

Joint digital-optical design of imaging systems for grayscale objects

M. Dirk Robinson and David G. Stork

Ricoh Innovations
2882 Sand Hill Rd, Suite 115
Menlo Park, CA 94025-7054

ABSTRACT

In many imaging applications, the objects of interest have broad range of strongly correlated spectral components. For example, the spectral components of grayscale objects such as media printed with black ink or toner are nearly perfectly correlated spatially. We describe how to exploit such correlation during the design of electro-optical imaging systems to achieve greater imaging performance and lower optical component cost. These advantages are achieved by jointly optimizing optical, detector, and digital image processing subsystems using a unified statistical imaging performance measure. The resulting optical systems have lower $F\#$ and greater depth-of-field than systems that do not exploit spectral correlations.

Keywords: spectral correlation, spectral coding, extended depth-of-field, image processing, optimization, digital imaging, end-to-end design

1. INTRODUCTION

As shown in recent work,¹ optimizing digital imaging systems by simultaneously designing both the optical and digital subsystems provides significant advantages over the traditional sequential imaging system design methods. Specifically, simple digital filters can restore the spatial contrast using spatial correlation information about source objects.¹ Such a joint design approach is based on analyzing the entire optical imaging system as a linear system. In this approach, the imaging system is modelled as

$$\mathbf{y} = \mathbf{H}(\Phi)\mathbf{x} + \mathbf{n}, \quad (1)$$

where \mathbf{H} represents the system's point spread function, Φ the collection of the optical design parameters (lens thickness, curvatures, glass types, etc), \mathbf{x} the ideally captured digital image, and \mathbf{n} the noise inherent to photodetection. End-to-end or joint optimization of the optical *and* digital system is achieved by minimizing the predicted mean-square-error (MSE) as defined by

$$\mathcal{E} \{ \|\mathbf{R}\mathbf{y} - \mathbf{x}\|^2 \}, \quad (2)$$

where \mathcal{E} represents the statistical expectation operation and \mathbf{R} represents the digital filtering subsystem. The statistical expectation operation considers the correlation of the random noise as well as the spatial correlation of the object. Under this framework, both the optical design parameters Φ and the digital filtering subsystem \mathbf{R} are varied to find the MSE-optimal imaging system. Basic monochromatic imaging systems designed in this way achieve better contrast at improved signal-to-noise ratios (SNR) while relaxing the optical requirements in terms of aberrations.¹

We extend this concept to include specialized imaging systems in which the objects of interest possess strong spectral correlations. For example, barcode images and many paper documents are typically printed

Send correspondence to:
E-mail: {dirkr,stork}@rii.ricoh.com

in black and white. The spectral reflectance of these objects are nearly perfectly correlated at every spatial location. In other words, the radiance distribution of the objects is very similar across a range of wavelengths. In the traditional design approach, a system designer would typically utilize a photodetector having a single spectral filter applied to all pixels uniformly since the goals of the imaging system is not include extracting color information. For instance, the imaging system designer might choose a single-filter or monochromatic CMOS or CCD detector array and apply only an infrared (IR) filter to capture a range of wavelengths. In this paper, we explore an alternate approach which applies different color filters to different pixels to segment the spectrum even though the final image is to be grayscale. We call our approach *spectrally-coded grayscale imaging*.

In this paper, we introduce a new end-to-end design methodology that considers this spectral correlation information during the design of both the optics and the image processing subsystems. Our approach relaxes the requirements on the optical aberrations and enhances imaging capabilities, such as extending the depth-of-field. First, in Sect. 2 we describe how specialized image processing utilizes the spectral correlation found in grayscale objects to extract information across multiple color channels. Second, in Sect. 3 we describe how to jointly optimize both the optical and the digital subsystems to improve imaging performance and enable new capabilities such as extended depth-of-field imaging. We also illustrate this method for a simple-three lens imaging system. We conclude with some speculations on further directions of this joint design.

