Automating the design of image processing pipelines for novel color filter arrays: Local, Linear, Learned (L^3) method

Qiyuan Tian[a], Steven Lansel[b], Joyce E. Farrell[a], and Brian A. Wandell[a,c]

[a]Electrical Engineering Department, Stanford University, Stanford, CA-94305, USA;
[b]Olympus Corporation of the Americas, Sunnyvale, CA-94085, USA;
[c]Psychology Department, Stanford University, Stanford, CA-94305, USA.

ABSTRACT

The high density of pixels in modern color sensors provides an opportunity to experiment with new color filter array (CFA) designs. A significant bottleneck in evaluating new designs is the need to create demosaicking, denoising and color transform algorithms tuned for the CFA. To address this issue, we developed a method(local, linear, learned or L^3) for automatically creating an image processing pipeline. In this paper we describe the L^3 algorithm and illustrate how we created a pipeline for a CFA organized as a 2×2 RGB/W block containing a clear (W) pixel. Under low light conditions, the L^3 pipeline developed for the RGB/W CFA produces images that are superior to those from a matched Bayer RGB sensor. We also use L^3 to learn pipelines for other RGB/W CFAs with different spatial layouts. The L^3 algorithm shortens the development time for producing a high quality image pipeline for novel CFA designs.

Keywords: Image processing pipeline, low light imaging, color filter array, image processing, local linear learned(L^3)

1. INTRODUCTION

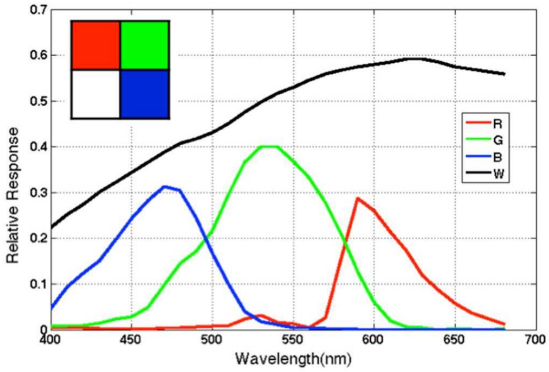
The drive to increase the number of megapixels in digital cameras has produced sensors with higher spatial resolution than is required for most imaging applications. This excess of pixels presents an opportunity to increase other aspects of sensor functionality. For example, there are opportunities to improve the dynamic range and spectral accuracy of the sensor by increasing the number of filters in a color filter array (CFA).

Most consumer digital cameras use the Bayer CFA[1] and color image processing pipelines have been optimized for this design. Cameras with other CFA designs require new image processing algorithms and the time and effort needed to develop these algorithms discourages experimentation with novel designs. In this paper, we describe a technology for automating the design of image processing pipelines for sensors with unconventional CFAs. The algorithm, first introduced by Lansel and Wandell,[2–4] is called L^3, based on the algorithm's three main principles: Local, Linear and Learned.

The L^3 algorithm uses simulated training data to learn a set of linear operators. Each operator is selected to optimize the output image for a specific type of pixel, light level, or contextual pattern. Output images are created by applying the appropriate learned and stored linear operators to the input data. Hence, the application of these precomputed local, linear operators is the image processing pipeline; the pipeline performs demosaicking, denoising, sensor conversion, and illuminant correction.

We illustrate the L^3 algorithm development for a 2×2 RGB/W CFA design (Figure 1 left).[5] The RGB/W design replaces one G filter with a clear (W) filter. The W filter transmits much more light than the G filter, enabling the sensor to record data at low light levels. The W pixels, like the human rod photoreceptors, are useful at very low levels of illumination, but they saturate at higher light levels.[5] The RGB pixels, like the human cone photoreceptors, respond well at higher light levels, but they are noisy and unreliable at low light levels.

Send correspondence to Qiyuan Tian. E-mail: QT (qytian@stanford.edu), SL (steven.lansel@olympus.com), JF (joyce_farrell@stanford.edu), BW (wandell@stanford.edu)



Pixel width/ height (μm)	2.2
Fill factor	0.45
Dark voltage (V/sec)	1 x 10^{-5}
Read noise (mV)	1.34
Dark signal nonuniformity (mV)	1.4
Photoreceptor nonuniformity (%)	2.2 x 10^{-3}
Conversion Gain (μV/e)	200
Voltage swing (V)	1.8
Well capacity (electrons)	9000
Analog gain	7.98

Figure 1. RGB/W CFA layout and spectral sensitivities (left) and sensor parameter values used for simulations (right).

