Scene Illuminant Estimation: Past, Present and Future

S. D. Hordley

School of Computing Sciences

University of East Anglia

Norwich, NR4 7TJ

UK

Abstract

This paper addresses the problem of colour constancy: how a visual system is able to ensure that the colours it perceives remain stable, regardless of the prevailing scene illuminant. Our aim is firstly to summarise and review the most important theoretical advances that have been made in this field. Second, we present a comparative analysis of algorithm performance which we use as the basis of a discussion of the current state of colour constancy research and of the important issues which future research in this field should address. Finally, we highlight some areas of recent research which we believe are important in the context of further improving the performance of colour constancy algorithms.

1 Introduction

The phenomenon of colour constancy: how a visual system is able to ensure that the colours it perceives remain stable, regardless of the prevailing illumination, has received considerable attention in the context of both human and computer vision. A visual system might achieve colour constancy by a variety of means and it is useful to classify approaches into colour invariant, or illuminant estimation procedures. In illuminant estimation

procedures (e.g. [47, 16, 53, 45, 56, 30, 2, 14, 25]), colour constancy is achieved by first obtaining an estimate of the illuminant in a scene from the image data recorded for that scene. Once the scene illuminant is known, the recorded image data is corrected to discount the colour of the scene light and thus render the image colour constant. Colour invariant approaches (e.g. [35, 33, 24, 34, 54, 8, 26]) on the other hand, achieve constancy without explicitly estimating the scene illuminant. Instead they derive, by algebraic manipulation of the image data, quantities which are invariant to the colour of the scene illumination. In some cases the algebraic procedure preserves the spatial structure of the image, so that at each pixel we obtain a correlate of the underlying object's reflectance characteristics. In other cases, image structure is lost and the image is summarised by just a few invariant quantities.

In this paper we focus on the illuminant estimation approach to colour constancy and moreover, we concentrate on estimation algorithms which might be employed in the context of a machine vision system or as part of a digital camera's processing pipeline. The extent to which our own visual system is colour constant and the mechanisms by which this constancy might be achieved are not the subject of this paper. However, the algorithms we discuss may be of interest in the context of human colour constancy in so far as they represent possible computational solutions to the problem. Our aims in this paper are threefold. First, we summarise and review the most important theoretical advances that have been made in illuminant estimation. Second, we present a summary of the relative performance of a number of different estimation algorithms. Finally, we discuss the strengths and weaknesses of existing solutions to the problem as well highlighting some shortcomings in algorithm evaluation. And, in the light of this discussion we suggest some directions for future research. In particular we briefly describe some recent research which addresses some of the issues raised in this work.

1.1 A Formal Statement of the Problem

The problem of colour constancy is most simply understood by considering a physical model of image formation. It is convenient to adopt a simple model in which the "colour" of an object is controlled by the interaction of light, surface and sensor. Throughout the paper, and in common with most previous literature, we assume that a scene is illuminated by a single light source with a spectral power distribution (SPD) $E(\lambda)$ which determines how much energy the source emits at each wavelength (λ) of the electromagnetic spectrum. A surface is characterised by its spectral reflectance function $S(\lambda)$ which defines the proportion of light incident upon it that it reflects, again on a per-wavelength basis. The light reflected from a given surface enters an imaging device to produce a colour response ρ_k which is defined as

$$\rho_k = \int E(\lambda)S(\lambda)Q_k(\lambda)\ d\lambda. \tag{1}$$

In Equation (1), $Q_k(\lambda)$ characterises the spectral response function of a given class of light sensor on the imaging plane of the device: it determines the proportion of the colour signal the sensor absorbs on a per-wavelength basis. In most imaging devices there are 3 distinct classes of sensor, so that the response to light at a given pixel is defined by a triplet of responses $\underline{\rho} = (\rho_1, \rho_2, \rho_3)^t$. Throughout the paper we assume a trichromatic imaging system and we will refer to the response of such a device to light from a given scene point either by $\underline{\rho}$ or using the notation RGB, as is common in much of the previous literature.

An inspection of Equation (1) makes the colour constancy problem apparent. The colour response $\underline{\rho}$ of a device to a given surface depends on both the reflectance properties of the surface, and the SPD of the prevailing scene illuminant. So, when illumination changes, so too do the colours recorded in an image. From a computer vision perspective this illumination dependence is problematic since ideally, the colours recorded by a device would tell us something about the intrinsic properties of the imaged surfaces. That this is not true, implies that using colour as a cue to help solve fundamental vision tasks such as scene segmentation, object recognition and tracking might run into problems if the scene illumination is changing. Solving for colour constancy is also an

important step in a digital camera's processing pipeline [48] since it is well established, that if image colours are not corrected to account for the scene illuminant colour, the image will look quite different to how an observer remembers the scene. This is a consequence of the fact that to some degree, our own visual system is itself colour constant. We note in passing, that the conditions under which our own visual system is colour constant (and the degree of constancy achieved) is as yet not clearly defined, and is the subject of much research (see for example [1, 9, 57, 11]).

If we pose the colour constancy problem as the problem of explicitly estimating the illuminant in a given scene, then in a theoretical sense, the problem can be formulated as how we invert Equation (1) to recover the scene illuminant SPD $E(\lambda)$. For a single surface we have only three measurements (the elements of $\underline{\rho}$), and both $E(\lambda)$ and $S(\lambda)$ are unknown. This implies that the problem is, in a strict sense, under-constrained, since both light and surface are continuous functions of wavelength. If we assume that $E(\lambda)$ is constant throughout an imaged scene, adding more surfaces provides us with additional constraints on the illuminant. However, it also introduces further unknowns: the reflectance functions of the additional surfaces. Thus, we cannot in this way equalise the mismatch between knowns and unknowns. Typically, light, surface and sensor are represented by their values at a discrete set of sample points so that the image formation equation becomes:

$$\rho_k = \sum_{j=1}^{n} E(\lambda_j) S(\lambda_j) Q_k(\lambda_j) \Delta\lambda \tag{2}$$

where $\Delta\lambda$ and n denote the fixed sampling interval and the number of sample points respectively. Note that n is typically much larger than 3 so that even in its discrete form, the image formation equation is under-constrained: there are fewer knowns (measurements) than unknowns. Given that the problem is fundamentally under-constrained, all progress towards its solution is founded on enforcing one or more additional constraints.

In many practical applications, an estimate of the SPD of the scene illuminant is not required. Rather, we are interested only in determining the mapping which transforms RGB responses of a device to surfaces viewed under an illuminant o, to the corresponding response under a second light c. That is, we seek the mapping $\mathcal{F}(\cdot)$

such that:

$$\underline{\rho}^c = \mathcal{F}(\underline{\rho}^o). \tag{3}$$

In this case, before we can design a colour constancy algorithm, we must specify the form that the mapping $\mathcal{F}$ takes. Most algorithms which have found practical application adopt (either implicitly or explicitly) a *diagonal model* of illumination change. Here RGB responses $\underline{\rho}^o$ and $\underline{\rho}^c$ of a device to the same surface viewed under two different lights are related by a mapping $\mathcal{F}(\cdot)$ which has the form of a diagonal matrix:

$$\begin{pmatrix} \rho_1^c \\ \rho_2^c \\ \rho_3^c \end{pmatrix} = \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix} \begin{pmatrix} \rho_1^o \\ \rho_2^o \\ \rho_3^o \end{pmatrix} \tag{4}$$

Equation (4) is strictly valid only for restricted classes of lights and sensors [27]. For example, it is exactly true if either lights or sensors are narrow-band (emit power, or are sensitive at only a single wavelength). In general, neither condition is true and for real lights and sensors Equation (4) holds only approximately. For broad-band sensors which have significant overlap, the approximation can be quite poor. It has been shown however, that if we allow the generalisation that sensor responses can be transformed by a fixed 3×3 linear transform prior to application of the model, then for most devices, capturing images under typical viewing illuminants, Equation (4) holds to a good approximation [28]. In this case we have a generalised diagonal model of illumination change:

$$\underline{\rho}^c = T_s^{-1} D T_s \underline{\rho}^o \tag{5}$$

where T_s is the 3×3 linear transform (often referred to as a *sharpening* transform) which is designed in effect, to make the underlying sensor sensitivity functions closer to ideal narrow-band sensors. D is the diagonal transform which accounts for the change of illumination.

