coefficients ensure that J remains a good approximation to the residual of the actual problem, $\|x_2(\boldsymbol{n}) - x_2^{\text{est}}(\boldsymbol{n})\|^2$. Note that Equation 9.14 does not imply $\sum_m g_2(\boldsymbol{n},\boldsymbol{m})c_2(\boldsymbol{n}-\boldsymbol{m}) = 0$. Instead, the contributions from x_1 and x_3 to the estimation of x_2 are limited to the frequency components in the band-pass region, whereas the contributions from x_2 are unrestricted.

In many cases, the existing image denoising techniques naturally extend to simultaneously solving the demosaicking and denoising problems. Let $\vec{x}$ (and similarly $\vec{\varepsilon},\vec{z},\vec{g}$) be a re-arrangement of $\{x(\boldsymbol{n}-\boldsymbol{m})|\boldsymbol{m}\in\eta(\boldsymbol{n})\}$ into a vector form. Then least-squares solution to Equation 9.11 often involves an inner product of the form $x^{\text{est}}(\boldsymbol{n}) = \vec{g}_{\text{LS}}^T(\vec{x}+\vec{\varepsilon})$, where

$$\vec{g}_{\text{LS}} = E\left[(\vec{x}+\vec{\varepsilon})(\vec{x}+\vec{\varepsilon})^T\right]^{-1} E\left[(\vec{x}+\vec{\varepsilon})^T x(\boldsymbol{n})\right]. \tag{9.16}$$

The inner product occurs often in Bayesian estimators, when the prior on the data are (conditionally) normally distributed (e.g., Laplace, Student's t, Gaussian mixture). If this prior on x is defined in the linear transform domain (such as on the wavelet coefficients), then the equivalent second-order statistics in the pixel domain are simply a linear transformation of the statistics in the transform domain.

Let $M = |\eta(\boldsymbol{n})|$ be the size of the neighborhood, $\eta(\boldsymbol{n})$. The band-pass constraint in Equation 9.14 may be imposed by asserting that $\vec{g}\in\mathbb{R}^M$ lives in a lower-dimensional subspace, $\text{span}\{\vec{v}\in\mathbb{R}^M|\vec{v}^T\vec{c}_1 = \vec{v}^T\vec{c}_3 = 0\}$, or $\vec{g} = \boldsymbol{G}\boldsymbol{s}$, where $\boldsymbol{G}\in\mathbb{R}^{M\times M-2}$ is an orthogonal matrix whose column vectors span this subspace. Then the constrained LS problem in Equation 9.15 can be rewritten as

$$J = \min_{\vec{g}=\boldsymbol{G}\boldsymbol{s}} E\left\|\vec{x}_2 - \vec{g}^T[\vec{x}_2+\vec{\varepsilon}]\right\|^2 = \min_{\boldsymbol{s}} E\left\|\vec{x}_2 - \boldsymbol{s}^T\boldsymbol{G}^T[\vec{x}_2+\vec{\varepsilon}]\right\|^2. \tag{9.17}$$

It is easy to verify that the solution to the above has the form $x_2^{\text{est}}(\boldsymbol{n}) = \vec{g}_{\text{CLS}}^T\vec{z}$, where

$$\vec{g}_{\text{CLS}} = \boldsymbol{G}\Big[\boldsymbol{G}^T E\left[(\vec{x}_2+\vec{\varepsilon})(\vec{x}_2+\vec{\varepsilon})^T\right]\boldsymbol{G}\Big]^{-1} E\left[\boldsymbol{G}(\vec{x}_2+\vec{\varepsilon})^T x_2(\boldsymbol{n})\right]. \tag{9.18}$$

Note that Equations 9.17 and 9.18 are minor alterations to Equations 9.11 and 9.16 using the same second-order statistics, respectively, and thus it is a straightforward exercise to

leverage existing monochromatic image denoising methods to a joint demosaicking and denoising scheme.

In order to design spatially adaptive filters similar to Equation 9.18 for estimating x_1 and x_3, we see that Equation 9.3 can be written alternatively as

$$\begin{aligned} y(\boldsymbol{n}) &= c_2(\boldsymbol{n})[x_2(\boldsymbol{n}) - x_1(\boldsymbol{n})] + c_3(\boldsymbol{n})[x_3(\boldsymbol{n}) - x_1(\boldsymbol{n})] + x_1(\boldsymbol{n}) \\ &= c_1(\boldsymbol{n})[x_1(\boldsymbol{n}) - x_3(\boldsymbol{n})] + c_2(\boldsymbol{n})[x_2(\boldsymbol{n}) - x_3(\boldsymbol{n})] + x_3(\boldsymbol{n}). \end{aligned}$$

It follows that the appropriate constraints on filter coefficients $g_1(\cdot,\cdot)$ and $g_3(\cdot,\cdot)$ are

$$\begin{aligned} &\sum_{\boldsymbol{m}\in\eta(\boldsymbol{n})} g_1(\boldsymbol{n},\boldsymbol{m})c_2(\boldsymbol{n}-\boldsymbol{m}) = 0, \qquad \sum_{\boldsymbol{m}\in\eta(\boldsymbol{n})} g_1(\boldsymbol{n},\boldsymbol{m})c_3(\boldsymbol{n}-\boldsymbol{m}) = 0 \\ &\sum_{\boldsymbol{m}\in\eta(\boldsymbol{n})} g_3(\boldsymbol{n},\boldsymbol{m})c_1(\boldsymbol{n}-\boldsymbol{m}) = 0, \qquad \sum_{\boldsymbol{m}\in\eta(\boldsymbol{n})} g_3(\boldsymbol{n},\boldsymbol{m})c_2(\boldsymbol{n}-\boldsymbol{m}) = 0. \end{aligned}$$

