The Chromagenic Colour Camera and Illuminant Estimation

Graham D. Finlayson, Peter Morovic and Steven D. Hordley, School of Computing Sciences University of East Anglia, Norwich, United Kingdom

Abstract

A chromagenic camera takes two pictures of each scene. The first is taken as normal but a specially chosen coloured filter is placed in front of the camera when capturing the second image. The chromagenic filter is chosen so that the combined image makes colour constancy, or white point estimation easier to solve. The chromagenic illuminant estimation algorithm is very simple. We compute the expect relations, currently implemented as 3x3 matrix transforms, between unfiltered and filtered RGBs for a range of typical lights. These relations are tested in situ for a given chromagenic image and the one that best predicts the image data is used to designate the illuminant colour.

However, in experiments we found that a 3x3 matrix transform, while generally quite accurate, can fail to model the relationship between filtered and unfiltered RGBs for some colours (e.g. saturated colours) and so, the chromagenic algorithm which works very well on average can nevertheless, on occasion, work poorly. In this paper we assume that convex combinations in local areas of RGB space are translated to the same convex combinations for corresponding filtered RGBs and use this insight to relate filtered and unfiltered RGBs. These locally convex relations model the image data more accurately. Testing these relations in situ in images and choosing the one which best models the data provides surprisingly effective illuminant estimation algorithm.

Experiments demonstrate that the chromagenic colour constancy algorithm provides superior illuminant estimation compared with conventional approaches (Gamut mapping, color by correlation, max RGB etc). This result holds across many different data sets. The method is also demonstrated to work on real images. The plausibility of the chromagenic approach for human vision is also discussed.

1. Introduction

About 10 years ago there was much interest in the so called multiple illuminant estimation algorithm. If we see the same scene under two or more lights is it easier to estimate the lights and so estimate the surface colours i.e. is it easier to solve for colour constancy? To answer this question D'Zmura and Iverson[3] showed that if we had p measurements per pixel and s surfaces, elights and light and surfaces were described by M and N dimensional linear models then colour constancy could be solved (in many cases) so long as $pse >= sM + eN - 1$. Unfortunately, the DZmura and Iverson approach works very poorly in practice (even when there is many more data points than model parameters to solve for). The reason for the failure is that the solution method they propose is highly non linear and numerically unstable (small changes to the input data can lead to large changes in the estimated model parameters).

On one level this paper is about getting the multiple light approach to work in practice. We begin by applying a well known sleight of hand in order to make the idea of multiple lights seem a more plausible starting premise for illuminant estimation: to a first approximation the image formed by placing a coloured filter in front of the camera is the same as changing the illumination impinging on the scene. So, we take an image, place a filter in front of the camera, take a second image and it is as if we have taken the same scene under two different light colours.

Our first contribution, and is the reason we use the term chromagenic is to recognize that the choice of coloured filter matters.[9] For example, placing a neutral density filter in front of a camera would result in RGBs which were the same as in the original image albeit scaled to be slightly dimmer. As such this filtering cannot add new information into the process. As an alternative one might be tempted to choose a filter that turns the 3D camera into 6D (as far as that is possible) so that we measure more degrees of freedom in the data. This too turns out to be the wrong approach: a spiky filter can lead to independent measurements but this reduces the sensitivity of the camera and increases noise into the systems. Smooth filters cannot increase (at least not by much) the effective capture dimensionality. We instead choose a filter so that for a given light the filtered RGBs are, as close as possible, a linear transform from unfiltered counterparts. Moreover, the filter is also chosen so that the linear transform changes with different lights. A filter which makes the relationship between filtered and unfiltered RGBs depend on and vary with illumination *chromagenic*.[*]

Our second contribution lies in the nature of the illumination estimation algorithm. In the prior two light estimation algorithms [], the image data is used to define the model parameters. As an example, the 6 measurements made by a chromagenic camera will mostly lie on a 3 dimensional plane (in 6 space) and the position of this plane depends on illumination. One solution strategy therefore would be to take our image data, fit the best plane and then look at the position of this plane to determine illuminant colour. Unfortunately, determining the best plane based on the data is problematic. First, the process depends on the data and is sensitive to outliers (e.g. data points which do not lie close to the true plane). Second, the mapping from plane position to illuminant colour is highly non linear: small perturbation in the plane position results in large changes in the estimates. These two problems are the main reason why conventional equation solving approaches to multi light illuminant estimation problem do not work.

