The Generalised Reproduction Error for Illuminant Estimation

Graham FINLAYSON, Roshanak ZAKIZADEH
School of Computing Sciences, University of East Anglia, UK

ABSTRACT

In a recent publication "Reproduction Angular Error: An Improved Performance Metric for Illuminant Estimation", British Machine Vision Conference (2014), it was argued that the commonly used Recovery angular error – the angle between the RGBs of the actual and estimated lights- is flawed when it is viewed in concert with how the illuminant estimate is used. Almost always, we use the illuminant estimate to make an image reproduction where the colour bias due to illumination is removed or reduced. It was shown that, when a single algorithm was used to estimate the light for a fixed scene viewed under a range of illuminants and where similar reproductions were produced when the estimate was 'divided out', the recovery angular error would, counterintuitively, vary widely. The Reproduction angular error introduced in that paper remedies this flaw by measuring the angle between a true white patch and the white that is reproduced when an illuminant estimate is made. In this paper we generalize the reproduction error concept to consider how well a range of colours are reproduced. We show how an illuminant estimate can be used to map the colours in a Macbeth colour checker for the actual illumination to reference lighting conditions. Then we evaluate the error of reproduction using the mean CIE Delta E. This new Generalised Reproduction error metric is used to compare the performance of a variety of different algorithms. Significantly, the rank-order of the reproduction angular error is quite similar to that established with the generalized reproduction error. Based on our experiments we propose that the simpler reproduction angular error can be used as a proxy to our generalised metric to assess the performance of illuminant estimation algorithms.

1. INTORODUCTION

The colours in an image captured by a digital camera are affected by the illuminant under which the scene is captured. Unlike human visual system which is able to perceive the colours constant regardless of illumination, the sensors of a camera capture a colour signal which is confounded by the illumination. To make the colours pleasant and usable by many computer vision tasks, the illuminant of the scene is estimated by reasoning about the distribution of colours in the image. In a second step, the colour of the illuminant is divided out from the colours of the image thereby removing the colour bias due to the illumination.

Illuminant estimation algorithms range from simple statistics-based methods to algorithms that use more complex statistics to learning-based methods. The *recovery error* which is the angle between the estimated and the ground-truth illuminant is commonly used to evaluate the performance of an illuminant estimation algorithm. The average (mean or median) recovery error, for a set of training set of images, is used as an index to compare and rank different algorithms.

However, in recent work (Finlayson & Zakizadeh 2014), a problem with the recovery error was identified. It was shown that the same scene viewed under two different colours of light where the same algorithm is used to estimate the illuminant can result in two very different angular errors. This is a problem because when each of the estimated lights are divided out from their respective images almost the same image reproduction results. Finlayson and Zakizadeh argued that the performance of illuminant estimation algorithms should be tied to how illuminant estimates are used. They are used to discount the colour bias die to illumination in making an image reproduction. Their new metric, called Reproduction Angular Error, measures the angle between the colour (RGB vector) of a white surface corrected using the estimated illuminant and the one corrected using the ground-truth illuminant (resulting in a true white patch). Significantly, this reproduction error provides a stable error for the same scene viewed under different lights and this gels with the fact that the corresponding reproductions look similar. Moreover, the new metric while broadly ranking algorithms the same as recovery angular error introduced several local changes in algorithm rank.

In this paper, we seek to measure the difference between a range of colours (not just a white patch) which are reproduced by the estimated and ground-truth illuminants. This is not as easy as it first sounds as when image data sets are compiled we often have the image and the measured white point but not the appearance of the scene under a ground truth illuminant. In out approach we first show how to make a synthetic set of Macbeth colour checkers for different illuminants for a known camera. Second we show how to make reproductions of the Macbeth colour images when the illuminant colour is discounted using the illuminants estimated by different algorithms (for the algorithm estimates we use the data provided by Gijsenij et al. 2011). Then these reproductions are compared with the actual colours in a Macbeth checker for the reference lighting condition. The CIE Lab ΔE (Sharma et al. 2005) is used to measure the colour difference between the colours of the checker reproduced by the estimated lights and those under reference lighting condition. If the correct illuminant is estimated then a very small ΔE would result. Conversely, poor estimates result in large average ΔEs.

