

Gamut Constrained Illuminant Estimation

G. D. Finlayson and S. D. Hordley
School of Computing Sciences
University of East Anglia
Norwich NR4 7TJ, UK
{graham, steve}@sys.uea.ac.uk

I. Tastl
Hewlett Packard Laboratories
Hewlett Packard Incorporated
Palo Alto, CA 94306, USA
I.Tastl@hpl.hp.com

Abstract

This paper presents a novel solution to the illuminant estimation problem: the problem of how, given an image of a scene taken under an unknown illuminant, we can recover an estimate of that light. The work is founded on previous gamut mapping solutions to the problem which solve for a scene illuminant by determining the set of diagonal mappings which take image data captured under an unknown light to a gamut of reference colours taken under a known light. Unfortunately a diagonal model is not always a valid model of illumination change and so previous approaches sometimes return a null solution. In addition, previous methods are difficult to implement. We address these problems by recasting the problem as one of illuminant classification: we define a priori a set of plausible lights thus ensuring that a scene illuminant estimate will always be found. A plausible light is represented by the gamut of colours observable under it and the illuminant in an image is classified by determining the plausible light whose gamut is most consistent with the image data. We show that this step (the main computational burden of the algorithm) can be performed simply, quickly, and efficiently by means of a non-negative least-squares optimisation. We report results on a large set of real images which show that it provides excellent illuminant estimation, outperforming previous algorithms.

1. Introduction

Colour has been found to be a useful aid in solving a variety of classical computer vision problems such as object recognition and tracking as well as image segmentation. Importantly, colour can only help in solving these problems provided that it is a stable feature of an object or surface. This implies that cameras must record colours which are constant regardless of the colour of the prevailing light. In practice a camera's response changes as the

illumination changes, so it is necessary to correct camera data to account for the colour of the prevailing light. This procedure is a two-stage process which begins by estimating the colour of the scene illumination and subsequently correcting the recorded image to account for the estimated illuminant colour. The second of these tasks – correcting the image for the colour of a known illuminant – is relatively straightforward. However, recovering an estimate of the scene illuminant from an arbitrary image has proven to be a very difficult problem, and despite much research [16, 5, 18, 13, 8, 14, 9, 2, 3] is still today very much an active area of research. In this paper we present a novel solution to the illuminant estimation problem, a solution which builds on the strengths of the *gamut mapping* algorithm of Forsyth [13] while addressing many of its weaknesses.

Early approaches to illuminant estimation sought to simplify the problem by placing certain constraints on the composition of a scene. For example the constraint that the scene contain a “white” (maximally reflective) surface [16] or that the average of all surfaces in the scene is neutral [5]. Other authors modelled lights and surfaces by low-dimensional linear models and derived algebraic solutions to the problem [18]. It is easy to understand why such approaches do not work in practice: the constraints they place on scenes are too strong. By adopting weaker constraints a variety of more successful algorithms have been developed [13, 14, 9] and these approaches can often give reasonable illuminant estimation. In particular there are three classes of algorithm which show promise and are worthy of further research: neural network based methods [6], gamut mapping methods [13, 8, 11, 1] which share the common root of the original gamut mapping work of Forsyth [13] and correlation based approaches such as Finlayson *et al's* [9] “Color by Correlation”. The neural network approach is unattractive for a number of reasons. First, it provides a black box solution to the problem which casts no light on the nature of the problem which is being studied. Second, neural networks are non-trivial to implement, and third and

most crucially they rely on training data which properly reflects the statistics of the world: in practice neural networks often do not generalise well. More promising is the correlation matrix algorithm [9] which has the advantage of being simple both in its implementation and in terms of the computations which must be carried out to estimate the illuminant. Set against these advantages is the fact that this approach also relies on having a reasonable statistical model of lights and surfaces: i.e. it requires accurate training data to work in practice. A second disadvantage of the approach is that it gains its simplicity in part from the fact that it works not with 3-d *RGB* camera data but with 2-d chromaticity data: i.e. brightness information is discarded. While brightness information can be incorporated into the algorithm [4], doing so destroys the simplicity of the approach. We believe that brightness information is potentially useful in the estimation process and so would like an algorithm which is able to make use of it. In light of the weaknesses of these two classes of algorithm we focus in this work on the third class of algorithms: the gamut mapping methods. In particular we describe in this paper a new illuminant estimation algorithm which is inspired by those methods but which addresses many of the weaknesses in the algorithms proposed so far.

