A Comparison of Primary and Complementary Color Filters for CCD-based Digital Photography

Richard L. Baer, William D. Holland, Jack Holm, Poorvi Vora

Hewlett-Packard Laboratories, Palo Alto, CA 94304

ABSTRACT

This paper presents a comparison between primary (RGB) and complementary (CYMG) CCD color filters arrays, as applied to digital photography. Our analysis is based upon the measured spectral characteristics of the primary and complementary color versions of the Matsushita MN3776 CCD. The important role of the color correction matrix on the quality of the image is considered both in terms of noise and color saturation. Our calculations show that there is a tradeoff between color saturation and ISO speed, when complementary filters are used. Complementary color filters only gain an ISO speed advantage when the color saturation is low. When the color correction matrix is chosen to make the ISO speeds of the two filter systems equivalent, the well capacity of the complementary CCD must be significantly higher because of the higher overall transmission of its color filters. Our comparison includes ISO speed calculations and plots of the color gamut for primary and complementary color filters with various color correction matrices. We conclude that primary color filters are superior for digital photography.

Keywords: CCD, digital photography, color filter array, mosaic color, ISO noise-based speed

1. INTRODUCTION

Modern electronic cameras detect light with an array of photosensitive pixels. The pixels respond to brightness variations in the scene, but they do not directly measure color. A number of different methods have been devised to measure color, including sequential image capture with different color filters, the use of multiple image sensors with different color filters and the use of mosaic color filter arrays[1] on the surface of a single image sensor. The most common approach is the use of the mosaic filter array because it is the least expensive.

While many different color filter array systems have been devised, the most common systems are complementary color, which utilizes cyan, yellow, magenta and green color filters, and primary color, which utilizes red, green and blue color filters. Most electronic camcorders use complementary color filters. The advantage of the complementary system is that it simplifies the derivation of separate luminance and chrominance signals in CCD-based cameras. Signals from adjacent color pixels in the same column can be summed together in the CCD and used to derive luminance and chrominance values by simple additions and subtractions of pairs of pixels at the output[2].

The earliest digital still cameras used complementary color CCDs that were originally designed for camcorders. As digital cameras have evolved, many manufacturers have opted for CCDs with primary color filters. However cameras with both filter types can be still found on the market. The purported advantage of primary color filters is better color reproduction, as a consequence of the better separation of the color channels. The purported advantage of complementary color (in still photography) is better sensitivity, as a consequence of the broader filter responses. Our conclusion is that both claims are correct: primary color filters provide better color fidelity, but complementary filters can provide higher sensitivity *if* the color fidelity is compromised.

This paper consists of two main sections. The theory section presents our methods for determining the appropriate color correction matrix based upon the quantum efficiency curves of the detector, and our method for determining the effective ISO speed potential of the sensor based upon the quantum efficiency, pixel area, read noise and color correction matrix. The experimental results section describes the data that we collected with our experimental test camera and the subsequent analysis. Our conclusions are presented in the final section.

Part of the IS&T/SPIE Conference on Sensors, Cameras, and Applications for Digital Photography • San Jose, California • January 1999
SPIE Vol. 3650 • 0277-786X/99/$10.00

2. THEORY

The purpose of our analysis is to determine what effect the choice of primary or complementary color filters has on the noise-based ISO speed that an electronic camera can achieve. The ISO speed depends on the signal processing algorithms used to prepare the image, as well as the characteristics of the image capture device. For example, a blur filter can decrease noise and increase speed, at the expense of resolution. One of the signal processing steps that has a profound effect on the noise, and varies depending on the spectral response of the sensor, is color correction. Color correction is usually implemented as a matrix multiplication that transforms the color values from the sensor into a more accurate representation of the colors in the scene.

This section is divided into two parts. In the first, the calculation of the color correction matrix and tradeoffs between saturation and noise amplification are described. In the second, the method for calculating the ISO noise-based speed potential of a digital camera described.

2.1. Color Correction Matrix Calculation

A color correction matrix is usually used to transform the raw color image obtained from color image sensors to adjust for factors such as variations in illumination and deviations of the actual filter characteristics from the ideal. The color correction matrix is usually calculated to obtain in a least squares fit for the expected data set. See reference 3 for ways of calculating the color correction matrix in the presence of measurement noise.

