Generalized Assorted Pixel Camera:

Post-Capture Control of

Resolution, Dynamic Range and Spectrum

Fumihito Yasuma[*†], Tomoo Mitsunaga[†] , Daisuke Iso[†] , and Shree K. Nayar[††]

†Image Sensing Technology Department, Sony Corporation

††Department of Computer Science, Columbia University

Abstract

We propose the concept of a generalized assorted pixel (GAP) camera, which enables the user to capture a single image of a scene and, after the fact, control the trade-off between spatial resolution, dynamic range and spectral detail. The GAP camera uses a complex array (or mosaic) of color filters. A major problem with using such an array is that the captured image is severely under-sampled for at least some of the filter types. This leads to reconstructed images with strong aliasing. We make four contributions in this paper: (a) We present a comprehensive optimization method to arrive at the spatial and spectral layout of the color filter array of a GAP camera, (b) We develop a novel anti-aliasing algorithm for reconstructing the under-sampled channels of the image with minimal aliasing, (c) We demonstrate how the user can capture a single image and then control the trade-off of spatial resolution to generate a variety of images, including monochrome, high dynamic range (HDR) monochrome, RGB, HDR RGB, and multispectral images and (d) Finally, the performance of our GAP camera has been verified using extensive simulations that use multispectral images of real world scenes. A large database of these multispectral images is being made available at `http://www1.cs.columbia.edu/CAVE/projects/gap_camera/` for use by the research community.

Index Terms - Color filter array, assorted pixels, sub-micron pixels, demosaicing, color reproduction, dynamic range, signal to noise ratio, multispectral imaging, skin detection.

1 Introduction

Most color image sensors use a color mosaic which is an assortment of different spectral filters. A color mosaic usually consists of three primary colors (e.g., RGB). One reason for the use of tri-chromatic filter arrays is that tri-chromatic sensing is near-sufficient in terms of colorimetric color reproducibility. It is also commonly assumed that this pixel assortment is the only practical way to sense color information with a single semiconductor image sensor[1].

Recently, new image sensing technologies have emerged that use novel pixel assortments to enhance image sensing capabilities. For high dynamic range (HDR) imaging, a mosaic of neutral density filters with different transmittances have been used [9] [8]. A new approach to high sensitivity imaging builds upon the standard Bayer mosaic by using panchromatic pixels [6] that collect a significantly larger proportion of incident radiation. Color filter arrays (CFAs) with more than three colors have been proposed to capture multispectral images [14] [2].

In this paper, we introduce the notion of a generalized assorted pixel (GAP) camera, that uses a mosaic with a richer assortment of filters and enables a user to produce a variety of image types from a single captured image. Each filter type in an assortment can serve to enhance a specific attribute of image quality. Examples of attributes are color reproduction, spectral resolution, dynamic range, and sensitivity. We propose a comprehensive framework for designing the spatial layout and spectral responses of the color filter array of a GAP camera. The following are the main contributions of our work: (a) We develop an optimization method to arrive at the spatial and spectral layout of the color filter array of a GAP camera. The cost function that we optimize includes terms related to colorimetric/spectral reproduction, dynamic range and signal-to-noise ratio (SNR), (b) We develop a novel anti-aliasing algorithm for reconstructing the under-

---

[1]The Foveon X3 sensor [7] is an exception.

sampled channels of the image with minimal aliasing. Our approach is to use a sub-micron pixel size to avoid aliasing for some of the channels. The high frequency content from these channels are then used to remove aliasing from the remaining (under-sampled) channels, (c) We have developed software that enables a user to capture a single image and then control the trade-off of spatial resolution to generate a variety of images. The output image can be monochrome, HDR monochrome, RGB, HDR RGB, or multispectral and (d) Finally, the performance of our GAP camera has been verified using extensive simulations that use multispectral images of real world scenes. The multispectral images are used to emulate GAP camera images and results computed from the GAP images are compared with the original multispectral images. We have released a large database of high quality multispectral images (at `http://www1.cs.columbia.edu/CAVE/ projects/gap_camera/`) for use by the research community.

The trend in manufacturing has been towards producing sensors with increasing numbers of smaller pixels. The time is therefore ripe for exploring more interesting pixel assortments than the ones used in the past. Furthermore, each of the previously proposed mosaics have been used to generate one specific type of output image. In contrast, our goal is to create a mosaic that lends itself to post-capture control over the output image. Since sensor fabrication is a very expensive endeavor, we have used high quality multispectral data as our ground truth as well as to verify our optimized mosaic and reconstruction algorithm. Given the high quality of results we have obtained, we have begun to pursue the fabrication of a GAP sensor.

2   Assorted Pixel Imaging with Sub-Micron Pixels

The resolution of an optical imaging system may be limited by multiple factors, but the dominant factors are *diffraction* and *aberration*. While aberrations can be corrected for

during lens design, diffraction is a fundamental limitation that cannot be avoided. Therefore, we assume an aberration-corrected optical system and focus on only diffraction. The 2-D diffraction pattern of a lens with a circular aperture is called the *Airy disk*. The width of the Airy disk determines the maximum resolution limit of the system and is given by: $I(\theta) = I_0\{2J_1(x)/x\}^2$, where I_0 is the intensity in the center of the diffraction pattern, J_1 is the Bessel function of the first kind of order one, and θ is the angle of observation. $x = \frac{\pi q}{\lambda N}$, where q is the radial distance from the optical axis in the observation plane, λ is the wavelength of incident light, and N is the f-number of the system. In the case of an ideal lens, this diffraction pattern is the Point Spread Function (PSF) for an in-focus image. The Fourier transformation of the PSF is used to characterize the resolution of the optical imaging system. This quantity is referred to as the Modulation Transfer Function (MTF). The MTF of an ideal optical system can be calculated directly from the wavelength λ of incident light and the f-number N, and is denoted as $MTF_{opt}(\lambda, N) = \mathcal{F}(I(\theta))$, where, $\mathcal{F}(\cdot)$ denotes the Fourier transformation.

Pixels typically have a rectangular shape, and their finite size contributes to the resolution characteristics of the imaging system. The MTF of an image sensor can be approximated as the Fourier transformation of a rectangular function which is denoted by $MTF_{sensor}(p) = \mathcal{F}(x(t))$, where

$$
x(t) = \begin{cases} 1 & |t| \leq \frac{p\zeta}{2} \\ 0 & |t| > \frac{p\zeta}{2} \end{cases}, \tag{1}
$$

is the rectangular function, p is pixel size, and ζ is the pixel aperture ratio, which is assumed to be 1 due to the use of on-chip microlenses. The total fundamental optical resolution limit of a camera (lens and sensor) is denoted in the frequency domain as $MTF = MTF_{opt}(\lambda, N) \cdot MTF_{sensor}(p)$. To compute an MTF, we use the values $\lambda =$
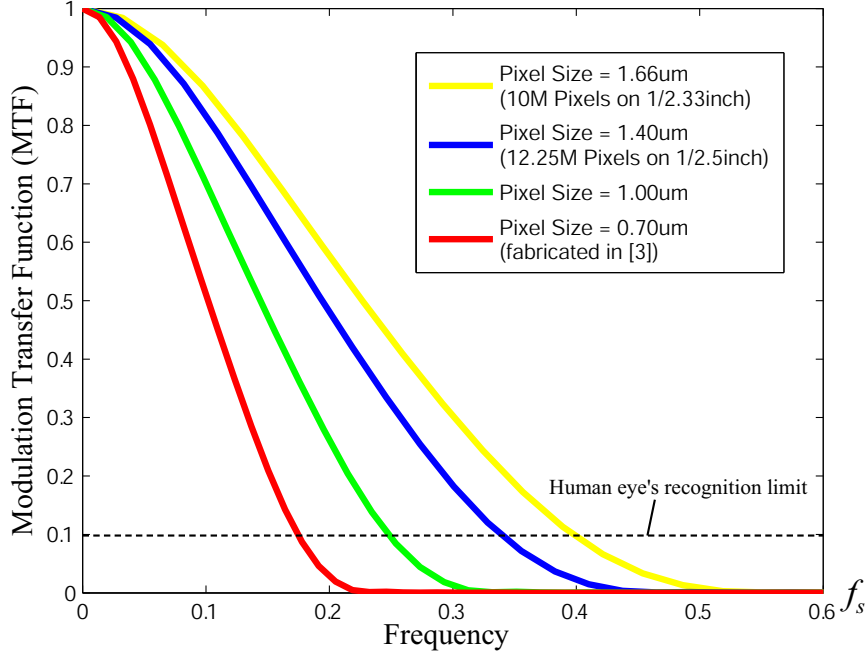
Figure 1: The optical resolution limits (MTFs) corresponding to different pixel sizes ($\lambda = 555nm$ and $N = f/5.6$). The MTF for pixel size $p = 1.0\mu m$ is 0.1 at about $0.25f_s$ (f_s is the image sensor's sampling frequency). We consider the optical resolution limit of an image sensor with $p = 1.0\mu m$ pixel size to be half of the image sensor's Nyquist frequency. The resolution performance of a sensor with sub-micron pixels exceeds the optical resolution limit.

