Preprint. Final version appears in

Proceedings Eighteenth IS&T Color Imaging Conference, San Antonio, Nov. 2010

The Rehabilitation of MaxRGB

Brian Funt and Lilong Shi; Simon Fraser University; Vancouver, B.C., Canada

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, the results presented here show that in fact MaxRGB works surprisingly well when tested on a new dataset of 105 high dynamic range images, and also better than previously reported when some simple pre-processing is applied to the images of the standard 321 image set [1]. The HDR images in the dataset for color constancy research were constructed in the standard way from multiple exposures of the same scene. The color of the scene illumination was determined by photographing an extra HDR image of the scene with 4 Gretag Macbeth mini Colorcheckers at 45 degrees relative to one another placed in it. With preprocessing, MaxRGB's performance is statistically equivalent to that of Color by Correlation [2] and statistically superior to that of the Greyedge [3] algorithm on the 321 set (null hypothesis rejected at the 5% significance level). It also performs as well as Greyedge on the HDR set. These results demonstrate that MaxRGB is far more effective than it has been reputed to be so long as it is applied to image data that encodes the full dynamic range of the original scene.

Introduction

MaxRGB is an extremely simple method of estimating the chromaticity of the scene illumination for color constancy and automatic white balancing based on the assumption that the triple of maxima obtained independently from each of the three color channels represents the color of the illumination. It is often used as a foil to demonstrate how much better some newly proposed algorithm performs in comparison. However, is its performance really as bad as it has been reported [1,3-5] to be? Is it really any worse than the algorithms to which it is compared?[1]

The prevailing belief in the field about the inadequacy of MaxRGB is reflected in the following two quotations from two different anonymous reviewers criticizing a manuscript describing a different illumination-estimation proposal:

"*Almost no-one uses Max RGB in the field (or in commercial cameras). That this, rejected method, gives better performance than the (proposed) method is grounds alone for rejection*."

"*The first and foremost thing that attracts attention is the remarkable performance of the Scale-by-Max (i.e. White-Patch) algorithm. This algorithm has the highest performance on two of the three data sets, which is quite remarkable by itself*."

We hypothesize that there are two reasons why the effectiveness of MaxRGB may have been underestimated. One is that it is important not to apply MaxRGB naively 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 is that MaxRGB generally has been applied to 8-bit-per-channel, non-linear images, for which there is both significant tone-curve compression and clipping of high intensity values.

To test the pre-processing hypothesis, the effects of pre-processing by median filtering, and resizing by bilinear filtering, are compared to that of the common pre-processing, which simply discards pixels 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 $B=2^n-1$). To test the dynamic-range hypothesis, a new HDR dataset for color constancy research has been constructed which consists of images of 105 scenes. For each scene there are HDR[2] (high dynamic range) images with and without Macbeth mini Colorchecker charts, from which the chromaticity of the scene illumination is measured. This data set is now available on-line[3].

MaxRGB is a special and extremely limited case of Retinex [6]. In particular, it corresponds to McCann99 Retinex [7] when the number of iterations is infinite, or to path-based Retinex [8] without thresholding but with infinite paths. Retinex and MaxRGB both depend on the assumption that either there is a white surface in the scene, or there are three separate surfaces reflecting maximally in the R, G and B sensitivity ranges. 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 digital counts. As a result, the maximum R, G and B digital counts from an image generally do not faithfully represent the corresponding maximum scene radiances. Barnard et al. [9] present some tests using artificial clipping of images that show the effect that lack of dynamic range can have on various illumination-estimation algorithms.

To determine whether or not MaxRGB is really as poor as it is report to be in comparison to other illumination-estimation algorithms, we compare the performance of several algorithms on the new image database. We also find that two simple pre-processing strategies lead to significant performance improvement in the case of MaxRGB. Tests described below show that MaxRGB performs as well on this new HDR data set as other representative and recently published algorithms. We also find that two simple pre-processing strategies lead to significant performance improvement. The results reported here extend those of an earlier study [10] in a number of ways: the size of the dataset

[1] Paper's title inspired by Charles Poynton, "The Rehabilitation of Gamma," *Proc. of Human Vision and Electronic Imaging III SPIE 3299*, 232-249, 1998.

[2] Note that the scenes were not necessarily of high dynamic range. The term HDR is used here to mean simply that that full dynamic range of the scene is captured within the image.

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

is more than doubled, tests now involve HDR images, comparison is made to other algorithms, and the effects of preprocessing methods are studied.

