Removing Outliers in Illumination Estimation

Brian Funt and Milan Mosny; School of Computing Science, Simon Fraser University, Burnaby, British Columbia, Canada

Abstract

A method of outlier detection is proposed as a way of improving illumination-estimation performance in general, and for scenes with multiple sources of illumination in particular. Based on random sample consensus (RANSAC), the proposed method (i) makes estimates of the illumination chromaticity from multiple, randomly sampled sub-images of the input image; (ii) fits a model to the estimates; (iii) makes further estimates, which are classified as useful or not on the basis of the initial model; (iv) and produces a final estimate based on the ones classified as being useful. Tests on the Gehler colorchecker set of 568 images demonstrate that the proposed method works well, improves upon the performance of the base algorithm it uses for obtaining the sub-image estimates, and can roughly identify the image areas corresponding to different scene illuminants.

Introduction

There are two types of outliers that create problems for illumination-estimation algorithms. In the illumination-estimation context, an outlier is an observation that does not fit the illumination model well. One type of outlier arises from noise in the image data created, for example, by a speck of dust on the imaging sensor, or by clipping of high digital counts. A second, more interesting, type of outlier arises from scenes that do not fit the expected model of the scene illumination—for example, a predominantly indoor scene with some light also coming through a window. In this paper, we apply the RANSAC (random sample consensus) [5] technique to randomly sampled subwindows of the input image as a means of handling outliers of both these types. The proposed algorithm was evaluated on Shi's reprocessed version of Gehler's original 'colorchecker' set of 568 images [7, 14] and found to reduce the high-percentile angular errors by roughly 30%.

Illumination estimation is the crucial step in standard automatic white balancing or 'color constancy'. The goal in illumination estimation is to determine the chromaticity of the overall scene illumination. The accuracy of the estimate is often measured in terms of the angular difference in degrees between the estimated chromaticity and the actual chromaticity when treated as vectors in 3-space. Many illumination-estimation algorithms have been proposed and are surveyed by Barnard *et al.* [1], Hordley *et al.* [10] and Gijsenij *et al.* [8].

Noting that outliers of the two types mentioned above will mislead most illumination-estimation algorithms, we propose a novel technique to improve upon any given algorithm or set of algorithms by explicitly taking outliers into account. The proposed method divides the input image into smaller sub-images, runs the algorithm(s) on each of the parts independently and then combines the resulting estimates. Outliers are identified and eliminated as part of the process of combining the estimates.

The rational behind the proposed method is that many illumination-estimation algorithms rely on information that can be significantly influenced by a small part of an image. For example, MaxRGB [7] and retinex [11] both can be influenced by a single pixel having a spuriously high R, G or B value. Gamut mapping algorithms such as Forsyth's [6] can be influenced by a single erroneous pixel that happens to stretch the convex hull of the gamut significantly in the wrong direction. For a single-illuminant scene, the estimates from multiple sub-images should be consistent. Those that are inconsistent can be identified as outliers and eliminated.

Sub-images, of course, do not carry as much information as the whole image. Therefore, the performance of an algorithm on each of the parts is likely to be worse than its performance on the full image; that is, assuming the full image contains no outliers. However, it is often the case that even a fairly small sub-image contains enough information for the underlying illumination-estimation algorithm to work reasonably well. As an example, consider the images in Figure 1. It is likely that quite a few of the vertical or horizontal slices will cover a sufficient proportion of the complete set of image colors to make gamut mapping algorithms work. Similarly, there is a good chance that they contain the necessary high R, G or B digital counts that MaxRGB requires, or are sufficiently textured for Edge-based Color Constancy [15] to succeed.

Illumination-estimation algorithms generally assume there is a single illuminant lighting the imaged scene, or at least that even if there is more than one illuminant then there is only one dominant illuminant. It is expected that white balancing the image relative to the dominant illuminant will suffice. In terms of the second type of outlier—those related to multiple illuminants—the information from a sub-image is likely to be more reliable, not less, than that from the image as a whole, because by being smaller the sub-image is more likely to involve only a single illuminant.