The linear minimum mean square error (LMMSE) estimate of the tristimulus values from the measurements using a specified set of color filters for zero-mean signal and zero-mean signal-uncorrelated, independent, identically distributed noise is[3]:

$\mathbf{C = A'RM(MRM + \sigma^2 I)^{-1}g}$

where **A** is the matrix whose columns are the CIE matching functions for the specified viewing illuminant; **R** the correlation matrix of the data; **M** the matrix whose columns are the combination of the camera's spectral response and the viewing illuminant; σ^2 the variance of the noise; and **g** the camera measurements which include the effect of measurement noise. The color correction matrix, **C** is:

$\mathbf{C = A'RM(MRM + \sigma^2 I)^{-1}}$

The incorporation of non-zero signal and noise means does not change the basic results of the analysis.

Filters with large overlap in spectral response (such as typical CMYG filters) tend to result in color correction matrices with high condition number[4] (ratio of maximum to minimum singular value) and hence tend to amplify noise and reduce the signal to noise ratio[4,5]. Increasing the value of sigma in the expression for the color correction matrix decreases the condition number and hence the noise amplification for matrices with a high condition number, but also decreases color saturation[6]. For typical CMYG filters, hence, it is possible to reduce the large noise amplification typical of these filters, but this can be done only at the cost of color saturation. RGB filters, on the other hand, do not have as much spectral overlap and hence do not result in color correction matrices with high condition number and hence do not require a decrease in noise amplification. With RGB filters, hence, there is also no decrease in color saturation.

2.2. ISO Speed Model

The ISO speed model is based upon International Standard ISO 12232, Photography – Electronic still-picture cameras – Determination of ISO speed[7]. Even though an ISO speed can not be defined for a particular CCD because of the dependence on image processing operations, the model can provide useful relative speed information. In order to compare the potential speeds of different devices, one must assume a common set of image processing operations. The common image processing steps that have been included in this analysis are white balancing and color correction.

Required inputs to the model include the quantum efficiency versus wavelength curves, the pixel dimensions and the read noise characteristics of the CCD. The ISO speed potential of the CCD increases with quantum efficiency and pixel area. The speed potential decreases with increased read noise, although the effect is less pronounced at high signal levels, where the shot noise of the photon flux is dominant.

The ISO noise-based speed is defined as:

$$S_{noise\,x} = 10 / H_{s/Nx}$$

where $S_{noise x}$ is the ISO noise speed at a mid-tone signal to noise ratio of x, and $H_{s/Nx}$ is the corresponding focal plane exposure in lux-seconds. In order to determine the speed, we solve for the level of illumination required to obtain a specified signal to noise ratio and derive the focal plane exposure from it.

The analysis begins by modeling the illumination source as a Planckian radiator. The CCD response is calculated by multiplying the quantum efficiency by the transmission of a standard infrared filter and the spectral distribution of the source and integrating. An infrared filter is a necessary component because the color filter dyes are transparent in the IR and the IR must be blocked in order to obtain adequate color accuracy. The CCD response is then normalized to the focal plane illumination, which is determined by integrating the photopic response with the spectral distribution of the source. The photopic response is zero in the IR, which is another reason that an IR filter must be included in the model to make the results meaningful. The result is a response vector, with one element for each unique color in the color filter array. The units of the response vector are electrons/(lux·sec·m^2).

The image processing steps that the model includes are white balancing and color correction. White balancing is achieved by selecting multipliers to equalize the responses of all the channels (since the color of the illuminant defines the white point at any color temperature). The derivation of the color correction matrix was described in the previous section. The response vector is multiplied by the white balance and color correction matrices, and then converted to luminance in order to determine the signal level.

The incoming photon flux obeys Poisson statistics, so the variance of the photon count in a specified interval is equal to the mean value. Consequently the response vector can also be used as the input to the noise calculation by interpreting it as noise power rather than mean signal. The square of the read noise (in electrons) can be added to the shot noise terms to get the total noise power. The inclusion of read noise only makes a significant difference in the noise-based speeds calculated for low mid-tone signal to noise ratios. Otherwise the shot noise of the photon flux is dominant. The noise power terms are multiplied by the square of the white balance, color correction and RGB to total noise matrices in order determine the noise power in the image. In the case where only luminance noise is considered, the RGB to total noise matrix is replaced with the RGB to luminance matrix. The noise level of the image is determined by taking the square root of these terms.

If read noise is ignored, the image signal to noise ratio varies as the square root of the focal plane exposure. In this case, the expression may be easily inverted to determine the focal plane exposure (and noise speed) as a function of signal to noise ratio. If read noise is included, numerical methods can be used to invert the expression.

