



# Color Image Processing Pipeline

A general survey of digital still camera processing

**D**igital still color cameras (DSCs) have gained significant popularity in recent years, with projected sales in the order of 44 million units by the year 2005. Such an explosive demand calls for an understanding of the processing involved and the implementation issues, bearing in mind the otherwise difficult problems these cameras solve. This article presents an overview of the image processing pipeline, first from a signal processing perspective and later from an implementation perspective, along with the tradeoffs involved.

## IMAGE FORMATION

A good starting point to fully comprehend the signal processing performed in a DSC is to consider the steps by which images are formed and how each stage affects the final rendered image. There are two distinct aspects of image formation: one that has a colorimetric perspective and another that has a generic imaging perspective, and we treat these separately.

In a vector space model for color systems, a reflectance spectrum  $r(\lambda)$  sampled uniformly in a spectral range  $[\lambda_{\min}, \lambda_{\max}]$  interacts with the illuminant spectrum  $L(\lambda)$  to form a projection onto the color space of the camera  $RGB_c$  as follows:

$$\mathbf{c} = \mathcal{N}(\mathbf{S}^T \mathbf{L} \mathbf{M} \mathbf{r} + \mathbf{n}), \quad (1)$$

where  $\mathbf{S}$  is a matrix formed by stacking the spectral sensitivities of the  $K$  color filters used in the imaging system column-wise,  $\mathbf{L}$  is a diagonal matrix with samples of the illuminant spectrum along its diagonal,  $\mathbf{M}$  is another diagonal matrix with samples of the relative spectral sensitivity of the charge-coupled device (CCD) sensor,  $\mathbf{r}$  is a vector corresponding to the rela-

tive surface spectral reflectance of the object, and  $n$  is an additive noise term (in fact, noise may even be multiplicative and signal dependent, making matters mathematically much more complex) [1], [2].  $\mathcal{N}$  corresponds to the nonlinearity introduced in the system. A similar algebraic setting exists for color formation in the human eye, producing tristimulus values where  $A$  is used to denote the equivalent of matrix  $S$ , and  $t$  denotes the tristimulus values in the CIEXYZ space [3]. Let us denote by  $f$  an image with  $M$  rows,  $N$  columns, and  $K$  spectral bands.

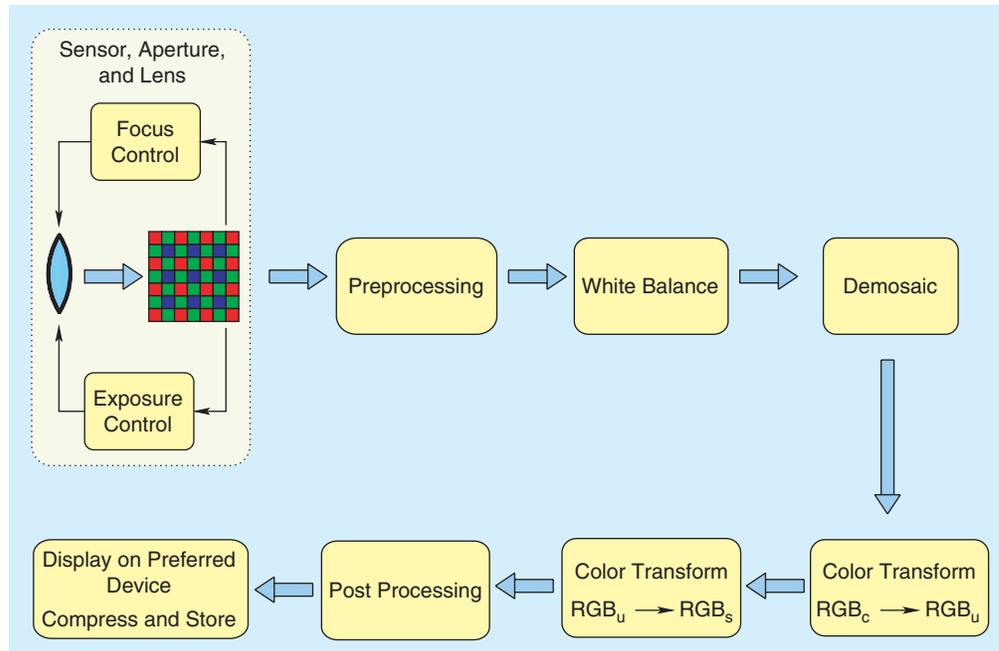
DSCs typically use  $K = 3$ , although one may conceive of a sensor with a more spectral bands. In some recent cameras, four sensors are used: red, green, blue, and emerald or cyan, magenta, yellow, and green. (Depending on the choice of the spectral sensitivities, the colors captured form color gamut of different sizes; for DSCs, it is typically important to capture skin tones with reasonable accuracy.) The image may also be considered as a two-dimensional (2-D) array with vector-valued pixels. Each vector-valued pixel is formed according to the model in (1), with values determined by the reflectance and illumination at the three-dimensional (3-D) world point indexed by 2-D camera pixel position. The image formed is then further modeled as follows:

$$\mathbf{g} = \mathcal{B}\{\mathbf{H}\mathbf{f}\}, \quad (2)$$

where  $\mathcal{B}$  is a color filter array (CFA) sampling operator,  $\mathbf{H}$  is the point spread function (a blur) corresponding to the optical system, and  $\mathbf{f}$  is a lexical representation of the full-color image in which each pixel is formed according to (1).

Although there is an overlap between color processing problems for other devices, such as scanners and printers, working with DSCs is complicated by problems stemming from the manner in which the input image is captured; the spatial variation in the scene lighting, the nonfixed scene geometry (location and orientation of light source, camera, and surfaces in the scene), varying scene illuminants (including combination of different light sources in the same scene), or the use of a CFA to obtain one color sample at a sensor location, to name a few.

Further resources on DSC issues include an illustrative overview of some of the processing steps involved in a DSC by Adams et al. [4]. A recent book chapter by Parulski and



[FIG1] Image processing involved in a DSC.

Spaulding details some of these steps [5]. An excellent resource for DSC processing is a book chapter by Holm et al. [6]. The multitude of problems that still remain unsolved is a fascinating source of research in the community.

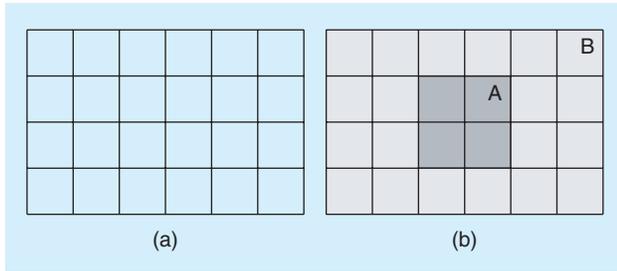
#### PIPELINE

The signal flowchart shown in Figure 1 briefly summarizes DSC processing. It should be noted that the sequence of operations differs from manufacturer to manufacturer.

