A PERCEPTUALLY BASED DESIGN METHODOLOGY FOR COLOR FILTER ARRAYS

Manu Parmar, Stanley J. Reeves

Department of Electrical and Computer Engineering
Auburn University
Auburn, AL-36849

ABSTRACT

We consider the problem of selection of samples during image acquisition in a color digital camera. Due to the high cost of imaging sensors, digital cameras do not ordinarily employ separate sensors for each color channel. Instead, a single sensor overlaid with a mosaic of color filters is used such that a single color channel is sampled at each photosite. The full-color image must be reconstructed before the image may be displayed. In this paper, we propose a perceptually based sample selection scheme for color filter arrays. Regularization techniques are used to formulate an error criterion that reflects the error between the original image and the reconstructed image when viewed through the human visual system. A sequential selection algorithm is used to select sample locations that minimize the error criterion. The resulting array is found to compare favorably with existing array patterns in terms of the error criterion.

1. INTRODUCTION

To be able to display a color image in print or on a display device, we need full information about at least three color primaries at each pixel location. Ideally a color digital camera would sample all color channels from an imaged scene at each pixel. In practice, in the interests of cost and size, typical digital color cameras only sample one color sample at a particular pixel. A *mosaic* of color filters is overlaid on the imaging sensor to achieve sparse sampling. Missing samples are reconstructed in a post-processing step commonly referred to as demosaicking. In this paper we propose a method of sample selection based on the nature of the human visual system.

Many different sampling configurations have been proposed by researchers and applied commercially. The first such array pattern was disclosed by Bayer in 1975 [1]. The Bayer pattern (Fig. 1) is the most commonly used color filter array (CFA) pattern, and most reconstruction algorithms proposed in the literature assume Bayer sampling. The pattern described by Bayer samples the additive primaries red, green, and blue and is an RGB CFA. The array is configured such that the green channel is sampled at every other pixel. The green samples are staggered by one pixel in adjacent rows, and blue and red channels are used in alternate rows to complete the *mosaic*. In [2], the inventors describe an array pattern that addresses the problem of saturation in the green channel of the Bayer array. This array has alternating green and luminance samples in a row adjacent to alternating red, luminance and blue samples, and improves color reproduction at the cost of spatial resolution. The inventors in [3] address saturation issues and propose an array that contains luminance samples in addition to the color primaries red, green, and blue. Hamilton et al. in [4] propose a CFA that uses the subtractive primaries cyan, magenta, and yellow in addition to green filters to address the issue of photon acquisition in low-light conditions. The authors in [5] propose a CFA pattern based on random arrays derived from a blue noise pattern in an attempt to reduce aliasing artifacts that result from high-frequency periodic patterns in the image.

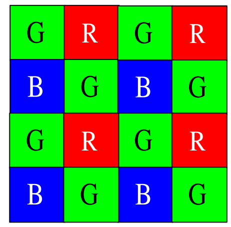

Fig. 1. The Bayer Array

In this paper, we propose a design methodology for selecting color samples in an RGB CFA. We use a simple model of the human visual system to characterize the perceptual error in an image reconstructed from a sub-sampled CFA. A sequential algorithm is used to select samples that minimize an error criterion that incorporates the effect of the human visual model.

2. MATHEMATICAL MODEL

We model the sub-sampled image as a linear transformation that maps the full-color image to an image that contains only one color value at a particular pixel location. The sub-sampled image is represented as

$$y_i = A_i x_i + u_i, \qquad i = \text{Red}, \text{Green}, \text{Blue}, \tag{1}$$

where x_i, $(mn \times 1)$ and y_i, $(mn \times 1)$ are the Red, Green and Blue channels of the original and the sub-sampled $m \times n$ images arranged in a column-ordered form, and u_i, $(mn \times 1)$, are the similarly arranged noise terms. The matrices A_i are the *sampling matrices*. For the fully-sampled case, A_i are identical to the $mn \times mn$ identity matrix. For the sub-sampled case, the matrices A_i contain only the rows corresponding to a sampled pixel location. We assume that the image and noise are uncorrelated.

We form a regularization functional for each channel that contains an energy bound on the residual $A_i x - y_i$ and a penalty on the roughness as:

$$\Phi_i = \|A_i x_i - y_i\|_2^2 + \mu_i L_i x_i{}^2. \tag{2}$$

The estimate of x_i found on minimizing the constrained least squares problem in (2) is

$$\hat{x}_i = (A_i^H A_i + \mu_i L_i^H L_i)^{-1} A_i^H y_i, \tag{3}$$

where A^H is the Hermitian transpose of A.