3. EXPERIMENTAL RESULTS

3.1. Experimental Apparatus

In order to investigate the difference between primary and complementary filters, we built a test camera based upon the Matsushita MN3776. The MN3776 is a VGA resolution interline-transfer CCD that is available in monochrome, primary color and complementary color versions. Figure 1 displays a block diagram of the test camera. A Data Translation DT3157 frame grabber was used to interface the camera to a PC.

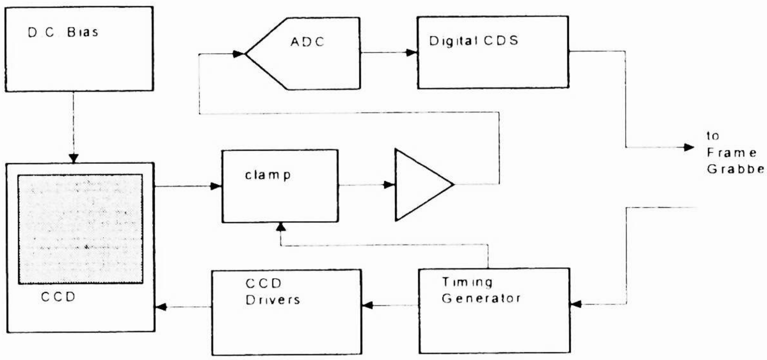


Figure 1. – Test Camera

An Oriel monochrometer with a tungsten halogen lamp, a Labsphere integrating sphere with a separate monitoring port and an NRC photodiode power meter were used to measure the quantum efficiency of the CCD as a function of wavelength. The CCD incorporates microlenses with an effective f-number of about f/2.8 (along the horizontal axis). The measurement system has an f-number of about f/1.5. Consequently the angular spectrum of the illuminant is too broad for the CCD. The resulting quantum efficiency curves therefore underestimate the values that would be measured under collimated illumination. Since we were only concerned with the *relative* performance of the different color filters, we made no attempt to correct for this effect.

The measured quantum efficiency curves for the MN3776 are shown in Figures 2, 3 and 4. All three versions of the MN3776 have about the same quantum efficiency at 800 nm in the infrared, where the color filter dyes are transparent. This supports our assumption that the characteristics of the microlens are not strongly dependent on the nature of the color filter.