Each of these blocks may be fine-tuned to achieve better systemic performance [7], e.g., introducing a small amount of blur using the lens system increases the correlation between neighboring pixels, which in turn may be used in the demosaicking step. Let us now consider each block in Figure 1.

#### SENSOR, APERTURE AND LENS

Although there is a need to measure three (or more) bands at each pixel location, this requires the use of more than one sensor and, consequently, drives up the cost of the camera. As a cheaper and more robust solution, manufacturers place a CFA on top of the sensor element. Of the many CFA patterns available, the Bayer array is by far the most popular [8]. Control mechanisms interact with the sensor (shown in Figure 1 as a red-green-blue checkered pattern, the CFA), to determine the exposure (aperture size, shutter speed, and automatic gain control) and focal position of the lens. These parameters need to be determined dynamically based on scene content. It is conventional to include an infrared blocking filter called a “hot mirror” (as it reflects infrared energy) along with the lens system, as most of the filters that are used in CFAs are sensitive in the near-infrared part of the spectrum, as is the silicon substrate used in the sensor.



**[FIG2]** Illustration of the division of the image into various blocks over which the luminance signal is measured for exposure control. (a) The image is divided into 24 blocks. (b) Sample partitioning of the scene for contrast measurement.

Exposure control usually requires characterization of the brightness (or intensity) of the image: an over- or underexposed image will greatly affect output colors. Depending on the measured energy in the sensor, the exposure control system changes the aperture size and/or the shutter speed along with a carefully calibrated automatic gain controller to capture well-exposed images. Both the exposure and focus controls may be based on either the actual luminance component derived from the complete RGB image or simply the green channel data, which is a good estimate of the luminance signal.

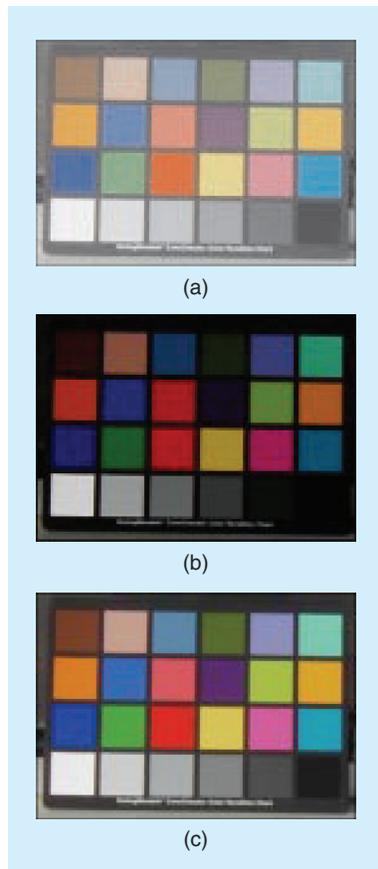
The image is divided into blocks [9], [10] as shown in Figure 2(a). The average luminance signal is measured in each one of these blocks that later is combined to form a measure of exposure based on the type of scene being imaged: backlit or frontlit scene or a nature shot. In a typical image, the average luminance signal is measured and is compared to a reference level, and the amount of exposure is controlled to maintain a constant scene luminance. Backlit or frontlit scenes may be distinguished by measuring the difference between the average luminance signal in the blocks, as shown in Figure 2(b). If the image is excessively frontlit, the average energy in region A will be much higher than that in region B, and vice versa in the case of a backlit scene. The exposure is controlled so as to maintain the difference between the average signals in these two areas, an estimate of the object contrast. Figure 3(a)–(c) illustrates an underexposed, overexposed, and well-exposed image, respectively.

Outdoor images (and many indoor ones as well) taken with typical cameras suffer from the problem of limited dynamic range in the case of an excessively backlit or frontlit scene. Dynamic range refers to the

contrast ratio between the brightest pixel and the darkest pixel in an image. The human visual system (HVS) can adapt to about four orders of magnitude in contrast ratio, while the sRGB system and typical computer monitors and television sets have a dynamic range of about two orders of magnitude. This leads to spatial detail in darker areas becoming indistinguishable from black and spatial detail in bright areas become indistinguishable from white. To address this problem, researchers have used the approach of capturing multiple images of the same scene at varying exposure levels and combining them to obtain a fused image that represents the highlight (bright) and shadow (dark) regions of an image in reasonable detail [11]. A detailed discussion of this topic is beyond the scope of this article, however, we refer the interested reader to the appropriate references [12]–[14]. Another interesting approach to this problem is to squeeze in two sensors in the location of what was formerly one sensor element; each with a different sensitivity to light, capturing two images of different dynamic ranges, hence effectively increasing the net dynamic range of the camera. Commercial cameras are beginning to incorporate high dynamic range (HDR) imaging solutions into their systems, either in software through image processing routines or in hardware by modifying the actual sensor, to facilitate the capture of excessively front or backlit scenes.

Focus control may be performed by using one of two types of approaches, active approaches that typically use a pulsed beam of infrared light from a small source placed near the lens system (called an auto-focus assist lamp) to obtain an estimate of the distance to the object of interest or passive approaches that make use of the image formed in the camera to determine the best focus. Passive approaches may be further divided into two types, ones that analyze the spatial frequency content of the image and ones that use a phase detection technique to estimate the distance to the object.

Techniques that analyze the spatial frequency content of the image typically divide the image into various regions (much like in the case of exposure control), and the position of the lens is adjusted to maximize the high-spatial-frequency content in the region(s) of interest (region labeled A in the case of a portrait image). In other words, given that sharp edges are preferred over smooth edges, in cases when the object of interest is in the central region (A) the position of the lens is adjusted so as to maximize the energy of the image gradient in this region. A digital high-pass filter kernel is used to measure



**[FIG3]** Images of a Macbeth ColorChecker chart showing exposure levels: (a) an underexposed image, (b) an overexposed version of the same image, and (c) a well-exposed image. These images were taken using a manually controlled lens control.

