Multispectral Colour Constancy

Milan Mosny, Brian Funt; Simon Fraser University; Burnaby, BC, Canada

Abstract

Does extending the number of channels from the 3 RGB sensors of a colour camera to 6 or 9 using a multispectral camera enhance the performance of illumination-estimation algorithms? Experiments are conducted with a variety of colour constancy algorithms (Maloney-Wandell, Chromagenic, Greyworld, Max RGB, and a Maloney-Wandell extension) measuring their performance as a function of the number of sensor channels. Although minor improvements were found with 6 channels, overall the results indicate that multispectral imagery is unlikely to lead to substantially better illumination-estimation performance.

Introduction

Illumination-estimation methods for colour constancy have been generally based on analysing the RGB values from a 3-channel colour image; however, the Chromagenic algorithm [1] is a recent exception in that it uses a total of 6 channels. The Maloney-Wandell [3] is another notable exception. It is defined for an arbitrary number of channels. It might be expected that the more channels, the better, but is this the case? In other words, will using a multispectral camera potentially lead to better illumination estimates than a standard color camera?

We address this question by testing a number of illumination-estimation algorithms on 3-channel, 6-channel and 9-channel multispectral image data, and comparing their accuracy as the number of sensors and their spectral sensitivity functions are varied. Since there are many avenues to explore, we restrict our attention to synthetic multispectral images; however, the synthesis is based on real reflectance and illumination spectra and variations on real camera sensitivities. We test the Greyworld, Max RGB, Chromagenic, and Maloney-Wandell algorithms. In addition, we introduce a new modified version of the Maloney-Wandell algorithm called Subspace Testing that works surprisingly well.

Illumination estimation plays a central role in colour constancy, automatic white balancing, and colour cast removal. For a scene that is illuminated by light of a spectral power distribution that remains constant, except for variations in intensity, throughout the whole scene, the goal of illumination-estimation is to recover the colour (up to a scaling for intensity) of that light as it would be seen by a camera in the camera's RGB colour space when it is reflected off an ideal-white surface. Since intensity is not recovered, it is natural to use chromaticity space for which we use *rgb*-chromaticity defined as *rgb* with $r=R/(R+G+B)$ etc. Once the illuminant chromaticity is known, a simple scaling of the colour channels derived from the *rgb* suffices to 'white balance' the image and remove any colour cast the image may have had.

The paper is organized as follows. First, the error measures for performance evaluation are discussed. Second, the standard algorithms to be tested are described along with the changes required in extending them beyond 3 channels. Third, the Subspace Testing algorithm is introduced. Fourth, the procedures for introducing the additional camera sensitivity functions and for synthesizing multispectral images are described. Fifth, the results are given followed, sixth, by the somewhat surprising conclusion they lead to.

Performance Measures

The angle between the *rgb* of the actual illumination and the estimated *rgb* provides a good measure of the performance of an illumination-estimation algorithm and has been widely used [1][5]. We use it here as well, but in two new ways. The first is simply to extend the definition to N dimensions, where N is the number of channels, and measure the angle between two N-dimensional vectors. The second is to convert N-dimensional illumination estimates into the *rgb* chromaticity space of a given 3-channel camera and then measure the angle in 3-space. We used the sensor sensitivities of Sony DXC-930 for the conversion.

The advantage of converting the results into *rgb* space is that it allows for fairer comparison across results achieved in spaces of different dimensions. Conversion also helps determine whether the additional information that comes from higher dimensions will help achieve better white balance for the purposes of displaying the results on an RGB monitor. The conversion from N-dimensions to 3 is done by identifying which illuminant from a database of known illuminants it is most similar to, and using that illuminant's *rgb* as the conversion value. Associated with each illuminant in the database is its 'colour' in N-dimensions. The N-dimensional colour to be converted is compared to those in the database to find the one with the smallest angular difference.

