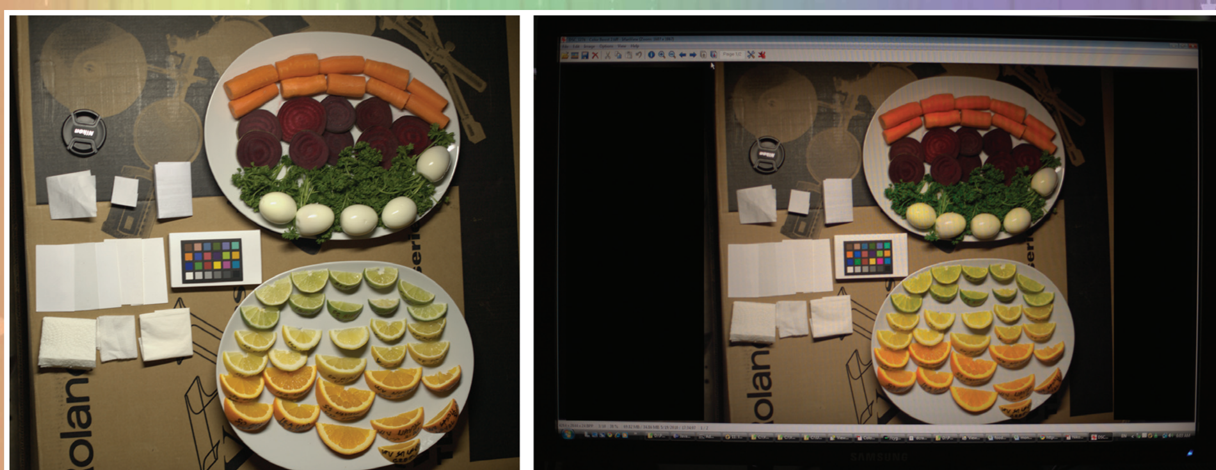
REPRINTED FROM:

JIST

JOURNAL OF IMAGING SCIENCE AND TECHNOLOGY

VOL. 56, NO. 2 | MARCH/APRIL 2012



IS&T

imaging.org

Journal of Imaging Science and Technology® 56(2): 020501-1–020501-10, 2012.

MaxRGB Reconsidered

Brian Funt▲ and Lilong Shi
School of Computing Science, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada
E-mail: funt@sfu.ca

Abstract. *The poor performance of the MaxRGB illumination-estimation method is often used in the literature as a foil when promoting some new illumination-estimation method. However, MaxRGB has usually been tested on images of only 8-bits per channel, where clipping of high radiances is likely to have occurred. The question arises as to whether the method itself is inadequate, or rather whether it has simply been tested on data of inadequate dynamic range or with inadequate preprocessing. In particular, is MaxRGB's underlying assumption that there is a white or white-equivalent surface present in every scene too strong? This question is explored here in two ways. The first avenue of investigation is based on a new database of 105 sets of multiple-exposure images. High-dynamic range images are constructed from these sets as well. The color of the scene illumination is determined by taking an extra image of the scene containing four Gretag Macbeth mini-Colorcheckers placed at an angle to one another. MaxRGB is found to perform surprisingly well when tested on either the multiple-exposure or the high-dynamic range images. The second avenue of investigation is to add some simple preprocessing to the basic MaxRGB algorithm. By removing clipped pixels followed by median filtering, MaxRGB also performs better than previously reported when tested on test images of common color constancy test sets, specifically the Simon Fraser University 321-image indoor set. In particular, the Wilcoxon signed-rank test indicates that MaxRGB outperforms the most recent bright-pixel variant of color by correlation on the 321 set. MaxRGB is also competitive against the recent Edge-Based algorithm and significantly better than the computationally intensive Bayesian method on the Grayball set and the Colorchecker set. Overall, the results presented demonstrate that MaxRGB is far more effective than it has been reputed to be. © 2012 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2012.56.2.020501]

INTRODUCTION

MaxRGB is a very simple method of estimating the chromaticity (i.e., $[r,\ g,\ b] = [R/(R+G+B),\ G/(R+G+B),\ B/(R+G+B)]$) of the scene illumination for color constancy and automatic white balancing based on the assumption that the triple of the maximal values obtained independently from each of the three color channels represents the color of the illumination. It has been much maligned in the literature; however, is its performance really as bad as it has been reported[1–4] to be?

MaxRGB is a special and extremely limited case of retinex.[5] In particular, it corresponds to McCann99 Retinex[6] when the number of iterations is infinite, or to path-based retinex[7] without thresholding but with infinite paths. Retinex and MaxRGB both depend on the assumption that one of the following holds as: (1) there is a white surface in the scene, (2) there is a shiny surface in the scene creating a specular highlight having the same chromaticity as the incident light, and (3) there are one or more separate surfaces reflecting maximally in the *R*, *G*, and *B* sensitivity ranges. Condition (3) is similar to condition (1) except that for condition (1), the three maxima occur at a single spatial location because an ideal white surface reflects maximally across all three sensitivity ranges. For condition (3), the maximal reflectances for the different color channels are allowed to come from multiple locations. MaxRGB also depends on the assumption, common to many illumination-estimation methods,[1–4,8–12] that the chromaticity of the illumination is constant throughout the scene.

MaxRGB estimates the chromaticity of the scene illumination as the chromaticity of the triplet (maximal *R*, maximal *G*, and maximal *B*). In practice, most digital still cameras are incapable of capturing the full dynamic range of a scene and use exposures and tone reproduction curves that clip or compress high values. As a result, the maximum *R*, *G*, and *B* digital counts from an image generally do not faithfully represent the corresponding maximum scene radiances. Funt et al.[13] present some tests using artificial clipping of images that show the effect that lack of dynamic range can have on various illumination-estimation algorithms.

We hypothesize that there are two reasons why the effectiveness of MaxRGB may have been underestimated. One is that it is important not only to apply MaxRGB as the simple maximum of each channel but rather it is necessary to preprocess the image data somewhat before calculating the maximum; otherwise, a single bad pixel or spurious noise will lead to the maximum being incorrect. The second, and perhaps more important reason, is that MaxRGB generally has been applied to 8-bit-per-channel, nonlinear images, for which there is likely to be significant tone-curve compression as well as clipping of high values, both of which affect the chromaticity of the maximal values of *R*, *G*, and *B*.

To test the first hypothesis, the effects on performance of three different preprocessing strategies are tested. To test the second hypothesis, a new dataset for color constancy research has been constructed that consists of images of

▲IS&T Member.

