Pixel Binning for High Dynamic Range Color Image Sensor Using Square Sampling Lattice

Jiachao Zhang, Jie Jia, *Student Member, IEEE*, Andong Sheng, and Keigo Hirakawa, *Senior Member, IEEE*

***Abstract*— We propose a new pixel binning scheme for color image sensors. We minimized distortion caused by binning by requiring that the superpixels lie on a square sampling lattice. The proposed binning schemes achieve the equivalent of 4.42 times signal strength improvement with the image resolution loss of 5 times, higher in noise performance and in resolution than the existing binning schemes. As a result, the proposed binning has considerably less artifacts and better noise performance compared with the existing binning schemes. In addition, we provide an extension to the proposed binning scheme for performing single-shot high dynamic range image acquisition.**

***Index Terms*— Color filter array, pixel binning, single-shot HDR.**

I. Introduction

DIGITAL camera is ubiquitous, as they are widely used in cell phones, vehicles, security and computer vision systems. In many applications, sensor resolution has exceeded optical resolution, meaning that increasing the pixel density will not necessarily result in image quality gains. However, low light imaging remains to be an unsolved problem. Noise remains serious impediment to image quality for modern image sensors, particularly for these that are high resolution. Techniques involving color filter with higher transmittance [1], [2], backside illuminated pixel architecture [3], [4], and pixel binning [5]–[7] have been used to reduce noise in hardware.

Binning is a process of combining the electrical charges of multiple neighboring pixels into a single "superpixel" [5]–[13]. The combined signal will be transformed into a digital signal by an analog-to-digital converter. This technique reduces the impact of read noise on the combined signal even if the individual pixel values are small. However, the noise performance improvements come at the price of spatial resolution loss. Binning was first introduced as a way to exploit CCD pixel sensor architecture. Today, a CMOS binning combines electrical charges by transistor switches that connect the photodiodes together. Binning has the advantage that it can be enabled only as needed. That is, the sensor may still operate at high resolution in ample light, and binning is used only when requiring better noise performance.

Besides the improved readout circuits designs [9]–[13], the spatial sampling of combined pixels has significant impact on the binning performance. As illustrated in Figure 1, existing binning scheme such as Kodak PIXELUX [6] and PhaseOne [7] combine neighboring four pixels to yield an signal gain of 4 times. But our prior work in [8] proved that the spatial and color artifacts caused by their binning patterns are due to aliasing stemming from the implied non-uniform sampling.

In this paper, we propose a new pixel binning design for color image sensors aimed at minimizing the resolution loss while improving noise performance. Our scheme combines neighboring pixels to yield superpixels arranged in a square sampling lattice. We achieve equivalence to 4.42 times signal strength improvement with image resolution loss of 5 times by ensuring that the binning pattern results in a square (i.e. uniform) sampling lattice. Contrast this to the existing binning scheme such as Kodak PIXELUX [6] and PhaseOne [7] that has a signal gain of 4 times but an effective resolution[1] loss of 16 (See Section II-A). As a result, we have considerably less spatial and color artifacts and better noise performance compared with the existing binning schemes. In addition, we provide an extension to the proposed binning scheme for performing a single-shot high dynamic range image acquisition.

II. Background and Review

A. Binning for Color Image Sensor

Binning is considerably more complicated for color image sensors. Color images are typically captured using color filter array (CFA), or a spatial multiplexing of red, green, and blue filters. Only a single color intensity is measured at each individual pixel sensor, and a post-capture interpolation step called "demosaicking" recovers full a color image based on the surrounding measured pixel component values. The most commonly used CFA pattern is called the Bayer pattern [14],

Manuscript received March 8, 2017; revised November 11, 2017; accepted January 3, 2018. Date of publication January 23, 2018; date of current version February 12, 2018. The work was supported in part by the National Natural Science Foundation of China under Grant 61702265, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20170856, and in part by the CCF-Tencent Open Research Fund. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Denis Kouame. *(Corresponding author: Jiachao Zhang.)*

J. Zhang and A. Sheng are with the Department of Automation, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: 311102240@mail.njust.edu.cn; shengandong@mail.njust.edu.cn).

J. Jia was with the University of Dayton, Dayton, OH 45469 USA. He is now with Mura Inc., San Jose, CA 95117 USA (jie.jia@muravision.com).

K. Hirakawa is with the Department of Electrical and Computer Engineering, University of Dayton, Dayton, OH 45469 USA (e-mail: khirakawa1@udayton.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2018.2795741

[1]The notion of "resolution" in this paper refers to the ability to discriminate fine image details. This differs from the "pixel count" or "pixel density" that may be misleadingly high, even if the "resolution" may be low.

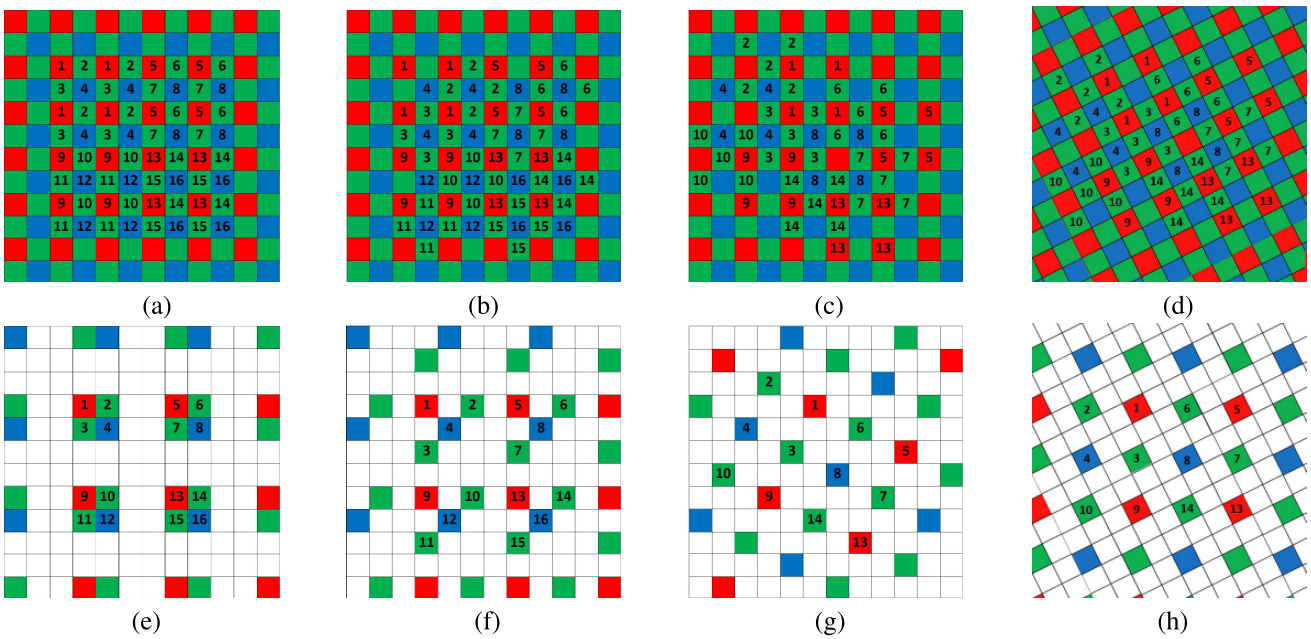


Fig. 1. Binning schemes for a Bayer pattered CFA sensor. (a)-(d) The numbers over high resolution Bayer pattern indicate the pixels combined together. (e)-(h) The numbers over resultant superpixel Bayer pattern denote the relative locations of combined pixels.

shown Figure 1(a)-(c), where the sampling density of green is twice that of red and blue. The CFA subsampled data at pixel location $\boldsymbol{n} = (n_x, n_y)^T \in \mathbb{Z}^2$ has the form:

$$X(\boldsymbol{n}) = \begin{cases} R(\boldsymbol{n}) & \text{if } n_x \text{ and } n_y \text{ are odd} \\ B(\boldsymbol{n}) & \text{if } n_x \text{ and } n_y \text{ are even} \\ G(\boldsymbol{n}) & \text{else,} \end{cases} \tag{1}$$

where $R(\boldsymbol{n})$, $G(\boldsymbol{n})$ and $B(\boldsymbol{n})$ represent the red, green and blue values in the sensor color space, respectively. With the read noise $N(\boldsymbol{n})$, the recorded CFA sensor data values are:

$$Y(\boldsymbol{n}) = X(\boldsymbol{n}) + N(\boldsymbol{n}). \tag{2}$$

The presence of CFA makes it difficult to implement binning. A common approach has been to combine the charges corresponding to the pixels of same CFA color—combine multiple red pixels to form a red super pixel, for example. Binning combines several electric charges of neighboring pixels by summing the photo current or charge accumulation, and hence each of the pixel cannot be shared by more than one superpixel. As such, the options for binning schemes are fairly limited. The two well known binning schemes are Kodak PIXELUX [6] and PhaseOne [7], shown in Figures 1(a) and 1(b). The resultant superpixel form another Bayer pattern, as shown in 1(e) and 1(f) respectively. After binning, a demosaicking algorithm is applied to the superpixel Bayer pattern to yield the final lower resolution image with less noise. See Figure 2(b). Compared to the ordinary pipeline, the binning is far less sensitive to noise. This is true even if demosaicked images are downsampled, as shown in Figure 2(a).