Evaluation of Preprocessing for MaxRGB

Three preprocessing methods are considered: (1) removal of clipped pixels; (2) 5x5 median filtering; and (3) resizing to 64x64 pixels using bicubic interpolation as implemented in Matlab's imresize function. Resizing has the effect of smoothing the image data via weighted averaging, but has the advantage that it also leads to fewer pixels. The choice of 64x64 is based on experimentation.

Since many authors have used the set of 321 indoor images from the Simon Fraser University image database created by Barnard et al. [1], we use it here for comparison to show how significantly MaxRGB improves with simple preprocessing. The results are tabulated in Table 1. 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. [3] along with results of the van de Weijer et al. Matlab implementation of MaxRGB [11]. Barnard's method involved smoothing by a uniform averaging. Van de Weijer's implementation removes the 3x3 neighborhood around each clipped pixel. Also included are the results for the do-nothing method (the illumination for all images is estimated to have chromaticity r=g=b=1/3), Greyworld, and Greyedge. The Greyedge method is included as representative of the performance of the majority of current illumination-estimation algorithms, since as van de Weijer et al. write "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." [3] (p. 2213). The results of other algorithms such as Color by Correlation [2] on this same dataset are given in van de Weijer's Table II [3] (page 2211) where the minimum reported error is for Gamut Constrained Illumination Estimation [12] with a median of 2.6 degrees. GCIE, however, benefits slightly from the possible illuminants being included as a subset of the complete training set of illuminants and so is not considered further here.

Table 1 compares the performance of the various algorithms where it can be seen that preprocessing improves MaxRGB substantially. In fact, the median and mean angular errors actually drop below that of Color by Correlation. The Sign Test and Kolmogorov-Smirnov (K-S) Test (Matlab implementation signtest and kstest2 [13] both find the performance of MaxRGB and Color by Correlation to be statistically equivalent (null hypothesis rejected at the 5% significance level) on the 321 dataset.

The HDR Image Dataset

The HDR Dataset consists of images of 105 scenes captured using a Nikon D700 digital still camera. The camera's auto-bracketing was used to capture up to 9 images of exposures with 1 EV (exposure value) difference between each 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 value less than 10321. During bracketing, the camera was set to allow it to adjust the shutter speed and/or the aperture setting automatically between frames in order to change the exposure by 1EV. In other words, the f-stop setting was not fixed. All images were recorded in Nikon's NEF raw data format [14]. The raw images were first processed to create almost-raw, 16-bit-PNG images from the NEF data, one image per exposure value. We will refer to these 16-bit PNGs as the 'base images'. These base images were used to create a set of HDR (high dynamic range) as described below. Two sets of base images are taken for each scene. One set includes 4 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 Colorchecker frame was in focus. For the second set, the focus was optimized for the scene overall. Figure 1 shows an example of a scene with and without the Colorcheckers.

Table 1: Performance of MaxRGB with various forms of preprocessing evaluated in comparison to the Greyworld, Do-nothing, Greyedge [3,11], and Color by Correlation [2] algorithms on Barnard's [1] set of 321 linear (gamma=1) images of indoor scenes. Boldface indicates the minimum in the respective column. Abbreviations: Mdn (median), Avg (average), RMS (root mean square), Max (maximum), CbyC (Color by Correlation), CbyC Bright (Color by Correlation using bright pixels only [15] and tested on 310 of the 321 images).

Methods tested on	Angular Difference (degrees)				L2 Distance			
321 image set	Mdn	Avg	RMS	Max	Mdn	Avg	RMS	Max
Do-Nothing	16	17	21	37	10	12	13	26
Greyworld	7.1	9.8	14	37	5.7	7.9	11	35
MaxRGB Barnard [1]						5.3		
MaxRGB code of [11]	6.5	9.1	12	36	4.5	6.3	8.2	25
Greyedge code of [11]	3.7	6.1	**8.5**	28	2.6	4.3	6.0	19
MaxRGB	6.5	9.2	12	36	4.5	6.3	8.3	25
MaxRGB (5x5 Median)	3.4	5.8	9.0	31	2.3	4.1	6.1	21
MaxRGB (bicubic)	**3.1**	**5.6**	8.6	**27**	**2.2**	**3.9**	**5.8**	**18**
CbyC from Table V of [1]						6.1		
CbyC Bright Table 7 [15]	3.2	6.6	10					

The Colorcheckers are placed in the scene at a point where the illumination incident on it is expected to be representative of the color of the overall scene illumination. While all scenes contain some variation in the illumination color because of interreflections, scenes that clearly have 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 [16]. 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 to16383. The raw images contain 4284x2844 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 2142x1422 RGB image.