Received May 23, 2011; accepted for publication Apr. 4, 2012; published online May 22, 2012.
1062-3701/2012/56(2)/020501/10/$20.00.

Table I. Performance comparison using Barnard et al.[1,18] set of 321, 637-by-468-pixel, 16-bit-linear (gamma = 1) images of indoor scenes. The algorithms compared are as follows: Do-Nothing, Grayworld, Shades-of-Gray[4], Edge-Based[2] (first and second order) (implementation from Ref. 17, MaxRGB without preprocessing[13], MaxRGB after 5 × 5 median filtering, MaxRGB after van de Weijer's[17] clipped-pixel removal, MaxRGB after van de Weijer's clipped-pixel removal followed by 5 × 5 median filtering, the Neural Network approach[11], Forsyth's CRULE[12], and CbyC[1] as reported by various authors. The results report by Hordley and Finlayson[19] were based on a subset consisting of 310 of the 321 images.[20] Gijsenij et al.[3] trained on part of the data set and tested on the remaining 290 images. Boldface indicates the minimum in the respective column. The Do-Nothing error is the error in simply assuming the scene illumination is always white (i.e., estimating its chromaticity as $r = g = b = 1/3$).

	Angular difference				L2 distance × 100			
Methods tested on 321 linear image set	Median	Mean	RMS	Max	Median	Mean	RMS	Max
Do-Nothing	15.6	17.9	20.5	37.0	10.0	11.6	13.7	25.8
GW (our code)	6.9	9.7	13.7	36.1	5.6	7.8	10.8	32.7
SoG[4] (our code)	4.1	6.3	9.0	28.7	2.9	4.4	6.2	19.8
EB1[2] (code from Ref. 17)	3.7	6.1	**8.5**	27.7	2.6	4.3	6.0	19.0
EB2[2] (code from Ref. 17)	4.5	6.8	9.1	35.6	3.1	4.7	6.2	23.8
MaxRGB w/o preprocessing	6.5	9.2	12.3	36.2	4.5	6.3	8.3	25.0
MaxRGB with uniform averaging (Barnard et al.[1] Table V)						5.3		
MaxM	3.4	5.8	9.0	31.0	**2.3**	4.1	6.1	21.0
MaxRGB clipped removal (MaxC)	6.5	9.1	12.2	36.2	4.5	6.3	8.2	25.0
MaxRGB clipped removal plus median filter (MaxCM)	**3.1**	**5.6**	**8.5**	**25.7**	**2.3**	**3.9**	**5.8**	**16.7**
Neural Network[11] (Barnard et al.[1] Table V)						6.0		
CbyC[9] (Barnard et al.[1] Table V)						6.1		
CRULE[12] (Barnard et al.[1] Table V)						4.3		
ECRULE-MV (Barnard variation on CRULE) (Barnard et al.[1] Table V)						4.0		
CbyC Finlayson et al.[10]		9.9						
CbyC I (bright pixels only) Hordley and Finlayson[19] Table VII (310 out of 321)	3.2	6.6	10.1					
CbyC Gijsenij et al.[3] (290 out of 321)	6.8	9.9						
CbyC Fredembach and Finlayson[21]	6.0	6.3						

105 scenes. For each scene, there are images in both high-dynamic-range and multiple-exposure formats, and images with and without Macbeth mini-Colorchecker charts, from which the chromaticity of the scene illumination is measured. This data set is now available online.[14] Tests described below show that MaxRGB performs as well on this data set as other representative and recently published algorithms. The results reported here significantly extend, combine, compare, and further analyze those of two earlier studies[15,16] in terms of the algorithms compared, the databases used, and the preprocessing strategies tested. Based on these new tests, we now able to conclude that MaxRGB works as well or better than the well-known Color by Correlation[9] method.

EVALUATION OF PREPROCESSING FOR MaxRGB

Three preprocessing methods are considered as follows: (1) removal of each clipped pixel along with those in its surrounding 3-by-3 neighborhood;[17] (2) median filtering; and (3) the combination of condition (1) followed by condition (2). A clipped pixel is defined as one for which at least one channel is maximal (i.e., for n-bit images when $R = 2^n - 1$ or $G = 2^n - 1$ or $B = 2^n - 1$). Since many authors have used the set of 321 indoor images from the Simon Fraser University image database created by Barnard et al.[18] (which includes the true illumination as measured with a spectroradiometer), we use it here for comparison to show how significantly MaxRGB improves with simple preprocessing. The results are tabulated in Table I. The table includes the results of MaxRGB with and without preprocessing along with the corresponding results published by Barnard et al.,[1] and by van de Weijer et al.[2] along with results of the van de Weijer et al. MATLAB implementation.[17] Barnard's method involved smoothing by a uniform averaging. Van de Weijer's implementation removes clipped pixels and, as well, the pixels in the 3×3 surrounding neighborhood of each clipped pixel. Also, included are the results for the Do-Nothing method (i.e., the illumination for all images is estimated to have chromaticity $r = g = b = 1/3$), Grayworld (GW), Edge-Based, and Color by Correlation.

The Edge-Based method is included in Table I as being representative of the performance of the majority of current illumination-estimation algorithms. It also has the advantage that code is available online[17] and it does not require training. Hence it is less susceptible to variations in how it is applied. Furthermore, van de Weijer et al. conclude with respect to the Edge-Based method, "The experimental results show that the newly proposed simple color constancy algorithms obtain similar results as more complex state-of-the-art color constancy methods"[2] (Page 2213). As another standard for comparison, Table I also includes published results concerning the widely varying performance of the well-known Color by Correlation (CbyC) algorithm applied to this same data, as quoted from various sources.[1,3,10,21]

For consistency, all the methods are applied to exactly the same images without any *additional* preprocessing. The assumption is that these methods in their original form without any additional preprocessing is what their originators intended to be used. Whether or not they might be improved by additional preprocessing is a question that is not addressed here. However, for several methods (e.g., Grayworld, Shades of Gray (SoG), Edge-Based) preprocessing is unlikely to make much difference, because they are based on averaging the results from many image locations. In the case of MaxRGB, preprocessing is important because each maximum is derived from a single spatial location. The tests are based on the implementations of the methods that are available online. The source of each implementation is identified in the tables of results.