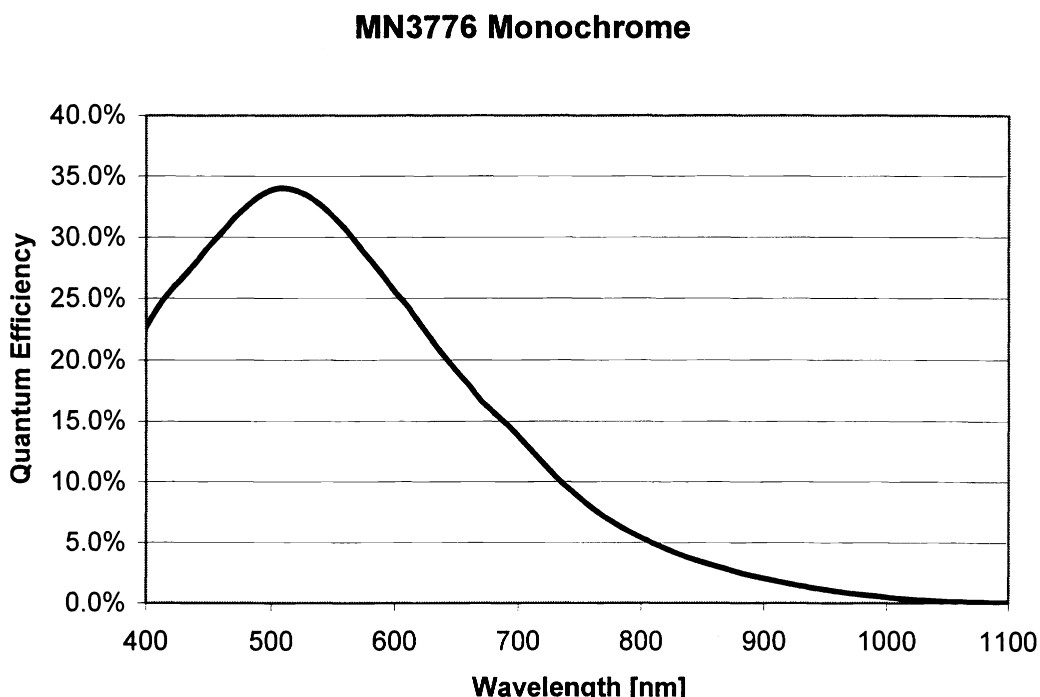


Figure 2 – Monochrome CCD Quantum Efficiency

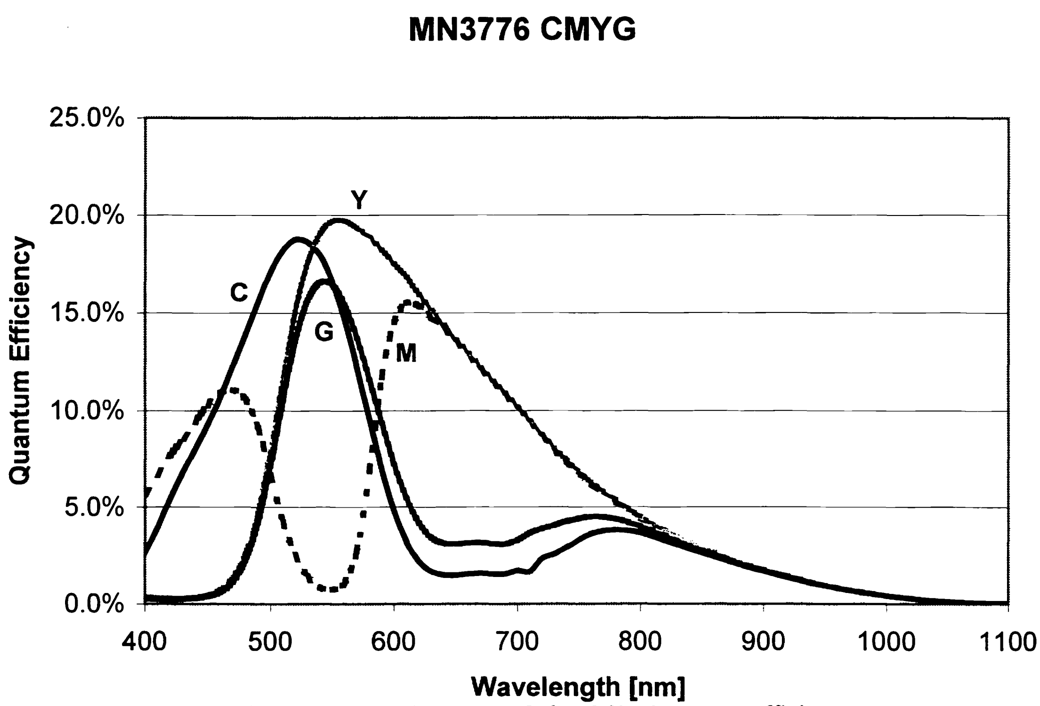