An HDR image was constructed from each set of base images. The base images require alignment, which was done by the simple Median Threshold Bitmap approach [17]. After applying a 3x3 median filter to the base images, the Matlab function makehdr from the Matlab Image Processing Toolbox [13] 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 13004 or less than 30 were excluded. Matlab's makehdr function requires the relative exposure (RE) value of each base image, which is calculated as

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

where $EV = \log_2 \frac{N_0^2}{t}$.

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 can be chosen arbitrarily here 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 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 4 white patches from the Colorcheckers of the base images. Each measurement is the average RGB of the 3x3 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 4 different angles of incidence. Not surprisingly these measurements do not always agree. For the tests described below, the average of the illumination chromaticities from the 4 Colorcheckers is used as the ground truth, but the average is a compromise. Taken over the 105 scenes, the median, mean, and maximum angular difference between the RGBs of each of the 4 patches and their collective median is given in the last row in Table 2 for the linear (gamma=1) case, and in the last row in Table 3 for the non-linear (gamma=2.2) case. Since we cannot expect the performance of an illumination-estimation method to surpass that of direct measurement of the illumination, and given that all 4 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.

***Figure 1**. One image from each of the two bracketed sets for a single scene. The upper image includes the frame holding the 4 Gretag Macbeth mini Colorcheckers, the lower one right excludes it. The Colorcheckers on the top and sides are at 45 degrees with respect to the middle one.*

Tests of MaxRGB on HDR Images

MaxRGB [9], Greyworld [19], Shades-of-Grey [4], and Greyedge [3] were run on the HDR images of the scenes without the Colorcheckers in them. The performance of these algorithms is measured in terms of the difference between the median of the

Table 2: Performance of Do-Nothing, MaxRGB [9], Greyworld [19], Shades-of-Grey [4] and Greyedge [3] evaluated on 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. 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). MaxRGB median uses 5x5 median filtering. MaxRGB bicubic resizes to 64x64. The row labeled "Checkers" gives the statistics of the difference between the RGBs of each of the 4 Colorcheckers' whites and their collective median calculated over all 105 scenes. Abbreviations as in Table 1.

Methods on HDR linear	Angular Difference (degrees)				L2-Distance x 100			
	Mdn	Avg	RMS	Max	Mdn	Avg	RMS	Max
Do-Nothing	15	15	16	30	15	14	15	22
MaxRGB (median)	4.3	6.3	8.4	23	3.0	4.8	6.3	16
GW	7.3	7.9	9.6	23	4.8	5.7	7.0	22
SoG	4.0	6.0	8.1	25	2.9	4.4	5.8	18
Greyedge	3.9	6.0	8.1	25	2.9	4.5	6.0	18
MaxRGB (bicubic)	3.9	6.3	8.6	28	3.0	4.6	6.2	19
Checkers	0.93	1.9	3.0	15	0.75	1.6	2.6	11

measured illumination chromaticity from the 4 Colorchecker white patches and that estimated by each method. The chromaticity difference is evaluated both in terms of angular difference and Euclidean distance. The results are shown in Table 2 and Table 3 for linear and non-linear HDR image data, respectively. The Sign Test evaluates the performance ranking of SoG, Greyedge and MaxRGB as statistically equivalent.

Table 3: Performance for non-linear (gamma=2.2) HDR images. Labels as in Table 2.

Methods on HDR non-linear	Angular Difference (degrees)				L2-Distance x 102			
	Mdn	Avg	RMS	Max	Mdn	Avg	RMS	Max
Do-Nothing	7.0	7.4	7.8	18	6.8	6.8	6.9	11
MaxRGB (median)	2.0	3.2	4.4	13	1.5	2.4	3.2	9.1
GW	4.0	4.4	5.2	13	2.7	3.0	3.6	11
SoG	2.3	3.3	4.4	13	1.6	2.3	3.0	8.4
Greyedge	2.6	3.4	4.5	14	1.9	2.6	3.3	9.8

The errors for the non-linear case are smaller than for the linear case. It is important to note that this is mainly due to the fact that gamma compresses the range of RGB values, and hence the errors as well, rather than because the methods actually work any better with non-linear image data. Unfortunately, it is not uncommon to find in the literature the performance of various algorithms compared across linear versus non-linear image test sets without taking into account the effect that gamma has on the resulting error measures.