The performance of the algorithms is measured in terms of the difference between the measured illumination chromaticity and that estimated by each method. The chromaticity difference is evaluated both in terms of angular difference and Euclidean distance (see Appendix). Of course, in terms of correcting the image colors in response to a change in the illuminant (i.e., for "color constancy" or white balancing), estimating the illuminant is only the first step in a two-step process. Given the chromaticity of the illuminant, a common approach is to adjust all colors by scaling the R, G, B channels based on the ratio of the chromaticity of the target illuminant to the chromaticity of the scene illuminant. This method is referred to variously as the diagonal model, the coefficient rule, or the von Kries model.[22] However, the magnitude of the error in estimating the illuminant tends to dominate the error in applying the diagonal model. Since all the methods compared in the tests below generate an estimate of the chromaticity of the illumination, which would then be followed by an identical second color correction step, we evaluate the error in the estimates themselves rather than in terms of the differences in the resulting adjusted images.

In Table II, the Wilcoxon signed-rank test is used to evaluate the statistical significance of the performance differences of the algorithms reported in Table I. Table II shows that MaxRGB with the proposed preprocessing performs better than all the other methods (for which the necessary data are available) listed in Table I to a statistically significant degree. For the remaining ones, which include Forsyth's CRULE,[12] ECRULE-MV (a variation on Forsyth's CRULE[12], and the neural network approach,[11] all have mean L2 (Euclidean) distances listed in Table I exceeding those of MaxCM (MaxRGB with clipped-pixel removal followed by median filtering). Figure 1 graphs the errors reported in Table I.

Table II. Wilcoxon signed-rank test (confidence level 95%) comparison of the algorithms listed in Table I. A "+," "−," and "=" indicate that the algorithm listed in the row is statistically better, worse, or equivalent, respectively, than the one in the corresponding column. Labels are as follows: GW, SoG, EB1, EB2, CbyC, MaxP (MaxRGB plain without preprocessing), MaxM (MaxRGB with median filtering), MaxC (MaxRGB after clipped-pixel removal), and MaxCM (MaxRGB after clipped-pixel removal and then median filtering). Note that MaxCM outperforms all the other non-MaxRGB methods from Table I. These Wilcoxon signed-rank test results are based on illumination estimates obtained by CbyC as provided by Hordley.[20] For compatibility with Hordley's results, this table is based on 310 of the 321 images (excludes jersey_syl-wwf, munsell2_syl-50MR16Q, munsell2_syl-wwf, munsell3_syl-50MR16Q, munsell4_syl-wwf, books-1_ph-ulm, books-1_solux-3500 + 3202, books-1_solux-4700, books-4_ph-ulm, books-4_syl-cwf, plastic-1_solux-3500 + 3202).

	GW	SoG	EB1	EB2	C-by-C	MaxP	MaxM	MaxW	MaxCM
GW	=								
SoG	+	=							
EB1	+	=	=						
EB2	+	−	−	=					
CbyC	+	=	=	+	=				
MaxP	=	−	−	−	−	=			
MaxM	+	+	+	+	−	+	=		
MaxC	=	−	−	−	−	=	−	=	
MaxCM	+	+	+	+	+	+	=	+	=

Another existing color constancy test data set is the gray ball set consisting of the 11,346 images Ciurea obtained using a digital video camera with a gray ball attached.[23] The gray ball provides the chromaticity of the illumination for each image. Clipping and tone-curve compression effects are even more likely in these images than those from the 321 set, since they are single frames extracted from digital video. In contrast, the images in the 321 set were obtained by averaging a large number of frames so as to reduce noise. The 321 set is also linear, whereas the 11,346 set is nonlinear, although the actual gamma or tone curve used is unknown. Table III compares the errors across the various methods, which include all those from Table I plus the more recent Leave-N-out n-jet method.[3] A graphical representation of the errors in Table III is provided in Fig. 1. Overall, preprocessed MaxRGB can be seen to be competitive with all the other methods listed. Table IV tabulates the results of the Wilcoxon signed-rank test, which support this conclusion, especially for the cases of MaxRGB with clipped-pixel removal (MaxC) and MaxRGB with clipped removal followed by median filtering (MaxCM).

Another test data set is 568-image "colorchecker" dataset provided by Gehler et al.[8] The images in it were taken with two digital single lens reflex cameras (Canon 5D and Canon1D), with all settings in automatic mode. All images were saved in Canon RAW format. The dataset also includes TIFF versions created from the RAW images using the automatic mode of the Canon DIGITAL PHOTO PROFESSIONAL program to convert the images into TIFF images. Each image contains a Macbeth colorchecker for reference. The

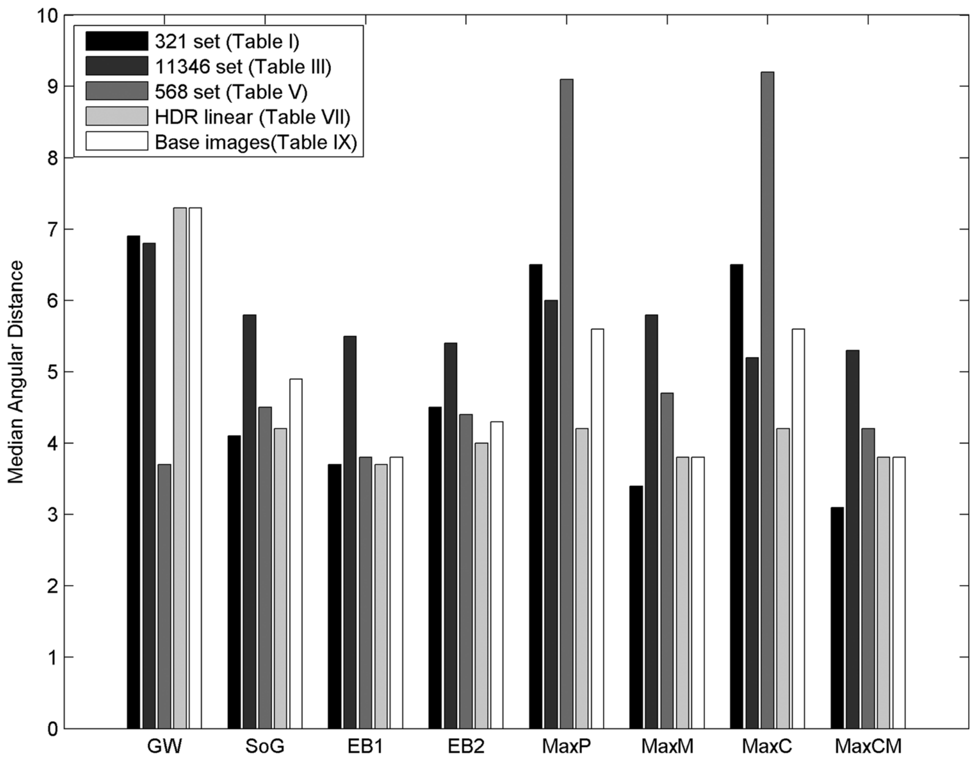


Figure 1. Performance comparison in terms of median angular on the different data sets of the methods GW, SoG, Edge-Based first order (EB1), Edge-Based second order (EB2), MaxP, MaxM, MaxC, and MaxCM. The data for the plots are taken from Tables I (321 set), III (11346 set), V (568 Colorchecker set), VII (HDR linear), and IX (base images).

