A Hybrid Strategy For Illuminant Estimation Targeting Hard Images

Roshanak Zakizadeh
University of East Anglia
r.zakizadeh@uea.ac.uk

Michael S. Brown
National University of Singapore
brown@comp.nus.edu.sg

Graham D. Finlayson
University of East Anglia
g.finlayson@uea.ac.uk

Abstract

Illumination estimation is a well-studied topic in computer vision. Early work reported performance on benchmark datasets using simple statistical aggregates such as mean or median error. Recently, it has become accepted to report a wider range of statistics, e.g. top 25%, mean, and bottom 25% performance. While these additional statistics are more informative, their relationship across different methods is unclear. In this paper, we analyse the results of a number of methods to see if there exist 'hard' images that are challenging for multiple methods. Our findings indicate that there are certain images that are difficult for fast statistical-based methods, but that can be handled with more complex learning-based approaches at a significant cost in time-complexity. This has led us to design a hybrid method that first classifies an image as 'hard' or 'easy' and then uses the slower method when needed, thus providing a balance between time-complexity and performance. In addition, we have identified dataset images that almost no method is able to process. We argue, however, that these images have problems with how the ground truth is established and recommend their removal from future performance evaluation.

1. Introduction

Images captured by a digital camera are often corrupted by a color cast caused by the scene illumination. This color cast can make the image look unpleasant and hinder subsequent processing for computer vision tasks. Illuminant estimation algorithms attempt to estimate the scene illumination from the input image and apply a correction to remove the color cast.

Illuminant estimation algorithms can be roughly classified into two types: statistical-based methods and learning-based techniques (see [19] for a detailed survey). Statistical-based methods (representative examples include [25, 4, 27, 13, 20, 7]) directly estimate the illumination from statistics computed from the input image. These methods are fast and work irrespective of the type of camera

hard image   hard image   easy image

Figure 1: Examples of images from the Gehler-Shi dataset [15, 26] considered hard and easy based on our analysis of the performance of 12 different methods on the entire dataset.

used. Their performance, however, is generally not as good as learning-based methods. Learning-based methods (representative examples include [14, 12, 16, 15, 18, 5, 11, 23]) exploit the availability of training images that have labelled ground truth illumination. These methods use image features to train regressors to predict the illumination based on the input image and associated training-data. Learning-based methods generally give superior results over statistical methods, but at the cost of higher running-times and the need to be trained per camera. The selection of an illumination estimation method is generally guided by the need for performance vs. time-complexity, e.g. most onboard camera white-balance algorithms still use statistical-based methods.

There are several benchmark datasets [2, 9, 26] that are used to evaluate the performance of illumination estimation algorithms. These datasets generally have a large number of images. As a result, only aggregate performance values are given over the whole dataset. Early work on illuminant estimation often used the mean error as a common statistic [1]. However, later it was shown [21] that the mean error does not provide an accurate statistical summary of the underlying error distribution and the median error was proposed instead. In a perceptual evaluation [17, 19] of illuminant estimation algorithms the trimean error was introduced which gives the additional values of the error distribution (e.g. the top and bottom 25%).

The routine reporting of these additional statistics provides more insight to a method's performance across an entire dataset. Interestingly, however, none of the prior works have examined if there is any commonality in these statistics across the images in the dataset. For example, it is unclear if the bottom 25% results have shared images across different methods. This would be interesting finding as it would indicate the existence of images that multiple methods consistency perform poorly on. We term these images as 'hard images'. This lack of analysis serves as the impetus for our work.

Contribution In this paper, we describe an analysis on 12 leading illumination estimation algorithms belonging to both statistical- and learning-based methods. In particular, we enumerate over all combinations of five methods out of 12 to find the set of images where at least the majority (three or more) methods fail. We consider these images to be 'hard' for this subset of methods. Our findings indicate that there are, indeed, sets of hard images for different subsets (e.g. see Figure 1). More importantly, these subsets can be grouped depending if their methods belong to statistical-based or learning-based. To this end, we found that there are a number of 'hard' images for the fast statistical-based methods that can be handled by more complex learning-based approaches (Section 2). This led us to develop a hybrid estimation approach that classifies the image as hard or easy depending on the results of the statistical-based methods (Section 3). In the case an image is categorised as hard, it is likely that the results of the simple camera on-board white balancing algorithms are incorrect. Such hard images can be saved as RAW on the camera for later off-line processing by slower, but more accurate, learning-based methods, such as the exemplar-based method [23, 24]. This leads to better overall illumination estimation performance while reducing the overall time-complexity. Our analysis also has found that certain images in a well established benchmark dataset are hard for all methods. On closer examination we found that these images have issues that makes establishing the ground truth difficult and advocate for their removal for future evaluation (Section 4).

