

Composite Target for Camera-Based Document/Object Capture System*

Yang Lei[▲]

School of Electrical and Computer Engineering, Purdue University West Lafayette, Indiana
E-mail: lei3@purdue.edu

Peter Majewicz, Kurt R. Bengtson, and Lisa Li

Hewlett-Packard Company, Boise, Idaho

Jan P. Allebach[▲]

School of Electrical and Computer Engineering, Purdue University West Lafayette, Indiana

Abstract. *The use of mobile phone camera technology in systems for capture of documents or three-dimensional objects is becoming increasingly popular. Compared to conventional flatbed or sheetfed scanner technology, such systems pose special image quality concerns due to the fact that the underlying camera and illumination technologies are inherently low-cost. Thus, each unit is designed to operate closer to the threshold of acceptable quality. So there are higher levels of noise and greater non-uniformity across the image plane, as well as greater unit-to-unit variability. Consequently, there is a need for a solution that can provide a detailed assessment of the spatial non-uniformity of key image quality metrics within the field of view of the image capture system, and which can be robustly and efficiently computed. This article proposes a specially designed composite target and an automatic analysis tool for image quality inspection of camera-based document/object capture systems. The target and analysis tool can be used during product development, as well as on the manufacturing line for real-time inspection.*

©2013 Society for Imaging Science and Technology.

[DOI: 10.2352/J.ImagingSci.Technol.2013.57.3.030404]

INTRODUCTION

With the nearly universal inclusion of one or more image capture devices in mobile phones, low-cost digital cameras have become a pervasive part of the technology landscape. These devices have found a wide range of applications in addition to their use in mobile phones. One such application is as a replacement for the traditional flatbed and/or sheetfed scanner found on top of typical multifunction printers, also known as all-in-one devices, that provide the capability to scan, copy, and print. Figure 1 shows an example of such a product. The top surface of the device contains a platen or platform on which documents or objects may be placed for capture. The arm above the platen contains a camera and one or more sources of illumination. The illumination

sources, which may be LEDs, are turned on one at a time to provide illumination of the object(s) on the platen from several different angles. A separate image is captured under each condition of illumination. The captured images can then be fused by image processing to reduce the glare and shadows that would be visible in an image captured with just one condition of illumination.¹ Compared to a traditional multifunction printer with a flatbed and/or sheetfed scanner, such a device offers greater convenience and/or flexibility when copying single page documents or multiple pages from a bound publication. In addition, this device enables an entire new range of applications, since 3D objects may also be captured.

However, there are a number of challenges associated with the design and manufacture of a camera-based MFP (Multi-Function Printer). Users who are accustomed to using traditional flatbed or sheetfed scanners will expect image quality that is comparable to that provided by such capture devices. However, it is difficult for a low-cost and highly miniaturized camera designed for use in mobile telephones to provide the same uniformity of image quality throughout its field of view as that offered by a flatbed or sheetfed scanner. In addition, whereas a flatbed or sheetfed scanner can provide a highly controlled and very uniform source of illumination for the document being imaged, the highly localized lamps mounted in the top arm shown in Fig. 1 cannot illuminate the platen with this degree of uniformity. This necessitates an approach to capture system image quality assessment² that provides a more detailed view of the spatial variation of image quality across the field of view of the capture system. Due to the low cost of the capture system components and the less robust configuration of these components than is typically the case for a flatbed or sheetfed scanner, greater unit-to-unit variability will be experienced during manufacture as well. Thus, it is desirable to individually inspect each unit during the manufacturing process to assure that it meets all specifications for image quality throughout the field of view of the capture system.

* Research supported by the Hewlett-Packard Company. Research partially conducted while Yang Lei was an intern with the Hewlett-Packard Company in Boise, Idaho.

[▲] IS&T Members.

Received Jan. 9, 2013; accepted for publication Jul. 29, 2013; published online May 1, 2013. Associate Editor: M.R.V. Sahyun.

1062-3701/2013/57(3)/030404/14/\$20.00

Several testing targets have been proposed to assess image quality for mobile phone cameras. The target proposed by Artmann and Wueller³ is relatively complex, consisting of many color and gray patches, radial bar charts, a natural photograph, pictures of different textures, and other features. Wueller⁴ proposed a test chart with 20 gray levels, combining the testing chart of ISO 14524 and the noise patches of ISO 15739. Williams and Burns⁵ proposed a test target with duplicate slanted-edge features. Baxter and Murray⁶ also described the use of ISO Visual Noise for assessment of visual quality provided by cell phone cameras. The targets and analysis methods that have been proposed for assessment of image quality for a low-cost camera-based capture system, such as that shown in Fig. 1, do not provide a detailed picture of how image quality varies across the field of view of the capture system. In addition, due to the complexity of the targets that have been proposed, these systems may not be well suited for fast, highly automated image quality assessment that is simple to complete with minimal operator training, and which is robust to the relatively hostile environment of the factory floor.

In this article, we propose a novel composite target that consists of an array of identical subpatterns that are repeated periodically across the field of view of the capture system. Each subpattern has a simple structure that supports the robust computation of several key image quality metrics. The entire target is printed on a single sheet of paper. We refer to it as a composite target because it consists of an array of identical subpatterns. The periodic repetition of these subpatterns across the field of view provides a highly detailed picture of the spatial variation of each image quality metric within the field of view. We also describe in detail the image analysis procedures that are applied to the captured image of the target to locate each subpattern, extract the regions of interest within it, and compute the key image quality metrics for that subpattern. Finally, for each key image quality metric, we define summary statistics that characterize the overall variation of that metric within the field of view. All these processing steps are designed to be computationally efficient, and robust to unit-to-unit variations in the image of the composite target that is captured by the camera system. A brief description of our work can be found in Ref. 7. In this article, it is illustrated with much more details.

The rest of this article is organized as follows. In the next section, we discuss the design of the composite target. Then, we provide an overview of the embedded camera capture pipeline and calibration procedure. Next, we describe the algorithms used to extract the various regions of interest from within each subpattern of the composite target. Following that, we describe the processing required to compute each of the key metrics of interest, and show experimental results for the target system for which the composite target and analysis procedure were developed. Finally, we provide a summary and conclusions at the end of the article.



Figure 1. Example multifunction printer or all-in-one device that replaces the flatbed or sheetfed scanner by a low-cost camera and illumination system, both mounted on an arm above the platen or capture platform.

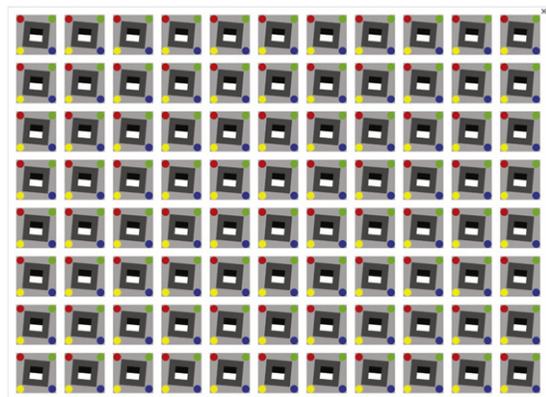


Figure 2. Digital frame of the composite target.

TARGET DESIGN

In the composite target proposed in this article, shown in Figure 2, one subpattern repeats on an 8×11 grid throughout the page, which covers the largest common area of A4 and Letter size pages. Designed around the slanted-edge target,⁸ each subpattern includes four gray level regions, four color regions, and four slanted edges.

We summarize the design of each subpattern as follows.

1. Four slanted edges between dark and light gray regions are used to compute the modulation transfer function (MTF). The angle between the slanted edge and the horizontal axis or vertical axis is -5° , as specified in ISO 12233.⁹
2. Four gray level (0, 33, 66, and 100% nominal reflectance) patches are designed to measure the accuracy of tone reproduction. These levels were chosen to span the entire reflectance range that could be printed with the chosen substrate and inkjet printer, in order to capture as much of the dynamic range of the sensor as possible, and to allow robust segmentation of the target into separate regions.
3. Four color dots (red, green, blue, and yellow) are designed to measure color accuracy.
4. White and black regions and four color dots are used to measure noise.

5. The centroid of each subpattern on the 8×11 grid is used to measure geometric distortion.

6. The variation of each of the five types of statistics discussed above, from subpattern to subpattern, is used to measure the uniformity of the capture system performance throughout its field of view.

The complete design of the composite target is guided by the need to support effective and autonomous machine analysis of the captured target. First, the structure of each subpattern is identical, and they are all arranged on a uniform grid. Second, within each subpattern, the elements are arranged in a regular manner. They are well separated from each other, and have constant values that generally contrast well with those of neighboring elements. Third, each element is sufficiently large so that when captured, a portion of its region that is well away from the edges of that region can be reliably extracted, and this portion contains a sufficient number of pixels to allow robust calculation of the desired metrics. Fourth, each element of the subpattern serves multiple purposes in the image analysis. Fifth and finally, the entire computational process must be sufficiently efficient to support the analysis of high resolution (2970×2159) pixel color images in a production environment.

In practice, the printing of the target and the handling of it after it is printed should be done very carefully. In our application, all hard copies were printed at 1200 dots per inch with the same high-quality HP Designjet 130nr Inkjet printer (Hewlett-Packard Company, Palo Alto, CA 94304) using identical settings. Every hard copy is attached to a thin plastic backing board to make sure it remains flat during the inspection process. The camera in the target system has an 8 megapixel (3280×2460 pixels) CMOS sensor. The lamps in the capture system itself are used to illuminate the target during the capture. When captured from the platen of our target system, each subpattern is approximately 227×227 pixels in size, and the four color dots at the corners of the subpattern have a diameter of approximately 44 pixels.