Table III. Performance comparison on the set of 11,346 images from the Ciurea[23] data set. These images are nonlinear, but the actual gamma value is unknown. The original images are 360×240 but are cropped to 240×240 in order to eliminate the gray ball. The Ciurea data set is divided into 15 separate groups of images based on location. The result of Gijsenij et al.[3] for Color by Correlation is based on training on 14 of the 15 groups and testing on the single remaining group, repeated for all 15 groups. The Leave-N-Out n-jet result is the best of those reported by Gijsenij et al.[3] Labels are as in Table I. Because these images are lower resolution than those in the 321 set, the median filter's size was reduced to 3-by-3.

	Angular-distance				L2 distance × 100			
Methods tested on 11,346 image set	Median	Mean	RMS	Max	Median	Mean	RMS	Max
Do-Nothing	6.7	8.3	11	**27**	4.6	5.9	7.9	22
GW (our code)	6.8	7.8	9.5	45	5.1	6.0	7.5	42
SoG[4] (our code)	5.8	6.7	**8.1**	36	4.2	**4.9**	**5.9**	28
EB1[2] (code from Ref. 17)	5.5	7.0	8.9	32	4.0	5.2	6.5	23
EB2[2] (code from Ref. 17)	5.4	7.1	9.1	34	3.8	5.2	6.7	25
CbyCfrom Table IV Gijsenij et al.[3]	6.5	8.1						
Leave-N-Out n-jet[3] Table IV	5.5	**6.5**						
MaxRGB (MaxP)	6.0	7.7	10.1	**27**	4.3	5.5	7.4	22
MaxM	5.8	7.6	9.8	28	4.2	5.5	7.2	22
MaxRGB clipped removal (MaxC)	**5.2**	6.7	8.7	**27**	**3.7**	**4.9**	6.4	**21**
MaxRGB clipped removal + median filter (MaxCM)	5.3	6.9	8.8	28	3.8	5.0	6.5	21

image coordinates (measured by hand) of each colorchecker square are provided with the dataset.[8]

Because the TIFF images in the colorchecker dataset were produced automatically, they contain clipped pixels, which are nonlinear (i.e., have gamma or tone-curve correction applied) and demosaiced as well as including the effect of the camera's white balancing. Since all these aspects may lead to problems for the various illumination-estimation methods, we chose to reprocess the raw data and created almost-raw 12-bit Portable Network Graphics

Table IV. Wilcoxon signed-rank test result for the algorithms from Table III tested on the 11,346-image set. A "+" means the algorithm listed in the corresponding row is better than the one in corresponding column; a "−" indicates the opposite; an "=" indicates that the performance of the respective algorithms is statistically equivalent. These Wilcoxon signed-rank test results are based on illumination estimates obtained using CbyC and Leave-N-Out provided by Gijsenij[24]. Labels as listed in Table II.

	GW	SoG	EB1	EB2	C-by-C	N-jet	MaxP	MaxM	MaxC	MaxCM
GW	=									
SoG	+	=								
EB1	+	+	=							
EB2	+	+	+	=						
CbyC	=	−	−	−	=					
N-jet	+	+	=	−	+	=				
MaxP	+	−	−	−	+	−	=			
MaxM	+	=	−	−	+	−	+	=		
MaxC	+	=	+	+	+	+	+	+	=	
MaxCM	+	=	+	+	+	+	+	+	−	=

(PNG) format (lossless compression) images from the Canon RAW format data by decoding them using dcraw.[25] To preserve the original digital counts for each of the RGB channels, demosaicing was not enabled. Both cameras output 12-bit data per channel so the range of possible digital counts is 0 to 4095. The raw images contain 4082×2718 (Canon 1D) and 4386×2920 (Canon 5D) 12-bit values in an RGGB pattern. To create a color image, the two *G* values were averaged, but no further demosaicing was done. This results in a 2041×1359 (for Canon 1D) or 2193×1460 (for Canon 5D) linear image (gamma = 1) in camera RGB space. The Canon 5D has a black level of 129, which was then subtracted. The Canon 1D's black level is zero.

The colorchecker has six achromatic squares. We used the median of the RGB digital counts from the brightest achromatic square containing no digital count > 3300 as the ground-truth measure of the illumination's chromaticity. This choice of threshold eliminates any clipping or possible nonlinearity that may occur for digital counts near 4095 (the maximum possible for a 12-bit image). Since there are six achromatic squares to choose from, this threshold means a bright, but not overexposed square, is used. The median was chosen instead of the mean because the median automatically excludes any of the black pixels surrounding each square that might have been incorrectly included in the square due to the inexactness in the hand labeling of the colorchecker's position.

Table V compares the errors across the various methods, which include all those from Table III. In addition, we also include the results of Gamut Mapping (GM)[10] and Bayes-GT[8] (based on three-fold validation). A graphical representation of the errors in Table V is included in Fig. 1. The corresponding results based on Wilcoxon test are shown in Table VI. To give the Gamut Mapping and the two N-jet methods the best possible training data, the full set of 568 images with the colorcheckers was used. Testing was then done on images with the colorcheckers removed. Sometimes the Gamut Mapping and N-jet methods fail to provide an illumination estimate (four times for Gamut Mapping, 1 for 1-jet). In such cases, we assign the illumination estimate as white with chromaticity (1/3, 1/3). Since this dataset consists of images from two different camera models, training and testing were done separately for each of the two corresponding image subsets. The results were then combined. For Bayes-GT,[8] it would also have been interesting to be able to include leave-one-out results in Tables V and VI as well, but unfortunately the method is computationally intensive and not practical. (Bayes-GT requires 4 min per image, so even the three-fold validation required 2 days of computation. It seems doubtful that it could ever be sped up enough to be of any more than theoretical interest.) Overall, preprocessed MaxRGB can be seen to be competitive with the other methods. Table VI tabulates the results of the Wilcoxon signed-rank test, which support this conclusion, especially for the case of MaxRGB with clipped removal followed by median filtering (MaxCM). On this dataset, simple Grayworld outperforms all the other methods.