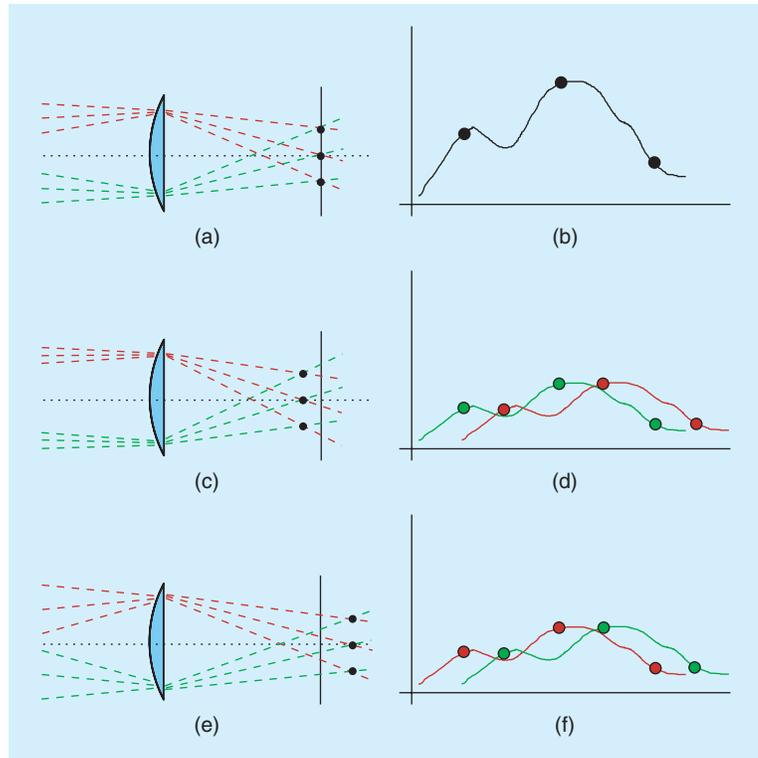
the resulting energy as a measure of focus. There are other “smart” measures of focus discussed in the open literature [15], [16] and as intellectual property [17]. Clearly, such techniques require that the scene have high contrast, which is not always the case. Also, such techniques have high computational demands on the camera.

Techniques that use phase detection utilize the phase difference between the energy measured from the two halves of the lens, much like in the case of a split-image rangefinder used in film-based, single-lens-reflex (SLR) cameras. The majority of this work is part of intellectual property owned by Honeywell [18]. This principle is easily explained through the use of ray diagrams, shown in Figure 4. Figure 4(a) shows a ray diagram when the lens is in good focus, and (b) shows the intensity profile corresponding to this lens position. When the object is moved farther away, the rays from the upper and lower halves of the lens no longer intersect at the same locations, and the measured energy from the two halves of the lenses are out-of-phase [Figure 4(b)–(c)] and requires the lens to move relative to the image plane to compensate for this defocus; in this case, towards the image plane. A similar argument may be made for the case when the object moves closer to the lens [Figure 4(e)–(f)].

Figure 5 shows one such image in two out-of-focus positions of the lens and one in-focus position. A focus measure is plotted in Figure 5(d). In-focus positions of the lens maximize the measure. Cheaper camera alternatives avoid this mechanism by using a fixed-focal-length lens.

We should note at this point that aliasing (due to the CFA sampling) causes highly objectionable artifacts in the output image. As a means of reducing these artifacts, most cameras use an antialiasing filter using one of two approaches: using a birefringent material that uses polarization to spread a beam of light over multiple pixels [19] or using phase delay techniques [4] that shift neighboring rays out of phase, consequently attenuating high-frequency content. This helps reduce Moiré patterns that may occur due to the sampling process involved, while aiding the demosaicking process by increasing the correlation between neighboring pixels.

The sensor element (CCD/CMOS sensor with a CFA) records the image by using photosensitive elements that convert light energy to electrical energy. The use of the CFA has the advantage that, although we are recording three color signals, we require only one sensor. Of course, this has the drawback that, at each sensor location, we measure only one color and the other two need to be estimated. This estimation process is called demosaicking. Another drawback with such a sampling process is that each channel of the image is usually sampled at one-half or one-quarter of the resolution of the sensor. In the Bayer array, the green channel is sampled at one-half, and the



**[FIG4]** Ray diagrams illustrating the basis for phase-detection techniques. (a) Case when the object is in good focus, with rays from the two halves of the lens intersecting on the image plane. (b) The corresponding intensities on the image plane. (c) The object is farther from the lens than in case (a) with the rays intersecting in front of the image plane. (d) The corresponding intensities on the image plane, showing the lower intensities of the two halves of the lenses separately. (e) The object is closer to the lens than in case (a) with the rays intersecting behind the image plane. (f) Corresponding intensities on the image plane with a condition similar to that in (d). In cases (c) and (e), the image appears blurred. In this figure, lines in red and green on the images on the right, respectively, denote the intensity from the upper and lower halves of the lens. Adapted from [18].

red and blue are sampled at one-quarter of the sensor resolution, which, after demosaicking, introduces artifacts. In such a scenario, the antialiasing filter is important, although there is a tradeoff with image sharpness.

In an exciting new development, the invention of entirely new sensor arrays [20] that more or less emulate the capture process of slide film to produce three color values at each and every pixel can eliminate the demosaicking step entirely and lead to crisper images.

Let us briefly look at the filters used in the sensor. Figure 6 shows a sample set of filter sensitivities. Notice that the filters have sensitivities that are nonzero, even outside what is typically accepted as the visible part of the spectrum (400–700 nm). This problem may be alleviated, as mentioned earlier, by using an IR-blocking filter in conjunction with the lens apparatus. Comparing the CIE standard observer sensitivities [21], [3, Figure 3] with the sensitivities of the CFA, we can anticipate that the color space of the camera is going to be very different from that of the human observer. Notice also that there is a remarkable amount of overlap in the spectral sensitivities of the

three channels in the camera, which, along with the blur, further reduces artifact problems associated with demosaicking.

### PREPROCESSING

The raw data obtained from the sensor needs to be processed before producing a full-color image to remove noise and other artifacts, along with a color processing pipeline to produce an accurate or, alternatively, a pleasing representation of the captured scene. One commonly used preprocessing step is defective pixel correction. It is possible that certain photo-elements in the sensor are defective and, if not corrected, show up as confetti-like errors after demosaicking. These missing or defective pixels are estimated by interpolating the accurately recorded data in their neighborhood.

### LINEARIZATION

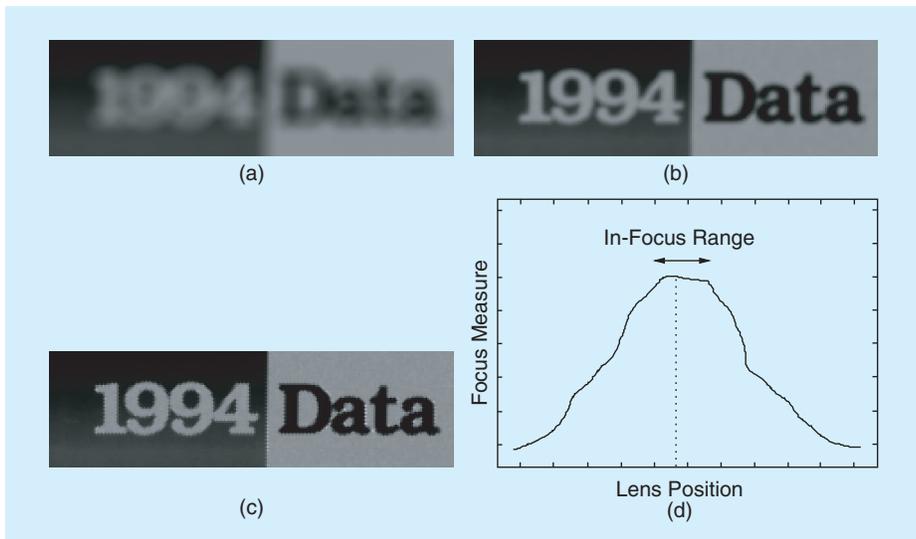
Some cameras require that the data be linearized since the captured data resides in a nonlinear space, mostly due to the electronics involved. However, most CCD sensors typically have a linear response. Cameras that include correction for nonlinear data use an opto-electronic conversion function (OECF) that relates the input nonlinear data to an output linear space. ISO standard 14,524 describes methods for measuring the OECF [22]. This correction transforms the raw measured data (typically with an 8-bit precision) into linear space (of higher bit precision, typically 12-bit). Relating to (1), the OECF is the inverse of the term denoted  $\mathcal{N}$ .