2. SPECTRALLY-CODED GRAYSCALE IMAGING

The Spectrally-coded grayscale imaging is a method of *encoding* image information in different parts of the spectrum and then *decoding* this information by a combination of spectral segmentation using optical filters and digital processing. Suppose that we begin with an ideal two-dimensional grayscale image represented by the vector \mathbf{x}_s . We assume that the object has the same spatial intensity distribution throughout a range of wavelengths (e.g., visible spectrum). Furthermore, we assume that the detector has a set of color filters to segment this spectrum. In this paper, we will explore the common tri-color filter which segments the visible spectrum into Red, Green, and Blue channel images, but the extension to other filters and a greater number of filters is straightforward. A similar concept has been applied in confocal microscopy to simultaneously image at multiple object depths.²

The captured image for each color channel is related to the unknown ideal grayscale image according to

$$\mathbf{y}_c = \mathbf{H}_c(\Phi)\mathbf{x}_s + \mathbf{n}_c, \quad c \in [R, G, B], \quad (3)$$

where Φ again represents the collection of optical design parameters associated with the imaging system such as the lens curvatures, glass types, or element spacings. The $\mathbf{H}_c(\Phi)$ term represents the sampled point-spread function for the c th color channel image. The term \mathbf{n} represents the additive noise having standard deviation σ . For the time being, we assume that the noise power is uniform over the different color channel images. The goal of the image processing is to estimate the unknown high-resolution image \mathbf{x}_s from the three noisy and blurry color images, represented by $\{\mathbf{y}_c\}$.

The MSE-optimal estimate of the ideal grayscale image \mathbf{x}_s is given by

$$\hat{\mathbf{x}}_s = \left(\sum_c \mathbf{H}_c^T(\Phi)\mathbf{H}_c(\Phi) + \sigma^2\mathbf{C}_x^{-1} \right)^{-1} \sum_c \mathbf{H}_c^T(\Phi)\mathbf{y}_c, \quad (4)$$

where \mathbf{C}_x represents the correlation matrix which captures the prior information about the spatial correlation or smoothness of the unknown signal. This form is a variant of the single-color Wiener filter model described in our previous work on digital-optical optimization.¹ Equation 4 shows that the estimate of the grayscale image is a weighted average of the sharpened color channel images, and thus this estimation strategy requires

only simple linear digital filters. Such filtering requires minimal computational resources. Estimating the grayscale image using the linear filtering of Eq. 4 produces images having an MSE given by

$$MSE(\Phi) = Tr \left[\left(\frac{1}{\sigma^2} \sum_c \mathbf{H}_c(\Phi)^T \mathbf{H}_c(\Phi) + \mathbf{C}_s^{-1} \right)^{-1} \right]. \quad (5)$$

In the traditional approach, a grayscale image is captured by means of a single color channel. To acquire a high quality image, the optical system must either use a very narrow-band spectral filter and sacrifice SNR due to lost photons, or ensure that color aberrations are minimized. The latter approach is common when the system has limited control over the radiance of the object. In such a case, the captured image is equivalent to Eq. 3 while having only a single channel. In this case, minimizing axial chromatic aberration or axial color becomes very important. Axial chromatic aberration describes the inability of the optical system to bring different wavelengths of light to focus at a single focal plane.³ For example, while the red wavelength image is well focussed at the focal plane, the blue wavelengths image is out of focus due to the dispersive nature of refractive lenses. When imaging broadband sources, the standard optical design method attempts to bring the collection of wavelengths to a single focus and thereby ensuring high contrast images across the visible spectral range of the object.

Minimizing axial color ensures that the effective point spread function, the point spread function after integrating over the range of wavelengths, yields sharp images. For such a case, the eigenvalues of the system matrix \mathbf{H}_s , locally the modulation transfer function (MTF) values, are large and so preserve image contrast throughout the range of spatial frequencies. The standard practice to minimize axial chromatic aberrations involves choosing lens materials with suitable dispersions to balance the aberrations. For example, in a triplet lens system, the first and third lens elements (positively powered elements) are made of Crown glasses (high Abbe numbers) while the second negative lens element is made of Flint glass (low Abbe numbers). In this way, the opposing chromatic aberrations are balanced.