THE MULTIEXPOSURE AND HDR IMAGE DATASETS

In order to determine how much the quality of the image data affects MaxRGB's performance, a new test data set of images of 105 scenes captured using a Nikon D700 digital still camera were created. The camera's autobracketing was used to capture a sequence of nine images with 1 EV (exposure value) difference between each image in the sequence. The rate of capture was 5 frames per second. The exposure range was set to ensure that in each set there would be at least one image with maximum digital count less than 10,321. During bracketing, the camera was set to allow it and to adjust the shutter speed and/or the aperture setting automatically between frames in order to change the exposure by 1 EV. In other words, the f-stop setting was not fixed. Although this is not standard practice when creating high-dynamic range (HDR) images, because varying the aperture causes a variation in depth of focus, our goal was to capture the widest range of exposures without resorting to long exposure times. The resulting images are not necessarily of optimal sharpness for HDR viewing, but they are of much higher resolution and much sharper than the images typically used for testing illumination-estimation methods. The results are reported below for both the HDR images and the individual single-exposure images from which they were assembled.

All images were recorded in Nikon's electronic (NEF) raw data format.[26] The raw images were then processed in two ways: The first was to create almost-raw, 16-bit PNG format (lossless compression) images from the NEF data, one image per exposure value. We will refer to these 16-bit PNGs as the "base images." The second was to create a set of HDR images from the base images. Both the multiple-exposure base images and the HDR images are used in

Table V. Performance comparison on the set of 568 images from the colorchecker[8] data set. These images are linear. Labels are as listed in Table III. The median filter's size is 5×5.

	Angular-distance				L2 distance $\times$ 100			
Methods tested on the colorchecker set	Median	Mean	RMS	Max	Median	Mean	RMS	Max
Do-Nothing	4.8	9.3	13	37	3.1	6.5	9.3	30
GW (our code)	3.7	4.8	6.2	25	2.6	3.4	4.5	20
SoG[4] (our code)	4.5	6.4	8.7	36	3.5	5.4	7.5	37
EB1[2] (code from Ref. 17)	3.8	6.6	9.4	38	3.0	5.5	8.0	40
EB2[2] (code from Ref. 17)	4.4	7.2	10	47	3.5	6.1	8.7	50
GM (code from Ref. 3)	4.3	6.2	8.4	32	3.2	5.0	6.8	24
N-jet complete 1-jet (1-jet) (code from Ref. 3)	4.2	6.0	8.2	32	3.2	4.8	6.5	24
N-jet complete 2-jet (2-jet) (code from Ref. 3)	4.1	5.9	8.0	32	3.1	4.6	6.3	24
Bayes (three-fold) (BAY) (code from Ref. 8)	5.8	7.1	8.9	34	5.0	5.9	7.3	28
MaxRGB (MaxP)	9.1	10	13	51	7.8	9.0	12	55
MaxM	4.7	7.7	11	42	3.6	6.4	9.1	44
MaxRGB clipped removal (MaxC)	9.2	10	13	51	7.8	9.0	12	55
MaxRGB clipped removal + median filter (MaxCM)	4.2	7.2	10	41	3.2	6.0	8.6	37

Table VI. Wilcoxon signed-rank test result for the algorithms from Table V tested on the colorchecker set of 568 images. A "+" means the algorithm listed in the corresponding row is better than the one in corresponding column; a "−" indicates the opposite; an "=" indicates that the performance of the respective algorithms is statistically equivalent. The labels "1-jet" and "2-jet" stand for N-jet (complete 1-jet) and N-jet (complete 2-jet), respectively. "BAY" stands for Bayes-GT[8] tested using three-fold validation. The rest of the labels are as given in Table IV.

	GW	SoG	EB1	EB2	GM	1-jet	2-jet	BAY	MaxP	MaxM	Max64	MaxC	MaxCM
GW	=												
SoG	−	=											
EB1	−	+	=										
EB2	−	+	−	=									
GM	−	=	−	+	=								
1-jet	−	=	−	+	+	=							
2-jet	−	+	−	+	+	+	=						
BAY	−	−	−	=	−	−	−	=					
MaxP	−	−	−	−	−	−	−	−	=				
MaxM	−	−	−	−	−	−	−	=	+	=			
MaxC	−	−	−	−	−	−	−	−	=	−	−	=	
MaxCM	−	+	−	=	+	=	−	=	+	+	−	+	=

testing MaxRGB against other illumination-estimation algorithms.

Two sets of base images were captured for each scene. One set includes four Gretag Macbeth mini-colorcheckers positioned at different angles with respect to one another. The second set contained images of the same scene, but without the colorcheckers. Between taking the two image sets, the camera was refocused and possibly moved slightly. For the first set, the focus was adjusted, so the colorcheckers were in focus. For the second set, the focus was optimized for the scene overall.

The colorcheckers were placed in the scene at a point where the illumination incident on them was expected to be representative of the color of the overall scene illumination. While all scenes contain some variation in the illumination color because of inter-reflections, scenes that clearly had strong variations in illumination color were avoided. For example, a room with interior tungsten lighting mixed with daylight entering through a window would be excluded.

To create the base images, the raw NEF images were decoded using dcraw.[25] To preserve the original digital counts for each of the RGB channels, demosaicing was not enabled. The camera outputs 14-bit data per channel, so the range of possible digital counts is 0 to 16,383. The raw images contain 4284×2844 14-bit values in an RGGB pattern. To create a color image, the two G values were averaged, but no further demosaicing was done. This results in a 2142×1422 RGB image.

An HDR image was also constructed from each set of base images. The base images require alignment, which was done by the simple Median Threshold Bitmap approach.[27] After applying a 3×3 median filter to the base images, the MATLAB function *makehdr* from the MATLAB Image Processing Toolbox[28] was used to combine them into one HDR image. To ensure the reliability of the pixel values, all base image pixels having values greater than 13,004 or less than 30 were excluded. MATLAB's *makehdr* function requires the relative exposure (RE) value of each base image, which is calculated as

$$RE = 2^{-EV} \cdot \frac{N_0^2}{t_0} \cdot \frac{S}{S_0}, \qquad (1)$$

Table VII. Performance comparison of illumination-estimation methods evaluated on the linear (gamma = 1) HDR image data in terms of the angular error and Euclidean distance metrics between the measured and estimated chromaticities of the illumination. Labels are as given in Table I. The row labeled colorchecker gives the statistics of the difference between the RGBs of each of the four colorchecker's whites and their collective median calculated over all 105 scenes. Since these images are much higher resolution than those in the 321 set, the median filter's size was increased to 15-by-15.