9.6 Missing Data

The statistical modelling of image signals in a linear transform domain is primarily motivated by the correlation structures that exist within the transform coefficients of image signals. These models, which require a complete observation of image data, are not easily generalizable to the digital camera context, as the observation of color image data is incomplete at the sensor interface. That is, processing with missing or incomplete pixels is difficult because a linear transformation takes a linear combination of the pixel values, and thus *all* of the noisy transform coefficients are unobserved. Yet, it is still convenient or desirable to apply the sophisticated statistical modelling techniques even when none of the transform coefficients are observable.

This section explicitly addresses the issue of combining the treatment of missing data and the wavelet-based modeling [49]. Bayesian hierarchical modelling is used to capture the second-order statistics in the transform domain. We assume a general model form and couple the EM algorithm framework with the Bayesian models to estimate the hyper- and nuisance parameters via the marginal likelihood; that is, we adopt the empirical partial Bayes approach. Within this framework, problems with missing pixels or pixel components, and hence unobservable wavelet coefficients, are handled simultaneously with image denoising.

In order to extend the complete image modelling strategy to incomplete data, let $\boldsymbol{w}_i^x(\boldsymbol{n}) = [w_i^{x_1}(\boldsymbol{n}), w_i^{x_2}(\boldsymbol{n}), w_i^{x_3}(\boldsymbol{n})]^T$ correspond to wavelet coefficients corresponding to x_1,x_2,x_3 at i-th level, and assume that

$$\boldsymbol{w}_i^x(\boldsymbol{n}) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\Sigma_i).$$

Then the distribution of the neighboring pixel values is also jointly normal, as linear transformation of a multivariate normal vector is also normal. Because the wavelet transform is unitary,

$$\boldsymbol{w}^{x+\epsilon}(\boldsymbol{n})|\boldsymbol{w}^x(\boldsymbol{n}) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\boldsymbol{w}^x(\boldsymbol{n}),\sigma_\epsilon^2\boldsymbol{I}).$$

To summarize, $\theta = \{\Sigma_i, \sigma_\epsilon^2\}$ are the hyper- and nuisance parameters, respectively.

If the θ is known, the regression of the missing pixels on the known clean pixel-component measurements $y(n)$, $E[x(n)|y(n),\theta]$, serves as a demosaicking method based on the LS estimator and has a straightforward implementation. Conditioned on the incomplete and noisy measurement of pixel-components $z(n)$, $E[x(n)|z(n),\theta]$, is an interpolated and denoised image signal, where

$$x^{\text{est}}(n) = E[x(n)|z(n),\theta] = E\left[E[x(n)|y(n),z(n),\theta]\Big|z(n),\theta\right].$$

The nested expectation operator has an intuitive interpretation: the inner expectation, $E[x(n)|y(n),z(n),\theta] = E[x(n)|y(n),\theta]$, is an interpolator, and the outer expectation is the denoiser. Conversely, the same formula can equivalently be written as:

$$x^{\text{est}}(n) = E[x(n)|z(n),\theta] = E\left[E[x(n)|(x+\epsilon)(n),z(n),\theta]\Big|z(n),\theta\right],$$

where the inner expectation operator, $E[x(n)|(x+\epsilon)(n),z(n),\theta] = E[x(n)|(x+\epsilon)(n),\theta]$ is a denoiser, and the outer expectation is the interpolator. Conditioned on θ, therefore, a design of simultaneous demosaicking and denoising method is straightforward.

The posterior mean estimate, x^{est}, is sensitive to the choice of parameters θ; and given only a subset of the noisy pixel components $z(n)$, we are left with estimating θ from the data when the wavelet coefficients are not observable. In particular, we solve for the θ that maximizes the marginal log-likelihood $\log p(z|\theta)$, and estimate x as its posterior mean conditioned on $\hat{\theta}$ (where $\hat{\theta}$ is obtained from the maximal likelihood estimate above). The direct maximization of $\log p(z|\theta)$ is very difficult because of the missing pixel values. The EM algorithm circumvents this problem by iteratively maximizing the much easier *augmented-data* log-likelihood, $\log p(x,\epsilon|\theta)$, where $\{x,\epsilon\}$ are the augmented data.