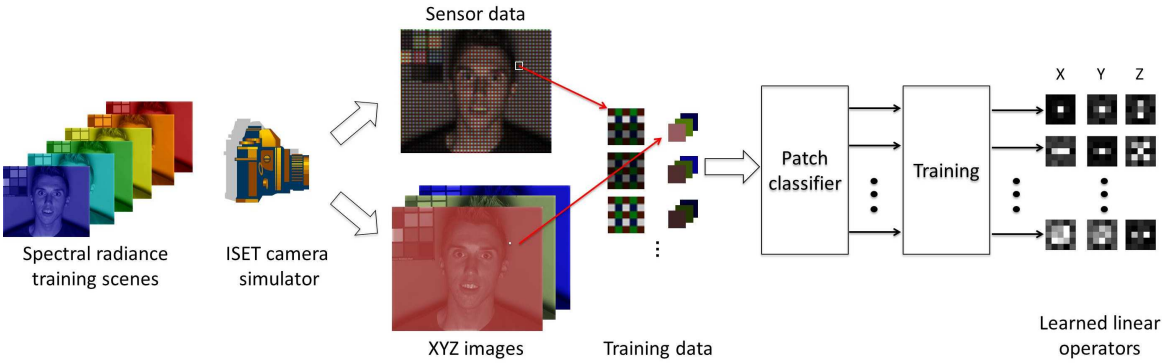


Figure 2. Overview of the L^3 method.

The RGB/W design poses several image processing challenges. In addition to the new spatial pattern, the RGB/W design contains a large sensitivity mismatch between the W and the RGB color pixels. The image processing pipeline must adapt to this sensitivity mismatch. While several groups developed algorithms for RGB/W arrays,[6–12] only one[7] specifically addressed the sensitivity mismatch. At high intensity levels the system should use the RGB data and avoid the saturated W pixels. As the image intensity transitions to lower levels, the system should smoothly replace the unreliable RGB data with information from the W pixel values. We show that the L^3 method meets the challenge of learning a pipeline that smoothly transitions between different illumination levels. The following sections describe the L^3 method in detail and present experimental results.

2. METHOD

Figure 2 provides an overview of the L^3 method, consisting of training data generation, response image-patch classification and linear operator learning. We explain each component in the following sections.

2.1 Training Data Generation

The training data are composed of scene selection, scene statistics, and sensor data. In general, the scenes and scene statistics should be representative of the intended application. In this paper, we illustrate L^3 for the goal of rendering natural images for a display. Thus the scenes are natural images and scene statistics are the CIEXYZ values. The sensor data are determined by a combination of camera properties(optics, sensor properties) and selected training scenes, which should be similar to the set-up in practice.

Camera simulation: We used the Image Systems Engineering Toolbox (ISET)[13,14] to simulate digital cameras with novel CFAs to generate both the training and experimental scenes. Camera simulation technology allows us to compute both the sensor data and the scene statistics (CIEXYZ values). ISET models camera

systems by simulating scene radiance, optics, sensor electronics and image processing pipelines. Here, the camera lens is modeled as an f/4 lens with diffraction limited optics and focal length of 3mm. Figure 1 illustrates the arrangement and the spectral sensitivity of the color filters in the RGB/W CFA (left) and the sensor specifications(right).

Training scenes: We used multispectral scenes from a public database[15,16] for training. A specific set of training scenes can be used to optimize the linear operators for specific applications, such as rendering of black and white text. Here we focused on common natural scenes for consumer photography. All of our training scenes contain human faces. One example is shown in Figure 2. The illuminant for these scenes is D65 with spatially-varying intensity. The L^3 training can use different illuminants, in which case the learned filters include illuminant correction.

A pixel's image neighborhood is extracted from the sensor image and the pixel's CIEXYZ values are calculated by ISET. A patch size of 9×9 usually works well. Many thousands of sensor-patch and pixel-XYZ examples make up the training data.

