EURASIP Journal on Advances in Signal Processing
a SpringerOpen Journal

RESEARCH Open Access

# Analysis and processing of pixel binning for color image sensor

Xiaodan Jin and Keigo Hirakawa*

**Abstract**

Pixel binning refers to the concept of combining the electrical charges of neighboring pixels together to form a superpixel. The main benefit of this technique is that the combined charges would overcome the read noise at the sacrifice of spatial resolution. Binning in color image sensors results in superpixel Bayer pattern data, and subsequent demosaicking yields the final, lower resolution, less noisy image. It is common knowledge among the practitioners and camera manufacturers, however, that binning introduces severe artifacts. The in-depth analysis in this article proves that these artifacts are far worse than the ones stemming from loss of resolution or demosaicking, and therefore it cannot be eliminated simply by increasing the sensor resolution. By accurately characterizing the sensor data that has been binned, we propose a post-capture binning data processing solution that succeeds in suppressing noise and preserving image details. We verify experimentally that the proposed method outperforms the existing alternatives by a substantial margin.

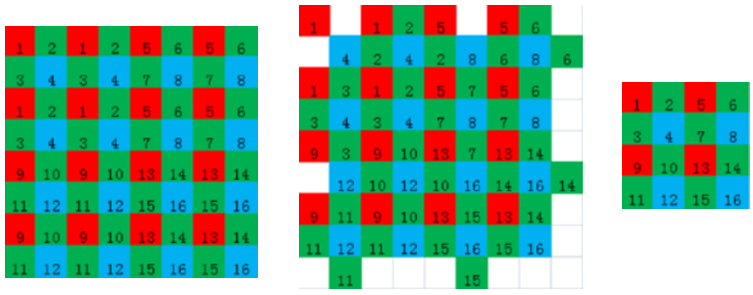
## 1 Introduction

Recent progress on digital camera technology has had extraordinary impact on numerous electronic industries, including mobile phones, security, vehicle, bioengineering, and computer vision systems. In many applications, sensor resolution has exceeded the optical resolution, meaning that the additional hardware complexity to increase pixel density would not necessarily result in large image quality gains. The significant improvement in sensor sensitivity has allowed cameras to operate in lighting conditions that were unthinkable with film cameras.

Despite increased sensitivity, however, noise remains a serious problem in modern image sensors. Available technologies for reducing noise in hardware include backside illuminated architecture [1,2], color filters with higher transmittance [3,4], and pixel binning [5-7]. Processing techniques at our disposal include image denoising [8-10], joint denoising and demosaicking [11-14], image deblurring [15,16] (long shutter to compensate for light), and single-shot high dynamic range imaging [17].

The goal of this article is to provide a comprehensive characterization of the pixel binning for color image sensors, and propose post-capture signal processing steps aimed at eliminating the binning artifacts. Binning refers to the concept of combining the electrical charges of neighboring pixels together to form a *superpixel*. The combined signal will then be amplified by a source follower and converted into digital values by an analog-to-digital converter. The main benefit of this technique is that the combined charges would overcome the read noise, even if the individual pixel values are small. The improved noise performance comes at the price of spatial resolution loss, however. Binning in color image sensors is complicated by the presence of color filter array (CFA). Data are typically obtained via a single CCD or CMOS sensor with a CFA spatial subsampling procedure, a physical construction whereby each pixel location measures only a single color. Figures 1a,b show the most well known CFA scheme called the Bayer pattern, which involves red, green, and blue filters. To maintain the fidelity of color, binning in *color* image sensors are performed by combining neighboring pixels with the same color filter. As evidenced by the two well known binning configurations shown in Figures 1a,b, the resultant superpixel form a Bayer pattern, as shown in Figure 1c. The subsequent demosaicking algorithm—the process of interpolating to recover the full RGB representation of the image from the CFA subsampled sensor data—yields the final, lower resolution, less noisy image.

*Correspondence: khirakawa1@udayton.edu
Electrical and Computer Engineering, University of Dayton, Dayton, Ohio, USA



(a) Kodak PIXELUX [7] (b) PhaseOne [5] (c) Output

**Figure 1 Commonly used binning schemes.** Binning refers to the concept of combining the electrical charges of neighboring pixels together to form a *superpixel*. (**a–b**) The numbers over the high resolution Bayer pattern indicate which pixels are combined together. (**c**) The resultant superpixel Bayer pattern, where the numbers indicate the relative locations of the combined pixels (for [7] and [5]).

However, it is a common knowledge among the practitioners and camera manufacturers that binning introduces pixelization artifacts. An example is shown in Figure 2. As will be made clear in the sequel, these artifacts differ from the ones stemming from loss of resolution, and therefore it cannot be eliminated simply by increasing the sensor resolution. In-depth analysis of the sampling scheme implied by the binning proves that gross mismatch between binning and demosaicking results is at fault for the severe pixelization. Hence the *right* way to correct this problem is to design a binning-aware demosaicking algorithm. The proposed method still draws from the established demosaicking principles, but with profound differences in the way spatially high frequency components are handled. To the best of the knowledge of the authors, this is the first major article to examine pixel binning problem in color image sensors from the signal processing perspective, and to provide post-capture processing solution to correct for the pixelization artifacts.

The remainder of this article is organized as follows. We begin by briefly reviewing CFA sampling and demosaicking in Section 2. Section 3 provides a rigorous analysis of binning. A novel binning-aware demosaicking technique is developed in Section 4. We experimentally verify its effectiveness in 5 before making concluding remarks in Section 6.

## 2 Background

### 2.1 CFA sampling

Thanks to the seminal work of [18] and further investigations by [19-21], CFA sampling is well characterized and understood. The key insight is the two dimensional Fourier analysis of CFA sampled sensor data, which reveals that the signal is preserved by an efficient space-color representation. Specifically, let $\boldsymbol{x} : \mathbb{Z}^2 \rightarrow \mathbb{R}^3$, where $\boldsymbol{x}(\boldsymbol{n}) = [x_r(\boldsymbol{n}), x_g(\boldsymbol{n}), x_b(\boldsymbol{n})]^T$ correspond to the RGB tri-stimulus value at location $\boldsymbol{n} \in \mathbb{Z}^2$. Then the CFA subsampled data has the following form:

$$\begin{aligned} y(\boldsymbol{n}) &= \boldsymbol{c}(\boldsymbol{n})^T \boldsymbol{x}(\boldsymbol{n}) \\ &= \boldsymbol{c}(\boldsymbol{n})^T \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \boldsymbol{x}(\boldsymbol{n}) \\ &= \begin{bmatrix} 1 & c_\alpha(\boldsymbol{n}) & c_\beta(\boldsymbol{n}) \end{bmatrix} \begin{bmatrix} x_g(\boldsymbol{n}) \\ x_\alpha(\boldsymbol{n}) \\ x_\beta(\boldsymbol{n}) \end{bmatrix}, \end{aligned} \tag{1}$$

(a) No binning (b) Kodak PIXELUX (c) Phase One

**Figure 2 Binning vs. no binning.** Compared to no binning, binning succeeds in reducing noise. However, the pixelization and zippering artifacts deteriorate the image quality. (**a**) Reconstruction from full resolution CFA; (**b**) reconstruction from Kodak PIXELUX scheme of Figure 1a; (**c**) reconstruction from PhaseOne scheme of Figure 1b.

where $\boldsymbol{c} : \mathbb{Z}^2 \rightarrow [0,1]^3$ denotes the translucency of CFA at location $\boldsymbol{n}$. The advantage to the representation is that the difference images $x_\alpha = x_r - x_g$ and $x_\beta = x_b - x_g$ enjoy rapid spectral decay and can serve as a proxy for chrominance. On the other hand, the "baseband" green image $x_g$ can be taken to approximate luminance. As our eventual image recovery task will be to approximate the true color image triple $\boldsymbol{x}(\boldsymbol{n})$ from acquired sensor data $y(\boldsymbol{n})$, note that recovering either representations ($\{x_r, x_g, x_b\}$ or $\{x_g, x_\alpha, x_\beta\}$) are equivalent. Moreover, the representation of (1) allows us to re-cast the pure-color sampling structure in terms of sampling structures $c_\alpha$ and $c_\beta$ associated with the difference channels $x_\alpha$ and $x_\beta$. For more extensive investigation on the bandlimitedness assumptions of $\{x_g, x_\alpha, x_\beta\}$, see [18-20].

Denote by the uppercase letters the discrete space Fourier transforms and $\boldsymbol{\omega} = (\omega_1, \omega_2)^T \in \{\mathbb{R}/(2\pi)\}^2$ ($\mathbb{R}/(2\pi)$ denotes the quotient group of $\mathbb{R}$ by the subgroup $2\pi\mathbb{Z}$) the two dimensional Fourier index. Then the Fourier analysis of CFA is:

$$C_\alpha(\boldsymbol{\omega}) = \sum_{\boldsymbol{\lambda} \in \pi\mathbb{Z}_2^2} \frac{\delta(\boldsymbol{\omega} - \boldsymbol{\lambda})}{4},$$

$$C_\beta(\boldsymbol{\omega}) = \sum_{\boldsymbol{\lambda} \in \pi\mathbb{Z}_2^2} e^{j\{\boldsymbol{\lambda}^T \binom{1}{1}\}} \frac{\delta(\boldsymbol{\omega} - \boldsymbol{\lambda})}{4},$$

where $\delta(\cdot)$ is the Dirac delta function, and $\mathbb{Z}_2$ denotes the cyclic group of order 2. Note that the phase shift term in $C_\beta$ arises due to the relative position of blue pixels relative to the red (the origin is assumed to be on a red pixel). The corresponding Fourier analysis of the sensor data $y$ takes the following form:

$$\begin{aligned} Y(\boldsymbol{\omega}) &= X_g(\boldsymbol{\omega}) + C_\alpha(\boldsymbol{\omega}) \star X_\alpha(\boldsymbol{\omega}) + C_\beta(\boldsymbol{\omega}) \star X_\beta(\boldsymbol{\omega}) \\ &= X_g(\boldsymbol{\omega}) + \sum_{\boldsymbol{\lambda} \in \pi\mathbb{Z}_2^2} \frac{X_\alpha(\boldsymbol{\omega} - \boldsymbol{\lambda}) + e^{j\{\boldsymbol{\lambda}^T \binom{1}{1}\}} X_\beta(\boldsymbol{\omega} - \boldsymbol{\lambda})}{4}, \end{aligned} \quad (2)$$

where $\star$ denotes convolution. The Fourier support of the resultant sensor signal is shown in Figure 3.