It is a common knowledge that binning schemes introduce artifacts. For example, it was proven in [8] that PIXELUX binning scheme that combines four pixels to form a superpixel

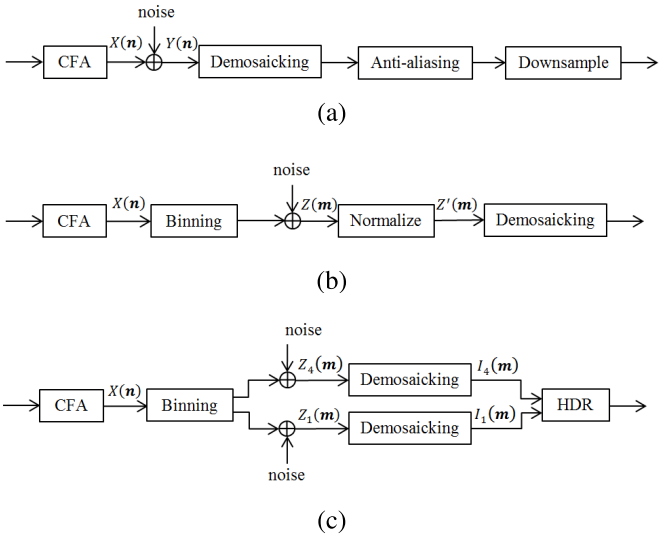


Fig. 2. Schematic diagram of processing CFA data. (a) Demosaicking and downsampling. (b) Binning with demosaicking. (c) Binning based high dynamic range imaging.

with four times the signal actually suffers 16 times resolution loss (ie. not four times loss as commonly believed) because of sever aliasing. On the other hand, PhaseOne binning scheme suffers from color artifacts. As detailed in Section IV, the red, green, and blue pixels are shifted in a way that edges in the three color components are not colocated. The artifacts (as shown in Figure 6 (b) and (c)) in PIXELUX and PhaseOne binning schemes stem from the non-uniform sampling—the implied locations of the resultant superpixels (centroids of all combined pixels) are not on evenly spaced grids.

For correcting the pixelization artifacts of Kodak PIXELUX binning scheme, a binning-aware demosaicking algorithm was

proposed in [8]. It preserves image details better by explicitly canceling the aliasing causing the "16 times" resolution loss, in order to recover an image with only "4 times" resolution loss. While this approach drastically improves the spatial details of the demosaicked image, the noise performance is not as robust. The reason is that although the main signal has four times signal strength of the original, the signal components that are being used to cancel the aliasing only has twice the signal strength of the original.

B. Review: Sampling Lattice Theory

We briefly review the lattice theory below. Consider a two dimensional integer grid $\mathbb{Z}^2$. A new lattice $\Lambda \subset \mathbb{R}^2$ can be formed using a generative matrix $M \in \mathbb{R}^2$ as

$$\Lambda = M\mathbb{Z}^2 = \{v \in \mathbb{R}^2 | v = Mw \text{ such that } w \in \mathbb{Z}^2\}. \quad (3)$$

A unitary matrix of the form

$$M = \begin{bmatrix} cos(\theta) & sin(\theta) \\ -sin(\theta) & cos(\theta) \end{bmatrix} \quad (4)$$

rotates $\mathbb{Z}^2$ by angle $\theta \in [0, 2\pi)$ so that the resultant lattice also forms a square grid. To form a square lattice that lies on an integer grid, we need

$$M = k \begin{bmatrix} cos(\theta) & sin(\theta) \\ -sin(\theta) & cos(\theta) \end{bmatrix} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \quad (5)$$

where k is a constant, ensuring that $a = k\ cos(\theta) \in \mathbb{Z}$ and $b = k\ sin(\theta) \in \mathbb{Z}$. Its determinant $\det(M) = a^2 + b^2 = k^2(cos(\theta))^2 + sin(\theta)^2) = k^2$ determines the "magnification" factor of Λ relative to $\mathbb{Z}^2$. For example, a quincunx lattice is formed by $a = b = 1$, or equivalently $k = \sqrt{2}$ and $\theta = \frac{\pi}{4}$. Its magnification factor is $k^2 = a^2 + b^2 = 2$. A square lattice of $a = 2$ and $b = 0$ on the other hand has a magnification of $k^2 = a^2 + b^2 = 4$. Note that a Bayer pattern in Figure 1(a) is a combination of quincunx and square lattices of Figures 3(b) and 3(c), respectively.

III. Proposed Binning Design

A. Grid Design

Recall that the artifacts in PIXELUX and PhaseOne binning schemes are caused by the non-uniform sampling. We propose a new binning scheme for Bayer Pattern CFA that has uniform sampling—i.e. superpixel have equal spacing. To ensure uniform sampling, the binning scheme has an implied superpixel sampling on the grid $\Lambda = M\mathbb{Z}^2$, where M takes the form as in (5). As already shown in Figure 3, this guarantees uniform sampling.

The elemental combinations of a and b are shown in Table I. We seek the values of a and b such that $|a| + |b|$ is an odd number (rows with bold fonts in Table I). This ensures that the post-binning pattern is a super Bayer pattern, because M is a "relative prime" of the quincunx and square lattice generation matrices of Figures 3(b) and 3(c) [15].

Recall that $\det(M) = a^2 + b^2$ determines the "magnification" factor of Λ relative to $\mathbb{Z}^2$. We seek M such that $\det(M)$ is not too high, as we do not want to sacrifice resolution too much. Therefore, we choose $a = 2$, $b = 1$, and

$$M = \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}$$

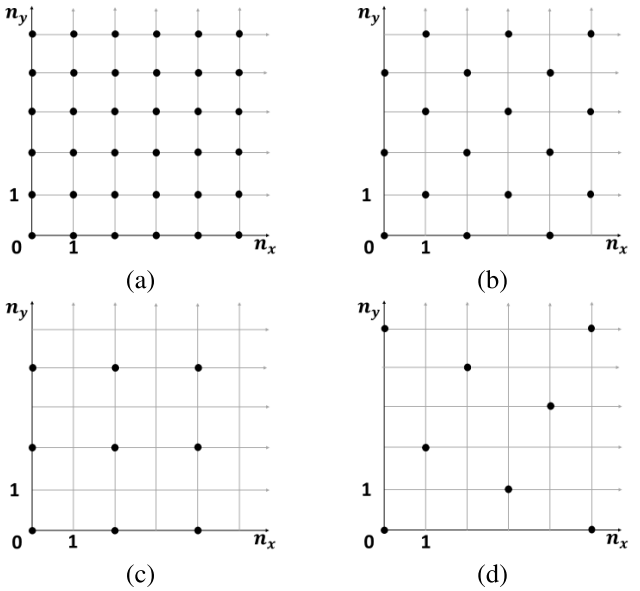


Fig. 3. Example of square lattices generated by matrices in (5). (a) $\mathbb{Z}^2$. (b) Quincunx ($a = b = 1$). (c) Square ($a = 2,\ b = 0$). (d) Square ($a = 2,\ b = 1$).

TABLE I

The Combinations of a and b for Lattice Generative Matrix M in (5)

a	b	$\lvert a\rvert + \lvert b\rvert$	$\det(M)$
1	1	2	2
2	**1**	**3**	**5**
3	1	4	10
4	**1**	**5**	**17**
2	2	4	8
3	2	5	13
3	**0**	**3**	**9**

which has the smallest resolution loss—the loss of $\det(M) = 5$ times. See Figure 3(d).