Since the illuminants in the database will not cover the N-dimensional space continuously some error is introduced through conversion. However, conversion also introduces an additional constraint on the estimated illumination; namely, the conversion constrains the results to fall into the space of known illuminants. Color-by-correlation [2] employs this constraint. We find here that it improves the performance of some other algorithms, Maloney-Wandell in particular.

For a set of test images, the following error measures are calculated and reported in the Tables below.

Median Angular Error in Image Space (Image). This is a median of all angles between the illuminant estimate in N-dimensional space and the actual illuminant in the N-dimensional space.

Median Angular Error in rgb Space (rgb). Some algorithms, for example Maloney-Wandell, can output *rgb* vector of the estimated illuminant directly, without using the lookup in the illuminant database. This measure is a median of all angles between illuminant estimate in *rgb* space (as output directly by the algorithm) and the actual *rgb* vector of the illuminant.

Median Angular Error for Illuminant Estimates Converted to rgb Space (Lookup rgb). This measure is obtained by converting the N-dimensional vectors output by the algorithms into *rgb* space

by the lookup in the illuminant database described above. The median of the angles between converted illuminant estimates (which are now in *rgb* space) and *rgb* vectors of the actual illuminants is computed.

Mean of the Top 0.5% Angular Errors of Converted Illuminant Estimates in rgb Space (Max Lookup rgb). This measure is obtained similarly to the "Lookup *rgb*" measure. The N-dimensional illumination estimates output by the algorithms are converted to *rgb* space by lookup in the illuminant database as described above. The mean of the top 0.5% of the angles between converted illuminant estimates (which are now in *rgb* space) and *rgb* vectors of the actual illuminants is computed. This measure shows the magnitude of the largest errors produced by the algorithms.

Median Error of the Lookup Conversion in the Image Space (Error Lookup) As mentioned above, the illuminant database used for conversion from N-dimensional image space into *rgb* space does not cover the N-dimensional space completely. This measure shows the median of angles between the illumination estimates produced by the algorithms and the closest match in the illumination database.

Algorithm Summaries

We include the **Greyworld** and **Max RGB** algorithms in our experiments since they bridge across results reported by different authors. For an RGB image, the Greyworld algorithm computes mean R, mean G and mean B values. These mean values are assumed to represent the illuminant colour. In N dimensions, the process is the same except carried out over N channels instead of 3. The Max RGB algorithm finds the pixel with highest R value, a potentially different pixel with highest G value and a pixel (which is potentially different from previous two) with the highest B value. These R, G and B values determine the answer. The process for an image with N colour dimensions is the same except using N channels.

We also include results for the '**do nothing'** algorithm, which simply assumes the scene illuminant is always white. The do-nothing angular error provides a measure of the variation in incident illumination and establishes a benchmark with which to compare other algorithms.

The **Chromagenic** algorithm [1] is somewhat more complex than Greyworld and Max RGB. The Chromagenic algorithm requires a 6-channel image, but the 6 channels are related to each other in a specific way. One way to acquire a 6-channel Chromagenic image is to take two registered images of the same scene. The first is taken in the usual way using a standard colour camera with 3 RGB sensors. The second is taken using the same camera, but with a filter placed in front of its lens. Clearly, the second 3 channels are related to the first 3. The Chromagenic algorithm exploits the relationship between them.

The Chromagenic algorithm aims to identify the scene illuminant as one member of a set of known illuminants. It requires a training set of images with numerous images taken under each different illuminant. During the training phase, the Chromagenic algorithm computes a linear mapping from filtered RGB responses to non-filtered RGB responses for each of the illuminants. The mapping is computed as follows [1]:

Let P_j and P_j^F denote Nx3 matrices of RGBs and filtered RGBs for N surfaces measured under the *j*-th of *M* training illuminants. The linear mapping T_j for *j*-th illuminant is computed as

$$T_j = [P_j^F]^+ P_j \quad (1) \tag{1}$$

where + denotes pseudo-inverse.