Given the $[t]$-th iterate hyper- and nuisance parameter estimate, $\theta^{[t]} = \{\Sigma_i^{[t]}, \sigma_\epsilon^{2[t]}\}$, the $[t+1]$-st iteration of the EM algorithm first calls for

$$Q\left(\theta;\theta^{[t]}\right) = E\left[\log p(x,\epsilon|\theta)\Big|z,\theta^{[t]}\right].$$

A celebrated result of EM algorithm [50] states that

$$\log p(z|\theta) - \log p(z|\theta^{[t]}) \geq Q\left(\theta;\theta^{[t]}\right) - Q\left(\theta^{[t]};\theta^{[t]}\right),$$

where $\log p(z|\theta)$ is the log-likelihood of θ based on the actual observed data, $z(n)$. Thus, the choice of θ that maximizes $Q(\theta;\theta^{[t]})$, that is, the next iterate $\theta^{[t+1]}$, increases our objective function:

$$\log p\left(z\Big|\theta^{[t+1]}\right) \geq \log p\left(z\Big|\theta^{[t]}\right).$$

Consequently, maximizing $Q(\theta;\theta^{[t]})$ is the same as maximizing $\log p(z|\theta)$, but with augmented-data sufficient statistics. Given $[t]$-th iterate hyper- and nuisance parameters $\theta^{[t]}$, the explicit formula for $Q(\theta;\theta^{[t]})$ is in the closed form:

$$\begin{aligned} Q\left(\theta;\theta^{[t]}\right) &= E\left[\log p(x,\epsilon|\theta)\Big|z,\theta^{[t]}\right] \\ &= \sum_{i,n} E\left[\log p(w_i^x(n)|\Sigma_i) + \log p(w_i^\epsilon(n)|\sigma_\epsilon^2)\Big|z,\theta^{[t]}\right]. \end{aligned} \tag{9.19}$$

It is then easy to verify that the maximizer of $Q(\theta;\theta^{[t]})$ is the weighted least squares estimate [50]:

$$\begin{aligned} \Sigma_i^{[t+1]} &= \frac{1}{N_i}\sum_n E\left[w_i^x(n)w_i^{xT}(n)\Big|z,\theta^{[t]}\right] \\ \Sigma_\epsilon^{[t+1]} &= \frac{1}{3\sum_i N_i}\sum_{i,n} E\left[w_i^{\epsilon T}(n)w_i^\epsilon(n)\Big|z,\theta^{[t]}\right], \end{aligned} \tag{9.20}$$

where N_i is the number of wavelets samples in the i-th subband. In each iteration, the computation of the sufficient statistics in Equation 9.19 is often called expectation- or E-step, whereas the process of carrying out Equation 9.20 to find $\theta^{[t+1]}$ is referred to as maximization- or M-step. Carrying out the math to find $E[w_i^x(n)w_i^{xT}(n)|z,\theta^{[t]}]$ and $E[w_i^{\epsilon T}(n)w_i^\epsilon(n)|z,\theta^{[t]}]$ in E-step is rather cumbersome, and the derivation is omitted in this chapter. Interested readers are encouraged to refer to References [49] and [50] for more details.

As was the case in the previous section, it is worth noting that wavelet coefficients are often modelled with heavy-tailed distributions (e.g., Laplace, Student's t, Gaussian mixture). Distributions belonging to an exponential family can be rewritten as a scalar mixture of Gaussian random variables, and thus are conditionally Gaussian. The EM algorithm developed above is therefore generalized to a heavy-tailed distribution via the integration over the mixture variable in the posterior sense.

9.7 Filterbank Coefficient Estimation

Computational efficiency and elegance of shrinkage or thresholding estimators and theoretical properties amenable to spatial inhomogeneities have contributed to the immense popularity of wavelet-based methods for image denoising. However, typical denoising techniques assume complete grayscale or color image observation, and hence must be applied *after* demosaicking. In the previous section we showed that it is possible to model noisy color images in the wavelet domain directly by taking advantage of the statistical framework of missing data. However, the computational burden of doing so is severe, and the energy compaction arguments put forth in Section 9.4 suggest an alternative approach by choosing to work with wavelet coefficients of the noisy subsampled data directly. In this section, we propose necessary changes to a *complete image* wavelet coefficient model such that it is amenable to the direct manipulation of $w_i^y(n)$, [17].

Given that the difference images are sufficiently low-pass, simplification in Equation 9.9 reveals that there is a surprising degree of similarity between $w_i^y(n)$ and $w_i^\ell(n)$. Specifically, $w_i^y(n) \approx w_i^\ell(n)$ for the majority of subbands — the exceptions are the subbands that are normally considered high-frequency, which now contain a strong mixture of the low-frequency (or scaling) coefficients from the difference images, α and β. Operating under the premises that the filterbank transform decomposes image signals such that subbands are approximately uncorrelated from each other, the posterior mean estimate of $w_i^\ell(n)$ takes the form

$$\{w_i^\ell\}^{\text{est}}(\boldsymbol{n}) = E\left[w_i^\ell(\boldsymbol{n}) \middle| w_i^z\right] \approx E\left[w_i^\ell(\boldsymbol{n}) \middle| w_i^{\ell+\varepsilon}\right]$$

for all subbands that meet the $w_i^y(\boldsymbol{n}) \approx w_i^\ell(\boldsymbol{n})$ approximation. Since the wavelet shrinkage function $f: \mathbb{R} \to \mathbb{R}$, $f(w_i^{\ell+\varepsilon}) = E(w_i^\ell(\boldsymbol{n})|w_i^{\ell+\varepsilon})$ is a well studied problem in the literature, we can leverage existing image denoising methods to the CFA image context. In a simple special case where $w_i^\ell(\boldsymbol{n}) \sim \mathcal{N}(0, \sigma_{\ell,i}^2)$, the L^2 estimator is

$$f(w_i^z) = \frac{\sigma_{\ell,i}^2}{\sigma_{\ell,i}^2 + \sigma_\varepsilon^2} w_i^z(\boldsymbol{n}).$$