B. Binning Design

Based on the lattice sampling and grid design criteria above, we propose a binning scheme shown in Figure 1(c). The proposed binning scheme combines five neighboring green pixel values together to form one green superpixel; and four red/blue pixels are combined to form a red/blue one. Specially, the superpixel at location $\boldsymbol{m} = (m_x, m_y) \in \mathbb{Z}^2$ is formed as

$$Z(\boldsymbol{m}) = \begin{cases} R\left(M\boldsymbol{m} + \binom{1}{1}\right) + R\left(M\boldsymbol{m} + \binom{1}{-1}\right) \\ +R\left(M\boldsymbol{m} + \binom{-1}{1}\right) + R\left(M\boldsymbol{m} + \binom{-1}{-1}\right) + N(\boldsymbol{m}) \\ \quad \text{(if } m_x \text{ and } m_y \text{ are odd)} \\ B\left(M\boldsymbol{m} + \binom{1}{1}\right) + B\left(M\boldsymbol{m} + \binom{1}{-1}\right) \\ +B\left(M\boldsymbol{m} + \binom{-1}{1}\right) + B\left(M\boldsymbol{m} + \binom{-1}{-1}\right) + N(\boldsymbol{m}) \\ \quad \text{(if } m_x \text{ and } m_y \text{ are even)} \\ G(M\boldsymbol{m}) + G\left(M\boldsymbol{m} + \binom{1}{1}\right) + G\left(M\boldsymbol{m} + \binom{1}{-1}\right) \\ +G\left(M\boldsymbol{m} + \binom{-1}{1}\right) + G\left(M\boldsymbol{m} + \binom{-1}{-1}\right) + N(\boldsymbol{m}) \\ \quad \text{(else)}, \end{cases} \quad (6)$$

where $N \sim \mathcal{N}(0, \sigma^2)$ is the read noise stemming from transforming the combined signals into a digital signal by an analog-to-digital converter. Unlike the existing binning

schemes of Kodak and PhaseOne, the resultant superpixels of proposed binning are arranged on uniform grid as shown in Figure 1(g). That is, the symmetry summation in (6) guarantees that the "implied" locations of the superpixels lie on $\Lambda = M\mathbb{Z}^2$.

Due to the fact that green superpixels have greater signal strength than red/blue superpixels, the acquired sensor data $Z(\boldsymbol{m})$ is first processed by a normalization, as follows

$$Z'(\boldsymbol{m}) = \begin{cases} \frac{1}{5} Z(\boldsymbol{m}) & \text{if } m_x + m_y \text{ is odd} \\ \frac{1}{4} Z(\boldsymbol{m}) & \text{if } m_x + m_y \text{ is even.} \end{cases} \tag{7}$$

The demosaicking is then applied to $Z'(\boldsymbol{m})$ (and not to $Z(\boldsymbol{m})$). Consider the noise performance of $Z'(\boldsymbol{m})$. The red/blue and green superpixels are subject to a read noise of $1/4$ and $1/5$ times the strength compared with no binning, respectively. The noise variance of individual signal strength for red/blue and green color components are $\sigma^2/16$ and $\sigma^2/25$, respectively, meaning that the average equivalent noise variance of a normalized superpixel $Z'(\boldsymbol{m})$ is

$$\frac{\frac{\sigma^2}{16} \times 2 + \frac{\sigma^2}{25} \times 2}{4} = \left(\frac{\sigma}{4.42}\right)^2. \tag{8}$$

We conclude that the proposed binning superpixel achieves equivalent of 4.42 times signal strength improvement with image resolution loss of 5 times.

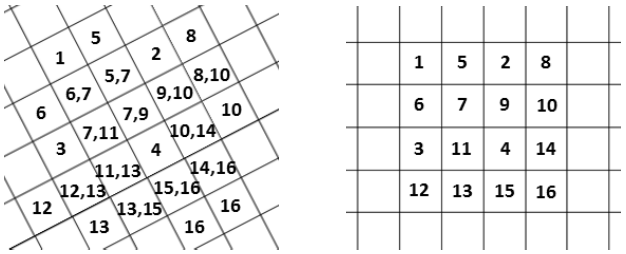
Notice that the proposed square sampling lattice has an angle of 26.6 degrees between binning signal and the original grid, introduced by the unitary transformation M. The output image of a camera processing pipeline applied to the proposed binning data is therefore slanted also. When rotating the final output back to original image grid, we incur additional penalties from interpolation.

C. Slanted Grid CFA Sensor for Video Cameras

In the recent years, consumer-level video cameras have advanced, operating at high level frame rates. Noise and pixel count are the limiting factors in these cameras—reduced exposure time for each frame introduces noise, while the increased data throughput pose challenges to the processing hardware. As such, it is very common for video cameras to be equipped with high resolution sensors, but operate at a low resolution as a default mode. Only when using digital zoom or when taking still photos with a video camera would the high resolution mode be used.

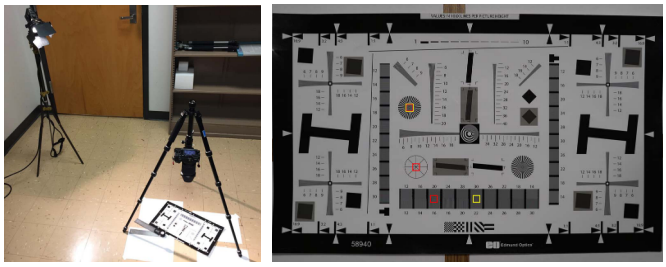
With this in mind, consider a new design in Figure 1(d). This slanted configuration has the advantage that when binning is enabled, the resultant superpixels lie on a non-rotated square grid. Hence there is no interpolation penalty stemming from rotation. See Figure 1(h). When operating in high resolution mode, however, the pixels are located in a rotated grid. In devices such as video cameras, where the low resolution is the default mode of operation, the unusual sensor configuration in Figure 1(d) is appropriate.

During the occasional usage of the high resolution image, the penalty stemming from rotation is low (when compared to rotating a low resolution image) because of the dense sampling rate. We propose an efficient technique to rotate the high



(a) Slanted sensor (b) Rotated sensor

Fig. 4. Example of post-demosaicking rotation in (9). The resultant rotated pixels in (b) are the averaging of the pixels labeled by the same number in (a). Since this is not binning, each pixel may be used to compute mutiple interpolated pixels. (a) Slanted sensor. (b) Rotated sensor.



(a) (b)

Fig. 5. Set up for resolution analysis. (a) Capture setting. (b) Resolution chart.

resolution image with minimal resolution penalty. Refer to the demosaicked image of the full resolution CFA data $Y(\boldsymbol{n})$ as $I(\boldsymbol{n})$ in Figure 2(a). The rotated image $J(\boldsymbol{m})$ of $I(\boldsymbol{n})$ is computed as follows:

$$\begin{aligned} J(2\boldsymbol{m}) &= I(M\boldsymbol{m}) \\ J\left(2\boldsymbol{m} + \begin{pmatrix}1\\0\end{pmatrix}\right) &= \frac{1}{2}\left(I\left(M\boldsymbol{m} + \begin{pmatrix}1\\0\end{pmatrix}\right) + I\left(M\boldsymbol{m} + \begin{pmatrix}1\\-1\end{pmatrix}\right)\right) \\ J\left(2\boldsymbol{m} + \begin{pmatrix}0\\1\end{pmatrix}\right) &= \frac{1}{2}\left(I\left(M\boldsymbol{m} + \begin{pmatrix}0\\1\end{pmatrix}\right) + I\left(M\boldsymbol{m} + \begin{pmatrix}1\\1\end{pmatrix}\right)\right) \\ J\left(2\boldsymbol{m} + \begin{pmatrix}1\\1\end{pmatrix}\right) &= \frac{1}{4}\left(I\left(M\boldsymbol{m} + \begin{pmatrix}1\\0\end{pmatrix}\right) + I\left(M\boldsymbol{m} + \begin{pmatrix}1\\1\end{pmatrix}\right)\right. \\ &\quad \left. + I\left(M\boldsymbol{m} + \begin{pmatrix}2\\0\end{pmatrix}\right) + I\left(M\boldsymbol{m} + \begin{pmatrix}2\\1\end{pmatrix}\right)\right). \end{aligned} \tag{9}$$

The pixels in resultant rotated sensor locate as shown in Figure 4(b). The rotated image $J(\boldsymbol{m})$ lies on a non-integer grid of

$$\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \times \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \mathbb{Z}^2.$$

Its sampling density is

$$\det\left(\begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \times \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}^{-1}\right) = 1.25,$$

meaning the sampling grid is slightly less dense than the original sampling grid. In terms of resolution, this simplistic approach is effective, as our experimental results show. The linear combination in Figure 4 results in some attenuation of high spatial frequency (but not completely eliminated; hence