Given an input image under unknown illumination, the Chromagenic algorithm considers each of the training illuminants in turn as a candidate. The candidate illuminant whose linear mapping best maps filtered RGB responses of the input image to non-filtered RGB responses of the input image is returned as the estimate of the unknown illumination.

The Chromagenic algorithm is extended to N-dimensions by simply dividing the set of channels into two sets. Two sets represent "filtered" and "unfiltered" image. The number of channels in the two sets need not to be the same. However, for Chromagenic algorithm to work successfully, the "filtered" channels may not be completely independent from the "unfiltered" channels. The algorithm then computes and uses a mapping from "filtered" to "unfiltered" channels as described above.

The **Maloney-Wandell** algorithm [4][3] assumes that the subspace of reflectances of all surfaces is linear and of reasonably small dimension. Dimensionality of the reflectances, M, is assumed to be smaller than the number of sensors, N, used to obtain the image. Hence the sensor responses for the surfaces under a single illuminant fall within a linear subspace of the same dimensionality. The illuminant defines the specific subspace. The algorithm proceeds by fitting an M-dimensional hyperplane through the origin to the sensor responses from the N-channel input image. The orientation of the plane determines the illuminant under which the image was taken.

The Maloney-Wandell algorithm produces spectral power distribution of the illuminant (up to a scaling). This is one of the algorithms that can output its illuminant estimate directly in *rgb* space. The spectral power distribution of the illuminant can be converted into N-dimensional images space (by using the spectral characteristics of the N sensors) or directly into *rgb* space (by using spectral characteristics of the RGB sensors).

We also introduce a modification to the Maloney-Wandell algorithm called the **Subspace Testing** algorithm. Subspace testing begins with the Maloney and Wandell's observation that the image colours (we extend the term 'colour' to N dimensions from 3) will lie in a linear subspace of the N-dimensional image input space. The modification is in the way in which the subspace is identified. Firstly, the Subspace Testing algorithm is trained on a set of illuminants with many images under each. For each illuminant, an M-dimensional subspace is fitted over the sensor responses from its entire collection of images. This hyperplane represents the gamut of possible colours (N-dimensional) under that illuminant. To identify an image's illuminant, Subspace Testing—rather than fitting a subspace over the responses present in the image as Maloney-Wandell does—tests the fit of the image colours to each of the training-derived hyperplanes. In particular, the hyperplane for which the root mean square distance between it and the input image's colours is smallest is selected. Since this hyperplane is from the training set, it has an illuminant associated with it. The colour of that illuminant is then output as the estimate of the colour of the image's illuminant.

Data Preparation

All experiments are based on images synthesized using the reflectance spectra, illuminant spectra and spectral sensitivity functions contained in the Simon Fraser University dataset (http://www.sfu.ca/~colour/data)[6]. The dataset contains spectral distributions of 1995 surface reflectances and 287 illuminants. Of the illuminant spectra 87 are used for training. All 287 illuminants are used for testing.

Multispectral images are synthesized for the case of 3, 6 and 9 sensor classes. For the 3-sensor case, we use the sensitivity functions Sony DXC-930 and normalize them to have equal total sensitivity. To extend the sensor set to 6, we add the same Sony DXC-930 RGB sensors shifted by a fixed wavelength towards the red part of the spectrum. To extend the sensor set to 9, we add the same camera sensors twice; one copy shifted towards red and the second towards blue. We experiment with different amount of shifting. Figure 1 shows the spectral sensitivity functions for 9 sensors based on shifts of ±16nm.

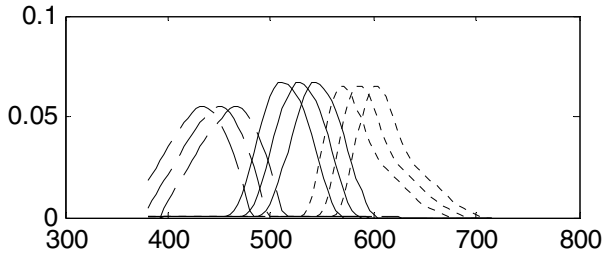

Figure 1. Multispectral sensitivity functions for the 9-sensor case. Sensitivity is plotted as a function of wavelength. The normalized Sony DXC-930 camera RGB sensors are the middle of each group with copies shifted by +16 nm and –16 nm.