When employing multiple color channels, the optical system for a grayscale image need not provide high quality images across the entire range of wavelengths, but the *collection* of color channels must provide all information needed to reconstruct the grayscale image. For example, if one color channel provides strong tangential contrast but weak sagittal contrast and another color channel provides strong sagittal contrast but weak tangential contrast, the combination of the two color channels can provide all the information required to estimate the grayscale image accurately. This approach thus relaxes the traditionally expensive constraints on optical subsystem performance and enables new classes of imaging systems.

Combining the information using Eq. 4 requires knowledge of the system's point-spread function for each color channel. In applications where the object distance d is not fixed, however, this depth information is difficult to obtain. In such a case, the image systems PSF matrix is a function of the unknown object distance d and expressed in $\mathbf{H}_c(\Phi, d)$. Combining the information across the multiple color channels requires knowledge of the object depth d .

Estimating object depth from a single image is a notoriously difficult problem to solve.⁴ What makes the problem difficult is the typical lack of a contrast signature revealing the object's depth. The effect of depth-dependent defocus manifests as a blurry or soft images, but attributing image softness to either defocus or merely a spatially-smooth object radiance map is difficult. One standard approach to estimating depth involves acquiring multiple images focused at different depth planes after which the depth can be estimated.⁴ The multiple images provide the necessary information to distinguish the image signal from the unknown depth-dependent defocus.

Analogous to the depth-from-defocus approaches,⁴ we propose optimizing the wavelength-dependent point-spread functions across multiple color channels to encode the object depth. In this way, we infer the object's depth by analyzing its associated axial color aberrations. The depth-dependent blur associated with

the multiple color channel images allows us to estimate the object depth. The maximum likelihood approach to estimating the object depth requires maximizing the function

$$J(d) = \left(\sum_c \mathbf{H}_c^T(\Phi) \mathbf{y}_c \right)^T \left(\sum_c \mathbf{H}_c^T(\Phi) \mathbf{H}_c(\Phi) + \sigma^2 \mathbf{C}_x^{-1} \right)^{-1} \left(\sum_c \mathbf{H}_c^T(\Phi) \mathbf{y}_c \right). \quad (6)$$

The accuracy of these methods hinges on the variation of the system matrix \mathbf{H}_c across the multiple color channels. We verify the depth-dependent variation of this sharpness measure with respect to the object depth in the next section. After estimating the object depth d , we obtain the system matrices $\mathbf{H}_c(\Phi, d)$ and can estimate the grayscale image via Eq. 4.

In the case of extended depth-of-field imaging, often the out-of-focus color channel provides little information about the grayscale image because the severe defocus eliminates the image signal. A simple approximation to Eq. 4 is to use only the sharpest color channel image. Similar to autofocus algorithms, we estimate image sharpness by filtering each color channel image with a high-pass spatial filter (e.g. standard Laplacian filter⁵) and compute the energy of the filtered images. A reasonable estimate of object depth d is obtained by fitting the relative sharpness of the different color channel images to a model. We find that this approach can provide reasonable estimates of the grayscale image as long as the axial color aberration is not too severe.

Enabling extended depth-of-field imaging requires that we provide good MSE performance over a range of object depths. The average MSE performance over the desired range of depths is computed by sampling the MSE at K different depth points and then averaging according to

$$P(\Phi) = \frac{1}{K} \sum_i MSE(\Phi, d_i), \quad (7)$$

where d_i represents samples within the depth range indexed by i . We choose the set of depths to correspond to equal depth ranges in terms of diopters. This can be approximated by evenly sampling in the depth-of-focus space. For our current implementation, we assume that the signal's spatial correlation is fractal in nature and so \mathbf{C}_x does not depend on the object distance. Other signals, such as bar codes, may have a correlation structure which changes according to the object depth according to the effective magnification at different object distances.