2. Analysing Results on a Common Dataset

The Gehler-Shi dataset [15, 26] has become the standard image dataset for illumination estimation performance evaluation. While newer datasets exist (e.g. the NUS 9 camera dataset [7]), the Gehler-Shi dataset remains the most commonly tested dataset in the color constancy literature. Gijsenij et al. [19] performed a thorough evaluation of 15 methods on the Gehler-Shi dataset. Their work provided results for each of these 15 methods for each image in dataset. We use this comprehensive results for our analysis in this paper.

From Gijsenij et al. [19], we select 12 algorithms that have received the greatest attention in the published literature. We divide them into two groups. *Statistical-based methods* including: S1 = shades of grey [13], S2 = grey world [4], S3 = 1st order grey edge [27], S4 = 2nd order grey edge and S5 = white-patch [25]. *Learning-based methods* including: L1 = exemplar-based [23], L2 = color constancy using natural image statistics [16], L3 = edge-based gamut, L4 = pixel-based gamut, L5 = intersection-based gamut [14, 18], L6 = Bayesian method [15] and L7 = spatial correlation [5].

As previously mentioned, Gijsenij et al. [19] provides the complete results (estimated illumination) by each 12 method for each image in the Gehler-Shi dataset. This dataset contains a total of 568 images involving two cameras, a Canon 1D (86 images) and a Canon 5D (482 images). Because the learning-based methods are trained per-camera, we focus on the Canon 5D given that it has the most images. This gives us a total of 482 images with 12 results, corresponding to the associated methods S1-5 and L1-7.

Our analysis is intended to find images that are collectively hard for multiple methods. In this case, 'hard' images are those where multiple methods are unable to estimate the illumination within some error threshold. In this paper, we use nine degrees error as this threshold, meaning that the estimated illumination has at least nine degrees (or more) angular difference from the ground truth illumination. Nine degrees is used as it represents a threshold that categorises typical error of the bottom 25% for most methods as reported by Gijsenij et al. [19]. Thus, we are comparing the images that are reported to give the worse performances for the 12 methods.

When we examine which images in the dataset that have at least nine degrees of error for all 12 methods, we found there are only a few images (this finding is discussed in more detail in Section 4). This means that there is significant variation in the images that different methods perform poorly on. To provide a more manageable grouping, we consider all combinations of 5 methods from the 12 total (i.e. 12 choose 5). In particular, we enumerate all five combinations of the 12 methods which gives total of 792 combinations. Among these combinations, we are interested in those for which at least three out of five methods introduce errors higher than our threshold. This is illustrated in Figure 2 which shows one out of the 792 combinations. The columns in Figure 2 represent a unique image in the dataset. The rows represent the five different methods tested. A white-box means a method has failed for this particular image (i.e. produces a high error). A black-box means the method is successful. Three or more empty boxes for a particular column represents an image where the majority of methods has failed. This is considered a 'hard' image for this particular combination of methods. For the