According to this new *generalised reproduction error* we can rank the performance of different algorithms. Crucially, we show that the ranking provided is almost the same as the recently introduced – and much simpler to calculate – reproduction angular error. This paper further validates the usefulness of the reproduction angular error metric.

2. BACKGROUND

The most commonly used metric for evaluating illuminant estimation algorithm is the recovery angular error:

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

which is the angle between the estimated RGB of illuminant $\underline{E}_{act}$ and the ground-truth RGB illuminant $\underline{E}_{est}$. Recently, this recovery angular error was shown to have the problem of introducing a wide range of errors when a given algorithm estimates the illuminant for a given scene (Finlayson & Zakizadeh 2014) viewed under a wide range of illuminants. This behaviour is problematic because when the different illuminant estimates are 'divided out' similar reproductions result. It is these reproduced images that are respectively assessed in photography or used in computer vision. To solve this problem, the Reproduction Angular

Error was proposed The Reproduction angular error is defined to be the angle between the RGB of a white surface when the actual and the estimated illuminations are 'divided out'.

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

3. GENERALISED REPRODUCTION ERROR

The Reproduction angular error assesses the performance of illuminant estimation algorithms according to how well white is reproduced when the colour bias due to illumination is removed. Here we wish to generalise this idea to consider how a range of colours are reproduced. Our idea is to provide a method for synthesising the RGB image of a Macbeth colour checker under an actual light and then use the RGB estimate of the illuminant – made by an algorithm – to correct the image colours (to remove the colour bias due to illumination). This corrected Macbeth checker is then compared with the actual reproduction (when the true illuminant is used).

In constructing our model, we use the set of spectra for 24 Macbeth colour checker patches and the 23 lights from the SFU dataset (Barnard et al. 2002). For camera sensitivity functions we use the Sony DXC-930 CCD (Barnard et al. 2002) but the sensitivities of any particular camera can be used in the problem formulation. Equation (3) teaches that the camera response (ρ^k) whose spectral sensitivities are denoted $R^k(\lambda)$ for the surface spectra ($S(\lambda)$) and the illuminant spectra ($E(\lambda)$) is calculated as:

$$\rho^k = \int_{380}^{700} R^k(\lambda)S(\lambda)E(\lambda)d\lambda \qquad (3)$$

For all numerical calculations, we assume the visible spectrum runs from 380 to 780 Nanometres and we use a 4 Nanometres sampling interval. For each of the 23 lights we, using (3), generate 24 RGBs. These 23 synthetic checker images encapsulate our understanding of how the checker appears under different lights. We wish to generalise this understanding so that we could, given the RGB of any target light, synthesise the appearance of the checker for any illuminant. Denoting the 24x3 RGBs for a Macbeth colour checker as *M*, we model *M* as a linear sum of three basis Macbeth colour checkers:

$$M \approx \sum_{i=1}^{3} M_i m_i \qquad (4)$$

In (4), m_i denotes a scalar weight and the optimal basis in a least-squares sense are found using Characteristic Vector Analysis (Maloney 1986) (in this case of the 23 synthetic Macbeth checker images). Crucially, we found the best basis models our data extremely well with the actual and 3-basis approximation being visually almost the same in appearance.

We chose a 3-dimensional linear model because illumination is defined by 3 numbers: the RGB of the light or the RGB of the estimated light. Let us place the RGB for the white reflectance in the Macbeth checker for each basis term M_i in the 3 columns of a calibration matrix Ω. Denoting an RGB of a light as $\underline{E}$, the linear combination of the columns of Ω defines the weights $\underline{m}$ used in Equation (4):

$$\underline{m} = \Omega^{-1}\underline{E} \qquad (5)$$

In (5), the illuminant vector $\underline{E}$ could be the actual light or the estimate made by an algorithm. Figure 1a shows one input image and four synthetic checkers. The image is from the SFU (Barnard et al. 2002) database. For this scene a white patch was also measured. Algorithms such as pixel-based gamut mapping will attempt to infer an estimate which ideally will be close to the measured light.

With the measured and actual RGBs of the light in hand, we generated from our linear model (4) and using (5) to find our model coefficient the appearance of the actual checker (Fig. 1b) and the one that pixel based gamut mapping infers (Fig. 1c). The third checker is the correct answer (Fig. 1d). The white patch is equal to [1,1,1]. All 3 images are scaled so that the brightest pixel value across all the colour channels is 1 and a gamma of .5 is applied.