3. GRAYSCALE TRIPLET IMAGING SYSTEM

In the previous sections, we explained how prior information about spectral correlation allows the use of spectrally encoding depth information to extended the imaging system's depth-of-field. In this section, we analyze a triplet lens design based on this spectral-coding design principle. Specifically, we compare the performances of a triplet imaging systems designed using the traditional, single color filter architecture with one designed using the digital-optical design framework and spectral coding.

3.1. Triplet Specifications

The triplet specifications correspond roughly to a 40 degree field-of-view (FOV) VGA web-camera specifications and a 1/5" sensor as shown in Table 1. The triplet system comprises two glass spherical elements and a third plastic aspheric element which corrects field errors. The plastic element is defined by even-ordered aspheric surfaces up to the 8th order rotationally-symmetric polynomial. The 20 optical design variables include the spherical lens curvatures, the aspheric terms, the lens and air thicknesses, and the glass or plastic types.

sensor size	$\frac{1}{5}$ "
resolution	VGA
pixel pitch	$4.5 \mu\text{m}$
spectral range	0.47...0.63
focal length	4.75 mm
FOV	40°
Glass Types	Glass,Glass,Plastic
Max. Chief Ray Angle	16°
Track Length	$\leq 7 \text{ mm}$
Max. $F\#$	3.0
Max. Distortion	3.0 %

Table 1. Triplet system specifications

3.2. Traditional Triplet

First, we used a traditional methods to optimize the triplet system focussing three test wavelengths (0.48, 0.54, 0.62 nm) onto a single focal plane at a working distance of 750 mm. We achieved this by balancing glass types in order to minimize chromatic aberration. The merit function used to optimize the optical system was based on the RMS optical path difference (OPD) wavefront error. The upper left side of Fig. 1 shows this aberration minimizing design. After global optimization, the design form followed a traditional positive-negative-positive triplet form in crown-flint-crown glass. After optimization using the Schott catalog the glass types are N-FK51A and N-SF10. The plastic is a high index, low dispersion COC type plastic E48R.

We find that the design provides acceptable performance at $F\#$ 3.0. Below this f-number, the lens begins to suffer from a loss in contrast over the range of wavelengths. The curves below the graph in Fig. 1 show the field curvature for the three different RGB test wavelengths. The curve shows that the optical system does a reasonable job of focussing all three color channels onto a single focal plane. To achieve this, however, the system suffers from a bit of astigmatism.

3.3. Spectral Coding Triplet

In the second design approach, we assume that the sensor uses a standard set of RGB color filters to segment the spectrum. We optimized the optical design using a merit function based on the average MSE over a range of seven depth locations according to Eq. 7. We use a simple spatial covariance model where the covariance between neighboring pixels is given by 0.9^k where k is the spatial separation in pixels.⁵ We assume the systems's SNR is 40 dB. We achieve joint optimization of both the optical and digital processing subsystems by using the user-defined operand capability of Zemax, a commercially-available lens design software tool. We created a user defined operand to compute the predicted MSE according to Eq. 7. In this fashion, we can leverage the optimization capabilities of the Zemax lens design software.¹

The depths were chosen to uniformly sample the depth-of-focus range for a nominal object distance of 750 mm. The depth locations used during optimization were infinity, 2000, 1000, 750, 380, 255, 190, and 150 mm. Again, we performed global optimization over the optical design parameters using the traditional triplet as the starting design.

The resulting design is shown in the upper right side of Fig. 1. The triplet form again follows a positive-negative-positive design form. The glass types, however, are both high index flints N-SF6 and LASF32. The plastic is also a low Abbe number polycarbonate plastic which is much less expensive than the high Abbe number COC plastic used in the traditional design. The design achieves increased light gathering capacity (1.5X) over the traditional design. The curves in the bottom right of Fig. 1 show the field curvature for the three color wavelengths. The field curvature plots show the strong separation between the three wavelengths focal plane due to strong axial color aberrations in the spectrally-coded system.