The appearance of the reconstructed image depends fundamentally on the characteristics of the human visual system (HVS). Color and spatial variances in the image are processed by the HVS to give the perceived image. This motivates the use of an HVS model to evaluate the performance of color image reconstruction. The next section elaborates on the HVS model used in this treatment.

To obtain the best estimate for the perceived image, we minimize the discrepancy in the reconstructed image when viewed through the HVS. Let the matrices H_i, i = Red, Green, Blue, represent the filtering effect corresponding to the point spread functions (PSFs) of the red, green and blue channels of the HVS respectively. We form a discrepancy function for one channel (dropping the subscript) as

$$d = E\{\|Hx - H\hat{x}\|_2^2\}, \tag{4}$$

where $\mathrm{E}\{.\}$ represents Expectation, and $\|.\|_2$ denotes the Frobenius matrix norm.

$$\begin{aligned} d &= \mathrm{E}\left\{\|Hx - H(A^H A + \mu L^H L)^{-1} A^H A x\|_2^2\right\} \\ &\quad + \mathrm{E}\left\{\|H(A^H A + \mu L^H L)^{-1} A^H n\|_2^2\right\} \\ &= \mathrm{E}\left\{\|H(A^H A + \mu L^H L)^{-1} \mu L^H L x\|_2^2\right\} \\ &\quad + \mathrm{E}\left\{\|H(A^H A + \mu L^H L)^{-1} A^H n\|_2^2\right\}. \end{aligned} \tag{5}$$

Let $P = (A^H A + \mu L^H L)$, such that

$$d = \mathrm{E}\left\{\|HP^{-1}\mu L^H L x\|_2^2\right\} + \mathrm{E}\left\{\|HP^{-1}A^H n\|_2^2\right\}. \tag{6}$$

Now, $E\{\|HP^{-1}A^H n\|_2^2\}$

$$\begin{aligned} &= \mathrm{E}\left\{\mathrm{tr}\left(n^H A P^{-H} H^H H P^{-1} A^H n\right)\right\} \\ &= \mathrm{tr}\left(\mathrm{E}\left\{A P^{-H} H^H H P^{-1} A^H n n^H\right\}\right) \\ &= \mathrm{tr}\left(A P^{-H} H^H H P^{-1} A^H R_n\right), \end{aligned} \tag{7}$$

where R_n is the correlation matrix for n and is described by the relation $R_n = \mathrm{E}\left\{nn^H\right\}$. We assume that the noise is independent, identically distributed such that $R_n = \mu I$. Also, P is symmetric and $P^H = P$. Thus, Eq. (7) reduces to

$$\mathrm{E}\left\{\|HP^{-1}A^H n\|_2^2\right\} = \mu\,\mathrm{tr}\left(AP^{-1}H^H H P^{-1} A^H\right). \tag{8}$$

Also, $\mathrm{E}\left\{\|HP^{-1}\mu L^H L x\|_2^2\right\}$

$$\begin{aligned} &= \mathrm{E}\left\{\mathrm{tr}\left(x^H \mu L^H L P^{-1} H^H H P^{-1} \mu L^H L x\right)\right\} \\ &= \mu^2\,\mathrm{tr}\left(\mathrm{E}\left\{L^H L P^{-1} H^H H P^{-1} L^H L x x^H\right\}\right) \\ &= \mu^2\,\mathrm{tr}\left(L^H L P^{-1} H^H H P^{-1} L^H L R_x\right), \end{aligned} \tag{9}$$

where R_x is the correlation matrix for x and is described by the relation $R_x = \mathrm{E}\left\{xx^H\right\}$. From Eqs. (8) and (9), we have

$$d = \mu\,\mathrm{tr}\left(P^{-1}H^H H P^{-1}\left(A^H A + \mu L^H L R_x L^H L\right)\right). \tag{10}$$

For $L = R_x^{-\frac{1}{2}}$, $L^H L = R_x^{-1}$, and Eq. (10) reduces to

$$d = \mu\,\mathrm{tr}\left(P^{-1}H^H H\right). \tag{11}$$

We define an error function as a weighted sum of the channel discrepancy functions as

$$e = \sum_i \kappa_i d_i = \kappa_i \sum \mu_i\,\mathrm{tr}\left((A_i^H A_i + \mu_i R_{x_i}^{-1})^{-1} H_i^H H_i\right), \tag{12}$$

where κ_i are scaling factors that reflect the perceptual importance of the fidelity in a particular channel.