However, in the subbands that contain a mixture of w_i^ℓ, w_i^α, w_i^β, and w_i^ε, we must proceed with caution. Let $w_i^\alpha(\boldsymbol{n}) \sim \mathcal{N}(0, \sigma_{\alpha,i}^2)$, $w_i^\beta(\boldsymbol{n}) \sim \mathcal{N}(0, \sigma_{\beta,i}^2)$. Consider the case such that $i_0 > I - \hat{I}$ and $i_1 < \hat{I}$, and define $\boldsymbol{j} = (i_0, I - i_1)$, $\boldsymbol{k} = (I - i_0, I - i_1)$. Then $w_i^z(\boldsymbol{n}), w_{\boldsymbol{j}}^z(\boldsymbol{n}), w_{\boldsymbol{k}}^z(\boldsymbol{n})$ are highly correlated due to their common components in their mixture, $w_{i'}^\alpha$ and $w_{i'}^\beta$, where $\boldsymbol{i}' = (I - i_0, i_1)$. Thus the L^2 estimates for $w_i^\ell(\boldsymbol{n}), w_{\boldsymbol{j}}^\ell(\boldsymbol{n}), w_{\boldsymbol{k}}^\ell(\boldsymbol{n})$ are

$$\begin{bmatrix} \{w_i^\ell\}^{\text{est}}(\boldsymbol{n}) \\ \{w_{\boldsymbol{j}}^\ell\}^{\text{est}}(\boldsymbol{n}) \\ \{w_{\boldsymbol{k}}^\ell\}^{\text{est}}(\boldsymbol{n}) \end{bmatrix} = E\left(\begin{bmatrix} w_i^\ell(\boldsymbol{n}) \\ w_{\boldsymbol{j}}^\ell(\boldsymbol{n}) \\ w_{\boldsymbol{k}(\boldsymbol{n})}^\ell \end{bmatrix} \middle| \begin{bmatrix} w_i^z(\boldsymbol{n}) \\ w_{\boldsymbol{j}}^z(\boldsymbol{n}) \\ w_{\boldsymbol{k}(\boldsymbol{n})}^z \end{bmatrix} \right)$$

$$= E\left(\begin{bmatrix} w_i^\ell(\boldsymbol{n}) \\ w_{\boldsymbol{j}}^\ell(\boldsymbol{n}) \\ w_{\boldsymbol{k}(\boldsymbol{n})}^\ell \end{bmatrix} \begin{bmatrix} w_i^z(\boldsymbol{n}) \\ w_{\boldsymbol{j}}^z(\boldsymbol{n}) \\ w_{\boldsymbol{k}(\boldsymbol{n})}^z \end{bmatrix}^T \right) E\left(\begin{bmatrix} w_i^z(\boldsymbol{n}) \\ w_{\boldsymbol{j}}^z(\boldsymbol{n}) \\ w_{\boldsymbol{k}(\boldsymbol{n})}^z \end{bmatrix} \begin{bmatrix} w_i^z(\boldsymbol{n}) \\ w_{\boldsymbol{j}}^z(\boldsymbol{n}) \\ w_{\boldsymbol{k}(\boldsymbol{n})}^z \end{bmatrix}^T \right)^{-1} \begin{bmatrix} w_i^z \\ w_{\boldsymbol{j}}^z \\ w_{\boldsymbol{k}}^z \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{\ell,i}^2 & & \\ & \sigma_{\ell,\boldsymbol{j}}^2 & \\ & & \sigma_{\ell,\boldsymbol{k}}^2 \end{bmatrix} \begin{bmatrix} \sigma_{\ell,i}^2 + \sigma_\varepsilon^2 + \frac{\sigma_{\alpha,i'}^2 + \sigma_{\beta,i'}^2}{16} & \frac{\sigma_{\alpha,i'}^2 - \sigma_{\beta,i'}^2}{16} & \frac{\sigma_{\alpha,i'}^2 + \sigma_{\beta,i'}^2}{16} \\ \frac{\sigma_{\alpha,i'}^2 - \sigma_{\beta,i'}^2}{16} & \sigma_{\ell,\boldsymbol{j}}^2 + \sigma_\varepsilon^2 + \frac{\sigma_{\alpha,i'}^2 + \sigma_{\beta,i'}^2}{16} & \frac{\sigma_{\alpha,i'}^2 - \sigma_{\beta,i'}^2}{16} \\ \frac{\sigma_{\alpha,i'}^2 + \sigma_{\beta,i'}^2}{16} & \frac{\sigma_{\alpha,i'}^2 - \sigma_{\beta,i'}^2}{16} & \sigma_{\ell,\boldsymbol{k}}^2 + \sigma_\varepsilon^2 + \frac{\sigma_{\alpha,i'}^2 + \sigma_{\beta,i'}^2}{16} \end{bmatrix}^{-1} \begin{bmatrix} w_i^z \\ w_{\boldsymbol{j}}^z \\ w_{\boldsymbol{k}}^z \end{bmatrix}.$$