### DARK CURRENT COMPENSATION

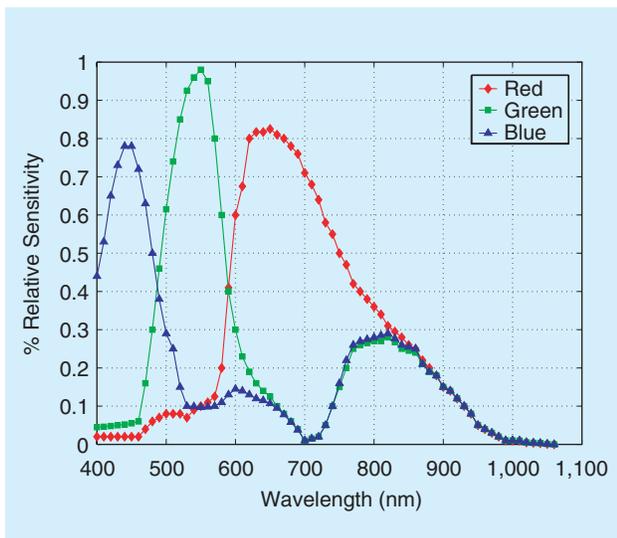
Even with the lens cap on, a dark current signal is recorded, which is due to thermally generated electrons in the sensor substrate. To account for this, two strategies are used; place an opaque mask along the edges of the sensor to give an estimate of intensity due to dark current alone, or capture a dark image for the given exposure time. In the first case, the mean dark current is subtracted from the entire image, and in the second, the dark image itself is subtracted from the captured data. It should be noted that dark current is a function of both exposure time and the ambient temperature. Dark current is one contributing component in the noise term,  $n$  in (1). It becomes increasingly important in low-exposure images, as the noise energy may be comparable to signal energy. Other contributing factors to the noise in the signal include “one-over-f” (flicker) and “reset” ( $kT/c$ ) noise that arise from the readout clock system and from thermal noise, respectively.

### FLARE COMPENSATION

Flare compensation is of significance in images where a bright source of light is in the field of view. Light entering the optics of the camera gets scattered and reflected, causing a nonuniform shift in the measured energy. Flare compensation techniques are mostly proprietary. A naive means of compensating for flare light is to subtract from the whole image a percentage of the mean measured signal energy in a channel. Another, more adaptive technique is to subtract a fixed percentage of the mean signal energy in the pixel’s neighborhood. The first technique, however, will make dark regions of the image darker blindly, and possibly negative, when in fact there would have been less flare in the darker regions in any case. Flare light is another contributing component in the noise term,  $n$  in (1).



[FIG5] Sample images showing focus control. (a) An out-of-focus image block at a certain lens position. (b) The same image with lens position closer to the required focal length. (c) An in-focus image block. (d) A plot of the focal measure versus the lens position. These are simulated images.



[FIG6] Sample spectral sensitivities of filters used in a digital color camera.

## WHITE BALANCE

The HVS has the ability to map “white” colors to the sensation of white, even though an object has different radiance when it is illuminated with different light sources. In other words, if you were shown a sheet of white paper under fluorescent lighting or under incandescent lighting or even under natural daylight, you would say that it was white, although the actual irradiated energy produces different colors for different illuminations. This phenomenon is called color constancy. While a pinkish-white light impinging on a white wall may be perceived as white by the human eye-brain system, white light reflected from a rose is not seen as white, this even though camera pixel values may be identical in the two cases. However, DSCs do not have the luxury of millions of years of evolution and hence need to be taught how to map white under the capture illuminant to white under the viewing illuminant (and other colors accordingly). All this needs to be done automatically, without knowledge about the capture illuminant. Under flash photography, however, this is simpler, as we know the illuminant relatively well. One means of performing white balance is to assume that a white patch must induce maximal camera responses in the three channels. In other words, if  $R$ ,  $G$ , and  $B$  denote the red, green, and blue channels of the image, the white-balanced image has signals given by  $R/R_{\max}$ ,  $G/G_{\max}$ , and  $B/B_{\max}$ . However, the maximum in the three channels is very often a poor estimate of the illuminant.

The next level of complication as an approach to color constancy (although an admittedly naive one) is the “gray world” assumption: assume all colors in an image will average out to gray,  $R = G = B$ . In other words, the channels are scaled based on the deviation of the image average from gray (keeping, say, the green channel gain, fixed). In this scheme, the white-balanced image has signals given by  $k_r R$ ,  $G$ , and  $k_b B$ , where  $k_r = G_{\text{mean}}/R_{\text{mean}}$  and  $k_b = G_{\text{mean}}/B_{\text{mean}}$ , the premise being that all off-balance neutral colors will get mapped to balanced neutrals.

This technique fails when most of the image is highly chromatic, e.g., a close-up image of a brightly colored flower. There are heuristics around this problem, though, that perform the scaling only for less chromatic colors. In [23], Kehtarnavaz et al. present a technique called “scoring” to perform automatic white balance. In [24], Barnard et al. present white balance techniques in an historical perspective and illustrate the

**OUTDOOR IMAGES TAKEN WITH TYPICAL CAMERAS SUFFER FROM THE PROBLEM OF LIMITED DYNAMIC RANGE IN THE CASE OF AN EXCESSIVELY BACKLIT OR FRONTLIT SCENE.**

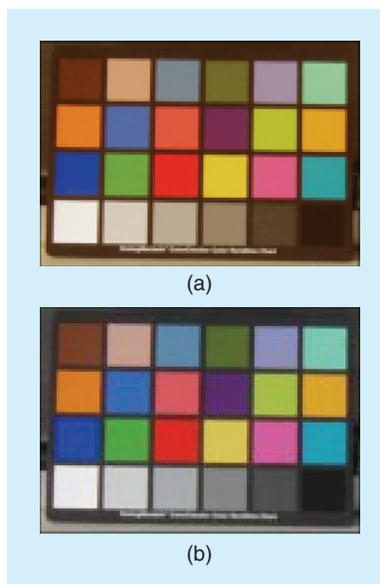
performance of the algorithms on synthetic data. It is not always the case that white balance must be done automatically. Some cameras, particularly the more expensive ones, put more emphasis on preprogrammed white balance or manual white balance than their automatic counterparts. Figure 7 shows a rendering of an image captured under an incandescent illuminant (with dominant energy near the red part of the spectrum). Notice that the white patch shows up with a yellow-red tint. Figure 7 shows the result of a gray-world assumption white balance performed to correct for this shift. More sophisticated techniques for illuminant estimation (e.g., [25]) are slowly entering the industry.