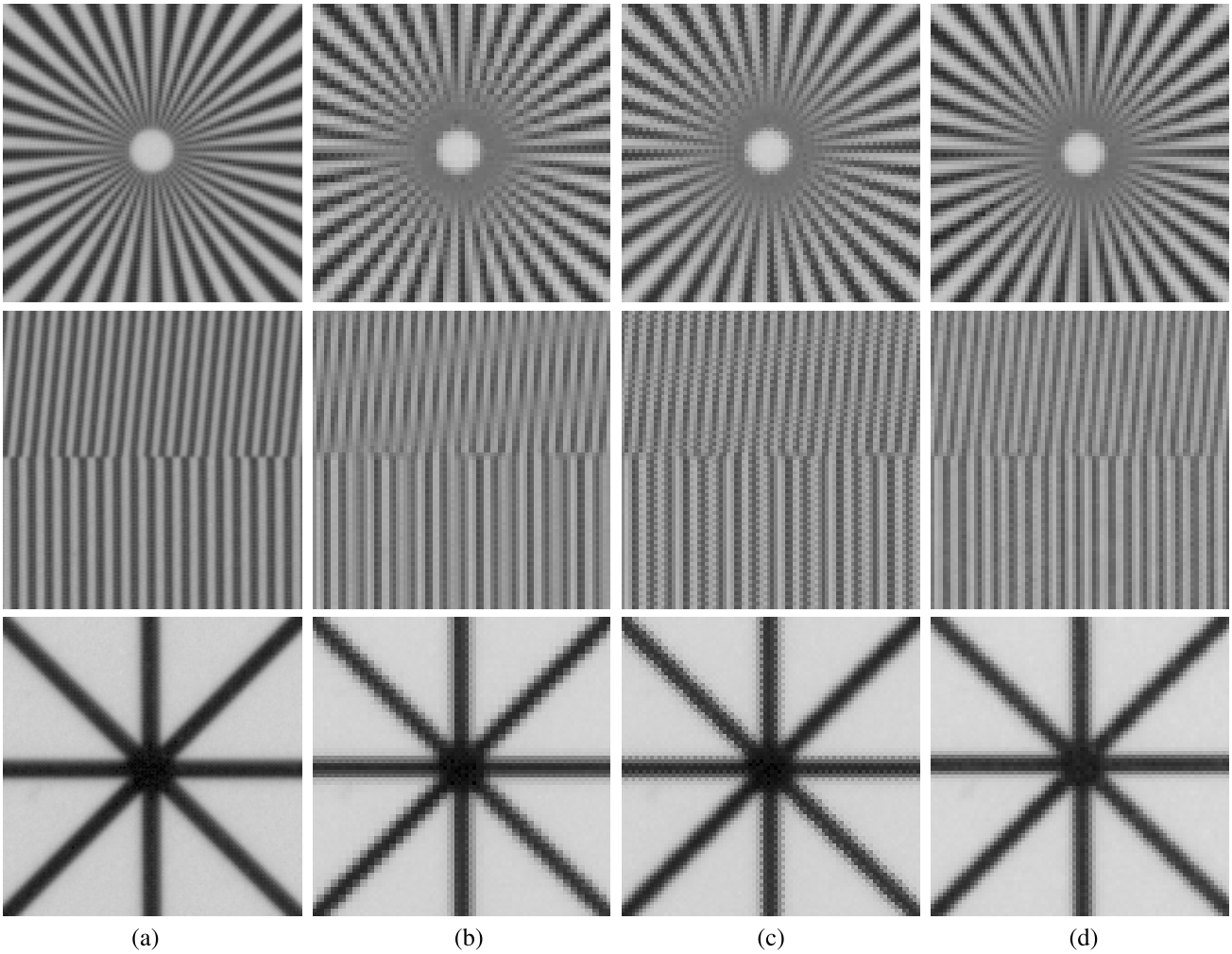


Fig. 6. Binning output, displayed without demosaicking, cropped from Figure 5(b) (red boxes). (a) A raw sensor data of a resolution chart. (b)-(d) Simulated binning data of the raw sensor data in (a). Since the resolution chart is monochromatic, raw sensor data can be displayed as a "grayscale image" with a properly executed white balance. This allows us to assess the resolution loss stemming from binning, independently of the demosaicking performance. The cropped image size of (a)-(d) are 164×164, 82×82, 82×82 and 74×74 respectively.

not a resolution "loss", technically). For example, one can show that at $(\frac{4}{5}\pi, 0)$, which is the highest horizontal spatial resolution representable by the rotated pixel grid, the effective frequency response is 0.6764. Hence we expect resolution loss by factor of 1.25, albeit with slight softening of high spatial frequencies.

IV. Experimental Results

A. Resolution Analysis

We first evaluate the resolution loss of the binning schemes independently of demosaicking and noise performance. We captured 14-bits raw sensor data of a resolution chart shown as in Figure 5(b) by Sony α6000 camera. Recall that gray scale images have equal red, green and blue intensities. Hence with appropriately executed white balance, a monochromatic resolution chart such as the one shown in Figure 5(b) can be displayed as a grey scale image without demosaicking. See Figure 6(a). To simulate the slanted sensor in Figure 1(d), we captured the resolution chart by tilting the target 26.6 degrees relative to the camera sensor. The data acquisition setup is shown in Figure 5(a).

In Figure 6, we show cropped portions of the resolution chart that are labeled in Figure 5(b). To minimize the influence of optical chromatic aberration far from the center of the image sensor [16], we captured each cropped region by centering the camera directly above them. Figure 6(d) shows the result of applying normalization in (7) to the proposed binning scheme in (6), simulated from the actual raw sensor in Sony α6000. We compared this to the full-resolution raw senor capture of the same scene (without rotating target by 26.6 degrees) as shown in Figure 6(a) as well as the simulated Kodak PIXELUX and PhaseOne in Figure 6(b) and 6(c).

As evidenced by the top two rows of Figure 6, the resolution of the proposed binning scheme in Figure 6(d) is slightly lower than the full resolution sensor data in Figure 6(a). The resolution loss is considerably less than Kodak PIXELUX and PhaseOne, however. As noted by our earlier paper in [8], the aliasing in Kodak PIXELUX binning effectively limits the resolution of the superpixels in Figure 6(b) to $1/16$ of the

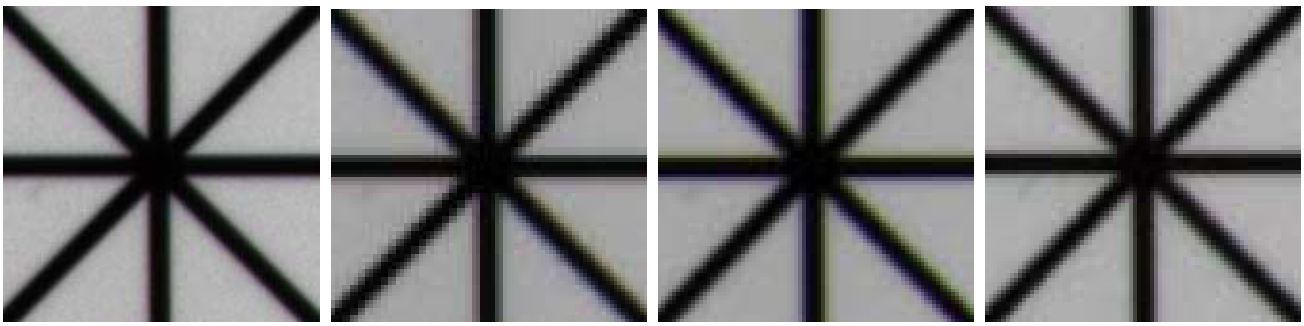


Fig. 7. Demosaicking result of the raw sensor data in Figure 6 (bottom row). The color artifacts in PhaseOne can be attributed to the zippering artifacts in Figure 6(c). (a) Original CFA+ [16]. (b) Kodak PIXELUX + [16]. (c) PhaseOne + [16]. (d) Proposed + [16].