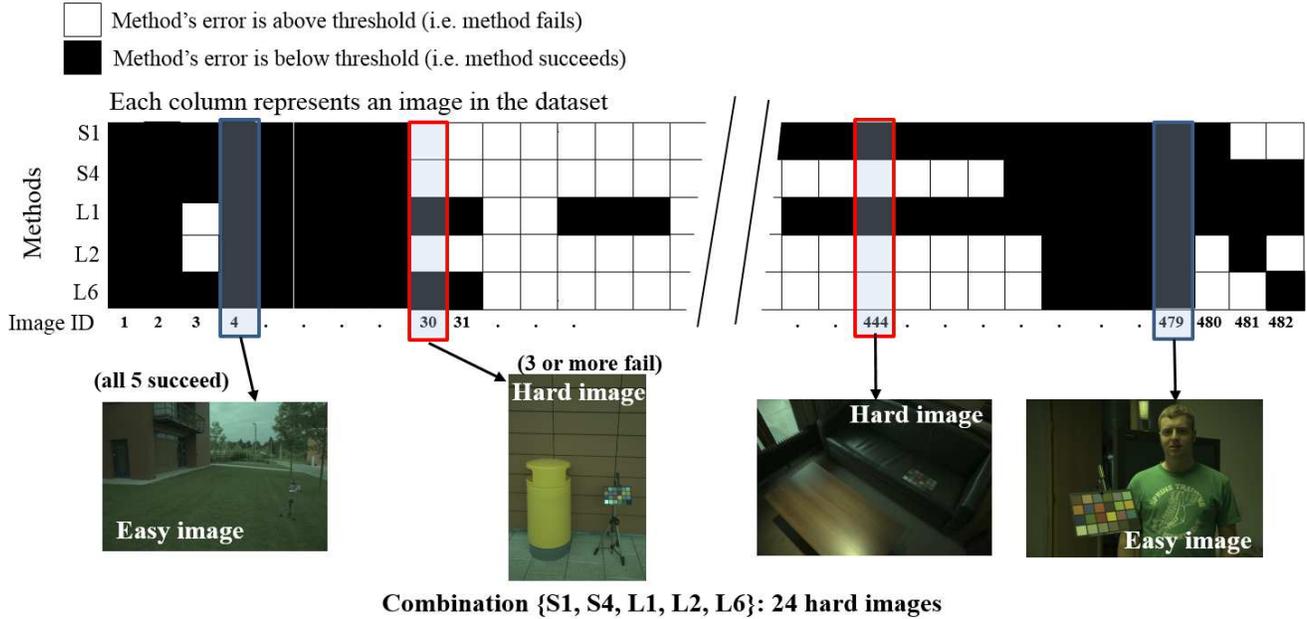# A combination of five illuminant estimation algorithms

Figure 2: A combination of five illuminant estimation algorithms. This combination results in 24 hard images out of 482 images of Canon5D from Gehler-Shi dataset [15, 26].

Combinations with **most** 'hard' images			Combinations with **least** 'hard' images		
Methods	failed images	Time (m)	Methods	failed images	Time (m)
S2 S3 S5 L3 L7	84	1.5	L1 L2 L4 L6 L7	31	12.6
S2 S3 S5 L3 L6	80	9.8	S4 L1 L2 L5 L7	27	3.7
S2 S3 S5 L3 L4	78	1.8	S1 S4 L1 L2 L6	24	11.2
S2 S3 S4 S5 L3	73	1	S2 L1 L2 L6 L7	22	11.7
S1 S2 S3 S4 S5	69	0.36	S2 S4 L1 L2 L7	19	2.87
S1 S2 S4 S5 L4	64	1.2	S2 S4 L1 L6 L7	18	11.1

Table 1: The five combinations out of 12 illuminant estimation algorithms in terms of number of images they fail for. We have highlighted the fastest (on the left) and slowest (on the right) combinations. Running times given are per image.

example shown, the combination are methods (S1, S4, L1, L2, L6), and this set results in 24 hard images.

This procedure is performed for all combinations of 5 methods out of the 12. For each combination, we record the number of hard images per combination and sort the list of combinations based on the number of hard images. Table 1 includes the combinations with **most** and **least** 'hard' images. Almost all combinations with **most** 'hard' images include three or more simple statistical-based algorithms. The combinations with **least** 'hard' images are mostly dominated by learning-based methods.