## DEMOSAICKING

Demosaicking is, by far, the most computationally intensive step in the processing pipeline. The techniques used are usually either proprietary or covered by patents. All demosaicking techniques make use of pixel neighborhood information to estimate the values of the pixel colors that were not measured. As one might imagine, this process will introduce artifacts that the remainder of the pipeline needs to remove. There are many techniques available for demosaicking, some of which are outlined in [26]. Popular techniques for demosaicking in open literature may be found in [27]–[31]. More details about demosaicking techniques may be found in a companion article by Gunturk et al. [32]. At this stage, the image is in a K-band form, residing in the color space of the camera.

## COLOR TRANSFORMATION—UNRENDERED COLOR SPACES

Unrendered color spaces are so called because they are not designed for an output medium but rather for a convenient storage or calculation medium. This advantage comes with the added drawback of storage requirements. Given that the spectral sensitivities of the camera are not identical to the human color matching functions, it is clear that cameras are not colorimetric. (The need to use colorimetric values lies in the fact that the color matching functions for the CIEXYZ system have been determined through careful visual experiments and are known to be a good estimate of the colors perceived by the HVS. However, recent advances have shown that the CIEXYZ space may be replaced with a scene-referred color space, such as the ISO-RGB color space [33] or the RIMM-RGB color space [34].) The data captured by the sensor is in the color space of the camera



**[FIG7]** (a) Image captured under an incandescent illuminant. (b) Same image after white balance using the gray-world assumption.

and has little to do with colorimetric (human) values. One means of obtaining colorimetric accuracy in the process of capturing an image is to transform the image from the sensor's spectral space to the CIEXYZ color space. In other words, given a camera measurement  $c$  at a pixel and the corresponding colorimetric value  $t$ , one may attempt to find the optimal color characterization transformation that maps the camera color space measurements to the CIEXYZ color space, i.e., find the (linear, for simplicity and efficiency of implementation) transformation  $B$  that minimizes the function

$$\begin{aligned} J &= \sum_{i=1}^N \|\mathbf{B}\mathcal{N}^{-1}\mathcal{N}(S^T\mathbf{L}\mathbf{M}\mathbf{r}_i + \mathbf{n}_i) - \mathbf{A}^T\mathbf{L}\mathbf{r}_i\|^2 \\ &= \sum_{i=1}^N \|\mathbf{B}\mathcal{N}^{-1}\mathbf{c}_i - \mathbf{t}_i\|^2, \end{aligned} \quad (3)$$

where  $i$  is the index of the colors that need to be mapped. This, however, does not guarantee white point preservation; white in the camera's color space need not get mapped to white in the CIEXYZ space. To this end, Finlayson and Drew propose a white point preserving least squares technique [35] that adds a constraint to the previous function to give

$$J = \sum_{i=1}^N \|\mathbf{B}\mathcal{N}^{-1}\mathbf{c}_i - \mathbf{t}_i\|^2 + \Lambda \|\mathbf{B}\mathcal{N}^{-1}\mathbf{c}_w - \mathbf{t}_w\|^2, \quad (4)$$

where  $\Lambda$  is a Lagrange multiplier and  $c_w$  and  $t_w$  are the camera measurements and tristimulus values of the white point, consequently preserving the white-point. A simple extension is to include the neutral colors as approximate constraints in this optimization. Although the transform matrix  $\mathbf{B}$  resulting from the second technique may have larger overall error, it guarantees that distinguished points are mapped error free or nearly so [35]. This matrix now maps the data from the camera's color space to the CIEXYZ color space (which is also unrendered). Other techniques for generating a transform matrix using higher order measurements are also used to obtain better estimates [35]. Typically, a Munsell ColorChecker chart [36] is used to obtain colors that need to be mapped, although other standard charts are also available.

Spaulding et al. [37] refer to the stages a raw digital camera image goes through as "image states," and refer to the unrendered (scene-referred) and rendered (output-referred) versions of the image. We had alluded to the fact that the filter sensitivities of color cameras are different from the CIEXYZ functions and that this is intentional. One of the main reasons for this lies behind the need not for accurate scene reproduction but rather for pleasing reproductions of the scene. These aims are not the same. This does not change the manner in which the aforementioned equations are set up for optimization, but does affect the manner in which the scene is encoded. The sRGB

color space [38] is one popularly used, unrendered color space that has been shown to be well suited to digital color camera operations, especially with regard to white-point conversions [39]. It was designed with the sRGB color space [40] as a candidate output-referred color space and is thereby based on the ITU-R BT.709-3 color primaries (basic RGB colors) and set to a D65 (standard daylight at "correlated color temperature"  $6,500^\circ$ ) white point. Another popularly used scene-referred (unrendered) color space is RIMMRGB [34], a wide-gamut color space, well suited for mathematical manipulations. This space is based on imaginary primaries and a D50 white point. Interchanging between the unrendered spaces and CIEXYZ color spaces is via simple matrix multiplications (linear transforms), included in these standards.

### COLOR TRANSFORMATION—RENDERED COLOR SPACES

Rendered color spaces are designed for output purposes and have a limited gamut, unlike their unrendered counterparts that are scene based. Rendered color spaces are produced from the image data in the unrendered color space. This process involves a loss in dynamic range (rendered spaces are limited to 8 b, while their unrendered counterparts have 12–16-b dynamic ranges). The most common rendered space is the sRGB color space [40], which has become a common cross-platform standard for multimedia. Another common rendered space is defined in ITU-R BT.709-3 [41], a standard originally devised for high definition televisions. The sRGB standard adopts the primaries defined by the ITU-R BT.709-3 standard. These standards define the transformation process from unrendered spaces (mostly CIEXYZ) to 8-b values required by most output media. In [42], Ssstrunk et al. discuss the advantages and disadvantages of a variety of such color spaces. The interested reader is referred to the work by Holm [43] for a study of considerations in color space standards design. With the data now in a standard color space for reproduction, the processing chain is almost complete, modulo possible post-processing steps that make the images look better.

### POSTPROCESSING

Different camera manufacturers use different proprietary steps subsequent to all the aforementioned processing, aimed at image appearance enhancement. Postprocessing is necessary as each of the previous steps may introduce objectionable artifacts. For example, the demosaicking step may introduce a zipper artifact along strong intensity edges. A few of the common post-processing steps employed are color-artifact removal, edge enhancement, and coring. These techniques are mostly heuristic based and require considerable fine-tuning.