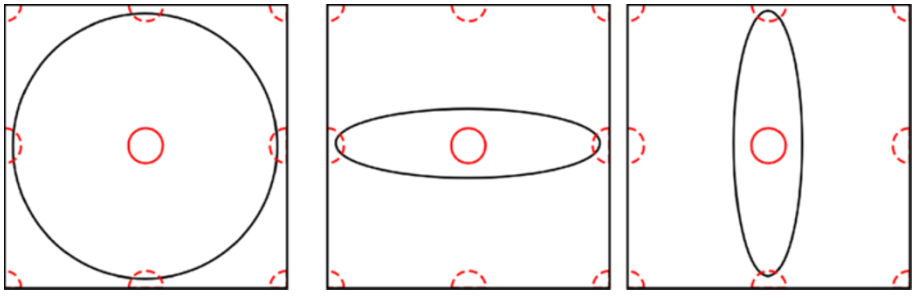
### 2.2 Demosaicking

Most demosaicking algorithms described in the literature make use (either implicitly or explicitly) of correlation structure in the spatial frequency domain, often in the form of local sparsity or directional filtering [14,19,21-23]. As noted in our earlier discussion, the set of carrier frequencies induced by $c_\alpha$ and $c_\beta$ include $[\pi, 0]^T$ and $[0, \pi]^T$, locations that are particularly susceptible to aliasing by horizontal and vertical edges. Figures 3b,c indicates these scenarios, respectively; it may be seen that in contrast to the radially symmetric baseband spectrum of Figure 3a, chrominance–luminance aliasing occurs along one of either the horizontal or vertical axes. However, successful reconstruction can still occur if a noncorrupted copy of this chrominance information is recovered, thereby explaining the popularity of (nonlinear) directional filtering steps [19,21-23]. We can, therefore, view the CFA design problem as one of spatial-frequency multiplexing, and the CFA demosaicking problem as one of demultiplexing to recover subcarriers, with spectral aliasing given the interpretation of "cross talk" [19].

In order to carry out this demultiplexing, signal-adaptive demosaicking methods take the scenarios of Figure 3a–c into account. Typically, this is carried out by first filtering in both horizontal and vertical directions to yield reconstructions $\hat{\boldsymbol{x}}_h$ and $\hat{\boldsymbol{x}}_v$, respectively. Taking their convex combination to yield the final result:

$$\hat{\boldsymbol{x}}_\tau(\boldsymbol{n}) = \tau(\boldsymbol{n})\hat{\boldsymbol{x}}_h(\boldsymbol{n}) + (1 - \tau(\boldsymbol{n}))\hat{\boldsymbol{x}}_v(\boldsymbol{n}), \quad (3)$$

where $\tau \in [0,1]$ is a set of weights. Based on models of a "natural image" behavior, various *policies* for determining the appropriate weights have been developed



(a) Radially symmetric (b) Vertical feature (c) Horizontal feature

**Figure 3 Idealized spectral support of a color image acquired under the Bayer pattern.** In each figure, the horizontal and vertical axes span $[-\pi, \pi)^2$ of Fourier index, and the DC is located at the center of the figure. Solid lines indicate the baseband signals, while replicated spectra with the dashed lines arises as a result of CFA sampling. Black and red lines correspond to the support of luminance and chrominance images, respectively. Alias-inducing chrominance replications are shown with **(a)** Radially symmetric luminance, **(b)** vertical feature luminance, **(c)** horizontal feature luminance.

[14,19,21-23]. For example, the weight combination should maximize the homogeneity $u_{\hat{x}}(\boldsymbol{n})$—defined as a percentage of pixels in the neighborhood of $\boldsymbol{n}$ (denoted $\eta(\boldsymbol{n})$) that are similar to $\boldsymbol{x}(\boldsymbol{n})$ [22]:

$$u_{\hat{x}}(\boldsymbol{n}) = \frac{\#\{\boldsymbol{m} \in \eta(\boldsymbol{n}) : d(\hat{\boldsymbol{x}}(\boldsymbol{n}), \hat{\boldsymbol{x}}(\boldsymbol{m})) < \epsilon\}}{\#\{\eta(\boldsymbol{n})\}} \quad (4)$$

where $d(\cdot,\cdot)$ is some distance metric and $\epsilon$ is a tolerance parameter.

## 3 Analysis of binning

Let us rigorously analyze the effects that binning has on the acquired sensor data. We begin in Section 3.1 with a brief review of the signal-to-noise ratio (SNR) gains that binning is expected to improve [24]—the main motivation behind binning. An in-depth analysis in Section 3.2 will prove that a combination of binning and demosaicking results in a loss of resolution that is far worse than commonly believed. Section 3.3 offers an alternative perspective that paves a path towards recovering artifact-free images.

### 3.1 Signal measurement uncertainty

There are at least three types of noise that contribute to the overall error. "Shot noise" is due to the stochasticity of the photon arrival process, and it is well modeled by Poisson distribution. The dark current stemming from in-circuit electron excitation results in "thermal noise," whose power is proportional to the exposure time. Finally, the source follower and analog-to-digital converter introduce the homoscedastic noise that is known as the "read noise." The overall SNR of captured image is well modeled by:

$$\mathrm{SNR}_{\mathrm{pix}} := 20\log_{10}(Q\cdot t\cdot y) - 10\log_{10}(Q\cdot t\cdot y + D\cdot t + N), \quad (5)$$

where $t$ is the exposure time, $Q$ is the quantum efficiency constant, $D$ is the dark current constant, and $N$ is the read noise power.

Owing to the fact that the image sensor resolution exceeds the optical resolution in many applications, binning is an attractive way to trade off the excess spatial resolution for gains in SNR. It is instructive first to consider summing $M$ pixel values digitally, post-acquisition. The signal $y$ is boosted $M$-fold while the noise power increases $M$ times, resulting in an overall $10\log_{10}(M)$ dB gain:

$$\begin{aligned}\mathrm{SNR}_{\mathrm{sum}} :=& 20\log_{10}(M\cdot Q\cdot t\cdot y) \\ &- 10\log_{10}(M\cdot Q\cdot t\cdot y + M\cdot D\cdot t + M\cdot N) \\ =& \mathrm{SNR}_{\mathrm{pix}} + 10\log_{10}(M) \geq \mathrm{SNR}_{\mathrm{pix}}.\end{aligned} \quad (6)$$

Combining electrical charges of neighboring pixels to form a superpixel in hardware offers advantages over simply summing pixels digitally. The main difference is that when the electrical charges are combined *before* source follower and analog-to-digital converter, the uncertainty due to read noise remains constant. The corresponding SNR is:

$$\begin{aligned}\mathrm{SNR}_{\mathrm{bin}} =& 20\log_{10}(M\cdot Q\cdot t\cdot y) \\ &- 10\log_{10}(M\cdot Q\cdot t\cdot y + M\cdot D\cdot t + N) \\ \geq& \mathrm{SNR}_{\mathrm{sum}}.\end{aligned} \quad (7)$$

As illustrated by the example in Figure 4, the differences between $\mathrm{SNR}_{\mathrm{bin}}$ and $\mathrm{SNR}_{\mathrm{sum}}$ are more noticeable when the signal intensity $y$ becomes small and read noise $N$ become dominant—meaning that binning is most effective in the low light ranges.

### 3.2 Binning "sampling"

Due to the fact that binning combines $M$ electric charges of neighboring pixels, each pixel cannot be shared by more than one superpixel. Moreover, the charges can be combined by summation only (i.e. no fractional combinations). As such, the options for binning schemes are fairly limited. Furthermore, the superpixels produced by pixel binning in color image sensors form a Bayer pattern that requires the additional step of demosaicking to recover the full color low resolution image. We will show that superpixel Bayer pattern suffers from many problems that the pixel-level Bayer pattern does not, leading to the conclusion that combining pixel binning and demosaicking is the wrong approach.

Consider Kodak PIXELUX, the most widely used binning scheme illustrated in Figure 1a,c [7]. It combines *four* neighboring pixel values together to form one superpixel.

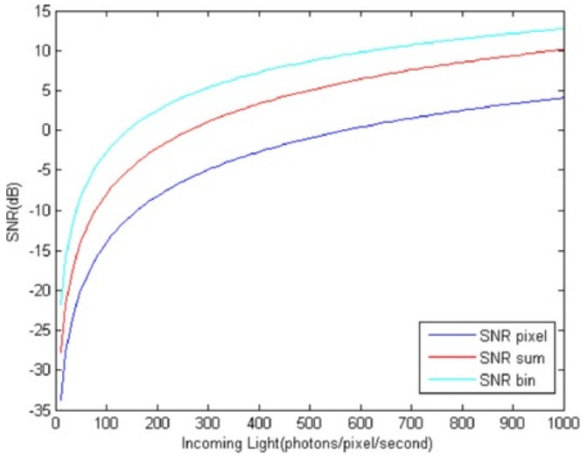



**Figure 4 SNR as a function of signal intensity.** Here, $M = 4$, $Q = 0.70$, $t = 1/100$ s, $D = 0.1$ electrons/pixel/second, and $N = 10$ electrons rms/pixel [24]. See (5-7).

This process of combining neighboring pixels to form a single superpixel is equivalent to applying a convolution operator followed by downsampling:

- *filtering*: let $h_{\text{bin}}$ denote the filter coefficients

$$\begin{aligned} h_{\text{bin}}(\boldsymbol{n}) =& \Delta\left(\boldsymbol{n}-\begin{pmatrix}1\\1\end{pmatrix}\right)+\Delta\left(\boldsymbol{n}-\begin{pmatrix}1\\-1\end{pmatrix}\right)\\ &+\Delta\left(\boldsymbol{n}-\begin{pmatrix}-1\\1\end{pmatrix}\right)+\Delta\left(\boldsymbol{n}-\begin{pmatrix}-1\\-1\end{pmatrix}\right), \end{aligned} \tag{8}$$

  where $\Delta(\cdot)$ denotes the Kronecker delta function. Then the charge summation in PIXELUX is

$$y_{\text{bin}}(\boldsymbol{n}) = y(\boldsymbol{n}) \star h_{\text{bin}}(\boldsymbol{n}).$$

- *downsampling*: to yield the superpixel Bayer pattern data $s$, do

$$\begin{aligned} s(2\boldsymbol{n}) =& y_{\text{bin}}(4\boldsymbol{n})\\ s\left(2\boldsymbol{n}+\begin{pmatrix}0\\1\end{pmatrix}\right) =& y_{\text{bin}}\left(4\boldsymbol{n}+\begin{pmatrix}0\\1\end{pmatrix}\right)\\ s\left(2\boldsymbol{n}+\begin{pmatrix}1\\0\end{pmatrix}\right) =& y_{\text{bin}}\left(4\boldsymbol{n}+\begin{pmatrix}1\\0\end{pmatrix}\right)\\ s\left(2\boldsymbol{n}+\begin{pmatrix}1\\1\end{pmatrix}\right) =& y_{\text{bin}}\left(4\boldsymbol{n}+\begin{pmatrix}1\\1\end{pmatrix}\right). \end{aligned} \tag{9}$$