Each method examined has a time complexity associated with it. The work by Gijsenij et al. [19] did not report this time-complexity, however, more recent work has examined most of the same methods and reported the running-time [7]. The only exception is that of the exemplar-based method (L1). For this method, we estimate its time to take approximately twice that of the gamut-based methods based on the running-time reported by the author [22]. The fastest and slowest combinations are highlighted in Table 1. The statistical-based methods in general have a much faster running-time than the learning-based methods. For exam-
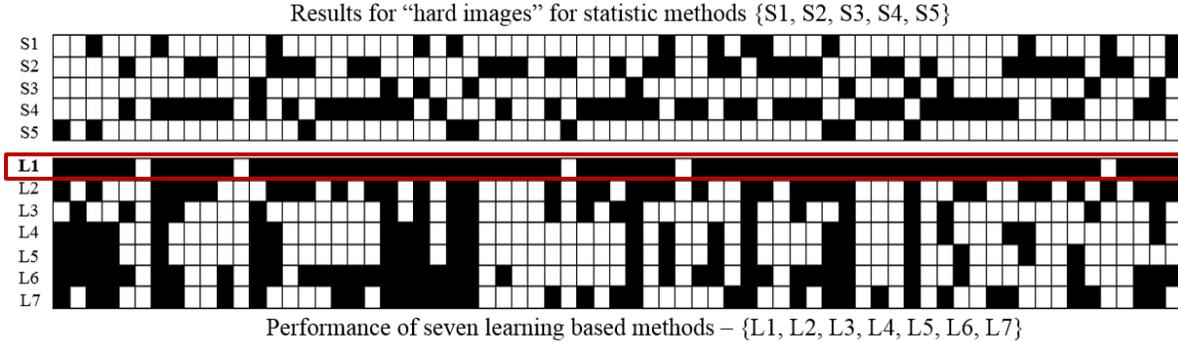
Figure 3: The hard images from the Gehler-Shi dataset [15, 26] for the five statistical-based methods. L1-L7 rows are the performance of the learning-based methods.

ple, the highest number of hard images is 84 that is achieved using the combination in the first row of section 'combinations with **most** hard images' in Table 1. This set of methods requires roughly 1.5 minute per image to run all 5 methods. The overall run-time is mainly attributed to the two learning-based techniques: (L3) edge-based gamut [20] and (L7) spatio spectral [5].

The fifth largest number of failure images (out of 792 combinations) is for the set of the five statistical methods (S1-S5). This is highlighted on the left in Table 1. This only requires approximately 0.36 minutes per image and is the fastest of all the combinations. This is a very interesting finding. It shows that the statistical-based methods tend to collectively fail on the same images in the dataset. This means that we have a chance to examine these images to see if we can build a classifier that can predict if an image is 'hard' or 'easy' for this set of methods. The question now is can we find a method that performs well on the hard images for the statistical-based approaches.

Given the combination of five statistical-based methods and their associated hard images, we examine the performance of the learning-based methods. Figure 3 shows the results. The diagram shows all 69 of the hard images (where at least three or more of the learning-based methods fail). The rows below show the results of the L1-L7 learning-based methods. It is interesting to note that there are some images considered hard for the statistical-based method that all learning-based method are successful on. Overall, however, the L1 (exemplar-based [23]) method does particularly well for the hard images, able to produce a better result on all except a few of the images.

Based on the analysis in this section, we have developed a hybrid method that first applies the statistical based approaches. As discussed in the next section, from this we can classify if the image is hard or easy. For images that are classified as hard, we propose that they are saved as RAW

(on the camera) for later to be processed off-line by learning based methods such as the exemplar-based (L1).

3. Hybrid Method for Targeting Hard Images

In this section, we describe our framework to classify images as hard or easy and then process them accordingly. As discussed in Section 2, an image is labelled as hard if at least three out of five simple statistical-based algorithms have an error beyond nine degrees. We label an image as easy if all five methods succeed, i.e have an error below the threshold. We set the threshold for easy images as eight degrees which is slightly lower than the hard images threshold. We use these labelled images as training data to build a classifier.

3.1. Features and Classifier

We have experimented with several image features to be used in designing a classifier to label a new input image as either hard or easy. One feature commonly used in learning-based color constancy methods is the rg chromaticity values ($[r, g] = [R, G]/(R + G + B)$). These are typically used to compute a histogram over the r and g values as features. We found, however, that the distribution of the rg values had little correlation to image being labelled hard or easy. We also examined the chromaticity values with respect to the rg chromaticity curve of the ground-truth illuminants (i.e. the locus of ground-truth illuminants in chromaticity space). Again, we found that these had little correlation to whether an image was labelled as hard or easy.