Gamut mapping illuminant estimation was first proposed by Forsyth [13] who derived the algorithm on the basis of a sound theoretical analysis of the illuminant estimation problem. His solution is founded on the observation that the set of camera *RGBs* observable under a given illuminant is a bounded convex set. This observation follows directly from physical constraints on the nature of a surface (all surfaces reflect between 0 and 100% of light incident upon them) and from the linear nature of image formation. Forsyth coined the term *canonical gamut* to describe the set of *RGBs* which can be possibly observed under some canonical illuminant. Now, an image whose illuminant is to be estimated can similarly be represented by a second set, an *image gamut*, which contains the *RGBs* observed in the image. Estimating the illuminant then becomes the problem of finding a mapping which takes the image gamut into the canonical gamut. Forsyth proposed that the mapping should take the form of a diagonal matrix: i.e. *RGBs* in the image are mapped to the canonical gamut by three scale factors applied independently to each of the *R*, *G*, and *B* responses. He pointed out that under such an assumption the mapping taking image gamut to canonical gamut would in general not be unique: i.e. there are many mappings which map the image gamut inside the canonical gamut. This amounts to saying that many possible illuminants are consistent with the observed image data. Forsyth's strategy was to first find the set of all feasible mappings and then to choose a single feasible mapping as the illuminant estimate.

The algorithm has been found to be reasonably success-

ful but it does suffer from a number of inherent limitations. One of the algorithm's greatest drawbacks is that its implementation is difficult and computationally complex. These difficulties are compounded by the fact that the operation of the algorithm depends on being able to accurately intersect a number of convex polyhedra to determine the feasible set of illuminants. Failure to perform this step accurately can sometimes result in the algorithm returning a null intersection: i.e. the algorithm is unable to estimate the illuminant in a scene. Another weakness of the algorithm is the fact that it is founded on the explicit assumption that illumination is well modelled by a diagonal matrix transform. While under certain circumstances a reasonable assumption, it cannot be said to be generally true. In particular many real scenes will contain an illuminant, or more commonly, one or more surfaces, for which the model is not adequate and in such situations the algorithm can return an unrealistic set of feasible illuminants, or worse, no solution at all. Finally, even when the algorithm works correctly it can sometimes return a feasible set which contains illuminants which will never be encountered in the world, thus it fails to incorporate all the available prior knowledge about the world.

In this paper we present an algorithm which avoids all these problems while still retaining the spirit of the original work. We avoid the problems of modelling illumination change by making no explicit model of illumination change. Rather, we observe that the range of illuminations we encounter in the world is quite restricted so that we can choose *a priori* and independently of any image data a representative set of plausible scene illuminants. Next, we characterise the gamut of possible device *RGBs* observable under each of these plausible lights. Given an image whose illuminant we wish to estimate we first determine its gamut. We can then get a measure of the degree to which each plausible illuminant is consistent with the image data by determining how well the image gamut matches the gamut for a plausible light. This is the core of the algorithm and importantly we are able to cast it in terms of a non-negative least squares problem - a problem for which there exist, simple, fast and easily implemented algorithms. Given a measure of gamut consistency for each plausible illuminant we can estimate the scene illuminant in a number of different ways. It follows that our method is guaranteed to provide an illuminant estimate for any set of image data we encounter.

The proposed algorithm also has a number of other advantages over previously proposed methods. First, unlike neural network or correlation matrix based approaches, it is not so dependent on accurate knowledge of the statistics of surfaces and illuminants. Furthermore the algorithm is essentially the same regardless as to whether we use 3-d *RGB* data or 2-d chromaticity data thus we are not forced to make decisions on which data we use on the basis of im-

plementational issues. Finally results on real images show that the algorithm yields excellent illuminant estimation.

The next section contains a discussion of illuminant estimation problem itself and summarises Forsyth's proposed *Gamut Mapping* solution. Then, in Section 3 we address the weaknesses of this approach by presenting a modified algorithm which we call *Gamut Constrained Illuminant Estimation* (GCIE). In Section 4 we evaluate the performance of the new algorithm on a large set of real images and we also discuss how best to set some of the free parameters in the algorithm's implementation. We conclude the paper in Section 5 with a brief summary.

2. Background

To understand the illuminant estimation problem we must first consider the process of image formation. In common with previous work, we adopt a simple Lambertian model [15] of image formation. We assume further that our camera samples the light incident on its image plane with three types of sensor, whose sensitivities are defined by $R_k(\lambda)$ ($k = 1, 2, 3$) – a function of wavelength λ . Light from an object is characterised by the colour signal $C(\lambda)$ which, under the assumptions of the Lambertian reflectance model, can be written as: $C(\lambda) = E(\lambda)S(\lambda)$ where $E(\lambda)$ is the spectral power distribution (SPD) of the ambient illumination, assumed constant across the scene, and $S(\lambda)$ is the surface reflectance function of the object.