In the next section we consider algorithms which, implicitly or explicitly, are founded on a diagonal model of illumination change. That is, they recover (at most) a 3-parameter estimate of the scene illuminant. However, before we consider specific algorithms, it is important to understand that even when we adopt this simplified

model of illumination change, the illuminant estimation problem is under-constrained. Implicitly, this model assumes that light and surface can both be described by 3 parameters. Thus, if we assume an image with m surfaces viewed under a single illuminant we still have fewer knowns ($3m$ pixel values) than unknowns ($3m + 3$) so that further constraints are required to obtain a unique solution. We discuss in the next section the various constraints that have been proposed.

2 Illuminant Estimation Algorithms

Early solutions to the problem [47, 16, 17] focused on attempting to reduce the discrepancy between the number of knowns and unknowns in Equation (1) by adopting linear model representations [46] of lights and surfaces. These approaches set out the theoretical conditions under which the colour constancy problem can be solved uniquely. Unfortunately, for the case of a trichromatic device, the required conditions on lights and surfaces are not satisfied in most typical images and so, this avenue does not lead to practically applicable algorithms. In general, estimation algorithms can be broadly classified as being either *statistical* or *physics-based*. We classify as statistical, those approaches which obtain an estimate of the scene illuminant based on some kind of statistical assumption about the distribution of colours in a scene: for example the *Grey-World* algorithm (see below) assumes that the average of all surface reflectances in an image is grey. By contrast, physics-based approaches are those that obtain an illuminant estimate by exploiting some artefact of the underlying physics of image formation. For example, it is well known that specular reflections often have the colour of the scene illuminant and this observation is exploited in a number of algorithms (also discussed below) to obtain an estimate of the scene illuminant. Of course, all algorithms are, to a certain extent, physics-based, since they are all founded on some kind of model of image formation. Similarly, physics-based approaches can also exploit statistical information. Nevertheless, the classification is useful, at least in as much as it gives a convenient way to order a discussion of previous works.

2.1 Simple statistical methods

If we adopt the diagonal model of illumination change characterised by Equation (4) above, solving the colour constancy problem amounts to estimating the RGB response of a device to a white surface. We call this the *white point* of the scene light and denote it $\underline{\rho}_w$. We denote by $\hat{\underline{\rho}}_w$ an estimate of this quantity. Two simple algorithms which are sometimes used in practice are the *Max RGB* [44] and *Grey-World* [12] methods. In the first case an estimate of the scene illuminant is calculated:

$$\hat{\underline{\rho}}_w = \left[\max_{\{i\}}(\rho_{1,i}), \ \max_{\{i\}}(\rho_{2,i}), \ \max_{\{i\}}(\rho_{3,i}) \right] \tag{6}$$

where $\{i\}$ is a labelling of all pixels in the image. This method will be successful providing that a scene contains either a single surface which is maximally reflective throughout the range of the sensitivity of the imaging device(i.e. a white surface), or a number of surfaces which are maximally reflective throughout the range of each of the three imaging sensors individually.

The *Grey-world* approach is founded on a similar constraint. In this case it is assumed that the average of all surface reflectances in a scene is a neutral (grey) reflectance so that an estimate of the scene illuminant can be found by calculating the mean sensor response in an image:

$$\hat{\underline{\rho}}_w = \left[\frac{1}{n}\sum_i \rho_{1,i}, \ \frac{1}{n}\sum_i \rho_{2,i}, \ \frac{1}{n}\sum_i \rho_{3,i} \right] \tag{7}$$

A variation of this method is the so called *Database Grey-world* [5] method. Here, it is assumed that the scene illuminant white point is the mean of the image RGBs, multiplied by a triplet of independent scale factors:

$$\hat{\underline{\rho}}_w = \left[\frac{\alpha}{n}\sum_i \rho_{1,i}, \ \frac{\beta}{n}\sum_i \rho_{2,i}, \ \frac{\gamma}{n}\sum_i \rho_{3,i} \right] \tag{8}$$

It is assumed in this case, that the scalars α, β and γ are determined based on some prior knowledge about the average of all reflectances which might be observed in the world.

It is easy to think of situations in which all three of these methods will fail. For example both variants of the grey-world method will fail on any image whose colour distribution is significantly different to the assumed

average: e.g. an image of a natural scene consisting largely of different shades of green will probably have an average reflectance quite different to the average calculated over all images. On the other hand, the *Max RGB* method can fail simply because there is no white surface in the scene. Nevertheless, in practice these methods can often give reasonable performance and this, combined with their simplicity, means that they are often used in practice and so are worthy of consideration in the algorithm evaluation which follows in the next section. More successful algorithms can be designed by adopting better founded statistical assumptions about images. There are three broad classes of statistical algorithms which are worthy of study: *Gamut Mapping* methods [30, 20, 40, 2, 29], *Bayesian* methods [18, 10, 50, 25, 7, 55, 49], and *Neural Networks* [14].

2.2 Neural Networks

In the Neural Network solution [14] it is attempted to learn the relationship between observed image data and the scene illuminant by training a network on a large set of images for which the scene illuminant is known. The trained network can be seen as a representation of information about the statistical structure of images. Funt *et al* [14] trained a network using *chromaticity* information rather than image RGBs. Chromaticity co-ordinates provide an intensity independent representation of an image RGB and are calculated thus:

$$\underline{c} = [c_1, \;\; c_2] = \left[\frac{\rho_1}{\rho_1 + \rho_2 + \rho_3}, \;\; \frac{\rho_2}{\rho_1 + \rho_2 + \rho_3} \right] \tag{9}$$

Given a trained network, an estimate of the scene illuminant chromaticity is calculated:

$$\hat{\underline{c}}_w = \text{network}\left(\{ \underline{c}_i \mid i = 1, \ldots m \} \right) \tag{10}$$

where network$(\cdot)$ represents the trained network which takes as input the m image chromaticities. Provided that the information encoded in a network accurately reflects the statistical structure of images that are encountered in practice, then a neural network ought to be a good solution to the problem. Unfortunately, it is difficult to train a neural network such that it generalises well: i.e. so that it is able to accurately predict the scene illuminant in

previously unseen images and for this reason (as is shown in Section 3) neural networks do not provide as good illuminant estimation as we might expect.

2.3 Gamut Mapping

The approaches we have discussed so far deliver a single estimate of the scene illuminant, either by placing additional constraints on images (e.g. requiring the presence of a white surface, or making the assumption that surface reflectances average to grey) or, in the Neural network example, by selecting the "most likely" colour of the scene illuminant based on the image data, where "most likely" is determined with respect to a set of training examples. By contrast, Gamut Mapping solutions (first proposed by Forsyth [30]) place no additional constraints on the problem but rather accept the fact that because the problem is under-constrained, the solution will not in general be unique. That is, there are typically many illuminants consistent with a given set of image data.