The lack of success with chromaticity values led us to examine features defined in the full 3D RGB space. In particular, we looked at the mean (centroid) location of the five estimated illuminants provided by the statistical methods (S1-S5). Figure 4 (top) shows the distribution of these centroid of the estimated illuminants for a set of hard (red) and easy (blue) images from Gehler-Shi dataset. We can see

Feature	Overall accuracy	Hard image accuracy	Easy image accuracy
1. Centroid	93.6%	85%	96.6%
2. Median from centroid	86.7%	68.3%	94.3%
3. Standard deviation (std)	82%	42.3%	95.9%
4. Centroid + std	89.7%	68.1%	95.9%
5. Median + std	85.4%	59.2%	94.7%

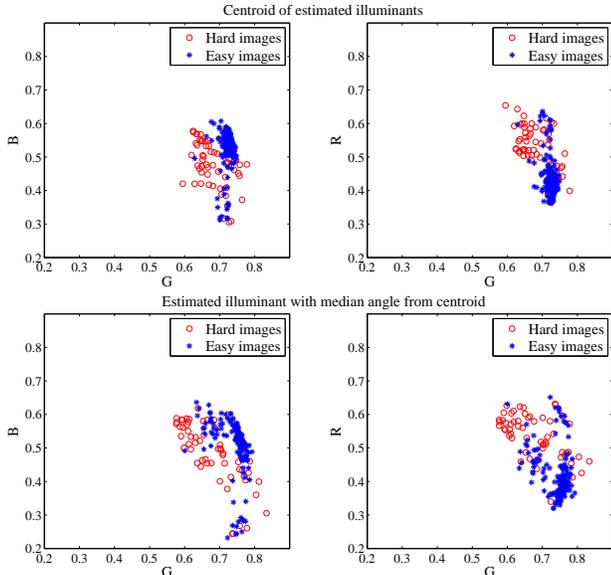Table 2: Performance of the SVM classifier with different features.

Figure 4: The top two diagrams show the centroids of five estimated illuminants for hard and easy images. The bottom two diagrams are the selected illuminants out of five estimations with median angle from the centroid. The features for easy images form a cluster in both cases.

that these form two distinct clusters of points. We also calculated the angle between each of five estimated illuminants and the centroid of the estimates. Among five estimates we selected the one with the median angle from the centroid. The points in the last two plots of Figure 4 (bottom) belong to the selected estimated illuminants. While there is a discernable pattern in the data, it is not as distinguishable as that with the the clusters of centroid.

Based on the observation in Figure 4, we experimented with classifiers using five different features: 1) The centroid of the five estimated illuminants; 2) the estimated illuminant selected out of the five estimates with the median angle from the centroid; 3) feature 1 and the standard deviation of the five estimated illuminants; 4) feature 2 and the standard deviation of the five estimated illuminants; and 5) the standard deviation of the five estimated illuminants. The

features were used to train a support vector machine (SVM) [10] classifier based on the implementation of Chang and Lin [6].

Table 2 shows the overall accuracy of the SVM classifier with all the features as well as how accurate the model classifies hard and easy images. We found that the simple centroid feature produced the best results over all the five features and use it in our overall framework.

3.2. Overall Procedure

The overall framework of our hybrid strategy can be seen in Figure 5. For a given input image, its illumination is estimated by the five statistical-based methods (S1-S5). The centroid (mean) of the five estimates is calculated and used with the SVM to predict if the image is hard or easy. If the image is classified as hard, we use a learning-based method such as the exemplar-based method [23] to process the image to obtain the final illumination estimate.

If the image is classified as easy, we have five estimates to choose from. A straight-forward option would be to use the average of these estimations. This is reported in our experiments in the next section. However, another option is to use this information to get a better prediction of the illuminant. In particular, the recent 'corrected moments' work by Finlayson [11] showed that a correction matrix can be pre-computed using the ground-truth illuminants from the training-data to correct the estimates of the existing simple derivative-based statistical methods. In this case, we can use the result of the two derivative-based methods S3 and S4 (1st grey edge and 2nd grey edge) to build the correction matrix. We found this approach gives notably better results over using the average of the S1-S5 scores. This is also reported in the experiments in the following section.