This colour signal is incident at a point on the image plane and the response of the k^{th} sensor, p_k is given by:

$$p_k = \int_{\omega} E(\lambda)S(\lambda)R_k(\lambda)d\lambda \quad (1)$$

where the integral is taken over ω , the visible spectrum. The response of a colour camera at a point on its image plane is thus: $\underline{p} = (p_1, p_2, p_3)^t$: a triplet of sensor responses. We will refer to this triplet either as \underline{p} , or as *RGB*. Equation (1) makes it clear that a camera's sensor responses strongly depend on $E(\lambda)$: the ambient illumination.

We can pose the illuminant estimation problem as that of inverting Equation (1) to recover $E(\lambda)$. However, since light is a continuous function of wavelength, and since a camera samples the incoming light at only three (broad) bands of the spectrum, solving Equation (1) is impossible without additional constraints. Fortunately, in most applications we are interested not in the spectrum of the illumination *per se*, but rather in determining what a scene imaged under that illumination would look like when rendered under a different light. Let us represent an image taken under an unknown light o , by a set \mathcal{I}^o , of n sensor responses $\mathcal{I}^o = \{\underline{p}_1^o, \underline{p}_2^o, \dots, \underline{p}_n^o\}$. For this image we would like to determine the corresponding set of data \mathcal{I}^c which would be

observed under a reference, or *canonical light* c . Defining the problem in this way has the advantage that explicit knowledge of $E(\lambda)$ is not required. Rather, solving this problem amounts to determining a mapping $\mathcal{F}()$ such that:

$$\underline{p}_i^c = \mathcal{F}(\underline{p}_i^o), \quad i = 1 \dots n \quad (2)$$

Forsyth [13] formulated the illuminant estimation problem in this way, and developed algorithms to solve it. His solutions are founded on the observation that due to physical constraints on the nature of lights (a light can emit no less than no energy at each wavelength) and surfaces (a surface can reflect no less than no light incident upon it, and no more than all incident light), the set of image colours observable under a given light is restricted. That is, we can define for our reference light c , a *canonical gamut* representing all the image colours that can be observed under that light. The canonical gamut is a bounded region of sensor *RGB* space. A change of illumination implies a different gamut of observable image colours and this new gamut is related to the first by the mapping \mathcal{F} . That is, if \mathcal{C} denotes the gamut of possible image colours under the canonical light c , and \mathcal{O} , the gamut under the second light o then:

$$\underline{p}^o \in \mathcal{O} \iff \mathcal{F}^{o,c}(\underline{p}^o) \in \mathcal{C} \quad (3)$$

where $\mathcal{F}^{o,c}$ is the mapping taking colours under light o , to their corresponding colours under light c .

Forsyth showed that the gamut of image colours possible under a given light form a bounded convex set; for a three sensor device this set is a convex polyhedron. To specify an algorithm to find this mapping we must first establish what form the mapping \mathcal{F} should take. The algorithm proposed and implemented by Forsyth is founded on the assumption that the mapping takes the form of a *diagonal matrix*. That is:

$$\underline{p}^o \in \mathcal{O} \iff D^{o,c}\underline{p}^o \in \mathcal{C} \quad (4)$$

where D^o is a 3×3 diagonal matrix. (later work [10] has shown that such a model is well justified for a large class of devices and illuminants). Estimating the scene illuminant now becomes the problem of finding the three non-zero entries of $D^{o,c}$. Forsyth developed a two-stage algorithm to perform this task. The algorithm begins with a set of image colours \mathcal{I}^o recorded under an unknown light o , from which it is desired to recover the mapping $D^{o,c}$. However, the set of colours, \mathcal{I}^o , in an image will in general be a subset of all possible colours (i.e. $\mathcal{I}^o \subset \mathcal{O}$) and the mapping taking the image gamut \mathcal{I}^o inside the canonical gamut \mathcal{C} may therefore not be unique. Thus, in the first stage of the algorithm a *feasible set* of mappings, \mathcal{D} is determined. Every mapping in this set maps the image gamut to the canonical gamut. The feasible set characterises the set of illuminants which are consistent with the image data. To complete the solution a single mapping is chosen from this set as an estimate of the

actual mapping $D^{o,c}$. Of course if many lights are feasible, choosing one of them may result in a wrong illuminant estimate. Thus, when selecting an illuminant care must be taken to choose an answer which is broadly representative of the whole set.

