

A Psychophysical Analysis of Illuminant Estimation Algorithms

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

Abstract

Illuminant estimation algorithms are often evaluated by calculating recovery angular error which is the angle between the RGB of the ground truth and the estimated illuminants. However, the same scene viewed under two different lights with respect to which the same algorithm delivers illuminant estimates and then identical reproductions - and so, the practical estimation error is the same - can, in fact and counterintuitively, result in quite different recovery errors. Reproduction angular error has been recently introduced as an improvement to recovery angular error. The new metric calculates the angle between the RGB values of a white surface corrected by the ground truth illuminant and corrected by the estimated illuminant. Experiments show that illuminant estimation algorithms could be ranked differently depending on whether they are evaluated by recovery or reproduction angular error. In this paper a psychophysical experiment is designed which demonstrates that observers choices on ‘what makes a good reproduction’ correlates with reproduction error and not recovery error.

Introduction

Colors in an image captured by the camera often do not look as they appear naturally to the human observer. This is due to the effect of light on the colors captured by the camera’s sensors. To discard this effect, the illuminant in the scene is often estimated by an illuminant estimation algorithm and the estimated RGB of the light is divided out from the RGB values of the image. Many illuminant estimation algorithms have been proposed over the years [1]. Evaluating and comparing the performance of illuminant estimation is of great importance. Illuminant estimation algorithms are often evaluated by measuring the errors that they introduce over a set of images captured by a single camera (a benchmark dataset). Usually, an average (median, mean or even 95% quantile and max [2].) of the error over all the images of the dataset is reported as the final error and the algorithms are ranked accordingly [3, 4].

Many error measurement methods have been recommended such as the angle between the RGB values of the estimated and the ground truth illuminants or the mean square error between the two values. As well as the improved angular error - reproduction angular error - which measures the angle between the RGB value of a white surface corrected by the estimated and the ground truth illuminants which is introduced recently [5, 6].

In this paper, we wish to study with which angular error- recovery or reproduction- the human observers agree more. To this end a psychophysical experiment is conducted where two scenes color corrected by two different illuminant estimation algorithms are shown to the observers along with their corrected version by the ground truth illuminants. The recovery and reproduction errors for the two algorithms estimating the illuminant of the two scenes is different resulting in opposite ranking of the two algo-

gorithms by the two metrics. Put it another way, suppose A and B are respectively the reproduction of a scene for an (arbitrary) illuminant estimation algorithm and B is the same scene reproduced with the actual illuminant. A second pair of images are C (scene reproduced by estimate made by second arbitrary illuminant estimation algorithm) and D (the ‘correct reproduction’). In terms of recovery error the estimate of the light that leads to the reproduction A has a larger error than that which results in C. Conversely, the reproduction error for C is larger than A. The observers are asked that the result of which algorithm is better which will ultimately show with which metric they agree more. The results of the experiment show that in most cases observers agree with the choice of reproduction angular error of the better illuminant estimation algorithm.

The psychophysics experiment conducted in this paper is also new in essence, unlike most studies carried out in the past, the images are given a photographic look being passed through a simplified in-camera color processing pipeline which is simulated by a learned look-up table (or lattice).

In Section 2, we provide a background on the performance evaluation of illuminant estimation algorithms. Section 3 explains the experimental setup and procedure. The results are discussed in Section 4. The paper is concluded in Section 5.

Background: Evaluation of Illuminant Estimation Algorithms

The most well know evaluation method in illuminant estimation is recovery angular error [7] which is defined as:

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

where, \underline{E}_{act} is the RGB of the ground truth and \underline{E}_{est} is the RGB of the estimated illuminant.

Later it was shown in [5] and [6] that recovery angular error introduces a wide range of error for the same algorithm applied on the same scene when only illuminant is changing even though the illuminant estimates that are made result in an identical image reproduction. As a remedy to this flaw, a new metric - reproduction angular error [5, 6] - was introduced as an improvement to recovery angular error:

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

The new metric measures the angle between the color of a white surface corrected by the estimated and the ground truth illuminants.

Reproduction angular error is shown to be more stable towards the changes of light. Although, there is a correlation between the two metric [8], but in some cases the ranking of algorithms switches when the algorithms are evaluated using reproduction angular error instead of recovery angular error. Whether

the switches in the ranking of illuminant estimation algorithms are predicted by human observers and to what extent they agree with the ranking by reproduction angular error needs to be tested in a psychophysical experiment.