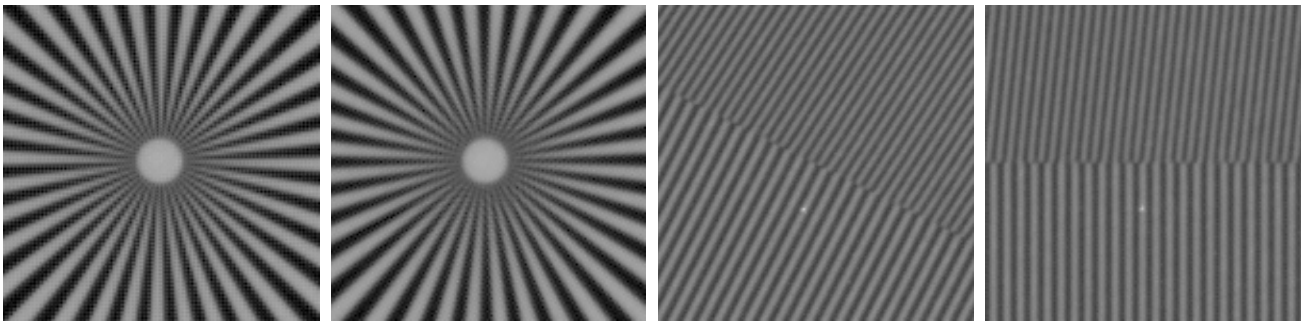


Fig. 8. Rotated high resolution images, cropped from Figure 5(b) (yellow boxes) and displayed without demosaicking. (a,c) A raw sensor data of a resolution chart. (b,d) Resultant rotated image of the raw sensor data in (a,c), respectively. Since the resolution chart is monochromatic, raw sensor data can be displayed as a "grayscale image" with a properly executed white balance. The pixel counts are (a,c) 164×164 and (b,d) 147×147.

full resolution in Figure 6(a). The pixelization artifacts in top row image, as well as the classical aliasing artifacts in the second row clearly support this. On the other hand, the distortions in Figure 6(c) resemble "zippering" artifacts commonly observed in demosaicking. Emphasizing that Figure 6 is shown without demosaicking, however, the "zipping" in superpixel Bayer pattern actually indicates color-dependent shift of edges. Indeed, as shown by the demosaicking outputs in Figure 7, zippering in Bayer pattern coincides with the color artifacts in demosaicking (i.e. color artifacts caused by binning, not demosaicking). Note that we repeated the experiment with various demosaicking algorithms (e.g. [17], [18]), and they all exhibited the same problem. (See also Table II-IV.) We conclude that the proposed binning scheme is superior to the Kodak PIXELUX and PhaseOne binning scheme in terms of resolution and color fidelity, despite lower pixel count overall.

We also demonstrate the effectiveness of the image rotation algorithm in (9). Shown in Figure 8(a) is a resolution chart that was rotated 26.6 degrees. Figure 8(b) shows the result of rotating this image algorithmically. There is little or no noticeable loss of resolution stemming from this rotation.

B. Demosaicking and Noise Performance

Binning techniques are aimed at reducing the impact of read noise on the combined signal. In this section, the binning performance after demosaicking is assessed, while varying the levels of noise. For this study, we captured natural scenes using Sony $\alpha 6000$ in raw sensor mode twice—first with camera held horizontally, and second by rotating the camera by 26.6 degrees to simulate the slanted grid sensor in Figure 1(d). Kodak PIXELUX, PhaseOne, and proposed binning schemes were simulated, demosaicked, color corrected, white balanced, and gamma corrected before being displayed in Figure 9-11. In absense of noise, proposed binning with PSDD demosaicking [16] yielded images with finer details than Kodak PIXELUX with the same demosaicking; and considerably less color artifacts than PhaseOne binning. Kodak PIXELUX with binning-aware demosaicking [8] yielded a result comparable to the proposed scheme.

When noise is added, the proposed binning scheme (with 4.42 times signal strength) had slightly less noisy appearance than Kodak PIXELUX and PhaseOne (with 4 times the signal strength) with the same demosaicking. Even at the highest noise level, however, the pixelization and color artifacts of Kodak PIXELUX and PhaseOne are clearly visible, respectively. Compared to the binning-aware demosaicking method in [8], our noise performance was better—not entirely unexpected because of the aliasing cancellation component that is more sensitive to noise.

For comparison, we also applied demosaicking to the full resolution CFA image, and then downsampled after averaging pixels in 2×2 neighborhood for anti-aliasing and noise suppression. With no noise, this yielded a sharper image than our binning results. But despite post-demosaicking averaging, this approach was unacceptably sensitive to noise.

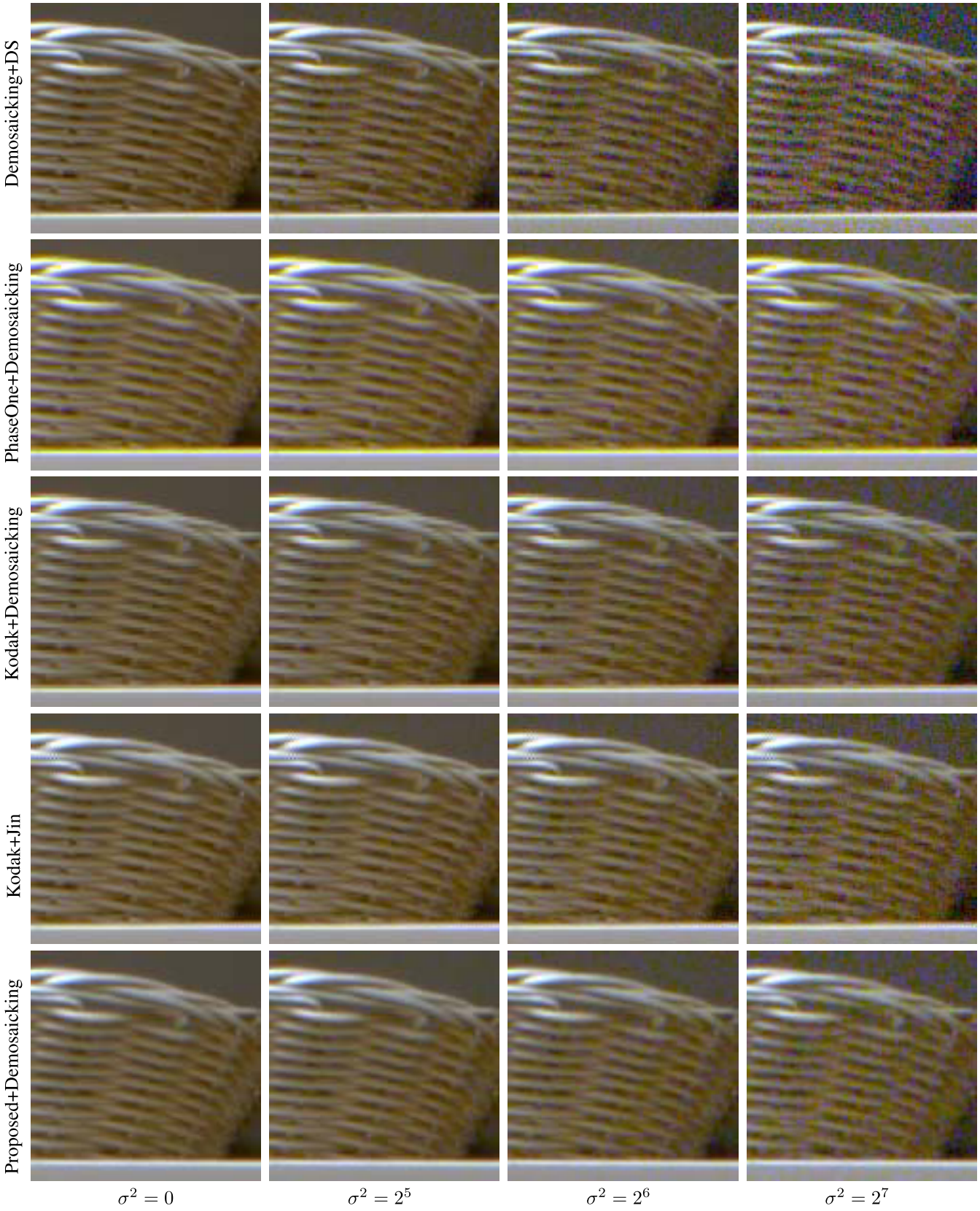


Fig. 9. Reconstructed images with various noise levels.