In the chromagenic algorithm we compute the relations mapping RGBs to filtered counterparts before hand in a preprocessing step.[8] Not only are these relations based on reliable clean data but they can also be applied in situations where the image is rank deficient (the image responses are 1 or 2-dimensional rather than the required 3). Assuming we wish to discriminate between Nlight colours we compute the N 3x3 linear transforms that best map RGBs to filtered counterparts. Now suppose we have a chromagenic image where the light colour is unknown. To estimate the light, we simply apply each of the precomputed relations and find the one which best models the underlying data. The best relation defines the illuminant estimate. This simple canonical algorithm was shown to perform as well or better than a large set of other algorithms including gamut mapping and colour by correlation.[8]

In this paper we propose to improve the performance of the chromagenic algorithm still further. We begin by considering the relationship between filtered and unfiltered RGBs. While a linear transform works well, we find it can fail for some colours (especially those that are highly saturated). However, experiments indicated that local convex combinations of RGBs tended to map to the same convex combinations in the filtered counterparts. Thus, we propose that the best locally convex mapping from RGB to filtered RGBs be used to drive chromagenic illuminant estimation.

Using the Simon Fraser testing protocol we found that the chromagenic algorithm worked significantly better than all others (including gamut mapping, colour by correlation and the linear transform chromagenic approach). Moreover, we found that it delivered continued good performance when a variety of data sets (outside the protocol) were tested. And, here there was a very large improvement observed when the locally convex mapping model is used. In the final paper, the technique will also be validated on real images using a set of Hyperspectral outdoor images.[14]

In section 2 we present the mathematical background and the basic canonical chromagenic algorithm. Section 3 presents the local convex relation model. Experimental data is reported in section 4.

2. Illuminant Estimation

Let us denote light, surface reflectance and device spectral sensitivity as $E(\lambda)$, $S(\lambda)$ and $\underline{X}_k(\lambda)$ where k indexes R,G,B. For simple, Lambertian surfaces, image formation can be written as:

$$x_k = \int_\omega \underline{X}_i(\lambda)E(\lambda)S(\lambda)d\lambda \tag{1}$$

where the integral is evaluated over ω the visible spectrum. It is useful to combine the triplet of responses x_k into a single vector which we denote x and refer to as the tristimulus values (underscoring denotes a vector quantity throughout the paper).

Now, let us introduce linear models for light and surface:

$$E(\lambda) \approx \sum_{i=1}^{M} \epsilon_i E_i(\lambda) \quad S(\lambda) \approx \sum_{j=1}^{N} \sigma_j S_j(\lambda) \tag{2}$$

and the image formation equations (Eq. (1)) can be written as:

$$\underline{x} = \Lambda(\underline{\epsilon})\underline{\sigma} \tag{3}$$

where $\Lambda(\underline{\varepsilon})$ is a $3\times N$ matrix mapping reflectance weights to RGB responses. The kjth term of this Lighting matrix is equal to:

$$\Lambda(\underline{\epsilon})_{kj} = \int_\omega \underline{X}_k(\lambda)[\sum_{i=1}^{M} \epsilon_i E_i(\lambda)]S_j(\lambda)d\lambda \tag{4}$$

One formulation of the illuminant estimation problem is that given a set of measured responses $\underline{x}$ how can we recover the illumination characteristics i.e. recover $\underline{\varepsilon}$?

The Linear models for light and reflectance, used in (2), are generally determined using Principal Component Analysis[15] or Characteristic Vector Analysis[12] in which case the model dimensions M and N are found to be 3 (for daylights) and 6 to 8 for reflectances. Given that there are only 3 measurements at each point, these large model dimensions cast doubt on the solubility of illuminant estimation. However, looking at (3) we see that image formation is in reality predicated on a (light dependent) Lighting Matrix multiplying a reflectance weight vector. While we have no access to $E(\lambda)$and $S(\lambda)$we see that the linearity of (1) is preserved: if we add two lights together we add the respective lighting matrices. It follows that the dimensionality of light and surface, viewed from the perspective of image formation, depends on how well a set of M, $3 \times N$ Lighting matrices interacting with $N \times 1$weight vectors model tristimuli. By reasoning in this way Marimont and Wandell[13] demonstrated that a very good model of image formation is possible with $M = 3$ (three lighting matrices) and $N = 3$ (three degrees of freedom in reflectance).