2.2 Patch Classification

The overall relationship between sensor responses and scene statistics (XYZ values) is nonlinear, but it is locally linear. Therefore it is critical to classify training patches into small categories based on various criteria such that within each category there is a simple linear relationship. Defining these classes is part of the L^3 procedure, and one might decide to organize the classes and parameters somewhat differently depending on the sensor and intended application. Using automatic classification is also possible. Figure 3 demonstrates the image patch classification we used for the RGB/W CFA based on the following four criteria.

Pixel type: The RGB/W CFA has four possible central colors, corresponding to the four filter types. The Bayer RGB also has four possible central colors, because we separate the two green filters. In general there are as many pixel classes as there are pixels in the repeating block of the CFA pattern.

Pixel saturation: The high sensitivity of the W pixel extends the sensor's dynamic range. The cost of this feature is that the W pixel is saturated when the RGB pixels are well exposed. Saturated pixels provide little useful information and violate linearity. Therefore, the learning algorithm is designed to ignore saturated pixels by setting their filter values to zero. In this implementation, if any of the pixels of a particular color, say W, are saturated we set the filter values corresponding to all of the W pixels in the patch to zero. We divide the image patches into three types of pixel saturation classes: no saturation, W-saturation, and W&G-saturation.

Response level: The signal-to-noise ratio (SNR) of the sensor data varies greatly with the sensor response level. It is important to separate classes by sensor response level so that the linear operator can account for the large variation in SNR as a function of response level. The level is calculated by finding the average response within the unsaturated channels, and then averaging the channel responses. We create separate classes for different response levels, and the lower response levels are more densely sampled because the SNR changes rapidly in that range.

Spatial contrast: Finally, image patches are divided into flat and textured classes. Flat patches are relatively uniform image areas that contain only low spatial frequencies. Texture patches contain higher frequencies, typically near an edge. The classification is based on the sum of mean absolute deviations from the mean for all non-saturated channels. This method works well in high light levels, but suffers from unavoidable errors for dark images where measurement noise is similar to texture.

There can be a rather large number of classes. In the example illustrated in Figure 3, the classes account for the four central pixel types (RGB/W), pixel saturation (no saturation, W saturation and W&G saturation), mean response level (20 levels), and spatial contrast (uniform and textured). Hence, there are 4×3×20×2=480 classes. Approximately half are not used because of impossible combinations such as a patch with W saturation and low response level. About 700 operators are learned; this is equal to the number of patch classes (about 240) times the three output values (CIE XYZ).

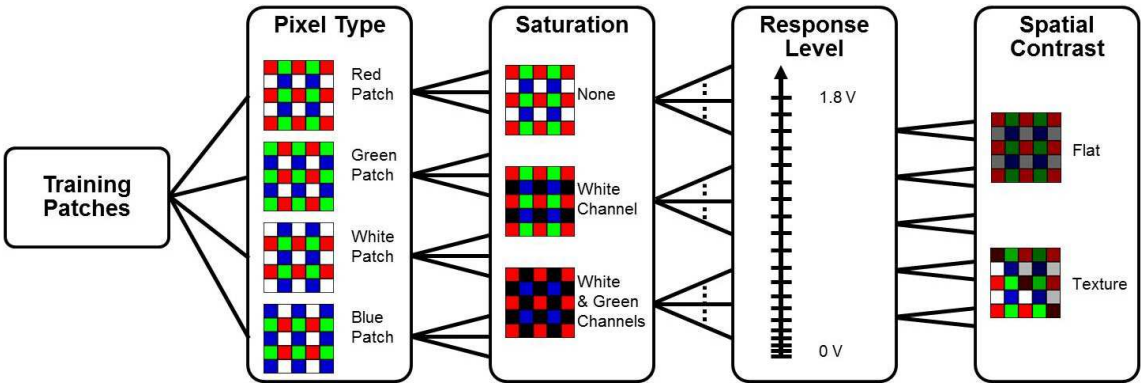


Figure 3. Image patch classification.