In its original form Forsyth's algorithm can often give a good estimate of the scene illuminant and it performs better on average than a number of other approaches [2, 3]. The algorithm does though suffer from a number of limitations. First, its success depends on the degree to which illumination change can be modelled by a diagonal matrix transform. Second, even when the properties of a device's sensors imply that such a model is a good one, real images often contain features (for example specularities) which confound the model. In either of these two cases, and due in part to the manner in which the algorithm is implemented, the algorithm will on occasion return no answer at all. Third, the method can return an illuminant estimate which is "implausible". That is, while the feasible set characterises all mappings consistent with the image data, not all of these mappings will correspond to lights which occur in the real world. Finally, the implementation of the algorithm is computationally complex and quite sensitive to the input data: determining the feasible set requires the intersection of a number of convex polyhedra and this intersection can at times, due to numerical instability, wrongly result in a null intersection - that is no estimate of the scene illuminant.

Since Forsyth's original work a number of authors have proposed various modifications to the gamut mapping algorithm which aim to make the algorithm more robust to diagonal model failure [8, 1], to ensure that the feasible set of illuminants is restricted to real world illuminants [8] or to improve the performance of the algorithm by more carefully selecting an illuminant from the feasible set [11, 1]. However, none of these algorithms address the most important limitations of the original approach: all algorithms employ essentially the same implementation and can thus be confounded by numerical accuracy problems or diagonal model failure, sometimes resulting in a null solution. To address these limitations we describe our new gamut mapping based illuminant estimation algorithm in the next section.

3. Gamut Constrained Illuminant Estimation

We begin by defining *a priori* a set of M plausible scene illuminants. This plausible set reflects our prior knowledge about the range of illuminants we expect to encounter and we note that the number and range of illuminants in this set will vary depending on how much prior knowledge of the illumination we have. Next we determine for each plausible illuminant the gamut of device *RGBs* observable under it. In theory, characterising this gamut for illuminant i requires us to determine the response of our device to all pos-

sible surface reflectance functions imaged under that light. In practice we will choose a subset of these reflectances. It will become clear that the success of the algorithm will be dependent on how this set is selected, but we leave a discussion of this issue until Section 4. Let us denote the N reflectances by $\{S^j(\lambda), j = 1 \dots N\}$. Given $E^i(\lambda)$: the SPD of the i^{th} plausible illuminant, and the spectral sensitivities of a device we can use Equation (1) to calculate the sensor responses \underline{p}_j^i to each surface.

Of course, an illuminant can have arbitrary brightness so we should really allow any scalar multiple of \underline{p}_j^i to be a possible response. However, given only sensor responses there is an ambiguity about the brightness of the illuminant, since a sensor response given by Equation (1) could equally well be accounted for by an illuminant with spectral power $sE^i(\lambda)$ and surface reflectance $\frac{1}{s}S^j(\lambda)$. Rather than trying to resolve this ambiguity, we instead ignore the overall intensity of the illumination and attempt to recover only its relative spectral power. Or, treating a sensor response as a vector in three-dimensional sensor space, consider only the direction of this vector and not its magnitude. In this case, a sensor response can be represented as a 3-vector \underline{q}_j^i , by applying a transform to factor out intensity information, for example:

$$q_{j,1}^i = \frac{p_{j,1}^i}{\sum_{k=1}^3 p_{j,k}^i}, \quad q_{j,2}^i = \frac{p_{j,2}^i}{\sum_{k=1}^3 p_{j,k}^i}, \quad q_{j,3}^i = \frac{p_{j,3}^i}{\sum_{k=1}^3 p_{j,k}^i} \quad (5)$$

The first two elements of \underline{q}_j^i are called the chromaticity coordinates of the sensor response and the set of all \underline{q}_j^i lie on a plane in three-dimensional sensor space. We can thus define the response of a device to the set of surface reflectances imaged under illuminant i as:

$$G^i = \{\underline{q}_1^i, \underline{q}_2^i, \dots, \underline{q}_N^i\} \quad (6)$$

It can be shown [12] that if all of G^i are observable under illuminant i , then so is any linear combination (with positive coefficients) \underline{q}_k^i :

$$\underline{q}_k^i = \sum_{j=1}^N \alpha_j \underline{q}_j^i, \quad \forall \alpha_j \geq 0 \quad (7)$$

The gamut of the i^{th} illuminant, which we denote \mathcal{G}^i , is thus an infinite cone in sensor space whose vertex is the origin and whose extreme rays are defined by the convex hull of G^i . This cone defines the set of observable chromaticities, but places no restriction on how bright a colour can be.