Figure 3 – Complementary Color CCD Quantum Efficiency

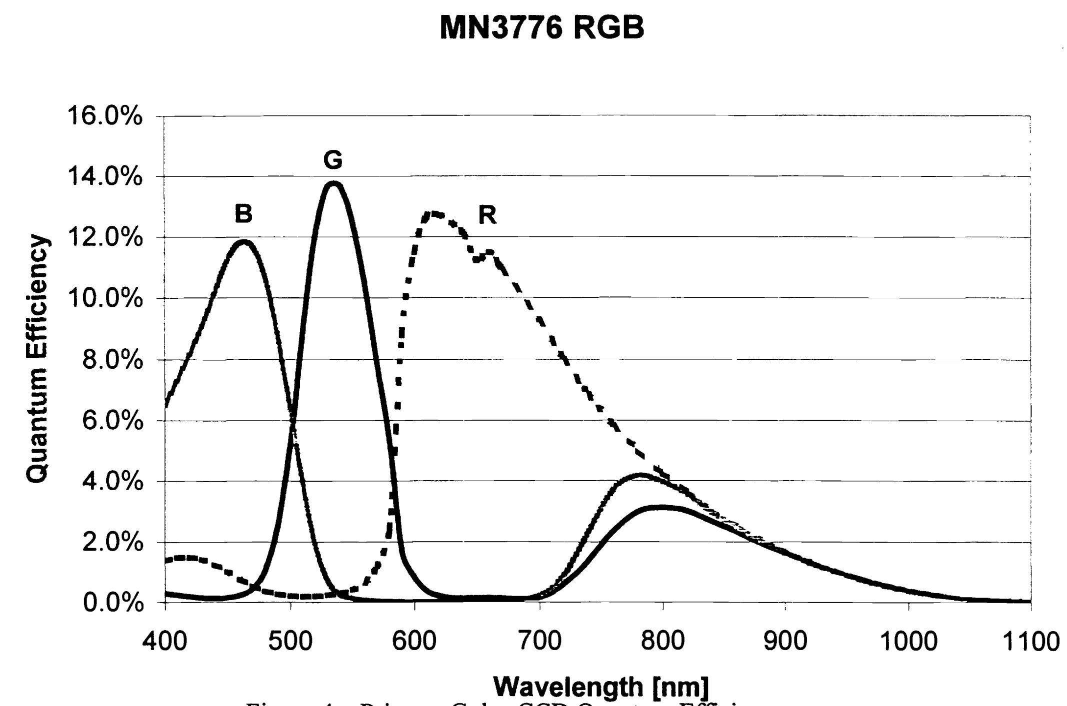
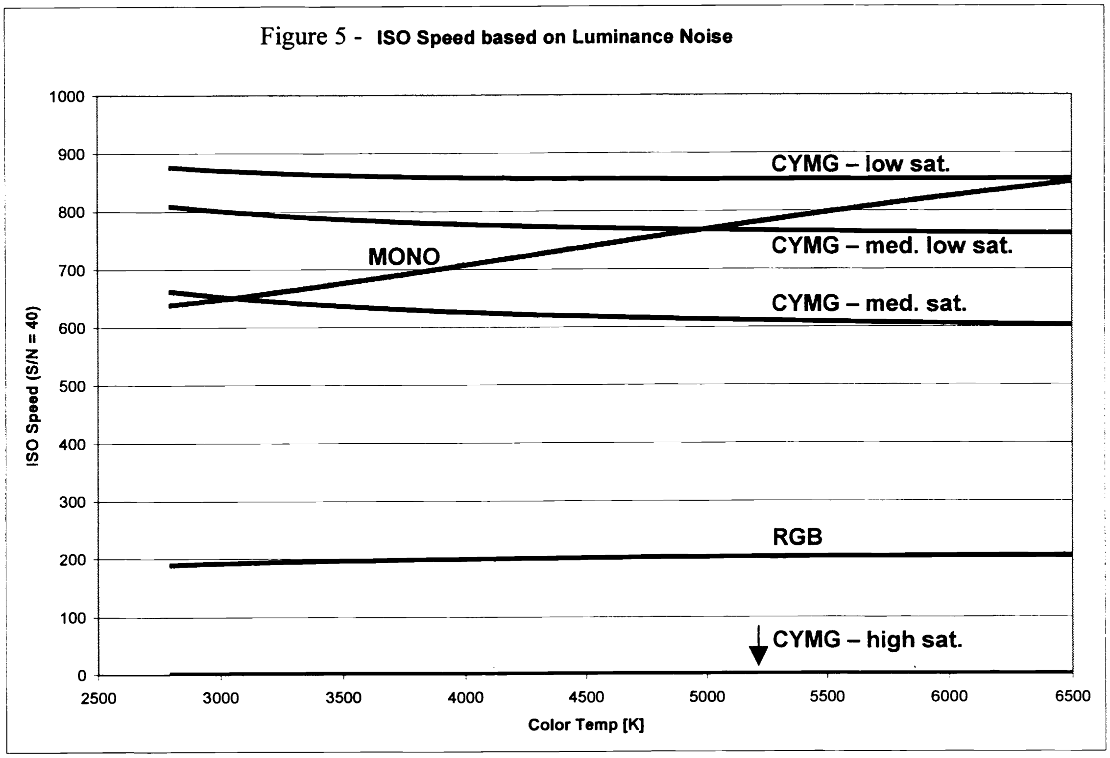