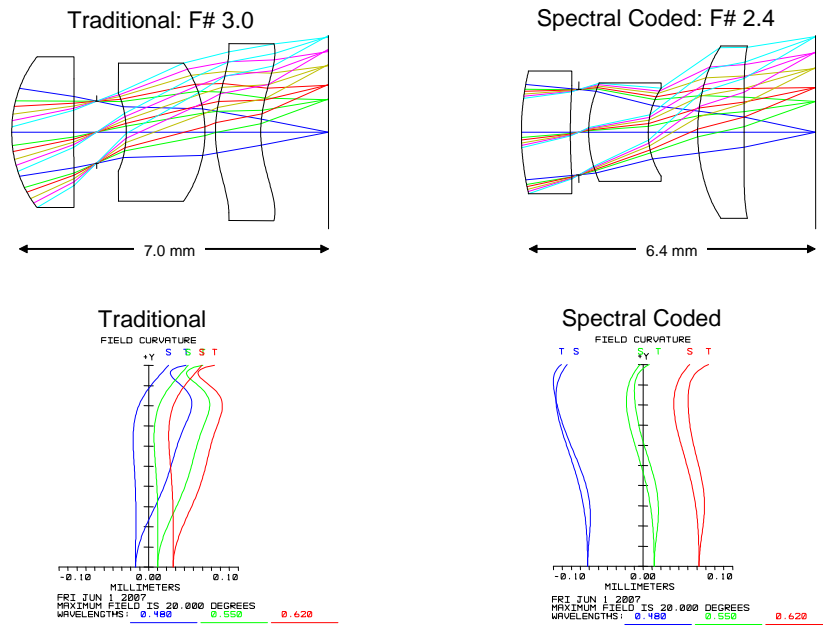


Figure 1. The lens on the left represents the traditional optical design approach which focusses all three wavelengths at a single focal plane. The bottom left curve shows the field curvature plots for the three test wavelengths (RGB). The traditional optical system brings the three color planes into focus at nearly the same focal plane. The system does, however, suffer from a bit of astigmatism. The spectral-coding design on the right achieves increased light gathering capacity (1.5X). The field curvature plots show the strong separation between the three wavelengths focal plane due to strong axial color aberrations.

3.4. Depth-of-field Comparison

We compare the effective depth-of-field performance of our traditional, single channel, triplet and our spectrally-coded triplet. In grayscale imaging the final image quality depends on the spectral sensitivity of the detectors. The top graphs of Fig. 2 show the spectral sensitivities for the single channel detector (left) and the spectrally-coded system (right). Both systems cover the same spectral range and reflect the typical sensitivities of commercially-available sensors combined with IR cutoff filters.

The curve in the bottom left of Fig. 2 shows the through-focus polychromatic MTF for the traditional system using a set of nine wavelengths weighted by spectral sensitivity of the single channel sensor. The through-focus polychromatic MTF shows the MTF at 50 lp/mm spatial frequency for the on-axis field point. As we would expect, the MTF falls off as we move away from the focal plane due to defocus. The system shows a maximum depth-of-focus of about 120 μm . While this design provides reasonable quality at the proper focal distance, the limited depth-of-field shows that the imaging system will provide defocused images at a working distance of about 250 mm which would require a focal shift of about 150 μm . Unfortunately, a fixed-focus lens system designed under this constraint will work only within a particular depth range around the chosen object distance. Furthermore, the depth-of-field decreases with decreased $F\#$ (and hence light sensitivity) creating an undesirable tradeoff.

The curves on the bottom right show the polychromatic MTF curves for the spectrally-coded system. The MTF again reflects a polychromatic average over different sets of nine spectrally-weighted wavelengths. As expected, the different color channels focus at different depth planes. The depth-of-focus for the spectrally-coded system is extended to about 240 μm . Over the focal range, however, at least one of the color channels

provides strong contrast. Furthermore, for every depth plane, at least one of the wavelengths has significantly poor contrast suggesting the ability to infer object depth using color channel image sharpness. Also, the spectrally-coded triplet has an increased light gathering capacity of $F\#$ 2.4.