Gamut mapping algorithms proceed in two stages. First, the set of all illuminants consistent with a given set of image data is recovered and then a single illuminant is selected from this set as the scene illuminant estimate. In gamut mapping algorithms a diagonal model of illumination change (Equation (4)) is adopted so that an illuminant o with white point $\underline{\rho}_w^o$ can be characterised by a diagonal matrix, D^o that maps sensor responses under illuminant o to their corresponding responses under a reference, or canonical illuminant c. Importantly, it can be shown [30] that the set of all RGB sensor responses observable under this reference illuminant is a closed convex subset of sensor RGB space. This set (denoted Γ) is referred to as the *canonical gamut*.

Given a set of image data: a collection of RGB responses $\mathcal{I}$, imaged under an arbitrary unknown illuminant, the aim is to find plausible scene illuminants. A given illuminant a is said to be plausible if its corresponding diagonal transform D^a maps all the observed sensor responses somewhere within the canonical gamut. That is, if:

$$D^a \underline{\rho^o} \in \Gamma \quad \forall \quad \underline{\rho^o} \in \mathcal{I}. \tag{11}$$

For any given set of image data, there will in general be many matrices D^a that satisfy Equation (11) and Forsyth proposed a method for determining this set of feasible illuminants $\mathcal{D}$. Once this set has been determined, an estimate of the scene illuminant is selected from it according to some selection criterion. For example, Forsyth [30] suggested choosing the diagonal matrix $D \in \mathcal{D}$ whose trace is maximum whilst other authors have suggested using the mean [2] or the median [40] of the feasible set.

The gamut mapping solution is one of the most well founded solutions to the colour constancy problem and can give good practical performance. Its weaknesses are the fact that it is founded on the diagonal model of illumination change: a model which doesn't always hold in practice, and the fact that algorithm implementation is non-trivial and not always robust to real image data. Both of these problems sometimes result in the failure of the algorithm to give any estimate of the scene illuminant. For example if an image does not fully satisfy the diagonal model, the algorithm can sometimes return an empty feasible set of illuminants. Barnard [4] has proposed a method that somewhat ameliorates the problem of diagonal model failure, essentially by increasing the volume of the canonical gamut. Alternatively, the solution can be re-formulated in a convex programming framework [29] which addresses some of the implementation difficulties. However, the lack of robustness to diagonal model failure remains.

2.4 Probabilistic methods

The final class of statistical algorithms we consider are so called Bayesian or Probabilistic algorithms [18, 10, 50, 25, 7, 55, 49]. These approaches share in common the fact that they attempt to capture information about the likelihood of observing a given RGB response under possible scene lights in the form of a statistical prior. Then, given an RGB image whose illuminant is to be classified, they calculate a measure of the likelihood that each possible scene illuminant was the illuminant in the given image. Typically, the illuminant with the highest likelihood is selected as the estimate of the scene illuminant.

The most simple implementation of a Bayesian approach to colour constancy is the so-called *Color By Correlation* method [25]. In this algorithm, prior information about plausible scene lights is encoded by restricting scene illuminants to one of a discrete finite set of possible lights, $\mathcal{I}_c$. Each of these possible scene lights can be assigned an arbitrary probability of being observed, though in practice, all lights are usually assigned an equal probability. Prior information about surfaces is encoded by defining a distribution of image chromaticities for each of the possible scene lights. This chromaticity distribution can either capture information about the gamut of possible surfaces (by declaring that a given chromaticity either is or is not observable under a given light) or it can encode a measure defining how likely it is that a given chromaticity value will occur under each light. In the first case, the algorithm is quite similar to a gamut mapping approach but with the addition of a restriction on the set of possible scene lights. In the second, more general case, we obtain a measure of the likelihood that each of the plausible illuminants $\underline{\rho}_w$ is the scene illuminant:

$$L\left(\underline{\rho}_w\right) \propto \sum_{i=1}^{n} \Pr\left(\underline{\rho}_i \mid \underline{\rho}_w\right) \tag{12}$$

The estimated scene illuminant is the light whose likelihood value is maximum:

$$\hat{\underline{\rho}}_w = \text{argmax}_{\underline{\rho}_x \in \mathcal{I}_c}\left[L(\underline{\rho}_x)\right] \tag{13}$$

All Bayesian approaches operate according to the same underlying principle but they differ both in terms of how they encode prior knowledge about lights and surfaces, and also in terms of their implementation details and complexity. The strength of Bayesian methods is the fact that they incorporate as much prior knowledge as possible into the problem formulation and this knowledge is employed in a statistically well founded way. A potential weakness of the Color By Correlation algorithm is the fact that it assumes that the probability of observing a given RGB in an image is independent of the other observed RGBs. This may not be true in practice and so the algorithm might be improved by accounting for any dependency. Another weakness of Bayesian methods is that, like the neural network approach discussed earlier, their success is limited by the degree to

which real image data conforms to the statistical priors encoded by the algorithms. We leave further discussion of this point to the empirical evaluation of algorithm performance in Section 3.

2.5 Physics-based methods

Physics-based approaches (e.g. [31, 43, 56, 45]) usually adopt a more general model of image formation than Equation (1) and seek to estimate the scene illuminant by exploiting knowledge about the physical interaction between light and surfaces. For example, it is known that for many surfaces, light is not reflected according to the Lambertian model of Equation (1), but rather it is better modelled using the *dichromatic* model. In this model, light reflected from a surface is assumed to be composed of two parts: a body reflection, modelled according to Equation (1), and a specular reflection which is independent of the surface and depends only on the scene illuminant. We can write the sensor response for a general dichromatic surface as:

$$\rho_k = C_b \int E(\lambda)S(\lambda)Q_k(\lambda)\mathrm{d}\lambda + C_s \int E(\lambda)Q_k(\lambda)\mathrm{d}\lambda \tag{14}$$

where C_b and C_s are independent scale factors which model the relative amounts of body and specular reflectance contributing to the overall light reflected from a given point in a scene. A simple approach to estimating the scene illuminant under the dichromatic assumption is to locate pixel values for which $C_b = 0$ since these correspond to points in the scene where the reflected light is purely specular. That is, it has the colour of the prevailing scene illuminant. In practice specular reflections are typically much brighter than body reflections, thus if present in an image, they will be located by the *Max RGB* algorithm defined in Equation (6). Unfortunately, pixel values corresponding to specularities are very often clipped in real images, and moreover, in many images purely specular reflections do not exist, thus such a naïve algorithm will rarely produce good results.