Proposed Algorithm

The proposed algorithm combines illumination estimates obtained from sub-images using RANSAC as a method of eliminating outliers. The core idea of RANSAC is to determine which observations are inliers and which are outliers and to base the final result only on the inliers. The data is assumed to fit some underlying model defined by some parameters (e.g., a model could be a straight line with the slope and intercept being the parameters). The process works by: (1) randomly selecting some observations; (2) determining the model parameters that best fit those observations; (3) testing all the remaining observations and classifying them as inliers or outliers based on how well they conform to the model; (4) checking to see that a sufficient number of inliers remain, and if not, discarding the model; (5) re-computing the model parameters based on the complete set of inliers. The algorithm repeats the steps (1) to (5) to generate many possible models. The model that fits the observations the best is returned as the result.

RANSAC is used to deal with the two types of outliers mentioned above. The underlying model of the image data that RANSAC fits is different for each of them. To handle multiple sources of illumination, the model is that the scene contains 3 distinct illuminants. Any additional illumination chromaticities that are observed (i.e., calculated from a sub-image by an illumination-estimation algorithm) will be classed as outliers. Applying the RANSAC steps in this case means obtaining illumination estimates from 3 sub-images of random size and location (the 'observations'), sorting the remaining estimates as either inlier or outlier, checking that a sufficient number of inliers has been found, and re-computing the estimates for each of the 3 clusters obtained this way based on the final set of inliers. For this last step, the inlier estimates within each cluster are simply averaged.

In terms of the outliers of the pixel-noise variety, the scene model is that there is only a single illuminant, and hence all the sub-image estimates should conform to this. Those that do not will be considered to be outliers. Note that this definition of outlier will also deal with secondary illuminants to a certain extent. If a secondary illuminant only lights a small portion of the scene then the estimates from the corresponding sub-windows will be excluded as outliers and therefore not influence the estimate of the dominant illuminant.

Related Work

Combining estimates from multiple illumination-estimation algorithms applied to whole images has been explored before, but generally not to sub-regions of images. One simple method of combining estimates is to use the arithmetic mean [3, 2]. Other strategies are to use a weighted average with weights determined by a training stage using least squares, or to use a trained neural network or support vector regression [12] to combine the estimates [3].

Shades of Grey [4] and Edge-based Color Constancy [15] combine MaxRGB-type and Greyworld-type clues. In both cases, the results are proportional to the Minkowski or p-norm of the form

$$C = \left(\frac{\sum_k C_k^p}{p} \right)^{\frac{1}{p}}$$

where C stands for R, G, or B values of individual pixels in the case of Shades of Gray, or of for the n^{th} derivative of a Gaussian-smoothed image in case of Edge-based Color Constancy. P-norm returns an arithmetic mean of values for $p = 1$. It returns the maximum of values c_i for $p = infinity$. By varying the parameter p, the algorithm returns results between grey world and MaxRGB, and in this sense it combines the two methods.