This is encouraging because the model numbers are small. However, they are still not small enough to enable us to decouple light and reflectance. To see why, suppose we have a single illuminant and S reflectances, providing us with $3S$tristimuli and $3S + 3$ unknowns. Even by observing that there is a scalar indeterminacy between surface lightness and illuminant brightness so that the unknowns number $3S + 2$, this is still less than the number of known quantities: $3S < 3S + 2$. Suppose now that we observe the S surfaces under two lights. We now have $6S$ measurements and $6S > 3S + 5(5 = 6 - 1 =$ two lights multiplied by 3 minus the brightness indeterminacy). Indeed, a number of authors[3,5,17] have presented algorithms which can algebraically solve the colour constancy problem in this case. However, performance in practice is poor.[3,4]

Rather than capturing a scene under two different lights we might instead assume a second, prefiltered image. We can write the new filtered tristimuli as:

$$x_k^F = \int_\omega \underline{X}_k(\lambda)F(\lambda)E(\lambda)S(\lambda)d\lambda, \quad k = R, G, B \tag{5}$$

We define a filtered illuminant

(6)

and (5) becomes

$$x_k^F = \int_\omega \underline{X}_k(\lambda)E^F(\lambda)S(\lambda)d\lambda, \quad k = R, G, B \tag{7}$$

where throughout the paper the superscript *F* denotes dependence on a coloured filter. From a naive equation counting perspective we now have enough knowns to solve for our unknowns: we simply assume two pictures of every scene: one filtered, one not. Next we investigate how we can use these two images to estimate the scene illuminant. We begin by defining the canonical form for a chroma-genic lighting matrix:

$$\begin{bmatrix} \Lambda \\ \Lambda^F \end{bmatrix} \Rightarrow \begin{bmatrix} \mathcal{I} \\ \Lambda^F \Lambda^{-1} \end{bmatrix} = \begin{bmatrix} \mathcal{I} \\ \mathcal{T} \end{bmatrix} \tag{8}$$

The top 3 rows of this canonical form (the right hand side of (18)) are equal to the Identity matrix and the bottom 3 are equal to the linear transform which takes unfiltered tristimuli to filtered tristimuli. Clearly, all we have done here is post-multiply by Λ^{-1} and so the canonical form must span the same 3 dimensional plane in 6-space. Indeed, all 3dimensional planes (embedded in 6-d) can be written as a linear transform P from their canonical form:

$$\begin{bmatrix} \mathcal{P} \\ \mathcal{T}\mathcal{P} \end{bmatrix} \tag{9}$$

The structure of Eq. (9) provided the inspiration of the chromagenic algorithm. Let us write the responses for *n* surfaces as a 6 × *n* matrix:

$$\begin{bmatrix} \mathcal{X} \\ \mathcal{X}^F \end{bmatrix} = \begin{bmatrix} \Lambda \\ \Lambda^F \end{bmatrix} \Omega \tag{10}$$

where Ω here is 3 × *n* (3 weights for *n* surfaces). From (Eqs. (8) and (9) we expect:

$$\mathcal{T}\mathcal{X} = \mathcal{X}^F \tag{11}$$

The canonical form of chromagenic illuminant estimation works as follows. For a database of mlights $E_i(\lambda)$ and *n* surfaces $S_j(\lambda)$ we precalculate $\mathcal{T}_i \approx X_i^F X_i^+$ where X_i and X_F represent the matrices of unfiltered and filtered tristimuli to the *n* surfaces under the *i*th light and + denotes pseudoinverse.[16] Given Psurfaces in an image we have 3 × *P* matrices *X* and X^F. Then the estimate of the scene illuminant is $E_{est}(\lambda)$ where