Note that downsampling implied by (9) is non-uniform—the spatial relationships between samples are changed by the different relative shifts applied to each super pixels (contrast this to (11) below). The Fourier transform of $s$ is (derived in Appendix 2):

$$\begin{aligned} S(\boldsymbol{\omega}) \approx& \sum_{\boldsymbol{\lambda}\in\pi\mathbb{Z}_2^2}\Bigg(X_\alpha\left(\frac{\boldsymbol{\omega}-\boldsymbol{\lambda}}{2}\right)\\ &+\underbrace{e^{j\left(\frac{\boldsymbol{\omega}}{2}\right)^T\binom{1}{1}}}_{\text{unwanted filter}}\ e^{j\left(\frac{\boldsymbol{\lambda}}{2}\right)^T\binom{1}{1}}X_\beta\left(\frac{\boldsymbol{\omega}-\boldsymbol{\lambda}}{2}\right)\Bigg)\\ &+\underbrace{\sum_{\boldsymbol{\theta}\in\mathbb{Z}_2^2}\frac{e^{j\left(\frac{\boldsymbol{\omega}}{2}\right)^T\theta}H_{\text{bin}}\left(\frac{\boldsymbol{\omega}}{2}\right)}{16}X_g\left(\frac{\boldsymbol{\omega}}{2}\right)}_{\text{unwanted filter}}\\ &+\sum_{\boldsymbol{\lambda}\in\frac{\pi}{2}\mathbb{Z}_4^2\setminus\binom{0}{0}}\underbrace{\sum_{\boldsymbol{\theta}\in\mathbb{Z}_2^2}\frac{e^{j\left(\frac{\boldsymbol{\omega}}{2}+\boldsymbol{\lambda}\right)^T\theta}H_{\text{bin}}\left(\frac{\boldsymbol{\omega}}{2}-\boldsymbol{\lambda}\right)}{16}}_{\text{antialias filter}}\underbrace{X_g\left(\frac{\boldsymbol{\omega}}{2}-\boldsymbol{\lambda}\right)}_{\text{aliasing}}. \end{aligned} \tag{10}$$

The corresponding Fourier support of $S(\boldsymbol{\omega})$ is shown in Figure 5. Note that the unwanted filter will boost $X_g$ to 16 at the DC. The approximate relation above is admitted by the bandlimitedness assumptions of $X_\alpha$ and $X_\beta$:

$$\begin{aligned} H_{\text{bin}}(\boldsymbol{\omega})X_\alpha(\boldsymbol{\omega}-\boldsymbol{\lambda}) \approx& 4X_\alpha(\boldsymbol{\omega}-\boldsymbol{\lambda})\\ H_{\text{bin}}(\boldsymbol{\omega})X_\beta(\boldsymbol{\omega}-\boldsymbol{\lambda}) \approx& 4X_\beta(\boldsymbol{\omega}-\boldsymbol{\lambda}). \end{aligned}$$

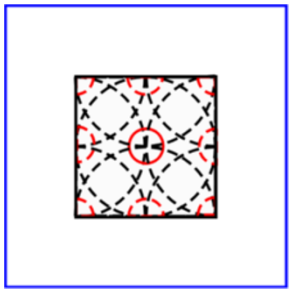


**Figure 5 Idealized spectral support of binning sampled data $S(\boldsymbol{\omega})$ in (10), corresponding to Figure 1.** As before, solid lines indicate the baseband signals, while spectra with the dashed lines arises as a result of CFA sampling. Black and red lines correspond to the support of luminance and chrominance images, respectively. The blue box represents the original sampling rate.

The main advantage of binning in (9) over (2) is that the signal strength of the baseband $X_g$ and the chrominance components $X_\alpha$ and $X_\beta$ are boosted by four times—consistent with the SNR analysis in the previous section. As evidenced by Figure 5a, the Fourier support of (9) closely resembles the Bayer pattern of Figure 3a. Superpixel Bayer pattern data in (10) is far from an ideal Bayer pattern representation of the true image $\boldsymbol{x}(\boldsymbol{n})$ we hope to recover from $s(\boldsymbol{n})$, however. One distortion we see is the unwanted filtering term $\sum_{\boldsymbol{\theta}\in\mathbb{Z}_2^2} e^{j\boldsymbol{\omega}^T\boldsymbol{\theta}/2}$ that degrades the baseband luminance/green signal $X_g(\boldsymbol{\omega})$. Another complication is that the antialiasing is only *partially* effective, allowing aliasing to corrupt the baseband $X_g(\boldsymbol{\omega})$ near $\boldsymbol{\omega} = \pm[0, \frac{\pi}{4}]^T, \pm[\frac{\pi}{4}, 0]^T, \pm[\frac{\pi}{4}, \frac{\pi}{4}]^T, \pm[\frac{\pi}{4}, \frac{-\pi}{4}]^T$.

Contrary to the popular belief that Kodak PIXELUX binning results in $2 \times 2$ reduction in resolution, the main conclusion we draw from (9) is that the "Nyquist rate" of this binning scheme is $\pi/4$ due to high risk of aliasing—implying that the *actual* resolution loss is $4 \times 4$, far worse than the presumed $2 \times 2$. *Even if* this Nyquist rate did not cause problems (e.g. increase sensor resolution), $s$ does not escape the unwanted filtering term in (9)—this cannot be eliminated simply by increasing sensor resolution. Hence when a demosaicking algorithm is applied to the superpixel Bayer pattern data $s$, what is expected is a filtered and aliased image that we have already seen in Figure 2.

### 3.3 Binning "subsampling"

Below, we offer an alternative perspective to the analysis of Section 3.2. The analytical results contained herein

will provide the basis for the proposed binning-aware demosaicking algorithm. Continuing with the analysis of PIXELUX, consider Figure 6a which displays data equivalent to the superpixels of Figure 1c. The superpixels are placed at the center of the four averaged pixels, denoting the implied superpixel positions. Other locations are given 0 value. This data can be represented by applying a convolution operator followed by subsampling, as follows:

- *filtering*: The charge summation in PIXELUX is

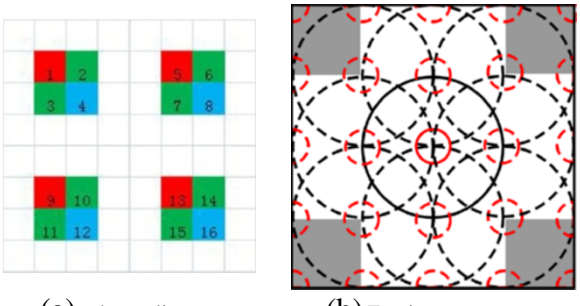
$$y_{\text{bin}}(\boldsymbol{n}) = y(\boldsymbol{n}) \star h_{\text{bin}}(\boldsymbol{n}).$$

- *subsampling*: to yield the binning subsampling data $t$, do

$$\begin{aligned} t(4\boldsymbol{n}) &= y_{\text{bin}}(4\boldsymbol{n}) \\ t\left(4\boldsymbol{n}+\begin{pmatrix}0\\1\end{pmatrix}\right) &= y_{\text{bin}}\left(4\boldsymbol{n}+\begin{pmatrix}0\\1\end{pmatrix}\right) \\ t\left(4\boldsymbol{n}+\begin{pmatrix}1\\0\end{pmatrix}\right) &= y_{\text{bin}}\left(4\boldsymbol{n}+\begin{pmatrix}1\\0\end{pmatrix}\right) \\ t\left(4\boldsymbol{n}+\begin{pmatrix}1\\1\end{pmatrix}\right) &= y_{\text{bin}}\left(4\boldsymbol{n}+\begin{pmatrix}1\\1\end{pmatrix}\right) \\ t(\boldsymbol{n}) &= 0 \text{ otherwise.} \end{aligned} \tag{11}$$

With arithmetic, the Fourier transform of $t$ is deduced to:

$$\begin{aligned} T(\boldsymbol{\omega}) \approx & \sum_{\boldsymbol{\lambda}\in\frac{\pi}{2}\mathbb{Z}_4^2} \frac{X_\alpha(\boldsymbol{\omega}-\boldsymbol{\lambda}) + e^{j\{\boldsymbol{\lambda}^T\binom{1}{1}\}} X_\beta(\boldsymbol{\omega}-\boldsymbol{\lambda})}{4} \\ & + \sum_{\boldsymbol{\lambda}\in\frac{\pi}{2}\mathbb{Z}_4^2} \sum_{\boldsymbol{\theta}\in\mathbb{Z}_2^2} \frac{e^{j\{\boldsymbol{\lambda}^T\boldsymbol{\theta}\}} H_{\text{bin}}(\boldsymbol{\omega}-\boldsymbol{\lambda}) X_g(\boldsymbol{\omega}-\boldsymbol{\lambda})}{16}. \end{aligned} \tag{12}$$




(a) subsampling (b) Fourier support

**Figure 6 Binning subsampling is an alternative interpretation to the binning sampling in Figure 1.** (**a**) Subsampled data $t(\boldsymbol{n})$ in (11) equivalent to the superpixel Bayer pattern of Figure 1c. (**b**) Idealized spectral support of binning subsampled data $T(\boldsymbol{\omega})$ in (12). The baseband signal $X_g$ is *free of aliasing* in the shaded region. As before, solid lines indicate the baseband signals, while spectra with the dashed lines arises as a result of CFA sampling. Black and red lines correspond to the support of luminance and chrominance images, respectively.

Note that the summation over $\lambda$ suggests 16 modulations. However, except $\boldsymbol{\lambda} \in \frac{\pi}{2}\mathbb{Z}_3^2$, other $\lambda$ results in $\sum_{\boldsymbol{\theta}\in\mathbb{Z}_2^2} e^{j\{\boldsymbol{\lambda}^T\boldsymbol{\theta}\}}$ is 0, as shown in Figure 7. The support of this transform is illustrated in Figure 6b.[a] As evidenced by this figure, the modulated baseband signal components $X_g(\boldsymbol{\omega}-\boldsymbol{\lambda})$ overlap each other almost entirely—that is, they are aliased. However, the shaded regions of Figure 6b are still free of aliasing. Indeed, this uncorrupted portion of the Fourier support is the key to post-binning processing that is the subject of next section.

## 4 Binning-aware demosaicking

Motivated by the analysis of pixel binning subsampling in (12), we now present a novel binning-aware demosaicking aimed at recovering full-color image $\boldsymbol{x}$ without introducing binning artifacts. We accomplish this in three stages.