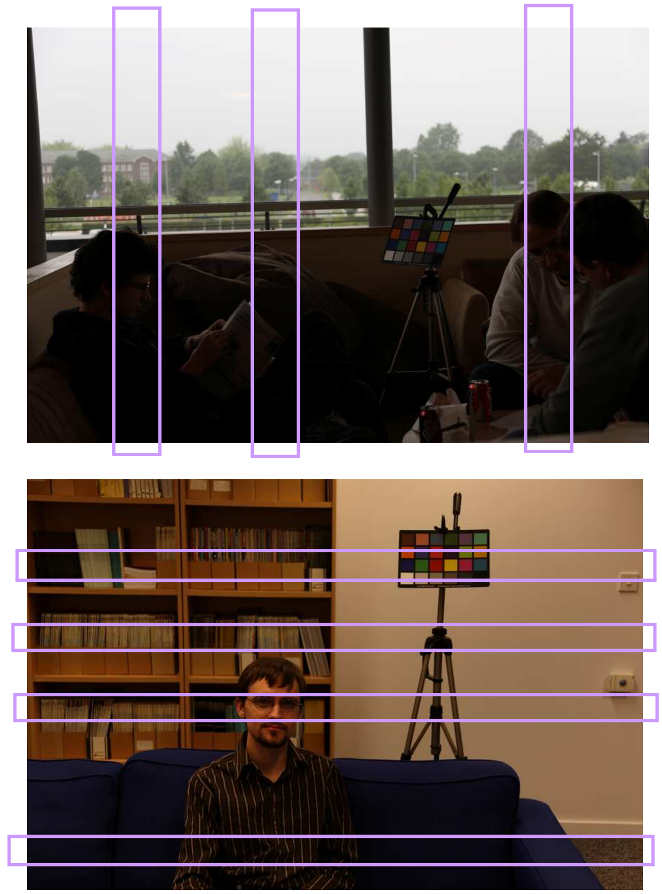

Figure 1 *Example images from the Gehler-Shi colorchecker set [7, 14] where vertical (top) and horizontal (bottom) sub-images may include almost as many clues about the illumination as the full image. However, some sub-images, such as the lowest horizontal one in the bottom image, will not contain enough information for a good estimate to be made.*

A related approach is that of Gijsenij *et al.* [9] in which MaxRGB and Greyworld are combined by applying MaxRGB not to the raw image data, but instead to a Gaussian-smoothed version of it. In either case, MaxRGB selects the maximum R, maximum G and maximum B of the image (smoothed or not smoothed). The Gijsenij *et al* algorithm is equivalent to MaxRGB when the scale parameter for Gaussian smoothing is set to 0, Greyworld when σ is set to infinity, and a combination of the two for values in between.

Various methods of computing the results based on a consensus between a given set of algorithms were explored by Bianco *et al*. [2] including picking the median, mean of the two closest results, the mean value of the results of the algorithms with relative distances below (100+N)% of the distance of the two closest estimates, and "No-N-Max" combination where the estimates are sorted by the sum of distances from other algorithms' estimates and the mean of the estimates excluding the N having the highest distance is returned.

The algorithm proposed here differs from the above-mentioned methods in that it combines the estimates from many image sub-regions as opposed to combining estimates from many

© 2012 Society for Imaging Science and Technology

algorithms applied to the whole image. To the best of our knowledge, RANSAC [5] has not been used previously as a means of combining illumination estimates.

Implementation Details

The proposed algorithm entails splitting the image into multiple parts, running a selected illumination-estimation algorithm on each of the parts and then combining the results in some fashion.

Four distinct ways of splitting the image into parts were implemented. These include splitting the image into overlapping vertical slices, overlapping horizontal slices, regular overlapping rectangles and random (possibly overlapping) rectangles. Each slice covers one tenth of the image. Slices are obtained by moving an initial slice at the left or top of the image one thirtieth of the image width or height to the right or down. Each slice (except for the left-most or top-most) covers two-thirds of the previous slice. This way, the image is split into 28 vertical slices or 28 horizontal slices. Overlapping regular rectangles are generated in a similar manner. Each rectangle is one tenth of the image wide and one tenth of the image tall. Rectangles were placed in the corners of a regular 30 by 30 grid. The size of the random rectangles varied from a thirtieth to a half the image width, and between a thirtieth and a half of the image height. They were placed on a random point of a regular 30 by 30 grid covering the input image. In total, 100 random rectangles where generated.

MaxRGB was used as the algorithm to run on each of the image parts but other choices are certainly possible and will be explored in the future. On its own, MaxRGB performs reasonably well on the Gehler-Shi colorchecker set so it provides a good, simple baseline with which to compare any improvement that the removing outliers might make.

RANSAC is applied as described above using 5 degrees of angular difference between its current model and a candidate estimate as the threshold for determining which estimates to classify as outliers. In other words, for the single-illuminant case, a sub-image illuminant estimate is considered to be an inlier if the angle between the illuminant rgb estimated for the sub-image and the rgb of the current candidate model is less than 5 degrees. The candidate model that fits the highest number of sub-regions is selected and the mean of all subwindow illuminant estimates that fit the model is returned as the final estimate.