$555nm$ (corresponding to the peak of the sensitivity of the human eye) and $N = f/5.6$ (which is a pupil size commonly used in consumer photography). With these numbers fixed, the fundamental MTF is determined only by pixel size p. The MTFs for various pixel sizes are shown in Figure 1. In this figure, the minimum pixel size we use is $0.7\mu m$, which is the pixel size of the fabricated detector described in [3]. Note that the MTF for pixel size $p = 1.0\mu m$ is 0.1 at about $0.25f_s$, where, f_s is the image sensor's sampling frequency. Generally, the human eye cannot recognize contrast when the MTF is less than 0.1. Thus, we can consider the optical resolution limit of an image sensor with $p = 1.0\mu m$ pixel size to be half of the image sensor's Nyquist frequency. From this, we can conclude that the resolution performance of a sensor with sub-micron pixels exceeds the optical

5
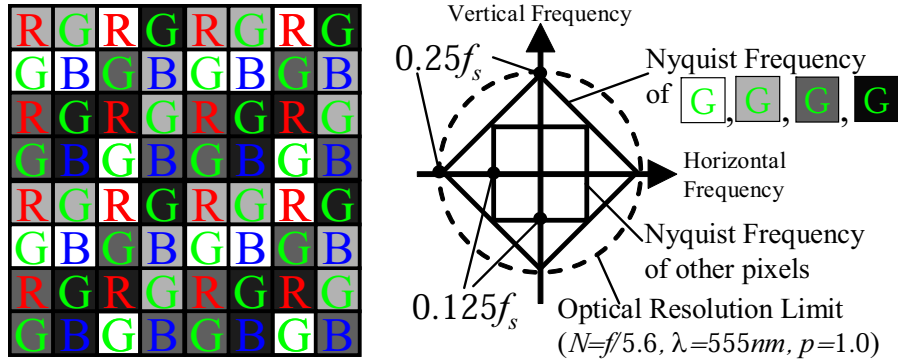
(a) 3 colors and 4 exposures CFA in [8] and its Nyquist Limits
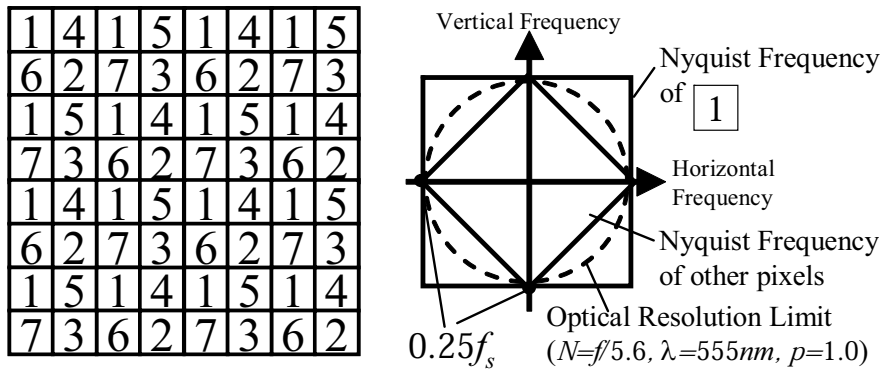
(b) 7 colors and 1 exposure CFA in [2] and its Nyquist Limits

Figure 2: Nyquist Limits of previous assorted designs used with sub-micron pixel image sensors (pixel pitch $p = 1.0nm$).

resolution limit.

Figure 2 shows the Nyquist limits when the CFA patterns of previous assorted pixels are used with the sub-micron pixel size image sensor. When the highest frequency of the input signal is lower than the Nyquist limit, aliasing does not occur, according to the sampling theorem. Therefore, aliasing is not generated at pixels marked '1' in Figure 2(b).
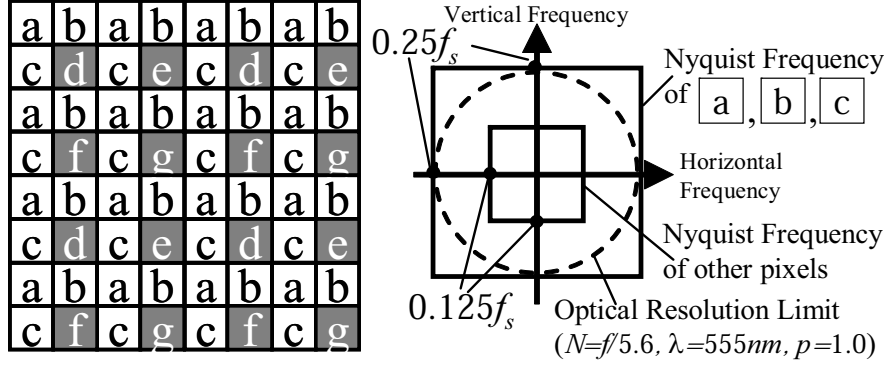
Figure 3: Our proposed GAP mosaic (7 colors and 2 exposures) used with a sub-micron pixel image sensor (pixel pitch $p = 1.0nm$) and its Nyquist Limits.

3 Spatial Design of GAP Mosaic

For a $1.0\mu m$ pixel size, the combined MTF due to both diffraction by the lens aperture and averaging by pixels, leads to an optical resolution limit of almost $1/4$ of the sampling frequency $f_s = 1/\Delta s$, where Δs is the sampling pitch. To exploit this property, we propose a novel CFA to be used in conjunction with a sub-micron image sensor, which is shown in Figure 3. The pixels marked 'a', 'b' and 'c' in the CFA capture three different spectral images on a rectangular grid with sampling pitch $\Delta s_{a,b,c} = 2p$. Thus, the Nyquist frequency for 'a', 'b' and 'c' pixels is $f_{n a,b,c} = f_{s a,b,c}/2 = f_s/4$. Note that '$a$', '$b$', and '$c$' pixels *do not* suffer from aliasing because the optical resolution limit is $0.25f_s$ due to diffraction. These aliasing-free pixels are used for reconstruction of high resolution images and we refer to their filters as the *primary filters* of the mosaic. Note that we have increased the number of aliasing-free pixels from one color in conventional mosaics to three colors in our GAP design (See Figure 2). This change results in significantly better anti-aliasing for our non-primary filters, as described next.

Pixels 'd', 'e', 'f', and 'g' each sample the incident image on rectangular grids through different spectral filters, and are referred to as the *secondary filters*. The sampling pitch for each of these filters is $\Delta s_{d,e,f,g} = 4p$, and the Nyquist frequency is

$f_{nd,e,f,g} = f_{sd,e,f,g}/2 = f_s/8$. These Nyquist frequencies are lower than the optical resolution limit, which can lead to aliasing artifacts. However, this aliasing can be removed by using high frequency information from the demosaiced images obtained using the primary filters.

Due to the nature of the cost function used in our optimization procedure, the primary filters end up with spectral responses that closely resemble the red, green and blue filters commonly used in color image sensors. As a result, the primary filters can be used to compute RGB images which essentially cover the entire visible wavelength spectrum. In other words, images captured by the secondary filters, irrespective of their spectral responses, are guaranteed to be highly correlated with the images obtained using the primary filters. Consequently, the demosaiced image obtained using the primary filters can be used to perform anti-aliasing of the images produced by the secondary filters. Furthermore, our cost function also results in the secondary filters having lower exposures than the primary ones. Hence, by using all of the primary and secondary filters, we can obtain high dynamic range information. Finally, since the 7 filters (primary and secondary) have different spectral responses, their reconstructed images can be used to obtain smooth estimates of the complete spectral reflectance distribution of each scene point, i.e., a multispectral image.

4   Spectral Responses of GAP Filters

Our GAP mosaic allows a variety of image characteristics to be captured simultaneously, but a trade-off must be made with respect to the fidelity of each characteristic. Monochrome and RGB images are reconstructed at high resolution from the primary filters. For HDR images, the dynamic range is improved by utilizing the secondary filters and decreasing the spatial resolution. We can obtain multispectral images due to low

resolution images captured by secondary filters. We now develop a cost function that incorporates several terms, including quality of color reproduction, reconstruction of reflectance and dynamic range.