OVERVIEW OF THE EMBEDDED CAMERA CAPTURE PIPELINE AND CALIBRATION PROCEDURE

Our composite target analysis tool essentially treats the image capture system as a black box from the input hardcopy target shown in Fig. 2 to the captured image of the composite target. However, it is important to know the steps that are part of the image processing pipeline embedded in the capture system in order to understand the impact that these steps have on the results of the image quality analysis that we will describe later in this article. The sensor produces a Bayer matrix encoded image.¹⁰ Each of the red, green, and blue color channels is calculated by extracting the corresponding pixels of the captured image based on the Bayer matrix. Then a gradient-corrected bilinear interpolation method¹¹ is applied to fill in the missing pixels of each color channel and to obtain a full resolution color image. These data are then processed through a color management pipeline to generate an image in the sRGB¹² color space.

During the calibration process, the image quality concerns mentioned at the beginning of this article are corrected in different ways. For example, the camera used in the capture system has a relatively wide field of view, which results in greater geometric distortion in the captured image. This distortion is estimated using pre-production data, and the same correction is part of the embedded image processing pipeline in each unit built on the manufacturing line. Another example is the non-uniform lighting from the three LEDs that are used to sequentially illuminate the field of view. There are hot spots and shading in the image captured with each different LED turned on. To correct this, the white platen is used to find the distribution of the light from each LED, and the distribution is used to compensate the non-uniform illumination. Finally, the image is cropped to remove regions outside the document/object.

REGION EXTRACTION

The location of the regions of interest within each subpattern, including the four gray level patches and the four color dots, must be precisely extracted so that we can have input information that is sufficiently accurate for the next step, which is the evaluation of the key metrics. Based on the observation that the white rectangle within each subpattern contrasts highly with its surroundings, we extract all the white rectangles first. Once we have the location of each white rectangle, the location of the remaining regions of interest in the subpattern containing that white rectangle can be determined by their offset from the centroid of the white rectangle.

We first compute the reflectance of the captured RGB image. The target product is calibrated to generate sRGB.¹² To represent the gray level in terms of luminance, i.e., CIE Y, we transform the gamma-corrected sRGB values to linear sRGB values first according to Eq. (1).¹³ The input I_{sRGB} is any one of the three sRGB channel images, which are the red channel image R_{sRGB} , the green channel image G_{sRGB} , and the blue channel image B_{sRGB} . Then we transform from linear sRGB to CIE XYZ based on Eq. (2).¹³ The reflectance is computed by normalizing the luminance CIE Y by 255, which represents white.

$$I_{\text{linear}} = \begin{cases} \frac{I_{sRGB}/255}{12.92}, & (I_{sRGB}/255) \leq 0.03928, \\ \left(\frac{I_{sRGB}/255 + 0.055}{1 + 0.055} \right)^{2.4}, & (I_{sRGB}/255) > 0.03928, \end{cases} \quad (1)$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R_{\text{linear}} \\ G_{\text{linear}} \\ B_{\text{linear}} \end{bmatrix}. \quad (2)$$

A preset global threshold of 50% reflectance is used to isolate each white rectangle from its surroundings. However,

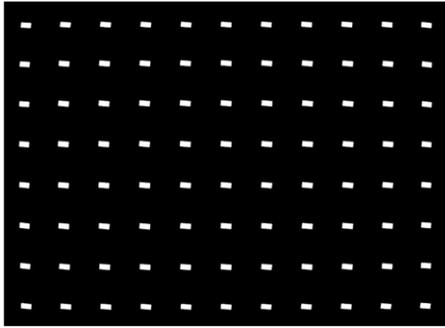


Figure 3. White rectangles extracted from the image captured by our target system. In this binary image, the bright areas correspond to the locations of interest, which are the white rectangles. We call this the white rectangular region map.

many irrelevant regions remain after this thresholding step, such as the white margins of the target. So we perform a connected component analysis on the resulting image, and record the number of pixels in each connected component. The nominal number of pixels in the white rectangle A_{white} is determined first. Then we set the lower threshold $T_{\text{lower}} = 0.85A_{\text{white}}$ and the upper threshold $T_{\text{upper}} = 1.15A_{\text{white}}$. If the actual number of pixels of a connected component falls between these two thresholds, we accept the corresponding region as a white rectangle. Otherwise, we reject it. These criteria are determined based on the data of pre-production units, and can be adjusted if needed. The result of extracting the white rectangular regions is shown in Figure 3.

Extract Gray Regions

Once we have the locations of all the white rectangles, it is easy to crop all subpatterns from the target since the size of each subpattern is relatively consistent throughout the page. The color and grayscale versions of the subpattern in row 4, column 1 of the composite target page captured using our target system are shown in Figures 4(a) and (b), respectively. Then, for the other three gray level regions, two methods are used to extract them: location approximation based on offsets and a thresholding step. The results from these two methods are combined pixel-by-pixel using a logical AND operation. Two post-processing steps are then used in order to remove irrelevant areas and noise. Finally, by applying safe region erosion to the boundary of the region maps, pixels that are actually in the region are selected. As an example, we take the dark gray region of the cropped patch in Fig. 4(b) to show the results of the steps mentioned above, which are outlined in Figure 5. In the remainder of this subsection, we discuss these steps in detail, and illustrate the results of each step in Fig. 4. Note that Fig. 4(c) and (e)–(h) are binary masks.

The location approximation method is quite straightforward. Based on the relative displacement between each region of interest and the centroid of the white rectangle, we set a rectangular area where that region of interest is most likely to be located. For example, each black rectangle is located right above a white rectangle. Fig. 4(c) shows the location approximation results for the dark gray region in

the subpattern in Fig. 4(b). As we can see from Fig. 4(c), the results from the location approximation may contain other regions, such as white and black rectangles in this case. Thus, the thresholding step is necessary. However, the uniformity of the captured image is not perfect. Also, the image quality varies from unit to unit. Experiments show that preset global thresholds do not work well for all units, not even within the entire image of the test page captured by a single unit. Therefore, the thresholds must be determined adaptively for each subpattern. The key idea of our solution is to first perform local histogram equalization within each subpattern, as in Fig. 4(d), then apply preset global thresholds, the result of which is shown in Fig. 4(e). This is equivalent to applying locally adaptive thresholds. Then a logical AND operation is performed between Fig. 4(c) and (e) to find the common region, which is shown in Fig. 4(f).

As we can see from Fig. 4(f), the result still contains some irrelevant areas, such as the black rectangular region, and noise. Therefore, we perform a logical XOR operation between Fig. 4(f) and the extracted black rectangular region. Also, a morphological opening operation is applied to remove noise. A hexagon-shaped structuring element with sides that have length 5 pixels is used for this operation. Equivalently, the structuring element can be viewed as a disk with radius 7.7 pixels that includes all pixels with distance less than or equal to 7.7 pixels from the center of the structuring element. These are the post-processing steps for dark gray region extraction. Fig. 4(g) shows the result of post-processing.

To guarantee accurate calculation of the key metrics, safe region erosion is applied to the dark gray region map that we obtained according to the post-processing procedure described above. We perform a morphological erosion operation on each dark gray region, in order to assure that the key metrics for each dark gray region are only calculated using pixels that actually belong to that region. A disk-shaped structuring element is used in the morphological erosion process. This structuring element consists of all pixels whose centers are no greater than r pixels away from the origin. The radius r is determined according to the width W_{gray} of the thinnest part of the region map, so that the width will become $0.6W_{\text{gray}}$ after morphological erosion. In other words, we set $r = \frac{(1-0.6)W_{\text{gray}}}{2}$. Different W_{gray} values are used for different region maps. For our system, $W_{\text{dark gray}} = 34$ pixels. Fig. 4(h) shows the result of safe region erosion for the dark gray region in the selected subpattern. The result for the dark gray regions of all subpatterns after safe region erosion is shown in Figure 6.