C. Mean Square Error Test

We also evaluate the performance in terms of MSE. Unlike data acquisition in Section IV-A and IV-B, we use exactly the same inputs (i.e. not rotated) for all binning algorithms to ensure impartial comparison. The 14 bit raw sensor data we used are shown in Figure 12, taken at random from

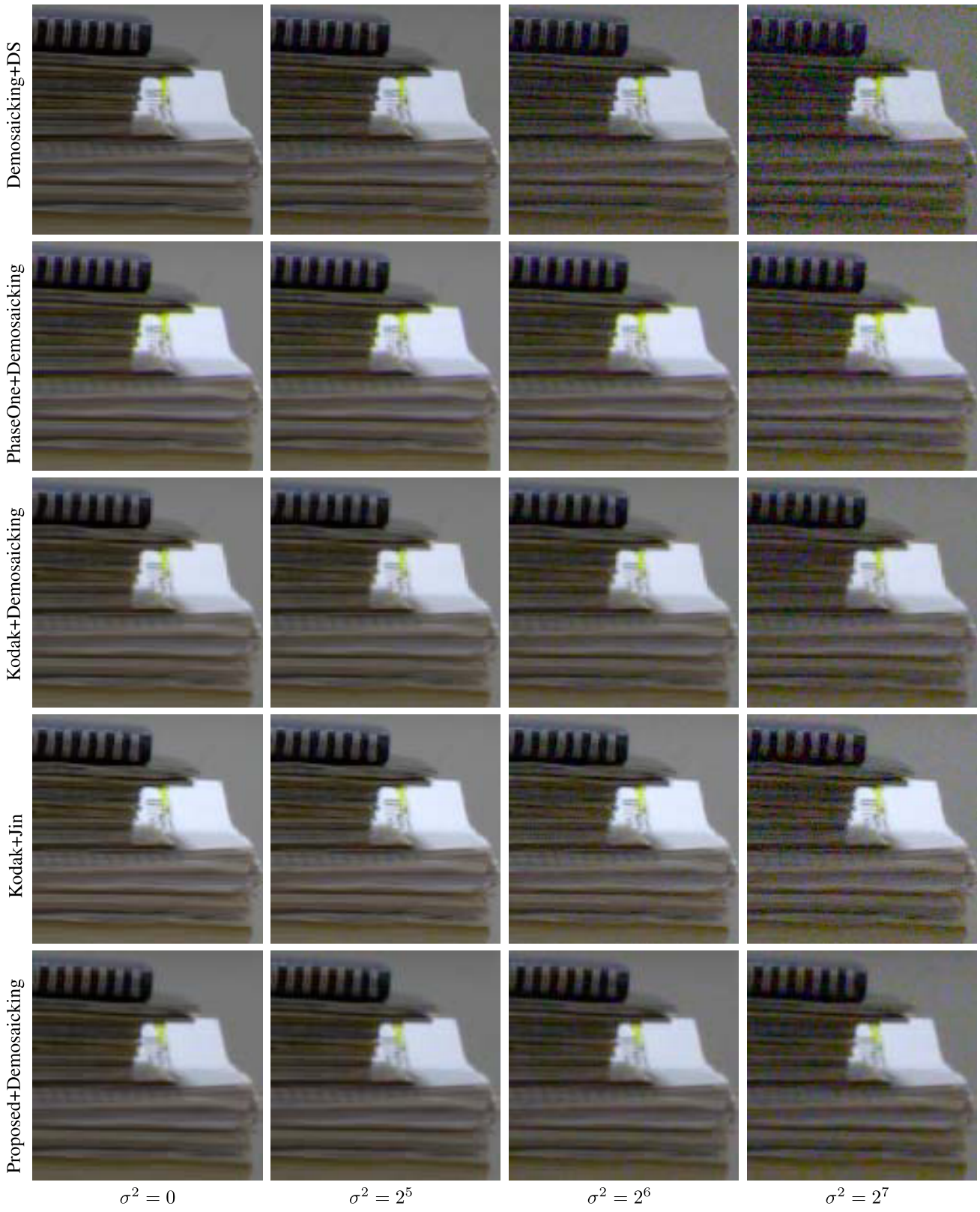


Fig. 10. Reconstructed images with various noise levels.

www.imaging-resource.com. We used Nikon D810 and Sony α7RII—these cameras are challenging to work with because they lack optical low pass filters used in most cameras. For computing the MSE, the reference image for the proposed binning scheme was found by first taking the demosaicking of the full-resolution CFA data with no noise; then applying an anti-aliasing filter of the form

$$\begin{bmatrix} 0 & 1/5 & 0 \\ 1/5 & 1/5 & 1/5 \\ 0 & 1/5 & 0 \end{bmatrix}$$

Fig. 11. Reconstructed images with various noise levels.

to the demosaicked image; then downsample by $M\mathbb{Z}^2$. Similarly, for computing the MSE of the Kodak PIXELUX and PhaseOne binning scheme, the reference image used anti-aliasing filter of the form

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

on the demosaicked full resolution image with no noise, and downsampled by 2×2.

The MSE and S-CIELAB [19], [20] scores are reported in Table II-IV. The proposed binning scheme has the lowest MSE and S-CIELAB for noisy and no noise images, regardless of the choices of demosaicking technique [16]–[18]. With no

TABLE II

THE AVERAGE MSE AND S-CIELAB DIFFERENCE PERFORMANCE OF 10 IMAGES COMPARING 6 SCHEMES. WE USED [16] ON NOISE-FREE SENSOR DATA TO GENERATE THE HIGH RESOLUTION REFERENCE IMAGE. IT WAS ANTIALIASED AND DOWNSAMPLED TO GENERATE REFERENCE IMAGES FOR THE (NOISY AND NOISE-FREE) BINNING RESULTS. "DS" DENOTES DOWNSAMPLING

Cam	Nikon D810						Sony A7RII					
Bin. Demos.	DS2 $\times$ 2 [16]	DS by M [16]	PhaseOne [16]	Kodak [16]	Kodak [8]	Proposed [16]	DS2 $\times$ 2 [16]	DS by M [16]	PhaseOne [16]	Kodak [16]	Kodak [8]	Proposed [16]
σ^2	MSE						MSE					
0	-	-	8638	10125	7607	5678	-	-	3218	3386	2124	1864
2^8	23066	18730	11216	12723	11779	7918	22984	18701	5732	5907	5984	4048
2^9	91719	74459	18876	20418	24012	14544	91704	74596	13216	13392	17361	10508
2^{10}	365431	296588	49270	50874	72228	40686	366149	297767	43009	43177	62605	36169
σ^2	S-CIELAB						S-CIELAB					
0	-	-	14.59	8.26	5.94	3.07	-	-	19.66	8.96	6.23	3.30
2^8	167.06	76.92	59.55	55.86	52.47	28.86	854.97	395.27	268.20	263.84	245.53	138.21
2^9	461.57	209.47	131.07	128.72	121.07	66.88	2236.17	1030.97	657.09	654.67	657.09	341.27
2^{10}	1356.39	611.31	344.72	342.65	320.84	178.24	5728.73	2658.84	1715.19	1711.09	1588.10	888.43

TABLE III

THE AVERAGE MSE AND S-CIELAB DIFFERENCE PERFORMANCE OF 10 IMAGES COMPARING 6 SCHEMES. WE USED [17] ON NOISE-FREE SENSOR DATA TO GENERATE THE HIGH RESOLUTION REFERENCE IMAGE. IT WAS ANTIALIASED AND DOWNSAMPLED TO GENERATE REFERENCE IMAGES FOR THE (NOISY AND NOISE-FREE) BINNING RESULTS. "DS" DENOTES DOWNSAMPLING

Cam	Nikon D810						Sony A7RII					
Bin. Demos.	DS2 $\times$ 2 [17]	DS by M [17]	PhaseOne [17]	Kodak [17]	Kodak [8]	Proposed [17]	DS2 $\times$ 2 [17]	DS by M [17]	PhaseOne [17]	Kodak [17]	Kodak [8]	Proposed [17]
σ^2	MSE						MSE					
0	-	-	8589	11182	8888	4090	-	-	3237	3782	2330	1899
2^8	20430	15597	12061	14642	13065	7034	20743	15743	6830	7382	6193	5003
2^9	80885	61731	22437	24988	25317	18727	82816	62838	17599	18132	17575	14247
2^{10}	322630	246138	63783	66186	73556	50152	331601	251420	60738	61245	62823	51181
σ^2	S-CIELAB						S-CIELAB					
0	-	-	14.25	9.82	5.07	4.69	-	-	17.99	9.59	5.58	4.90
2^8	117.34	57.22	40.98	37.99	52.21	19.33	583.60	285.28	152.75	147.74	245.33	76.42
2^9	311.55	149.92	79.99	77.67	120.92	35.52	1525.48	742.78	356.45	352.13	609.10	180.27
2^{10}	899.59	430.87	190.72	188.79	322.09	95.10	4013.78	1962.61	907.47	902.28	1589.24	460.23