4.1 The Cost Function

The value x_m measured at a pixel in the mth channel ($m \in \{a, b, c, d, e, f, g\}$) is given by

$$x_m = \int_{\lambda_{min}}^{\lambda_{max}} i(\lambda)r(\lambda)c_m(\lambda)d\lambda, \tag{2}$$

where $i(\lambda)$ is the spectral distribution of the illumination, $r(\lambda)$ is the spectral reflectance distribution of the scene point, and $c_m(\lambda)$ is the spectral response of the camera's m^{th} color channel. When the wavelength λ is sampled at equally-spaced L points, Eq.2 becomes a discrete expression:

$$x_m = \sum_{l=1}^{L} i(\lambda_l)r(\lambda_l)c_m(\lambda_l). \tag{3}$$

If we rewrite Eq.3 in matrix form, we obtain

$$\mathbf{X} = \mathbf{C^T I R}, \tag{4}$$

where $\mathbf{X} = [x_a, x_b, ...x_g]^{\mathbf{T}}$, $\mathbf{C} = [c_m(\lambda_l)]$, $\mathbf{I}$ is a diagonal matrix made up of the discrete illumination samples $i(\lambda_l)$, and $\mathbf{R} = [r(\lambda_l)]$. Our goal now is to determine the 7 spectral response functions in $\mathbf{C}$. We do this by dividing up our cost function into several parts, each of which is described below.

9

4.1.1 Cost 1: Color Reproduction of RGB Image

To obtain HDR RGB images, a high exposure RGB image is reconstructed from the primary filters, and a low exposure image is reconstructed from the secondary filters. The spectral responses of all the filters must ideally yield the highest color reproduction. A variety of filter rating indices have been proposed to evaluate the color reproduction characteristics of a filter [13][12]. These indices use a cost function that minimizes the difference between the measured color of a reference material and its known color. To calculate this difference, we use the CIE 1931 XYZ color space, which is based on direct measurements of human visual perception. The calculation of sRGB tristimulus values (which are employed in many digital cameras and color monitors) from the CIE XYZ tristimulus values uses a linear transformation. The CIE XYZ tristimulus values are defined as $\mathbf{Y} = \mathbf{A^T IR}$, where $\mathbf{Y}$ represents the true tristimulus values, and $\mathbf{A}$ is a matrix of CIE XYZ color matching functions $[\bar{x}\,\bar{y}\,\bar{z}]$. The estimated CIE tristimulus values corresponding to the primary filters $\hat{\mathbf{Y}}'$ can be expressed as an optimal linear transformation: $\hat{\mathbf{Y}}' = \mathbf{T'X'}$, where $\mathbf{X'} = [x_a, x_b, x_c]^\mathbf{T}$. The transformation $\mathbf{T'}$ is determined so as to minimize the color difference: $\min \|\mathbf{Y} - \mathbf{T'X'}\|^2$. The estimated CIE tristimulus values corresponding to the secondary filters are denoted as $\hat{\mathbf{Y}}'' = \mathbf{T''X''}$, where $\mathbf{X''} = [x_d, x_e, x_f, x_g]^\mathbf{T}$.

The average magnitude of the color difference between the true color $\mathbf{Y}$ and the estimate $\hat{\mathbf{Y}}$ over a set of N real-world objects may be used as a metric to quantify the camera's color reproduction performance. The color reproduction errors corresponding to the primary and secondary filters can therefore be written as:

$$E'(\mathbf{C}) = \min_{\mathbf{T'}} \sum_{n=1}^{N} \|\mathbf{Y}_n - \mathbf{T'X'}_n\|^2, \tag{5}$$

$$E''(\mathbf{C}) = \min_{\mathbf{T''}} \sum_{n=1}^{N} \|\mathbf{Y}_n - \mathbf{T''X''}_n\|^2. \tag{6}$$

10

4.1.2 Cost 2: Reconstruction of Spectral Reflectance

In this paper, we use the model-based spectral reconstruction method described in [10]. Fortunately, the spectral reflectance distribution of most real-world surfaces can be well-approximated using a low-parameter linear model. The linear model we use is the set of orthogonal spectral basis functions $b_k(\lambda)$ proposed by Parkkinen *et al* [11]:

$$r(\lambda) = \sum_{k=1}^{K} \sigma_k b_k(\lambda), \tag{7}$$

where σ_k are scalar coefficients and K is the number of basis functions. By substituting Eq.7 in Eq.2, we get a set of equations:

$$x_m = \sum_{k=1}^{K} \sigma_k \int_{\lambda_{min}}^{\lambda_{max}} b_k(\lambda)i(\lambda)c_m(\lambda)d\lambda. \tag{8}$$

These equations can be written as $\mathbf{X} = \mathbf{F} \cdot \boldsymbol{\sigma}$, where $\mathbf{F}$ is a $M \times K$ matrix: $\mathbf{F} = \int_{\lambda_{min}}^{\lambda_{max}} b_k(\lambda)i(\lambda)c_m(\lambda)d\lambda$, M is number of color filter channels (in our GAP mosaic, $M = 7$), and $\boldsymbol{\sigma} = [\sigma_k]$. The spectral reflectance distribution is reconstructed by minimizing $\|\mathbf{F} \cdot \sigma - \mathbf{X}\|^2$. Note that the spectral reflectance distribution of most real-world materials is known to be smooth and must be positive [10]. Thus, the reconstruction problem can be posed as a constrained minimization as follows:

$$\hat{\sigma} = \arg\min_{\sigma} \|\tilde{\mathbf{F}} \cdot \sigma - \tilde{\mathbf{X}}\|^2, \text{ subject to } \mathbf{B} \cdot \sigma \geq \mathbf{0}, \tag{9}$$

where $\tilde{\mathbf{F}} = \left[\mathbf{F^T}\ \alpha\mathbf{P^T}\right]^{\mathbf{T}}$, $\mathbf{P}_{lk} = \partial^2 b_k(\lambda_l)/\partial\lambda^2$ is a smoothness constraint, α is a smoothness parameter, $1 \geq l \geq L$, $1 \geq k \geq K$, $\tilde{\mathbf{X}} = \left[\mathbf{X^T}\ \mathbf{0}\right]^{\mathbf{T}}$, and $\mathbf{B} = [b_k(\lambda_l)]$. This regularized minimization can be solved using quadratic programming. The multispectral

image's mean squared reconstruction error $R(\mathbf{C})$ is given by

$$R(\mathbf{C}) = \sum_{n=1}^{N} \|\sigma_n - \hat{\sigma}_n\|^2, \qquad (10)$$

where σ_n represents the actual coefficients of the nth object and $\hat{\sigma}_n$ are the reconstructed coefficients. In our implementation, the number of basis functions K is 8 and the smoothness parameter α is set to 64.0 [10].

4.1.3 Cost 3: Dynamic Range and SNR

To achieve HDR imaging, our secondary filters have lower transmittances than the primary filters, as mentioned earlier. This may cause deterioration of the signal-to-noise ratio (SNR) for the secondary filters. This trade-off can be controlled based on the ratio of the exposures of the primary and secondary filters: $\beta = \frac{e_{max}}{e_{min}}$, where e_{max} is the average exposure of the primary filters and e_{min} is the average exposure of the secondary filters. Therefore, β is determined by $\mathbf{C}$ in Eq.4. Our goal here is to determine the value of β that best balances extension of dynamic range versus reduction of SNR.

Dynamic range is often defined as $DR = 20 \log_{10} \frac{V_{full}}{N_r}$, where V_{full} represents the full-well capacity of the detector, and N_r is the total noise of the detector. The total noise of the detector is defined as $N_r = \sqrt{N_{shot}^2 + N_{dark}^2}$, where $N_{shot} = \sqrt{V}$ is the shot noise, V is the signal, and N_{dark} is the dark noise [4]. In the case of a GAP camera, we do not change N_r, but the maximum detectable level becomes $\beta\, V_{full}$ [9]. Hence, the dynamic range of a GAP camera is

$$DR_{GAP} = 20 \log_{10} \frac{\beta\, V_{full}}{N_r}. \qquad (11)$$

The SNR can be written as: $SNR = 20 \log_{10} \frac{V}{N}$, where V is the signal and N is the noise. The signal corresponding to a secondary filter can be expressed using the exposure

12

ratio β as $V'' = V'/\beta$, where V' is a signal due to a primary filter. When the signal due to the primary filter is not saturated, the signal due to the secondary filter can be determined from the primary signal. The SNR for a secondary filter when the primary signal is saturated is the worst-case SNR of the GAP mosaic:

$$SNR_{GAP} = 20\log_{10}\frac{V_{full}/\beta}{N_{max}}, \tag{12}$$

where $N_{max} = \sqrt{N''^2_{shot} + N^2_{dark}}$, and $N''_{shot} = \sqrt{\frac{V_{full}}{\beta}}$.