Extract Color Dot Regions

In this subsection, we describe the procedure for extracting the color dot regions in the subpattern shown in Fig. 4(a). Figure 7 contains a block diagram for the overall procedure. Figure 8 illustrates the results from each individual step. Figure 9 shows the final result. These three figures are analogous to Figs. 5, 4 and 6 for extraction of the gray regions

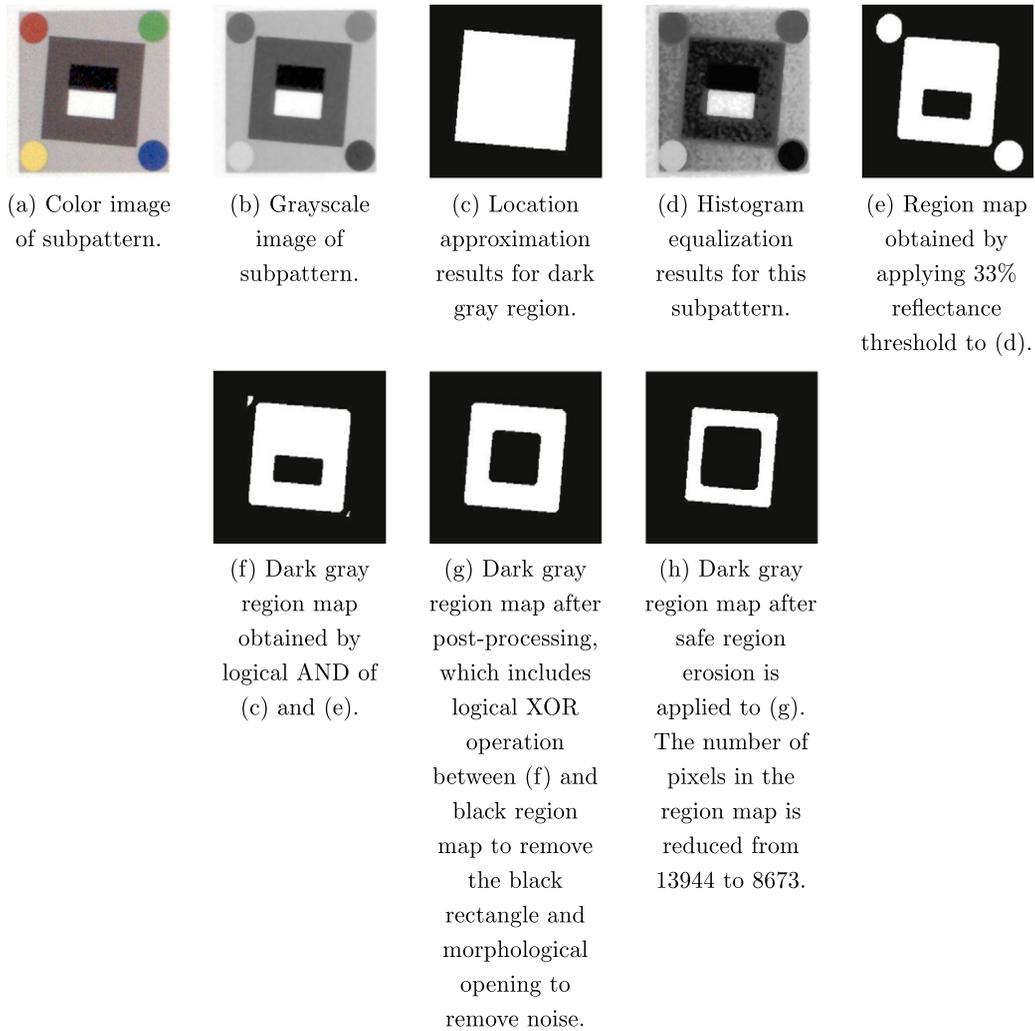


Figure 4. Illustration of the steps for determining the gray region maps to be used for extracting the gray regions. The specific example shown here is for the dark gray region. Parts (c) and (e)–(h) are binary masks.

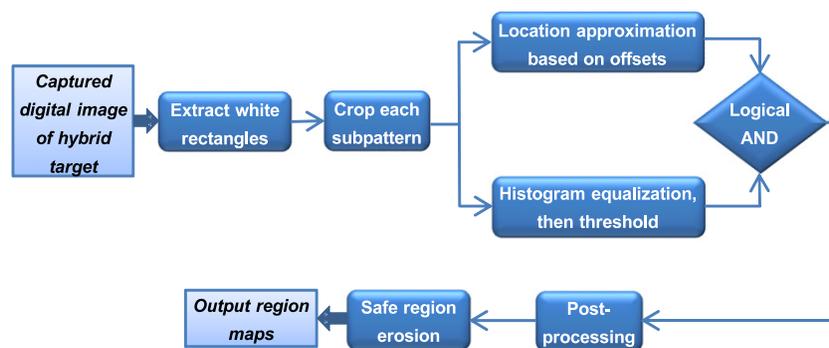


Figure 5. Block diagram of the region extraction algorithm for gray regions.

in each subpattern. Similar to the manner in which gray regions are extracted, location approximation is also used to approximately identify the color dot regions of interest in each subpattern. However, the thresholding step is a little different. Since within each subpattern, all color dot regions

are isolated from each other and are surrounded by the light gray region, the saturation is quite different between the color dot regions and the surrounding light gray region. In the color dot regions, the saturation value of each pixel is high, while that of the surrounding light gray region is much lower.

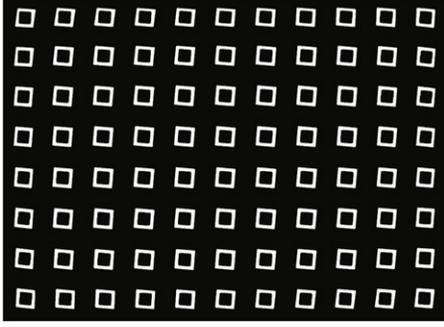


Figure 6. Final dark gray region map extracted from the composite target image captured with our target system.

Thus, we convert the RGB image to HSV (Hue, Saturation, and Value) space,¹⁴ and compute the saturation image I_S according to Eqs. (3)–(5). Note that $0 \leq I_S \leq 1$. Fig. 8(c) shows the resulting saturation image.

$$I_S = \begin{cases} 0, & I_C = 0, \\ \frac{I_C}{I_V}, & \text{otherwise,} \end{cases} \quad (3)$$

where

$$I_C = \max(I_R, I_G, I_B) - \min(I_R, I_G, I_B), \quad (4)$$

$$I_V = \max(I_R, I_G, I_B). \quad (5)$$

We apply a global threshold of 20% to the saturation value of each pixel in I_S , followed by a morphological opening operation with the same structuring element that was used for the post-processing step in gray region extraction. The result of this thresholding step is shown in Fig. 8(d). We then perform a logical AND operation between the region map based on location approximation shown in Fig. 8(b) and the result of the thresholding step shown in Fig. 8(d) to obtain the initial yellow color dot region map, which is shown in Fig. 8(e). For the post-processing step, another morphological opening operation is used to remove potential noise. The hexagon-shaped structuring element used for the gray region maps is again used for this opening.

Safe region erosion is applied to the resulting color dot region maps, as it was for extraction of the gray region maps. A disk-shaped structuring element is again used for the morphological erosion. The same rule $r = \frac{(1-0.6)D_{\text{color dot}}}{2}$ is used to determine the radius r of the structuring element, where $D_{\text{color dot}}$ is the diameter of the color dot. We calculate $D_{\text{color dot}}$ based on the number of pixels $A_{\text{color dot}}$ contained within the initial color dot region map, according to $D_{\text{color dot}} = 2\sqrt{\frac{A_{\text{color dot}}}{\pi}}$. For our system, $D_{\text{color dot}} = 44$ pixels. Fig. 8(f) shows the result of applying safe region erosion to Fig. 8(e). The final yellow color dot region map of all subpatterns is shown in Fig. 9.

At this point, accurate region maps have been obtained for all regions of interest, along with the centroid of each subpattern. All this information will be used to calculate and evaluate the key image quality metrics.

KEY METRICS EVALUATION

After getting the location maps of all regions of interest, all the key metrics, including image sharpness (Modulation Transfer Function), tone reproduction, color accuracy, noise level, geometric distortion, and uniformity of the image quality metrics across the captured composite target, can be measured, and evaluated numerically.

Modulation Transfer Function

Several methods and corresponding targets are available to use for measuring the Modulation Transfer Function (MTF) of image capture systems. In most cases, several types of targets can be used for MTF measurement. Examples include a sinusoidal wave target,¹⁵ a grille pattern, and a slanted-edge target.⁹ Zhang et al.¹⁶ examined three different approaches to computing MTF based on the three targets mentioned above, and explored the relationship among them. Jang and Allebach¹⁷ developed a comprehensive procedure, including a new test page, for the characterization of printer MTF. Bonnier and Lindner¹⁸ compared three MTF measurement methods, and proposed an improvement to the procedure developed by Jang and Allebach¹⁷. In the present article, our composite target is designed based on the slanted-edge target, and the corresponding analysis algorithm proposed by Burns⁸ will be used.

The MTF of the camera in the capture system is very important, and is strictly controlled by the camera manufacturer. However, the camera may be damaged before or while it is installed in the target product. To be consistent with the usage of the capture system, we perform the MTF analysis on the luminance (CIE Y) channel of final output color image. The subpattern in row 3, column 5 of the luminance channel of the captured composite target test page is shown in Figure 10(a). To calculate the MTF for this subpattern, we use the four slanted edges between the dark gray and light gray regions, which are highlighted by the red boxes in Fig. 10(b). To find these slanted edges, we first extract the white, black, and dark gray regions shown in Fig. 10(a). To do this, we follow the same process as shown in Fig. 5. Then, we fill the hole inside the dark gray region map with the white and black rectangular region maps to obtain the region map for the MTF calculation. The result is shown in Fig. 10(c).

The four slanted edges in the subpattern shown in Fig. 10(b) need to be located next to carry out the MTF calculation.⁸ The number of pixels $A_{\text{MTF region}}$ in the MTF region map shown in Fig. 10(c) is used to calculate the length of the four edges $L_{\text{MTF region}}$ of the MTF region map according to $L_{\text{MTF region}} = \sqrt{A_{\text{MTF region}}}$, since the MTF region map is square. Then, we estimate the locations of the four slanted edges according to their offsets from the centroid of the MTF region map, and determine four $0.7L_{\text{MTF region}} \times 0.3L_{\text{MTF region}}$ rectangular areas to crop. The four red rectangles in Fig. 10(b) are the cropped regions. The Slant Edge Analysis Tool sformat¹⁹ 2.0 package developed by Burns is used for the MTF calculation. For each cropped region, the algorithm⁸ fits the edge with a second order

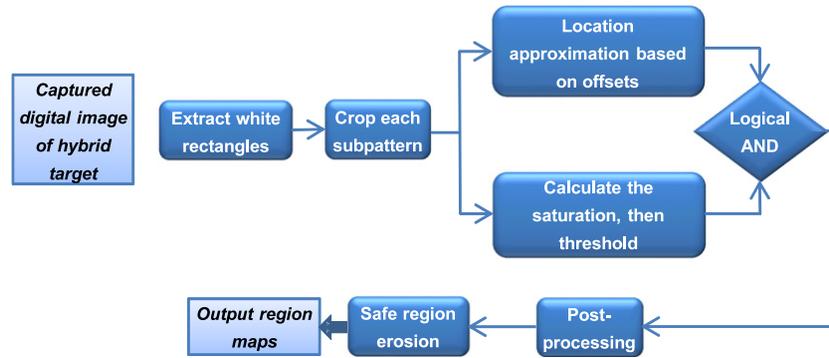


Figure 7. Block diagram of the region extraction algorithm for color dot regions.