$$est = \min_i(err_i) \quad (i = 1, 2, \cdots, m) \tag{12}$$

and

$$err_i = ||\mathcal{T}_i X - \mathcal{X}^F||$$

There is a subtlety in the algorithm which distinguishes it from other multiple light approaches. Specifically, the relationship between unfiltered and filtered responses under each of a set of candidate lights is calculated in a preprocessing step and then each of these relationships are tested to see how well they account for the data. In previous approaches the data itself is used to define the model parameters.[3,12] For example, we could find the best transform based on our data (by solving for $\mathcal{T}$ using least squares) and then see how this compares with precomputed transforms. Unfortunately in this approach the transform is based on the quality of the data (which may be poor).

The failure is easiest to understand in the case where $\mathcal{X}$ is rank deficient. In this case the pseudo inverse is numerically highly unstable and so will be sensitive to small changes in the responses including small amounts of noise. Of course the precomputed relations that we use will still be useful in discriminating between lights. This rank deficient case can occur when an scene has small colour complexity and it is precisely these scenes which conventional algorithms struggle to solve.

It has been shown[8] that this canonical algorithm performs statistically significantly better than other methods including color by correlation and gamut mapping. However, chromagenic illuminant estimation suffers also from outliers that result in large estimation errors. This is due to the fact that while a simple linear transform captures the characteristic of a particular illuminant quite well, it does not do so uniquely.

3. Convex Relation Model

A linear model of illuminant change is not perfect and in fact it is possible for the same linear mapping to correspond to very different illuminants and hence result in large estimation errors. Another way of looking at the relationship between unfiltered and filtered RGBs is by a locally linear mapping approach. More specifically we can express a filtered RGB as a weighted average of it's neighboring RGBs, such that these weights are arrived at from the unfiltered RGBs. In this way we achieve a significantly more accurate mapping that also results in a reduced estimation error.

Suppose we have an RGB in an image. Given a set of training RGBs and their corresponding filtered RGBs, we find the closest three RGBs from this training set to the image RGB. Then we can express this RGB as a weighted linear combination of the three training RGBs, with weights that are nonnegative and less than or equal to one and that sum to one. Such combinations are referred to as convex combinations. Let $\underline{x}$ be an image RGB and $\underline{x}^F$ it's corresponding filtered counterpart. Suppose the training set is denoted as $\underline{y}^i$ and y^F, then we write:

$$\underline{x} = \alpha_1 \underline{y}_i + \alpha_2 \underline{y}_j + \alpha_3 \underline{y}_k$$

where $\underline{y}_i$, $\underline{y}_j$, $\underline{y}_k$ are the three closest RGBs to x. Assuming the training RGBs and image RGBs correspond to the same illuminating light source, also:

$$\underline{x}^F = \alpha_1 \underline{y}_i^F + \alpha_2 \underline{y}_j^F + \alpha_3 \underline{y}_k^F$$

with the same weights α_1, α_2, α_3. However given that we don't know the illuminant, this will not hold but for the correct light source.

Thus the training procedure for this method is simply to store corresponding sets of RGBs and filtered RGBs. Then, given a set of image RGBs and filtered RGBs, each of the image RGBs is expressed as a convex combination of the training RGBs for each training light in turn. The sets of convex weights α_1, α_2, α_3 are then applied to the filtered RGBs corresponding to the same RGBs which were closest to the image RGB. The result is a predicted filtered RGB. The distance between the predicted and the actual filtered RGBs is then a measure the likelihood of the particular illuminant. The smaller the fitting error, the better the representation of the training set.

4. Experiments

We follow the Simon Fraser synthetic colour constancy evaluation.[1] The reflectance set consists of 1995 meausured reflectances which broadly represent most typical surfaces. There are 87 measured illuminants (including Daylights, fluorescents and incandescents) which are used to train the algorithms. Estimated SONY DXC900 sensitivities are used for our camera sensors. To evaluate a colour constancy algorithm there is a set of 287 test lights in the Simon Fraser set. This set comprises the original 87 together with an additional 200 which are convex combinations of these. Thus we expect chromagenic colour constancy to deliver imperfect constancy as it must answer with one of the 87 test lights. We now randomly select 2, 4, 8, 16, 32 and 64 surfaces and generate the 6 corresponding images (one image per surface set) for each of the 287 lights. For each of the 6*287 images we run our algorithm and recover an estimate of the illuminant. Error of recovery is calculated in RGB space: we calculate the RGB for the estimated illuminant, $\hat{\underline{q}}^E$, and compare it with the the RGB for the actual light $\underline{q}^E$. Error is defined to be the angle between these two vectors. This process is repeated for 1000 times.