Since the camera has a high performance in SNR and dynamic range when SNR_{GAP} and DR_{GAP} are both large, we use the following as our cost function:

$$D(\mathbf{C}) = \frac{1}{DR_{GAP}} \cdot \frac{1}{SNR_{GAP}}. \tag{13}$$

In our implementation, we have used $V_{full} = 3500e^-$ and $N_{dark} = 33e^-$ (see [3]).

4.1.4 Total Cost Function

Since each of the above cost functions represents a specific dimension of image quality, our final cost function is a weighted sum of the individual costs:

$$G = w_1\{E' + E''\} + w_2R + w_3D. \tag{14}$$

The weights (w_1, w_2, w_3) must be determined according to the image quality requirements of the application for which the GAP camera is manufactured. Since all camera filters must have positive spectral responses ($\mathbf{C}$ must be positive), the optimization of $\mathbf{C}$ can be written as

$$\mathbf{C} = \arg\min_{\mathbf{C}} G, \quad \text{subject to } \mathbf{C} \geq \mathbf{0}. \tag{15}$$

4.1.5 Initial Guesses for Filter Spectral Responses

Note that in the absence of additional constraints, our goal of finding the 7 spectral response functions in $\mathbf{C}$ is an intractable optimization problem. Therefore, we assign initial guesses to the filter responses. These filter guesses are driven by two factors: (a) They are selected from a set of 177 commercially available optical band pass filters [1] and on-chip filters [7]. (b) The commercial filters are assigned to the 7 channels based on only one of our cost functions, namely, color reproduction. That is, we find the primary filters $\mathbf{C}_0'$ and secondary filters $\mathbf{C}_0''$ such that:

$$\min_{\mathbf{C}_0'} E(\mathbf{C}_0') \quad (\mathbf{C}_0' \in \mathbf{C}_0), \tag{16}$$

$$\min_{\mathbf{C}_0''} E(\mathbf{C}_0'') \quad (\mathbf{C}_0'' \in \mathbf{C}_0), \tag{17}$$

where $\mathbf{C}_0$ is the set of 177 commercial filters. Once the 7 assignments are made in this way, they are used as initial guesses in the final stage of the optimization. This final stage is a constrained non-linear minimization of Eq.15 which requires the used of an iterative algorithm. In our implementation, we used the `fmincon` routine of Matlab. For the weights, we have used $w_1 = w_2 = w_3 = 1.0$. As mentioned earlier, these weights can be chosen differently to meet the needs of the application.

4.2 Results of GAP Filter Optimization

Using the above optimization, we obtain the optimal filter spectra shown in Figure 4. We use the spectral reflectance distribution of color patches in the Macbeth color chart and the Munsell color book as the known references $r(\lambda)$, and the illuminance spectrum of D65 for $i(\lambda)$. Three observations are worth making. First, as a result of the color reproduction term in the cost function, the primary filters are close in their responses to red, green and blue filters. Second, due to the spectral reconstruction term, the computed filters
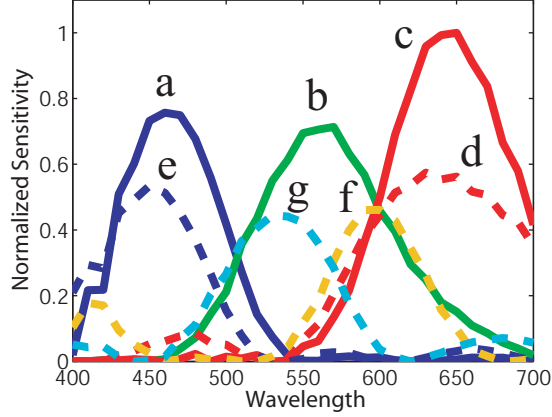
Figure 4: The spectral responses of the seven optimized filters. The secondary filters ('d', 'e', 'f', 'g') have lower exposures than the primary ones ('a', 'b', 'c'). Hence, using the primary and secondary filters, we can obtain high dynamic range information. Since the 7 filters (primary and secondary) have different spectral responses, their reconstructed images can also be used to obtain smooth estimates of the complete spectral reflectance distribution.

nicely sample the visible spectrum, which enables the GAP camera to produce reliable multispectral images. Third, due to the nature of our cost function, the primary filters have higher transmittances than the secondary filters. This enables the capture of high dynamic range images.

Table 1 shows the errors in the color reproduction and spectral reconstruction components of our cost function G, the estimated dynamic range, and the SNR of the initial and final (optimized) set of seven filters. Note that all the errors except SNR are reduced as

Table 1: Optimization Accuracy

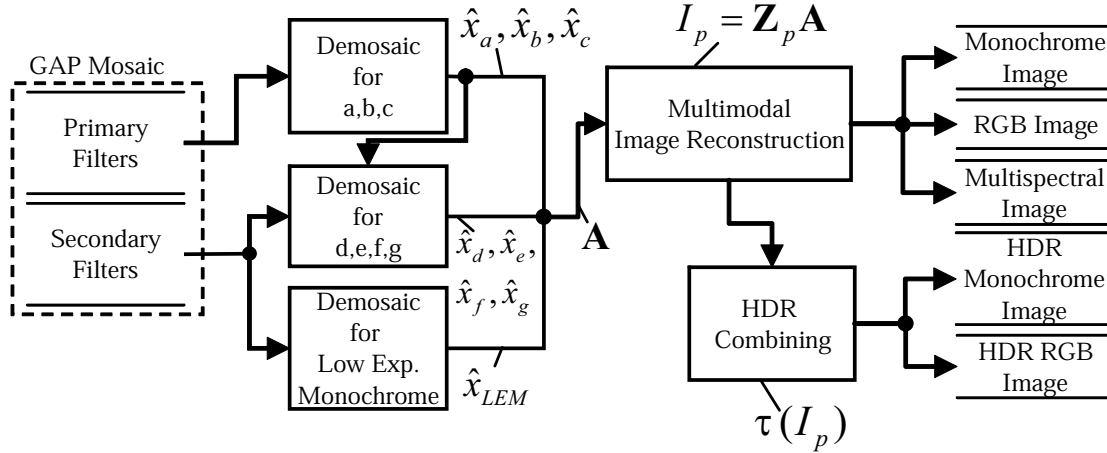|	Initial filters	Optimal filters	Bayer
$\Delta E'(\mathbf{C})$	0.0497	0.0429	0.0490
$\Delta E''(\mathbf{C})$	0.0100	0.0055	N/A
$\Delta R(\mathbf{C})$	0.0624	0.0610	0.0709
DR_{GAP}	58.2970	62.9213	56.9020
SNR_{GAP}	34.7069	32.3694	N/A

15

Figure 5: An overview of the proposed multimodal image reconstruction.

a result of the optimization. The deterioration of SNR is kept low at around 2.3dB while the dynamic range is improved by about 4.6dB. To illustrate the superiority of our GAP mosaic, the results corresponding to a set of 3 optimized color filters in a Bayer mosaic pattern are also included in Table1. The errors of the low exposure RGB image's color reproduction $\Delta E''$ and SNR all involved spectral responses of secondary filters. As a result, there is no corresponding value for these in the Bayer mosaic. Note that all performances are higher for our GAP mosaic.

5   Post-capture Control of Image Types

At each pixel of the GAP mosaic in Figure 3, there is only one color measurement, which means that the other colors must be estimated from neighboring pixels in order to produce interpolated output images (irrespective of their type). This process is commonly referred to as "demosaicing."