Better founded approaches have been proposed by a number of authors [56, 45, 38], and at their heart is the observation that the sensor responses from different points on a given surface will all fall on a plane in sensor RGB space. That is, for a fixed surface $S(\lambda)$ and illuminant $E(\lambda)$, responses defined according to Equation (14)

are constrained to a two-dimensional subspace (a dichromatic plane) of 3-d RGB space. It follows that two planes, defined by surfaces with different reflectance functions $S(\lambda)$, will intersect in a line determined by the prevailing scene illuminant $\underline{\rho}_w$. So, if V and W are two 3×2 matrices whose columns span the dichromatic planes for two distinct surfaces, then an estimate of the scene illuminant can be determined by solving:

$$V\underline{v} = W\underline{\omega} \tag{15}$$

for $\underline{v}$ and $\underline{\omega}$. So, the estimate of the scene illuminant white point is:

$$\hat{\underline{\rho}}_w = V\underline{v} \tag{16}$$

In theory, Equations (15) and (16) define an elegant solution to the illuminant estimation problem, however, it has proven difficult to make the approach work in practice. One major difficulty with this approach is the fact that there is an implicit assumption that an image can be reliably segmented into regions corresponding to the underlying surfaces, since without such a segmentation, determining the dichromatic planes is problematic. Moreover, even if dichromatic planes can be determined, intersecting them to obtain a reliable estimate of the illuminant white point is often difficult in practice. In general, these and other difficulties have restricted the practical application of physics-based algorithms. More recently however, a robust physics based approach has been proposed [19] which delivers good performance on real images. We discuss this algorithm further in Section 5.

This review of estimation algorithms is by no means exhaustive, it does however, provide a summary of all the major theoretical approaches to the problem that have been set forth in the literature. In the next section we examine the relative performance of these different approaches to estimation.

3 Algorithm Evaluation

The experimental procedure for evaluating algorithm performance is straightforward. We take an image whose scene illuminant is known and use this image data to obtain an estimate of the scene illuminant using the algorithm we wish to evaluate. Then, we compare the recovered scene illuminant estimate to the actual scene illuminant to obtain a measure of accuracy. In practice, the performance of all algorithms is image dependent so that to obtain a meaningful measure of algorithm accuracy we must look at the "average" performance over a representative set of images. It follows that there are three important factors which must be addressed if we are to obtain a reliable measure of algorithm performance. First, we must choose an appropriate set of test images. Second, we must define suitable error measures for comparing an algorithm's estimate of the scene light to the actual scene light. Third, we must consider how best to summarise these errors over the set of test images.

Algorithms are typically evaluated (see e.g. [5, 6, 41]) using both synthetic and real images. Synthetic images are essentially a collection of RGB values rendered according to a model of image formation such as that in Equation (2) and using sets of measured reflectances and illuminants. Testing on synthetic images means that algorithm performance can be evaluated over many thousands of images and makes it easy to investigate algorithm performance as a function of image complexity (i.e. as a function of the number of distinct RGB values in an image). In addition, testing on synthetic images ensures that the statistical information encoded by the various algorithms precisely matches the statistics of the test images. In this sense, synthetic image testing gives a best case measure of algorithm performance. Testing on real images gives a more realistic view of an algorithm's typical performance as well as enabling an evaluation of how robust an algorithm is to noise, and perhaps most importantly, to a mismatch between its training data and the image data it is tested on. However, obtaining real images suitable for algorithm evaluation is non-trivial since it requires that the scene illumination is known and controllable and in addition, it requires access to a well calibrated imaging device or access to raw image data, to which no processing has been applied.

Algorithm performance is most commonly measured using an *angular error*:

$$e_{Ang} = \text{acos}\left(\frac{\underline{\rho}_w{}^t \hat{\underline{\rho}}_w}{\|\underline{\rho}_w\|\|\hat{\underline{\rho}}_w\|}\right) \tag{17}$$

that is, the angle between the actual white point of the scene illuminant $\underline{\rho}_w$ and an algorithm's estimate of that white point $\hat{\underline{\rho}}_w$. This error measure is independent of the intensity of the scene illuminant and can thus be used to evaluate algorithms which estimate only the scene illuminant chromaticity. The accuracy of a given algorithm is typically expressed in terms of an "average" angular error over a set of test images.

Figure 1 summarises the performance of the six statistical algorithms discussed in Section 2 over a set of synthetic images. These results were obtained following the experimental procedure defined by Barnard *et al* [5] in which images are synthesised according to the model of image formation defined by Equation (2) by selecting surface reflectance functions from a set of 1995 measured reflectances and a single scene illuminant from a set of 287 lights (87 measured illuminants together with another 200 illuminants synthesised by interpolating between the measured illuminants). We used a convex programming implementation of the Gamut Mapping algorithm as described in [29]. The Bayesian algorithm is the *Color By Correlation* approach described in [25] and the Neural Network was designed and trained following (as far as possible) the procedure defined in [14]. Where an algorithm required training data, we used the same 1995 surface reflectance functions from which the test images were constructed and 87 illuminants (the 87 measured lights used to synthesis image data).

Figure 1 shows algorithm performance as a function of the number of surfaces in an image. Performance is measured over 6000 different images with between 2 and 64 surfaces and two different summary statistics are used to summarise performance: Root Mean Square (RMS) (left-hand plot) and median angular error (right-hand plot) . This figure highlights a number of important points about algorithm performance. First, as might be expected, all algorithms (except for Grey-World) tend towards very good performance as we increase the number of surfaces in an image. However, algorithm performance is on average quite poor for scenes with low colour complexity. Second, the results support the theory that incorporating more (and more realistic) statistical

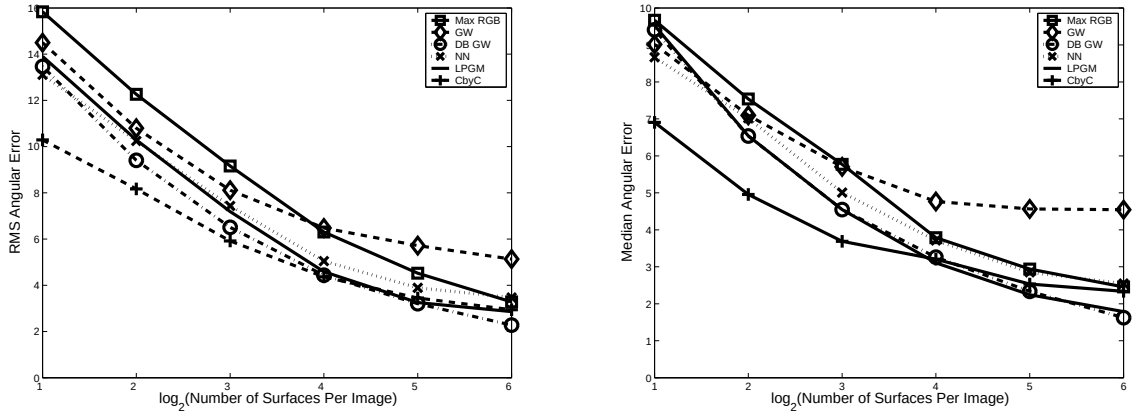

Figure 1: Root Mean Square (left) and Median (right) angular error as a function of the number of surfaces in an image.

information about image content leads to better algorithm performance. So, for example, the Bayesian algorithm gives the most accurate performance, with the Neural Network and Gamut Mapping methods also performing reasonably well. It is also noticeable that the Bayesian approach significantly out performs all other methods for images with few surfaces. Finally, the figures highlight an important point about algorithm evaluation: the relative performance of algorithms is dependent on the choice of summary statistic. While the overall trends are similar regardless of whether we judge algorithms according to RMS or median angular error, looking at these two error measures for a fixed number of surfaces per image leads to quite different conclusions about the relative performance of algorithms.

This point is clearly highlighted in Table 1 where algorithms are ranked according to their average performance over all 6000 test images using a variety of different summary statistics. Each of the summary statistics leads to different conclusions about the relative performance of algorithms. It is shown in [41] that the error distributions of colour constancy algorithms tend to be strongly skewed so that measures such as the mean and RMS error are less appropriate than the median statistic when comparing relative algorithm performance. That work also makes the point that it is important to consider the statistical significance of the differences in errors for each

Table 1: Rankings over all test images, regardless of the number of surfaces, (a total of 6000 images) according to a variety of different summary statistics.