3.3. Experiments

We have tested our hybrid strategy on the Gehler-shi [15, 26] dataset using different features mentioned in Section 3.1. To generate a set of labelled data we categorise hard and easy image based on their thresholds (here we set eight degrees for easy images and nine degrees for hard images as explained in Section 2). Out of 482 images of *Canon 5D* from *Gehler-shi* [15, 26] dataset, this results in 233 labelled images. The sets of training and test images are are made
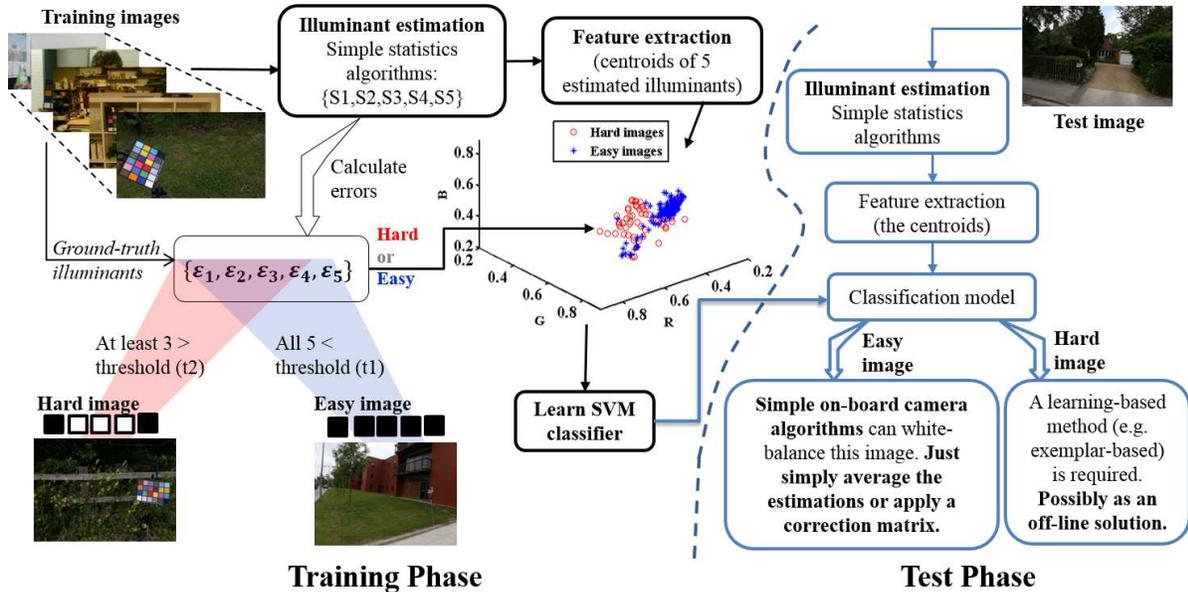
Figure 5: The proposed framework which focuses on illuminant estimation for hard images. The classifier categorises images into hard and easy. The easy images can be treated using fast statistical-based techniques. Images classified as hard are processed using a slower, but more accurate learning-based method.

by 3-fold cross validation, i.e. each fold has 155 training and 78 test images. The SVM classifier based on the 'centroid' feature is built on the training set and the accuracy of it is examined on the test images. The model's performance on this set of 78 test images showed an accuracy of 93.6% with 85% for classifying hard images and 96.6% for classifying easy images. Table 3 shows the result of the model applied on a set of unlabelled images.

The performance of the five statistical methods (S1-S5) for all images are shown in the first row of the Table 3. The L1 column shows the error of exemplar-based method for all images. The exemplar-based has an overall good performance for all images but is significantly slower than the S1-S5 methods combined.

In our hybrid algorithm, we use our SVM to classify the input images. In the first column of the proposed section of Table 3 the average of statistical-based methods is used as our estimate. By excluding hard images we have avoided the high error of S1-S5 that is obtained when applied to all images. It is interesting to note that median of the average of S1-S5 is less than the median of the individual methods. As previously mentioned, we also use the corrected-moment illuminant estimation method [11] to further improve the results. This method uses a cross validation procedure to build a correction matrix that takes the results from the S3 and S4 estimates and refine the result based on the ground-truth illuminants of training data. Table 3 shows the (corrected) algorithm performance. This allows us to get an additional

gain on the performance of the statistical based methods. Note that the approach in [11] still has trouble on the hard images and the use of the exemplar-based method is significantly better and therefore necessary for the hard images.