Denoting Λ_m as the set of pixel locations (i, j) for filter $m \in \{a, b, c, d, e, f, g\}$, a

16
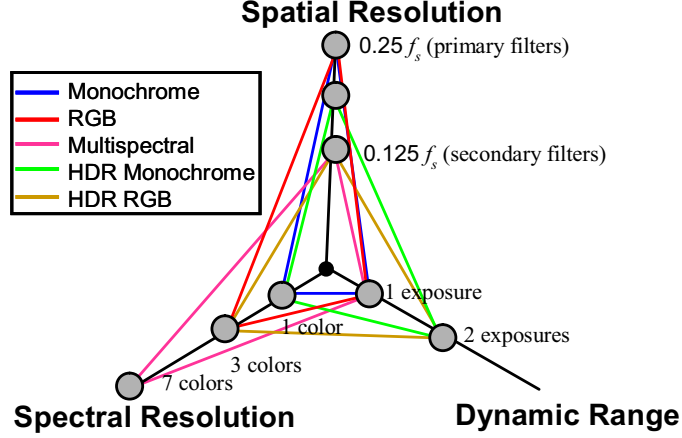
Figure 6: The diagram of control shows the trade-off between spatial resolution, dynamic range and spectral detail. Our proposed GAP mosaic lends itself to post-capture control over the output image, including monochrome, HDR monochrome, RGB, HDR RGB, and multispectral images.

mask function for each filter can be defined as

$$
W_m(i,j) = \begin{cases} 1 & (i,j) \in \Lambda_m \\ 0 & otherwise \end{cases}.
\tag{18}
$$

In the GAP mosaic, there are seven types of color channels: a, b, c, d, e, f, and g. There-fore, the observed data $y(i,j)$ is

$$
y(i,j) = \sum_{m \in \{a,b,c,d,e,f,g\}} W_m(i,j)x_m(i,j),
\tag{19}
$$

where x_m is m^{th} channel's full resolution image, given by Eq.2. Figure 5 shows the complete framework of our proposed multimodal image reconstruction. The interpolated image after demosaicing is denoted as $\hat{x}$. Different types of images are reconstructed from all interpolated images by simply changing the image reconstruction matrix:

$$
I_p = \mathbf{Z}_p \mathbf{A},
\tag{20}
$$

17

where I_p is the reconstructed image (which can be monochrome, HDR monochrome, RGB, HDR RGB, or multispectral), $\mathbf{Z}_p$ is an image reconstruction matrix, and $\mathbf{A}$ is an interpolated image set denoted as a vector: $\mathbf{A} = \begin{bmatrix} \hat{x}_a & \hat{x}_b & \hat{x}_c & \hat{x}_d & \hat{x}_e & \hat{x}_f & \hat{x}_g & \hat{x}_{LEM} \end{bmatrix}^{\mathbf{T}}$. The user can control the trade-off of spatial resolution to generate a variety of images by changing the image reconstruction matrix Z_p from a single captured image $y(i,j)$. Figure 6 shows the trade-off between spatial resolution, dynamic range and spectral detail of reconstructed images in the case of our GAP camera. We now describe the different processing operations of Figure 5.

5.1   Demosaicing for 'a', 'b', and 'c' Images

As described in section 3, images captured by the primary filters do not suffer from aliasing. Therefore, we can estimate the missing data using a simple interpolation. The 'a', 'b', and 'c' channels' images $\hat{x}_a$, $\hat{x}_b$, and $\hat{x}_c$ are reconstructed using just the data measured by the primary filters, to maintain high resolution. For interpolation we use a Finite Impulse Response (FIR) filter $F(i,j)$:

$$\hat{x}_\nu(i,j) = W_\nu(i,j)y(i,j) + \bar{W}_\nu(i,j)\left[F(i,j) * y(i,j)\right], \tag{21}$$

where $\nu = a, b,$ or c, $*$ denotes convolution, and $\bar{W}(i,j) = 1 - W(i,j)$. To minimize the loss of high frequencies due to interpolation, we used Matlab's `fir2` function to find FIR filters of size 30×30 that pass all frequencies.

5.2   Demosaicing for 'd', 'e', 'f', and 'g' Images

Interpolated secondary filter images – $\hat{x}_d$, $\hat{x}_e$, $\hat{x}_f$, and $\hat{x}_g$ – can be respectively computed using only the 'd', 'e', 'f', or 'g' pixels. However, this results in severe aliasing (see Figure 7(a)). In conventional demosaicing methods for RGB mosaics [5], an assumption

18

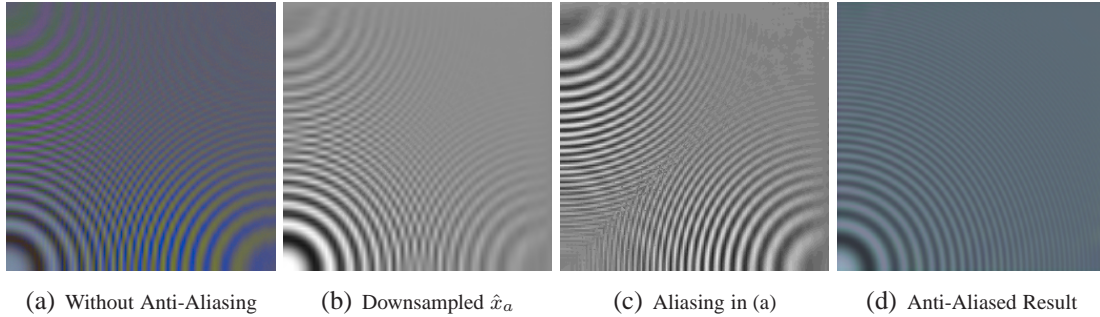(a) Without Anti-Aliasing   (b) Downsampled $\hat{x}_a$ (c) Aliasing in (a)   (d) Anti-Aliased Result

Figure 7: The anti-aliasing algorithm (simulated with $N = f/5.6$, p=1.0μm). (a) Low exposure RGB image computed from the secondary filters without anti-aliasing (false color artifacts caused by aliasing are observed). (b) Downsampled image $\Omega * \{W_e(i,j)\hat{x}_a(i,j)\}$ computed using the pixels with primary filter 'a'. (c) Aliasing $\Upsilon_e(i.j)$ estimated using (b) and the full resolution image for channel 'a' (brightness enhanced for visualization). (d) Low exposure RGB image obtained after anti-aliasing using the image in (c).

of strong positive inter-channel correlation is commonly used so as to suppress aliasing of the sparsely sampled channels (R, B) by estimating the amount of aliasing from the high frequency information of a densely sampled channel (G). However, this assumption often results in artifacts because the differences in the spectral responses of RGB filters cause the inter-channel correlation of RGB to be not always strongly positive. On the other hand, our anti-aliasing method can exploit the inherent inter-channel correlations within GAP mosaic. As shown in Figure 4, one primary filter color can be chosen for each secondary filter color in terms of similarity of the spectral response, with high expectation of strong positive inter-channel correlation due to strong overlap between the spectral responses of the chosen primary and secondary channels. For example, channel 'a' is chosen as a strongly correlated channel with channel 'e'. So we first sample the interpolated full resolution 'a' filter image $\hat{x}_a$ at all 'e' locations to estimate the aliasing of 'e' filter image. These samples are then used to compute a full resolution image for the 'e' filter: $\Omega * \{W_e(i,j)\hat{x}_a(i,j)\}$, where Ω represents an low-pass filter. We used a bilinear interpolation filter for Ω. Aliasing can be inferred by subtracting the original $\hat{x}_a$ image

19

from this interpolated one. To get the final estimate of aliasing in the 'e' channel, Υ_e, we need to scale this difference by Ψ_{ae} – the ratio of the filter transmittances of the 'a' and 'e' pixels – since the exposures of 'a' and 'e' pixels are different. We assume that the response of channel 'e' is proportional to the response of channel 'a' to the same incident light with a coefficient. The estimated aliasing Υ_e is given by:

$$\Upsilon_e(i.j) = \Psi_{ae} \left[\Omega * \{ W_e(i,j) \hat{x}_a(i,j) \} - \hat{x}_a(i,j) \right], \tag{22}$$

where $\Psi_{ae} = \frac{\|C_e\|}{\|C_a\|}$. The anti-aliased image $\hat{x}_e$ is obtained as:

$$\hat{x}_e(i,j) = \Omega * \{ W_e(i,j) y(i,j) \} - \Upsilon_e(i.j). \tag{23}$$

Since the same sampling W_e and low-pass filter Ω are used, the aliasing component in $\Omega * \{ W_e(i,j) y(i,j) \}$ and estimated aliasing $\Upsilon_e(i.j)$ are identical. The other anti-aliased secondary filter images – $\hat{x}_d, \hat{x}_f, \hat{x}_g$ – can be similarly computed. Figure 7 shows an example that illustrates the efficacy of this anti-aliasing technique.