Having defined the gamut for each illuminant the next step is to use these gamuts to estimate the scene illuminant given a set of image data. To this end let us first represent the image by the set of all its sensor responses:

$$\mathcal{I}^o = \{\underline{p}_1^o, \underline{p}_2^o, \dots, \underline{p}_n^o\} \quad (8)$$

Now let us hypothesise that the i^{th} illuminant is the scene illuminant. If this hypothesis is true then any image colour, \underline{p}_k^o will fall within the cone of observable image colours for the i^{th} illuminant. Or in mathematical terms:

$$\underline{p}_k^o = \sum_{j=1}^N \alpha_j \underline{q}_j^i \quad \alpha_j \geq 0, \quad \underline{q}_j^i \in \mathcal{G}^i, \quad \forall j \quad (9)$$

On the other hand, if the hypothesis is not true then an image colour might fall outside the gamut of the i^{th} illuminant so that there will be some error in representing an image colour in this way:

$$e_k^i = \left\| \underline{p}_k^o - \sum_{j=1}^N \alpha_j \underline{q}_j^i \right\|^2 \quad (10)$$

Independently of whether or not the hypothesis is true we can determine the “best” representation of the image colour with respect to Equation (10) by choosing the α_j which minimise the error e_k^i : a simple least-squares problem. However, we have the added constraint that all the α_j should be positive, so we have in fact a *non-negative* least squares problem - a problem for which a known, fast, and simple solution exists [17]. We point out that readers who are familiar with optimization will understand that minimizing (10) might be carried out using a variety of numerical algorithms. The non-negative least squares algorithm is chosen because of its very fast operation.

The total error in hypothesising the i^{th} illuminant as the scene light is given by: $e_{total}^i = \sum_{k=1}^n e_k^i$. If the image data falls completely in the gamut of a light then the corresponding error for this light will be zero and we can say that the hypothesis accounts perfectly for the data. If one or more of the colours in the image falls outside the gamut for a light then the corresponding error will be non-zero. The greater the error, the less well the hypothesis accounts for the data. Thus we can use the error to estimate the scene illuminant. For example, we can choose as our estimate the illuminant with minimum fitting error.

Clearly the algorithm we have just formulated has similarities to Forsyth’s gamut mapping approach, however our solution is posed in such a way that it avoids the limitations in the original method which we previously discussed. By hypothesising a set of candidate illuminants we avoid the need to adopt an explicit model of illumination change and we also ensure that the illuminant estimate we recover will correspond to a real, physically plausible light: a property absent from Forsyth’s formulation. Furthermore the algorithm is guaranteed to return a solution in contrast to Forsyth’s algorithm and other modifications of it. Finally, the algorithm is much simpler to implement than is Forsyth’s, being in essence a repeated application of the method of non-negative least squares for which there exist fast, robust and computationally simple algorithms.

3.1. Variations on GCIE

In the algorithm just described the gamut of an illuminant is defined to be an infinite cone in sensor *RGB* space. That is, we constrain the chromaticity of possible sensor responses but we allow their intensity to be arbitrary. However, when checking whether or not an image *RGB* is consistent with a particular illuminant, we retained the intensity of the *RGB* so that the error is given by Equation (10). Two variations of the algorithm suggest themselves immediately: we could discard intensity information both when defining the illuminant gamut and when checking image *RGB*s for gamut consistency. That is, we check not that an *RGB* lies within the illuminant’s gamut, but rather, whether or not an image *chromaticity* is consistent with the illuminant. A second variation of the algorithm is to retain intensity information both in the image data and when constructing the illuminant gamuts. Considering the second case first, let us represent the gamut of the i^{th} illuminant by a set \mathcal{G}^i , of sensor responses such that any convex combination of these responses can be observed under illuminant i . That is any possible response \underline{p}_k^i can be written:

$$\underline{p}_k^i = \sum_{j=1}^N \alpha_j \underline{p}_j^i \quad \underline{p}_j^i \in \mathcal{G}^i, \quad \alpha_j \geq 0, \quad \sum_{j=1}^N \alpha_j = 1 \quad (11)$$

It follows that the error in fitting the k^{th} response to the i^{th} illuminant gamut is given by:

$$e_k^i = \left\| \underline{p}_k^o - \sum_{j=1}^N \alpha_j \underline{p}_j^i \right\|^2 \quad (12)$$

Finding the “best” fitting α_j is again a constrained least-squares problem but now, in addition to the non-negativity constraint on α_j we have the additional constraint that the α_j sum to one: $\sum_{j=1}^n \alpha_j = 1$. This implies that we can no longer determine the α_j directly by non-negative least-squares.