For the multiple-illuminant case, the threshold is also set to 5 degrees. The candidate model is formed by picking 3 random sub-images and using their 3 separate estimates as the model of the 3 illuminants. For each input image, 400 candidate models are generated. The process returns the model that fits the estimates produced from the largest number of subwindows. A subwindow fits a candidate model if its illuminant estimate differs by no more than 5 degrees from at least one of the three rgb chromaticities of the model. The algorithm essentially clusters the estimates, while at the same time eliminating those that do not fit any cluster. This is in contrast to k-means clustering, which includes all initial points in the final clusters. The mean of the cluster that contains the most estimates is returned as the final illuminant chromaticity.

Test Results

The proposed outlier removal method in its numerous variations (horizontal vs. vertical slices vs. regular rectangles vs. random rectangles, single-illuminant RANSAC model vs. three-illuminant RANSAC model) was evaluated on the Gehler-Shi colorchecker set [7, 14]. For comparison, the sub-image estimates were also combined by simple averaging.

Images were pre-processed first. These filters were applied:

1. Dark pixel removal. Dark pixels, that is, those with $R + G + B < Threshold,$ are removed. The *Threshold* is set to an average of $R+G+B$ values collected from all pixels in the image.
2. Clipped pixel removal. Pixels whose *RGB* values are above the upper limit of the camera's dynamic range are removed. The threshold is set to 98% of the maximum *RGB* value of 255.
3. Even blocks pre-processing [13]. Size of the neighborhood N is set to 5.

The results are evaluated in terms of the angular error between the *rgb* chromaticity vector of the estimated illuminant and the chromaticity vector of the true scene illuminant as provided in the dataset. The error is reported in degrees. Figures 2, 3, 4 and Table 1 show the results.

The plots in Figures 2 and 3 show the angular error as a function of the percent of images having that error or less. The figures include MaxRGB on its own, the simple mean of the sub-image estimates, and 8 variations of the RANSAC method corresponding to vertical slices vs. horizontal slices vs. regular rectangles vs. random rectangles, and single-illuminant model versus three-illuminant model.

For the three-illuminant model, Figure 4 shows both the error based on the estimate from the most populous cluster (i.e., the algorithm's best guess) as well as the smallest error between the true illuminant and the estimate from any of the three clusters (labeled "Best of Three" in the figure). In essence, this last error measure is using the best of the algorithm's three guesses. As such, it will necessarily reduce the error. It is included in the results here for comparison because of a problem with the Gehler data set, which is that the colorchecker used for measuring the 'true' illumination often appears not to be ideally located. For example, in the upper image of Figure 1, the dominant illuminant is from the window, but the colorchecker is facing inwards and is therefore illuminated from the room light. This is a clearly a multi-illuminant scene for which a single colorchecker measurement is insufficient.

Figures 2, 3 and 4 show the distribution of errors over the test set and it can be used to determine angular error at a particular percentile. The outlier-removal strategy outperforms MaxRGB at the 50^{th} percentile and higher. At the 90^{th} percentile, the difference is significant. The vertical vs. horizontal variations are fairly similar to one another.

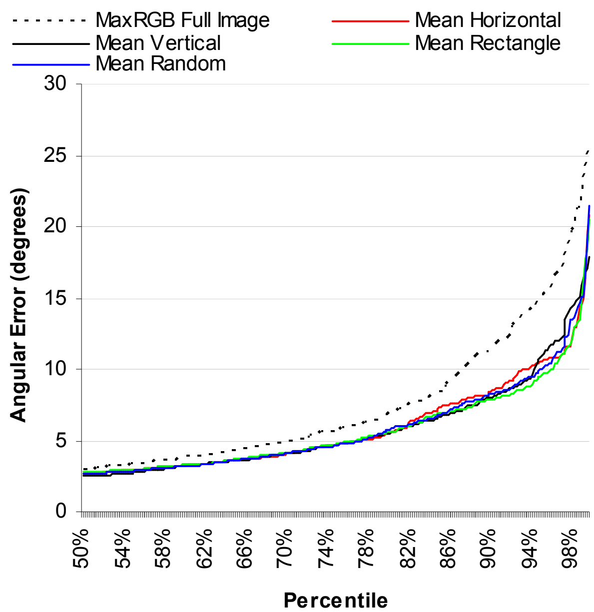

Figure 2 Performance comparison between MaxRGB using the entire image versus the mean of MaxRGB estimates from the various types of sub-windows. The y-axis is the angular error in degrees between the estimated illuminant and the true illuminant. The x-axis is the percentage of images for which the error in the illuminant estimate is less than or equal to the y-axis value

