The Correlation of Reproduction and Recovery Angular Errors for Similar and Diverse Scenes

Roshanak Zakizadeh and Graham D. Finlayson ; School of Computing Sciences, University of East Anglia, Norwich, UK

Abstract

Illuminant estimation algorithms are usually evaluated by measuring the angular error between the RGB vectors of the estimated illuminant and the ground-truth illuminant (recovery angular error). However, the recovery angular error reports a wide range of errors for a given illuminant estimation algorithm and a given scene viewed under multiple lights (despite the fact that when the estimated lights are divided out the reproductions are similar). Following this observation, the reproduction angular error was proposed which instead measures the angle between the RGBs of a white surface being color corrected by dividing out the estimated illuminant RGB and R=1, G=1 and B=1 (i.e. white if the illuminant is correctly discounted). In this work we look at the correlation between the two metrics for each individual algorithm applied on a set of images. We observe that where the images are from the same scene under different illuminations the recovery and reproduction errors are often uncorrelated. Whereas when the same algorithm is applied on the images of diverse scenes the two errors are highly correlated.

Introduction

The appearances of colors in a scene captured by an imaging device are affected by the illumination or illuminations prevailing the scene. To discard the color bias due to the illuminant and, thereby, produce a pleasant image as well as an applicable one for many computer vision tasks, the illuminant color is estimated using an illuminant estimation algorithm. In a second step the estimated RGB values of light is divided out from the RGBs of each pixel. The performance of an illuminant estimation algorithm is often evaluated by measuring the angle between the two RGB vectors of the ground-truth ($\underline{E}_{act}$) and the estimated ($\underline{E}_{est}$) illuminant using the recovery angular error:

$$err_{recovery} = \cos^{-1}\left(\frac{\underline{E}_{act} \cdot \underline{E}_{est}}{|\underline{E}_{act}||\underline{E}_{est}|}\right) \tag{1}$$

Recently, the conventional metric recovery angular error was shown, counterintuitively, to introduce a wide range of errors for the same scene and the same algorithm [1]. That is when the scene is fixed and only the color of the light changes and when an illuminant estimation algorithm provides estimates which when the colour bias due to illumination is 'divided out' produces the same reproduction (this turns out to be true for the majority of illuminant estimation algorithms) a large range of recovery errors can be reported (from 0 to over 40 degrees!)

To address this problem, a new metric [1] - *reproduction angular error* - which measures the angle between the color of a white surface after corrected by the estimated illuminant (i.e. the reproduced white) and the ground-truth white RGB (the actual light divided by itself) was proposed:

$$err_{reproduction} = \cos^{-1}\left(\frac{(\underline{E}_{act}/\underline{E}_{act}) \cdot (\underline{E}_{act}/\underline{E}_{est})}{\sqrt{3}|\underline{E}_{act}/\underline{E}_{est}|}\right) \tag{2}$$

Note, in Equation 2 the actual light divided by itself is simply R=1, G=1 and B=1 (white looks right).

Reproduction angular error was shown to be more stable towards changes in the illuminant when the scene is the same and the same algorithm is used. Using the new metric, the ranking of algorithms reported in the literature broadly remained unchanged but a pairs of consecutively ranked algorithms could swap their relative position. And, the best tuning parameters for individual algorithms was shown to depend on the metric used.

In this paper we study the correlation between the reproduction and recovery errors for a given algorithm. We notice the low correlation between the errors for the images where the scene is the same and the only difference is in the illuminant. This is expected as the premise of reproduction error is that it is stable to changes of illuminant (depends only on the algorithm) compared to recovery error (which depends both on the illuminant and the algorithm). On the other hand, we observe when the scenes are diverse the results of reproduction and recovery metrics for the same algorithm are very much correlated.

This is an important result as it establishes that the development of illuminant estimation algorithms is in good order. The best algorithms now are better than the antecedent methods whether the legacy recovery or the new reproduction error is used. However, the reproduction error adds to the state of the art because, in general, the same scene will be viewed under several lights.