The demosaicking step introduces objectionable artifacts referred to as zipper and confetti artifacts. It is important to keep these artifacts to a minimum while at the same time retaining image sharpness. The solution to this problem lies in a series of choices starting from the camera lens system to the size of the sensor to the demosaicking technique used, with implementation restrictions and tradeoffs con-

trolling these choices. For a given set of choices, further processing is used to reduce errors/artifacts. This is typically done by generating luminance-chrominance channels from the demosaicked image and performing spatial operations on the chrominance channels, especially at locations where artifacts are pronounced.

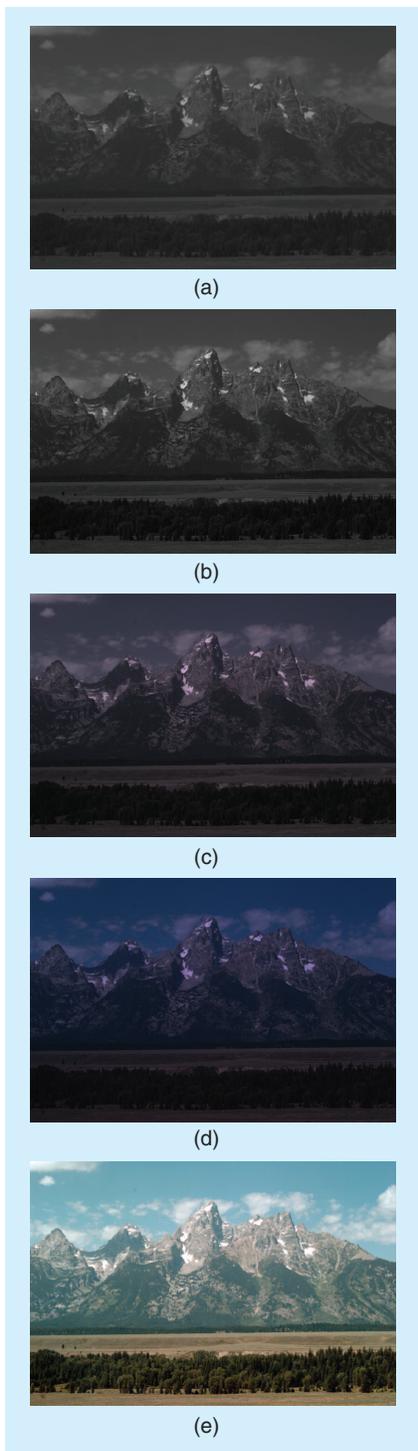
The human eye is known to be highly sensitive to sharp edges; we prefer sharp edges in a scene to blurred ones. Specifically, we are more sensitive to horizontal and vertical edges than diagonal ones, and even less sensitive to edges in other directions. Most camera manufacturers use an edge-enhancement step such as unsharp masking to make the image more appealing by reducing the low frequency content in the image.

Coring is used to remove detail information that has no significant contribution to image detail and behaves much like noise [44]. The term “coring” originates from the manner in which the technique is implemented: usually a representation of the data to be filtered is generated at various levels of detail (scales), and noise reduction is achieved by thresholding (or coring) the transform coefficients computed at the various scales. How much coring needs to be performed (how high the threshold needs to be set) is, again, heuristic.

#### DISPLAY/ARCHIVE

Depending on the reproduction medium (computer monitor, type of printer), the data needs to be further transformed into the appropriate color space. For example, in the case of a CRT monitor (an additive color system), the data needs to be transformed into an 8-b format, taking into consideration the display model used (the gamma values, offsets, and color temperature). In the case of a four-color, dye-sublimation printer (subtractive color), the data is transformed into a CMYK color space, and the appropriate color reproduction model is used.

For storage purposes, current implementations focus on two solutions, professional cameras, which have the freedom of much larger sensors and



**[FIG8]** Sample image as it flows through the image pipeline. (a) The raw image as stored in a TIFF/EP file. (b) After linearization, dark noise subtraction and a gray-world white balance. (c) The image after demosaicking. (d) The image after transformation to the sRGB color space. (e) The image after transformation to the sRGB color space and application of a  $3 \times 3$  median filter. All images in unrendered spaces are shown here using a min-max scaling for display purposes.

storage space, prefer to store the images in either a proprietary file format or in a recently developed file format called tagged image file format for electronic photography (TIFF/EP). TIFF/EP files are regular TIFF files, but additional tags in the file store information about the linearization function, spectral sensitivities, illuminant used, details of camera settings, and the color transform matrix, and usually store the image data in its mosaicked form (the demosaicking step has not yet been performed) [45], [46]. Consumer cameras tend to use the EXIF format [47], as it is highly compact and easy to implement in hardware [48]. The JPEG algorithm is another alternative, and usually baseline JPEG is used. In some cameras, audio that accompanies digital images can also be stored.

JPEG2000, a new international standard for image compression [49], provides better compression efficiency than the current JPEG standard, as well as a set of useful features for DSC applications such as accurate control of compressed data size and progressive coding for image quality and resolution. While these advantages make JPEG2000 a good candidate for the next generation compression standard for DSC, its computational complexity and memory requirements are limiting factors.

#### IMPLEMENTATION

Some of the more difficult problems faced by camera developers are

- from an implementation perspective, maintaining a fast shutter release, short shot-to-shot delay, low power consumption, speed of focus versus accuracy tradeoff, the sequence of color processing (when must the white balancing and other normalizations be performed, before or after demosaicking) and, above all, performing all of the mathematical calculations in fixed-point arithmetic on an embedded system
- from a color science and signal processing perspective, choice of the antialiasing filter, color filter sensitivities, CFA pattern and

consequently the demosaicking technique used, automatic focus, exposure and white balancing, improving image quality, and choosing a convenient color space (and the mathematics involved) for storing the images.

As an illustration of the various processing steps, we show in Figure 8 a summarized version of the image processing chain and how the image looks after each of the various stages.

Digital camera development is a highly market-driven field. As in any market-driven product, the user often determines the features and options that are implemented. From the perspective of a camera user, there are many goals to be met; high image quality, accurate color reproduction, low power consumption, and fast, yet conveniently reconfigurable operation. The market, and hence consumers, determine what is acceptable and what is not for given price and performance. Not simply science, but human tastes dictate the final steps in the pipeline: a “preferred reproduction” is required [50], rather than an accurate one.

### MEASURING CAMERA QUALITY

From the previous description, one observes that each camera manufacturer can potentially have their own intellectual property invested in a camera, which can change the resulting images. Each of the blocks described previously may be customized by choosing the size of the sensor, filter sensitivities, auto-focus and exposure technique, color transform, white balance algorithm, and postprocessing, image reproduction technique. To assess the quality of the various choices made and the resulting images, camera manufacturers use a host of visual observers. Evaluating image quality has been receiving much interest in the standards committees. It raises questions about the ability to quantify observer judgment. In a comprehensive book, Keelan and other contributors [51] have clearly described work performed in the Eastman Kodak Company, over the years, in quantifying image quality. The ISO/WD 20462/1.2-3.2 working-draft standard documents contain recent attempts at standardizing experimental methods used in evaluating image quality [52].