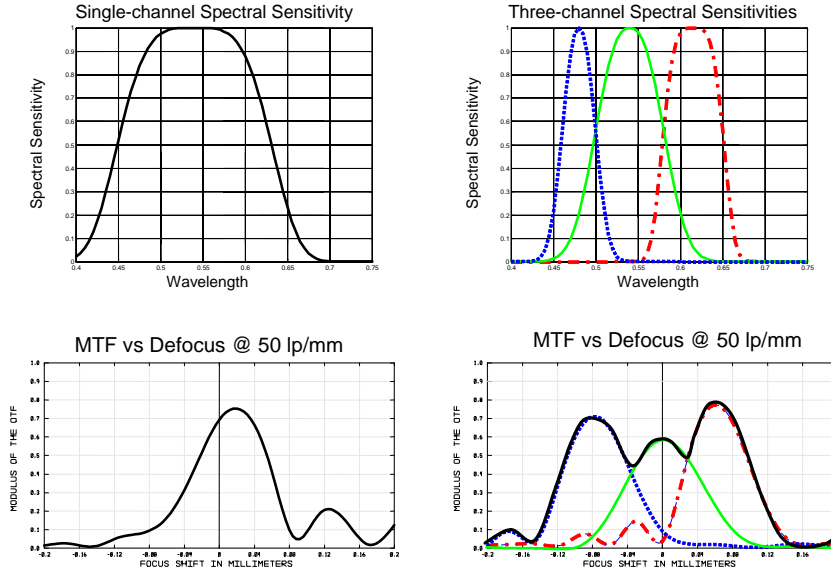


Figure 2. The top left curve shows the spectral sensitivity of the single-channel sensor used in the traditional imaging system. The bottom left curve shows the polychromatic MTF of the traditional triplet system using nine spectrally-weighted wavelengths for the on-axis field point. The system shows reasonable performance within about $\pm 60 \mu\text{m}$ from the nominal focus position. The top right curve shows the spectral sensitivity for the three color channel spectrally-coded triplet system. The three curves on the bottom right show the polychromatic MTF of the using different spectrally-weighted wavelengths according the color channel sensitivities for the on-axis field point. The three curves reveal the three different color focal planes. The thick line shows the effective MTF combining the best MTF among the three color channels. The system shows that at least one of the color channels is in focus within $\pm 120 \mu\text{m}$ of the nominal focal distance.

3.5. Image Simulation Results

We simulated images produced by these two systems using our imaging system simulation tool.¹ Our image simulation tool is similar to that described in⁶ with the extension of adding multi-spectral weighting according to the pixel spectral sensitivity. We use a traditional Air Forces resolution target as our simulated object, a binary target having either uniformly broadband radiance or none simulating a perfectly correlated object. When simulating the captured images, we use three spectral samples per color channel to simulate the spectral integration of the detector.

Figure 3 compares portions of the target at two different object depths. We show a cropped portion of the image so as to reveal the resolution properties of the image. The leftmost image column shows the target at 1.5 meters (top) and 130 millimeters (bottom) for the traditional single channel imaging system. The image shows good contrast at 1.5 meters, but very poor contrast at 130 millimeters due to depth-of-field limitations. The images in the second and third columns show the same images for the Red channel (middle) and Blue channel (right) at the two object depths. At 1.5 meters, the red image shows almost equivalent contrast to the single channel imaging system while the blue image shows very low contrast. Alternatively, the red image shows very low contrast when the object is at 130 millimeters, while the blue image shows sharp contrast. These images visualize the contrast predicted by the through focus MTF shown in Fig. 2.