In Figure 1 we plot number of surfaces against the median angular error for 6 algorithms: Max RGB,[11] greyworld.[2] Linear programming gamut mapping,[6] color by correlaton[7] the canonical chromagenic algorithm and the new convex combination approach. In grey world the rgb for the illuminant is the average image colour. In modified grey world a weighted average is used (to account for biases in the reflectance set). Max RGB returns the maximum R, G and B in an image as the illuminant estimate. Gamut mapping solves for colour constancy by imposing the constraint that the range of colours observable depends on illumination (the reddest red response cannot occur under blue light). Color by correlation is a probabilistic approach which returns the maximum likelihood solution. The latter two approaches deliver relatively much better colour constancy especially for small numbers of surfaces. However, this performance is achieved at the price of relatively higher algorithm complexity and the need for significant calibration.

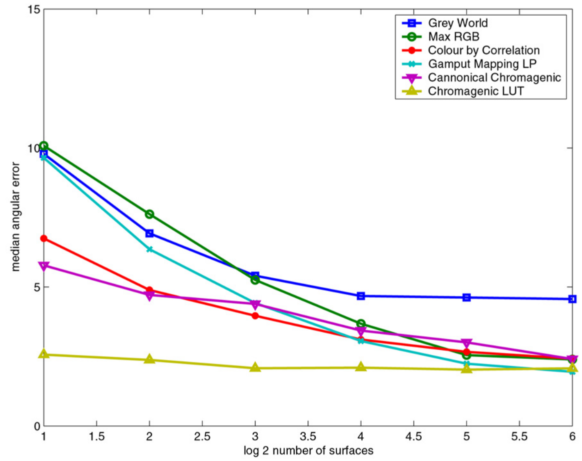

Figure 1. Colour constancy algorithm performance in terms of median angular error as a function of log 2 number of surfaces in a scene under the Simon Fraser training and testing setup.

The above results are remarkable in that regardless of the number of surfaces, the convex combination approach results in identical estimation error at a level to which the best algorithms converge for higher numbers of surfaces. This can be explained because our approach is optimal if training data represents the testing data well, which in the case of the Simon Fraser testing setup is the case.

In a second experiment we kept the training procedure the same as above, training on 87 illuminants and 1995 surfaces, but tested on a set of 99 daylights and 404 natural surface reflectances,[18] neither of which are in the training set. The results are plotted in Figure 2.

Errors are significantly higher here, indeed most algorithms simply fail this test resulting in very high estimation error, reflecting the substantial difference between training and testing data. Even algorithms such as Max RGB or Gray World, not affected by a training procedure, result in significantly higher error than in the above test. However, the good performance of the convex combination chromagenic approach relatively to the others is maintained, clearly outperforming all algorithms.

5. Conclusions

In this paper we presented the chromagenic illuminant estimation algorithm developing it further by introducing the convex combination model of relating RGBs and their filtered counterparts. We have introduced the canonical algorithm in detail and described the principle of the new idea to use a training set of corresponding RGBs and filtered RGBs and express each given image RGB and filtered RGB as a convex combination of the training data. Looking at the goodness of this fit we then deduce which of the possible training lights might have been the illuminating light source in a scene. We have conducted a set of preliminary experiments and have demonstrated that even when all other algorithms fail, this new approach maintains good performance. We will enrich the experimental part in the final paper by providing more analysis and also results from outdoor images using a set of hyperspectral data.