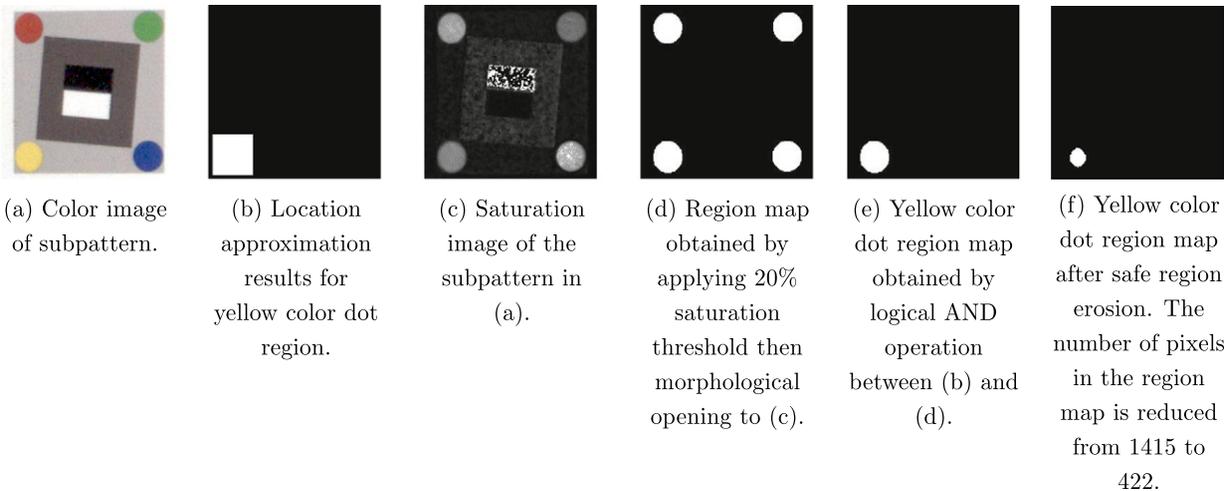


Figure 8. Illustration of the steps for determining the color dot region maps to be used for extracting the color dot regions. The specific example shown here is for the yellow color dot region. Parts (b) and (d)–(f) are binary masks.

polynomial to remove the effects of lens distortion, and compute the derivative in the direction normal to the edge. The discrete Fourier transform of the derivative signal is calculated, and normalized to get the spatial frequency response. For each slanted edge, the analysis provides MTF values across a range of spatial frequencies. Figure 11 shows the MTF response curves for the four edges in Fig. 10(b). Based on a captured image resolution of 245 dpi, the Nyquist cutoff frequency would be 4.9 lp/mm. We see that the four responses are similar, although the response in the vertical direction (upper and lower edges) is significantly stronger than the response in the horizontal direction (left and right edges).

Since it is printed with an inkjet printer, the hard copy of the composite target itself has variations in nominally smooth areas caused by the halftone texture that is used to render the tone in those smooth areas. Thus, the MTF response curve measured from the image of the composite target captured by our target camera system will consist of contributions from both the halftone texture in the printed target and the camera system itself. In order to separate these contributions, we scan the QA-62 slant edge target (Applied Image Inc., Rochester, NY 14609) with the same camera system, and analyze the image using the Slant Edge Analysis Tool sformat¹⁹ 2.0 package. We put the QA-62 target in the



Figure 9. Final yellow color dot region map extracted from the composite target captured with our target system.

center of camera’s field of view, and measured the MTF response curve for the four edges. The results are shown in Figure 12. The average MTF value at 2 lp/mm for the four edges is 0.6626. For the composite target, the average MTF value measured at two subpatterns in the center, i.e. rows 4 and 5, column 6, is 0.7145. Thus, we conclude that the printing process used to reproduce the composite target tends to increase the MTF measurement.

In order to show how image sharpness varies across the page, the specific spatial frequency 2 lp/mm (line pairs per millimeter) is chosen, which is the spatial frequency set by the

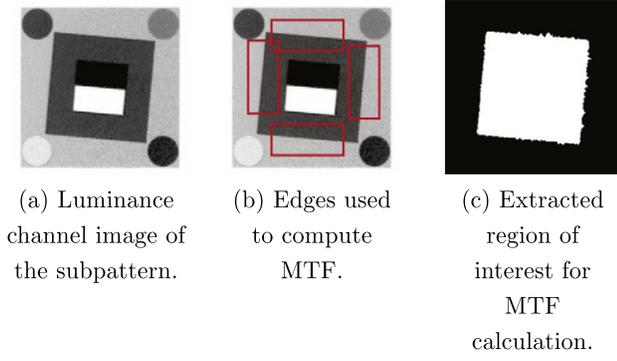


Figure 10. The subpattern in row 3, column 5 of the captured composite target test page image illustrating the information used for MTF calculation.

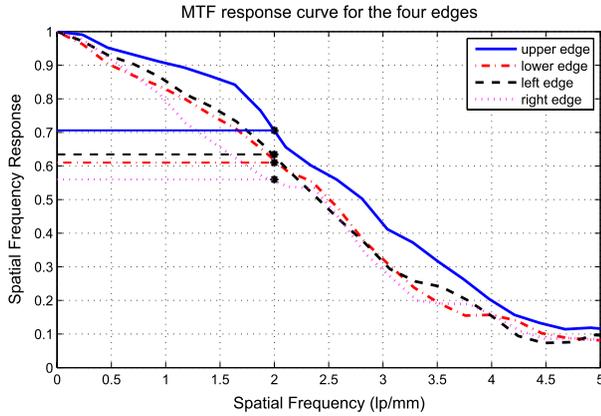


Figure 11. MTF response curves of our target capture system calculated with the composite target image, using the subpattern at the location (row 3, column 5) which is shown in Fig. 10(b). The horizontal axis is the spatial frequency from 0 to 5 lp/mm (line pairs per millimeter). The vertical axis represents the spatial frequency response ranging from 0 to 1. The response value at 2 lp/mm is selected for evaluation across the entire target.

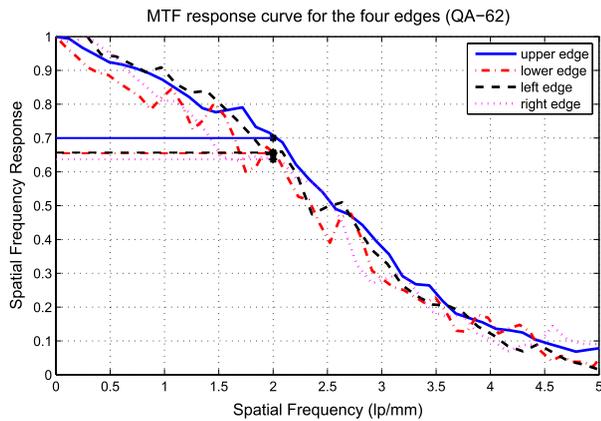


Figure 12. MTF response curves of our target capture system calculated with the QA-62 target positioned at the center of the camera's field of view. The horizontal axis is the spatial frequency from 0 to 5 lp/mm. The vertical axis represents the spatial frequency response ranging from 0 to 1.

camera manufacturer to evaluate the quality of their cameras. This is also the specification that the camera manufacturer agreed to meet for the manufacturer of the multifunction

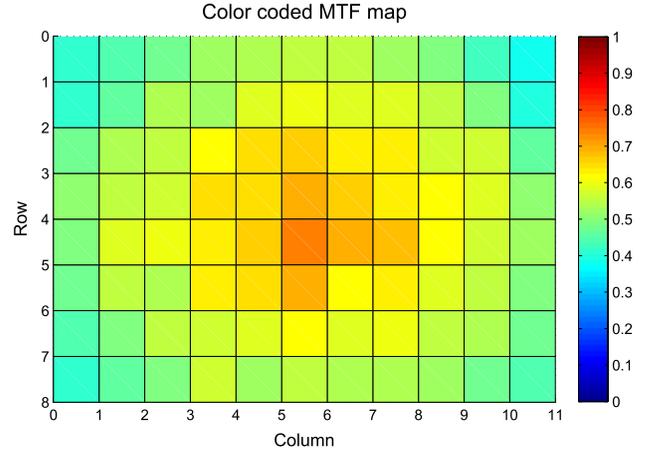


Figure 13. Color coded MTF map corresponding to the average MTF value at 2 lp/mm calculated from the four slant edges in the subpatterns of the composite target captured with our target system. Blue represents a low MTF value and low image sharpness. Red represents a high MTF value and high image sharpness.

printer unit. Let $MTF^{side}(k, l)$, where $side = upper, lower, left, or right$, denote the MTF response for the subpattern in row k , column l at spatial frequency 2 lp/mm. Then we compute the average $MTF_{avg}(k, l) = (MTF^{upper}(k, l) + MTF^{lower}(k, l) + MTF^{left}(k, l) + MTF^{right}(k, l))/4$ to represent the performance at that location. These MTF values can be color coded to give us a 2D view of how the MTF varies across the page, as shown in Figure 13. Inspection of Fig. 13 indicates, as would be expected, that the MTF response is largest in the center of the field of view, and decreases as we approach the edges of the field of view.