2.3 Linear Operator Derivation

Our goal is to estimate a linear mapping between the sensor measurements in the neighborhood of an image pixel (the patch) and each scene statistic at that pixel. Suppose the training data for a particular class have k noise-free patches, each containing $\sqrt{m} \times \sqrt{m}$ pixels. These are placed in the columns of a matrix $M \in R^{m \times k}$ and then corrupted by photon noise and sensor noise $N \in R^{m \times k}$. Each column of N is independent and identically distributed with mean 0 and autocorrelation A_n. We assume the noise is independent of M and D, which is not precisely true but simplifies the estimation. For each patch class, we find the linear transformation that best maps the noisy sensor measurements into the desired scene measurements $D \in R^{n \times k}$. We might define the optimal transformation as the one that minimizes the expected sum of squared errors between the desired output and the estimates, which is given by the Wiener filter[3] $W \in R^{n \times m}$:

$$W = DM^T(MM^T + kA_n)^{-1} \tag{1}$$

We find that under low-light conditions minimizing only the squared error produces output images with highly visible spatial noise[17] (Figure 6, top left). Thus, it is more pleasing to select a mapping that has a slight color bias but lower spatial variance and higher squared error. We can control the spatial noise by selecting an alternative filter using the formula in Equation 2:

$$W = DM^T(MM^T + \alpha kA_n)^{-1} \tag{2}$$

When α is greater than one, the spatial noise is reduced and the color-bias, measured by the sum of squared errors, is increased. This general form of linear estimation is referred to variously as ridge regression, Tikhonov regression, and Bayesian linear regression.[18]

When the linear transformation is implemented in XYZ space, reducing the variance has two different perceptual effects. The luminance signal becomes blurred and the chrominance signals become desaturated. When we minimize error and variance with respect to CIE XYZ values, the parameters do not separate these two effects. To separately control the luminance and chrominance effects, we solve for the linear transformation using different α values in a color space where the luminance channel and chrominance signals are explicitly represented.

We use an opponent color space WC_bC_r, optimized for rendering under the D65 illuminant, which is standard for sRGB. The luminance channel W is chosen as the projection of the D65 spectra to the three dimensional space spanned by the XYZ color matching function, which produces the metamer of D65 with minimal energy. The two chrominance channels are contained in the plane orthogonal to W in the XYZ subspace. In order to equally desaturate all hues, we choose $\alpha_{C_b} = \alpha_{C_r}$. Therefore any C_b and C_r vectors that span this orthogonal subspace give the same results. For specificity, we choose C_b and C_r by projecting CIE X and Z function to the orthogonal plane respectively. When using this color space for low light, desaturation goes to the white point given by D65. The derived basis functions for WC_bC_r are shown in Figure 4.

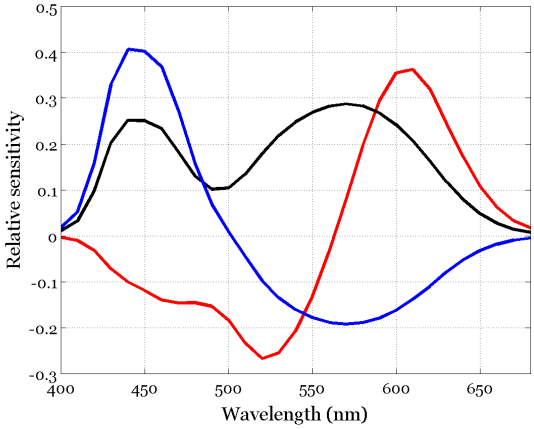


Figure 4. Basis functions for the opponent color space WC_bC_r optimized for rendering under the D65 illuminant.

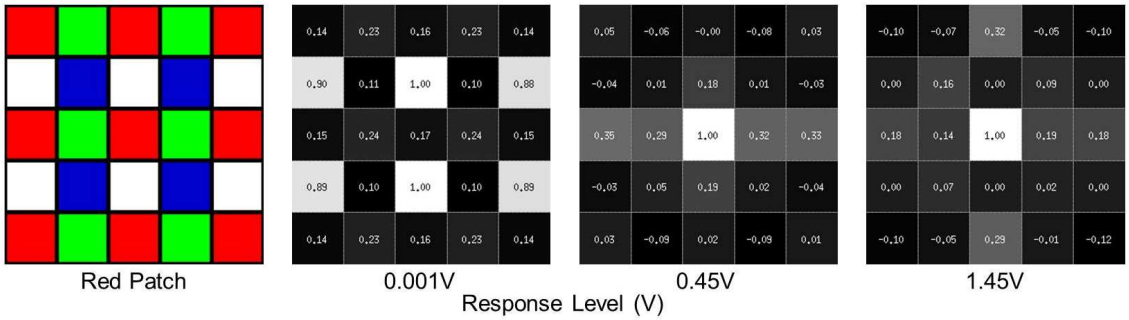


Figure 5. Learned linear operators to estimate X value at three different light levels for Red-centered patch.