The training data set consists of the 87 illuminant spectra and the 1995 reflectance spectra. The various algorithms require the data in different formats, so the data is optimized for each algorithm separately, but they are all restricted to the same set of spectra.

A set of test 'scenes' is constructed as follows. The test 'scenes' contain 2, 4, 8, 16, 32, or 64 surface reflectances. For each illuminant from the set of 287 test illuminants, and for each surface count, 5 different 'scenes' are created. Each scene contains a random selection of the specified number of surface reflectances that were illuminated by the illuminant. This results in 287x6x5=8610 scenes that are then converted to multispectral images based on the 'camera' sensitivities. Some experiments use only test images with a surface count of 16.

Each test image consists of 64 pixels independent of its surface count. Images are normalized so that the largest sensor response across all channels equals 1. This somewhat simulates adjustment of the camera's exposure settings. Gaussian additive noise is applied to the sensor values at each pixel. Unless specified otherwise, the standard deviation of the noise is set to 2%.

Results

The first set of results addresses the question as to how the performance of the Maloney-Wandell algorithm varies as a function of its subspace dimension. This subspace is restricted to being N-1 dimensional or less, but is it optimal to use the highest possible subspace dimension? The results are shown in Table 1 for the case of a 9-channel (based on 36nm shifts) images of scenes with 16 distinct surfaces. Perhaps surprisingly, the algorithm performs the best if the dimensionality of the reflectance subspace is equal to 1. This result holds in the 3-channel case as well where for subspace dimension 1, the error (Lookup *rgb*) is 3.24, while for dimension 2, it becomes 26.3.

Table 1. Maloney-Wandell Algorithm Performance.

Dim	Image	*rgb*	Lookup *rgb*	Lookup Error
1	5.29	4.49	3.15	3.90
2	27.80	27.62	11.05	13.97
3	20.41	20.70	9.22	10.73
4	29.38	27.15	10.32	17.67
5	22.40	21.55	9.44	12.49
6	44.54	40.35	11.58	32.60
7	50.62	43.71	11.83	37.18
8	52.26	48.60	12.30	40.09

The second set of results is from tests with the Subspace Testing algorithm. As with the original Maloney-Wandell algorithm, it also performs best when the dimensionality of the reflectance subspace is chosen to be 1, although there is less volatility in this case. The images were generated with 9 sensors and 16 surfaces per image. The results are given in Table 2.

Table 2. Subspace Testing Algorithm Performance

Dim	Image	Lookup *rgb*
1	3.20	2.70
2	3.87	3.61
3	3.69	3.59
4	4.40	4.05
5	6.26	5.73
6	9.86	8.24
7	13.11	10.85
8	16.58	13.53

The next set of results compares the performance of the various algorithms for 3, 6, and 9 sensors for scenes containing 2, 4, 8, 16, 32 and 64 surfaces. We did some initial experimentation to find the best amount of sensor shift and these are the sensors we then used. The best performance was with a sensor shift of 36nm. "Med" columns in tables 3 to 8 show the results in terms of the median angular error for illuminant estimates converted to *rgb* space (Lookup *rgb*). For each case, "Max" columns show the mean of the top 0.5% angular errors of converted illuminant estimates in rgb space (Max Lookup *rgb*) to provide a measure of how badly a method might fail. All Maloney-Wandell and Subspace Testing results in this table are based on a 1-dimensional subspace since it was shown above to be the best choice. We report results here for the Chromagenic algorithm but it is important to remember that the sensors here do not satisfy the algorithm's assumptions.