Similarly,

$$\begin{bmatrix} \{w_{i'}^\alpha\}^{\text{est}}(\boldsymbol{n}) \\ \{w_{i'}^\beta\}^{\text{est}}(\boldsymbol{n}) \end{bmatrix} = E\left(\begin{bmatrix} w_{i'}^\alpha(\boldsymbol{n}) \\ w_{i'}^\beta(\boldsymbol{n}) \end{bmatrix} \middle| \begin{bmatrix} w_i^z(\boldsymbol{n}) \\ w_{\boldsymbol{j}}^z(\boldsymbol{n}) \\ w_{\boldsymbol{k}(\boldsymbol{n})}^z \end{bmatrix} \right)$$

$$= E\left(\begin{bmatrix} w_{i'}^\alpha(\boldsymbol{n}) \\ w_{i'}^\beta(\boldsymbol{n}) \end{bmatrix} \begin{bmatrix} w_i^z(\boldsymbol{n}) \\ w_{\boldsymbol{j}}^z(\boldsymbol{n}) \\ w_{\boldsymbol{k}(\boldsymbol{n})}^z \end{bmatrix}^T \right) E\left(\begin{bmatrix} w_i^z(\boldsymbol{n}) \\ w_{\boldsymbol{j}}^z(\boldsymbol{n}) \\ w_{\boldsymbol{k}(\boldsymbol{n})}^z \end{bmatrix} \begin{bmatrix} w_i^z(\boldsymbol{n}) \\ w_{\boldsymbol{j}}^z(\boldsymbol{n}) \\ w_{\boldsymbol{k}(\boldsymbol{n})}^z \end{bmatrix}^T \right)^{-1} \begin{bmatrix} w_i^z \\ w_{\boldsymbol{j}}^z \\ w_{\boldsymbol{k}}^z \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\sigma_{\alpha,i'}^2}{16} & \frac{\sigma_{\alpha,i'}^2}{16} & \frac{\sigma_{\alpha,i'}^2}{16} \\ -\frac{\sigma_{\beta,i'}^2}{16} & \frac{\sigma_{\beta,i'}^2}{16} & -\frac{\sigma_{\beta,i'}^2}{16} \end{bmatrix} \begin{bmatrix} \sigma_{\ell,i}^2 + \sigma_\varepsilon^2 + \frac{\sigma_{\alpha,i'}^2 + \sigma_{\beta,i'}^2}{16} & \frac{\sigma_{\alpha,i'}^2 - \sigma_{\beta,i'}^2}{16} & \frac{\sigma_{\alpha,i'}^2 + \sigma_{\beta,i'}^2}{16} \\ \frac{\sigma_{\alpha,i'}^2 - \sigma_{\beta,i'}^2}{16} & \sigma_{\ell,\boldsymbol{j}}^2 + \sigma_\varepsilon^2 + \frac{\sigma_{\alpha,i'}^2 + \sigma_{\beta,i'}^2}{16} & \frac{\sigma_{\alpha,i'}^2 - \sigma_{\beta,i'}^2}{16} \\ \frac{\sigma_{\alpha,i'}^2 + \sigma_{\beta,i'}^2}{16} & \frac{\sigma_{\alpha,i'}^2 - \sigma_{\beta,i'}^2}{16} & \sigma_{\ell,\boldsymbol{k}}^2 + \sigma_\varepsilon^2 + \frac{\sigma_{\alpha,i'}^2 + \sigma_{\beta,i'}^2}{16} \end{bmatrix}^{-1} \begin{bmatrix} w_i^z \\ w_{\boldsymbol{j}}^z \\ w_{\boldsymbol{k}}^z \end{bmatrix}.$$

Once $\{w_i^\ell\}^{\text{est}}, \{w_i^\alpha\}^{\text{est}}, \{w_i^\beta\}^{\text{est}}$ are computed $\forall \boldsymbol{i}, \boldsymbol{n}$ as above, then $\boldsymbol{x}^{\text{est}}(\boldsymbol{n})$ is calculated by taking the inverse filterbank transform of $\{w_i^\ell\}^{\text{est}}, \{w_i^\alpha\}^{\text{est}}, \{w_i^\beta\}^{\text{est}}$ to find the estimates of $\ell(\boldsymbol{n}), \alpha(\boldsymbol{n}), \beta(\boldsymbol{n})$, which in turn is used to solve $\boldsymbol{x}^{\text{est}}$.

Practically, it should be noted that the actual implementation of this method should include cycle-spinning, a standard technique in filterbank and wavelet literature whereby a linear space-variant system can be transformed into linear space-invariant system via averaging over all possible spacial shifts. As with the previous sections, we note that the estimator naturally extends to multivariate normal or heavy-tailed distributions.

9.8 Conclusion

Given the inadequacies and model inconsistencies of treating the image denoising and demosaicking problem independently, we focused on the analysis and the techniques for processing (see Figure 9.8 and Figure 9.9) subsampled data. In particular, the Fourier and filterbank (wavelet-packet) analyses reveal a systematic aliasing structure in CFA images, where the observed data consists of a mixture of baseband luminance signal, spectrally-shifted difference images, and noise. The same analysis motivates a unified strategy to address demosaicking and denoising estimation problems by interpreting the sensor data as luminance image distorted by noise with some degrees of structure.