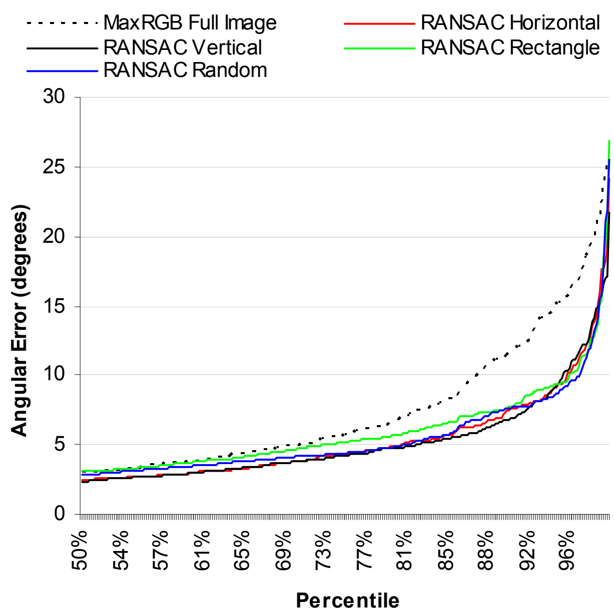

Figure 3 Performance comparison between MaxRGB using the entire image versus the proposed single-illuminant RANSAC model applied to the MaxRGB estimates from the various types of sub-windows. Axis labels as in Figure 2.

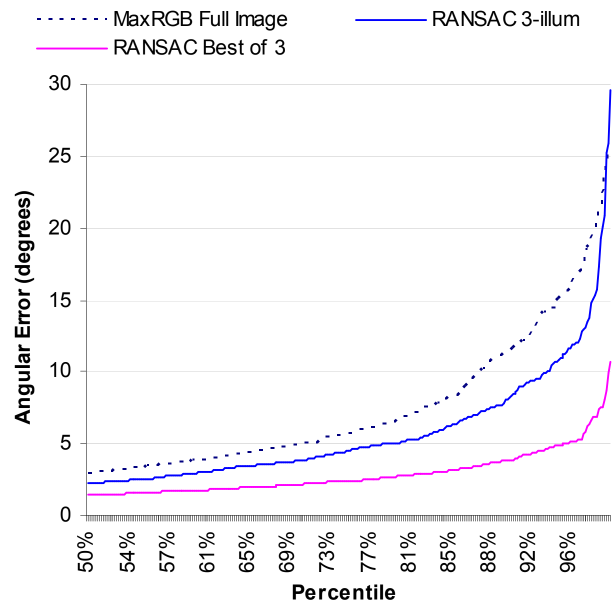

Figure 4 Performance comparison between MaxRGB using the entire image versus the proposed three-illuminant RANSAC model applied to the MaxRGB estimates from the random rectangular sub-windows. For Best of Three description see text. Axis labels as in Figure 2.

Table 1 Performance of the algorithms in terms of median (i.e., 50th percentile), 90th, 98th and maximum angular errors. Root mean square error and mean angular errors are reported as well.

	50th	Mean	RMS	Max	90th	98th
MaxRGB Full image	3.0	4.5	6.5	25.5	11.3	18.9
Mean Horizontal	2.7	3.6	4.8	20.8	8.4	**11.6**
Mean Vertical	2.5	3.5	4.8	**17.9**	8.0	14.0
Mean Rectangle	2.8	3.6	4.7	20.5	7.9	**11.6**
Mean Random	2.6	3.6	4.8	21.5	8.2	12.4
RANSAC Horizontal	2.5	3.4	4.6	24.2	7.4	12.5
RANSAC Vertical	**2.3**	**3.2**	**4.4**	21.7	**6.8**	12.6
RANSAC Rectangle	3.1	3.9	5.0	26.9	7.6	11.7
RANSAC Random	2.8	3.6	4.6	25.6	7.5	11.8
RANSAC 3-illum	**2.3**	3.5	5.1	29.6	8.1	13.8
RANSAC Best of 3	*1.4*	*1.9*	*2.4*	*10.7*	*3.8*	*6.3*

© 2012 Society for Imaging Science and Technology

Table 1 shows the performance of the algorithms in terms of median, 90th, 98th and maximum angular errors. Root mean square error and mean angular errors are reported as well. The outlier-removal strategy outperforms the base MaxRGB algorithm for all reported error measures.