	Angular-distance				L2 distance × 100			
Methods tested on HDR image set	Median	Mean	RMS	Max	Median	Mean	RMS	Max
Do-Nothing	14.7	15.1	15.5	29.8	14.5	14.2	14.5	21.5
GW (our code)	7.3	7.9	9.6	23.8	4.8	5.7	7.0	21.7
Shades of Gray[4] (our code)	4.2	6.0	8.1	24.7	2.9	4.4	5.9	17.4
EB1[2] (code from Ref. 17)	3.7	5.9	8.0	24.2	2.9	4.5	6.0	17.1
EB2[2] (code from Ref. 17)	4.0	5.9	7.9	22.9	3.1	4.5	6.0	16.1
Bayes (three-fold) (BAY) (code from Ref. 8)	5.9	8.0	10	29	4.1	5.9	7.6	20.0
MaxRGB (MaxP)	4.2	6.5	8.8	28.6	3.0	4.9	6.6	20.4
MaxM	3.8	6.3	8.5	25.5	3.1	4.7	6.3	17.8
MaxRGB clipped removal (MaxC)	4.2	6.5	8.8	28.6	3.0	4.9	6.6	20.4
MaxRGB clipped removal plus median filter (MaxCM)	3.8	6.3	8.5	25.5	3.1	4.7	6.3	17.8
Colorchecker	0.9	1.9	3.0	15.5	0.75	1.6	2.6	11.4

Table VIII. Performance tested on the nonlinear (gamma = 2.2) HDR images. Labels as in Table I. As in the linear case, the Wilcoxon signed-rank test on these results indicates that with the exception of Grayworld, which is worse, the other algorithms are statistically equal at the 95% confidence level.

	Angular-distance				L2 distance × 100			
Methods tested on HDR image set	Median	Mean	RMS	Max	Median	Mean	RMS	Max
Do-Nothing	7.0	7.4	7.8	17.8	6.8	6.8	6.9	11.5
GW (our code)	4.0	4.4	5.1	12.8	2.7	3.0	3.5	11.4
SoG[4] (our code)	2.3	3.3	4.4	13.4	1.5	2.3	3.0	8.4
EB1[2] (code from Ref. 17)	2.6	3.4	4.4	13.5	1.9	2.5	3.3	9.5
EB2[2] (code from Ref. 17)	2.6	3.2	4.2	13.5	1.9	2.5	3.1	9.6
MaxRGB (MaxP)	2.2	3.4	4.6	15.7	1.6	2.4	3.3	9.5
MaxM	2.1	3.2	4.3	13.1	1.5	2.3	3.1	9.0
MaxRGB clipped removal (MaxC)	2.1	3.2	4.3	12.3	1.5	2.4	3.1	8.2
MaxRGB clipped removal plus median filter (MaxCM)	2.2	3.4	4.6	15.7	1.6	2.4	3.3	9.5
Colorchecker	0.48	0.95	1.5	7.9	0.35	0.77	12.5	6.0

where $EV = \log_2(N^2/t)$. N is the relative aperture (f-number), t is the exposure time (shutter speed) in seconds, and S is the ISO. N_0, S_0, and t_0 are constants related to the camera, but which in the present case can be chosen arbitrarily since all that is required is the relative exposure. They were set ($N_0 = 16$, $t_0 = 1/8000$, $S_0 = 100$) such that the resulting REs are positive integers. The final HDR images may vary slightly in size due to possible cropping at the boundaries of the images as they are aligned.

MEASURING THE SCENE ILLUMINATION

The illumination chromaticity is determined by manually sampling the RGB digital counts from each of the four white patches from the colorcheckers of the base images. The brightest image from each set not containing any over-exposed pixels (i.e., all RGB $< 2^{14} - 1$) anywhere within the colorcheckers is used, and each measurement is the average RGB of the 3×3 neighborhood of a pixel near the center of the white patch. Since the colorcheckers differ in orientation, we obtain measurements of the scene illumination at four different angles of incidence. Not surprisingly these measurements do not always agree. For the tests described below, the median of the illumination chromaticities from the four colorcheckers is used as the ground truth.

Taken over the 105 scenes, the median, mean, and maximum angular difference between the RGBs of each of the four patches and their collective median is given in the last row of Table VII for the linear case, and in the last row of Table VIII for the nonlinear case. Since we do not expect the performance of an illumination-estimation method to surpass that of direct measurement of the illumination—and given that all four colorcheckers represent the chromaticity of the "true" illumination—these values represent a lower bound on the mean, median, and maximum illumination-estimation errors possible for any algorithm.

TESTS OF MaxRGB ON HDR IMAGES

MaxRGB,[13] Grayworld,[29] Shades-of-Gray,[4] and Grayedge[2] were run on the HDR images of the scenes without the colorcheckers. The results are shown in Tables VII and VIII for linear and nonlinear HDR image data, respectively. A graphical representation of the errors in Table VII is shown in Fig. 1. The errors for the nonlinear case are smaller than for the linear case, which is mainly due to the fact that gamma compresses the range of RGB values and hence the errors.

TESTS OF MaxRGB ON MULTIEXPOSURE IMAGES

Although each bracketed image has been assembled into a single HDR image, we also test the illumination-estimation methods on the individual base images from each bracketed image set (the ones without the colorcheckers in them) to measure their performance as a function of exposure. We evaluate the effect of clipping on MaxRGB performance by considering the sequence of bracketed images that have their maximum digital counts below a specified threshold. As shown in Figure 2, for images without clipping, the illumination-estimation error for MaxRGB remains low and relatively constant until a sharp rise at the point when the maximum scene radiance exceeds the 14-bit range of the camera. At that point, the high digital counts are clipped to the maximum value of 16,383 and MaxRGB fails. As more and more clipping occurs, we can expect MaxRGB eventually to approximate the Do-Nothing algorithm, which simply estimates the scene illumination as always being white, because the RGB maximum values in every 14-bit image will always be $R = G = B = 16{,}383$.