Conditioned on the *complete observation* image model of the digital camera designer's choosing, we proposed three design regimes for estimating the complete noise-free image signal of interest given a set of incomplete observations of pixel components that are corrupted by noise. First, well-understood DSP machineries were employed to design a spatially-adaptive linear filter whose stop-band contains the spectral copies of the difference images, and the pass-band suppresses noise. Second, coupling of the EM algorithm framework with the Bayesian models to estimate the hyper- and nuisance parameters via the marginal likelihood, and in turn, adopting the empirical partial Bayes approach for estimating the ideal color image data allowed us to apply heavy-tailed priors to the unobservable wavelet coefficients. Third, exploiting the reversed-order filterbank structure, a regression of luminance and difference image filterbank coefficients on the CFA image filterbank coefficients were simplified. The above estimation techniques were derived using second-order statistics for complete observation models.

Acknowledgments

The author would like to thank his wonderful collaborators, Dr. Thomas W. Parks in the Department of Electrical and Computer Science at Cornell University, Dr. Xiao-Li Meng in the Department of Statistics at Harvard University, and Dr. Patrick J. Wolfe in the School of Engineering and Applied Sciences at Harvard University, whose invaluable contributions

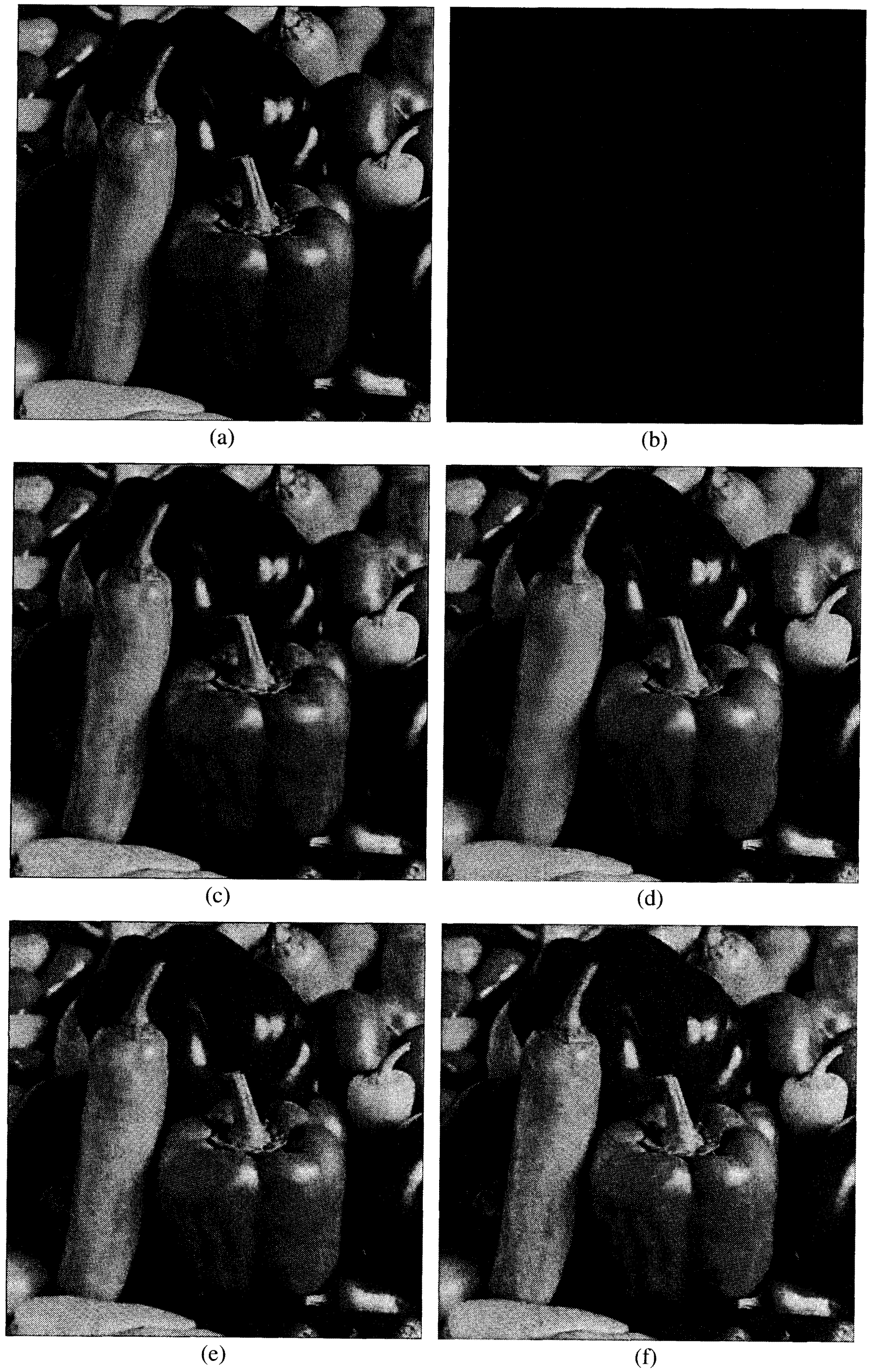

FIGURE 9.8

Reconstruction of the *Peppers* image given a simulated noisy sensor data: (a) noise-free original color image, (b) color version of simulated noisy sensor data, (c) estimated with demosaicking method in Reference [7] and denoising method in [23], (d) estimated with the approach in Section 9.5, (e) estimated with the approach in Section 9.6, and (f) estimated with the approach in Section 9.7.