The problem though can be converted to a non-negative least-squares optimisation in the following way. Let us convert a three-dimensional sensor response \underline{p} into a four-dimensional vector \underline{r} :

$$\underline{r} = (p_1, p_2, p_3, W)^t \quad (13)$$

where W is some constant. If we do this, for all sensor responses in both the illumination gamut and the image, then we can define a new error term \hat{e}_k^i :

$$\hat{e}_k^i = \left\| \underline{r}_k^o - \sum_{j=1}^n \alpha_j \underline{r}_j^i \right\|^2 \quad (14)$$

and clearly:

$$\hat{e}_k^i = e_k^i + W \left(1 - \sum_{j=1}^n \alpha_j \right) \quad (15)$$

If we choose W to be sufficiently large so that the term $W \left(1 - \sum_{j=1}^n \alpha_j \right)$ is the dominant one in Equation (15), then minimising the error in that equation by the method of non-negative least squares will lead to a set of α_j whose sum is approximately one. In this way we can use the method to determine the convex combination of points in \mathcal{G}^i which best represent the sensor response p_k^o . In all other ways the algorithm is the same as that presented in the previous section.

We could employ a similar approach to handle the variation in which we discard intensity information both from the image data and when constructing the illuminant gamuts. In this case we represent camera responses as two-dimensional chromaticity co-ordinates \underline{c} :

$$c_{j,1} = \frac{p_{j,1}}{p_{j,1} + p_{j,2} + p_{j,3}}, \quad c_{j,2} = \frac{p_{j,2}}{p_{j,1} + p_{j,2} + p_{j,3}} \quad (16)$$

We then convert these chromaticities into 3-dimensional vectors \underline{s} :

$$\underline{s} = (c_1, c_2, W)^t \quad (17)$$

and follow the same minimisation procedure as in the 3-dimensional case.

This analysis shows that our algorithm is the same regardless as to whether we consider 2- or 3-dimensional data. Thus we are free to choose the dimensionality of our data on the basis of which gives the best results rather than on the basis of what makes the algorithm easy to implement.

In the next section we address the issue as to which version of the algorithm is best and discuss how to fix some of the free parameters in the algorithm by considering its performance in a set of illuminant estimation experiments.

4. Algorithm Performance

To test the performance of the new algorithm we followed an experiment of Barnard *et al* [3] who recently published a comparison of many different illuminant estimation algorithms, assessing their performance on a set of real images. The images consist of 32 scenes captured under 11 different lights¹ and were captured specifically for testing the performance of illuminant estimation algorithms².

¹In total 321 images were used in the experiment because a small number of the captured images were not suitable for the experiment.

²We thank Dr. Kobus Barnard *et al* for making this data publicly available from www.cs.sfu.ca/colour/data.

Scene content is varied (ranging from a standard photographic test chart to just a few green apples) and thus represents different levels of difficulty for illuminant estimation. The 11 illuminants encompass a wide range of illuminants typically encountered in the world.

Testing algorithm performance on this set of images allows us to perform a rigorous test of the new algorithm's performance and also to easily compare performance to that of many other algorithms reported in [3]. Specifically, in addition to GCIE, we report performance for the following algorithms: *Max-RGB*, and *Grey-World* which estimate the illuminant by the image maximum and image average respectively. A *neural-network* algorithm, a version of Finlayson *et al's* *Color by Correlation* algorithm, and a modified version of Forsyth's *gamut mapping* algorithm. Details of each algorithm's implementation can be found in [2]. Where different versions of an algorithm exist we report results for the best performing version.

The experimental paradigm is simple: for each image we obtain an estimate of the *RGB* response to a perfectly reflecting surface using each algorithm. For example, if GCIE selects the i^{th} illuminant as being most consistent with the image data we use the *RGB* value corresponding to white under illuminant i as our illuminant estimate. We compare this estimate to the *RGB* response of a white tile viewed under the actual illuminant³.