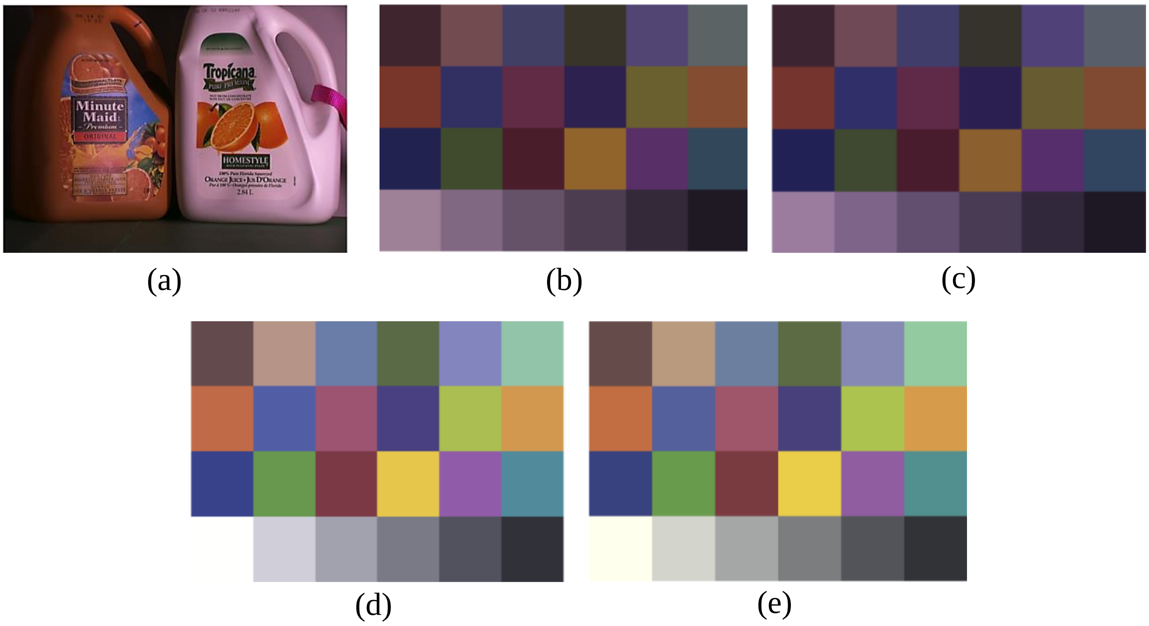

Figure 1: (a) Image from SFU dataset, (b) Synthetic colour checker under the ground truth light under which Fig. 1a was taken, (c) Synthetic colour checker under the estimation of the same ground truth light made by pixel-based gamut mapping algorithm, (d) Synthetic colour checker under the reference light and (e) Corrected colour checker by pixel-based gamut mapping algorithm.

So far, we have focussed on explaining how we synthesise the colours of the Macbeth colour checker for a target light. But, we ultimately seek to model the appearance of a checker under an actual light when it is corrected to the reference checker (fig. 1d) using the wrong illuminant estimate.

Denoting, respectively, the checker under the reference (white light), the actual coloured light and the estimated coloured light as M^{ref}, M^{act} and M^{est}, the estimated reproduction, $\widetilde{M}^{ref}$, is calculated as:

$$\widetilde{M}^{ref} = M^{act}T \quad , \quad T = [M^{est}]^{+}M^{ref} \qquad (6)$$

In (6), $[M^{est}]^{+}$ denotes the Moore-Penrose inverse. That is, T is the least-squares fit from the checker viewed under the estimated light to the reference lighting conditions. This 3x3 matrix T is then applied to the checker under the actual light.

The *Generalised Reproduction Error* for the ith Macbeth colour checker patch is:

$$err_i = \left\| f(\widetilde{M}_i^{ref}) - f(M_i^{ref}) \right\| \quad (7)$$

where f maps an RGB to CIE LAB. Note the function f must map the camera values to corresponding XYZs and then the standard CIE Lab formulae can be used.

In Fig. 1e we show an actual checker under a coloured light corrected using the estimated light of pixel gamut mapping and the procedure described in (6). Note the reproduction is reasonable but there remains a slight yellowish cast.