3. EXPERIMENTAL RESULTS

3.1 Learned Filters

Figure 5 illustrates three linear operators optimized for a pixel patch centered on a red pixel. These operators are designed to estimate the CIE X value at the center red pixel for uniform classes at three different sensor response levels. At the lowest levels, no channel is saturated. The operator mainly relies on the W pixels because the RGB values have close to zero signal. The W pixel responses are combined across the entire patch to reduce noise. As the mean response level increases, RGB pixel responses become more reliable. The operator weights concentrate on the center red pixel, because its spectral information is most informative about the X value. To reduce noise and take advantage of between channel correlations, some weight is also allocated to the W and G pixels. At the highest response levels, the W pixels are saturated and thus have zero weight. The X value is calculated as a weighted sum of the R, G, and B pixels that are in the center of the patch.

3.2 Color Bias and Spatial Variance Tradeoff for Low Light

Figure 6 demonstrates the bias and variance tradeoff for low-light conditions. As luminance weight α_W increases vertically, the image becomes blurred while the luminance noise is reduced. As chrominance weights α_{C_b} and α_{C_r} increase horizontally, the image loses color saturation but chrominance noise is reduced. Note that once the luminance weight α_W becomes too high, the blur in textured regions is unacceptable (notice the eyes in the bottom right image in Figure 6). Since patches are clustered into flat and texture classes, it's possible to operate the tradeoff separately. Flat regions convey little high frequency information and can be extensively blurred,

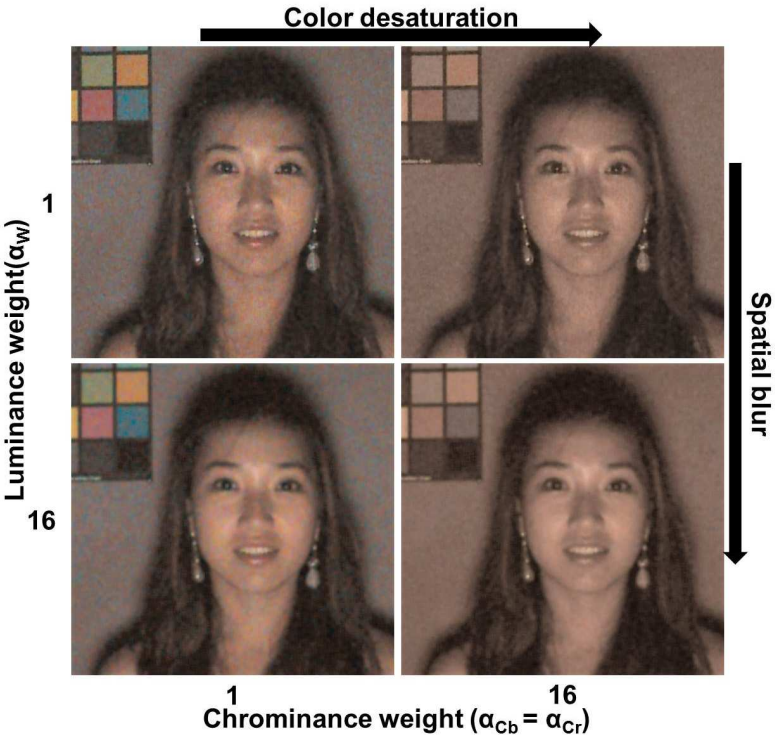


Figure 6. Color-bias and spatial-variance tradeoff for a low luminance scene. Scene luminance: $1cd/m^2$, exposure: 10 ms, aperture: f/4.

without reducing the image sharpness, while texture regions can be moderately blurred. In the resultant image (Figure 7, bottom left), the noise is reduced while the image details are preserved.