Psychophysical evaluation of illuminant estimation algorithms have been the subject of a number of studies before [9, 10]. For instance, in [9], the correlation between different numerical evaluation methods and human observer evaluation of illuminant estimation algorithm was studied and it was shown that among all the numerical methods of performance evaluation of illuminant estimation algorithms, recovery angular error is more correlated with the observers' opinion in comparison with other methods.

In this paper, we wish to study with which angular error-recovery or reproduction- the human observers agree more. In the next section we detail an experimental framework we use to answer this question. Then the results of our experiments are presented and the experimental approach we have adopted is discussed in detail

Experiment Setup

Data Preparation

The images for our experiment are from the 200 images of the Canon EOS 600D camera from the recently created NUS dataset [11].

In our experiment, illuminant estimates are 'divided out' from the raw images. Then, in a second step we apply a camera processing pipeline (in effect modelling color correction, gamut mapping and tone correction [12]). The color mapping process, i.e. mapping the raw sensor values to their corresponding RGB outputs, has been the subject of a number of studies (e.g. [13, 14, 15, 16]). However, most of those research are concerned with radiometric calibration, which is the process of recovering scene radiance from image intensities. For this experiment, we are interested in creating a camera output look for the white balanced images generated when we divide out the camera's estimate. Li *et al.* [17] suggest a calibrated (trained) sparse 3D lookup table (LUT), also known as lattice [18, 19], suffices to map the raw sensor values from a particular camera to their corresponding RGB outputs. Crucially, to a good approximation, the same lattice can be used independent of the white point [17]. Using the calibrated lattice for a specific camera we can render white balanced images as if they have been passed through an in-camera color processing pipeline like the simplified model in Figure 1.

Here for calibrating the lattice we have used a random selection of 50 raw images white balanced by the ground-truth illuminants and their corresponding output JPG images captured by Canon EOS 600D camera from the NUS dataset. We randomly select 50000 pixels from each image to generate the lattice. The above sampling results in 2500000 points for training the lattice. In our experiment, the dimension of the lattice is three (for R, G and B color channels), the boundaries of the grid is set to be between zero and one and the size of the grid is 35 nodes in each dimension. So, we solve for a $35 \times 35 \times 35 = 42875$ lattice. Figure 2 gives a visual illustration of calibrating the lattice for the Canon EOS 600D camera based on a set of NUS dataset images captured by the same camera. Where for simplicity a $9 \times 9 \times 9$ lattice is shown. The optimisation to derive the lattice is presented in detail in [19].

Figure 3 shows a few examples of images from Canon600D

camera which are white balanced by the shades of grey algorithm [20] and their equivalent JPG transformation by the calibrated lattice. The first row of this figure are the raw white balanced images, the images are raised to the $1/2.2$ gamma value to make them visible. The second row are the results of just applying the trained lookup table to the images in the first row with no gamma correction. The last row are the actual JPG outputs from Canon EOS 600D. Note that the images in the second row look better than those in the first. Often illuminant estimation experiments (in computer vision or psychophysics) use raw images with some gamma correction but this figure illustrates that applying the lookup table instead of gamma correction results in better images.

Monitor

The images are presented on a high resolution professional LCD Backlit monitor (an HP DreamColor LP2480zx) with 1920×1200 pixels resolution. The monitor uses both a true 30-bit panel and an RGB LED backlight, providing over one billion possible colors and a wide color gamut. According to ISO 3664 standards, the calibration of the monitors for the psychophysics experiment is necessary. The monitor was calibrated using Spyder4Elite prior to running the experiment. The calibration was carried out in the same environment in which the experiment would later take place. The experimental environment is a room illuminated with a dim light source (to avoid eye strain) provided behind the monitor to avoid glare.

Observers

All observers participated in the experiment have normal color vision and normal to corrected-to-normal visual accuracy (all observers were asked to declare any visual deficiency including color blindness). At the beginning of the experiment the observers are allowed to adapt their vision for 30 seconds by staring at a variegated grey screen. This adaptation period is necessary to allow the observer's vision to adjust to the viewing conditions. There were eight observers participating in this experiment with their age ranging from almost 25 to 65. The group of observers contained both male and female participants.