In Section 2 we discuss the assumptions underlying the reproduction error. A comparison between recovery and reproduction errors is being made in Section 3. In Section 4 we carry out very simple correlation analysis comparing reproduction and recovery errors. A short discussion is included in Section 5. We conclude the findings in Section 6.

Background: Formulation of Reproduction Angular Error

Image formation is often formulated as [2]:

$$\rho_k^{E,S} = \int_{\omega} R_k(\lambda)E(\lambda)S(\lambda)d\lambda \qquad k \in \{R,G,B\} \tag{3}$$

Here $E(\lambda)$ is the spectral power distribution of light striking the objects with the surface reflectance of $S(\lambda)$. The light reflected is proportional to $E(\lambda)$ and $S(\lambda)$ and it is sampled by a sensor with spectral sensitivity of $R(\lambda)$ over visible spectrum (ω).

It is obvious from Equation 3 that light and surface play symmetrical roles in image formation and Equation 3 can be simplified as [3]:

$$\rho_k^S = \int_\omega R_k(\lambda)S(\lambda)d\lambda \quad \rho_k^E = \int_\omega R_k(\lambda)E(\lambda)d\lambda \tag{4}$$

and

$$\rho_k^{E,S} = \rho_k^S * \rho_k^E \qquad k \in \{R,G,B\} \tag{5}$$

In color constancy we aim to recover $\underline{\rho}^S$ which is the color of the surface under uniform light $E(\lambda) = 1$.

According to this simplified image formation model, if the illuminant is estimated as $\underline{\rho}^{Est}$, then by 'dividing out' we can estimate the surface color:

$$\frac{\underline{\rho}^{E,S}}{\underline{\rho}^{Est}} \approx \underline{\rho}^S \tag{6}$$

Implication of the simple model is that response across illuminants are related by a diagonal matrix:

$$\underline{\rho}^{E',S} = diag(\underline{d}) * \underline{\rho}^{E,S} \quad \underline{d} = [\alpha\ \beta\ \gamma] \quad \alpha,\beta,\gamma > 0 \tag{7}$$

Also, if illuminant estimation is viewed as a kind of statistical moment of RGB values of an image with N pixels:

$$\underline{\rho}^{Est} = moment(\underline{\rho}^{E,S_1}, \underline{\rho}^{E,S_2}, ..., \underline{\rho}^{E,S_N}) \tag{8}$$

From Equation 7 we can write:

$$\underline{d} * \underline{\rho}^{Est} = moment(\underline{\rho}^{E',S_1}, \underline{\rho}^{E',S_2}, ..., \underline{\rho}^{E',S_N}) \tag{9}$$

Equation 9 teaches that if two lights ($\underline{\rho}^E$ and $\underline{\rho}^{E'}$) are related by three scaling factors d then the corresponding illuminant estimates ($\underline{\rho}^{Est}$ and $\underline{\rho}^{Est'}$) are similarly related.

According to Equation 9 reproduction angular error (Equation 2) is stable against changes in the illumination. Since it is clear that the three scaling factors d relating the two illuminants are cancelled in Equation 2.

Clearly, this is not the case for recovery angular error and this metric is highly affected by the changes in the illumination.

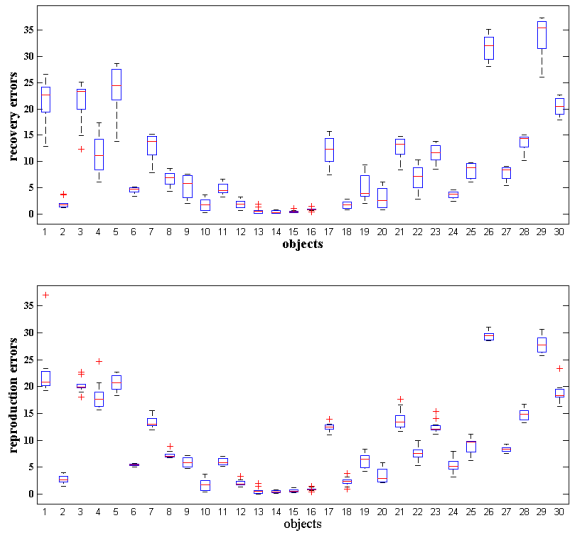

Figure 1: Box plots of recovery (top) and reproduction (bottom) angular errors for the 30 objects in SFU dataset.