3.3 Bayer and RGB/W comparison with L^3

We compare the pipeline results based on training for a Bayer CFA (Figure 7, top) and for an RGB/W CFA (7, bottom). In this example, the RGB/W pipeline has less noise under low light level conditions. The advantage accrues because the W pixel improves the low light SNR, and the L^3 pipeline learns linear operators that take advantage of the W pixel. In the high light conditions the images are similar. A detailed examination across a wide array of challenging test targets may uncover some minor resolution differences.

3.4 Automatic Design for Various RGB/W Layouts

Most demosaicking algorithms have been designed to optimize the Bayer pattern, and the development of image processing algorithms for novel CFAs requires considerable effort. The L^3 algorithm automates the work of creating an image processing pipeline for new CFA designs. Figure 8 shows rendered images for three different RGB/W CFA layouts using the image processing pipeline automatically developed with the L^3 method. We hope that simplifying the design of the image processing algorithm will enable more experimentation.

4. CONCLUSIONS

We developed a local, linear, learned (L^3) algorithm that automates the process of creating an imaging system's processing pipeline. The L^3 algorithm can be applied to a large range of sensor designs and imaging applications. This algorithm reduces the time and effort involved in creating an image processing pipeline for novel sensor designs that will produce high quality images.

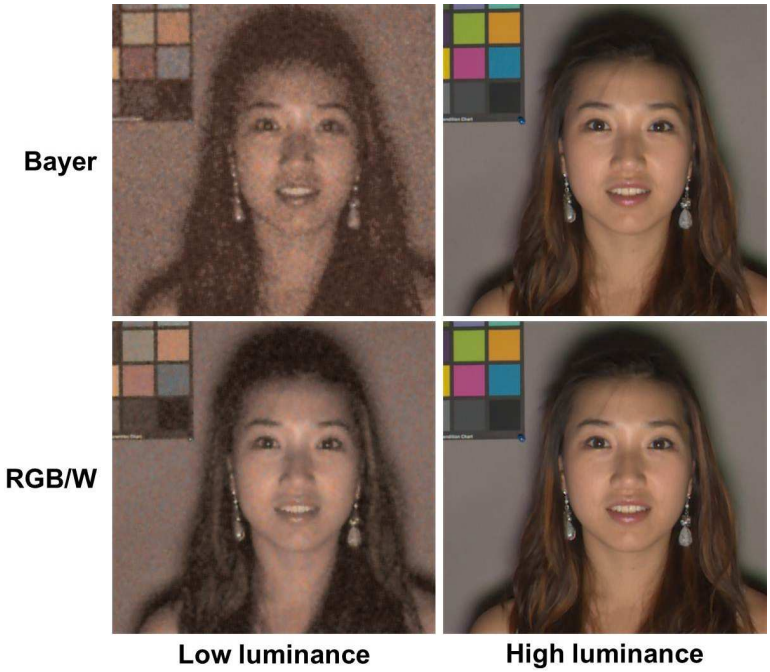


Figure 7. Rendered images for Bayer and RGB/W CFAs with the L^3 pipeline at low and high luminance. Separate bias and variance tradeoff on flat and texture regions is employed for RGB/W. Flat regions: $\alpha_W = 16$, $\alpha_{C_b} = \alpha_{C_r} = 4$; texture regions: $\alpha_W = 1$, $\alpha_{C_b} = \alpha_{C_r} = 4$.

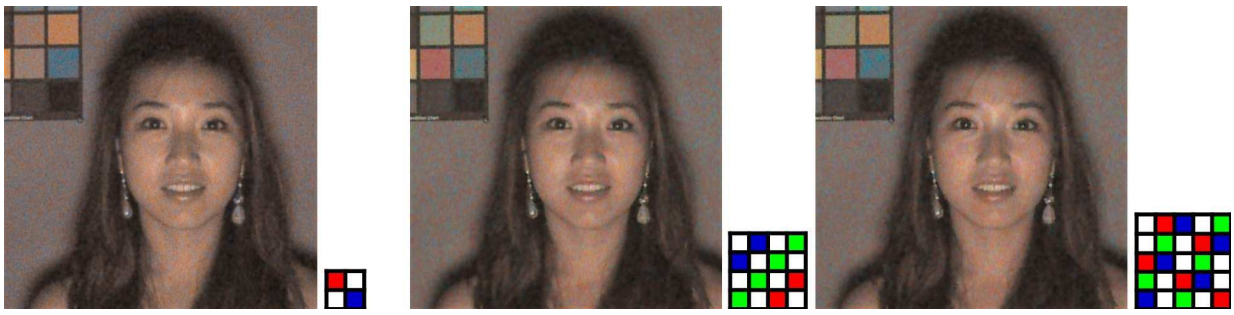


Figure 8. Rendered images for different RGB/W CFA layouts using L^3. From left to right: Kodak's,[8] Aptina's[19] and Wang et al.'s[6] CFA.