When assessing error, we discard intensity information (which we cannot recover without ambiguity) and look at error only in chromaticity. We do this by calculating the angular distance between an algorithm's estimate of the *RGB* ($\underline{\hat{p}}_w$) and the *RGB* of the actual illuminant (\underline{p}_w):

$$\text{Angular Error} = \cos^{-1} \left(\frac{\underline{\hat{p}}_w \cdot \underline{p}_w}{\|\underline{\hat{p}}_w\| \|\underline{p}_w\|} \right) \quad (18)$$

To completely specify the GCIE algorithm we must fix a number of free parameters in the algorithm design. First, we must decide on how many and which illuminants we deem to be plausible. In general this choice will depend on what prior knowledge we have about the scene illumination. We report results here for two extreme cases: in one case we use the 11 lights under which the scenes were captured and in a second case we use 87 lights which represent a uniform sampling (in chromaticity space) of a large number of measurements of real world illuminants. This allows us to test algorithm performance when we have maximal (11 lights) and minimal (87 lights) prior knowledge.

Next we need to specify the set of surfaces used to calculate the gamut for each plausible illuminant. This stage is crucial to the success of the algorithm since in choosing our

³This *RGB* was obtained from a white tile placed in each scene: images were obtained with and without the tile in the scene and testing was done on images without the tile.

gamuts we must ensure first that the gamuts for illuminants do indeed differ from one another. In developing the theory we insisted only that surface reflectance functions on which gamuts are based be physically realisable. This is a very weak constraint and by itself will lead to very large gamuts. For our purposes such gamuts are likely to be ineffectual since while these gamuts will differ at their extremes, these extremes are likely to correspond to reflectances which, while physically realisable, will occur only rarely in the world. Instead we base our gamuts on reflectance functions which are in a statistical sense “likely” to occur in scenes. We do this by basing gamuts on sets of measured surface reflectance functions. But rather than using all these reflectances we instead model their chromaticity distributions using a bi-variate Normal distribution and construct gamuts using only those reflectances which fall within some number of standard deviations of the mean chromaticity. Adopting such a model allows us, in a formal sense, to construct a gamut of statistically likely chromaticities and to control the size of a gamut by considering only reflectances whose chromaticities fall within a standard deviation ellipse of our choice.

A third issue is how we choose an illuminant estimate once we have a measure of consistency for each plausible light. We have experimented with two methods for doing this: either we choose the illuminant most consistent with the image data (minimum fitting error) or we take the average of a number of illuminants with small fitting error. We have found both methods to give similar results and so report here results for the minimum error approach.

The final issue which needs to be addressed is what pre-processing of the image data is performed prior to estimating the illuminant. Barnard *et al* [3] have found that pre-processing has a significant effect on performance for most estimation algorithms. In the case of GCIE pre-processing is important both in the context of algorithm performance but also in the interests of computational efficiency. The main computational step in the algorithm is to check the consistency of a set of image data with the gamuts of each of the candidate illuminants. Checking the consistency, while a simple step, would be excessively burdensome if repeated for all pixels in an image. Thus we seek a method of reducing the number of *RGBs* which need to be checked. The most principled way to do this is to segment the image into the underlying surfaces which constitute the scene and to use the *RGBs* of each segment as input into the algorithm. Since the number of surfaces in a scene is generally small - at least relative to the number of pixels in the image, this will lead to a large reduction in the number of *RGBs* which must be checked and has the added advantage that the *RGBs* correspond to real surfaces. Against this approach is the fact that the segmentation of an image is itself a computationally intensive task and the fact that existing

	Mean	Median
Max- <i>RGB</i>	8.9	-
Grey-World	11.7	-
Neural Network	9.1	-
Color by Correlation	9.9	-
Gamut Mapping	5.6	-
GCIE Version 1, 11 lights	6.8	5.2
GCIE Version 2, 11 lights	7.4	5.8
GCIE Version 3, 11 lights	4.9	2.3
GCIE Version 1, 87 lights	8.6	6.5
GCIE Version 2, 87 lights	9.0	7.9
GCIE Version 3, 87 lights	5.3	3.2

Table 1. Angular error (degrees) for the different algorithms averaged over 321 images. Algorithms in bold were presented in this paper.

algorithms do not always produce an accurate segmentation of the scene. However, even quite a crude segmentation of the scene is likely to be good enough for our purposes and the computational cost of the segmentation is still likely to be less than that associated with testing every pixel in the scene.