To gain a more compact and more quantitative appreciation of how the MTF varies within the page, we define several summary statistics:

$$MTF_{avg}^{side} = \frac{1}{88} \sum_{k=1}^8 \sum_{l=1}^{11} MTF^{side}(k, l), \quad (6)$$

$$MTF_{stddev}^{side} = \left(\frac{1}{87} \sum_{k=1}^8 \sum_{l=1}^{11} (MTF^{side}(k, l) - MTF_{avg}^{side})^2 \right)^{1/2}, \quad (7)$$

$$MTF_{min}^{side} = \min_{1 \leq k \leq 8, 1 \leq l \leq 11} MTF^{side}(k, l), \quad (8)$$

$$MTF_{max}^{side} = \max_{1 \leq k \leq 8, 1 \leq l \leq 11} MTF^{side}(k, l), \quad (9)$$

$$MTF_{range}^{side} = MTF_{max}^{side} - MTF_{min}^{side}. \quad (10)$$

Here, as previously, $side = upper, lower, left, or right$. These statistics are shown in Table I.

Table I shows that the MTF averaged over all the subpatterns and all four edges is relatively large at 0.556. The standard deviation for each edge, taken over all the subpatterns and then averaged over the four edges, is relatively small at 0.076. This suggests that the MTF performance of the capture system is reasonably consistent within the camera's field of view. Yet, the range averaged over all four edges is

Table 1. Summary statistics for the MTF of our target system.

Statistic	Slanted edge (<i>side</i>) in subpattern				Average over the four edges
	Upper	Lower	Left	Right	
MTF_{avg}^{side}	0.625	0.584	0.535	0.478	0.556
MTF_{stddev}^{side}	0.088	0.089	0.085	0.077	0.076
MTF_{min}^{side}	0.394	0.376	0.353	0.300	0.377
MTF_{max}^{side}	0.824	0.775	0.739	0.700	0.740
MTF_{range}^{side}	0.430	0.400	0.385	0.400	0.364

0.364. This relatively large value is due to the strong localized peak in the MTF in the center of the field of view, and the gradual roll-off in the MTF as we approach the corners of the field of view, which can be seen in Fig. 13.

Tone Reproduction

In this subsection, we describe a set of metrics that is intended to characterize the tone reproduction of the image capture system. For this purpose, we use the 1976 CIE $L^*a^*b^*$ uniform color space.²⁰ As mentioned earlier, our capture system is calibrated to yield the output image in sRGB coordinates. We apply Eqs. (1) and (2), followed by Eqs. (11)–(14) below, to convert the captured image pixel-by-pixel to CIE $L^*a^*b^*$.

$$L^* = 116(f(Y/Y_n)) - 16, \quad (11)$$

$$a^* = 500[f(X/X_n) - f(Y/Y_n)], \quad (12)$$

$$b^* = 200[f(Y/Y_n) - f(Z/Z_n)], \quad (13)$$

where

$$f(t) = \begin{cases} t^{\frac{1}{3}}, & \text{if } t > \left(\frac{6}{29}\right)^3, \\ \frac{1}{3} \left(\frac{29}{6}\right)^2 t + \frac{4}{29}, & \text{otherwise,} \end{cases} \quad (14)$$

and (X_n, Y_n, Z_n) are the tristimulus coordinates for the D65 white point.

Using the four region maps described earlier for the gray regions, we then compute the average CIE $L^*a^*b^*$ tristimulus coordinates $L^*a^*b^*$ for each of these regions. Here region type = black, dark gray, light gray, or white.

We examine the tone reproduction capabilities of our system in terms of the match between the CIE L^* values computed from the sRGB image capture system output, as described above, and the reference CIE L^* values measured with a GretagMacbeth Spectrolino spectrophotometer (GretagMacbeth, New Windsor, NY 12553-6148) directly from the printed composite target sheet. For this metric, we use the L^* values measured from the black, dark gray, light gray, and white regions. For each subpattern, this gives us four data points $(L_{captured}^{*region\ type}(k, l), L_{reference}^{*region\ type})$. Here region type = black, dark gray, light gray, or white, and (k, l) denotes the row and column in which the subpattern is located. Note that

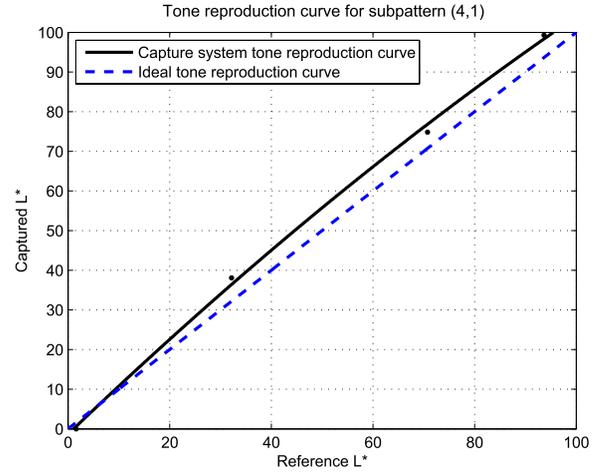


Figure 14. Tone reproduction curve for subpattern (4,1).

for each region type, a single reference L^* value is used for all subpatterns.

We then fit a second order polynomial to these four data points to obtain the tone reproduction curve for $L^*_{captured}$ as a function of $L^*_{reference}$. Figure 14 shows the resulting tone reproduction curve for the (4, 1)th subpattern in Fig. 4(a), plotted against the ideal tone reproduction curve for which $L^*_{captured} \equiv L^*_{reference}$.

In order to quantify the goodness of fit between the capture system tone reproduction curve and the ideal tone reproduction curve for each subpattern, we compute the root-mean-squared error between the captured and reference L^* values, averaged over the four region types.

$$\Delta L^{*rms}(k, l) = \left(\frac{1}{4} \sum_{\text{region type}} \left(L_{captured}^{*region\ type}(k, l) - L_{reference}^{*region\ type} \right)^2 \right)^{1/2}. \quad (15)$$

For the tone reproduction curve shown in Fig. 14, $\Delta L^{*rms}(4, 1) = 4.63$.

To see how the tone reproduction accuracy varies across the field of view, we color code the root-mean-squared error $\Delta L^{*rms}(k, l)$ for each subpattern and display it as a map, as we did in Fig. 13 for the MTF response. The result is shown in Figure 15. From Fig. 15, we see considerable variation

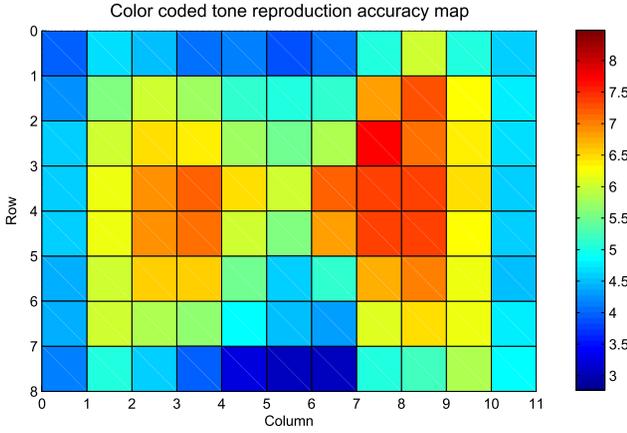


Figure 15. Color coded tone reproduction accuracy map corresponding to the root-mean-squared error $\Delta L^{*rms}(k, l)$ in the subpatterns of the composite target image captured by our target capture system. Blue represents a low value for $\Delta L^{*rms}(k, l)$ and a high accuracy. Red represents a high value for $\Delta L^{*rms}(k, l)$ and a low accuracy.

Table II. Summary statistics for the tone reproduction of our target system.

ΔL^{*rms}_{avg}	5.55
ΔL^{*rms}_{stddev}	1.12
ΔL^{*rms}_{min}	3.08
ΔL^{*rms}_{max}	7.71
ΔL^{*rms}_{range}	4.63

across the field of view in the accuracy of tone reproduction. Interestingly, the error is largest in the dumb-bell-shaped region near the center of the field of view and lower near the periphery of the field of view, especially at the center top and bottom.

Finally, to gain a summary view of the tone reproduction accuracy within the entire composite target, we define statistics, which are based on all subpatterns, that are analogous to those defined for the MTF in Eqs. (6)–(10). These are ΔL^{*rms}_{avg} , ΔL^{*rms}_{stddev} , ΔL^{*rms}_{min} , ΔL^{*rms}_{max} , and ΔL^{*rms}_{range} . These are shown in Table II.

We observe that although the range ΔL^{*rms}_{range} is relatively large at 4.63, as can be seen in Fig. 15, the standard deviation ΔL^{*rms}_{stddev} is much smaller at 1.12. Thus, the tone reproduction accuracy is relatively constant across much, but not all, of the captured composite target test page. This can also be seen in Fig. 15, since most of the squares are yellow, green, or light blue, corresponding to mid-range values.