### Step 1: Chrominance estimation

Drawing parallels to [19], we assume that *local* image features are either vertically or horizontally oriented (approximately). If this assumption holds, certain subsets of the modulated chrominances in (11) are assumed to be alias-free conditional under the vertically or horizontally oriented image features—this is illustrated in Figures 8a. For example, assuming *horizontal feature*, an amplitude demodulation recovers the desired chrominance images $x_\alpha$ and $x_\beta$:

$$\begin{bmatrix}\hat{x}_{\alpha,h}\\ \hat{x}_{\beta,h}\end{bmatrix} = \underbrace{\begin{bmatrix}1 & 1\\ 1 & -1\\ 1 & -j\\ 1 & j\end{bmatrix}^\dagger}_{\text{from (12)}} \underbrace{\begin{bmatrix} h_0(\boldsymbol{n}) \star \left(t(\boldsymbol{n})\cdot e^{j\{\boldsymbol{n}^T\binom{\pi}{\pi}\}}\right)\\ h_0(\boldsymbol{n}) \star \left(t(\boldsymbol{n})\cdot e^{j\{\boldsymbol{n}^T\binom{0}{\pi}\}}\right)\\ h_0(\boldsymbol{n}) \star \left(t(\boldsymbol{n})\cdot e^{j\{\boldsymbol{n}^T\binom{-\pi/2}{\pi}\}}\right)\\ h_0(\boldsymbol{n}) \star \left(t(\boldsymbol{n})\cdot e^{j\{\boldsymbol{n}^T\binom{\pi/2}{\pi}\}}\right)\end{bmatrix}}_{\text{demodulation}}, \tag{13}$$

where $(\cdot)^\dagger$ denotes a pseudo inverse matrix and $h_0$ is a lowpass filter whose passbands matches the support of $X_\alpha$ and $X_\beta$. The reconstruction of vertically oriented image

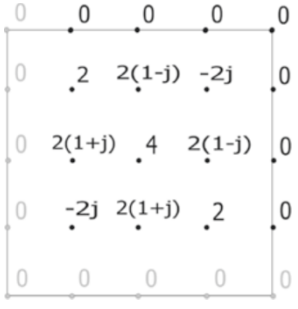



**Figure 7 Fourier transform of $\sum_{\boldsymbol{\theta}\in\mathbb{Z}_2^2} e^{j\{\boldsymbol{\lambda}^T\boldsymbol{\theta}\}} M_{\text{bin}}(\boldsymbol{\omega})$.**

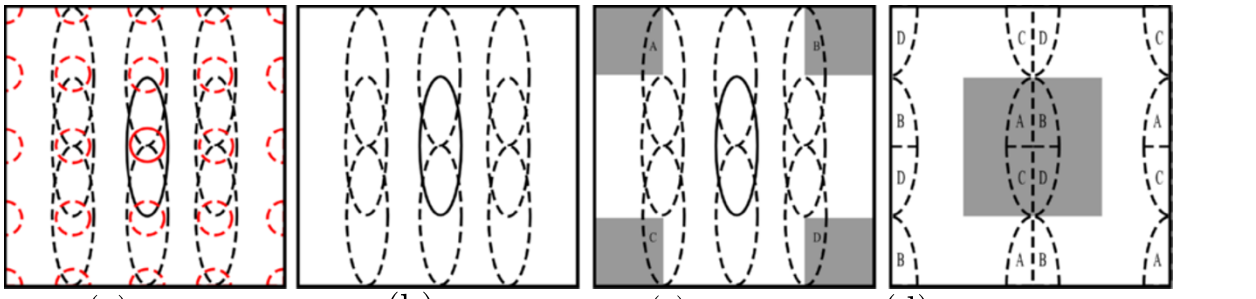



**Figure 8 Idealized spectral support of binning subsampled data, at various stages of binning-aware demosaicking.** Shaded regions denote filter support. See text.

feature (denoted $\hat{x}_{\alpha,v}, \hat{x}_{\beta,v}$) is same as (13) but with 90° rotation.

#### Step 2: Luminance filtering

Once the $\hat{x}_{\alpha,h}$ and $\hat{x}_{\beta,h}$ are recovered, we compute the green image $\hat{x}_{g,h}$. Subtracting

$$\sum_{\boldsymbol{m}\in 4\mathbb{Z}^2} \hat{x}_{\alpha,h}(\boldsymbol{m})\Delta(\boldsymbol{m}-\boldsymbol{n})+\hat{x}_{\beta,h}(\boldsymbol{m}+\begin{pmatrix}1\\1\end{pmatrix})\Delta(\boldsymbol{m}+\begin{pmatrix}1\\1\end{pmatrix}-\boldsymbol{n})$$

from subsampled binning data $t(\boldsymbol{n})$ results a Fourier transform that is comprised only of $x_g$: (from (12))

$$T_g(\boldsymbol{\omega}) = \sum_{\boldsymbol{\lambda}\in\frac{\pi}{2}\mathbb{Z}_4^2}\sum_{\boldsymbol{\theta}\in\mathbb{Z}_2^2} \frac{e^{j\{\boldsymbol{\lambda}^T\boldsymbol{\theta}\}}H_{\text{bin}}(\boldsymbol{\omega}-\boldsymbol{\lambda})X_g(\boldsymbol{\omega}-\boldsymbol{\lambda})}{16}. \tag{14}$$

This is illustrated in Figure 8b. To reconstruct the green image $\hat{x}_{g,h}$ from the unaliased (shaded in Figure 8b) portions of $t_g$, we carry out a standard demodulation, as follows:

$$\hat{x}_{g,h} = \underbrace{h_2(\boldsymbol{n}) \star \{\overbrace{f(\boldsymbol{n}) \cdot \{\underbrace{h_1(\boldsymbol{n}) \star t_g(\boldsymbol{n})}_{\text{isolate unaliased}}\}}^{\text{modulation}}\}}_{\text{isolate signal}},$$

where $h_1$ and $h_2$ are lowpass and highpass filters, respectively; and $f$ is a sum of sinusoids intended for modulation, as follows:

$$H_1(\boldsymbol{\omega}) = \begin{cases} 1 & \text{if } |\omega_1| > \frac{\pi}{2} \text{ and } |\omega_2| > \frac{\pi}{2} \\ 0 & \text{else} \end{cases}$$

$$H_2(\boldsymbol{\omega}) = \begin{cases} 1 & \text{if } |\omega_1| < \frac{\pi}{2} \text{ and } |\omega_2| < \frac{\pi}{2} \\ 0 & \text{else} \end{cases} \tag{15}$$

$$F(\boldsymbol{\omega}) = \sum_{\boldsymbol{\lambda}\in\left\{\binom{\pm\pi}{\pm\pi}/4\right\}} \frac{\delta(\boldsymbol{\omega}+\boldsymbol{\lambda})}{\sum_{\boldsymbol{\theta}\in\mathbb{Z}_2^2} e^{j\{\boldsymbol{\lambda}^T\boldsymbol{\theta}\}}}.$$

As illustrated in Figure 8c,d, the modulation by $f(\boldsymbol{n})$ not only shifts the spectrums, but also creates additional aliasing copies. Hence, the filter $h_2$ is needed to attenuate them. The same procedure can be used to find the green image $\hat{x}_{g,v}$ based on $\hat{x}_{\alpha,v}$ and $\hat{x}_{\beta,v}$.

#### Step 3: Directional selection

Once $\hat{\boldsymbol{x}}_h = \{\hat{x}_{g,h}, \hat{x}_{\alpha,h}, \hat{x}_{\beta,h}\}$ and $\hat{\boldsymbol{x}}_v = \{\hat{x}_{g,v}, \hat{x}_{\alpha,v}, \hat{x}_{\beta,v}\}$ are found, they must be combined to yield the final estimate, $\hat{\boldsymbol{x}}_t = \{\hat{x}_g, \hat{x}_\alpha, \hat{x}_\beta\}$ via the convex combination (3). As already mentioned, the directional selection variable $\tau$ has received considerable attention in research and many techniques are available. However, these studies often lack analysis under noise—although binning reduces noise considerably, most directional selection variables are nevertheless sensitive to random perturbations.

To address the problem of directional selection under noise, we modified the $\tau$ criteria used in the popular adaptive homogeneity directed (AHD) demosaicking method as follows:

$$\hat{\tau}(\boldsymbol{n}) = \arg\max_{\tau\in[0,1]} u_{\hat{x}_\tau}(\boldsymbol{n}) \tag{16}$$

$$\hat{\boldsymbol{x}}_t(\boldsymbol{n}) = \hat{\boldsymbol{x}}_{\hat{\tau}(\boldsymbol{n})}(\boldsymbol{n}), \tag{17}$$

where $\hat{\boldsymbol{x}}_\tau$ and $u_{\hat{x}}$ are as defined in (3) and (4), respectively. Contrast this to the original AHD formulation which selected either $\hat{\boldsymbol{x}}_h$ or $\hat{\boldsymbol{x}}_v$ (i.e. $\tau \in \{0, 1\}$ instead of $\tau \in [0, 1]$) as the final output $\hat{\boldsymbol{x}}_t$. The modified strategy of (16) behaves similarly to the original AHD near the edges of an image, but encourages averaging in the flat regions of the image. It was found empirically to be far more robust to directional selection under noise.

## 5 Experimental validation

### 5.1 Setup

The proposed binning-aware demosaicking $\hat{\boldsymbol{x}}_t(\boldsymbol{n})$ in (16) is compared to four available alternatives ($\hat{\boldsymbol{x}}_s(\boldsymbol{n})$, $\hat{\boldsymbol{x}}_p(\boldsymbol{n})$,

$\hat{\boldsymbol{x}}_y(\boldsymbol{n})$, and $\hat{\boldsymbol{x}}_{y'}(\boldsymbol{n})$). The first is a state-of-the-art demosaicking method [19] applied to superpixels $s(\boldsymbol{n})$ (i.e. output from PIXELUX binning):

$$\hat{\boldsymbol{x}}_s(\boldsymbol{n}) = \text{demosaicking}(s(\boldsymbol{n})).$$

The second is the same demosaicking method [19] applied to PhaseOne binning superpixels $p(\boldsymbol{n})$:

$$\hat{\boldsymbol{x}}_p(\boldsymbol{n}) = \text{demosaicking}(p(\boldsymbol{n})).$$

The third is the application of the same demosaicking method [19] to a full resolution CFA $y(\boldsymbol{n})$ (i.e. without binning):

$$\hat{\boldsymbol{x}}_y(\boldsymbol{n}) = \text{demosaicking}(y(\boldsymbol{n})).$$

The fourth is a simulation of a lower resolution sensor. Let $\boldsymbol{x}'(\boldsymbol{n})$ denote the downsampled version of the ideal lowpassed (antialiased) image:

$$\boldsymbol{x}'(\boldsymbol{n}) = \{h_2 \star \boldsymbol{x}\}(2\boldsymbol{n}). \tag{18}$$

The CFA subsampled data captured by this lower resolution sensor is then

$$y'(\boldsymbol{n}) = \boldsymbol{c}(\boldsymbol{n})^T \boldsymbol{x}'(\boldsymbol{n}),$$

where $\boldsymbol{c} : \mathbb{Z}^2 \rightarrow [0,1]^3$ is same the translucency of CFA used in (1). The application of the same demosaicking method [19] to lower resolution CFA $y'(\boldsymbol{n})$ is:

$$\hat{\boldsymbol{x}}_{y'}(\boldsymbol{n}) = \text{demosaicking}(y'(\boldsymbol{n})).$$

The output images from the proposed method ($\hat{\boldsymbol{x}}_t$) and the full resolution demosaicking ($\hat{\boldsymbol{x}}_y$) have the same size as the original image $\boldsymbol{x}$. On the other hand, the conventional binning processing are based on superpixel sampling, so the pixel density of $\hat{\boldsymbol{x}}_s$ and $\hat{\boldsymbol{x}}_p$ is just a quarter of the original image (same is true also for $\hat{\boldsymbol{x}}_{y'}$). Hence when we compare all results (Figures 9, 10, 11, 12, 13; Table 1), we downsample $\hat{\boldsymbol{x}}_t$ and $\hat{\boldsymbol{x}}_{y'}$ by $2 \times 2$ (in the same manner as (18)) such that all results have the same pixel density as the lower resolution image $\boldsymbol{x}'$.