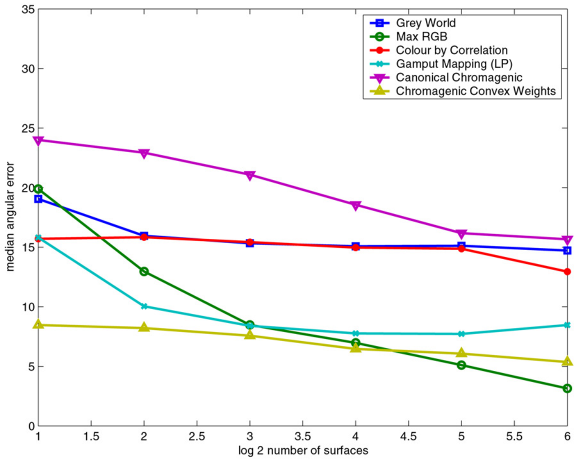

Figure 2. Colour constancy algorithm performance in terms of median angular error as a function of log 2 number of surfaces in a scene, training on Simon Fraser data and testing on 99 daylights and 404 natural reflectances.

Acknowledgements

This research has been supported by the EPSRC grant number GR/R65978/01. Graham Finlayson also gratefully acknowledges the support of the Leverhulme Trust.

References

* This work was inspired by the coloured Chromagen lenses which can be prescribed to colour blind observers which subjectively improve their vision.[10]

1. K. Barnard, V. Cardei, and B. Funt. A comparison of computational color constancy algorithms; Part One: Methodology and experiments with synthesized data. IEEE Transactions in Image Processing, 11(9):972 – 984, 2002.
2. G. Buchsbaum. A spatial processor model for object colour perception. Journal of the Franklin Institute, 310:1–26, 1980.
3. Dzmura and Iverson. Color Constancy I: Basic theory of twostage linear recovery of spectral descriptors of lights and surfaces. Journal of the Optical Society of America, 10:2148 – 2165, 1993.
4. Dzmura and Iverson. Color Constancy II: Results for twostage linear recovery of spectral descriptions for lights and surfaces. Journal of the Optical Society of America, 10:2166 –, 1993.
5. G. Finlayson, M. Drew, and B. Funt. Diagonal transforms suffice for color constancy. In IEEE International Proceedings Fourth International Conference on Computer Vision, Berlin, Germany, pages 164– 171, 1993.
6. G. D. Finlayson and S. D. Hordley. Improving gamut mapping color constancy. IEEE Transactions on Image Processing, 9:1774–1783, 2000.
7. G. D. Finlayson, S. D. Hordley, and P. Hubel. Color by Correlation: A simple, unifying framework of color constancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(11):1209 – 1221, 2001.
8. G. D. Finlayson, S. D. Hordley, and P. Morovic. Chromagenic colour constancy. In Proceedings of AIC Granada (Spain), 2005.
9. G. D. Finlayson, S. D. Hordley, and P. Morovic. Chromagenic filter design. In Proceedings of AIC Granada (Spain), 2005.
10. N. Burnett Hodd. Putting chromagen to the test. Optometry Today, pages 39–42, 1998.
11. E. Land. The retinex theory of color vision. Scientific American, 237:108 – 128, 1970.
12. Lawrence T. Maloney and Brian A. Wandell. Color constancy: a method for recovering surface spectral reflectance. Journal of the Optical Society of America A, 3:29–33, 1986.
13. David H. Marimont and Brian A. Wandell. Linear models of surface and illuminant spectra. Journal of the Optical Society of America, 11:1905–1913, October 1992.
14. S M. C. Nascimento, F. Ferreira, and D. H. Foster. Statistics of spatial coneexcitation ratios in natural scenes. Journal of the Optical Society of America A, 19:1484 – 1490, 2002.
15. J. P. S. Parkkinen, J. Hallikanen, and T. Jaaskelainen. Characteristic spectra of munsell colors. Journal of the Optical Society of America A, 6:318–322, 1989.
16. Gilbert Strang. Linear Algebra and its Applications. Saunders College Publishing, 1988.
17. M. Tsukada and Y. Ohta. An approach to color constancy using multiple images. In International Conference on Computer Vision, pages 385 – 389, 1990.
18. Stephen Westland, Julian Shaw, and Huw Owens. Colour statistics of natural and manmade surfaces. Sensor Review, 20(1):50–55, 2000.