Figure 4 – Primary Color CCD Quantum Efficiency

3.2. ISO Speed comparison

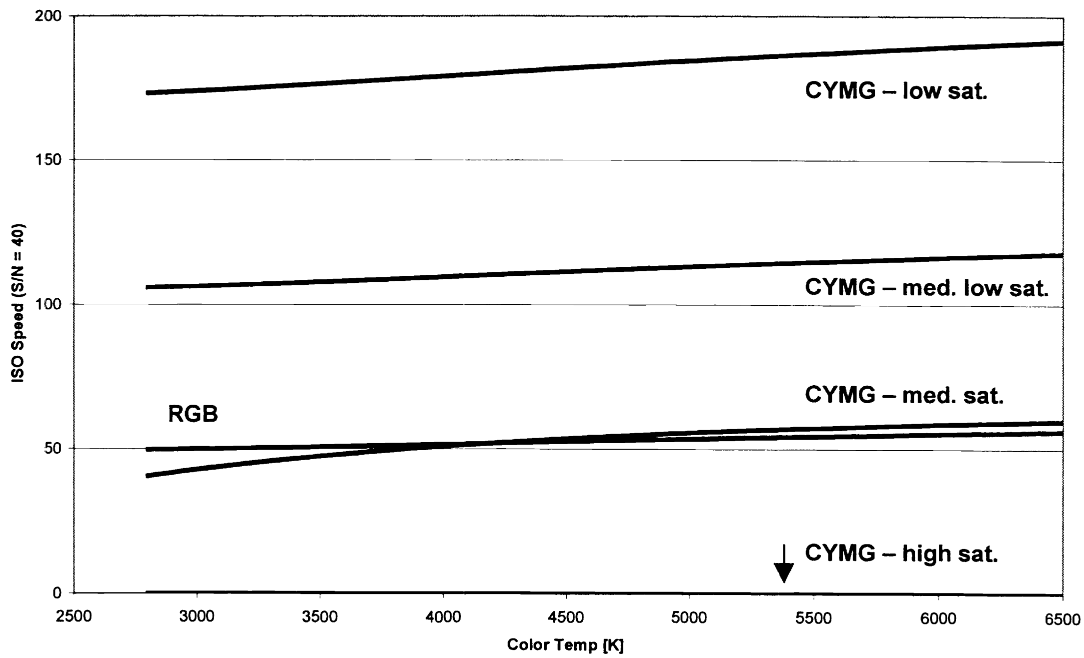
The effective ISO noise speed of the three versions of the MN3776, based only upon the noise of the luminance channel, is presented in Figure 5. In the complementary color case, we consider three different color correction matrices, corresponding to three different degrees of color saturation. The results are consistent with simple intuition: the higher the quantum efficiency and the broader the filter responses, the higher the speed. In the complementary color case we also observe that the speed varies inversely with the degree of saturation.

The fact that the speed of the monochrome sensor is lower that the speed of the complementary color sensor (for low saturation) appears paradoxical. It reflects the assumptions that are made about the demosaicing process in the speed model. When multiple color channels are present, the model assumes that pixels of each color are combined to form super-pixels of lower resolution. Hence the complementary color sensor has approximately the same ISO speed as the monochrome sensor, but at 1/4th the resolution. The RGB sensor has a lower noise speed than the CMYG sensor because three pixels contribute to each super-pixel rather than four, and because the RGB quantum efficiencies are lower and the filter pass-bands are narrower. The system with the lowest speed is the combination of the complementary sensor with the color correction matrix that produces full saturation. In this case the noise amplification is so severe that the effective ISO speed is greatly reduced.



The effective ISO noise speed of the three versions of the MN3776, based upon the noise of the luminance and chrominance channels, is presented in Figure 5. Only the primary and complementary color sensors are considered in this case, since the monochrome sensor doesn't collect any chrominance information. The speeds are reduced by approximately a factor of four, indicating that color noise is the largest component of the total noise. In this case, the color correction matrix has an even larger influence on the effective speed of the complementary color sensors. When a color correction matrix is chosen which yields low color saturation, the speed is several times higher than that of the primary color sensor. When a matrix is chosen that yields medium saturation, the speed of the complementary color sensor is about equal to that of the primary color sensor. If a matrix is chosen that yields high saturation, the amplification of the noise causes the speed to fall so low that it isn't even distinguishable in the figure.

Figure 6 - **ISO Speed based upon Luminance and Chrominance Noise**



The number of electrons required to reach a mid-tone SNR of 40 (excellent quality) is shown in figure 7, for the primary and complementary sensors. In the complementary case, four different color correction matrices with different levels of color saturation are considered. The complementary sensor requires more electrons in all but the lowest saturation case. This is a significant consideration because CCD pixel dimensions continue to shrink in order to increase resolution and decrease cost. This has the unfortunate side effect of lowering the charge capacity of the pixel.