Experimental Procedures

An observer is shown two pairs of images on a variegated grey background (like the one shown in Figure 4). The first pair contains a ground-truth (based on the physical white point) reproduction and that produced by algorithm *a* estimate. We, respectively, denote the two images in the first pair I_t^1 and I_a^1 . A second image pair is calculated in the same way. A ground-truth image I_t^2 is produced and that for a second illuminant estimation algorithm I_b^2 . Note the scene in the first image pair is different from the second and two different illuminant estimation algorithms are used.

Figure 4 shows an example from our experiment. Top left and right (I_a^1 and I_b^2) respectively are the reproduction delivered using the edge-based and pixel-based gamut mapping algorithms [22, 21]. The other images (I_t^1 and I_t^2) are the ground-truth reproduction. The images are selected carefully. Image I_a^1 is reproduced by the illuminant estimation algorithm *a* has a lower reproduction error than image I_b^2 reproduced by algorithm *b* (in this example the reproduction errors are, respectively: $3.76^\circ < 8.73^\circ$). Conversely, the recovery angular of algorithm

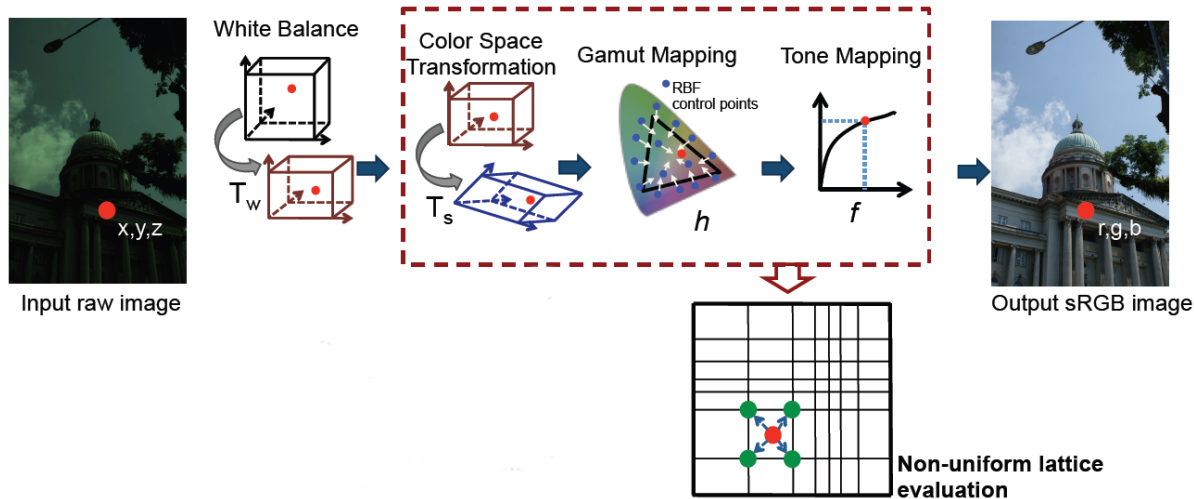


Figure 1: The simplified in-camera color processing pipeline (from [17] with a little modification). A single sparse LUT has replaced several steps in the imaging pipeline.

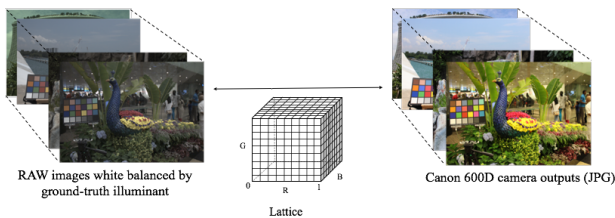


Figure 2: Lattice calibration for the Canon EOS 600D camera from the NUS dataset.

a for image I_a^1 is higher than the recovery error of algorithm b for image I_b^2 (in this case $4.15^\circ > 3.46^\circ$). That the pairs of images shows are ranked in opposition from one another in terms of recovery and reproduction error we call the ‘swapping property’. Seven pairs of images similar to the one in Figure 4 are chosen for each two algorithms.

In the experiment, the observer is then asked which image pair appears more similar. That is does I_a^1 look closer to I_t^1 compared with I_b^2 and I_t^2 (or vice versa). Note the observers do not know which image is corrected using the ground-truth illuminant and which by the estimate. We are interested in whether an observer judges I_a^1 to be closer to I_t^1 or I_b^2 to I_t^2 . If the former, the reproduction error correctly predicts image reproduction. If the latter, it is recovery error that predicts observer’s responses. The experiment is repeated for eight observers. Each image representation is repeated twice with the ‘a’ and ‘b’ pairs shown respectively left and right and the converse.