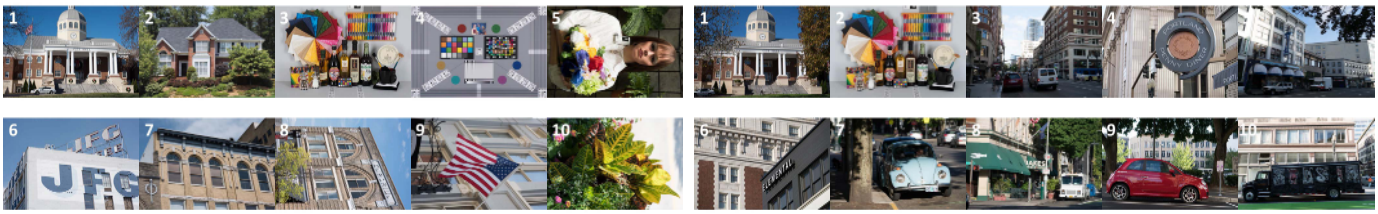
TABLE IV

THE AVERAGE MSE AND S-CIELAB DIFFERENCE PERFORMANCE OF 10 IMAGES COMPARING 6 SCHEMES. WE USED [18] ON NOISE-FREE SENSOR DATA TO GENERATE THE HIGH RESOLUTION REFERENCE IMAGE. IT WAS ANTIALIASED AND DOWNSAMPLED TO GENERATE REFERENCE IMAGES FOR THE (NOISY AND NOISE-FREE) BINNING RESULTS. "DS" DENOTES DOWNSAMPLING

Cam	Nikon D810						Sony A7RII					
Bin. Demos.	DS2 $\times$ 2 [18]	DS by M [18]	PhaseOne [18]	Kodak [18]	Kodak [8]	Proposed [18]	DS2 $\times$ 2 [18]	DS by M [18]	PhaseOne [18]	Kodak [18]	Kodak [8]	Proposed [18]
σ^2	MSE						MSE					
0	-	-	7969	9780	8532	3246	-	-	3174	3432	2468	1259
2^8	22555	17618	11153	13045	12704	6067	19753	15737	6157	6454	7607	3872
2^9	85365	67159	20581	22572	24953	14240	73716	59810	14604	14953	17711	11158
2^{10}	313873	250562	57082	59331	73167	15674	277783	230706	46049	46522	62967	38155
σ^2	S-CIELAB						S-CIELAB					
0	-	-	14.27	7.26	6.31	3.15	-	-	19.17	8.40	6.33	3.64
2^8	76.11	38.57	45.92	42.22	52.72	21.57	130.27	67.32	89.84	85.92	246.97	45.28
2^9	115.44	59.02	74.74	72.33	121.12	39.60	187.86	97.65	133.02	130.49	609.46	69.44
2^{10}	166.91	86.32	121.53	119.80	321.81	62.91	264.97	138.36	197.25	195.14	1589.66	105.14

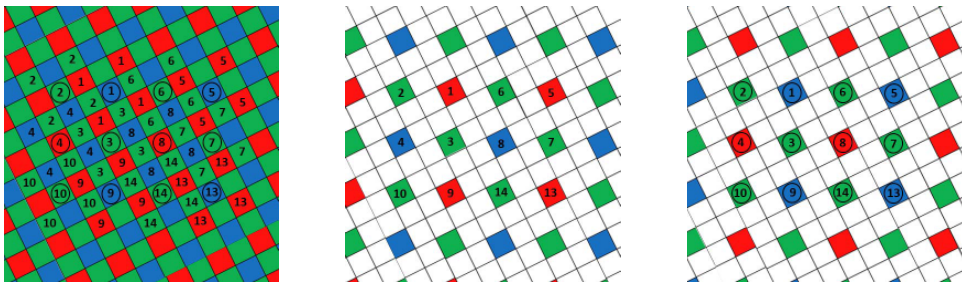
noise, Kodak PIXELUX binning with the binning-aware demosaicking [8] comes close in performance, but this approach is not robust to noise. With high noise, Kodak PIXELUX and PhaseOne had very similar performance to each other. With no binning, the noise performance was unacceptable.

We fully acknowledge the limitations to Table II-IV. First, in absence of full color image, the reference image is obtained from a demosaicked version of the full resolution CFA image. This is justifiable because in real-world applications, this is the "best possible" result we can obtain from a camera sensor. Hence, MSE values in Table II-IV represent the deviation from this desirable imaging which is also a meaningful quantity. Second, the MSE values are obviously demosaicking method-dependent. However the anti-aliasing procedure in producing the reference image makes the differences between the



Nikon D810 Sony A7RII

Fig. 12. Raw sensor test images used to evaluate MSE performance of binning schemes. Source: www.imaging-resource.com, reproduced with a written permission.



(a) (b) (c)

Fig. 13. Proposed binning-based single-shot HDR. (a) Slanted sensor. (b) Superpixels of Z_4. (c) Superpixels of Z_1.

demosaicking methods negligible; and in practice, we observed that the differences between binning schemes made a much bigger impact on the MSE than the demosaicking methods. Finally, the sampling densities of the Kodak PIXELUX and PhaseOne superpixels differ from the proposed binning scheme. There is simply no ways to make the sampling rate equal. Nevertheless, the Table II-IV captures the general trend that is consistent with the visual qualities of Figures 9-11.

V. Binning-Based Single-Shot High Dynamic Range Imaging

Inspired by the CFA-based single-shot HDR imaging of [21]–[23], we propose a binning-based single-shot HDR imaging scheme. HDR imaging is a technique to reproduce a greater dynamic range luminosity (e.g. a scene with direct sunlight and dark shadows) by merging multiple low dynamic range images. Recall that the proposed binning scheme combines 5 neighboring green pixels or 4 neighboring red/blue pixels to form superpixels. This means some of the red/blue pixels are not used in generating superpixels, as shown in Figure 1(d). We take advantage of this fact to yield a high dynamic range resolution.

Suppose we modify square lattice binning in (6) by combining four neighboring pixels of the red/green/blue color components, as follows:

$$\begin{aligned} Z_4(\boldsymbol{m}) = {} & X\left(M\boldsymbol{m} + \begin{pmatrix}1\\1\end{pmatrix}\right) + X\left(M\boldsymbol{m} + \begin{pmatrix}1\\-1\end{pmatrix}\right) \\ & \times X\left(M\boldsymbol{m} + \begin{pmatrix}-1\\1\end{pmatrix}\right) + X\left(M\boldsymbol{m} + \begin{pmatrix}-1\\-1\end{pmatrix}\right) \\ & + N(\boldsymbol{m}), \end{aligned} \tag{10}$$

where $N \sim \mathcal{N}(0, \sigma^2)$ and the subscript of $Z_4(\boldsymbol{n})$ denotes the four times the signal strength in superpixels. See Figure 13(b).

The remaining pixels that are not involved in generating superpixels Z_4 form a lower intensity image Z_1 with the same brightness level of original image $X(\boldsymbol{n})$:

$$Z_1(\boldsymbol{m}) = X(M\boldsymbol{m}) + N'(\boldsymbol{m}), \tag{11}$$

where $N' \sim \mathcal{N}(0, \sigma^2)$. See Figure 13(c).