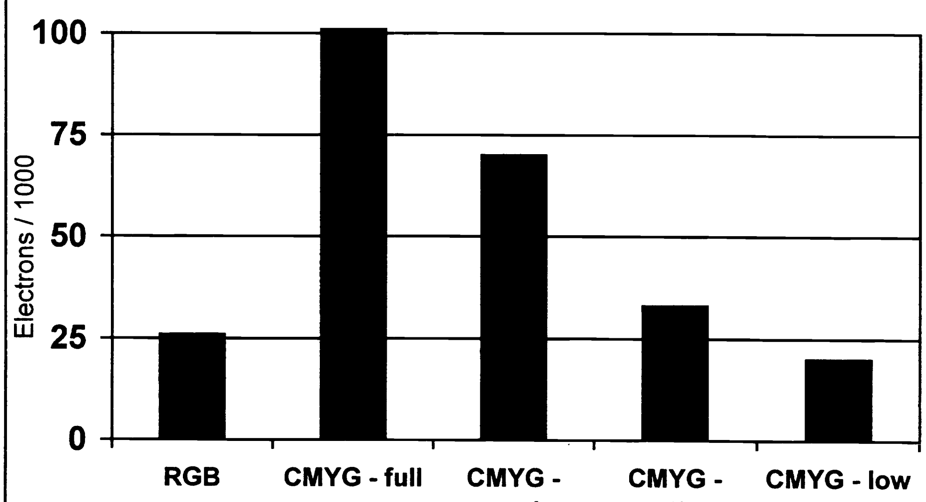


Figure 7. – Charge required to reach S/N 40

3.3. Color gamut comparison

We compared the color gamuts for the RGB filters with sigma = 0, and the CMYG filters with different values of sigma, and hence different levels of noise amplification and color saturation. For both RGB and CMYG filters we first simulated noise-free measurements for hypothetical spectral distributions with single non-zero values at the 31 wavelengths from 400 to 700 nm. at a 10 nm. spacing for recording illuminant CIE D65. Each of these spectral distributions corresponds to a saturated color. We then color corrected the measurements for viewing illuminant D65 with sigma = 0 (infinite SNR) for the RGB filters and with different values of sigma (corresponding to SNRs of infinity, 80, 25.5 and 18) for the CMYG filters. We plotted the resulting tristimulus estimates on a CIE chromaticity plot in Figure 8.

The CIE chromaticity plot is a two-dimensional plot of the values X/(X+Y+Z) vs. Y/(X+Y+Z), where (X, Y, Z) are the CIE tristimulus values. The boundary of the CIE chromaticity plot corresponds to the completely saturated colors - those with a single non-zero value in the spectrum. Lines joining two points in CIE tristimulus space remain lines joining the corresponding points in the chromaticity space. Hence, all spectra that are positive linear combinations of the completely saturated colors (i.e all physically realizable spectra) have chromaticity values that lie within this boundary. Closeness to the boundary is a measure of saturation, and the least saturated color is white (corresponding to X=Y=Z), represented by values (0.33, 0.33) on the plot.

The gamut of realizable spectra using a specified monitor depends on the spectra of the monitor phosphors. All realizable spectra are positive linear combinations of the monitor phosphor spectra, and hence, on the CIE chromaticity plot, all spectra realizable on the monitor lie inside the triangle defined by the chromaticities of the individual phosphors. The triangle is defined by a dashed line on the plot, and each vertex corresponds to one of the monitor phosphors.

Notice that the RGB chromaticity values are at the boundary of the triangle, and so are the CMYG chromaticity values for high and medium-high noise amplification. For these values, however, the noise amplification is visible, as can be seen from the color-corrected images. For medium-low and low saturation the CMYG chromaticity gamut shrinks inside the triangle defined by the monitor and it is clear that the saturation is compromised when decreasing noise amplification.

4. CONCLUSIONS

The ISO noise-based speed potential of a camera can be estimated from the quantum efficiency, pixel area and read noise of its sensor, using a simple model for the effect of subsequent image processing steps. The quantum efficiency curves enter into the calculations in two ways. First, they determine how much light will be collected and what the mid-tone signal to noise ratio will be at the output of the CCD. Second, they determine the requisite color correction matrix. The color correction matrix restores the colors to their proper values, but it also amplifies the noise. The effect is much more severe with complementary color filters than with primary color filters because the off-diagonal matrix elements are much larger.