The linear images used in this simulation study are a part of the collection of [25,26], examples of which are shown in Figure 9. Numerical scores in Table 1 and Figure 13 were obtained by averaging performance over 84 images. Noise is simulated by adding pseudorandom white Gaussian noise to the CFA data $y(\boldsymbol{n})$, the superpixel CFA data $s(\boldsymbol{n})$ and $p(\boldsymbol{n})$, and the lower resolution CFA data $y'(\boldsymbol{n})$. In the experiments, the 12 bit image data in [25,26] were renormalized to ranges 0–1—meaning noise standard deviation $\sigma_n = 0.01$ correspond to standard deviation of 40.96 in a 12 bit camera processing pipeline, etc. Considering the noise models in (5–7), one may ask if such a simplified noise model is appropriate. As evidenced by the analysis in (7), however, the difference between SNR and $\text{SNR}_{\text{sum}}$ is $M$ (the number of pixels combined together); and the difference between $\text{SNR}_{\text{sum}}$ and $\text{SNR}_{\text{bin}}$ is the read noise power $N$. Hence the SNR gains in binning is attributed only to the signal-independent portion of the noise, and *not* on the signal dependent portion. Furthermore, the read noise dominates in the low light regime. Hence simulated additive white Gaussian noise suffices for experimental verification. The binning subsample signal $t(\boldsymbol{n})$ represents the same data as $s(\boldsymbol{n})$ and is computed by upsampling $s(\boldsymbol{n})$ (insert zeros where necessary).

### 5.2 Results

Example outputs from four different methods ($\hat{\boldsymbol{x}}_{y'}, \hat{\boldsymbol{x}}_y, \hat{\boldsymbol{x}}_s, \hat{\boldsymbol{x}}_p, \hat{\boldsymbol{x}}_t$) are shown in Figures 10, 11 and 12. As expected, demosaicking applied to a full resolution CFA ($\hat{\boldsymbol{x}}_y$) has a noisy appearance due to low SNR of individual pixels. However, edges and image features are clearly defined even after downsampling thanks to the full resolution description. Demosaicking applied to superpixel CFAs ($\hat{\boldsymbol{x}}_s, \hat{\boldsymbol{x}}_p$), on the other hand, yields the opposite qualities—the noise is significantly reduced owing to high SNR of binning, but the image suffers from severe artifacts stemming from aliasing in (10). More specifically, the aliasing in Kodak PIXELUX binning manifests itself as a pixelization artifact, while PhaseOne binning results in zippering artifacts. However, one may argue that the aliasing artifacts in $\hat{\boldsymbol{x}}_p$ become less bothersome at the highest level of noise because the zippering and noise become less distinguishable. By contrast, the proposed binning-aware demosaicking method ($\hat{\boldsymbol{x}}_t$) succeeds in suppressing noise while preserving the image features. Of particular interest is the comparison between $\hat{\boldsymbol{x}}_s$ and $\hat{\boldsymbol{x}}_t$, since they both use Kodak PIXELUX binning but the proposed method

**Figure 9 Example of images used in experiment (zoomed).**

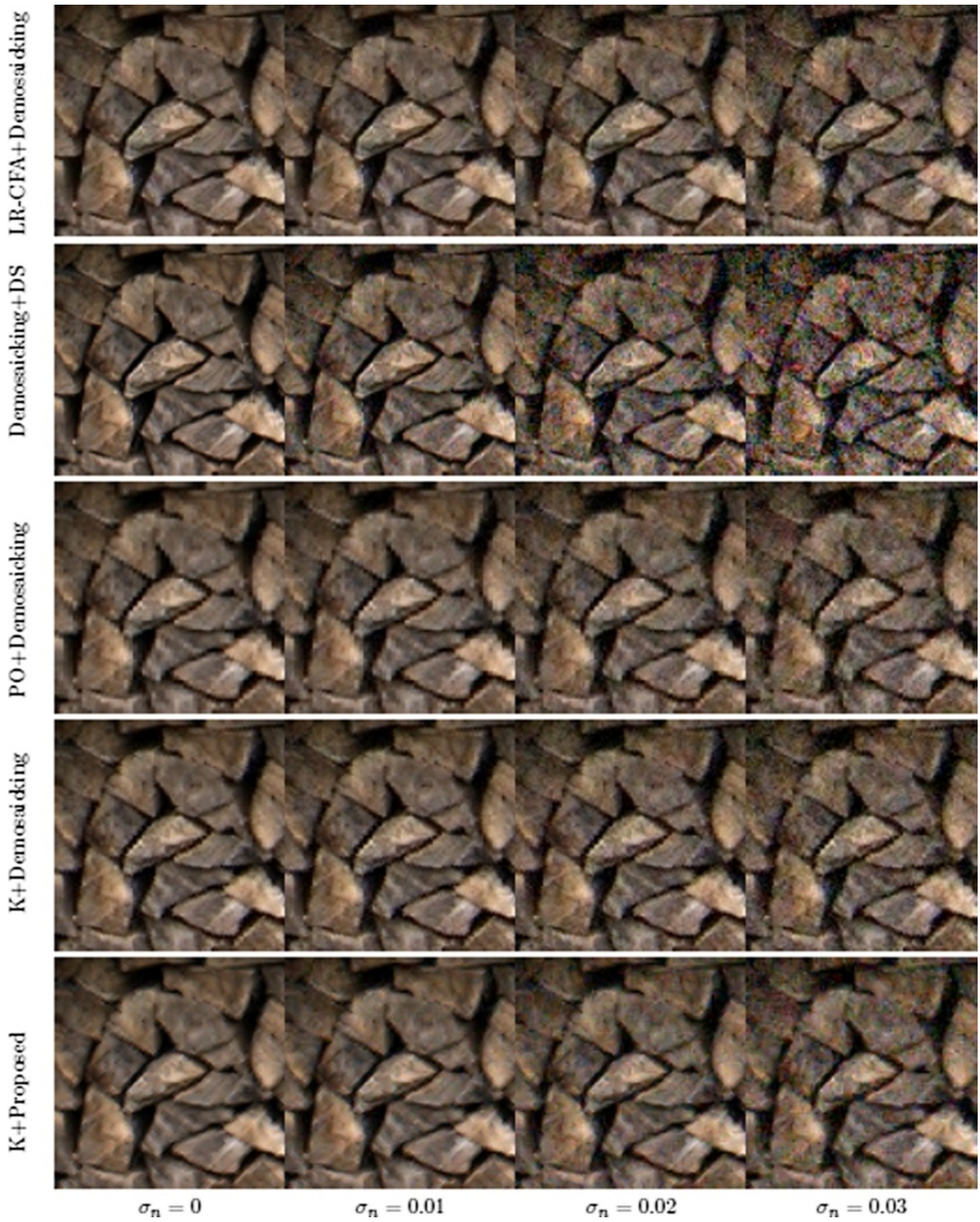



**Figure 10 Reconstructed images with various noise levels.** Demosaicking method used for comparison is that of [19]. Here, LR ≜ low resolution, DS ≜ downsample, PO ≜ PhaseOne [5], K ≜ Kodak PIXELUX [7].

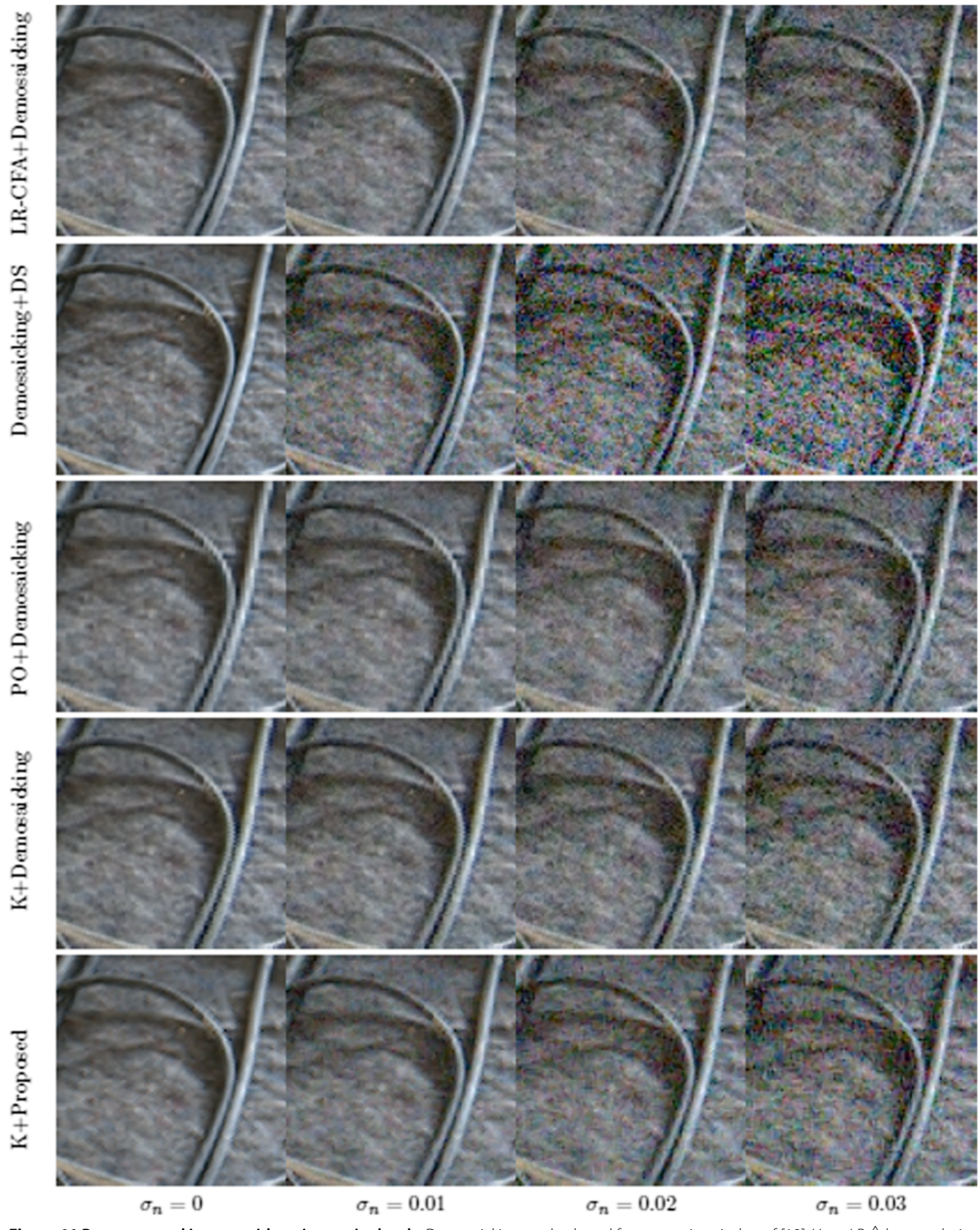



**Figure 11 Reconstructed images with various noise levels.** Demosaicking method used for comparison is that of [19]. Here, LR ≜ low resolution, DS ≜ downsample, PO ≜ PhaseOne [5], K ≜ Kodak PIXELUX [7].