Our results show that this strategy of using fast statistical-based methods can give us good performance on the easy images, while identifying the difficult images and passing them to a slower, but more accurate learning-based approach. While the overall running-time is slow due to the use of the learning-based method, our approach can reduce this by almost half while giving similar performance. Moreover, the results for easy images can be obtained in a matter of seconds.

4. False Hard Image Removal

As mentioned in Section 2 we found nine images that almost all methods failed on from the Gehler- Shi [15, 26] dataset. We were keen to see if there were some characteristics to the hard images that no method could resolve, however, on careful inspection of these images we realise it was due to the position of the color chart in the scene. Figure 6 shows an example.

In all of these images, the color checker board that is used to provide the ground truth illumination is placed under a different illumination than the rest of the scene. This means the scene is lit by two different illuminations, but in the cases of these nine images, the dominant illumination

	S1	S2	S3	S4	S5	L1	Proposed	
							average	corrected
All	4.37°	7.04°	4.81°	4.73°	6.46°	2.4°		
Easy	3.5°	6.9°	4.26°	4.7°	4.7°	2.1°	3.42°	**2.4°**
Hard	6°	7.04°	6.1°	4.8°	12.9°	2.91°	2.91°	**2.91°**
Time (per image)	3.4s	1.8s	6.8s	8s	1.85s	1.96m	21.9s + (1.96m per hard image)	
Time (total)	18.5m	9.8m	36.9m	43.5m	10m	10.7h	4.5h	

Table 3: The median errors of the proposed hybrid framework treating hard and easy images differently. In comparison we show the errors of fast statistical algorithms (S1 to S5), as well as time complexity of exemplar-based method [23] (L1).

Figure 6: Examples of images that all methods incorrectly estimate the illumination on from the Gehler-Shi dataset [15, 26].

arguable does not fall on the color checker board. These images do not represent fair test cases and should be removed as they introduce negative results for evaluation and are erroneously used by learning-based methods for training. We have provided an updated version of the Gehler-Shi [15, 26] which excludes these nine images and their measured ground truth illuminations[1].

5. Concluding Remarks

This paper has analysed the performance of multiple color constancy methods to examine if methods fail on the same images. As far as we are aware, this is the first work to examine the relations of the hard images across different color constancy methods.

Our analysis revealed that there are common 'hard' images for subsets of methods. One of these subsets with a large number of hard images is composed of all fast statistical-based color constancy methods. We also observed that there exist some learning-based methods that give excellent performance on this set of hard images, but at a significant cost in running-time. Based on these observations, we proposed a hybrid method that classifies an image as hard or easy and then processes it accordingly. This allows easy images to be processed quickly. Easy images white-balancing could even be performed onboard the camera itself. For the images classified as hard, learning-

based methods such as the exemplar-based method [23] are applied to give good results. We note that learning-based methods will continue to improve in terms of performance and speed. Recent work by Bianco et al. [3] and Cheng et al. [8] provided similar estimation performance to the exemplar-based method (L1) used in our work, but at a significantly faster running-times. These methods can be easily incorporated into our overall framework's running time, however, we note that learning-based methods will still need to be performed off-line and therefore require the determination of which images are 'hard' and require such off-line processing.

Our analysis has also identified nine images in the widely used Gehler-Shi [15, 26] dataset that were problematic for all 12 methods we examined. We have found that these images have problems with how the ground-truth is established and we recommend their removal from the dataset for the future studies.

For future work we are keen to extend our idea to additional color constancy datasets. Currently, we were only able to apply this approach to the Gehler-Shi dataset as it has sufficient number of images. More recent datasets (e.g. NUS 9-camera) has more overall images, but less images per camera (only around 200 images per camera). We did attempt to apply this approach to the older Grey-ball dataset [9] but found the dataset is inappropriate given that it is low-resolution video footage (320×240) and is not properly linearised. We also found that this dataset had a large number of hard images due to improper position of the Grey-ball used for the ground truth. This points to the need of additional datasets in the color constancy community and is an area we are focusing on for future work.