Color Accuracy

In this subsection, we describe a set of metrics that is intended to characterize the color accuracy of the image capture system. For this purpose, we again use the 1976 CIE $L^*a^*b^*$ uniform color space.²⁰ As discussed in the previous subsection, we start by converting the captured image of the composite target pixel-by-pixel to CIE $L^*a^*b^*$. Using the four region maps described earlier for the color dots, we then compute the average CIE $L^*a^*b^*$ tristimulus coordinates

$L^*a^*b^*$ for each of these regions. Here, color dot = red, green, blue, or yellow.

We examine the color accuracy of our system in terms of the match between the CIE $L^*a^*b^*$ values computed from the sRGB image capture system output, as described above, and the reference CIE $L^*a^*b^*$ values measured with a spectrophotometer directly from the printed composite target sheet. For this metric, we compute the color difference in ΔE_{94} units²¹ between the captured tristimulus coordinates and the reference values from the red, green, blue, and yellow color dots:

$$\Delta E_{94}^{color\ dot}(k, l) = \sqrt{\left(\frac{\Delta L^{*color\ dot}(k, l)}{k_L S_L}\right)^2 + \left(\frac{\Delta C^{*color\ dot}(k, l)}{k_C S_C}\right)^2 + \left(\frac{\Delta H^{*color\ dot}(k, l)}{k_H S_H}\right)^2}, \quad (16)$$

where

$$\Delta L^{*color\ dot}(k, l) = L^{*color\ dot}_{captured}(k, l) - L^{*color\ dot}_{reference}(k, l), \quad (17)$$

$$\Delta C^{*color\ dot}(k, l) = \sqrt{a^{*color\ dot}_{captured}(k, l)^2 + b^{*color\ dot}_{captured}(k, l)^2} - \sqrt{a^{*color\ dot}_{reference}(k, l)^2 + b^{*color\ dot}_{reference}(k, l)^2}, \quad (18)$$

$$\Delta H^{*color\ dot}(k, l) = \sqrt{(\Delta E_{ab}^{*color\ dot}(k, l))^2 - (\Delta L^{*color\ dot}(k, l))^2 - (\Delta C^{*color\ dot}(k, l))^2}, \quad (19)$$

$$\Delta E_{ab}^{*color\ dot}(k, l) = \sqrt{(\Delta L^{*color\ dot}(k, l))^2 + (\Delta a^{*color\ dot}(k, l))^2 + (\Delta b^{*color\ dot}(k, l))^2}, \quad (20)$$

$$\Delta a^{*color\ dot}(k, l) = a^{*color\ dot}_{captured}(k, l) - a^{*color\ dot}_{reference}(k, l), \quad (21)$$

$$\Delta b^{*color\ dot}(k, l) = b^{*color\ dot}_{captured}(k, l) - b^{*color\ dot}_{reference}(k, l), \quad (22)$$

$$k_L = 1, \quad (23)$$

$$S_L = 1, \quad (24)$$

$$S_C = 1 + 0.045 C^{*color\ dot}, \quad (25)$$

$$S_H = 1 + 0.015 C^{*color\ dot}, \quad (26)$$

$$C^{*color\ dot}(k, l) = \sqrt{\sqrt{a^{*color\ dot}_{captured}(k, l)^2 + b^{*color\ dot}_{captured}(k, l)^2} \cdot \sqrt{a^{*color\ dot}_{reference}(k, l)^2 + b^{*color\ dot}_{reference}(k, l)^2}}. \quad (27)$$

Here, color dot = red, green, blue, or yellow, and (k, l) denotes the row and column in which the subpattern is located. For each subpattern, this gives us four error values. Note that for each color dot, a single reference $L^*a^*b^*$ value is used for all subpatterns.

To obtain an overall measure of color accuracy for each subpattern, we compute the average of the ΔE_{94} values for the four color dots in the subpattern:

$$\Delta E_{94}(k, l) = \frac{1}{4}(\Delta E_{94}^{red}(k, l) + \Delta E_{94}^{green}(k, l) + \Delta E_{94}^{blue}(k, l) + \Delta E_{94}^{yellow}(k, l)). \quad (28)$$

To see how color accuracy varies across the field of view, we color code the color reproduction error $\Delta E_{94}(k, l)$ for each subpattern and display it as a map, as we did in Figs. 13

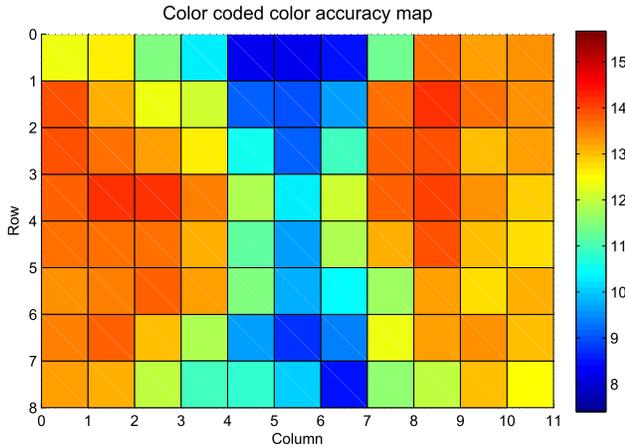


Figure 16. Color coded color accuracy map corresponding to the color error in ΔE_{94} units, averaged over the four color dots in each subpattern of the image captured by our target capture system. Blue represents a low value for $\Delta E_{94}(k, l)$ and a high accuracy. Red represents a high value for $\Delta E_{94}(k, l)$ and a low accuracy.

and 15. The result is shown in Figure 16. In Fig. 16, we see that, overall, the color accuracy is much worse than the accuracy of tone reproduction shown in Fig. 15. This is to be expected for a three-channel capture device that is not colorimetric—especially one that is relatively low-cost. The error is largest in a bow-tie-shaped region that spans the entire composite target, and smallest in the three central columns of the subpattern array.

Finally, to gain a summary view of the color accuracy within the entire composite target, we define statistics that are based on all subpatterns, and which are analogous to those defined for the MTF in Eqs. (6)–(10) and for tone reproduction above. These are $\Delta E_{94}^{\text{color dot avg}}$, $\Delta E_{94}^{\text{color dot stddev}}$, $\Delta E_{94}^{\text{color dot min}}$, $\Delta E_{94}^{\text{color dot max}}$, and $\Delta E_{94}^{\text{color dot range}}$. They are shown in Table III. Comparing Tables II and III, we note again that, on average, the color accuracy is much worse than the accuracy of tone reproduction. For tone reproduction, we observe an average value for ΔL^{*rms} of 5.55 in Table II, whereas in Table III, for $\Delta E_{94}^{\text{color dot}}$, the value averaged across all four color dots is 12.23. Note that from Eqs. (16), (23) and (24), it follows that $\Delta E_{94} = \Delta L^*$, when we are considering only errors in lightness. Interestingly, the two metrics show nearly the same values across the target for standard deviation, 1.50 and 1.67, respectively, and range, 6.17 and 6.02, respectively. Examining the accuracy for individual colors as shown in Table III, we see that yellow color dots have the largest average error, while blue color dots have the smallest average error.

Noise Level

In this subsection, we evaluate the noise level of the image captured by our capture system with the modified version of the Dooley–Shaw²² metric for image graininess shown in Eq. (29). It measures the granularity of a gray scale or RGB

Table III. Summary statistics for the color accuracy of our target system.

Statistic	Color dot				Average
	Red	Green	Blue	Yellow	
$\Delta E_{94}^{\text{color dot avg}}$	11.93	13.33	9.16	14.48	12.23
$\Delta E_{94}^{\text{color dot stddev}}$	2.62	1.42	2.76	0.98	1.67
$\Delta E_{94}^{\text{color dot min}}$	4.87	8.84	1.78	11.90	8.22
$\Delta E_{94}^{\text{color dot max}}$	14.94	15.48	13.58	16.77	14.24
$\Delta E_{94}^{\text{color dot range}}$	10.07	6.64	11.80	4.87	6.02

rectangular image patch with nominally constant tone.

$$G_c = \int \int \sqrt{W(u, v; c)} VTF(u, v; c) dudv, \quad (29)$$

where $c = L^*$, a^* , or b^* denotes the color channel in which the granularity is being computed, $W(u, v; c)$ is the Wiener noise power spectrum of the image in color channel c , and $VTF(u, v; c)$ is the Hirose²³ visual transfer function for color channel c , representing the sensitivity of the human visual system to different spatial frequencies in different color channels. Compared with Eq. (7) in Ref. 21, we have dropped the term $e^{-1.8D}$ in front of the integral. The rationale for doing this is that this term was intended to account for the non-uniform sensitivity of the human visual system as a function of density. However, as discussed below, we will perform the granularity calculation in CIE $L^*a^*b^*$, which already accounts for this non-uniform sensitivity.

Since it is printed with an inkjet printer, the hard copy of the composite target itself has variations in nominally smooth areas caused by the halftone texture that is used to render the tone in those smooth areas. Thus, the noise observed in the image of the composite target captured by our target camera system will consist of contributions from both the halftone texture in the printed target and the camera system itself. In order to separate these contributions, we scan the hard copy target with an Epson Expression 10000 XL flatbed scanner (Epson America, Inc., Long Beach, CA 90806) at 2400 dpi (dots per inch), and perform the same noise computations on the scanned image as on the captured image, assuming the same viewing distance of 10.8 inches, but adjusting for the different resolution of these two sources.