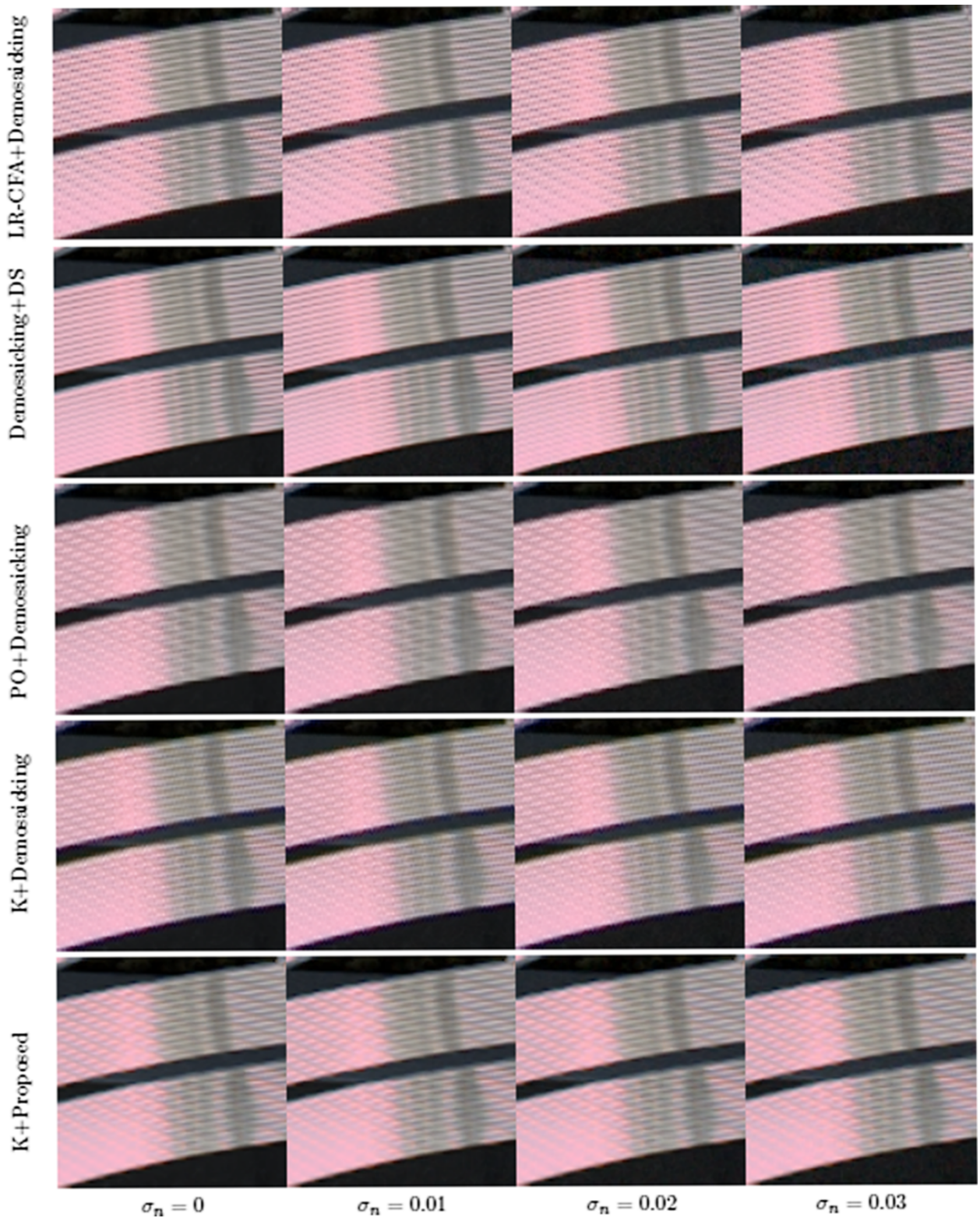



**Figure 12 Reconstructed images with various noise levels.** Demosaicking method used for comparison is that of [19]. Here, LR $\triangleq$ low resolution, DS $\triangleq$ downsample, PO $\triangleq$ PhaseOne [5], K $\triangleq$ Kodak PIXELUX [7].

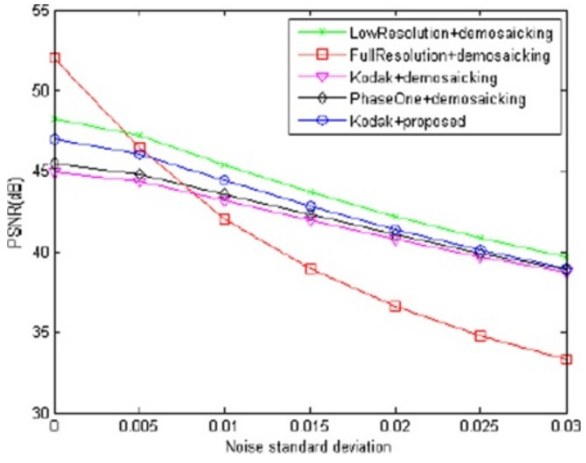



**Figure 13 PSNR (red/green/blue pixels are combined) of each methods, Kodak and PhaseOne refers to the binning methods of [7] and [5], respectively.** Demosaicking method used for comparison is that of [19].

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Figure 14 Binning sampling filter $m_{\text{bin}}(\boldsymbol{n})$.**

yield drastically improved outcomes. Overall, the proposed method has better visual quality than $\hat{\boldsymbol{x}}_s$ and $\hat{\boldsymbol{x}}_p$ for $\sigma_n < 0.03$; but proposed has a slightly noisier appearance at the highest level of noise ($\sigma_n = 0.03$). Finally, the output from the low resolution camera $\hat{\boldsymbol{x}}_{y'}$ is both robust to noise and aliasing. This is expected, as lower resolution CFA data $y'(\boldsymbol{n})$ does not share the problems that superpixel CFAs $p(\boldsymbol{n}), s(\boldsymbol{n}), t(\boldsymbol{n})$ have. However, $\hat{\boldsymbol{x}}_y$ has superior reconstruction over $\hat{\boldsymbol{x}}_{y'}$ without noise ($\sigma_n = 0$).

**Table 1 Reconstruction performance in PSNR with various noise levels**

| Noise ($\sigma_n$) | Color | LR-CFA | HR-CFA | HR-CFA + binning | | |
|---|---|---|---|---|---|---|
| | | [19] | [19] + DS | PO + [19] | K + [19] | K + Proposed |
| 0.000 | R | 48.1787 | 51.4671 | 45.2485 | 45.3275 | 47.8061 |
| | G | 51.8147 | 54.7110 | 46.5344 | 47.1429 | 48.6259 |
| | B | 46.5116 | 50.9532 | 43.6715 | 44.5752 | 45.5689 |
| 0.005 | R | 47.1788 | 46.5149 | 44.6596 | 44.7572 | 46.8994 |
| | G | 50.1223 | 48.5953 | 46.3633 | 41.5903 | 47.6607 |
| | B | 45.6221 | 45.2237 | 43.1426 | 43.9331 | 44.6091 |
| 0.010 | R | 45.4430 | 42.1931 | 43.4914 | 43.5929 | 45.2686 |
| | G | 47.6553 | 43.9223 | 44.5383 | 44.9318 | 45.9540 |
| | B | 44.0420 | 40.7293 | 42.0733 | 42.7001 | 42.9732 |
| 0.015 | R | 43.7677 | 39.1858 | 42.2368 | 42.3349 | 43.6632 |
| | G | 45.5264 | 40.7919 | 4.18183 | 43.4857 | 44.3119 |
| | B | 42.4968 | 37.6716 | 40.9172 | 41.4211 | 41.3987 |
| 0.020 | R | 42.2672 | 36.9118 | 41.0434 | 41.1293 | 42.2202 |
| | G | 43.7388 | 38.4621 | 41.9128 | 42.1503 | 42.8488 |
| | B | 41.0955 | 35.3795 | 39.8064 | 40,2189 | 39.9910 |
| 0.025 | R | 40.9430 | 35.0944 | 39.9375 | 40.0160 | 40.9391 |
| | G | 42.2200 | 36.6133 | 40.7530 | 40.9454 | 41.5568 |
| | B | 39.8450 | 33.5524 | 38.7712 | 39.1164 | 38.7417 |
| 0.030 | R | 39.7693 | 33.5816 | 38.9267 | 38.9959 | 39.7997 |
| | G | 40.9107 | 35.0803.9327 | 39.7008 | 39.8595 | 40.4102 |
| | B | 38.7310 | 32.0235.3426 | 37.8168 | 38.1076 | 37.6279 |

Demosaicking method used for comparison is that of [19]. Here, LR ≙ low resolution, HR ≙ high resolution, DS ≙ downsample, PO ≙ PhaseOne [5], K ≙ Kodak PIXELUX [7] (assumes pixel range [0,1]).

Figure 12 shows an example where none of the reconstruction methods produced a satisfactory output (except for $\hat{\boldsymbol{x}}_y$ under no noise).

The performance is evaluated also in terms of peak SNR, using the downsampled version of the ideal low-passed (antialiased) image $\boldsymbol{x}'$ in (18) as their reference. The results are summarized in Table 1. When there is no noise ($\sigma_n = 0$), ordinary demosaicking reconstruction $\hat{\boldsymbol{x}}_y$ and lower resolution sensor $\hat{\boldsymbol{x}}_{y'}$ yields the best results, as expected. However, the proposed $\hat{\boldsymbol{x}}_t$ is a very close third, yielding comparably satisfactory results. Binning result $\hat{\boldsymbol{x}}_s$ is worst by far due to binning artifacts.

When noise is taken into consideration, the quality of $\hat{\boldsymbol{x}}_y$ suffers greatly as expected. Even with noise variance as little as $\sigma_n = 0.005$, the performance of $\hat{\boldsymbol{x}}_y$ deteriorates significantly, while performance of $\hat{\boldsymbol{x}}_s$, $\hat{\boldsymbol{x}}_p$, $\hat{\boldsymbol{x}}_t$, and $\hat{\boldsymbol{x}}_{y'}$ in terms of PSNR are far less sensitive to noise. With moderate noise levels ($\sigma_n < 0.03$) the proposed binning-aware demosaicking clearly outperforms the artifact-plagued demosaicking of superpixels. With the largest noise level considered ($\sigma_n = 0.03$), PSNR performances of $\hat{\boldsymbol{x}}_s$, $\hat{\boldsymbol{x}}_p$, and $\hat{\boldsymbol{x}}_t$ are closer to each other because deteriorations in output images are dominated by noise (rather than by artifacts).

The analysis in Figures 10, 11, 12, 13 and Table 1 sheds a light on the decades-old debate about resolution versus noise. On one hand, the lower resolution sensor delivers consistent performance under noise ($\hat{\boldsymbol{x}}_{y'}$). However, Figure 11 shows that under no noise, extra sensor resolution is still desirable. Consider Figure 13. The comparison between green (low resolution) and red (high resolution) curves is consitent with the image quality of Figures 10 and 11. With the availability of pixel binning, we would compare the green curve with the "max function" over the red and blue (binning) curves in Figure 13. Hence one can think of binning as a way to narrow the gap between the red and green curves in noise, without making sacrifices to the advantages of higher spatial resolution.