### WHAT DOES THE FUTURE HOLD?

Interestingly, film cameras are the benchmark for digital color cameras. Users will feel compromised until DSCs match film cameras in aspects such as shot-to-shot delay, auto-focus speed, and picture quality. However, DSCs already exceed film cameras by providing features such as white-balance (without additional equipment or post-processing at the print stage) and ISO-sensitivity options and continue to differentiate DSCs from their film counterparts.

HDR imaging is becoming increasingly popular in recent implementations. One may well expect HDR imaging solutions to become mainstay in all cameras.

From a color processing perspective, the near-future holds promise of a standard for digital camera characterization, with care taken for white point preservation [35]. Work in this direction has led to the development of the ISO-RGB color space [33] and more recently, the scRGB color space [38]. The data in the

ISO-RGB and scRGB color space represents an estimate of the scene colorimetry and hence are able to adapt to varying imaging conditions.

Digital cameras can also function as digital video cameras with few additional hardware components, especially when they have programmable processing units to implement video coded algorithms. Traditional video cameras will be the benchmark for this functionality and will drive video signal processing requirements such as real-time auto white balance, exposure and focus control, and vibration blur compensation.

From the perspective of a consumer, the future holds larger sensor arrays, which will be accompanied by more powerful processors in digital cameras, with possibly lower power requirements and more complex operations. Image enhancement and editing features will be more common in future DSCs, and various image processing algorithms will be required to implement such features. Among many possibilities are histogram balancing, contrast enhancement, saturation enhancement, negative image, solarization, posterization, sharpening, and red-eye reduction, which are conventionally done on the computer after moving the picture files from the DSC.

### ACKNOWLEDGMENTS

The authors are indebted to Jack Holm for providing them with images and also for providing them with various documents and valuable insight that have helped improve the quality of this work.

### AUTHORS

*Rajeev Ramanath* received the bachelor of engineering degree in electrical and electronics engineering from Birla Institute of Technology and Science, Pilani, India in 1998. He obtained an M.S. and Ph.D. in electrical engineering from North Carolina State University in 2000 and 2003, respectively. After completing a one-year postdoctoral appointment with the Electrical and Computer Engineering department at North Carolina State University, he has taken a position with Texas Instruments, Inc. with their DLP products division. His research interests include computer vision, image and signal processing, demosaicking in digital color cameras, color science, and automatic target recognition.

*Wesley E. Snyder* received his B.S. in electrical engineering from North Carolina State University (NCSU) in 1968. He received the M.S. and Ph.D., also in electrical engineering, from the University of Illinois. In 1976, he returned to NCSU to accept a faculty position in electrical engineering, where he is currently a full professor. His research is in the general area of image processing and analysis. He is currently working on new techniques in mammography, inspection of integrated circuits, and automatic target recognition. He also has an appointment at the Army Research Office. He has just completed a new textbook on machine vision.

*Youngjun Yoo* received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1998. He is currently a member group technical staff and sen-

ior customer program manager in Cellular Media Business, Texas Instruments (TI), Dallas, which he joined in 1998. His research interests include image/video compression, color imaging, real-time imaging architecture and applications, and software engineering. He published numerous journal/conference papers and technical articles and holds several patents. He is a recipient of 1998 Samsung HumanTech Thesis Award and is a member of SPIE and IEEE.

**Mark S. Drew** is an associate professor in the School of Computing Science at Simon Fraser University (SFU) in Vancouver, Canada. He received his undergraduate education in engineering science at the University of Toronto, studied functional analysis in mathematics for his master's degree, and received the Ph.D. in theoretical physics at the University of British Columbia. His interests lie in the fields of multimedia, computer vision, especially focusing on image processing, color, photorealistic computer graphics, and visualization. He has published more than 80 refereed papers in journals and conference proceedings. He is the holder of a U.S. patent in digital color processing and a U.K. patent application in color computer graphics.