Importantly, the observers understand the concept of illuminant estimation and image reproduction. As such, given the scene content the observer - even though they are looking at a reproduction - can plausibly infer the color of the light a scene could have been observed. Further, on average, illuminant estimation algorithms ‘undershoot’. That is to say a very yellow light is estimated to be slightly less yellow than average and so the reproduction still - again on average - can display a colour bias in the direction of the illumination. This said, the observer can - albeit at a higher

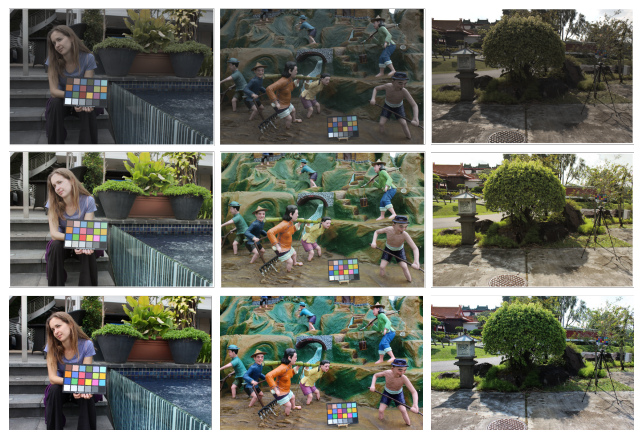


Figure 3: Examples of raw to JPG transformation using the calibrated lattice: The first row are the raw white balanced images by the shades of grey algorithm. The second row are the results of applied lattice. The last row are the actual JPG outputs from the camera (with the camera’s properly white balanced algorithm).

cognitive level - plausibly judge the reproductions in terms of the starting illuminant colour. And, so, could plausibly judge the pair where the recovery error is less as preferred ‘closer’.

Of course in proposing the above mental gymnastics we are not really expecting the observers to make this inference. Quite the opposite. Rather, we are pointing out the inference they would have to make if they were to choose image reproductions as appearing closer according to recovery error. Yet, if we as a community continue to use recovery error (where we use illuminant estimates to drive image reproduction) then we are, in effect, betting that people do follow this convoluted line of reasoning.

We compare four illuminant estimation algorithms in this experiment: 1st grey edge [23], 2nd grey-edge[23], shades of grey and pixel-based gamut mapping; here, denoted as GE1, GE2, SOG and GP respectively. Each algorithm is compared with the rest and in each pair of comparisons seven pairs of images are



Figure 4: Screen setup for the experiment.

used. For instance, to compare 1st grey edge and 2nd grey-edge algorithm seven pairs of images are shown to the observers where in each pair one image is corrected by the 1st grey edge and the second image is white balanced by 2nd grey edge. Each pair of pairs has the ‘swapping property’ previously described. That is pair ‘a’ compared to ‘b’ can have lower recovery error and in reverse pair ‘b’ compared to ‘a’ has lower reproduction error.

It should be noted the images for which illuminant estimation algorithms fail significantly (with recovery or reproduction errors higher than nine degrees) [2] are excluded in this experiment. Since, when the results of illuminant estimation algorithms are significantly poor it does not matter which algorithm has performed better and if fact it is almost impossible to tell.

Results and Discussion

The Chi-square test [24] is a statistical test commonly used to compare observed data with the expected data. Here, the expected data is the number of pairs in which the image corrected by algorithm *a* is better than the image corrected by algorithm *b* according to the chosen metric. For instance if according to reproduction angular errors of the seven pairs of images, algorithm *a* is predicted to be better than algorithm *b* then the expected value would be seven. The observed value is the number of pairs where algorithm *a* is preferred over algorithm *b* by the observer. Chi-square is the suitable measure of the “goodness to fit” between the observed and expected values.

The chi-square test is used here to attempt to reject the null hypothesis that the observed and the expected data won’t fit or in other words are independent.

With the expected (*e*) and the observed (*o*) values known, the Chi-square is calculated as the sum of the squared difference between:

$$\chi^2 = \frac{(o-e)^2}{e}. \quad (3)$$

It can be seen from the above calculation that it is intuitive to conclude that a large difference between the observed and expected values will result in accepting the null hypothesis which the independency of the two. If the observers agree with the results by reproduction or recovery angular error, then the difference should be small and the null hypothesis will be rejected. Clearly if χ^2 is zero then the expected and observed values are exactly the same and we can immediately reject the null hypothesis. In general, for small χ^2 we can reject the null hypothesis for some

criterion amount we will not be able to reject the null hypothesis. This criterion amount is found by consulting the statistical tables.