Overall, the results in Tables 3 to 8 indicate a slight improvement in performance as the number of channels increases from 3 to 6. A further increase from 6 to 9 sensors exhibits no additional improvement.

Wilcoxon test was performed on the results for 16 surfaces. For each algorithm, the performance with 6 sensors was compared to the performance with 3 sensors and the performance with 9 sensors was compared to performance with 6 sensors. Given a significance level = 0.01, we found that in all cases save for the 'do nothing' algorithm, the performance with 6 sensors was statistically better than the performance with 3 sensors. However, at that significance level, the performance with 9 sensors is not statistically better than the performance with 6 sensors.

This is contrary to what we expected to find, which was a significant and continuing improvement as the number of channels increased. One possible explanation might be simply that all the important information in the input spectra is captured with very few sensors. Principal component analysis (PCA) of the 1995 surface reflectances in our data set finds that 3 dimensions explain 93.7% of the variance, 6 explain 99.2% of the variance, and 9 explain 99.7% of the variance.

Table 3 Algorithm Performance for 2 Surfaces

Dimensions	3		6		9	
	Med	Max	Med	Max	Med	Max
Max RGB	6.42	32.08	5.52	33.57	5.52	32.14
Greyworld	6.43	31.03	5.42	31.82	5.30	30.61
Chromagenic	n/a	n/a	6.45	38.68	6.29	37.29
Subspace Testing	7.48	33.28	6.29	33.58	5.93	33.76
Maloney-Wandell	7.78	34.55	6.45	34.55	6.55	34.01
Do Nothing	17.44	29.00	17.44	29.00	17.44	29.00

Table 4 Algorighm Performance for 4 Surfaces

Dimensions	3		6		9	
	Med	Max	Med	Max	Med	Max
Max RGB	5.45	28.08	4.90	28.33	4.71	28.47
Greyworld	5.53	24.83	4.80	24.06	4.54	24.16
Chromagenic	n/a	n/a	5.74	36.88	5.72	33.99
Subspace Testing	5.80	27.74	5.01	28.60	4.80	27.69
Maloney-Wandell	6.02	29.62	5.18	28.70	5.02	29.34
Do Nothing	17.44	29.00	17.44	29.00	17.44	29.00

Table 5 Algorighm Performance for 8 Surfaces

Dimensions	3		6		9	
	Med	Max	Med	Max	Med	Max
Max RGB	3.95	21.51	3.66	22.56	3.60	23.74
Greyworld	4.59	18.38	4.03	18.55	3.88	18.15
Chromagenic	n/a	n/a	4.45	33.37	4.62	31.63
Subspace Testing	4.13	19.80	3.71	19.80	3.49	19.79
Maloney-Wandell	4.32	20.67	3.78	20.47	3.99	19.80
Do Nothing	17.44	29.00	17.44	29.00	17.44	29.00

Table 6 Algorighm Performance for 16 Surfaces

Dimensions	3		6		9	
	Med	Max	Med	Max	Med	Max
Max RGB	3.29	18.83	2.98	18.43	2.95	18.43
Greyworld	4.04	15.79	3.56	15.54	3.57	16.31
Chromagenic	n/a	n/a	3.80	27.86	3.95	26.38
Subspace Testing	3.11	14.88	2.71	14.86	2.70	14.38
Maloney-Wandell	3.24	15.36	2.88	14.42	3.15	14.59
Do Nothing	17.44	29.00	17.44	29.00	17.44	29.00

Table 7 Algorighm Performance for 32 Surfaces

Dimensions	3		6		9	
	Med	Max	Med	Max	Med	Max
Max RGB	2.53	12.53	2.40	13.44	2.39	13.68
Greyworld	3.68	12.66	3.35	14.20	3.42	14.55
Chromagenic	n/a	n/a	3.27	24.57	3.67	23.29
Subspace Testing	2.22	8.91	2.06	10.67	2.00	10.63
Maloney-Wandell	2.41	8.92	2.10	11.31	2.66	12.36
Do Nothing	17.44	29.00	17.44	29.00	17.44	29.00