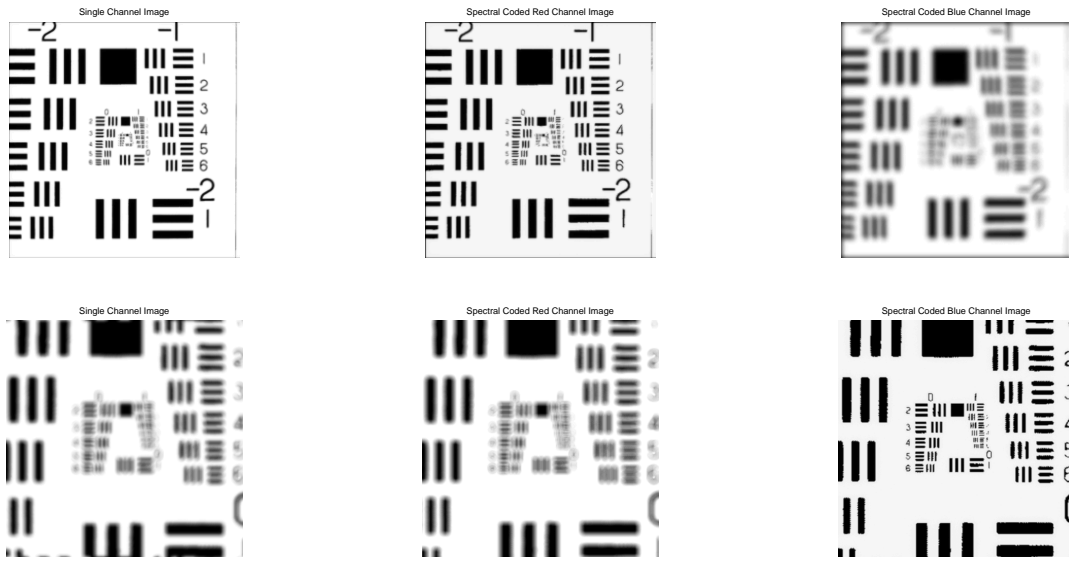


Figure 3. The left column shows the images of a resolution target produced by the traditional single channel imaging system for an object located at 1500 mm (top) and 130 mm (bottom). The system shows good contrast at 1500 mm but very low contrast at the short working distance of 130 mm due to limited depth-of-field. The second and third columns shows the red and blue channel images respectively. The red image shows good contrast at 1500 mm while the blue image shows good contrast at 130 mm.

To evaluate the ability to discern object depth using the spectrally-coded imaging system, we simulated imaging the resolution target located at 2 m, 1 m, 750 mm, 380 mm, 250 m, 190 mm, and 130 mm. We then applied a simple Laplacian sharpness filter to each of the color channel images and compared the relative magnitude of the filtered images for a small patch near the center of the image. To compute the relative magnitude, we first integrate the energy in the filtered images for a 50×50 pixel patch at the center of the image for three color channels. Then, we normalize the three color channel values so that the sum of the energies equals one. This approximates the percentages of the total high frequency image energy present in the three different color channel images. Figure 4 compares the relative sharpness for the three color channel images as a function of object depth. The images in the left and right columns show the magnitude of the Laplacian filtered images for the object located at 130 mm and 1.5 m respectively. The curves demonstrate the clear relationship between object distance and relative sharpness. The simplest application of the relative sharpness is to find the sharpest image over the collection of image planes to use as the captured image.

4. CONCLUSIONS AND FUTURE WORK

We presented a novel framework for analyzing and designing imaging systems for the class of grayscale objects where the different spectral bands have strong correlation. We demonstrated how such correlation information enables the system designer not only to relax strict requirements on optical aberrations, but also enables new imaging capabilities such as extended depth-of-field imaging through spectral coding. We used our new design philosophy to design an extended depth-of-field triplet imaging system verifying the increased depth-of-field through image simulation. Finally, we highlighted the additional advantage of this new approach to building simple object depth estimation using simple filter-based sharpness measures. The cost of such image processing is low enough to make this approach attractive for grayscale imaging systems.