	RMS	Median	Mean	Sign Test
MxRGB	6	5	5	5
GW	5	6	6	6
DB GW	2	3	2	2
LP GM	4	2	3	2
CbyC	1	1	1	1
NN	3	4	4	4

algorithm. Again, care must be taken to choose the appropriate statistical test and, given that the underlying error distribution cannot be well approximated by (for example) a Normal distribution, non-parametric tests such as the Sign Test [39] are the most reliable. In essence, the Sign Test tests the null hypothesis that the median of the difference in errors for two algorithms is zero against the alternative hypothesis the difference is non-zero. The last column of Table 1 ranks the six algorithms according to this test. We note that a potential weakness of the error analysis we have presented here (in common with other analyses presented elsewhere [41, 5]) is the fact that the "cost" of an error is considered to be directly proportional to its magnitude which might not be appropriate in some applications. For example in digital photography it is likely that an illuminant estimate which has an accuracy less than some threshold value is "good enough" so that the actual magnitude of an error within this threshold is unimportant. Similarly, once an error is sufficiently high that it is unacceptable for a given application, the actual magnitude of the error becomes unimportant. However, if we consider algorithm errors independently of a particular application, it is difficult to determine error costs thus, we assume that cost is proportional to error magnitude.

Table 2: RMS, median, and mean angular error together with Algorithm Ranking according to the Sign Test over the 321 real images.

	RMS Error	Median Error	Mean Error	Sign Test Rank
MxRGB	8.88	4.05	6.38	3
GW	14.52	8.94	11.48	6
DB GW	12.44	6.85	9.44	4
LP GM	6.85	3.71	5.00	1
CbyC	10.09	3.19	6.56	1
NN	11.04	7.78	9.18	4

Real image performance of algorithms is most usefully evaluated using the set of 321 real images gathered at Simon Fraser University [6]. Thirty different scenes were captured under up to 11 different lights using a well calibrated digital camera. The performance of the same six algorithms on this test set is summarised in Table 2. For the most part the trend of the real image results follows that for the synthetic images: algorithm performance is correlated with the level of statistical information incorporated into each method. There are some notable exceptions however. First, the Bayesian approach is no longer the best performing algorithm: judged using the Sign Test, its performance is equivalent to Gamut Mapping. Second, the *Max-RGB* algorithm performs better on real images than it does in synthetic image tests. Finally, there is again variation in the ranking of algorithm performance according to the choice of summary statistic. Indeed, this variation is more pronounced for the real images. For example, judged according to RMS error the Bayesian approach is ranked only third, while when the median error is used, it is the best performing algorithm. This suggests that while the algorithm performs well on average, there are some images in the test set for which it performs quite poorly. An inspection of the errors on an image-by-image basis reveals that there are indeed some images for which error is quite large however, there is

no obvious characteristic shared by such images. We further discuss the implication of these experimental results in the next section.

4 Discussion

The synthetic image results presented above provide useful information as to the best case performance of the tested algorithms and they suggest, as would be expected, that best performance is achieved by incorporating as much prior knowledge as possible of the statistical distribution of lights and surfaces in the world. The real image results follow a similar trend though performance of some algorithms (notably the Neural Network and Bayesian approaches) is somewhat degraded. However, the experimental evaluation suffers from a number of shortcomings. One important issue is the fact that while the results allow us to judge the relative performance of the tested algorithms, they provide little information as to the absolute performance of any given algorithm and thus it is difficult to determine whether or not the performance of any algorithm is "good enough". In fact this problem is difficult to answer since "good enough" colour constancy depends on the context in which the illuminant corrected images are to be applied. For example, in computer vision, illuminant estimation can be considered sufficiently accurate provided that subsequent visual tasks such as object recognition or tracking can be performed effectively. Unfortunately, there is no simple relationship between the ability to successfully perform these tasks and a particular value of illuminant estimation error.

A number of studies [32, 36, 2] have been conducted to investigate whether or not colour constancy algorithms are sufficiently accurate to enable colour based object recognition. The general conclusion from these studies is that the answer to this question is no. However, these studies, as well as the real image experiments reported in the previous section, are weak in the sense that they are based on quite small image sets. For example, in the real image experiments reported above, algorithms are tested on just 30 different scenes. Moreover, it is unclear whether the scenes are representative of the type of scenes which will be encountered in practical

applications: all the scenes consist of a small number of objects captured in an indoor environment. Thus, there are no images of people, no landscape scenes etc. As a result, it is difficult to say whether the test images represent an easy or difficult test for the algorithms. It has also been shown [6] that algorithm performance can be significantly changed by varying the processing applied to an image prior to estimating the illuminant and also, by changing the training regimes of the different algorithms. Another study [3] has tested illuminant estimation algorithms, in the context of an object recognition task, on a much larger set of images. However, in this case only two algorithms are tested (Grey-world and Max-RGB) and while illuminant estimation is shown to improve recognition performance, it is difficult to determine whether the level of performance achieved is "good enough".

Another shortcoming of the algorithm evaluation we have presented is the fact that the angular error measure we have used is a non-intuitive measure. That is, it is not immediately obvious, whether an angular error of 3 degrees represents good or bad illuminant estimation. In Figure 2, we provide an informal visualisation of different angular errors. To synthesise the images in this figure, we began by assuming that the true illuminant white point is the RGB triplet $[1, 1, 1]$. Next we generated new RGB triplets which are on average k degrees from the true illuminant white point. We then formed an image of square patches, such that each patch is assigned one of the generated RGB values. Figure 2 shows images for values of $k = 9, 7, 5, 3$ and 2. It is clear from this figure that an angular error of nine degrees represents quite poor illuminant estimation performance while an angular error of 2 degrees might be considered good enough. We also calculated the average $CIE - Lab$ [42] error between our assumed actual white (the RGB triplet $[1, 1, 1]$) and the estimated RGB triplets with a given angular error. $CIE - Lab$ error provides an approximately perceptually uniform measure of the difference between two colours. The average LAB errors we calculate for the case of $9, 7, 5, 3$ and 2 degrees angular error are $20.5, 15.9, 11.3, 6.8$ and 4.5. An LAB error of 1 corresponds to a just noticeable difference between two colours when they are viewed side by side in isolation. However, it has been suggested that $CIE - Lab$ errors of up to 6 are acceptable in complex images. This lends further support to the suggestion that an angular error

of 2 degrees represents “good enough” colour constancy performance. However, this informal analysis is by no means conclusive. Referring to Table 2, we see that the best performing algorithm has a median angular error of 3.17 degrees. So that, accepting the analysis above, even the best performing algorithm will produce significant visual errors for at least 50% of images.

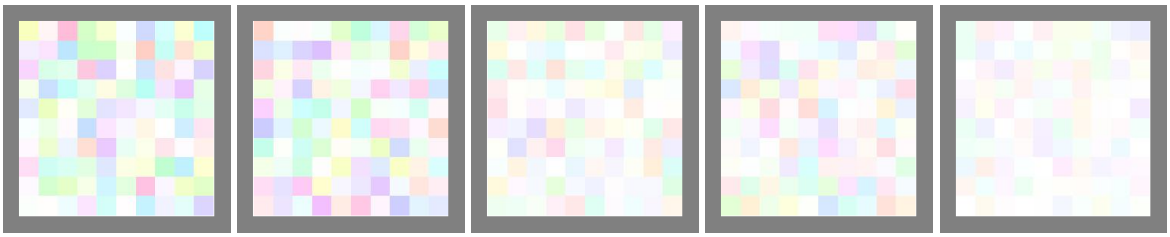

Figure 2: Patchwork images such that image patches are on average k degrees from an RGB of $[1, 1, 1]$. From left to right $k = 9, 7, 5, 3$ and 2 respectively.