Table 8 Algorighm Performance for 64 Surfaces

Dimensions	3		6		9	
	Med	Max	Med	Max	Med	Max
Max RGB	2.24	10.10	2.09	11.02	2.15	10.71
Greyworld	3.60	10.05	3.41	12.23	3.53	12.32
Chromagenic	n/a	n/a	3.04	25.53	3.52	21.28
Subspace Testing	1.85	6.50	1.80	9.66	1.78	9.77
Maloney-Wandell	2.01	7.28	1.86	8.61	2.49	10.66
Do Nothing	17.44	29.00	17.44	29.00	17.44	29.00

To see how the algorithms are affected by noise, we synthesized images with varying amounts of Gaussian additive noise. Tests were based on synthetic 3-, 6- and 9-channel images containing 16 distinct surfaces with noise of increasing standard deviation. Table 9 tabulates the Lookup *rgb* performance measure. As expected, the performance of the algorithms worsens as the standard deviation of the noise increases. Different algorithms seem to exhibit different levels of performance degradation. Max RGB, Greyworld, Maloney-Wandell and Subspace Testing degrade gracefully. The Chromagenic algorithm exhibits a steeper performance decrease.

To test the Chromagenic algorithm, we need an appropriate 6-channel sensor space. In particular the second set of 3 sensors derived from the first by the addition of a filter. We present results for two types of Chromagenic filter. The first type is a straight-line cooling/warming filter defined as follows.

The spectral transmittance of the cooling/warming filter is

$$F_c(\lambda) = 1 - s\,(\lambda - \lambda_{min}) / (\lambda_{max} - \lambda_{min}) \qquad (2)$$

where λ ranges from λ_{min} = 380 nm to λ_{max} = 780 nm. We varied the filter slope s over the set {-1, -0.1, -0.01, 0.01, 0.1, 1}. We found that the best results were obtained with s set to 1.0.Table

10 compares the results for this filter type across all the different algorithms.

Table 9. The effect of noise on algorithm performance

Noise STD	0	0.005	0.01	0.02	0.05	0.1
Dimensions=3						
Max RGB	3.43	3.36	3.36	3.29	3.53	4.96
Greyworld	4.03	4.03	4.01	4.04	3.91	3.88
Subspace Testing	3.16	3.15	3.13	3.11	3.17	3.43
Maloney-Wandell	3.29	3.28	3.27	3.24	3.26	3.53
Do Nothing	17.44	17.44	17.44	17.44	17.44	17.44
Dimensions=6						
Max RGB	3.13	3.03	2.99	2.98	3.15	4.25
Greyworld	3.55	3.55	3.56	3.56	3.49	3.55
Chromagenic	3.69	3.73	3.71	3.80	4.84	7.66
Subspace Testing	2.71	2.71	2.71	2.71	2.82	3.16
Maloney-Wandell	2.83	2.84	2.88	2.88	2.95	3.17
Do Nothing	17.44	17.44	17.44	17.44	17.44	17.44
Dimensions=9						
Max RGB	3.17	3.11	3.07	2.95	3.14	4.35
Greyworld	3.57	3.59	3.57	3.57	3.48	3.36
Chromagenic	3.57	3.59	3.57	3.95	6.63	9.57
Subspace Testing	2.64	2.66	2.67	2.70	2.72	2.84
Maloney-Wandell	2.73	2.82	2.91	3.15	4.21	5.36
Do Nothing	17.44	17.44	17.44	17.44	17.44	17.44