3. HUMAN COLOR VISION MODEL

The image processing flow for an image captured with a digital camera and viewed by an observer has multiple steps. To get an accurate description of the perceived image, the PSFs of the demosaicking process and the HVS must be known precisely. In this treatment, the authors use a rudimentary model for the PSFs of the three color channels based on a functional model of the low-contrast photopic modulation transfer function (MTF) of the HVS described by Sullivan et al. in [6]. We suppose that the MTF of the entire work-flow of the digital camera retains the dominating characteristics of the HVS in that:

1. it is more sensitive to spatial frequencies in the vertical and horizontal directions, and
2. the response of chrominance channels falls faster than the response of the luminance channel.

We also also assume that the green channel corresponds closely to the luminance response. Thus, this work does not rely on an exhaustive and precise human color vision model pertaining to the RGB space but uses a simple model as an example to propose for the problem of sample selection. The matrices H_i introduced in Section 2 represent the response of the entire chain.

The MTF of the green channel is obtained from the MTF as described by Sullivan as

$$V_{G_{ij}} = \begin{cases} a(b + c\bar{f}_{ij})\exp\left(-\left(c\bar{f}_{ij}\right)^d\right), & \text{if } \bar{f}_{ij} > f_{max} \\ 1.0, & \text{otherwise,} \end{cases} \tag{13}$$

where the constants a, b, c, and d are calculated from empirical data to be 2.2, 0.192, 0.114 and 1.1 respectively; $\bar{f}_{ij}$ is the radial spatial frequency in cycles/degree as subtended by the image on the human eye scaled for the viewing distance, and f_{max} is the frequency corresponding to the peak of V_{ij}. Since we need the MTF in terms of discrete linear frequencies along the vertical and horizontal directions (f_i, f_j), we must express (f_i, f_j) in terms of the radial frequency $\bar{f}_{ij}$.

The discrete frequencies along the horizontal and vertical directions depend on the pixel pitch Δ of the output device (print or display device) and the total number of frequencies M. A location

(i, j) in the frequency domain corresponds to the following f_i and f_j in cycles/mm:

$$f_i = \frac{i-1}{\Delta M},$$
$$f_j = \frac{j-1}{\Delta M}. \tag{14}$$

The linear frequencies are scaled for the viewing distance s and converted to radial frequency as

$$f_{ij} = \frac{\pi}{180 \arcsin\left(\frac{1}{\sqrt{1+s^2}}\right)} \sqrt{f_i^2 + f_j^2}. \tag{15}$$

The MTF is not uniform along all directions. The HVS is most sensitive to spatial variation along the horizontal and vertical directions. To account for this variation, the MTF is normalized by an angle dependent function $s(\theta_{ij})$ such that

$$\bar{f}_{ij} = \frac{f_{ij}}{s(\theta_{ij})}, \tag{16}$$

where

$$s(\theta_{ij}) = \frac{1-w}{2}\cos(4\theta_{ij}) + \frac{1+w}{2}, \tag{17}$$

with w being a symmetry parameter and

$$\theta_{ij} = \arctan\left(\frac{f_j}{f_i}\right). \tag{18}$$

The response obtained for the green channel for $w = 0.7$, and a viewing distance of 45 cm and a pixel pitch of 0.27 mm is shown in Fig. 2.

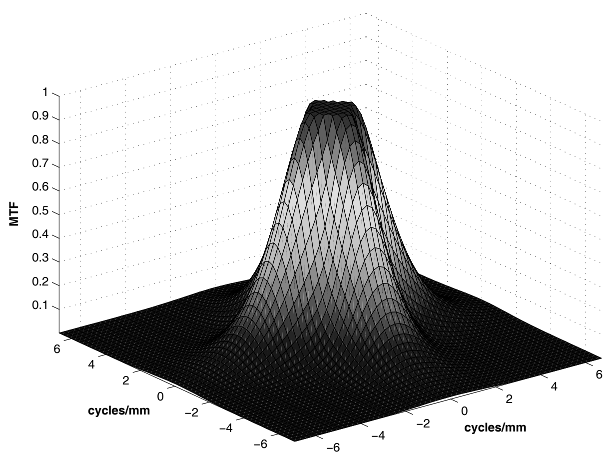

Fig. 2. HVS green channel MTF

The response of the HVS to chrominance, or the contrast sensitivity to spatial variations in the chrominance channels, falls off faster than the response to the luminance channel. A simple chrominance response model corresponding to a decaying exponential is chosen as a basis for the HVS response to the blue and red channels. The red and blue channel response is modelled as