The visualisation in Figure 2 provides some useful intuition as to the meaning of a given level of angular error, but the relationship between angular error and perceived colour accuracy is of course much more complex than this. For example. the perceptibility of a given angular error will depend on the particular colour being observed (RGB space is not perceptually uniform). Currently, the only reliable way to determine whether colour constancy algorithms afford good enough estimation accuracy to facilitate pleasing digital photographs is to use psycho-physical techniques.

In summary, we conclude that further testing on larger data sets is required to properly determine the absolute level of performance of existing algorithms and we should also bear in mind that “good enough” colour constancy depends on the purpose for which the illuminant corrected images will subsequently be used. It is important to understand that algorithms such as Gamut Mapping, Bayesian and Neural Network approaches do a good job, at least in a theoretical sense, of exploiting the statistical information available in images. Thus it is unlikely that further progress towards a solution will be made using statistical methods. This is not to say that further theoretical advances are not possible, and in the next section we discuss some possible directions in which such

advances might occur.

5 Future Research Directions

We briefly review two recent theoretical advances in algorithm development. The first is the so called *Chromagenic Colour Constancy* algorithm which obtains additional information about the scene illuminant by making additional measurements at each scene point. The second approach combines illuminant estimates from two different algorithms to produce a new, more robust estimate.

5.1 Chromagenic Colour Constancy

The so-called *Chromagenic*[*] *Camera* [21, 23] adopts a conventional RGB imaging paradigm except that two images of each scene are captured. The first is just a conventional RGB image, whilst the second is an RGB image, optically pre-filtered using a special chromagenic filter. An estimate of the scene illuminant is found by exploiting the fact that filtered and unfiltered image RGBs are related in a well defined way, which depends on the (fixed) known chromagenic filter and also on the scene illuminant which it is wished to determine. More specifically, it is assumed that filtered and unfiltered RGBs are related by an (illuminant dependent) 3×3 linear transform. An estimate of the scene illuminant is obtained by defining *a priori* a set of possible scene illuminants each of which is characterised by the transform mapping filtered RGBs under that light to corresponding unfiltered RGBs. Given a pair of chromagenic images whose scene illuminant it is wished to estimate, each illuminant transform is tried in turn and the illuminant whose transform best maps between the image pair is chosen as the scene illuminant. In general, there are two stages to the chromagenic algorithm: a pre-processing stage, performed once for a given chromagenic device, and an operation stage applied once for each image.

[*]The word chromagenic comes from the term *chromagen* a proprietary term for coloured contact lenses used in the treatment of colour deficient observers [37].

Pre-processing

We choose *a priori*, a set of N plausible scene illuminants $E_i(\lambda)$. In addition we select a set of M surface reflectances $S_j(\lambda)$ representative of the surfaces which occur in the world. Now, for the ith illuminant we define an $M \times 3$ matrix $\mathcal{Q}_i$ whose jth row contains (R_{ij}, G_{ij}, B_{ij}): the sensor response to the jth surface imaged under the ith illuminant. Similarly we define $\mathcal{Q}_i^F$, also an $M \times 3$ matrix whose jth row contains $(R_{ij}^F, G_{ij}^F, B_{ij}^F)$, the sensor response to the jth surface imaged under the ith illuminant and filtered by a chromagenic filter $F(\lambda)$. Then for each plausible illuminant we define a 3×3 transform matrix:

$$\mathcal{T}_i = \mathcal{Q}_i^F \mathcal{Q}_i^+ \tag{18}$$

where $+$ denotes pseudo-inverse. $\mathcal{T}_i$ is the transform which best maps unfiltered sensor responses to filtered sensor responses, in a least-squares sense. The set $\{\mathcal{T}_i\}$ characterises the possible scene illuminants. Note that it is not necessarily the case that $\mathcal{T}_i$ for a given illuminant is a perfect mapping from unfiltered to filtered responses. Indeed, in practice it is rarely the case. However, this is not important provided that the relationship is approximately true. What is more important is that the transform $\mathcal{T}_i$ for a given illuminant is a better mapping than is obtained using another transform $\mathcal{T}_j$ for an arbitrary illuminant j. Of course, if a linear transform is a very poor model of the change in RGB introduced by a filter, then the algorithm is likely to perform badly.

Operation

Given a chromagenic image pair, each consisting of p pixels, let $\mathcal{Q}$ and $\mathcal{Q}^F$ denote the $p \times 3$ matrices of unfiltered and filtered sensor responses respectively. For each plausible scene illuminant we calculate a fitting error $e_i = \|\mathcal{T}_i\mathcal{Q}_i - \mathcal{Q}_i^F\|$ under the assumption that $E_i(\lambda)$ is the scene illuminant. Then our estimate of the scene illuminant is $E_{est}(\lambda)$ where

$$est = \min_i(err_i) \;\; (i = 1, 2, \cdots, m). \tag{19}$$

To be successful the algorithm requires that the transforms $\mathcal{T}_i$ for each illuminant be different from one another and the degree to which this is the case is controlled in part by the choice of chromagenic filter. Good illuminant estimation performance can be obtained using a range of common photographic transmittance filters [23]. For example Figure 3 compares the performance of the chromagenic algorithm using a standard transmittance filter, to that of the Bayesian algorithm described in Section 2. The figure shows the median angular error performance of the two methods as a function of the number of surfaces per image based on the synthetic image experiment described in Section 3.

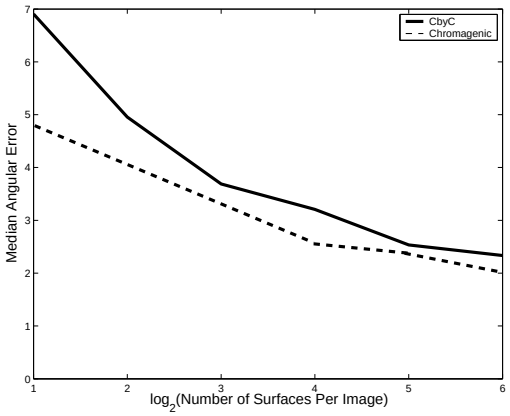

Figure 3: Median angular error as a function of the number of surfaces per image for the Bayesian algorithm (solid line) and the Chromagenic algorithm (dashed line).

It is clear that the algorithm out performs the Bayesian method (the best performing algorithm of those tested in Section 3) and the improvement is particularly pronounced for images with just a few surfaces. A further significant improvement in performance can be obtained by careful filter design [22]. These results suggest that the chromagenic camera is a promising area of future research: in the short-term, testing of the algorithm on real image data is required to investigate algorithm robustness to artefacts such as image noise. In addition, the major drawback of the Chromagenic approach is the fact that it requires a significant change in camera design.