Formally, to be able to accept or reject the null hypothesis, the calculated Chi-square value in Eq. 3 should be compared against the critical chi-square value in the corresponding table (e.g. [24]). The critical value is decided from the table of chi-square for a desired significance level (e.g. 5% or 0.05). If calculated chi-square value is greater than the critical chi-square value the null hypothesis is accepted and the observed and expected data will not fit. Otherwise, the null hypothesis is rejected and observers agree with the expected data. We have eight observers in our experiment, so the number of samples (observations) is eight. The critical chi-square value for seven degree of freedom with $p = 0.05$ is 14.07

In Table 1, the Chi-square values for the goodness of fitness between the observers’ data and the expected values by reproduction angular error can be seen. Here, each cell of table contains two values (*x*, *y*), where *x* represents the number of pairs of images for which algorithm *a* performs better than algorithm *b* according to reproduction angular error. The value *y* is the observers’ data, which shows the number of comparisons in which the observer has preferred algorithm *a* over algorithm *b*. For instance in the column indicated by GE1-GE2, the 1st grey edge algorithm is compared against the 2nd grey edge algorithm. Based on observer 1 for all seven pairs of images GE1 is better than GE2, or observer 3 has agreed with the GE1 superiority over GE2 only for four out of seven pairs. The expected number of pairs where algorithm *a* performs better than *b* according to the reproduction angular error is for all the seven pairs. However, we found that some observers were not consistent with their choices when they were shown the same pair for the second time. If that was the case, we excluded that pair from the calculation of Chi-square for that specific observer. An example of such occurrence can be seen for observer 2, who has been consistent with his choices only for three pairs when comparing GE1 and GE2 algorithms. Since we are comparing four algorithms: GE1, GE2, SOG and GP, there are six columns of data which is the number of possible combinations of two out of four algorithms.

A comparison between the critical chi-square value for eight observers (which is 14.07 with the significance of $p = 0.05$) and the ones calculated in Table 1 shows there is no reason to reject the null hypothesis that the observed and expected values match. In other words, the observers agree with the prediction of the quality of the reproduced images by reproduction angular error.

Table 2 reports the same result but for comparison of the observers’ data with the results by recovery angular error. Notice that the name of the algorithms in this table is switched, i.e. GE1-GE2 in Table 1 has changed to GE2-GE1 in Table 2. The high values of Chi-square in Table 2 for all six pairs of algorithms reject the null hypothesis that the observers’ data match recovery angular error’s prediction.

To analyse whether there is an agreement between the observers [25, 26] the individual difference from the mean of observations have been calculated. For each observer, the correlation coefficient of *x/y* ratio by which the observer has agreed that algorithm *a* is better than algorithm *b* with the average of the same ratio for all the observers is computed. For all the pairs in Table 1 and all the eight observers the correlation coefficients calculated vary from 0.7 to 0.9 with an average of 0.8. Also, the correla-

Table 1: Chi-square values for comparing the results by the observers and the reproduction angular error. Each (x, y) represents (reproduction error, observer’s data). (x, y) denotes there are x pairs for which the observer made a consistent judgment and for y ($y \leq x$) of these pairs the observer agreed with the error metric.

	GE1-GE2	GE1-SOG	GE1-GP	GE2-SOG	GE2-GP	SOG-GP
observer 1	(7,7)	(6,5)	(7,4)	(7,7)	(7,6)	(6,4)
observer 2	(3,2)	(4,1)	(7,7)	(7,6)	(6,5)	(7,6)
observer 3	(7,4)	(7,5)	(7,6)	(7,6)	(7,7)	(7,7)
observer 4	(5,4)	(5,4)	(6,5)	(5,5)	(4,4)	(5,4)
observer 5	(6,4)	(7,7)	(7,6)	(7,5)	(6,6)	(7,6)
observer 6	(7,4)	(6,5)	(5,5)	(5,5)	(6,5)	(3,2)
observer 7	(6,4)	(6,6)	(6,4)	(7,7)	(7,6)	(7,6)
observer 8	(4,2)	(5,4)	(3,3)	(6,4)	(7,7)	(6,4)
Chi-square	5.44	3.55	2.40	1.52	0.62	2.30

Table 2: Chi-square values for comparing the results by the observers and the recovery angular error. Each (x, y) represents (recovery error, observer’s data). (x, y) denotes there are x pairs for which the observer made a consistent judgment and for y ($y \leq x$) of these pairs the observer agreed with the error metric.