## 6 Conclusion

In this article, we proved via a rigorous analysis of *binning sampling* that Kodak PIXELUX binning scheme results in $4 \times 4$ reduction in image resolution—contrary to the popular belief that binning of four pixels should result in $2 \times 2$ reduction in resolution. We proposed a binning-aware demosaicking algorithm based on the Fourier analysis of *binning subsampling* to combine unaliased copies of the Fourier spectra together via the demodulation. The resultant method succeeds in reconstructing the color image with only $2 \times 2$ resolution loss—or increasing the resolution by $2 \times 2$ over the traditional approach of applying demosaicking to superpixels. The binning-aware demosaicking also succeeds in suppressing noise and preserving image details. We verified experimentally that the binning-aware demosaicking outperforms the alternatives.

## Appendix 1: Proof of Fourier Representation of binning subsampling

We provide the proof for Equation (12). Let $H_{\text{bin}}$ be the Fourier transform of (8). Then the combination of charges can be represented as:

$$\begin{aligned} X_{r_{\text{bin}}}(\boldsymbol{\omega}) &\triangleq X_r(\boldsymbol{\omega})H_{\text{bin}}(\boldsymbol{\omega}) \\ X_{g_{\text{bin}}}(\boldsymbol{\omega}) &\triangleq X_g(\boldsymbol{\omega})H_{\text{bin}}(\boldsymbol{\omega}) \\ X_{b_{\text{bin}}}(\boldsymbol{\omega}) &\triangleq X_b(\boldsymbol{\omega})H_{\text{bin}}(\boldsymbol{\omega}). \end{aligned}$$

Due to band limitedness of $X_\alpha$ and $X_\beta$, the following approximation hold:

$$\begin{aligned} X_{\alpha_{\text{bin}}}(\boldsymbol{\omega}) &\triangleq X_{r_{\text{bin}}}(\boldsymbol{\omega}) - X_{g_{\text{bin}}}(\boldsymbol{\omega}) = H_{\text{bin}}(\boldsymbol{\omega})X_\alpha(\boldsymbol{\omega}) \approx 4X_\alpha(\boldsymbol{\omega}) \\ X_{\beta_{\text{bin}}}(\boldsymbol{\omega}) &\triangleq X_{b_{\text{bin}}}(\boldsymbol{\omega}) - X_{g_{\text{bin}}}(\boldsymbol{\omega}) = H_{\text{bin}}(\boldsymbol{\omega})X_\beta(\boldsymbol{\omega}) \approx 4X_\beta(\boldsymbol{\omega}). \end{aligned} \tag{19}$$

where we used the fact that $H_{\text{bin}}(0) = 4$.

Define $m_{\text{bin}}(\boldsymbol{n}) = \sum_{\boldsymbol{\theta} \in \mathbb{Z}_4^2} \Delta(\boldsymbol{n} - \boldsymbol{\theta})$, as illustrated in Figure 14. The binning subsampling data $t(\boldsymbol{n})$ refers to the concept of combining the electrial charges of four neighboring pixels together to form a *superpixel*. The process is illustrated in Figure 6a. Mathmatically, $t(\boldsymbol{n})$ can be written as:

$$\begin{aligned} t(\boldsymbol{n}) =& m_{\text{bin}}(\boldsymbol{n})x_{r_{\text{bin}}}(\boldsymbol{n}) + m_{\text{bin}}\left(\boldsymbol{n} + \begin{pmatrix}1\\0\end{pmatrix}\right)x_{g_{\text{bin}}}(\boldsymbol{n}) \\ &+ m_{\text{bin}}\left(\boldsymbol{n} + \begin{pmatrix}0\\1\end{pmatrix}\right)x_{g_{\text{bin}}}(\boldsymbol{n}) + m_{\text{bin}}\left(\boldsymbol{n} + \begin{pmatrix}1\\1\end{pmatrix}\right)x_{b_{\text{bin}}}(\boldsymbol{n}) \\ =& m_{\text{bin}}(\boldsymbol{n})x_{\alpha_{\text{bin}}}(\boldsymbol{n}) + m_{\text{bin}}\left(\boldsymbol{n} + \begin{pmatrix}1\\1\end{pmatrix}\right)x_{\beta_{\text{bin}}}(\boldsymbol{n}) \\ &+ \sum_{\boldsymbol{\theta} \in \mathbb{Z}_2^2} m_{\text{bin}}(\boldsymbol{n} + \theta)x_{g_{\text{bin}}}(\boldsymbol{n}). \end{aligned}$$

In the Fourier domain, $t(\boldsymbol{n})$ can be expressed as

$$\begin{aligned} T(\boldsymbol{\omega}) =& M_{\text{bin}}(\boldsymbol{\omega}) \star X_{\alpha_{\text{bin}}}(\boldsymbol{\omega}) + (e^{j\{\boldsymbol{\omega}^T\binom{1}{1}\}}M_{\text{bin}}(\boldsymbol{\omega})) \star X_{\beta_{\text{bin}}}(\boldsymbol{\omega}) \\ &+ \sum_{\boldsymbol{\theta} \in \mathbb{Z}_2^2} e^{j\{\boldsymbol{\lambda}^T\boldsymbol{\theta}\}}M_{\text{bin}}(\boldsymbol{\omega}) \star X_{g_{\text{bin}}}(\boldsymbol{\omega}) \end{aligned}$$

where

$$M_{\text{bin}}(\boldsymbol{\omega}) = \sum_{\boldsymbol{\lambda} \in \frac{\pi}{2}\mathbb{Z}_4^2} \frac{\delta(\boldsymbol{\omega} - \boldsymbol{\lambda})}{16}.$$

With arithmetic and approximation of (19), the Fourier transform of $t(\boldsymbol{n})$ simplifies to:

$$\begin{aligned} T(\boldsymbol{\omega}) \approx& \sum_{\boldsymbol{\lambda} \in \frac{\pi}{2}\mathbb{Z}_4^2} \frac{X_\alpha(\boldsymbol{\omega} - \boldsymbol{\lambda}) + e^{j\{\boldsymbol{\lambda}^T\binom{1}{1}\}}X_\beta(\boldsymbol{\omega} - \boldsymbol{\lambda})}{4} \\ &+ \sum_{\boldsymbol{\lambda} \in \frac{\pi}{2}\mathbb{Z}_4^2} \sum_{\boldsymbol{\theta} \in \mathbb{Z}_2^2} \frac{e^{j\{\boldsymbol{\lambda}^T\boldsymbol{\theta}\}}H_{\text{bin}}(\boldsymbol{\omega} - \boldsymbol{\lambda})X_g(\boldsymbol{\omega} - \boldsymbol{\lambda})}{16}. \end{aligned}$$

The Fourier support of $T(\omega)$ is illustrated in Figure 6b. Note that the summation over $\lambda$ suggests that binning subsampling will result in 16 modulations. However, $\sum_{\boldsymbol{\theta}\in\mathbb{Z}_2^2} e^{j\{\boldsymbol{\lambda}^T\boldsymbol{\theta}\}}$ is 0 for many values of $\boldsymbol{\lambda}$, as shown in Figure 7. As a result, there are only nine *actual* modulations.

## Appendix 2: Proof of Fourier representation of binning sampling

We provide the proof for Equation (10). The binning sampling data $s(\boldsymbol{n})$ refers to the concept of combining the electrial charges of four neighboring pixels together to form a *superpixel* Bayer pattern. The process is illustrated in Figures 1a,c. Similar to binning subsampling (see Appendix 1, binning sampling $s(\boldsymbol{n})$ has the following representation (it is mathmatically convenint to consider $s(\frac{\boldsymbol{n}}{2})$ for $\boldsymbol{n}$ even, rather than $s(\boldsymbol{n})$ directly);

$$\begin{aligned}
s\left(\frac{\boldsymbol{n}}{2}\right) =& m_{\text{bin}}(\boldsymbol{n}) x_{r_{\text{bin}}}(\boldsymbol{n}) + m_{\text{bin}}\left(\boldsymbol{n}+\begin{pmatrix}2\\0\end{pmatrix}\right) x_{g_{\text{bin}}}\left(\boldsymbol{n}+\begin{pmatrix}1\\0\end{pmatrix}\right)\\
&+ m_{\text{bin}}\left(\boldsymbol{n}+\begin{pmatrix}0\\2\end{pmatrix}\right) x_{g_{\text{bin}}}\left(\boldsymbol{n}+\begin{pmatrix}0\\1\end{pmatrix}\right)\\
&+ m_{\text{bin}}\left(\boldsymbol{n}+\begin{pmatrix}2\\2\end{pmatrix}\right) x_{b_{\text{bin}}}\left(\boldsymbol{n}+\begin{pmatrix}1\\1\end{pmatrix}\right)\\
=& m_{\text{bin}}(\boldsymbol{n}) x_{\alpha_{\text{bin}}}(\boldsymbol{n}) + m_{\text{bin}}\left(\boldsymbol{n}+\begin{pmatrix}2\\2\end{pmatrix}\right) x_{\beta_{\text{bin}}}\left(\boldsymbol{n}+\begin{pmatrix}1\\1\end{pmatrix}\right)\\
&+ m_{\text{bin}}(\boldsymbol{n}) x_{g_{\text{bin}}}(\boldsymbol{n}) + m_{\text{bin}}\left(\boldsymbol{n}+\begin{pmatrix}2\\0\end{pmatrix}\right) x_{g_{\text{bin}}}\left(\boldsymbol{n}+\begin{pmatrix}1\\0\end{pmatrix}\right)\\
&+ m_{\text{bin}}\left(\boldsymbol{n}+\begin{pmatrix}0\\2\end{pmatrix}\right) x_{g_{\text{bin}}}\left(\boldsymbol{n}+\begin{pmatrix}0\\1\end{pmatrix}\right)\\
&+ m_{\text{bin}}\left(\boldsymbol{n}+\begin{pmatrix}2\\2\end{pmatrix}\right) x_{g_{\text{bin}}}\left(\boldsymbol{n}+\begin{pmatrix}1\\1\end{pmatrix}\right).
\end{aligned}$$