Table 10 Results for Chromagenic Filters

	Aggressive Filter		Cooling Filter		RGB Sensors Only	
Dimension	6		6		3	
	Med	Max	Med	Max	Med	Max
Max RGB	2.97	18.92	3.26	18.44	3.29	18.83
Greyworld	3.69	15.90	3.94	16.07	4.04	15.79
Chromagenic	6.91	41.88	17.72	42.57	n/a	n/a
Subspace Testing	2.93	14.82	3.16	14.91	3.11	14.88
Maloney-Wandell	2.97	15.13	3.39	15.18	3.24	15.36
Do Nothing	16.15	27.62	17.44	29.00	17.44	29.00

We also considered a more aggressive Chromagenic filter that cuts off large portions of the original sensor spectra. The filter has 3 narrow bandpass windows, one for each of the original sensors. Each window has Gaussian shape with standard deviation of 10 nm. The windows are centered 24 nm beyond each of the respective sensor's peaks. The filter spectral characteristics and resulting sensors are shown in figure 2.

Results using this Chromagenic filter are also tabulated in table 10.

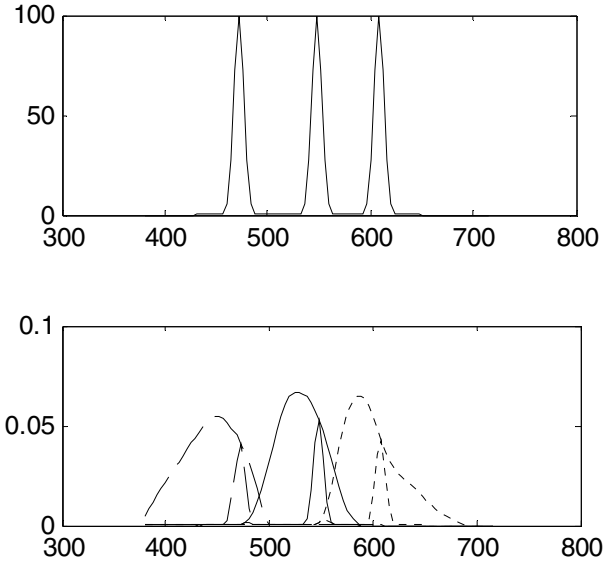

Figure 2. Top: percent spectral transmittance of the aggressive Chromagenic filter. Bottom: resulting 6 sensor sensitivity functions. X-axis is wavelength in nm.

Conclusion

The experiments reported indicate that multispectral imagery is unlikely to be of much benefit to illumination-estimation algorithms for colour constancy and automatic white balancing. This conclusion goes against the intuition that the additional information provided by the additional sensor channels should be useful. Of course, this conclusion is only valid for the algorithms tested and conditions under which they were tested. It is demonstration of what we can expect in general, not a theoretical analysis or proof. In addition to exploring the value of multispectral imagery for illumination estimation, a new Subspace Testing algorithm, an extension of the principles embodied in the Maloney-Wandell algorithm, was introduced that obtains very competitive results.

Acknowledgement

This research was funded by the Natural Sciences and Engineering Research Council of Canada.

References

[1] G.D. Finlayson, P.M. Hubel, S.D. Hordley, Chromagenic Colour Constancy, Proc. AIC, (2005).

[2] G.D. Finlayson, S.D. Hordley, P.M. Hubel, "Color by Correlation: A Simple, Unifying Framework for Color Constancy", IEEE Trans. Pattern Anal. Mach. Intell., 23, 1209 (2001).

[3] B.A. Wandell, "The synthesis and analysis of color images." IEEE Trans. Pattern Anal. Mach. Intell.. 9, 2 (1987).

[4] L.T. Maloney, Brian A. Wandell, "Color constancy: a method for recovering surface spectral reflectance", JOSA A, 3, 29 (1986)

[5] K. Barnard, B. Funt, V. Cardei, "A Comparison of Computational Color Constancy Algorithms, Part One; Theory and Experiments with Synthetic Data",IEEE Trans. Im. Proc., 11, 972 (2002).

[6] K. Barnard, L. Martin, B.V. Funt, A. Coath, "A Data Set for Color Research," Color Research and Application, 27, 148 (2002)