	GE2-GE1	SOG-GE1	GP-GE1	SOG-GE2	GP-GE2	GP-SOG
observer 1	(7,0)	(6,1)	(7,3)	(7,0)	(7,1)	(6,2)
observer 2	(3,1)	(4,3)	(7,0)	(7,1)	(6,1)	(7,1)
observer 3	(7,3)	(7,2)	(7,1)	(7,1)	(7,0)	(7,0)
observer 4	(5,1)	(5,1)	(6,1)	(5,0)	(4,0)	(5,1)
observer 5	(6,2)	(7,0)	(7,1)	(7,2)	(6,0)	(7,1)
observer 6	(7,3)	(6,1)	(5,0)	(5,0)	(6,1)	(3,1)
observer 7	(6,2)	(6,0)	(6,2)	(7,0)	(7,1)	(7,1)
observer 8	(4,2)	(5,1)	(3,0)	(6,2)	(7,0)	(6,2)
Chi-square	22.44	31.55	34.40	40.52	42.62	32.30

tion coefficient between the x/y ratios for the individual observer ranges from 0.6 to 0.9. The highest agreement between the observers was for the GE2-GP pair of algorithms and the lowest correlation was for the GE1-GE2 pair. This is expected as the 1st and 2nd order grey edge algorithms (GE1-GE2) are instances of the same algorithm and their performances are very close in many cases which makes the choice difficult for the observers.

Conclusion

Evaluation of illuminant estimation algorithms using the reproduction and recovery angular errors shows there are sometimes disagreements between the two metrics regarding the ranking of a pair of algorithms. In this paper, a psychophysical study was conducted to investigate which of the two error metrics predicted the image preference judgments made by human observers.

The results of the experiments show that in most cases the observers agree with the evaluation by reproduction angular error. In other words, where according to reproduction angular error algorithm a is performing better than b , in most cases the observers make the same choice. Although, there are cases where the observers disagree with the reproduction angular error’s evaluation. However, the overall statistical analysis of the results using the Chi-square test shows the observers data highly agree with the results by reproduction angular error.

Perceptual analysis of images in terms of accuracy of reproduced colors is a difficult task since it could depend on many factors other than the accuracy of colors, such as content, etc. In

digital photography the aim is not always reproducing the colors which are colorimetrically accurate but a reproduction of preference is sometimes more desired. To this end, in the experiment performed in this paper, we also aimed to create a more photographic look for the raw images by passing them through an actual camera pipeline. This will provide the observers in the experiment with more natural photographic-look images and makes the task of comparison easier for them. To our knowledge, this is the first time in a psychophysics experiment concerning the quality of the color corrected images that the images are rendered to a photographic look before the experiment.

References

- [1] Arjan Gijzenij, Theo Gevers, and Joost V.D.Weijer. Computational color constancy: Survey and experiments. *IEEE Transactions on Image Processing*, 20, 9, pp. 2475-2489. (2011).
- [2] Roshanak Zakizadeh, Michael S. Brown and Graham D. Finlayson. The IEEE International Conference on Computer Vision (ICCV) Workshops, pp. 16-23. (2015).
- [3] Steven D. Hordley. Scene illuminant estimation: past, present, and future. *Color Research & Application* 31, 4, pp. 303-314. (2006).
- [4] Steven D Hordley and Graham D Finlayson. Reevaluation of color constancy algorithm performance. *Journal of the Optical Society of America A*, 23, 5, pp. 10081020. (2006).
- [5] Graham D. Finlayson and Roshanak Zakizadeh, Reproduction angular error: an improved performance metric for illuminant estimation. *Proceedings of British Machine Vision Conference (BMVC)*, (2014).