Recovery versus Reproduction Errors

We have used the grey-world [4] estimations [5] of 11 illuminants for 30 objects in the SFU data set [6] to illustrate the degree of deviation of recovery errors from one illuminant to the other for a single object. In the SFU dataset the same object is captured under 11 different lights. The box plots in Figure 1 show the range of reproduction and recovery angular errors for the 30 objects in SFU dataset. We can see the range of errors according to recovery angular error (top box plot) is much wider than the range of reproduction angular errors (bottom box plot).

We also calculate the standard deviation of the recovery error per object and the per object standard deviation for the reproduction error. We plot (for all 30 objects) the standard deviation of recovery against reproduction standard deviations in Figure 2. Clearly, the reproduction error is much more stable than the recovery error.

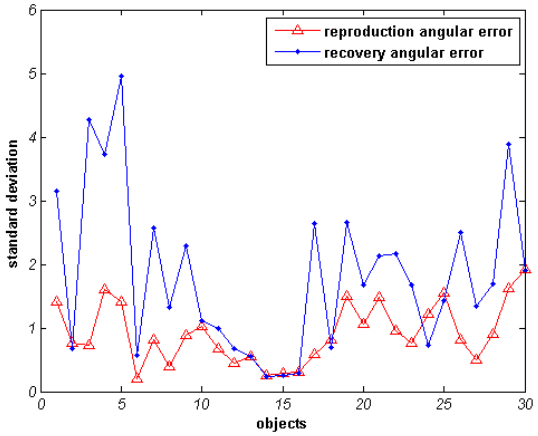

Figure 2: Standard deviation of recovery and reproduction angular errors for the 30 objects in the SFU dataset.

Correlation between Reproduction and Recovery Angular Errors

In our first experiment, we study the correlation of the two metrics for the SFU data set [6] (multiple objects each being viewed under multiple lights) for a range of algorithms. Our expectation here is that, recovery and reproduction errors while correlated, this correlation will be less for a data set where the same object is viewed under multiple lights.

The second correlation is tested for the Gehler colour checker data set [7] which comprises a wide variety of scenes viewed under a single light.

Similar scenes with different illuminants

The SFU set [6] of 321 images is one of the benchmark datasets used in color constancy. The dataset consists of 30 different sets of objects. Images are captured by the Sony DXC-930 CCD camera [6] with respect to a large range of different illuminants (10 to 11 types of illuminants). Figure 3 shows an example of the same object being captured under different illuminations.

Now, for each image the illuminant is estimated using six algorithms [4, 8, 9, 10] (see first column of Table 1). We assess the correlation of the algorithms using both the recovery and reproduction angular errors. In Table 1, we tabulate Pearson's r