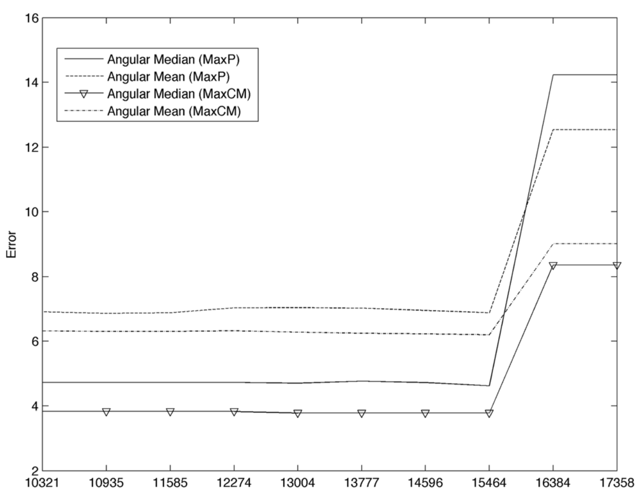


Figure 2. Median, mean, root mean square, and maximum of the angular error between the estimated illumination chromaticity and the measured illumination chromaticity as a function of the threshold on maximum digital count allowed within the image for MaxP (MaxRGB without preprocessing) and MaxCM applied to the base images. The sharp increase in error occurs at the point at which the radiance exceeds the 14-bit range of the camera. In the case of MaxCM, its preprocessing helps to reduce the error, but clipping still leads to a significant increase.

Although the plot in Fig. 2 shows the error to be relatively constant under 16,000, there is a slight dip around 11,000. Using this as a cutoff, MaxRGB was tested on images with a maximum *R*, *G*, or *B* digital count of 11,000 or less. The results are tabulated in Table IX for linear image data. A graphical representation of the angular errors in Table IX is shown in Fig. 1.

CONCLUSION

MaxRGB was found to work well when clipping was avoided and the full dynamic range of the scene was preserved. Two different methods of capturing and representing the scene's full dynamic range were used. The first was to capture a set of images of different exposure and then to run MaxRGB on the brightest image, whose exposure did not exceed a specified threshold. If the 14-bit exposure range of the camera was exceeded and clipping occurred, it will have significant negative impact to MaxRGB's performance. The second method of preserving the dynamic range was to construct an HDR image from the set of multiple exposures and then apply MaxRGB to the result. In this case, MaxRGB worked as well or better than other representative algorithms. However, as was the case with other illumination-estimation methods, MaxRGB relies on the assumption that the chromaticity of the illumination is constant throughout the scene. If there are multiple sources of illumination, especially a light source of different chromaticity appearing directly in the image, MaxRGB may fail.

Tests also showed that simple preprocessing of the image data by removing clipped pixels and applying median filtering can significantly reduces the MaxRGB error on Barnard's[1] standard 321-image test set. Similarly, on the large Ciurea[23] data set, it reduces the error to the point that MaxRGB becomes very competitive with other methods, including Color by Correlation[9] and Leave-N-Out n-jet.[7]

Overall, the results presented here show that the poor performance of MaxRGB previously reported in the literature may have more to do with the consequences of the limited exposure range (8 to 14 bits per channel) of digital still cameras than the failure of MaxRGB's fundamental assumption that every scene contains some region, or combination of regions, that reflects maximally in each of the *R*, *G*, and *B* sensitivity ranges. This contradicts the widely held view expressed, for example, by Fredembach and Finlayson[30] that "…one has, in general, no control as to which reflectances are present in a scene. This uncertainty is the reason why simple estimation methods such as gray-world and MaxRGB are unreliable (if every scene contained a 'white-like' patch, MaxRGB would be very accurate)." Although it certainly is true that scenes without white-like patches can occur, the ill-posed nature of the illumination-estimation problem means that all illumination-estimation algorithms are based on assumptions of one sort or another that may

Table IX. Performance tested on the linear (gamma = 1) base images. Labels as in Table I. The Wilcoxon signed-rank test on these results indicates that Grayworld is the worst and that the other algorithms are statistically equal. The results for MaxRGB with and without clipping followed by median filtering are identical because in the multiple-exposure set there are no clipped pixels, so no pixels are in fact removed.

	Angular-distance				L2 distance × 100			
Methods tested on base images	Median	Mean	RMS	Max	Median	Mean	RMS	Max
Do-Nothing	14.7	15.1	15.5	29.8	14.5	14.2	14.5	21.5
GW (our code)	7.3	7.8	9.5	23.3	4.6	5.6	7.0	21.3
SoG[4] (our code)	4.9	6.2	8.2	24.1	3.3	4.7	6.1	17.0
EB1[2] (code from Ref. 17)	3.8	5.9	8.0	23.4	3.0	4.4	5.9	16.5
EB2[2] (code from Ref. 17)	4.3	6.0	8.0	22.6	3.0	4.5	6.0	15.8
Bayes (three-fold) (BAY) (code from Ref. 8)	6.6	8.3	11.0	33.0	4.1	6.2	8.2	24.0
MaxRGB (MaxP)	5.6	7.8	10.1	31.2	4.5	6.3	8.2	28.0
MaxM	3.8	6.3	8.5	25.4	3.0	4.8	6.3	17.8
MaxRGB clipped removal (MaxC)	5.6	7.9	10.1	31.2	4.5	6.3	8.2	28.0
MaxRGB clipped removal plus median filter (MaxCM)	3.8	6.3	8.5	25.0	3.0	4.8	6.3	17.8
Colorchecker	0.93	1.9	3.0	15.5	0.75	1.6	2.6	11.4

from time-to-time be violated. The results reported here show that the MaxRGB assumption is no more likely to be violated than the assumptions underlying the other representative algorithms that were tested.

ACKNOWLEDGMENTS

Funding was provided by the Natural Sciences and Engineering Research Council of Canada. The authors thank Steve Hordley and Arjan Gijsenij for providing the output data from their algorithms.