$$V_{B,R}(f_{ij}) = e^{(-0.15 f_{ij})}, \tag{19}$$

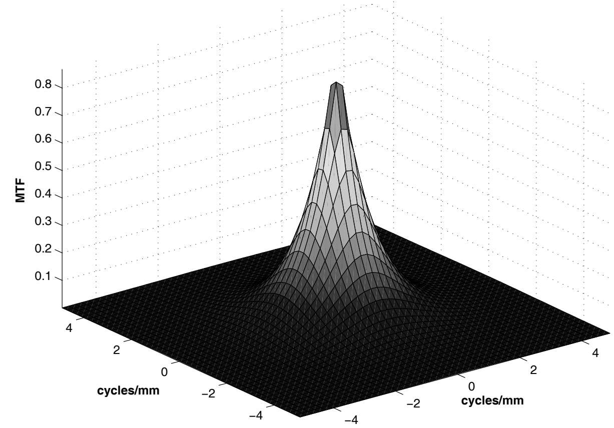

Fig. 3. HVS red and blue channel MTFs

The response obtained for the red and blue channels is shown in Fig. 3.

The HVS point spread functions h_i for i = Red, Green, Blue are obtained as

$$h_G = \mathscr{F}^{-1}\{V_G(i,j)\},$$
$$h_{R,B} = \mathscr{F}^{-1}\{V_{R,B}(i,j)\}. \tag{20}$$

The matrices H_i are constructed from h_i such that multiplication of a column-ordered image by H_i yields the 2-D convolution of the image by the point spread function h_i.

4. SAMPLING STRATEGY

The goal is to sample only one color channel at each sample location. Thus, we have to select mn samples from a set of $3mn$ samples. The error criterion defined in (12) may be used to optimize the selection procedure. The criterion does not depend on the scene being imaged and may be used for sub-sampling a general scene if the statistical properties (R_x and R_n) of the fully sampled image are defined accurately.

Each row in the matrices A_i in (12) corresponds to a sample in the respective channel. The error criterion defined in (12) may be used to obtain the row that when eliminated would cause the least error in the reconstructed signal when viewed through the HVS. An exhaustive optimization would require the computation of the error criterion for all combinations of eliminated rows, and would require $\frac{(3mn)!}{(2mn)!\,(mn)!}$ computations of the error criterion. For a reasonably sized array, this computation would require immense resources.

The authors in [7] use a greedy algorithm for sequential backward selection (SBS) of samples for signal reconstruction. The sequential backward selection algorithm can not be guaranteed to provide optimal results, but the authors in [8] have shown that the algorithm consistently provides good results with a relatively tight upper bound on the error criterion. We devise an SBS scheme for optimizing the criterion as follows. We start with a fully sampled image with all mn samples in each channel. The error criterion is computed after eliminating one row from one of the matrices A_i, and the row that gives the least value for the criterion is eliminated.

In the next step, The matrix A_i from which the row is eliminated is of dimension $(m-1) \times n$. The error criterion is computed again after eliminating one row from A_i, and rows of A_i are successively eliminated with the constraint that the three channels are sampled in a mutually exclusive manner.

Computation of the error criterion requires the computation of the inverse of the matrix P for each eliminated row. For an $m \times n$ array, P is of dimension $mn \times mn$, and the inversion requires considerable computation even for small arrays. The error criterion may be simplified using the Sherman-Morrison matrix inversion formula such that we need find only an update term after each elimination. Also, the matrices H_i are circulant block-circulant and the matrix products involving H_i may be computed using DFTs. In spite of these simplifications, the computation of the criterion is cumbersome since in the form of (12), it requires the storage of at least the three $mn \times mn$ initial matrices P_i^{-1}.

5. EXPERIMENTS

The power spectral density of a random process is given by the Wiener-Khinchine relation, $S_x(j\omega) = \mathscr{F}\{R_x\}$. We obtained an R_x representative of a general scene imaged by a digital camera from the mean, S_{avg}, of the power spectrums of a large number of images reflecting various image types as $R_x = \mathscr{F}^{-1}\{S_{avg}\}$. The images used to obtain S_{avg} span a wide range of categories including natural scenes, landscapes, portraits, and a few color test images obtained from the USC-SIPI image database available at *http://sipi.usc.edu/services/database/Database.html*.