In Fourier domain,

$$\begin{aligned}
S(2\boldsymbol{\omega}) =& M_{\text{bin}}(\boldsymbol{\omega}) * X_{\alpha_{\text{bin}}}(\boldsymbol{\omega}) + \left(e^{j\left\{\boldsymbol{\omega}^T\binom{2}{2}\right\}} M_{\text{bin}}(\boldsymbol{\omega})\right)\\
&* \left(e^{j\left\{\boldsymbol{\omega}^T\binom{1}{1}\right\}} X_{\beta_{\text{bin}}}(\boldsymbol{\omega})\right) + M_{\text{bin}}(\boldsymbol{\omega}) * X_{g_{\text{bin}}}(\boldsymbol{\omega})\\
&+ \left(e^{j\left\{\boldsymbol{\omega}^T\binom{2}{0}\right\}} M_{\text{bin}}(\boldsymbol{\omega})\right) * \left(e^{j\left\{\boldsymbol{\omega}^T\binom{1}{0}\right\}} X_{g_{\text{bin}}}(\boldsymbol{\omega})\right)\\
&+ \left(e^{j\left\{\boldsymbol{\omega}^T\binom{0}{2}\right\}} M_{\text{bin}}(\boldsymbol{\omega})\right) * \left(e^{j\left\{\boldsymbol{\omega}^T\binom{0}{1}\right\}} X_{g_{\text{bin}}}(\boldsymbol{\omega})\right)\\
&+ \left(e^{j\left\{\boldsymbol{\omega}^T\binom{2}{2}\right\}} M_{\text{bin}}(\boldsymbol{\omega})\right) * \left(e^{j\left\{\boldsymbol{\omega}^T\binom{1}{1}\right\}} X_{g_{\text{bin}}}(\boldsymbol{\omega})\right)\\
=& \sum_{\boldsymbol{\lambda}\in\frac{\pi}{2}\mathbb{Z}_4^2} \frac{X_{\alpha_{\text{bin}}}(\boldsymbol{\omega}-\boldsymbol{\lambda}) + e^{j\left\{(\boldsymbol{\omega}+\boldsymbol{\lambda})^T\binom{1}{1}\right\}} X_{\beta_{\text{bin}}}(\boldsymbol{\omega}-\boldsymbol{\lambda})}{16}\\
&+ \sum_{\boldsymbol{\lambda}\in\frac{\pi}{2}\mathbb{Z}_4^2}\sum_{\boldsymbol{\theta}\in\mathbb{Z}_2^2} \frac{e^{j\{(\boldsymbol{\omega}+\boldsymbol{\lambda})^T\theta} X_{g_{\text{bin}}}(\boldsymbol{\omega}-\boldsymbol{\lambda})}{16}.
\end{aligned}$$

Separating the $X_{g_{\text{bin}}}(\boldsymbol{\omega}-\boldsymbol{\lambda})$ to two parts, $\boldsymbol{\lambda}=\binom{0}{0}$ and $\boldsymbol{\lambda}\neq\binom{0}{0}$ and downsampling ($2\boldsymbol{\omega}\mapsto\boldsymbol{\omega}$), we have

$$\begin{aligned}
S(\boldsymbol{\omega}) =& \sum_{\boldsymbol{\lambda}\in\pi\mathbb{Z}_2^2} \frac{\left(X_{\alpha_{\text{bin}}}\left(\frac{\boldsymbol{\omega}-\boldsymbol{\lambda}}{2}\right) + e^{j\left\{\left(\frac{\boldsymbol{\omega}+\boldsymbol{\lambda}}{2}\right)^T\binom{1}{1}\right\}} X_{\beta_{\text{bin}}}\left(\frac{\boldsymbol{\omega}-\boldsymbol{\lambda}}{2}\right)\right)}{4}\\
&+ \sum_{\boldsymbol{\theta}\in\mathbb{Z}_2^2} \frac{e^{j\left(\frac{\boldsymbol{\omega}}{2}\right)^T\theta} X_{g_{\text{bin}}}\left(\frac{\boldsymbol{\omega}}{2}\right)}{16}\\
&+ \sum_{\boldsymbol{\lambda}\in\frac{\pi}{2}\mathbb{Z}_4^2\setminus\binom{0}{0}}\sum_{\boldsymbol{\theta}\in\mathbb{Z}_2^2} \frac{e^{j\{\left(\frac{\boldsymbol{\omega}}{2}+\boldsymbol{\lambda}\right)^T\theta} X_{g_{\text{bin}}}\left(\frac{\boldsymbol{\omega}}{2}-\boldsymbol{\lambda}\right)}{16},
\end{aligned}$$

where the 1/4 term on $X_{\alpha_{\text{bin}}}$ and $X_{\beta_{\text{bin}}}$ comes from exchanging $\mathbb{Z}_4^2$ with $\mathbb{Z}_2^2$. With arithmetic and approximation of (19), the Fourier transform of $s(\boldsymbol{n})$ simplifies to:

$$\begin{aligned}
S(\boldsymbol{\omega}) \approx& \sum_{\boldsymbol{\lambda}\in\pi\mathbb{Z}_2^2}\Bigg(X_\alpha\left(\frac{\boldsymbol{\omega}-\boldsymbol{\lambda}}{2}\right)\\
&+ \underbrace{e^{j\left(\frac{\boldsymbol{\omega}}{2}\right)^T\binom{1}{1}}}_{\text{unwanted filter}} e^{j\left(\frac{\boldsymbol{\lambda}}{2}\right)^T\binom{1}{1}} X_\beta\left(\frac{\boldsymbol{\omega}-\boldsymbol{\lambda}}{2}\right)\Bigg)\\
&+ \underbrace{\sum_{\boldsymbol{\theta}\in\mathbb{Z}_2^2} \frac{e^{j\left(\frac{\boldsymbol{\omega}}{2}\right)^T\theta} H_{\text{bin}}\left(\frac{\boldsymbol{\omega}}{2}\right)}{16}}_{\text{unwanted filter}} X_g\left(\frac{\boldsymbol{\omega}}{2}\right)\\
&+ \sum_{\boldsymbol{\lambda}\in\frac{\pi}{2}\mathbb{Z}_4^2\setminus\binom{0}{0}} \underbrace{\sum_{\boldsymbol{\theta}\in\mathbb{Z}_2^2} \frac{e^{j\left(\frac{\boldsymbol{\omega}}{2}+\boldsymbol{\lambda}\right)^T\theta} H_{\text{bin}}\left(\frac{\boldsymbol{\omega}}{2}-\boldsymbol{\lambda}\right)}{16}}_{\text{antialias filter}} \underbrace{X_g\left(\frac{\boldsymbol{\omega}}{2}-\boldsymbol{\lambda}\right)}_{\text{aliasing}}
\end{aligned}$$

## 7 Endnote

[a]Filter $h_{\text{bin}}$ is a combination of highpass and lowpass. However, binning takes advantage of the fact that the sensor resolution exceeds optical resolution, meaning $h_{\text{bin}}$ is taken to be a lowpass/antialiasing filter on $x_g$.

**Competing interests**
The authors declare that they have no competing interests.

**Acknowledgement**
This work was funded in part by Texas Instruments.

Received: 11 October 2011 Accepted: 29 May 2012
Published: 21 June 2012

**References**

1. H Yamanakam, Method and apparatus for producing ultra-thin semiconductor chip and method and apparatus for producing ultra-thin back illuminated solid-state image pickup device. US Patent 7,521,335 (2006)
2. T Edwards, R Pennypacker, Manufacture of Thinned Substrate Imagers. US Patent 4,226, 334 (1981)

3. J Compton, J Hamilton, Image sensor with improved light sensitivity. US Patent 2007/0024931 (2007)
4. U Barnhofer, J DiCarlo, B Olding, B Wandell, in *Proceedings of the SPIE* . Color estimation error trade-offs, (2003), pp. vol. 5027, 263–273
5. W Borchenko, Phase One Patent Pending Sensor+Explained. http://www.phaseone.com/Digital-Backs/P65//media/Phase%20One/Reviews/Review%20pdfs/Backs/Phase-One-Sensorplus.ashx
6. Z Zhou, B Pain, E Fossum, Frame-transfer CMOS active pixel sensor with pixel binning. IEEE Trans. Electron. Dev. **44**(10), 1764–1768 (1997)
7. F Chu, Improving CMOS image sensor performance with combined pixels (2005). http://www.eetimes.com/design/embedded/4013011/Improving-CMOS-image-sensor-performance-with-combined-pixels
8. K Dabov, A Foi, V Katkovnik, K Egiazarian, Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans. Image Process. **16**(8), 2080–2095 (2007)
9. J Portilla, V Strela, M Wainwright, E Simoncelli, Image denoising using scale mixtures of Gaussians in the wavelet domain. IEEE Trans. Image Process. **12**(11), 1338–1351 (2003)
10. K Hirakawa, F Baqai, P Wolfe, in *Proc. SPIE , Electronic Imaging*, vol. 7246. Wavelet-based Poisson rate estimation using the Skellam distribution , (2009)
11. L Zhang, R Lukac, X Wu, D Zhang, PCA-based spatially adaptive denoising of CFA images for single-sensor digital cameras. IEEE Trans. Image Process. **18**(4), 797–812 (2009)
12. K Hirakawa, T Parks, Joint demosaicing and denoising. IEEE Trans. Image Process. **15**(8), 2146–2157 (2006)
13. L Zhang, X Wu, D Zhang, Color reproduction from noisy CFA data of single sensor digital cameras. IEEE Trans. Image Process. **16**(9), 2184–2197 (2007)
14. K Hirakawa, X Meng, P Wolfe, in *IEEE International Conference on Acoustics, Speech and Signal Processing 2007. ICASSP 2007*. A framework for wavelet-based analysis and processing of color filter array images with applications to denoising and demosaicing, (2007), pp. vol. 1, pp. I–597
15. R Fergus, B Singh, A Hertzmann, S Roweis, W Freeman, Removing camera shake from a single photograph. ACM Trans. Graph. (TOG). **25**(3), 787–794 (2006)
16. A Levin, P Sand, T Cho, F Durand, W Freeman, in *ACM SIGGRAPH 2008 papers, ACM*. Motion-invariant photography, (2008), pp. 1–9
17. K Hirakawa, P Simon, in *IEEE International Conference on Computer Vision*. Single-shot high dynamic range imaging with conventional camera hardware, (2011), p. vol. 1
18. D Alleysson, S Susstrunk, J Hérault, Linear demosaicing inspired by the human visual system. IEEE Trans. Image Process. **14**(4), 439–449 (2005)
19. E Dubois, Frequency-domain methods for demosaicking of Bayer-sampled color images. IEEE Signal Process. Lett. **12**(12), 847–850 (2005)
20. P K Hirakawa, Wolfe, Spatio-spectral color filter array design for optimal image recovery. IEEE Trans. Image Process. **17**(10), 1876–1890 (2008)
21. J Gu, P Wolfe, K Hirakawa, in *2010 17th IEEE International Conference on Image Processing (ICIP)*. Filterbank-based universal demosaicking, (2010), pp. 1981–1984
22. K Hirakawa, T Parks, Adaptive homogeneity-directed demosaicing algorithm. IEEE Trans. Image Process. **14**(3), 360–369 (2005)
23. L Zhang, X Wu, Color demosaicking via directional linear minimum mean square-error estimation. IEEE Trans. Image Process. **4**(12), 2167–2178
24. T Fellers, K Vogt, M Davidson, CCD signal-to-noise ratio. http://www.microscopyu.com/tutorials/java/digitalimaging/signaltonoise/
25. P Gehler, C Rother, A Blake, T Minka, T Sharp, in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Bayesian color constancy revisited, (2008)
26. L Shi, B Funt, Re-processed Version of the Gehler Color Constancy Dataset of 568 Images. http://www.cs.sfu.ca/colour/data/

doi:10.1186/1687-6180-2012-125
**Cite this article as:** Jin and Hirakawa: **Analysis and processing of pixel binning for color image sensor.** *EURASIP Journal on Advances in Signal Processing* 2012 **2012**:125.