Finally, it is worth noting that the method has some similarity to the work reported in [15] where an estimate of

the illuminant in a scene is also determined by considering two images of the scene. Here, the two images differ in the fact that one of the pair is captured in a conventional manner using only the ambient scene illumination while the second image is captured under a combination of scene illumination and camera flash. Since the flash is a known illuminant it is possible to derive a scheme to discover the scene illuminant given the image pair. While this method is similar to the chromagenic approach in the sense that it is based on a pair of images differently illuminated, it is important to understand that the illumination change exploited in the two algorithms is quite different. In the work of DiCarlo *et al* the image captured using flash is illuminated by the sum of two distinct illuminants: the flash and the ambient scene illuminant. In the chromagenic case, the filtered scene is effectively captured by a filtered version of the scene illuminant (a multiplicative change to the scene illuminant spectrum). This implies that the two methods are fundamentally different in terms of how they estimate the scene illuminant given an image pair. In addition, the flash based approach is somewhat limited in terms of the scenes to which it can be applied since it relies on scene objects reflecting the camera flash illumination back to the camera. This will not be the case in, for example, landscape images.

5.2 Combining Illuminant Estimates

An alternative approach to improving estimation accuracy is to combine illuminant estimates from different algorithms to produce a new, hopefully more robust estimate. One example of such an approach is the work of Funt *et al* [13] where estimates from the Max-RGB, Grey-World and Neural Network algorithms are combined in a variety of ways, leading to improved estimation performance. Clearly, such an approach will only lead to improved performance if the errors for individual algorithms are uncorrelated. That is, if two algorithms perform equally well or badly on a given image, combining them is unlikely to improve performance. One promising approach which has recently been proposed [52] is to combine the estimates of a statistical algorithm and a physics-based method. In this case, the two algorithms exploit quite different scene information so that there

is a good chance that their errors are uncorrelated. The statistical algorithm employed in this work is the Color By Correlation approach, described in Section 2, while the physics-based method estimates the scene illuminant using image specularities. Full details of the physics-based approach can be found in [51] but in essence the method is founded on the idea of intersecting dichromatic planes discussed in Section 2. The algorithm first identifies possible dichromatic planes based on a naïve segmentation of the image data. Then, each pair of dichromatic planes is intersected producing a set of illuminant estimates. Each of these illuminant estimates then "votes" for one of a set of possible candidate scene illuminants which are defined in advance. An illuminant estimate "votes" for the candidate light it is closest to. This results in a likelihood measure for each of the candidate lights which we denote $L_{Spec}(\underline{\rho}_i)$. Similarly, the statistical algorithm provides a likelihood measure $L_{Stat}(\underline{\rho}_i)$ for each candidate light. These likelihood measures are then combined to produce a single measure of likelihood for each light:

$$L(\underline{\rho}_i) = \alpha L_{Stat}(\underline{\rho}_i) + (1-\alpha) L_{Spec}(\underline{\rho}_i) \tag{20}$$

where α is a weighting parameter whose value is determined empirically. The candidate light with maximum likelihood is then chosen as the estimate of the scene illuminant.

Table 3: Median and RMS error over the set of 321 real images for the statistical, physics-based and combined algorithms.

	RMS Error	Median Error
Statistical	10.09	3.19
Physics-Based	9.70	3.00
Combined Algorithm	5.20	2.20

Table 3 summarises the performance of the statistical, physics-based and combined algorithms on the set of 321 real images used in the experiments in Section 3. Individually the statistical and physics-based approaches

provide a similar level of performance but by combining them a significant improvement is obtained. Most notably, the number of images with very high illuminant error is reduced (as evidenced by the fact that the RMS error is significantly smaller for the combined algorithm) and the median performance of the combined approach is close to the angular error value of 2 degrees which, the visualisation presented in the previous section suggests may be representative of "good enough" colour constancy performance. However, these results can only be considered preliminary, since as was discussed in the previous section, the set of test images is quite restricted. In addition, future work in this area ought to consider alternative ways of combining the estimates of the different algorithms, as well as investigating the effect of combining evidence from other statistical and physics-based algorithms.

6 Conclusions

We have presented in this paper a review of the most important theoretical contributions towards a solution of the colour constancy problem. An empirical evaluation of these algorithms showed that, while some of these algorithms are often capable of providing good colour constancy performance, the available evidence suggests that no algorithm can be considered to afford "good enough" performance in all cases. Thus we conclude that further work is required, both in terms of algorithm evaluation and training on real image data, and in terms of delivering further theoretical insights into the problem. We highlighted two possible directions in which these insights might be made: first, by making additional measurements at the time of image capture, and second by combining estimates from two or more different algorithms.

Acknowledgements

The author wishes to thank the EPSRC for their support of this work under grant number GR/R65978/01.

References

[1] Lawrence Arend and Adam Reeves. Simultaneous color constancy. *Journal of the Optical Society of America, A*, 3(10):1743–1751, 1986.

[2] K. Barnard. *Practical Colour Constancy*. PhD thesis, Simon Fraser Univ., School of Computing Science, 2000.

[3] K. Barnard and P. Gabbur. Color and color constancy in a translation model for object recognition. In *Proceedings of the Eleventh Color Imaging Conference*, pages 364–368. IS&T/SID, 2003.

[4] Kobus Barnard. Improvements to gamut mapping colour constancy algorithms. In *6th European Conference on Computer Vision*, pages 390–402. Springer, June 2000.

[5] Kobus Barnard, Vlad Cardei, and Brian Funt. A comparison of computational color constancy algorithms; part one: Methodology and experiments with synthetic images. *IEEE Transactions on Image Processing*, 11(9):972–984, 2002.

[6] Kobus Barnard, Lindsay Martin, Adam Coath, and Brian Funt. A comparison of computational color constancy algorithms; part two: Experiments with image data. *IEEE Transactions on Image Processing*, 11(9):985–996, 2002.

[7] Kobus Barnard, Lindsay Martin, and Brian Funt. Colour by correlation in a three dimensional colour space. In *6th European Conference on Computer Vision*, pages 275–289. Springer, June 2000.

[8] Daniel Berwick and Sang Wook Lee. A chromaticity space for Specularity, Illumination color- and illumination pose-invariant 3-D object recognition. In *Sixth International Conference on Computer Vision*, pages 165–170. Narosa Publishing House, 1998.

[9] David H. Brainard, Wendy A. Brunt, and Jon M. Speigle. Color constancy in the nearly natural image. I. Asymmetric matches. *Journal of the Optical Society of America, A*, 14(9):2091–2110, 1997.

[10] David H. Brainard and William T. Freeman. Bayesian Method for Recovering Surface and Illuminant Properties from Photosensor Responses. In *Proceedings of the IS&T/SPIE Symposium on Electronic Imaging Science & Technology*, volume 2179, pages 364–376, 1994.

[11] David Bramwell. *Colour Constancy in simple and complex scenes*. PhD thesis, University of Newcastle, 1997.

[12] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310:1–26, 1980.

[13] V. C. Cardei and B. V. Funt. Committee-based color constancy. In *Proceedings of the IS&T/SID Seventh Color Imaging Conference: Color Science, Systems and Applications*, pages 311–313, 1999.

[14] Vlad C. Cardei, Brian Funt, and Kobus Barnard. Estimating the scene illuminant chromaticity by using a neural network. *Journal of the Optical Society of America, A*, 19(12):2374–2386, 2002.

[15] Jeffrey M. DiCarlo, Feng Xiao, and Brian A. Wandell. Illuminating illumination. In *Proceedings of the Ninth Color Imaging Conference*, pages 27–34. IS&T/SID, November 2001.

[16] M. M. D'Zmura and G. Iverson. Color constancy. I. Basic theory of two-stage linear recovery of spectral descriptions for lights and surfaces. *Journal of the Optical Society of America, A*, 10(10):2148–2165, 1993.