6. Acknowledgement

References

[1] http://colour.cmp.uea.ac.uk/datasets/GehlerFalse.html

[1] K. Barnard, V. Cardei, and B. Funt. A comparison of computational color constancy algorithms. I: Methodology and

experiments with synthesized data. *IEEE Transactions on Image Processing*, 11(9):972–984, 2002. 1

[2] K. Barnard, L. Martin, B. Funt, and A. Coath. A data set for color research. *Color Research & Application*, 27(3):147–151, 2002. 1

[3] S. Bianco, C. Cusano, and R. Schettini. Color constancy using CNNs. *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Deep Vision: Deep Learning in Computer Vision*, 2015. 8

[4] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1):1–26, 1980. 1, 2

[5] A. Chakrabarti, K. Hirakawa, and T. Zickler. Color constancy with spatio-spectral statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(8):1509–1519, 2012. 1, 2, 4

[6] C.-C. Chang and C.-J. Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011. 6

[7] D. Cheng, D. K. Prasad, and M. 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):1049–1058, 2014. 1, 2, 3

[8] D. Cheng, B. Price, S. Cohen, and M. S. Brown. Effective learning-based illuminant estimation using simple features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1000–1008, 2015. 8

[9] F. Ciurea and B. Funt. A large image database for color constancy research. 11th *IS&T/SID Color and Imaging Conference*, 2003(1):160–164, 2003. 1, 8

[10] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995. 6

[11] G. D. Finlayson. Corrected-moment illuminant estimation. *IEEE International Conference on Computer Vision (ICCV), 2013*, pages 1904–1911, 2013. 1, 6, 7

[12] G. D. Finlayson, S. D. Hordley, and P. M. Hubel. Color by correlation: A simple, unifying framework for color constancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1209–1221, 2001. 1

[13] G. D. Finlayson and E. Trezzi. Shades of gray and colour constancy. *Color and Imaging Conference*, 2004(1):37–41, 2004. 1, 2

[14] D. A. Forsyth. A novel algorithm for color constancy. *International Journal of Computer Vision*, 5(1):5–36, 1990. 1, 2

[15] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp. Bayesian color constancy revisited. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008. 1, 2, 3, 5, 6, 8

[16] A. Gijsenij and T. Gevers. Color constancy using natural image statistics. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007. 1, 2

[17] A. Gijsenij, T. Gevers, and M. P. Lucassen. Perceptual analysis of distance measures for color constancy algorithms. *Journal of the Optical Society of America A*, 26(10):2243–2256, 2009. 1

[18] A. Gijsenij, T. Gevers, and J. Van De Weijer. Generalized gamut mapping using image derivative structures for color constancy. *International Journal of Computer Vision*, 86(2-3):127–139, 2010. 1, 2

[19] A. Gijsenij, T. Gevers, and J. Van De Weijer. Computational color constancy: Survey and experiments. *IEEE Transactions on Image Processing*, 20(9):2475–2489, 2011. 1, 2, 3

[20] A. Gijsenij, T. Gevers, and J. Van De Weijer. Improving color constancy by photometric edge weighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5):918–929, 2012. 1, 4

[21] S. D. Hordley and G. D. Finlayson. Reevaluation of color constancy algorithm performance. *Journal of the Optical Society of America A*, 23(5):1008–1020, 2006. 1

[22] H. R. V. Joze. *Estimating the colour of the illuminant using specular reflection and exemplar-based method*. PhD thesis, Applied Sciences: School of Computing Science, 2013. 4

[23] H. R. V. Joze and M. S. Drew. Exemplar-based colour constancy. *British Machine Vision Conference, 2012*, pages 1–12, 2012. 1, 2, 4, 6, 8

[24] H. R. V. Joze and M. S. Drew. Exemplar-based color constancy and multiple illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(5):860–873, 2014. 2

[25] E. H. Land. The retinex theory of color vision. *Scientific American*, 237(6):108–128, 1977. 1, 2

[26] L. Shi and B. Funt. Re-processed version of the gehler color constancy dataset of 568 images. *Simon Fraser University*, 2010. 1, 2, 3, 5, 6, 8

[27] J. Van De Weijer, T. Gevers, and A. Gijsenij. Edge-based color constancy. *IEEE Transactions on Image Processing*, 16(9):2207–2214, 2007. 1, 2