For GCIE we used the segmentation method of Meer *et al* [7]. There was no reason for choosing this method beyond the fact that the authors had made available an implementation of their work. We found it to give reasonably accurate segmentations of all the images and it produced on average, around 15 regions per image. This implies that the process of illuminant estimation is then very quick since we need to check only a very small number of *RGBs*. Results for the algorithms tested in [3] are also obtained with segmented images. We point out that the segmentation method is not the same as that used for GCIE, however the segmentations ought to be close enough to allow us to compare illuminant estimation performance. We point out further, that the pre-processing was not chosen to optimise performance for any particular algorithm, thus it is possible that the performance of any tested algorithm could be significantly improved by choosing a more appropriate pre-processing procedure. Table 1 summarises results for all the algorithms tested in terms of average angular error in degrees. For the GCIE algorithms we also report median angular error, data which was unavailable for the other methods. We report results for three variants of the new GCIE algorithm. Version 1 ignores intensity when constructing gamuts, but retains the intensity of the image data, version 2 ignores intensity in both cases, and version 3 retains intensity in both cases.

The results reveal a number of important facts. First is that the gamut mapping algorithms, whether the original

gamut mapping or our new GCIE approach give the best performance. This suggests that algorithms which place only weak constraints on the world can outperform other more restrictive approaches such as Max-RGB or Grey-World. It is surprising therefore that other well founded approaches such as Neural Networks and Color by Correlation perform less well in these tests. Their poor performance is most likely to be related to the fact that their success relies upon having accurate statistical knowledge about the frequency of occurrence of surface reflectances. Inaccuracies in this knowledge manifest themselves in poor illuminant estimation.

Turning to the performance of the new algorithm we see that performance is very much dependent on which version of the algorithm we use. Specifically, the results suggest that it is important to consider intensity information both when constructing the gamuts of plausible lights and when checking image data for gamut consistency. Indeed, keeping full 3-d information leads to algorithm performance which is better than the previous best algorithm (the original gamut mapping algorithm) and in absolute terms an angular error of 4.9 represents excellent average performance. We further point out that median performance is also very good - 50% of images have an error 2.3 or less. Another interesting aspect of the results is the performance of GCIE when we use 11 lights as compared to 87. Interestingly, providing the maximum prior knowledge about the illuminants (restricting lights to the 11 scene lights) does not give too much better performance. We have found that gamut size has a small but significant effect on algorithm performance: the results presented here represent the best case performance where gamuts were constructed by using all reflectances whose chromaticities were within 2.5 standard deviations of the mean.

5. Conclusions

We have set forth in this paper a new illuminant estimation algorithm which can be viewed as a robust implementation of Forsyth's Gamut Mapping algorithm. We have improved upon Forsyth's work by removing some of the limitations inherent in that approach. Specifically we remove the constraint of a diagonal model of illumination change with the result that unlike the original work, the new method is guaranteed to always give an illuminant estimate which is furthermore constrained to be physically plausible. The simplicity of the implementation of the new approach is in contrast to the original work and leads to an algorithm which is both fast and robust to all imaging conditions. We have demonstrated the power of the new approach by testing it on a set of real images and achieving excellent illuminant estimation performance.

Acknowledgements

The authors acknowledge the support of SONY Corporation and in particular the support of Ted Cooper. Ingeborg Tastl was an employee of SONY US Research Laboratories when this work was undertaken.

References

- [1] K. Barnard. *Practical Colour Constancy*. PhD thesis, Simon Fraser Univ., School of Computing Science, 2000.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310:1–26, 1980.
- [6] Vlad Cardei. *A Neural Network approach to Colour Constancy*. PhD thesis, Simon Fraser Univ., School of Computing Science, 2000.
- [7] Dorin Comaniciu and Peter Meer. Mean shift analysis and applications. In *Proceedings of the 7th International Conference on Computer Vision*, pages 1197–1203. IEEE, 1999.
- [8] G. D. Finlayson. Color in Perspective. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10):1034–1038, 1996.
- [9] 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.
- [10] 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.
- [11] Graham D. Finlayson Steven Hordley. Improving gamut mapping color constancy. *IEEE Transactions on Image Processing*, 9(10):1774–1783, October 2000.
- [12] D. A. Forsyth. *Colour Constancy and its Applications in Machine Vision*. PhD thesis, Oxford University, Department of Engineering Science, 1988.
- [13] D. A. Forsyth. A Novel Algorithm for Colour Constancy. *International Journal of Computer Vision*, 5(1):5–36, 1990.
- [14] Brian V. Funt, Vlad Cardei, and Kobus Barnard. Learning color constancy. In *Proceedings of the Fourth Color Imaging Conference*, pages 58–60, November 1996.
- [15] Berthold K. P. Horn. *Robot Vision*. MIT Press, 1986.
- [16] Edwin H. Land. The Retinex Theory of Color Vision. *Scientific American*, pages 108–129, 1977.
- [17] Charles L. Lawson and Richard J. Hanson. *Solving Least Squares Problems*. Prentice Hall, 1974.
- [18] 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.