APPENDIX

Evaluation Measures

The illumination chromaticity $[r, g, b]$ is defined in camera sensor space as $r = R/(R+G+B)$, $g = G/(R+G+B)$, and $b = B/(R+G+B)$. Two different error measures are used to measure the accuracy of an illumination estimate. The first is the Euclidean distance between the r and g components of the measured illumination chromaticity $[r_a, g_a]$ and the estimated illumination chromaticity $[r_e, g_e]$

$$\mathrm{E}_{L2} = \left[(r_a - r_e)^2 + (g_a - g_e)^2\right]^{1/2} \tag{A1}$$

The second error measure is the angular difference in degrees between the measured chromaticity $[r_a, g_a, b_a]$ and the estimated chromaticity $[r_e, g_e, b_e]$ defined as

$$E_{angular} = \cos^{-1}\left[\frac{(r_a, g_a, b_a)\cdot(r_e, g_e, b_e)}{\sqrt{r_a^2 + g_a^2 + g_a^2}\cdot\sqrt{r_e^2 + g_e^2 + g_e^2}}\right]\cdot\frac{360}{2\pi}. \tag{A2}$$

If a method provides only the r and g components, the third component can be obtained as $b = 1 - r - g$. If an estimate includes a chromaticity component that is less than zero, it is set to zero. Note that Gijsenij et al.[31] found a good correlation between a perceptual error measure and the angular error used here.

For the distance error, we also compute the root mean square (RMS), mean, and median errors over a set of N test images. It has been argued that the median is the most appropriate metric for evaluating color constancy.[19] The standard RMS is defined as

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} E_i^2}, \tag{A3}$$

where E_i can be either L2 or angular distances.

REFERENCES

1. K. Barnard, L. Martin, A. Coath, and B. Funt, “A comparison of computational color constancy algorithms—Part II: Experiments with images”, IEEE Trans. Image Process. **11**, 985–996 (2002).
2. J. van de Weijer, T. Gevers, and A. Gijsenij, “Edge-based color constancy”, IEEE Trans. Image Process. **16**, 2207 (2214).
3. A. Gijsenij, T. Gevers, and J. van deWeijer, “Generalized gamut mapping using image derivative structures for color constancy”, Int. J. Comput. Vis. **86**, 127–139 (2010).
4. G. Finlayson and E. Trezzi, “Shades of gray and colour constancy”, *Proc. IS&T/SID 12th Color Imaging Conf.* (IS&T, Springfield, VA, 2004), pp. 37–41.
5. E. Land and J. McCann, “Lightness and retinex theory”, J. Opt. Soc. Am. A **61**, 1 (1971).
6. B. Funt, F. Ciurea, and J. J. McCann, “Retinex in matlab”, J. Electron. Imaging **13**, 68–72 (2004).
7. E. H. Land, “The retinex theory of color vision”, Sci. Am. **237**, 108–128 (1977).
8. P. V. Gehler, C. Rother, A. Blake, T. P. Minka, and T. Sharp, “Bayesian color constancy revisited”, *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (IEEE, Anchorage, Alaska, 2008), pp. 1–8.
9. G. Finlayson, S. Hordley, and P. Hubel, “Color by correlation: A simple, unifying framework for color constancy”, IEEE Trans. Pattern Anal. Mach. Intell. **23**, 1209–1221 (2001).
10. G. Finlayson, S. Hordley, and I. Tastl, “Gamut constrained illumination estimation”, Int. J. Comput. Vis. **67**, 93 (109).

[11] V. Cardei, B. Funt, and K. Barnard, "Estimating the scene illumination chromaticity using a neural network", J. Opt. Soc. Am. A **19**, 2374–2386 (2002).

[12] D. Forsyth, "A novel algorithm for color constancy", Int. J. Comput. Vis. **5**, 5–36 (1990).

[13] B. Funt, K. Barnard, and L. Martin, "Is machine colour constancy good enough?", *ECCV'98 5th European Conf. on Computer Vision* (ECCV, Freiburg, Germany, 1998), pp. 445–459.

[14] www.cs.sfu.ca/~colour/data/.

[15] B. Funt and L. Shi, "The effect of exposure on MaxRGB color constancy", Proc. SPIE **7527**, 75270Y (2010).

[16] B. Funt and L. Shi, "The rehabilitation of MaxRGB", *Proc. IS&T/SID 18th Color Imaging Conf.* (IS&T, Springfield, VA, 2010), pp. 256–259.

[17] J. van de Weijer, "MATLAB implementation of max-RGB, http://lear.inrialpes.fr/people/vandeweijer/software, accessed April 2, 2010.

[18] K. Barnard, L. Martin, B. Funt, and A. Coath, "A data set for color research", Color Res. Appl. **27**, 147–151 (2002).

[19] S. Hordley and G. Finlayson, "A re-evaluation of colour constancy algorithm performance", J. Opt. Soc. Am. A **23**, 1008–1020 (2006).

[20] S. Hordley, private communication (July 19, 2010).

[21] C. Fredembach and G. Finlayson, "Bright-chromagenic algorithm for illuminant estimation", J. Imaging Sci. Technol. **52**, 040906 (2008).

[22] G. Finlayson, M. Drew, and B.V. Funt, "Color constancy: Generalized diagonal transforms suffice", J. Opt. Soc. Am. A **11**, 3011–3020 (1994).

[23] F. Ciurea and B. Funt, "A large image database for color constancy research", *Proc. IS&T/SID 11th Color Imaging Conf.* (IS&T, Springfield, VA, 2003), pp. 160–164.

[24] A. Gijsenij, private communication (July 16, 2010).

[25] Code for dcraw obtained from http://en.wikipedia.org/wiki/dcraw, accessed July 24, 2009.

[26] Nikon Electronic Format (NEF), Nikon USA, http://www.nikonusa.com/Learn-And-Explore/Nikon-Camera-Technology/ftlzi4ri/1/Nikon-Electronic-Forma-NEF.html, accessed July 24, 2009.

[27] G. Ward, "Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures", J. Graphics Tools **8**, 17 (30).

[28] MATLAB Image Processing Toolbox 7.0, The Mathworks, Inc.

[29] G. Buchsbaum, "A spatial processor model for object colour perception", J. Franklin Inst. **310**, 337 (250).

[30] C. Fredembach and G. Finlayson, "The bright-chromagenic algorithm for illuminant estimation", J. Imaging Sci. Technol. **52**, 040906 (2008).

[31] A. Gijsenij, T. Gevers, and M. P. Lucassen, "Perceptual analysis of distance measures for color constancy algorithms", J. Opt. Soc. Am. A **26**, 2243–2256 (2009).

For information on reprints or reproduction contact
Donna Smith
Production Editor
The Journal of Imaging Science and Technology
Society for Imaging Science and Technology
7003 Kilworth Lane
Springfield, Virginia 22151 USA
703/642-9090 extension 107
703/642-9094 (fax)
dsmith@imaging.org
www.imaging.org