FIGURE 9.9

Reconstruction of the *Lena* image given a simulated noisy sensor data: (a) noise-free original color image, (b) color version of simulated noisy sensor data, (c) estimated with demosaicking method in Reference [7] and denoising method in [23], (d) estimated with the approach in Section 9.5, (e) estimated with the approach in Section 9.6, and (f) estimated with the approach in Section 9.7.

to the works in References [17], [40], [42], [49] are reflected in this chapter. His gratitude extends also to Dr. Bahadir Gunturk at Louisiana State University and Dr. Javier Portilla at Universidad de Granada for making their simulation codes available; and to Daniel Rudoy, Ayan Chakrabarti, and Prabahan Basu at Harvard University for their constructive criticisms.

References

[1] B.E. Bayer, "Color imaging array." U.S. Patent 3 971 065, July 1976.

[2] R. Lukac and K.N. Plataniotis, "Color filter arrays: Design and performance analysis," *IEEE Transactions on Consumer Electronics*, vol. 51, no. 4, pp. 1260–1267, November 2005.

[3] K. Parulski and K.E. Spaulding, *Digital Color Image Handbook*, ch. Color image processing for digital cameras, G. Sharma (ed.), Boca Raton, FL: CRC Press, 2002, pp. 727–757.

[4] S. Yamanaka, "Solid state camera." U.S. Patent 4 054 906, November 1977.

[5] M. Parmar and S.J. Reeves, "A perceptually based design methodology for color filter arrays," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Montreal, Canada, May 2004, vol. 3, pp. 473–476.

[6] K. Hirakawa and T.W. Parks, "Adaptive homogeneity-directed demosaicing algorithm," *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 360–369, March 2005.

[7] 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, September 2002.

[8] X. Wu and N. Zhang, "Primary-consistant soft-decision color demosaicking for digital cameras," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1263–1274, September 2004.

[9] B.K. Gunturk, J. Glotzbach, Y. Altunbasak, R.W. Schafer, and R.M. Mersereau, "Demosaicking: Color filter array interpolation in single chip digital cameras," *IEEE Signal Processing Magazine; Special Issue on Color Image Processing*, vol. 22, no. 1, pp. 44–54, January 2005.

[10] R. Lukac and K.N. Plataniotis, "Universal demosaicking for imaging pipelines with an RGB color filter array," *Pattern Recognition*, vol. 38, no. 11, pp. 2208–2212, November 2005.

[11] D. Alleysson, S. Süsstrunk, and J. Herault, "Linear demosaicing inspired by the human visual system," *IEEE Transactions on Image Processing*, vol. 14, no. 4, pp. 439–449, April 2005.

[12] L. Chang and Y.P. Tang, "Effective use of spatial and spectral correlations for color filter array demosaicking," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 2, pp. 355–365, May 2004.

[13] R. Kakarala and Z. Baharav, "Adaptive demosaicing with the principal vector method," *IEEE Transactions on Consumer Electronics*, vol. 48, no. 4, pp. 932–937, November 2002.

[14] W. Lu and Y.P. Tan, "Color filter array demosaicking: new method and performance measures," *IEEE Transactions on Image Processing*, vol. 12, no. 10, pp. 1194–1210, October 2003.

[15] D.D. Muresan and T.W. Parks, "Demosaicing using optimal recovery," *IEEE Transactions on Image Processing*, vol. 14, no. 2, pp. 267–278, February 2005.

[16] R. Ramanath and W.E. Snyder, "Adaptive demosaicking," *Journal of Electronic Imaging*, vol. 12, no. 4, pp. 633–642, October 2003.

[17] K. Hirakawa, X.L. Meng, and P.J. Wolfe, "A framework for wavelet-based analysis and processing of color filter array images with applications to denoising and demosaicking," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Honolulu, HI, USA, April 2007, vol. 1, pp. 597–600.

[18] M.S. Crouse, R.D. Nowak, and R.G. Baraniuk, "Bayesian tree-structured image modeling using wavelet-domain hidden Markov models," *IEEE Transactions on Image Processing*, vol. 46, no. 7, pp. 1056–1068, July 1998.

[19] D.L. Donoho and I.M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, September 1994.

[20] M. Jansen and A. Bultheel, "Empirical Bayes approach to improve wavelet thresholding for image noise reduction," *Journal of American Statistical Association*, vol. 96, no. 454, pp. 629–639, June 2001.

[21] I.M. Johnstone and B.W. Silverman, "Wavelet threshold estimators for data with correlated noise," *Journal of Royal Statistical Society - Series B*, vol. 59, no. 2, pp. 319–351, 1997.

[22] J. Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli, "Image denoising using scale mixture of Gaussians in the wavelet domain," Tech. Rep. TR2002-831, Comput. Sci. Dept., Courant Inst. Math. Sci., New York Univ., 2002.