To start the noise measurement process, we crop rectangular patches from the white and black gray regions and the four color dots. To simplify the identification of these rectangular patches, we do not use the irregularly shaped light gray and dark gray regions. For the composite target image captured by our target camera system, we take advantage of the region maps generated in the previous section and crop an image patch at the center of each of the region maps with a specified size. The image patches cropped from the white and black gray regions are 15×30 pixels. The image patches cropped from the four color dots are 15×15 pixels. For the reference target image captured by the flatbed scanner, we do not have region maps, so we find the

regions of interest using the displacement from subpattern to subpattern. In this case, because of the higher resolution of the scanned image, the sizes of the image patches are 150×300 pixels for the white and black gray regions, and 150×150 pixels for the four color dots. To calculate the visually weighted power spectrum for the image patches, we convert them pixel-by-pixel to the 1976 CIE $L^*a^*b^*$ uniform color space,²⁰ using Eqs. (1)–(2), and (11)–(14).

The Wiener noise power spectra for L^* , a^* , and b^* are then estimated by a 2D discrete fast Fourier transform, whose square root is scaled point-by-point with the Hirose visual transfer function. The integral is estimated by the summation of rectangular areas from 0 to the Nyquist frequency, which is about 4.9 lp/mm for the target capture system.

For the L^* , a^* , and b^* spaces, the respective Dooley–Shaw metrics G_{L^*} , G_{a^*} , and G_{b^*} can be calculated according to a discrete-parameter version of Eq. (29). The square root of the sum of the squares (srss) of the three granularity measures is then calculated for each region examined for noise level:

$$G_{\text{srss}}^{\text{region type}}(k, l) = \sqrt{(G_{L^*}^{\text{region type}}(k, l))^2 + (G_{a^*}^{\text{region type}}(k, l))^2 + (G_{b^*}^{\text{region type}}(k, l))^2}. \quad (30)$$

Here, region type = red, green, blue, yellow, white, or black, and (k, l) denotes the row and column in which the subpattern is located. For each subpattern, this gives us six granularity values. To obtain an overall measure of the noise level for each subpattern, we compute the average of the square-root-of-sum-of-squares granularity $G_{\text{srss}}^{\text{region type}}(k, l)$ for the six regions we used in the subpattern:

$$G_{\text{srss}}(k, l) = \frac{1}{6} (G_{\text{srss}}^{\text{red}}(k, l) + G_{\text{srss}}^{\text{green}}(k, l) + G_{\text{srss}}^{\text{blue}}(k, l) + G_{\text{srss}}^{\text{yellow}}(k, l) + G_{\text{srss}}^{\text{white}}(k, l) + G_{\text{srss}}^{\text{black}}(k, l)). \quad (31)$$

To see how the noise level varies across the field of view, we color code the square-root-of-sum-of-squares granularity $G_{\text{srss}}(k, l)$ for each subpattern and display it as a map, as we did in Figs. 13, 15 and 16 for the MTF, tone reproduction, and color accuracy, respectively. The results are shown in Figures 17 and 18, respectively, for the reference 2400 dpi image of the composite target obtained from the flatbed scanner and the composite target image captured by our target camera system. Examining Figs. 17 and 18 individually, we see no systematic variation of the noise across the composite target, although both maps do show one or two hot spots. With these exceptions, we can conclude that the noise level is relatively uniform over the composite target. Comparing Figs. 17 and 18 we see that the reference image obtained from the target by the flatbed scanner has a much lower and more uniform level of noise than does the composite target image captured by the camera system. This validates the concept of using the image of the composite target captured by the camera system to estimate the noise introduced by the camera system.

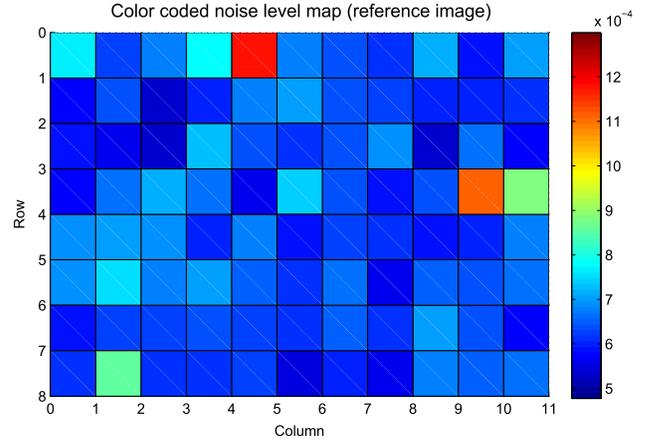


Figure 17. Color coded noise level map corresponding to the square-root-of-sum-of-squares granularity, averaged over the six selected regions in each subpattern of the reference composite target image obtained from the flatbed scanner. Blue represents a low value for $G_{\text{srss}}(k, l)$ and a high image quality. Red represents a high value for $G_{\text{srss}}(k, l)$ and a low image quality.

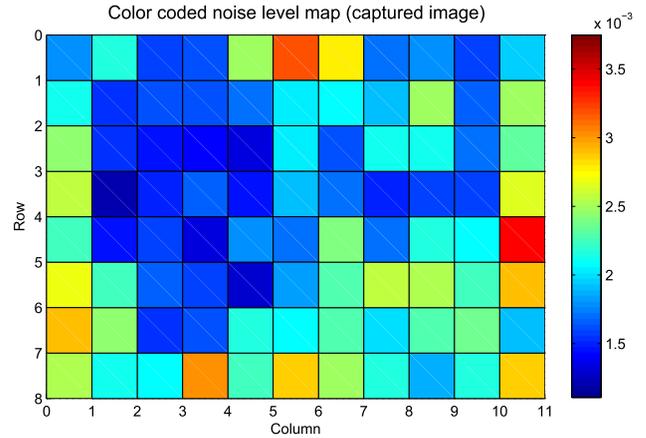


Figure 18. Color coded noise level map corresponding to the square-root-of-sum-of-squares granularity, averaged over the six selected regions in each subpattern of the composite target image captured by our target camera system. Blue represents a low value for $G_{\text{srss}}(k, l)$ and a high image quality. Red represents a high value for $G_{\text{srss}}(k, l)$ and a low image quality.

To gain a more quantitative understanding of these relationships, we define statistics that are based on all subpatterns, and which are analogous to those defined previously for the MTF in Eqs. (6)–(10), and, similarly, for tone reproduction and color accuracy. These are $G_{\text{srss}}^{\text{source}}$, $G_{\text{srss}}^{\text{source}}_{\text{stddev}}$, $G_{\text{srss}}^{\text{source}}_{\text{min}}$, $G_{\text{srss}}^{\text{source}}_{\text{max}}$, and $G_{\text{srss}}^{\text{source}}_{\text{range}}$, which are based on the square-root-of-sum-of-squares granularity $G_{\text{srss}}(k, l)$ for all the subpatterns. Here, source = flatbed scanner reference or camera system. The results are shown in Table IV. From Table IV, we see that on average, the granularity of the image captured by the camera system is more than three times greater than that of the reference image from the flatbed scanner. The standard deviation is almost five times greater for the camera image than for the reference image, and the range is more than three times greater. Thus, we conclude that the capture system is adding two times the average noise level, and increasing the

Table IV. Summary statistics for the noise level based on the reference composite target image from the flatbed scanner and the composite target image captured by our target camera system. Actual values are $\times 10^{-3}$.

Statistic	Source		Camera system minus flatbed scanner reference
	Camera system	Flatbed scanner reference	
$G_{srrs}^{source} _{avg}$	2.04	0.65	1.39
$G_{srrs}^{source} _{stddev}$	0.48	0.10	0.38
$G_{srrs}^{source} _{min}$	1.23	0.53	0.70
$G_{srrs}^{source} _{max}$	3.41	1.18	2.23
$G_{srrs}^{source} _{range}$	2.18	0.65	1.53

non-uniformity across the field of view by an even larger amount, relative to the reference image from the flatbed scanner. This is in agreement with what we observe in Figs. 17 and 18. It is important to note that the noise in the reference image from the flatbed scanner includes both noise that is intrinsic to the printed target and noise contributed by the flatbed scanner. The fourth column in Table IV shows the difference between the noise level in the camera image and the noise level in the reference image. It provides another way to compare the noise levels in the images captured by these two sources.

Geometric Distortion

Barrel distortion²⁴ is observed in the output images of the camera-based capture system. In this subsection, we describe a set of metrics to characterize the geometric distortion of the capture system. The centroids of all subpatterns form an 8×11 grid, and are used for this purpose. For each subpattern in row k , column l of the captured image, we have the $X_{capture}(k, l)$ and $Y_{capture}(k, l)$ coordinates of its centroid. The averaged $Y_{capture}(k)$ coordinates of subpatterns in the same row and averaged $X_{capture}(l)$ coordinates of subpatterns in the same column are used as the reference coordinates, and computed according to Eqs. (32) and (33).

$$Y_{reference}(k) = \frac{1}{11} \sum_{l=1}^{11} Y_{capture}(k, l), \quad (32)$$

$$X_{reference}(l) = \frac{1}{8} \sum_{k=1}^8 X_{capture}(k, l). \quad (33)$$

The vectors $\vec{D}(k, l) = (X_{capture}(k, l), Y_{capture}(k, l)) - (X_{reference}(l), Y_{reference}(k))$, $1 \leq k \leq 8$ and $1 \leq l \leq 11$ are calculated as the displacement errors. Figure 19 shows the displacement vectors for all centroids. From Fig. 19, we see that the distortion is smallest near the center of the composite target, and increases as we move toward the periphery of the composite target.