The sample selection procedure detailed in Section 4 was applied for fully-sampled RGB arrays of different sizes. The error criterion values obtained for a Bayer array (e_{Bayer}) and an array obtained by the SBS scheme (e_{SBS}) detailed in Sec. 4 are shown in Table 1. The weights on the individual channel errors are $\kappa_{Red} = 1$, $\kappa_{Green} = 1.6$, and $\kappa_{Blue} = 1$. The values of κ_i reflect the relative importance of the Green channel on image quality and precise values may be obtained through psychovisual experiments. An 8×8 array obtained using SBS is shown in Fig. 4.

Table 1. Comparison of error criterion values with a Bayer array

Array size	e_{Bayer}	e_{SBS}
8×8	28.8083	27.5952
12×12	46.0583	44.3362
16×16	74.9760	72.3530
32×32	218.4921	211.1279

6. CONCLUSIONS

In this paper we have devised a perceptually based sample selection procedure for CFAs. We have derived an error criterion that assumes reconstruction of the sub-sampled array via regularization. The criterion incorporates a simple model of the HVS color response to characterize the effect of the HVS in deciding the perceived quality of the reconstructed image. A sequential algorithm was proposed that optimizes the above criterion in the process of sample selection. Experimental results indicate that this procedure effectively selects sampling arrangements that give lower values of error as compared to the Bayer array. The methodology described in this paper is expected to yield better results in terms of

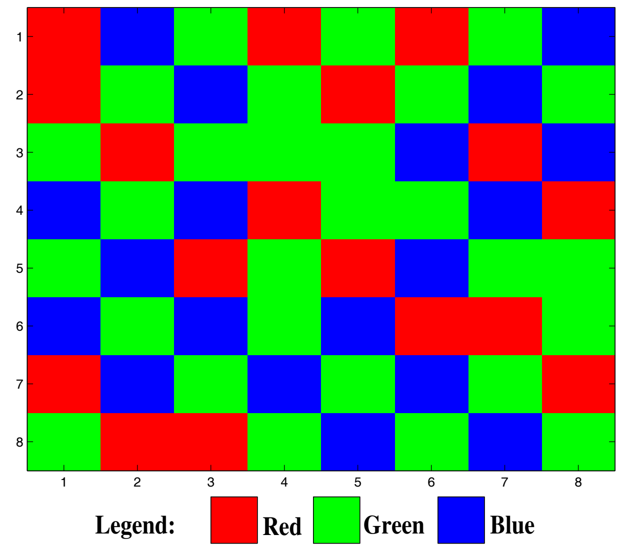

Fig. 4. An 8×8 array

the MSE in a perceptually uniform color space between a reconstructed sub-sampled image and a fully-sampled image if a more accurate characterization of the HVS response to color were used. Further research will be directed in this area.

7. REFERENCES

[1] B.E. Bayer, “Color imaging array,” U.S. Patent 3971065, July 1976.

[2] T. Yamagami, T. Sasaki, and A. Suga, “Image signal processing apparatus having a color filter with offset luminance filter elements,” U.S. Patent 5323233, June 1994.

[3] E.B. Gindele and A.C. Gallagher, “Sparsely sampled image sensing device with color and luminance photosites,” U.S. Patent 6476865 B1, Nov. 2002.

[4] J.F. Hamilton, J.E. Adams, and D.M. Orlicki, “Particular pattern of pixels for a color filter array which is used to derive luminanance and chrominance values,” U.S. Patent 6330029 B1, Dec. 2001.

[5] Wei Zhu, Kevin Parker, and Michael A. Kriss, “Color filter arrays based on mutually exclusive blue noise patterns,” *Journal of Visual Communication and Image Representation*, vol. 10, pp. 245–247, 1999.

[6] J. Sullivan, L. Ray, and R. Miller, “Design of minimum visual modlation halftone patterns,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 1, pp. 33–38, 1991.

[7] S. J. Reeves and L. P. Heck, “Selection of observations in signal reconstruction,” *IEEE Transactions on Signal Processing*, vol. 43, no. 3, pp. 788–791, March 1995.

[8] S. J. Reeves and Z. Zhao, “Sequential algorithms for observation selection,” *IEEE Transactions on Signal Processing*, vol. 47, no. 1, pp. 123–132, January 1999.