5.3 Demosaicing for Low Exposure Monochrome Images

In order to compute an HDR monochrome image, we need to first compute a low exposure monochrome image. We can construct this low exposure image using only the 4 secondary filters which have lower exposure and also collectively cover the entire visible spectrum (Figure 4). In Figure 3, we see that 4 different secondary pixels are arranged diagonally about each 'a' pixel. Therefore, the monochrome value at each 'a' pixel can be computed as the average of the measurements at the 4 neighboring secondary pixels:

$L(i, j) = W_a(i, j)\{\mathbf{Q}_D * y(i, j)\}$, where

$$\mathbf{Q}_D = \begin{pmatrix} \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} \end{pmatrix}. \tag{24}$$

Note that by adding four pixels in a diagonal neighborhood, aliasing caused by half-pixel phase shifts gets canceled out[2]. The values at the 'a' pixels are then interpolated for all other pixels to yield the low exposure monochrome image $\hat{x}_{LEM}$:

$$
\begin{aligned}
\hat{x}_{LEM}(i, j) &= L(i, j) + W_s\{\mathbf{Q}_D * L(i, j)\} \\
&+ W_b\{\mathbf{Q}_H * L(i, j)\} + W_c\{\mathbf{Q}_V * L(i, j)\},
\end{aligned} \tag{25}
$$

where $W_s(i, j) = \begin{cases} 1 & (i, j) \in \{d, e, f, g\} \\ 0 & otherwise \end{cases}$, and $\mathbf{Q}_H = \mathbf{Q}_V^T = [0.5\ 0\ 0.5]$.

5.4 Multimodal Image Reconstruction

As shown in Figure 3, the primary filters capture images at a higher sampling frequency than the secondary filters. Thus, the spatial resolution of $\hat{x}_a$, $\hat{x}_b$, and $\hat{x}_c$ is higher than that of $\hat{x}_d$, $\hat{x}_e$, $\hat{x}_f$, and $\hat{x}_g$. Although the aliasing of images reconstructed from secondary filters is reduced due to our anti-aliasing process, the usage of $\hat{x}_{LEM}$ and $\hat{x}_d$, $\hat{x}_e$, $\hat{x}_f$,$\hat{x}_g$ slightly degrades the spatial resolution of the reconstructed image. We now describe how each of the different output images can be reconstructed with the least loss in spatial resolution.

[2]Note that this anti-aliasing method can only be used for a monochrome image computed from the low exposure secondary filters and not for a color image computed from the same.

5.4.1 Reconstruction of Monochrome Image

Monochrome image I_M is reconstructed using just the data measured by the primary filters to maintain high resolution:

$$I_M = \mathbf{Z}_M \mathbf{A}, \tag{26}$$

where $\mathbf{Z}_M = \begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$.

5.4.2 Reconstruction of RGB Image

To construct the RGB image I_{RGB}, we use the color reproduction matrix $\mathbf{T}'$ (section 4.1.1) and $\mathbf{H}$ (linear transformation from CIE XYZ to sRGB) to combine the information in the $\hat{x}_a, \hat{x}_b,$ and $\hat{x}_c$ images computed using only the primary pixels (Eq.21):

$$I_{RGB} = \mathbf{Z}_{RGB} \mathbf{A}, \tag{27}$$

where $\mathbf{Z}_{RGB} = \begin{bmatrix} \mathbf{HT}' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$.

5.4.3 Reconstruction of HDR Monochrome and HDR RGB Image

We combine the monochrome image I_M and the anti-aliased low exposure monochrome image $\hat{x}_{LEM}$ to produce the HDR monochrome image I_{HDRM}:

$$I_{HDRM} = \tau(\mathbf{Z}_{HDRM} \mathbf{A}), \tag{28}$$

where $\mathbf{Z}_{HDRM} = \begin{bmatrix} 1/3 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$, and $\tau(\cdot)$ is the processing that combining the HDR image from different exposure images, and is based on the method

22

described in [9].

Similarly, we obtain the HDR RGB image I_{HDRRGB} from the RGB image I_{RGB} and the low exposure RGB image that is obtained by multiplying the secondary filter images by a color reproduction matrix $\mathbf{T}''$ and color space conversion $\mathbf{H}$:

$$I_{HDRRGB} = \tau(\mathbf{Z}_{HDRRGB}\mathbf{A}), \tag{29}$$

where $\mathbf{Z}_{HDRRGB} = \begin{bmatrix} \mathbf{HT}' & \mathbf{0} \\ \mathbf{0} & \mathbf{HT}'' \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$.

5.4.4 Reconstruction of Multispectral Image

For multispectral imaging, $\hat{x}_a, \hat{x}_b, \hat{x}_c$ and the anti-aliased $\hat{x}_d, \hat{x}_e, \hat{x}_f, \hat{x}_g$ images are used to reconstruct the spectral reflectance distribution of an object using the method given by Eq.9:

$$\mathbf{X} = \mathbf{Z}_{MS}\mathbf{A}, \tag{30}$$
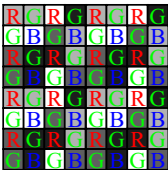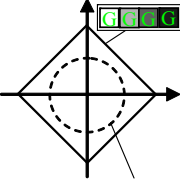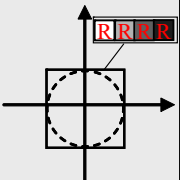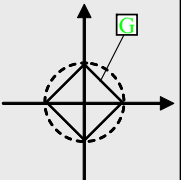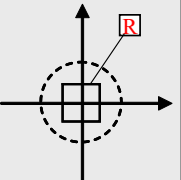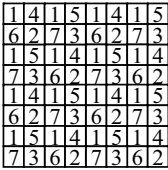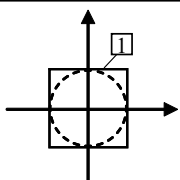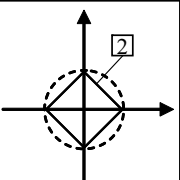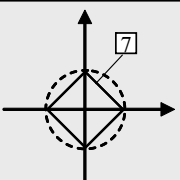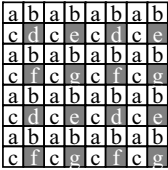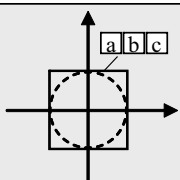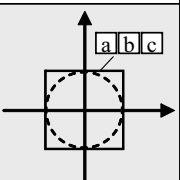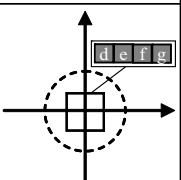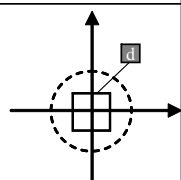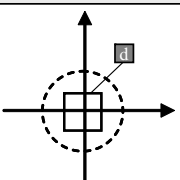
where $\mathbf{Z}_{MS} = \begin{bmatrix} \mathbf{I}_{7\times7} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$, and $\mathbf{I}_{7\times7}$ is 7 by 7 identity matrix.

6 Comparison with other Mosaics

The performance of demosaicing methods depends on the spatial layout and spectral responses of the color filters used, both of which vary from detector to detector. Moreover, previous CFA mosaics [8][2] were not designed for controlling the trade-off of spatial resolution to generate a variety of images. Therefore, a direct comparison of image qualities is difficult to perform. Instead, Table2 shows a qualitative comparison between the performances of our GAP mosaic, a previously proposed assorted pixel CFA [8] and a

Table 2: Comparison of the GAP camera with previous assorted pixels [8] and multispectral CFA [2]. The shaded CFA offers the best quality image for each image type.

Mosaic Pattern	Monochrome	RGB	HDR Monochrome	HDR RGB	Multispectral
Assorted Pixel	GGGG, Optical Resolution Limit	RRRR	G	R	N/A
Multispectral CFA	1	2	N/A	N/A	7
GAP Camera	a b c	a b c	d e f g	d	d

previously proposed multispectral CFA [2]. Note that when the difference of exposures for HDR imaging is disregarded in the case of assorted pixels, it is identical to the Bayer mosaic [5]. For monochrome images, although there is fundamentally no spatial resolution difference between the three mosaics, the monochrome image of the GAP mosaic is reconstructed from the 'a', 'b', and 'c' channels, which together cover all visible wavelengths (See Figure 4). Therefore, the GAP mosaic can reproduce monochrome images more accurately than other CFAs. The GAP mosaic can also create HDR RGB images of the same resolution as the assorted pixel array. However, because the assorted pixel array has 4 different exposures, it is more effective at extending dynamic range than the GAP camera.