The current work suggests numerous future research directions. In this report, we ignored the loss in

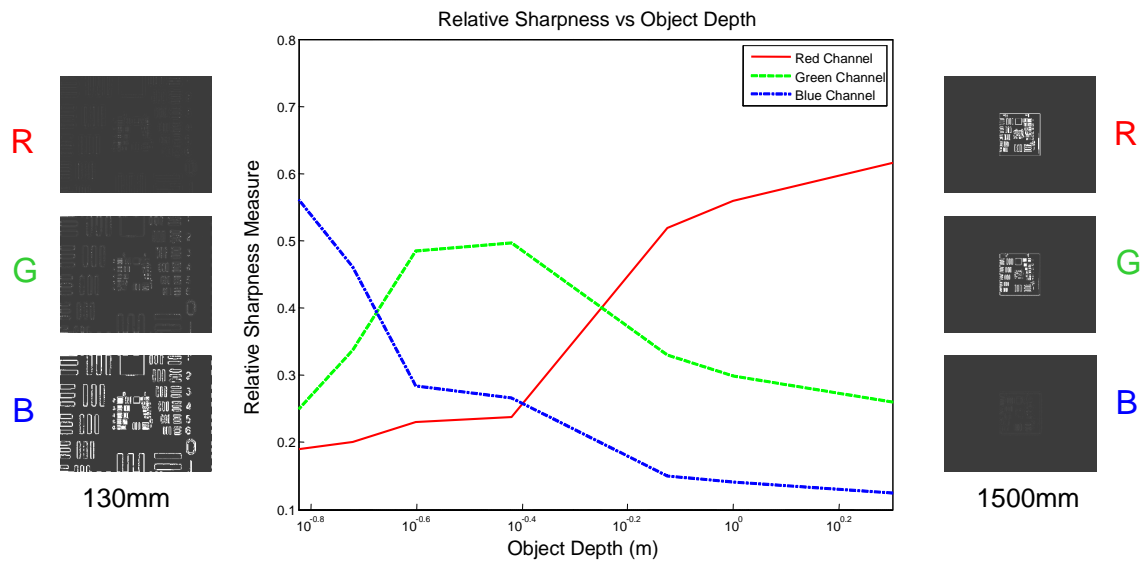


Figure 4. The curve shows the relative sharpness for at the center of each color channel image versus the distance of the object from the camera. The sharpness was computed as the average magnitude of the color channel filtered by a Laplacian filter. The column of images at the left and right visualize the magnitude of the Laplacian filtered images at 130 mm and 1.5 m respectively. The curves demonstrate the clear relationship between object distance and the computed sharpness metric.

spatial resolution due to spatial multiplexing of the color filters. The most practical application of the spectral coding will undoubtedly required such spatial multiplexing. Future research could address the processing required to restore resolution by combining the multiple color channel sub-images. In our work, we focussed on strongly correlated objects such as bar codes or grayscale documents. A general multispectral analysis of general images could reveal spectral correlations, albeit weaker, in general images. Future work might address methods for leveraging this weak spectral correlation information when designing general purpose imaging systems to improve $F\#$ and increase depth-of-field.

REFERENCES

1. D. G. Stork and M. D. Robinson, "Theoretical foundations for joint digital-optical analysis of electro-optical imaging systems," *Applied Optics*, April 2008.
2. H. Tiziani and H. Uhde, "Three-dimensional image sensing by chromatic confocal microscopy," *Applied Optics* **33**(10), pp. 1838–1844, 1994.
3. J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New York, NY, second ed., 1986.
4. S. Chaudhuri and A. Rajagopalan, *Depth from defocus: A real aperture imaging aproach*, Spinger Verlag, 1999.
5. A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, Englewood Cliffs, New Jersey, 1 ed., 1989.
6. P. Maeda, P. B. Catrysse, and B. A. Wandell, "Integrating lens design with digital camera simulation," *SPIEProceedings SPIE Electronic Imaging, San Jose, CA* **5678**, pp. 48–58, February 2005.