- [6] Graham D. Finlayson, Roshanak Zakizadeh and Arjan Gijsenij. The Reproduction Angular Error for Evaluating the Performance of Illuminant Estimation Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39, 7, pp. 1482 - 1488. (2017).
- [7] Graham D. Finlayson, Brian V Funt, and Kobus Barnard. Color constancy under varying illumination. In *Computer Vision, 1995. Proceedings., Proc. Fifth International Conference on*, pp. 720-725. (1995).
- [8] R. Zakizadeh and G. D. Finlayson. The correlation of reproduction and recovery angular errors for similar and diverse scenes. *Proc. Color and Imaging Conference, Society for Imaging Science and Technology*, (2015).
- [9] Arjan Gijsenij, Theo Gevers, and M. P. Lucassen. Perceptual analysis of distance measures for color constancy algorithms. *Journal of the Optical Society of America A*, 26, 10, pp. 2243-2256. (2009).
- [10] Javier Vazquez-Corral, C. Praga, Ramon Baldrich and Maria Vanelle. Color constancy algorithms: Psychophysical evaluation on a new dataset. *Journal of Imaging Science and Technology*, 53, 3, pp. 31105-1. (2009).
- [11] Dongliang Cheng, Dilip K. Prasad, and Michael S. Brown. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *Journal of the Optical Society of America A*, 31, 5, pp. 1049-1058. (2014).
- [12] Rajeev Ramanath, Wesley E Snyder, Youngjun Yoo, and Mark S Drew. Color image processing pipeline. *IEEE Signal Processing Magazine*, 22, 1, pp. 34-43. (2005).
- [13] Ayan Chakrabarti, Daniel Scharstein, and Todd Zickler. An empirical camera model for internet color vision. *Proc. BMVC*, pg. 4. Cite-seer, (2009).
- [14] Seon Joo Kim, Hai Ting Lin, Zheng Lu, Sabine Susstrunk, Stephen Lin, and Michael S Brown. A new in-camera imaging model for color computer vision and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34, 12, pp. 2289-2302. (2012).
- [15] Michael D Grossberg and Shree K Nayar. Determining the camera response from images: What is knowable? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25, 11, pp. 1455-1467. (2003).
- [16] Michael D Grossberg and Shree K Nayar. Modeling the space of camera response functions. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 26, 10, pp. 1272-1282. (2004).
- [17] Hai Ting Lin, Zheng Lu, Seon Joo Kim, and Michael S Brown. Nonuniform lattice regression for modeling the camera imaging pipeline. *Proc. European Conference on Computer Vision*, pg. 556-568. Springer, (2012).
- [18] Eric Garcia and Maya Gupta. Lattice regression. In *Advances in Neural Information Processing Systems*, pp. 594-602. (2009).
- [19] Eric Garcia, Raman Arora, and Maya R Gupta. Optimized regression for efficient function evaluation. *IEEE Transactions on Image Processing*, 21,9, pp. 4128-4140. (2012).
- [20] Finlayson, Graham D., and Elisabetta Trezzi. Shades of gray and colour constancy. *Proc. Color and Imaging Conference, Society for Imaging Science and Technology* (2004).
- [21] Forsyth, David A. A novel algorithm for color constancy. *International Journal of Computer Vision*, 5, 1, pp. 5-35. (1990).
- [22] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. Generalized gamut mapping using image derivative structures for color constancy. *International Journal of Computer Vision* 86, 2-3, pp. 127-139. (2010).
- [23] Joost Van De Weijer, Theo Gevers, and Arjan Gijsenij. Edge-based color constancy. *IEEE Transactions on image processing* 16, 9, pp. 2207-2214. (2007).
- [24] William Jay Conover. *Practical nonparametric statistics*, Third Edition. JohnWiley & Sons, New York, 1999. ISBN 0-471-16068-7.
- [25] Arjan Gijsenij, Theo Gevers, and Marcel P Lucassen. Perceptual analysis of distance measures for color constancy algorithms. *Journal of the Optical Society of America A*, 26, 10, pp. 2243-2256. (2009).
- [26] Richard L Alfvén and Mark D Fairchild. Observer variability in metameric color matches using color reproduction media. *Color Research & Application*, 22,3, pp. 174-188. (1997).

Author Biography

Roshanak Zakizadeh received her BSc in Computer Science from Shahid Bahonar University of Kerman and her PhD in Computer Science from the University of East Anglia. She also holds a Masters Degree in Color in Informatics and MEdia Technology (CIMET) from Université Jean Monnet and the University of Eastern Finland. She is currently a Research Engineer at V-Nova Ltd. in London.

Graham D. Finlayson obtained his BSc in Computer Science from the University of Strathclyde in 1989 and his MSc and PhD degrees from Simon Fraser University in 1992 and 1995 respectively. He spent the following years at the University of York in the UK as a Lecturer and later at the University of Derby as a Reader. In September 1999, he joined the University of East Anglia as a Professor.