[23] J. Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli, "Image denoising using scale mixture of Gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, November 2003.

[24] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy, "A joint inter and intrascale statistical model for Bayesian wavelet based image denoising," *IEEE Transactions on Image Processing*, vol. 11, no. 5, pp. 545–557, May 2002.

[25] L. Sendur and I.W. Selesnick, "Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency," *IEEE Transactions on Signal Processing*, vol. 50, no. 11, pp. 2744–2756, November 2002.

[26] L. Sendur and I.W. Selesnick, "Bivariate shrinkage with local variance estimation," *IEEE Signal Processing Letters*, vol. 9, no. 12, pp. 438–441, December 2002.

[27] J.L. Starck, D.L. Donoho, and E. Cande, "Very high quality image restoration," in *Proceedings of the SPIE Conference on Wavelet and Applications in Signal and Image Processing*, San Diego, CA, USA, July 2001, vol. 4478, pp. 9–19.

[28] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proceedings of the IEEE International Conference on Computer Vision*, Bombay, India, January 1998, pp. 839–846.

[29] G. Hua and M. T. Orchard, "A new interpretation of translation invariant denoising," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Montreal, Canada, May 2004, vol. 3, pp. 189–192.

[30] X. Li and M.T. Orchard, "Spatially adaptive image denoising under overcomplete expansion," in *Proceedings of the IEEE International Conference on Image Processing*, Vancouver, BC, Canada, September 2000, vol. 3, pp. 300–303.

[31] D.D. Muresan and T.W. Parks, "Adaptive principal components and image denoising," in *Proceedings of the IEEE International Conference on Image Processing*, Barcelona, Spain, September 2003, vol. 1, pp. 101–104.

[32] K. Hirakawa and T.W. Parks, "Image denoising for signal-dependent noise," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2730–2742, September 2006.

[33] D.D. Muresan and T.W. Parks, "Adaptively quadratic (AQua) image interpolation," *IEEE Transactions on Image Processing*, vol. 13, no. 5, pp. 690–698, May 2004.

[34] H. Tian, B. Fowler, and A.E. Gamal, "Analysis of temporal noise in CMOS photodiode active pixel sensor," *IEEE Journal on Solid State Circuits*, vol. 36, no. 1, pp. 92–101, January 2001.

[35] G.E. Healey and R. Kondepudy, "Radiometric CCD camera calibration and noise estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 3, pp. 267–276, March 1994.

[36] R. Ding and A.N. Venetsanopoulos, "Generalized homomorphic and adaptive order statistic filters for the removal of impulsive and signal-dependent noise," *IEEE Transactions on Circuits Systems*, vol. 34, no. 8, pp. 948–955, August 1987.

[37] A.V. Oppenheim, R.W. Schafer, and J.R. Buck, *Discrete-Time Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, 2nd Edition, 1999.

[38] P. Fryzlewicz and G.P. Nason, "A Haar-Fisz algorithm for Poisson intensity estimation," *Journal of Computational and Graphical Statistics*, vol. 13, no. 3, pp. 621–638, September 2004.

[39] P. Fryzlewicz and G.P. Nason, "Smoothing the wavelet periodogram using the Haar-Fisz transform," *Scientific Charge*, 2001.

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

[41] M. Raphan and E.P. Simoncelli, *Advances in Neural Information Processing System*, ch. Learning to be Bayesian without Supervision. Cambridge, MA: MIT Press, vol. 19, 2007.

[42] K. Hirakawa, "Signal-dependent noise characterization in wavelets domain," in *Proceedings of the SPIE Conference on Optics & Photonics*, San Diego, CA, USA, August 2007.

[43] D.R. Cok, "Signal processing method and apparatus for producing interpolated chrominance values in a sampled color image signal," U.S. Patent 4 642 678, February 1987.

[44] R. Kimmel, "Demosaicing: Image reconstruction from color CCD samples," *IEEE Transactions on Image Processing*, vol. 8, no. 9, pp. 1221–1228, September 1999.

[45] R. Lukac, K.N. Plataniotis, D. Hatzinakos, and M. Aleksic, "A novel cost effective demosaicing approach," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 1, pp. 256–261, February 2004.

[46] R. Lukac and K.N. Plataniotis, "Normalized color-ratio modeling for CFA interpolation," *IEEE Transactions on Consumer Electronics*, vol. 50, no. 2, pp. 737–745, May 2004.

[47] E. Dubois, "Filter design for adaptive frequency-domain Bayer demosaicking," in *Proceedings of the IEEE International Conference on Image Processing*, Atlanta, GA, USA, October 2006, pp. 2705–2708.

[48] M.J.T. Smith and T.P. Barnwell, "A procedure for designing exact reconstruction filter banks for tree structured subband coders," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, San Diego, CA, USA, March 1984, pp. 27:1.1–1.4.

[49] K. Hirakawa and X.L. Meng, "An empirical Bayes EM-wavelet unification for simultaneous denoising, interpolation, and/or demosaicing," in *Proceedings of the IEEE International Conference on Image Processing*, Atlanta, GA, USA, October 2006, pp. 1453–1456.

[50] G.J. McLachelan and T. Krishnan, *The EM Algorithm and Extentions*. New York: John Wiley & Sons, 1997.