Complementary color filters have a broader pass-band than primary color filters. Our measurements also indicate that their peak transmission is higher. These factors cause CCDs with complementary color filters to collect more light and achieve higher signal to noise ratios at the CCD output than CCDs with primary filters. However, the signal to noise ratio advantage at the CCD output is nullified by the noise amplifying properties of the color correction matrix. In order to gain a sensitivity advantage from complementary color, the color saturation and color gamut must be sacrificed.

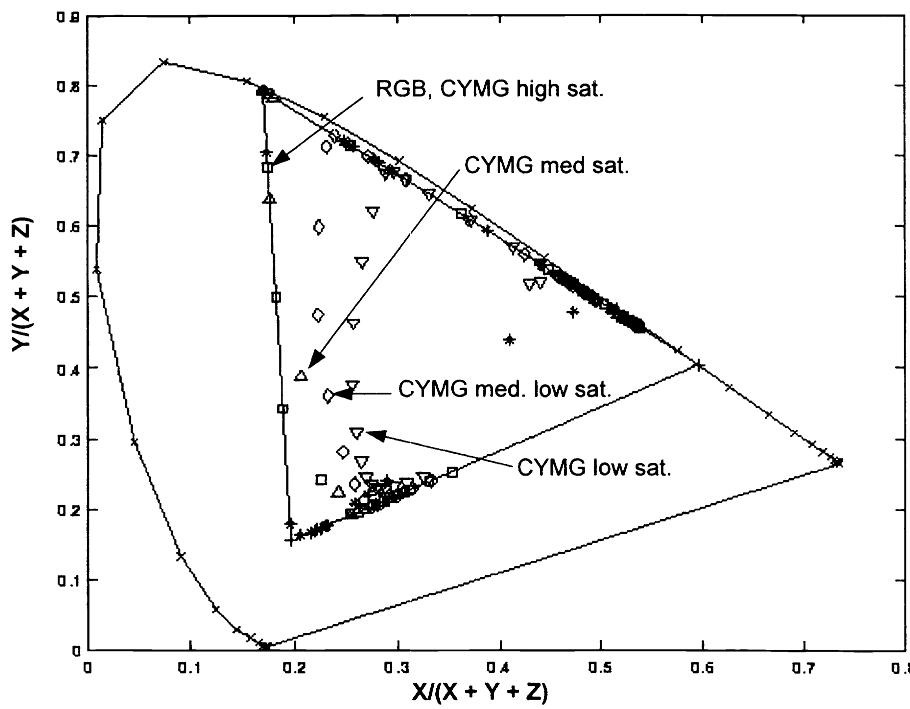


Figure 8. – Color Gamut Comparison

ACKNOWLEDGEMENTS

The authors wish to gratefully acknowledge Matsushita Electric Corporation, for providing the CCDs used in this investigation.

REFERENCES

1. D. Manabe, T. Ohta, Y. Shimidzu, "Color Filter Array for IC Image Sensor", Proc. CICC83, pp. 451-455, Rochester, May 23-25, 1983.
2. T. Watanabe, K. Hashiguchi, T. Yamano, J-I. Nakai, S. Miyatake, O. Matsui, K. Awana, "A CCD Color Signal Separation IC for Single-Chip Color Imagers", IEEE Trans. Elec. Dev., Vol. ED-31, No. 2, pp. 183-188, 1984.
3. M. J. Vrhel, and H. J. Trussell, `Filter Considerations in Color Correction', IEEE Transactions on Image Processing, Vol. 3, No. 2, pp. 147-161, Mar. 1994.
4. Poorvi L. Vora, `Inner products and orthogonality in the design of color recording filter design', Hewlett-Packard Technical Report, HPL-98-189, Nov. 1998.
5. M. J. Vrhel, and H. J. Trussell, `Optimal Color Filters in the Presence of Noise', IEEE Transactions on Image Processing, Vol. 4, No. 6, pp. 814-823, Jun. 1995.
6. Poorvi L. Vora and Cormac E. Herley, "Trade-offs between color saturation and noise sensitivity in image sensors", Proc. Int. Conf. Image Proc., Chicago, Oct. 1998.
7. International Organization for Standardization, ISO 12232:1998(E), 1998.

Correspondence: Email: baer@hpl.hp.com; Telephone: (650) 857-7505; FAX: (650) 857-4320