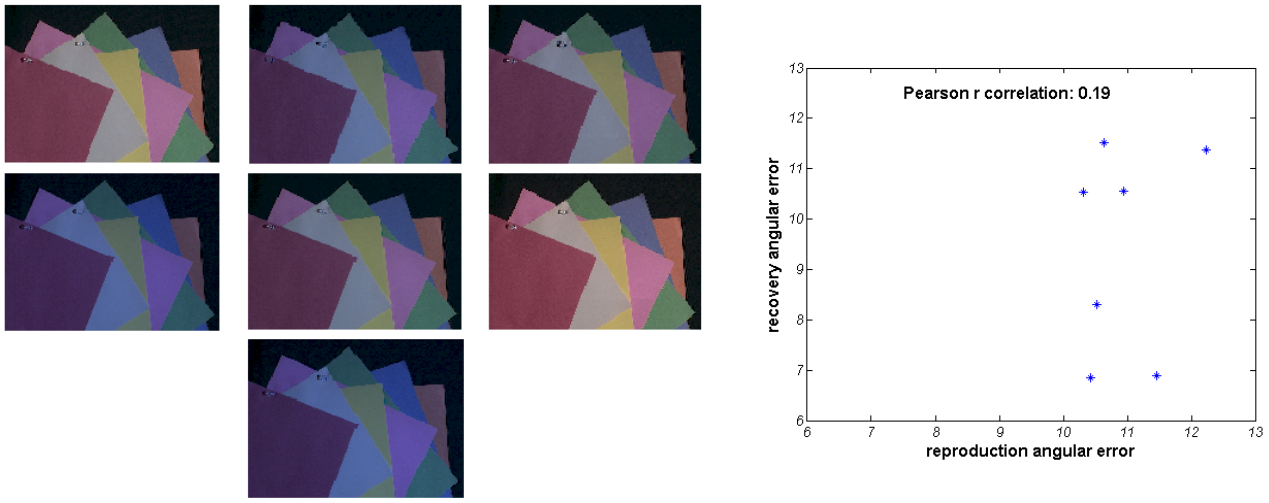

Figure 3: Correlation of reproduction and recovery angular errors for 1^{st} grey-edge (p-norm = 3, $\sigma = 3$) algorithm [8] applied on a set of images in the SFU dataset [6]. The number on the plot shows the Pearsons r correlation value between the two errors. The images are not color corrected.

coefficient of correlation [11]. A correlation of 1 means the errors would be proportional to one another, 0 no correlation and -1 maximum negative correlation. Interestingly, for the six algorithms tested there is a low correlation between the reproduction and recovery angular errors.

In Figure 3 the plot of correlation between the two errors for the 1^{st} grey-edge algorithm [8] can be seen. As you can see the error values are highly uncorrelated. As expected the reproduction error is stable but for given fairly constant reproduction error the recovery error varies widely.

Table 1: Results of Pearson's r correlation test for the reproduction and recovery errors for several algorithms on a set images from the SFU dataset [6].

Algorithm	Pearson's r
1^{st} grey-edge $(p = 3, \sigma = 3)$ [8]	0.19
2^{nd} grey-edge $(p = 4, \sigma = 2)$ [8]	0.04
grey-world [4]	0.55
Shades of grey $(p = 6)$ [9]	0.09
Edge gamut mapping $(\sigma = 7)$ [10]	0.29
Pixel gamut mapping $(\sigma = 8)$ [10]	0.21

Diverse scenes

The Gehler-Shi dataset [7] of 568 images comprises many individual scenes viewed under many lights (the same scene is not viewed under more than one light). In Figure 4 we show a few of different scenes. On the right side of Figure 4 the reproduction and recovery angular errors for the 1^{st} grey-edge algorithm for the Gehler-Shi dataset is shown.

In Table 2 the Pearson's r values is reported for a group of algorithms on all the images of Gehler-Shi dataset [7]. The correlation values are almost close to one for all the algorithms. This is a significant result as it shows that on average for typical viewing conditions the legacy recovery error can be used to rank algorithms. The flaw in its formulation while important and worth remedying does not invalidate the historical development, and ranking of algorithms using datasets such as Gehler-Shi and the recovery errors. That is the best algorithms today are better than those of five years ago and these in turn are better than the venerable grey-world [4] and MaxRGB [12] algorithms.

Table 2: Results of Pearson's r correlation test for the reproduction and recovery errors for several algorithms on all the images of Gehler-Shi dataset [7]

Algorithm	Pearson's r
1^{st} grey-edge $(p = 3, \sigma = 3)$ [8]	0.95
2^{nd} grey-edge $(p = 5, \sigma = 6)$ [8]	0.95
grey-world [4]	0.99
Shades of grey $(p = 5)$ [9]	0.98
MaxRGB [12]	0.97
Pixel gamut mapping $(\sigma = 5)$ [10, 13]	0.99
Edge gamut mapping $(\sigma = 3)$ [10]	0.96

However, it is also important to note that the correlation statistic is a 'broad brush'. While the correlation analysis gives us confidence that the results in the literature (reporting the relative performance of algorithms) are in good order, [1] showed that there are small changes in the overall rankings when the two error metrics are used.