In summary, when high spatial resolution is necessary, the GAP camera offers images with quality that is similar to, or better than, other CFA mosaics.
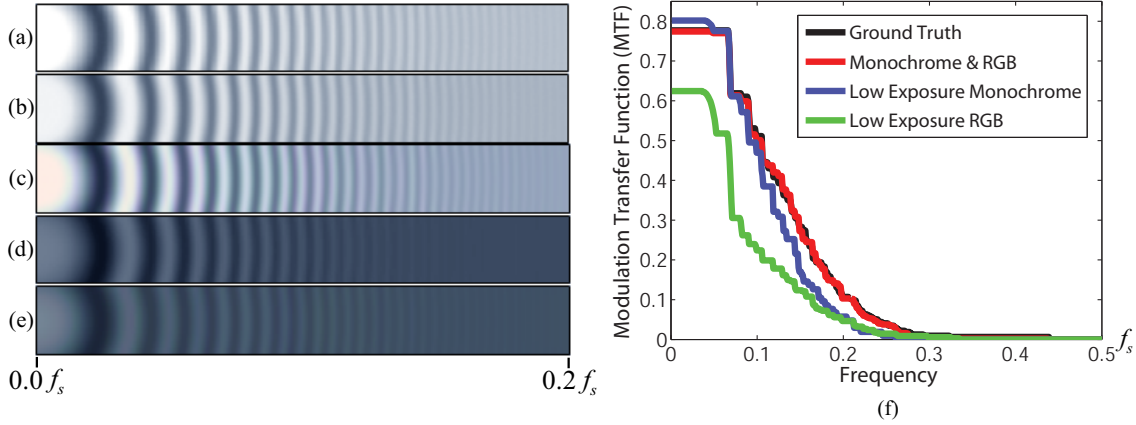
Figure 8: The result of multimodal demosaicing for a CZP chart. (a) Ground truth (simulated with $N = f/5.6$, $p = 1.0\mu m$). (b) Demosaiced monochrome image. (c) Demosaiced RGB. (d) Demosaiced (anti-aliased) low-exposure monochrome image. (e) Demosaiced (anti-aliased) low exposure RGB image. (f) MTFs of ground truth and demosaiced CZP images.
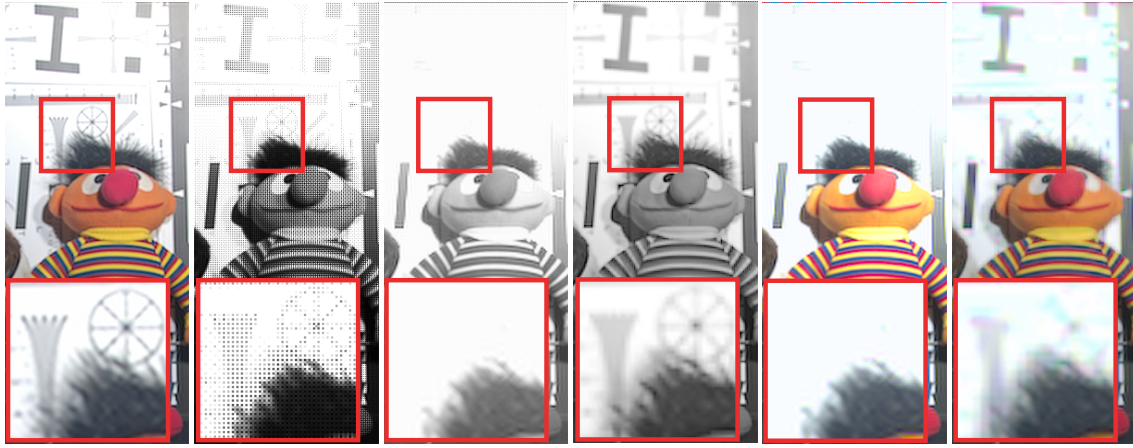
7 Experimental Results

7.1 Results for the CZP Chart

Figure 8(a) shows a synthesized Circular Zone Plate (CZP) image computed using a diffraction-limited model of a lens with an f-number of 5.6 and $1.0\mu m$ pixel size (without considering noise). This serves as the ground truth. Figures 8(b-e) show demosaiced images computed from a GAP mosaic image – (b) monochrome, (c) RGB, (d) low exposure monochrome, and (e) low exposure RGB. Figure 8(f) shows MTFs of these demosaiced images. The monochrome and RGB images computed using the primary filters are very close to the ground truth. The low exposure monochrome image has an MTF of 0.1 at $0.1754f_s$, while the low exposure RGB image's MTF is 0.1 at $0.1647f_s$. For standard monochrome and RGB images this occurs at $0.2125f_s$. This demonstrates that our GAP mosaic with multimodal demosaicing allows a user to control the trade-off between spatial resolution and radiometric details of the output image.

7.2 Experiments with Multispectral Images

We also captured 31-band multispectral images (400-700nm, at 10nm intervals) of several static scenes using a tunable filter (VariSpec^{TM} Liquid Crystal Tunable Filter) and a cooled CCD camera (Apogee Alta U260, 512×512 pixels). We have captured multispectral images for a wide variety of objects and materials, including, textiles, skin, hair, real and fake fruits and vegetables, candy, drinks, paints, etc. We believe this database could be valuable to researchers working in areas related to multispectral imaging. The database has been made publicly available at: `http://www1.cs.columbia.edu/CAVE/projects/gap_camera/`.
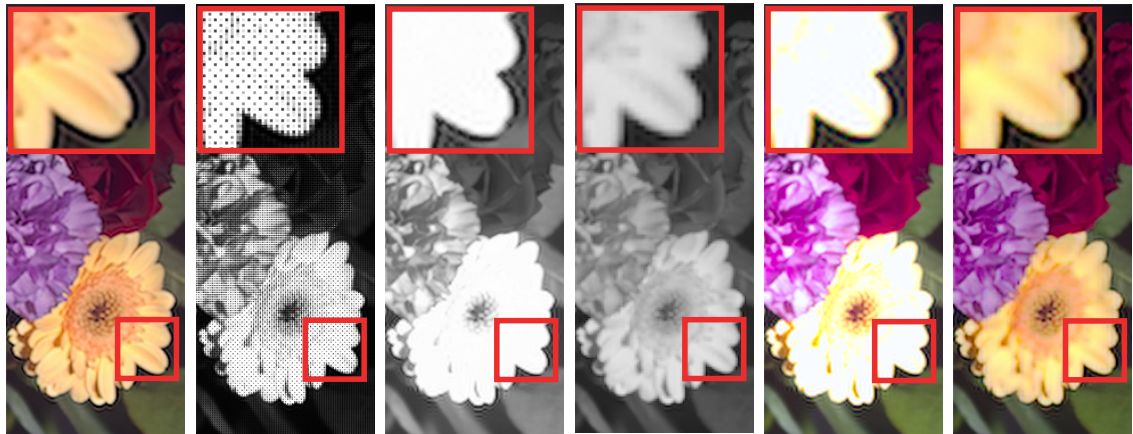
The multispectral images were used to simulate images captured with a GAP mosaic. Figure 9 and 10 shows these as well as our multimodal demosaicing results for two different scenes. For both scenes, the textures and colors of saturated regions in the monochrome and RGB images become visible in the corresponding HDR images. As expected, one can see more detail in the HDR monochrome images than in the HDR RGB images. We also experimented within skin detection using RGB image and multispectral data. Figure 10(h) shows the result of skin detection using an RGB image by using a simple correlation-based method. Figure 10(i) shows the result of skin detection applied to multispectral images computed from GAP images (See Figure 10(g)). Note that the scene shown in Figure 10 includes a real face (skin) on the right and a photo of the same face (printed paper) on the left. As seen in Figure 10(h), these two faces (real and fake) are difficult to distinguish using the RGB image – skin detection based on color analysis finds both the faces although only one of them is real. In contrast, skin detection applied to the multispectral image computed from the GAP image results in the desired result – only the real face is found as pixels within it have the spectrum of real skin (see Figure 10(i)).
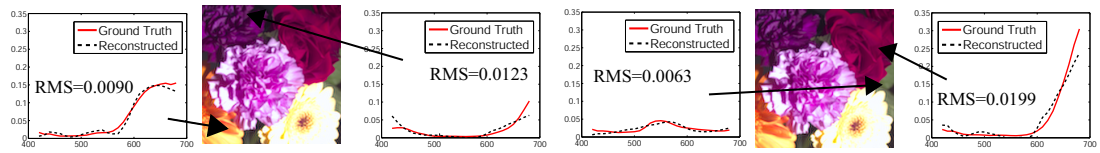
(a) Ground Truth   (b) GAP Mosaic   (c) Monochrome   (d) HDR Monochrome   (e) RGB   (f) HDR RGB

RMS=0.0110   RMS=0.0096   RMS=0.0069   RMS=0.0122
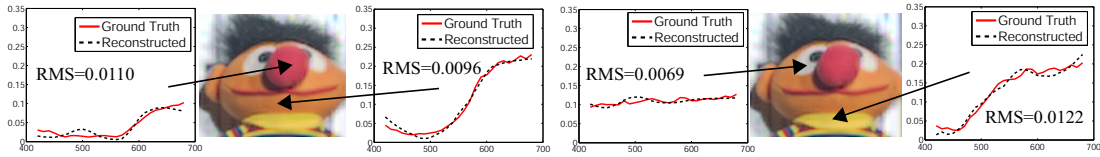
(g) Reconstructed Spectral Reflectance Distribution