4. RESULTS

Here we use the 321 images from the SFU dataset (Barnard et al. 2002). This data set has linear images and a variety of objects are imaged under 11 lights (ranging from quite yellowish to very blue). All images were captured with the SONY DXC-930. A variety of algorithms, including those listed in Table 1, were tested by Gijsenij et al. 2011 who makes all the estimated RGBs available to the community. We can thus calculate for all Macbeth colour checker images and the overall median generalised reproduction error. Then according to this global median we can rank the algorithms.

In Table 1 we list the algorithms and record the rank for the Recovery and Reproduction angular errors and the new Generalised Reproduction error.

Table 1. Comparison of ranking of algorithms based on reproduction angular errors and generalised reproduction errors.

	Recovery angular error		Reproduction angular error		Generalised Reproduction Error	
Method	Median error	Rank	Median error	Rank	Median error	Rank
Grey-world	7.0°	9	7.49°	9	7.02	9
MaxRGB	6.5°	8	7.44°	8	6.13	8
Shades-of-gray	3.7°	**7**	3.94°	6	3.26	6
1st grey-edge	3.2°	5	3.59°	5	3.12	5
2nd grey-edge	2.7°	4	3.04°	4	2.88	4
Pixel-based gamut	2.267°	**2**	2.83°	3	2.64	3
Edge-based gamut	2.278°	**3**	2.70°	2	2.59	2
Intersection-based gamut	2.09°	1	2.48°	1	2.46	1
Heavy tailed-based	3.45°	**6**	4.11°	7	3.74	7

While the rankings of all three metrics are almost similar it is clear recovery angular error ranks algorithms a little differently from reproduction angular error. Further in (Finlayson & Zakizadeh 2014) it was shown that the rankings are statistically different. And, this fact draws attention to the care the algorithm designer needs to take using the appropriate metric to assess their algorithm. The reproduction angular error assesses how well an algorithm reproduces white (i.e. when the estimated illuminant is divided out). Generalised reproduction error builds on this concept and accounts for the error for other surface colours. The ranks for the generalised reproduction error are almost identical to the reproduction angular error. Indeed – space prohibits us elucidating on this point here – the rankings are not statistically significantly different. We can conclude, for the data tested, that the simple reproduction error can be used as a proxy for the generalised reproduction error developed here.

5. CONCLUSION

Reproduction angular error measures the angle between a true white patch and the white patch that results when an algorithm's estimate is 'divided out' from the image. In this paper we generalised reproduction angular error to assess not only how white is reproduced but, instead, all the colours on a Macbeth colour checker. The generalised reproduction error is the CIE Lab colour difference of a reference checker and a reproduction that results when the same checker viewed under an actual coloured light is colour corrected using an estimate of that light supplied by an algorithm. Like the simple reproduction angular error the same algorithm/scene pair returns very similar error independent of the colour of the light to be estimated (because in all cases the resulting reproductions are similar). We observed that the ranking of a selection of algorithms based on the generalised reproduction error ΔEs are very similar to the ranks given by the simple reproduction angular metric. Thus, while the generalised reproduction error provides a finer grained summary of the 'goodness' of an illuminant estimation algorithm, the simpler reproduction angular error can be used to assess algorithm performance.

ACKNOWLEDGEMENTS

We thank EPSRC for supporting this research (grant: H022236).

REFERENCES

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

Finlayson, D. G. and R. Zakizadeh. 2014. Reproduction angular error: an improved performance metric for illuminant estimation. In *Proc. British Machine Vision Confernece 2014, Nottingham, UK, September 1-5, 2014.*

Gijsenij, A., T. Gevers, and J. Van De Weijer. 2011. Computational color constancy: Survey and experiments. *IEEE Transactions on Image Processing, 20(9):2475-2489.*

Maloney, L. T. 1986. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *J. Opt. Soc. America A, 3(10), 1673-1683.*

Sharma, G., W. Wu, and E. N. Dalal. 2005. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application, 30(1):21-30.*

Address: School of Computing Sciences, University of East Anglia,
Norwich Research Park, Norwich, NR4 7TJ, UK
E-mails: r.zakizadeh@uea.ac.uk, g.finlayson@uea.ac.uk