Discussion

One might also be interested in knowing whether the algorithms could be ranked differently by reproduction and recovery angular errors. It is shown in [1] that while the rankings of algorithms reported in the literature remain broadly similar, the new metric (reproduction angular error) can change the ranks of local pairs of algorithms. For instance pixel-based and edge-based gamut mapping [10] are assigned different ranks by the two metrics (see Table 3). Further the Kendall's rank correlation test [11]

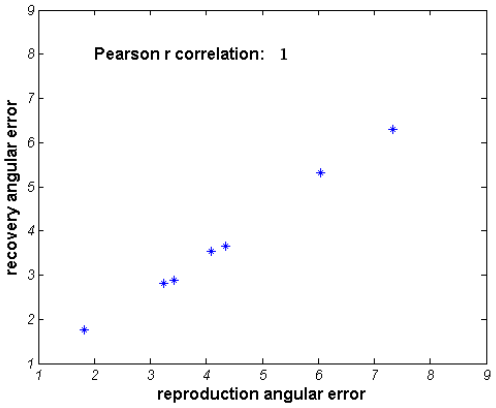

Figure 4: Correlation of reproduction and recovery angular errors for 1^{st} grey-edge (3, 3) algorithm applied on a set of images in Gehler-Shi dataset[7]. The number on the plot shows the Pearson's r correlation value between the two errors. The images are not color corrected.

the authors in [1] showed that in some cases this switching in ranking is statistically significant.

Table 3 shows the recovery and reproduction angular errors for a set of algorithms applied on the SFU dataset. We calculated the median errors over all 321 images of SFU dataset. You can see that in several cases the rankings of algorithms have switched (ranks in **bold**).

Table 3: Median recovery and reproduction errors of several colour constancy algorithms for the SFU dataset [6]. The changed ranks are in bold.

	Recovery		Reproduction	
Algorithm	Median	Rank	Median	Rank
grey-world [4]	7.0°	10	7.49°	10
MaxRGB [12]	6.5°	9	7.44°	9
Shades of grey [9]	3.7°	**8**	3.94°	**7**
1^{st} grey-edge [8]	3.2°	**6**	3.59°	**5**
2^{nd} grey-edge [8]	2.7°	4	3.04°	4
Pixel-based gamut [10, 13]	2.27°	**2**	2.83°	**3**
Edge-based gamut [10]	2.28°	**3**	2.7°	**2**
Intersection-based gamut [10]	2.09°	1	2.48°	1
Heavy tailed-based [14]	3.45°	**7**	4.11°	**8**
Weighted grey-edge [15]	3.09°	**5**	3.62°	**6**

Table 4 shows the same results for the 568 images of Gehler-Shi dataset [7]. Compared to the results for the SFU images (where the same scenes are captured under different illuminations), there are less switches in the rankings of algorithms for Gehler-Shi dataset.

For the algorithms whose ranks have changes in Table 3 the Spearsman's correlation coefficient is calculated as 0.85. The Spearman's coefficient being close to one shows a positive relationship between the rankings of algorithms by the two metrics. However, Spearsman's test is not suitable for quantifying the discrepancy between the number of concordant and discordant pairs. Kendall's test [11] is often used for such a comparison. The test is performed on the algorithms whose ranks have changed in Table 3 which is discussed by Finlayson and Zakizadeh in [1]. The Kendall's test result shows a significance switch in the rankings of algorithms for the SFU dataset.

We can infer from the reproduction and recovery errors in Tables 3 that for a set of images of similar scenes with diverse illuminations the legacy recovery angular error can result in different evaluation of algorithms.

Table 4: Median recovery and reproduction errors of several colour constancy algorithms for the Gehler-Shi dataset [7]. The changed ranks are in bold.