REFERENCES

[1] Bayer, B. E., "Color imaging array," (July 20 1976). US Patent 3,971,065.

[2] Lansel, S. and Wandell, B., "Local linear learned image processing pipeline," in [*Imaging Systems and Applications*], Optical Society of America (2011).

[3] Lansel, S. P., *Local Linear Learned Method for Image and Reflectance Estimation*, PhD thesis, Stanford University (2011).

[4] LANSEL, S., WANDELL, B., et al., "Learning of image processing pipeline for digital imaging devices," (Dec. 7 2012). WO Patent 2,012,166,840.

[5] Parmar, M. and Wandell, B. A., "Interleaved imaging: an imaging system design inspired by rod-cone vision," in [*IS&T/SPIE Electronic Imaging*], 725008–725008, International Society for Optics and Photonics (2009).

[6] Wang, J., Zhang, C., and Hao, P., "New color filter arrays of high light sensitivity and high demosaicking performance," in [*Image Processing (ICIP), 2011 18th IEEE International Conference on*], 3153–3156, IEEE (2011).

[7] Getman, A., Kim, J., and Kim, T.-C., "Imaging system having white-rgb color filter array," in [*Image Processing (ICIP), 2010 17th IEEE International Conference on*], 569–572, IEEE (2010).

[8] Kumar, M., Morales, E. O., Adams, J. E., and Hao, W., "New digital camera sensor architecture for low light imaging," in [*Image Processing (ICIP), 2009 16th IEEE International Conference on*], 2681–2684, IEEE (2009).

[9] Honda, H., Iida, Y., Egawa, Y., and Seki, H., "A color cmos imager with 4x4 white-rgb color filter array for increased low-illumination signal-to-noise ratio," *IEEE transactions on electron devices* **56**(11), 2398–2402 (2009).

[10] Honda, H., Iida, Y., Itoh, G., Egawa, Y., and Seki, H., "A novel bayer-like wrgb color filter array for cmos image sensors," in [*Electronic Imaging 2007*], 64921J–64921J, International Society for Optics and Photonics (2007).

[11] Honda, H., Iida, Y., and Egawa, Y., "High sensitivity color cmos image sensor with wrgb color filter array and color separation process using edge detection," in [*Proc. Int. Image Sens. Workshop*], 263–266 (2007).

[12] Luo, G., "A novel color filter array with 75% transparent elements," in [*Electronic Imaging 2007*], 65020T–65020T, International Society for Optics and Photonics (2007).

[13] Farrell, J. E., Xiao, F., Catrysse, P. B., and Wandell, B. A., "A simulation tool for evaluating digital camera image quality," in [*Electronic Imaging 2004*], 124–131, International Society for Optics and Photonics (2003).

[14] Farrell, J. E., Catrysse, P. B., and Wandell, B. A., "Digital camera simulation," *Applied Optics* **51**(4), A80–A90 (2012).

[15] Parmar, M., Lansel, S., and Farrell, J., "An led-based lighting system for acquiring multispectral scenes," in [*IS&T/SPIE Electronic Imaging*], 82990P–82990P, International Society for Optics and Photonics (2012).

[16] `http://imageval.com/scene-database/`.

[17] Barnhöfer, U., DiCarlo, J. M., Olding, B. P., and Wandell, B. A., "Color estimation error trade-offs," in [*Electronic Imaging 2003*], 263–273, International Society for Optics and Photonics (2003).

[18] Hastie, T., Tibshirani, R., Friedman, J., and Franklin, J., "The elements of statistical learning: data mining, inference and prediction," *The Mathematical Intelligencer* **27**(2), 83–85 (2005).

[19] `http://www.aptina.com/Aptina_ClarityPlus_WhitePaper.pdf`.