(h) Ground Truth   (i) GAP Mosaic   (j) Monochrome   (k) HDR Monochrome   (l) RGB   (m) HDR RGB

RMS=0.0090   RMS=0.0123   RMS=0.0063   RMS=0.0199

(n) Reconstructed Spectral Reflectance Distribution

Figure 9: Results for real scenes. (a,h) Ground truths (simulated with $N = f/5.6$, $p = 1.0\mu m$). (b,i) GAP mosaic (raw) images. (c,j) demosaiced monochrome images. (d,k) HDR monochrome images. (e,l) RGB images. (f,m) HDR RGB images. (g,n) multispectral images and examples of reconstructed spectral reflectance distributions.
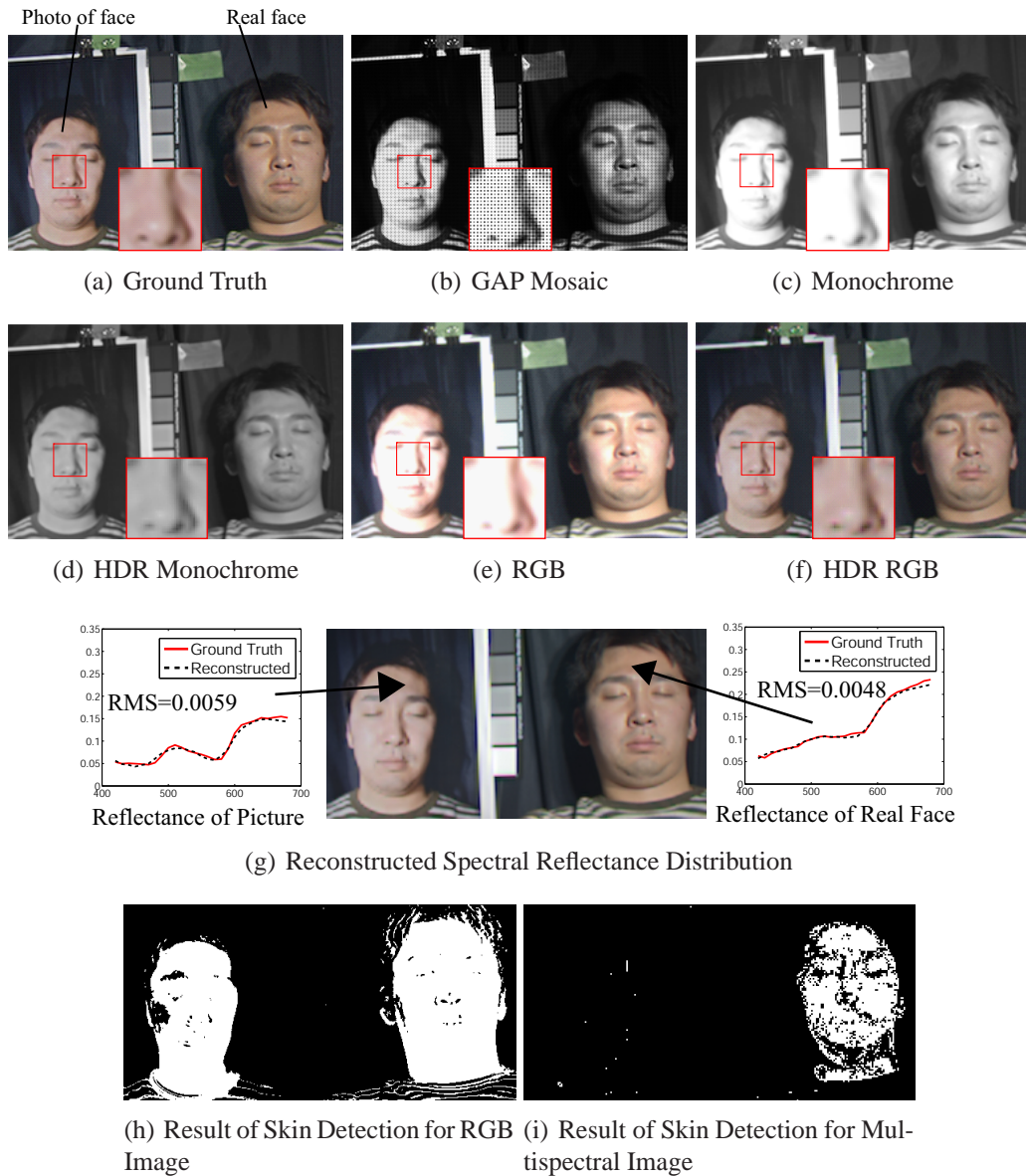
27

(a) Ground Truth

(b) GAP Mosaic

(c) Monochrome

(d) HDR Monochrome

(e) RGB

(f) HDR RGB

(g) Reconstructed Spectral Reflectance Distribution

(h) Result of Skin Detection for RGB Image

(i) Result of Skin Detection for Multispectral Image

Figure 10: Results for real scenes. (a) Ground truth (simulated with $N = f/5.6$, $p = 1.0\mu m$). (b) GAP mosaic (raw) image. (c) demosaiced monochrome image. (d) HDR monochrome image. (e) RGB image. (f) HDR RGB image. (g) Multispectral image and reconstructed spectral reflectance distributions. (h) Result of skin detection for RGB Image by using simple correlation-based method. (i) Result of skin detection applied to a multispectral image computed from the GAP image.

28

8 Conclusion

We have presented the concept of a generalized assorted pixel camera. We have developed a general framework for designing GAP cameras that can simultaneously capture extended dynamic range and spectral resolution. We have also proposed a demosaicing algorithm that includes anti-aliasing for spectral channels that are under-sampled.

This algorithm has some limitations that we plan to investigate in future work. For instance, when the primary pixels are saturated, anti-aliasing cannot be applied to the secondary pixels and the spectral reflectance distribution cannot be fully reconstructed in the saturated regions. Since the fabrication of an image detector is very expensive, the results in this paper were generated using simulations done using real multispectral images. Encouraged by our results, we are currently pursuing the fabrication of a GAP detector chip.

References

[1] http://www.asahi-spectra.com/.

[2] G. A. Baone and H. Qi. Demosaicking methods for multispectral cameras using mosaic focal plane array technology. In *Proceedings of SPIE*, volume 6062, Jan 2006.

[3] K. Fife, A. E. Gamal, and H.-S. P. Wong. A 3mpixel multi-aperture image sensor with 0.7μm pixels in 0.11μ cmos. In *IEEE ISSCC Digest of Technical Papers*, Feb. 2008.

[4] G.C.Holst. *CCD Arrays, Cameras and Displays 2nd Ed*. JCD Publishing, 1998.

[5] B. K. Gunturk, J. Glotzbach, Y. Altunbasak, R. W. Schafer, and R. M. Mersereau. Demosaicking:color filter array interpolation. In *IEEE Signal Processing Magazine*,

2005.

[6] J. F. H. JR and J. T. Compton. Processing color and panchromatic pixels. *US Patent: 2007/0024879 A1*, Feb 2007.

[7] R. Lyon and P. Hubel. Eyeing the camera: Into the nextcentury. In *ISandT/TSID 10th Color Imaging ConferenceProceedings*, pages 349–355, 2002.

[8] S. G. Narasimhan and S. K. Nayar. Enhancing resolution along multiple imaging dimensions using assorted pixels. In *IEEE Transactions on PAMI*, volume 27, Apr. 2005.

[9] S. Nayar and T. Mitsunaga. High Dynamic Range Imaging: Spatially Varying Pixel Exposures. In *IEEE Conference on CVPR*, volume 1, pages 472–479, Jun 2000.

[10] J. I. Park, M. H. Lee, M. Crossberg, and S. K. Nayar. Multispectral imaging using multiplexed illumination. In *Proceedings of IEEE ICCV*, Oct 2007.

[11] J. Parkkinen, J. Hallikainen, and T. Jaaskelainen. Characteristic spectra of munsell colors. In *Journal of the Optical Society of America*, volume 6, Feb 1989.

[12] S. Quan, N. Ohta, R. Berns, and X. Jiang. Unified measure of goodness and optimal design of spectral sensitivity functions. In *J. of Imaging Science and Technology*, 2002.

[13] G. Sharma and H. Trussell. Figures of merit for color scanners. In *IEEE Trans. Image Processing*, volume 6, July 1997.

[14] R. Shogenji, Y. Kitamura, K. Yamada, S. Miyatake, and J. Tanida. Multispectral imaging using compact compound optics. In *Optics Express*, pages 1643–1655, 2004.