An interesting feature of the three-illuminant RANSAC implementation is its ability to estimate secondary or tertiary illuminants. Figure 5 shows the rectangles fitting the most populous cluster in red, the second most populous cluster in green and the third cluster in blue. In the top image, the red rectangles identify the outdoor illuminant, whereas the green rectangles identify locations illuminated from indoors. The bottom image shows clusters corresponding to direct sunlight (red rectangles) and shade (green rectangles).

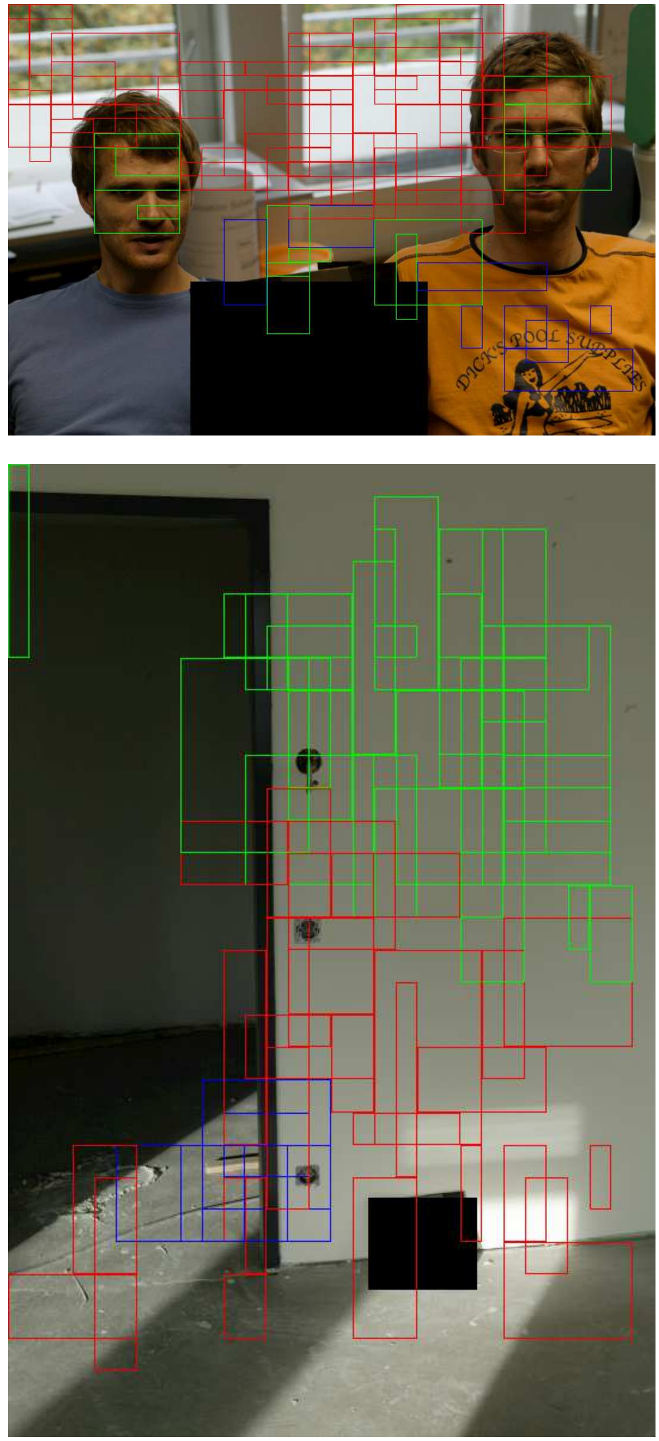

Figure 5. Two sample images processed with three-illuminant RANSAC algorithm. The images are overlaid with rectangles showing the sub-images that are not consider as outliers and are kept. Other sub-images were tested, but automatically excluded by RANSAC. Sub-images belonging to the most populous cluster are shown in red, the second most populous cluster in green, and the third in blue. In the upper image, the red rectangles correspond primarily to the outdoor illuminant seen through the window, and the green ones to the indoor illuminant. In the lower image, the red rectangles correspond primarily to sunlight and the green to shade.