Conclusion

MaxRGB was tested on high dynamic range images and found to work well when the full dynamic range of the scene was preserved. In addition, simple preprocessing of the image data with either a median filter or bicubic interpolation significantly reduces the MaxRGB error on the Barnard's [1] standard 321-image test set. Nonetheless, for many digital imaging applications, MaxRGB may still not provide a good enough estimate of the scene illumination. However, given appropriately pre-processed image data of adequate dynamic range, its performance is not very different from that of the other illumination-estimation algorithms tested.

MaxRGB is based on the assumption that there is a white or white-equivalent reflectance in each scene. All illumination-estimation algorithms make assumptions about image content that may be violated some of the time. The question is whether or not MaxRGB's assumption is any more likely to be violated than those of other algorithms. The competitive performance of MaxRGB on HDR images indicates that MaxRGB's past failures may be due more to the lack of dynamic range in standard 8-bit image data than due to violations of its fundamental assumption that a white (or white-equivalent) surface is present in every scene.

References

[1] K. Barnard, L. Martin, A. Coath, and B. Funt, "A Comparison of Computational Color Constancy Algorithms, Part 2; Experiments with Images," IEEE Transactions on Image Processing, 11(9), pg. 985-996 (2002).

[2] G. Finlayson, S. Hordley, and P. Hubel, "Color by correlation: A simple, unifying framework for color constancy," IEEE Trans. Pattern Anal. Machine Intell., 23 (2001).

[3] J. van de Weijer, T. Gevers and A. Gijsenij, "Edge-Based Color Constancy," IEEE Transactions on Image Processing, 16, 9 (2007)

[4] G. Finlayson and E. Trezzi, "Shades of gray and colour constancy," Proc. IS&T/SID 12th Color Imaging Conf., pg. 37–41, (2004).

[5] A. Gijsenij, T. Gevers, J. van deWeijer, "Generalized Gamut Mapping using Image Derivative Structures for Color Constancy," Int J Comput Vision, 86, 2/3 (2010).

[6] E. Land and J. McCann, "Lightness and Retinex Theory," Journal of the Optical Society of America, 61, 1 (1971).

[7] B. Funt, F. Ciurea, and J. J. McCann, "Retinex in Matlab," Journal of Electronic Imaging, 13, 1 (2004).

[8] E.H. Land, "The retinex theory of color vision," Scientific American, 237, 6 (1977).

[9] B. Funt, K. Barnard and L. Martin, "Is Machine Colour Constancy Good Enough?" Proc. ECCV'98 5th European Conference on Computer Vision", pg. 445-459. (1998).

[10] B. Funt and L. Shi, "The Effect of Exposure on MaxRGB Color Constancy," in Proc. SPIE Volume 7527 Human Vision and Electronic Imaging XV, San Jose, (2010).

[11] Matlab implementation http://lear.inrialpes.fr/people/vandeweijer/software, Accessed April 2, 2010.

[12] G. Finlayson and S. Hordley, "Gamut constrained illumination estimation," Int. J. Comput. Vis., 67(1), pp. 93–109. (2006).

[13] Matlab Image Procession Toolbox 7.0, The Mathworks Inc.

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

[15] S. Hordley and G. Finlayson, "A Re-evaluation of Colour Constancy Algorithm Performance," J. Opt. Soc. Am. A, 23, 5 (2006).

[16] http://en.wikipedia.org/wiki/Dcraw, accessed July 24, 2009

[17] G. Ward, "Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures," Journal of Graphics Tools, 8, 2 (2003).

[19] G. Buchsbaum, "A spatial processor model for object colour perception," Journal of the Franklin Institute, 310 (1980).

Author Biographies

Brian Funt received the Ph.D. degree in computer science from the University of British Columbia, Vancouver, BC, Canada, in 1976. He is a Professor with the School of Computing Science, Simon Fraser University, Vancouver, where he has been since 1980. He has published widely in color, color constancy, retinex, color calibration, and illumination estimation.

Lilong Shi received his M.Sc. and PhD degrees in computer science from Simon Fraser University, Vancouver, Canada, in 2005 and 2010. He is currently a researcher in Simon Fraser University, Vancouver, Canada. His research interests are illumination estimation, color constancy, and image processing.