Finally, to gain a summary view of the geometric distortion within the entire composite target, we define statistics for the magnitude of the displacement vectors

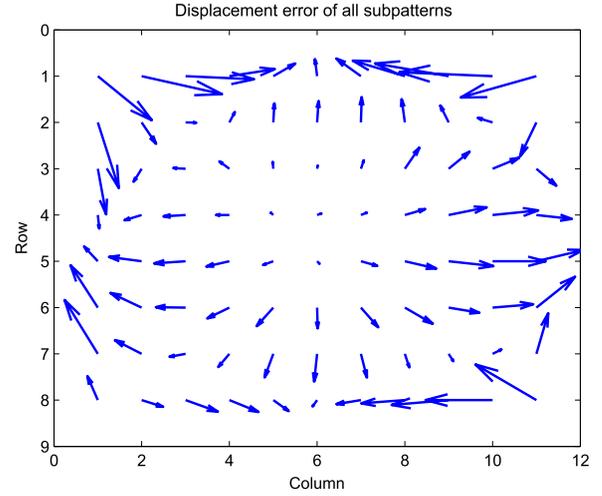


Figure 19. Displacement error map for our target capture system. The magnitude of the vectors expressed as a percentage of the width of the field of view (FOV) is indicated by the scale in the figure. The arrows point from the reference to the measurement location.

Table V. Summary statistics for geometric distortion of our target system.

$ \vec{D} _{avg}$	0.0948 [†]
$ \vec{D} _{stddev}$	0.0618
$ \vec{D} _{min}$	0.0101
$ \vec{D} _{max}$	0.290
$ \vec{D} _{range}$	0.280

[†]The units are expressed as a percentage of the width of the field of view of the camera.

$|\vec{D}(k, l)|$, which are based on all subpatterns, that are analogous to those defined previously for the MTF in Eqs. (6)–(10), and similarly for tone reproduction, color accuracy, and noise. These are $|\vec{D}|_{avg}$, $|\vec{D}|_{stddev}$, $|\vec{D}|_{min}$, $|\vec{D}|_{max}$, and $|\vec{D}|_{range}$. They are shown in Table V. From Table V, we see that the average error is almost 0.1% of the width of the field of view with a minimum error of one-tenth of that value and a maximum error that is nearly 30 times larger. The average error corresponds to a displacement of approximately 3 pixels. Together with a standard deviation that is approximately 2/3 of the average error, we conclude that there is a wide range in the geometric distortion throughout the field of view.

SUMMARY AND CONCLUSIONS

We have developed a novel composite target and associated image analysis tools that provide a quantitative assessment of the variation of several key image quality metrics within the field of view of a camera-based image capture system. Our intended application is a multifunction printer or all-in-one device that can scan, copy, and print, but which uses a low-cost mobile phone camera for copying and an external source for illumination, rather than a flatbed or sheetfed scanner. However, we believe that the target and image

analysis tools that we have developed are applicable to a broader range of camera-based capture systems.

The composite target consists of a regular array of periodically repeated identical subpatterns, each of which supports the computation of several key image quality metrics. By examining how each metric varies across the field of view, we gain a precise picture of the spatial non-uniformity of the capture system with respect to that metric. We also introduce a set of statistics for each key metric that provides a summary view of the degree of spatial non-uniformity.

The key metrics that we compute are modulation transfer function, accuracy of tone reproduction, color accuracy, noise, and geometric distortion. Experimental results for a typical capture system that is used in our intended application show significant and systematic spatial variation for all these metrics, except noise. This validates the fundamental premise underlying the development of the composite target and its analysis tool, which is to quantify precisely this spatial variation.

For our example capture system, we observed an average MTF of 0.556 at the target frequency of 2 lp/mm, an average root-mean-squared tone reproduction error of 5.55 ΔE units, an average root-mean-squared color accuracy of 12.23 ΔE units, a visually weighted average granularity of $2.04 \times 10^{-3} \Delta E$ units, and an average displacement of approximately 3 pixels over the field of view due to geometric distortion. From these statistics, color accuracy is the weakest characteristic of our example capture system.

The analysis tool is written in Matlab (MathWorks, Inc. Natick, MA 01760-2098). Running on a computer with a 2.4 GHz dual core Intel CPU, the complete analysis of a captured composite target 2970×2159 pixel image requires on average 80 s.

REFERENCES

- ¹ T. H. Ha, C. T. Wu, P. Majewicz, K. R. Bengtson, and J. P. Allebach, "Glare and shadow reduction for desktop digital camera capture systems," *Proc. SPIE* **8652**, 3–5 (2013).
- ² N. Burningham, Z. Pizlo, and J. P. Allebach, "Image quality metrics," *Encyclopedia of Imaging Science and Technology*, edited by J. P. Hornak (John Wiley and Sons, New York, 2002), pp. 598–616, ISBN: 0-471-33276-3.
- ³ U. Artmann and D. Wueller, "Improving texture loss measurement: spatial frequency response based on a colored target," *Proc. SPIE* **8293**, 22–26 (2012).
- ⁴ D. Wueller, "Proposal for a Standard procedure to test mobile phone cameras," *Proc. SPIE* **6069**, 15–19 (2006).
- ⁵ D. Williams and P. D. Burns, "Applying and extending ISO/TC42 digital camera resolution standards to mobile imaging products," *Proc. SPIE* **6494**, 28 (2007).
- ⁶ D. J. Baxter and A. Murray, "Calibration and adaptation of ISO visual noise for I3A's camera phone image quality initiative," *Proc. SPIE* **8293**, 22–26 (2012).
- ⁷ Y. Lei, P. Majewicz, K. R. Bengtson, L. Li, and J. P. Allebach, "Hybrid target for camera-based document/object capture system," *Proc. IS&T's NIP28: Int'l. Conf. on Digital Printing Technol. and Digital Fabrication 2012* (IS&T, Springfield, VA, 2012), pp. 245–249.
- ⁸ P. Burns, "Slanted-edge MTF for digital camera and scanner analysis," *Proc. IS&T's PICS: Image Processing, Image Quality, Image Capture, Systems Conf.* (IS&T, Springfield, VA, 2000), pp. 135–138.
- ⁹ ISO/TC42/WG18, Photography-Electronic Still-Picture Cameras—Resolution Measurement, Int. Std. ISO 12233 (2000).
- ¹⁰ B. E. Bayer, "Color imaging array," US Patent 3,971,065 (1976).
- ¹¹ H. S. Malvar, L. He, and R. Cutler, "High-quality linear interpolation for demosaicing of bayer-patterned color images," *Proc. IEEE Int'l. Conf. on Acoustics, Speech, and Signal Processing* (IEEE, Washington, DC, 2004), pp. III485–III488.
- ¹² M. Anderson, R. Motta, S. Chandrasekar, and M. Stokes, "Proposal for a standard default color space for the internet—sRGB," *Proc. IS&T/SID's Fourth Color Imaging Conf.* (IS&T, Springfield, VA, 1996), pp. 238–246.
- ¹³ M. Stokes, M. Anderson, S. Chandrasekar, and R. Motta, "A standard default color space for the internet: sRGB," Nov. 1996. <http://www.color.org/sRGB.xalter>.
- ¹⁴ G. H. Joblove and D. Greenberg, "Color spaces for computer graphics," *Proc. ACM's Fifth SIGGRAPH Annual Conf. on Computer Graphics and Interactive Techniques* (ACM, New York, NY, 1978), pp. 20–25.
- ¹⁵ K. J. Barnard, G. D. Boreman, A. E. Plogstedt, and B. K. Anderson, "Modulation-transfer function measurement of SPRITE detectors: sinewave response," *Appl. Opt.* **31**, 144–147 (1992).
- ¹⁶ X. Zhang, T. Kashti, D. Kella, T. Frank, D. Shaked, R. Ulichney, and J. P. Allebach, "Measuring the modulation transfer function of image capture devices: what do the numbers really mean?," *Proc. SPIE* **8293**, 22–26 (2012).
- ¹⁷ W. Jang and J. P. Allebach, "Characterization of printer MTF," *J. Imaging Sci. Technol.* **50**, 264–275 (2006).
- ¹⁸ N. Bonnier and A. J. Lindner, "Measurement and compensation of printer modulation transfer function," *J. Electron. Imaging* **19**, 011010 (2010) p. 20.
- ¹⁹ P. Burns, "Sfrmat 2.0 User's Guide," 2nd ed. (2003).
- ²⁰ G. Wyszecki and W. S. Stiles, *Color Science, Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. (John Wiley and Sons, New York, 1982), ISBN: 0-471-02106-7.
- ²¹ R. Balasubramanian, "Device characterization," *Digital Color Imaging Handbook*, edited by G. Sharma (CRC Press, 2002), pp. 269–384, ISBN: 978-0-8493-0900-7.
- ²² R. P. Dooley and R. Shaw, "Noise perception in electrophotography," *J. Appl. Photogr. Eng.* **5**, 190–196 (1979).
- ²³ Y. Hirose, "Perceptual characteristics of image noise in color hardcopy images," *Proc. SPIE* **1912**, 464–473 (1993).
- ²⁴ R. E. Fischer and B. Tadic-Galeb, *Optical System Design*, 2nd ed. (McGraw-Hill, New York, 2008), pp. 85–89, ISBN 9780071593588.