The two generated CFA images Z_4 and Z_1 make use of all sensor pixels. As illustrated in the schematic in Figure 2(c), Z_4 and Z_1 are demosaicked separately—we refer to the resultant images as I_4 and I_1 respectively. Note that I_4 and I_1 are two images of the same scene with very different characteristics. As shown by Figure 14(b), the demosaicked image I_4 has a high risk of saturation, owing to the increased signal strength by binning. By contrast, I_1 in Figure 14(c) has much lower risk of saturation, but it suffers from noise in dark regions of the image. Hence we propose to combine the demosaicked images of Z_4 and Z_1 by treating them as long and short exposure images, respectively. Specifically, the HDR image combined with the following form:

$$I(\boldsymbol{m}) = \begin{cases} \alpha I_4(\boldsymbol{m}) + \beta I_1(\boldsymbol{m}) & \text{if } I_4(\boldsymbol{m}) < \tau \\ I_1(\boldsymbol{m}) & \text{else,} \end{cases} \tag{12}$$

where τ is the saturation value (e.g. $\tau = 2^{14}$ in Sony α6000). The linear weights of $\alpha = 4/17$ and $\beta = 1/17$ for $I_4(\boldsymbol{m})$ and $I_1(\boldsymbol{m})$, respectively, is derived as follows. Assuming that noise in I_4 and I_1 are independent and identically distributed, the noise variance in $I(\boldsymbol{m})$ is $(\alpha^2 + \beta^2)\sigma^2$, which we minimize by

$$\underset{\alpha\ \beta}{\arg\min}\ (\alpha^2 + \beta^2)\sigma^2, \tag{13}$$

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

Fig. 14. Simulated HDR output from the proposed binning-based single-shot HDR in Figure 13. (a) Image w/o. HDR. (b) Demosaicked I_1. (c) Demosaicked I_4. (d) Combined HDR Image I.

subject to the constraint that $I(\boldsymbol{m})$ has the correct intensity value:

$$4\alpha + \beta = 1. \tag{14}$$

We simulated the HDR image $I(\boldsymbol{m})$ by combing the low dynamic range images $I_4(\boldsymbol{m})$ and $I_1(\boldsymbol{m})$ in Figures 14(b) and 14(c). An example HDR output image is shown in Figure 14(d). The sensor data for proposed HDR scheme are captured by tilting the camera sensor with angle of 26.6 degrees as in Section IV-B. Pseudorandom white Gaussian noise with variance $\sigma^2 = 2^8$ was added to the low dynamic range superpixel CFA data Z_4 and the low resolution, low intensity CFA data Z_1. The demosaicked image I_1 in Figure 14(b) has the light intensity and noise level comparable to the image with no HDR, shown in 14(a). However, the scene outside is properly exposed while too low for inside scene. On the other hand, the inside scene of demosaicked image I_4 in Figure 14(c) is acceptable in terms of noise while the outside scene is saturated because of four times light intensity of superpixels. The combined HDR image shown in Figure 14(d) effectively improves the performance—both the inside and outside scene are properly exposed, with little noticeable influence of noise and saturation.

VI. Conclusion

In this article, we proposed a new pixel binning scheme for color image sensors via square lattice sampling. The resultant binning scheme achieves equivalent to 4.42 times signal strength improvement with loss of 5 times in image pixel density. It is superior to the Kodak PIXELUX and PhaseOne binning schemes in terms of resolution and color fidelity despite the lower pixel count. The proposed binning has considerably less artifacts and better noise performance in contrast with the existing binning schemes. In addition, we provided an extension to the proposed binning scheme for performing single-shot high dynamic range image acquisition. The combined HDR image effectively improves the performance, with little noticeable influence of noise and saturation.

Acknowledgment

The authors thank the staff at Imaging Resource who have generously given them the permission to reproduce the raw images used in this study.

References

[1] J. T. Compton, J. F. Hamilton, Jr., and T. E. DeWeese, "Image sensor with improved light sensitivity," U.S. Patent 8 194 296, Jun. 5, 2012.
[2] U. Barnhöfer, J. M. DiCarlo, B. P. Olding, and B. A. Wandell, "Color estimation error trade-offs," *Proc. SPIE*, vol. 5017, pp. 263–273, May 2003.
[3] H. Yamanaka, "Method and apparatus for producing ultra-thin semiconductor chip and method and apparatus for producing ultra-thin back-illuminated solid-state image pickup device," U.S. Patent 7 276 429, Oct. 2, 2007.
[4] T. W. Edwards and R. S. Pennypacker, "Manufacture of thinned substrate imagers," U.S. Patent 4 266 334, May 12, 1981.
[5] Z. Zhou, B. Pain, and E. R. Fossum, "Frame-transfer CMOS active pixel sensor with pixel binning," *IEEE Trans. Electron Devices*, vol. 44, no. 10, pp. 1764–1768, Oct. 1997.
[6] F. Chu, "Improving CMOS image sensor performance with combined pixels," Kodak Image Sensor Solutions Group, Tech. Rep., 2005.
[7] T. Andersen, "Colour binning of a digital image to reduce the image resolution," U.S. Patent 7 929 807 B2, Apr. 19, 2011.
[8] X. Jin and K. Hirakawa, "Analysis and processing of pixel binning for color image sensor," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, pp. 1–15, Dec. 2012.
[9] W. J. Kindt and B. D. Segerstedt, "Pixel binning image sensor," U.S. Appl. Note 2006 0 077 269 A1, Apr. 13, 2006.
[10] N. E. Bock, "Apparatus and method for pixel binning in an image sensor," U.S. Patent 7 091 466 B2, Aug. 15, 2006.
[11] A. Krymski, "Method and apparatus for pixel signal binning and interpolation in column circuits of a sensor circuit," U.S. Patent 7 319 218 B2, Jan. 15, 2008.
[12] N. E. Bock, "Methods for pixel binning in an image sensor," U.S. Patent 7 402 789 B2, Jul. 22, 2008.
[13] A. Lahav and D. Cohen, "Color pattern and pixel level binning for APS image sensor using 2×2 photodiode sharing scheme," U.S. Patent 7 773 138 B2, Aug. 10, 2010.
[14] B. Bayer, "Color imaging array," U.S. Patent 3 971 065, Jul. 20, 1976. [Online]. Available: https://www.google.com/patents/US3971065

[15] P. E. Rivett and N. I. P. Mackinnon, "Prime matrices," *Math. Gazette*, vol. 70, no. 454, pp. 257–259, 1986.
[16] J. T. Korneliussen and K. Hirakawa, "Camera processing with chromatic aberration," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4539–4552, Oct. 2014.
[17] K. Hirakawa and T. W. Parks, "Adaptive homogeneity-directed demosaicing algorithm," *IEEE Trans. Image Process.*, vol. 14, no. 3, pp. 360–369, Mar. 2005.
[18] D. Kiku, Y. Monno, M. Tanaka, and M. Okutomi, "Beyond color difference: Residual interpolation for color image demosaicking," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1288–1300, Mar. 2016.
[19] X. Zhang and B. A. Wandell, "A spatial extension of CIELAB for digital color-image reproduction," *J. Soc. Inf. Display*, vol. 5, no. 1, pp. 61–63, 1997.
[20] J. Farrell, M. Okincha, M. Parmar, and B. Wandell, "Using visible SNR (vSNR) to compare the image quality of pixel binning and digital resizing," *Proc. SPIE*, vol. 7537, no. 1, pp. 75370–75378, Jan. 2010.
[21] S. K. Nayar and T. Mitsunaga, "High dynamic range imaging: Spatially varying pixel exposures," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Jun. 2000, pp. 472–479.
[22] S. K. Nayar and S. G. Narasimhan, "Assorted pixels: Multi-sampled imaging with structural models," in *Proc. Eur. Conf. Comput. Vis.*, 2002, pp. 636–652.
[23] K. Hirakawa and P. M. Simon, "Single-shot high dynamic range imaging with conventional camera hardware," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 1339–1346.

Jiachao Zhang received the B.S. degree in electrical engineering from the Nanjing University of Science and Technology, Nanjing, China, in 2011, and the M.S degree in electrical engineering from the University of Dayton, Dayton, OH, USA, in 2015. She is currently pursuing the Ph.D. degree with the Nanjing University of Science and Technology. Her research focuses on statistical modeling and color image processing.

Jie Jia received the B.S. degree in telecommunication engineering from the Beijing University of Posts and Telecommunications, in 2009, the M.S. and Ph.D. degrees in electrical engineering from the University of Dayton, Dayton, OH, USA, in 2011 and 2017, respectively. He was a Research Assistant with the Intelligent Signal Systems Laboratory, University of Dayton. He is currently with Mura Inc., as a Research Scientist. His current research interests include multispectral imaging, color image processing, and computer vision.

Andong Sheng received the B.S., M.S., and Ph.D. degrees form the School of Astronautics, Harbin Institute of Technology, in 1985, 1988, and 1990, respectively. He is currently a Professor with the College of Automation, Nanjing University of Science and Technology. His research interest covers multi-agent and the nonlinear estimation theory and its application.

Keigo Hirakawa (S'00–M'05–SM'11) received the B.S. degree (Hons.) in electrical engineering from Princeton University, Princeton, NJ, USA, in 2000, the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, USA, in 2003 and 2005, respectively, and the M.M. degree (Hons.) in jazz performance studies from the New England Conservatory of Music, Boston, MA, USA, in 2006. He was a Research Associate with Harvard University, Cambridge, MA, USA, from 2006 to 2009. He is currently an Associate Professor and the Program Director of computer engineering with the University of Dayton, Dayton, OH, USA. He also heads the Intelligent Signal Systems Laboratory, University of Dayton, where the group focuses on statistical signal processing, color image processing, and computer vision.