Discussion

The accuracy of illumination estimation methods is hampered by outliers. Outliers can be due to simple noise or to the presence of unexpected secondary illuminants. A method of detecting and removing these outliers based on the RANSAC algorithm was presented and shown to lead to significantly better illumination estimates on the colorchecker dataset of 568 images. The proposed method of removing outliers entails splitting the input image into smaller sub-images, obtaining illumination estimates on each of the sub-images, eliminating those estimates that do not fit the existing illumination model, then combining those that do fit to make a final estimate, or estimates. An advantage of this technique is that even when the illumination model assumes the scene should only contain a single illuminant, it is able to discard the distracting information from any secondary illuminants that happen to be present. On the other hand, when multiple illuminants are expected, the method identifies the image regions corresponding to the different illuminants.

Future work includes testing on other image sets, exploring different methods of selecting sub-images, finding the optimal number of sub-images to use, experimenting with base algorithms other than MaxRGB, and randomly choosing from a set of multiple base algorithms to use on a given sub-image.

Acknowledgement

The financial support of the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged.

References

[1] K. Barnard, V. Cardei, and B.V. Funt. A comparison of computational color constancy algorithms-part i: Methodology and experiments with synthesized data. *IEEE Transactions on Image Processing*, 11(9):972–984, September 2002.

[2] S. Bianco, F. Gasparini, and R. Schettini. Consensus-based framework for illuminant chromaticity estimation. *Journal of Electronic Imaging*, 17(2):023013, 2008.

[3] Vlad C. Cardei and Brian V. Funt. Committee-based color constancy. In *The Seventh Color Imaging Conference: Color Science, Systems, and Applications: Putting It All Together, November 16-19, 1999, Scottsdale, Arizona, USA*, pages 311–313. IS&T - The Society for Imaging Science and Technology, 1999.

[4] Graham D. Finlayson and Elisabetta Trezzi. Shades of gray and colour constancy. In *Color Imaging Conference*, pages 37–41. IS&T - The Society for Imaging Science and Technology, 2004.

[5] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24:381–395, June 1981.

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

[7] Peter V. Gehler, Carsten Rother, Andrew Blake, Thomas P. Minka, and Toby Sharp. Bayesian color constancy revisited. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), 24-26 June 2008, Anchorage, Alaska, USA*, 2008.

[8] A. Gijsenij, T. Gevers, and J. van de Weijer. Computational color constancy: Survey and experiments. *IEEE Transactions on Image Processing*, (in press), 2011.

[9] Arjan Gijsenij and Theo Gevers. Color constancy by local averaging. In *2007 Computational Color Imaging Workshop (CCIW'07), in conjunction with ICIAP'07, Modena, Italy*, pages 1–4, 2007.

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

[11] E.H. Land. The retinex theory of color vision. *Scientific American*, 237:108–129, 1977.

[12] Bing Li, Weihua Xiong, De Xu, and Hong Bao. A supervised combination strategy for illumination chromaticity estimation. *ACM Trans. Appl. Percept.*, 8(1):5:1–5:17, November 2010.

[13] Milan Mosny and Brian Funt. Cubical gamut mapping colour constancy. In *Proceedings of CGIV2010 IS&T Fifth European Conf. on Colour in Graphics, Imaging and Vision, Joensuu*, June 2010.

[14] Lilong Shi and Brian Funt. Re-processed version of the Gehler color constancy dataset of 568 images. http://www.cs.sfu.ca/ colour/data/shi_gehler/. accessed September 2011.

[15] J. van de Weijer, T. Gevers, and A. Gijsenij. Edge-based color constancy. *IEEE Transactions on Image Processing*, 16(9):2207–2214, September 2007.

© 2012 Society for Imaging Science and Technology