	Recovery		Reproduction	
Algorithm	Median	Rank	Median	Rank
grey-world [4]	6.3°	10	6.8°	10
MaxRGB [12]	5.7°	9	6.5°	9
Shades of grey [9]	3.9°	5	4.4°	5
1^{st} grey-edge [8]	4.3°	**6**	4.9°	**7**
2^{nd} grey-edge [8]	4.4°	**7**	4.8°	**6**
Pixel-based gamut [10, 13]	2.3°	1	2.7°	1
Edge-based gamut [10]	5.0°	8	5.8°	8
Intersection-based gamut [10]	2.3°	1	2.7°	1
Heavy tailed-based [14]	2.96°	**4**	3.47°	**3**
Cart-based selection [16]	2.91°	**3**	3.48°	**4**

Conclusion

In this paper we studied the correlation between the reproduction and recovery errors for a given algorithm on images of similar and diverse scenes. We noticed the low correlation between the errors in case of images of the same scene captured under different illuminations. Such result was expected as the premise of reproduction error is that it is stable to changes of illuminant compared to recovery error which is more dependent on the illuminant. On the other hand, we observed when the scenes are diverse the results of reproduction and recovery metrics for the same algorithm are very much correlated. This observation is important as it establishes that the development of illuminant estimation algorithms is in good order. However, since we expect to capture images of same scene as the illumination changes, we recommend the adoption of reproduction error.

Acknowledgements

This research was supported by EPSRC grant H022236.

References

[1] G. D. Finlayson and R. Zakizadeh, Reproduction angular error: an improved performance metric for illuminant estimation, Proceedings of British Machine Vision Conference (BMVC), 2014.

[2] B. A. Wandell, The synthesis and analysis of color images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 9(1), pp. 2-13, 1987.

[3] H. Y. Chong, S. J. Gortler, and T. Zickler, The von Kries hypothesis and a basis for color constancy, IEEE 11^{th} International Conference on Computer Vision (ICCV), pp. 1-8, 2007.

[4] G. Buchsbaum, A spatial processor model for object colour perception, Journal of the Franklin institute, 310(1), pp. 1-26, 1980.

[5] A. Gijsenij, T. Gevers, and J. V. D. Weijer, Computational color constancy: Survey and experiments, IEEE Transactions on Image Processing, 20(9), pp. 2475–2489, 2011.

[6] K. Barnard, L. Martin, B. Funt and A. Coath, A data set for color research, Color Research & Application, 27(3), pp. 147–151, 2002.

[7] L. Shi and B. Funt, Re-processed version of the Gehler color constancy dataset of 568 images, Simon Fraser University, 2010. URL http://www.cs.sfu.ca/ colour/data/.

[8] J. Van De Weijer, T. Gevers, and A. Gijsenij, Edge-based color constancy, IEEE Transactions on Image Processing, 16(9), pp. 2207-2214, 2007.

[9] G. D. Finlayson and E. Trezzi, Shades of gray and colour constancy, 12^{th} IS&T/SID Color and Imaging Conference, pp. 37-41, 2004.

[10] A. Gijsenij, T. Gevers, and J. V. De Weijer, Generalized gamut mapping using image derivative structures for color constancy, International Journal of Computer Vision, 86(2-3), pp. 127-139, 2010.

[11] P. Sprent and N.C. Smeeton, Applied Non-parametric Statistical Methods, Fourth Edition, Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis, 2007, ISBN 9781584887010.

[12] E. H. Land, The retinex theory of color vision, Scientific American, 237(6), pp. 108-128, 1977.

[13] D. A. Forsyth, A novel algorithm for color constancy, International Journal of Computer Vision, 5(1), pp. 5-36, 1990.

[14] A. Chakrabarti, K. Hirakawa, and T. Zickler, Color constancy with spatio-spectral statistics, IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(8), pp. 1509–1519, 2012.

[15] A. Gijsenij, T. Gevers, and J. V. De Weijer, Improving color constancy by photometric edge weighting, IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(5), pp. 918–929, 2012.

[16] S. Bianco, G. Ciocca, and C. Cusano, Color constancy algorithm selection using CART, Computational Color Imaging, Springer Berlin Heidelberg, pp. 31–40, 2009.

Author Biography

Roshanak Zakizadeh joined the School of Computing Sciences at the University of East Anglia in 2013 where she is pursuing her PhD under the supervision of Professor Graham Finlayson. Prior to that she obtained her Masters Degree in Colour in Informatics and MEdia Technology (CIMET) in 2013 from Université Jean Monnet, France and University of Eastern Finland, Finland.