## REFERENCES

- [1] H.J. Trussell, "DSP solutions will run the gamut for color solutions," *IEEE Signal Processing Mag.*, vol. 10, no. 2, pp. 8–23, 1993.
- [2] G. Sharma and H.J. Trussell, "Digital color imaging," *IEEE Trans. Image Processing*, vol. 6, no. 7, pp. 901–932, 1997.
- [3] H.J. Trussell, M.J. Vrhel, and E. Saber, "Color image processing," *IEEE Signal Processing Mag.*, vol. 22, no. 1, pp. 14–22, 2005.
- [4] J.E. Adams, K. Parluski, and K. Spaulding, "Color processing in digital cameras," *IEEE Micro*, vol. 18, no. 6, pp. 20–29, 1998.
- [5] K. Parulski and K. Spaulding, "Color image processing for digital cameras," in *Digital Color Imaging Handbook*, G. Sharma, Ed. Boca Raton, FL: CRC Press, 2003, pp. 727–757.
- [6] J. Holm, I. Tastl, L. Hanlon, and P. Hubel, "Color processing for digital photography," in *Colour Engineering: Achieving Device Independent Colour*, P. Green and L. MacDonald, Eds. New York: Wiley, 2002, pp. 179–220.
- [7] R. Bala and G. Sharma, "System optimization in digital color imaging," *IEEE Signal Processing Mag.*, vol. 22, no. 1, pp. 55–63, 2005.
- [8] B.E. Bayer, "Color imaging array," U.S. Patent 3 971 065, 1976.
- [9] A. Morimura, K. Uomori, Y. Kitamura, A. Fujioka, J. Harada, S. Iwanura, and M. Hirota, "A digital video camera system," *IEEE Trans. Consumer Electron.*, vol. 36, no. 4, pp. 866–875, 1990.
- [10] T. Haruki and K. Kikuchi, "Video camera system using fuzzy logic," *IEEE Trans. Consumer Electron.*, vol. 38, no. 3, pp. 624–634, 1992.
- [11] P. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. SIGGRAPH*, 1997, pp. 369–378.
- [12] G. Ward, "High dynamic range imaging," in *Proc. IS&T/SID 9th Color Imaging Conf.*, 2001, pp. 9–16.
- [13] T. Mitsunaga and S.K. Nayar, "High dynamic range imaging: Spatially varying pixel exposures," in *Proc. IEEE CVPR*, pp. 472–479, 2000.
- [14] S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High dynamic range video," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 319–325, 2003.
- [15] J. Lee, K. Kim, and B. Nam, "Implementation of passive automatic focusing algorithm for digital still camera," *IEEE Trans. Consumer Electron.*, vol. 41, no. 3, pp. 449–454, 1995.
- [16] M. Subbarao and J.K. Tyan, "Selecting the optimal focus measure for auto-focussing and depth-from-focus," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, no. 8, pp. 864–870, 1998.
- [17] S. Pregara, "Autofocus sensor," U.S. Patent 6 600 878 B2, 2003.
- [18] N.L. Stauffer, "Range determination system," U.S. Patent 4 185 191, 1980.
- [19] J. Greivenkamp, "Color dependent optical prefilter for the suppression of aliasing artifacts," *Appl. Opt.*, vol. 29, no. 5, pp. 676–684, 1990.
- [20] R. Lyon and P.M. Hubel, "Eyeing the camera: Into the next century," in *Proc. IS&T/SID 10th Color Imaging Conf.*, 2002, pp. 349–355.
- [21] G. Wyszecki and W.S. Stiles, *Color Science—Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. New York: Wiley, 1982.
- [22] *Photography—Electronic Still-Picture Cameras—Methods for Measuring Opto-Electronic Conversion Functions (OECFs)*, ISO 14524, 1999 [Online]. Available: <http://www.iso.org>
- [23] N. Kehtarnavaz, H.J. Oh, and Y. Yoo, "Development and real-time implementation of auto white balancing scoring algorithm," *J. Real Time Imaging*, vol. 8, no. 5, pp. 379–386, 2002.
- [24] K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms—Part I: Methodology and experiments with synthesized data," *IEEE Trans Image Processing*, vol. 11, no. 9, pp. 972–983, 2002.
- [25] G.D. Finlayson, S.D. Hordley, and I. Tastl, "Gamut constrained illuminant estimation," in *Proc. ICCV03*, 2003, pp. 792–799.
- [26] R. Ramanath, W.E. Snyder, G.L. Bilbro, and W.A. Sander, "Demosaiicking methods for the Bayer color array," *J. Electr. Imaging*, vol. 11, no. 3, pp. 306–315, 2002.
- [27] R. Kimmel, "Demosaiicking: Image reconstruction from color CCD samples," *IEEE Trans. Image Processing*, vol. 8, no. 9, pp. 1221–1228, 1999.
- [28] H.J. Trussell and R.E. Hartwig, "Mathematics for demosaicking," *IEEE Trans. Image Processing*, vol. 11, no. 4, pp. 485–492, 2002.
- [29] B.K. Gunturk, Y. Altunbasak, and R.M. Mersereau, "Color plane interpolation using alternating projections," *IEEE Trans. Image Processing*, vol. 11, no. 9, pp. 485–492, 2002.
- [30] R. Ramanath and W.E. Snyder, "Adaptive demosaicking," *J. Electr. Imaging*, vol. 12, no. 4, pp. 633–642, 2003.
- [31] N. Kehtarnavaz, H.J. Oh, and Y. Yoo, "Color filter array interpolation using color correlation and directional derivatives," *J. Electron. Imaging*, vol. 12, no. 4, 2003.
- [32] B.K. Gunturk, J. Glotzbach, Y. Altunbasak, R.M. Mersereau, and R.W. Schafer, "Demosaiicking: Color filter array interpolation," *IEEE Signal Processing Mag.*, vol. 22, no. 1, pp. 44–54, 2005.
- [33] *Graphic Technology and Photography—Colour Characterisation of Digital Still Cameras (DSCs)—Part 1: Stimuli, Metrology, and Test Procedures*, ISO/WD 17321-1, 1998.
- [34] ANSI/I3A IT10.7466, *Photography—Electronic Still Picture Imaging—Reference Input Medium Metric RGB Color Encoding (RIMM-RGB)*, Amer. Nat. Standards Inst., 2002.
- [35] G.D. Finlayson and M.S. Drew, "Constrained least-squares regression in color spaces," *J. Electr. Imaging*, vol. 6, no. 4, pp. 484–493, 1997.
- [36] C.S. McCamy, H. Marcus, and J.G. Davidson, "A color-rendition chart," *J. Appl. Photog. Eng.*, vol. 2, no. 3, pp. 95–99, 1976.
- [37] K.E. Spaulding, G.J. Woolfe, and G.J. Giorgianni, "Reference input/output medium metric RGB color encodings (RIMM/ROMM RGB)," in *Proc. IS&T/SID 8th Color Imaging Conf.*, 2000, pp. 155–163.
- [38] *Multimedia Systems and Equipment—Colour Measurement and Management—Part 2-2: Colour Management—Extended RGB Colour Space—scRGB*, IEC 61966-2-2, 2003.
- [39] G. Finalyson and M.S. Drew, "White-point preserving color correction," in *Proc. IS&T/SID 5th Color Imaging Conf.*, 1997, pp. 258–261.
- [40] *Multimedia Systems and Equipment—Colour Measurement and Management—Part 2-1: Colour Management—Default RGB Colour Space—sRGB*, IEC 61966-2-1 (1999-10), 1999 [Online]. Available: <http://www.srgb.com>
- [41] *Parameter Values for the HDTV Standards for Production and International Programme Exchange*, ITU-R BT.709-3, 1998.
- [42] S. Süsstrunk, R. Buckley, and S. Swen, "Standard RGB color spaces," in *Proc. IS T/SID 7th Color Imaging Conf.*, 1999, pp. 127–134.
- [43] J. Holm, "Issues relating to the transformation of sensor data into standard color spaces," in *Proc. IS&T/SID 5th Color Imaging Conf.*, 1997, pp. 290–295.
- [44] J.P. Rossi, "Digital techniques for reducing television noise," *J. SMPTE*, vol. 87, no. 3, pp. 134–140, 1978.
- [45] *Electronic Still-Picture Imaging—Removable Memory—Part 2: TIFF/EP Image Data Format*, ISO 12234-2, 2001.
- [46] TIFF-EP, "Image file format overview" [Online]. Available: <http://www.kodak.com/US/en/digital/dlc/book2/chapter4/digitalp1.shtml>
- [47] *Digital Still Camera Image File Format Standard (Exchangeable Image File Format for Digital Still Camera: Extif)*, Japan Electron. Industry Develop. Assoc., JEIDA-49-1998, version 2.1 ed., 1998.
- [48] K. Illgner, H.G. Gruber, P. Gelabert, J. Liang, Y. Yoo, W. Rabadi, and R. Talluri, "Programmable DSC platform for digital still camera," in *Proc. IEEE ICASSP*, 1999, pp. 2235–2238.
- [49] *JPEG 2000 Image Coding System*, ISO/IEC 15444-1, 2000.
- [50] P.M. Hubel, J. Holm, G.D. Finlayson, and M.S. Drew, "Matrix calculations for digital photography," in *Proc. IS&T/SID 5th Color Imaging Conf.*, 1997, pp. 105–111.
- [51] B.W. Keelan, *Handbook of Image Quality*. New York: Marcel Dekker, 2002.
- [52] *Photography—Psychophysical experimental methods to estimate image quality*, ISO/WD 20462-1.2-3.2, 2002.