[17] M. M. D'Zmura and G. Iverson. Color constancy. II. Results for two-stage linear recovery of spectral descriptions for lights and surfaces. *Journal of the Optical Society of America, A*, 10(10):2166–2179, 1993.

[18] M. M. D'Zmura and G. Iverson. Probabilistic Color Constancy. In M. M. D'Zmura, D. Hoffman, G. Iverson, and K. Romney, editors, *Geometric Representations of Perceptual Phenomena: Papers in Honor of Tarow Indow's 70th Birthday*, pages 187–202. Laurence Erlbaum Associates, 1994.

[19] G. Finlayson and G. Schaefer. Solving for colour constancy using a constrained dichromatic reflection model. *International Journal of Computer Vision*, 42(3):127–142, 2001.

[20] G. D. Finlayson. Color in Perspective. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10):1034–1038, 1996.

[21] G. D. Finlayson, S. D. Hordley, and P. M. Morovic. Chromagenic colour constancy. In *Proceedings of the 10th Congress of the International Colour Association*, pages 547–550, 2005.

[22] G. D. Finlayson, S. D. Hordley, and P. M. Morovic. Chromagenic filter design. In *Proceedings of the 10th Congress of the International Colour Association*, pages 1023–1026, 2005.

[23] G. D. Finlayson, S. D. Hordley, and P. M. Morovic. Colour constancy using the chromagenic constraint. In *Computer Vision and Pattern Recognition 2005*, pages 1079–1086. IEEE, 2005.

[24] G.D. Finlayson, S.S. Chatterjee, and B.V. Funt. Color angular indexing. In *The Fourth European Conference on Computer Vision (Vol II)*, pages 16–27. European Vision Society, 1996.

[25] 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.

[26] G.D. Finlayson, B. Schiele, and J. Crowley. Comprehensive colour image normalization. In *eccv98*, pages 475–490, 1998.

[27] Graham D. Finlayson, Mark S. Drew, and Brian V. Funt. Color constancy: generalized diagonal transforms suffice. *Journal of the Optical Society of America, A*, 11(11):3011–3019, 1994.

[28] Graham D. Finlayson, Mark S. Drew, and Brian V. Funt. Spectral Sharpening: sensor transformations for improved color constancy. *Journal of the Optical Society of America, A*, 11(5):1553–1563, 1994.

[29] Graham D. Finlayson and Ruixia Xu. Convex programming colour constancy. In *Workshop on Color and Photometric Methods in Computer Vision*, pages 1–7. IEEE, October 2003.

[30] D. A. Forsyth. A Novel Algorithm for Colour Constancy. *International Journal of Computer Vision*, 5(1):5–36, 1990.

[31] B. V. Funt, M.S. Drew, and J.Ho. Color constancy from mutual reflection. *International Journal on Computer Vision*, 6:5–24, 1991.

[32] Brian Funt, Kobus Barnard, and Lindsay Martin. Is machine colour constancy good enough? In *5th European Conference on Computer Vision*, pages 455–459. Springer, June 1998.

[33] Brian V. Funt and Graham D. Finlayson. Color Constant Color Indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):522–529, 1995.

[34] T. Gevers and A.W.M. Smeulders. Color based object recognition. *Pattern Recognition*, 32:453–464, 1999.

[35] T. Gevers and H. M. G. Stokman. Classifying color transitions into shadow-geometry, illumination highlight or material edges. In *International Conference on Image Processing*, pages 521–525, 2000.

[36] Graham. D. Finlayson, Steven Hordley, and Paul Hubel. Illuminant estimation for object recognition. *COLOR research and application*, 27(4):260–270, 2002.

[37] D. Harris and S. J. MacRow-Hill. Application of chromagen lenses in dyslexia: a double-masked placebo-controlled trial. *Journal of the American Optometry Association*, 70:629–640, 1999.

[38] Glenn Healey. Estimating spectral reflectance using highlights. In Glenn E. Healey and Steven A. Shafer and Lawrence B. Wolff, editor, *Color*, pages 335–339. Jones and Bartlett, Boston, 1992.

[39] Robert V. Hogg and Elliot A. Tanis. *Probability and Statistical Inference*. Prentice Hall, 2001.

[40] Graham D. Finlayson Steven Hordley. Improving gamut mapping color constancy. *IEEE Transactions on Image Processing*, 9(10):1774–1783, October 2000.

[41] S. D. Hordley and G. D. Finlayson. Re-evauating colour constancy. In *Proceedings of the 17th International Conference on Pattern Recognition*, pages 76–79. IEEE, August 2004.

[42] R.W.G. Hunt. *The Reproduction of Colour*. Fountain Press, 5th edition, 1995.

[43] Gudrun J. Klinker, Steven A. Shafer, and Takeo Kanade. A physical approach to color image understanding. *International Journal of Computer Vision*, 4:7–38, 1990.

[44] Edwin H. Land. Lightness and retinex theory. *Journal of the Optical Society of America, A*, 61:1–11, 1971.

[45] Hsien-Che Lee. Method for computing scene-illuminant chromaticity from specular highlights. In Glenn E. Healey and Steven A. Shafer and Lawrence B. Wolff, editor, *Color*, pages 340–347. Jones and Bartlett, Boston, 1992.

[46] Laurence T. Maloney. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *Journal of the Optical Society of America, A*, 3(10):1673–1683, 1986.

[47] Laurence T. Maloney and Brian A. Wandell. Color constancy: a method for recovering surface spectral reflectance. *Journal of the Optical Society of America, A*, 3(1):29–33, 1986.

[48] Rajeev Ramanath, Wesley E. Snyder, Youngjun F. Yoo, and Mark S. Drew. Color image processing pipeline in digital still cameras. *IEEE Signal Processing, Special Issue on Color Image Processing*, pages 34–43, 2004.

[49] Charles Rosenberg, Thomas Minka, and Alok Ladsariya. Bayesian colour constancy with non-gaussian models. In *NIPS*. IS&T/SID, 2003.

[50] Guillermo Sapiro. Bilinear voting. In *ICCV98*, pages 178–183, November 1998.

[51] G. Schaefer. Robust dichromatic colour constancy. In *Springer Lecture Notes in Computer Science, LNCS Vol. 3212 (Proc. Int. Conference Image Analysis and Recognition*, pages 257–264, 2004.

[52] G. Schaefer, S. Hordley, and G. Finlayson. A combined physical and statistical approach to colour constancy. In *Computer Vision and Pattern Recognition 2005*, pages 148–153, June 2005.

[53] S. A. Shafer. Using color to separate reflection components. *Color Research and Application*, 10:210–218, 1985.

[54] M. Stricker and M. Orengo. Similarity of color images. In *SPIE Conf. on Storage and Retrieval for Image and Video Databases III*, volume 2420, pages 381–392, 1995.

[55] S Tominaga, S. Ebisui, and B.A. Wandell. Scene illuminant classification - brighter is better. *Journal of the Optical Society of America, A*, 18(1):55–64, 2001.

[56] Shoji Tominaga and Brian A. Wandell. Standard surface-reflectance model and illuminant estimation. *Journal of the Optical Society of America, A*, 6(4):576–584, 1996.

[57] A. Valberg and B. Lange-Malecki. Color constancy in Mondrian patterns: A partial cancellation of physical chromaticity shifts by simultaneous contrast. *Vision Research*, 30(3):371–379, 1990.