Illuminant and Gamma Comprehensive Normalisation in Log RGB Space

Graham Finlayson and Ruixia Xu
School of Information Systems
University of East Anglia
Norwich NR4 7TJ
United Kingdom

graham@sys.uea.ac.uk

September 3, 2002

Abstract

The light reflected from an object depends not only on object colours but also on lighting geometry and illuminant colour. As a consequence the raw colour recorded by a camera is not a reliable cue for object based tasks such as recognition and tracking. One solution to this problem is to find functions of image colours that cancel out dependencies due to illumination. While many invariant functions cancel out either dependency due to geometry or illuminant colour, only the comprehensive normalisation has been shown (theoretically and experimentally) to cancel both. However, this invariance is bought at the price of an iterative procedure.

The first contribution of this paper is to present a non-iterative comprehensive normalisation procedure. Iteration is avoided by working with logarithms of RGB images rather than the RGBs themselves. We show that under certain simplifying assumptions, in log colour space two simple projection operators lead to invariance to geometry and light colour in a single step.

Although both comprehensive normalisation and the non-iterative normalisation work well in the context of colour based object recognition, neither of them accounts for all dependencies that might realistically be present in images. For example, a power (gamma) function is typically applied to image data as part of the coding process and this function can be device and even image dependent. Thus in a second part of the paper we ask whether we can also remove colour dependency due to gamma? We show that we can and furthermore, that invariance can be achieved by adding a single further step to the non-iterative normalisation procedure.

Finally we demonstrate the efficacy of these new normalisation procedures by conducting a series of object recognition experiments on sets of linear and non-linear images.

1 Introduction

Colour is an important cue for object recognition and is often used as a feature for indexing image databases [2, 6, 19, 22]. However, object recognition based on raw camera responses fails when the illumination conditions change. For example, the shading observed on objects is a function of the light source position: colour is dependent on lighting geometry. Also, as the illumination colour changes so too does the colour of the light reflected from objects and as a consequence different RGBs are recorded for the same surface viewed under different lights.

Two approaches are reported in the literature for dealing with these illumination dependencies. First, colour constancy algorithms [18, 15, 7, 10, 24] attempt to recover surface reflectance (or more precisely some correlate of surface reflectance) which is, by definition, independent of lighting conditions. Unfortunately colour constancy has proven to be a very hard problem to solve. The current state of the art can only deliver approximate constancy. Moreover, it has been shown that the constancy delivered by many of the algorithms [14] is insufficient to render colour a stable enough cue for object recognition.

An easier way to deal with the illuminant colour problem is the colour invariant approach [5, 8, 4, 21, 1, 16]. The goal here is to find functions of proximate image pixels which cancel out lighting dependencies. As an example the chromaticity [26] function, used extensively in colour science and computer vision, removes the intensity from an RGB response vector. Since shading in an image is an intensity artifact, it follows that the chromaticity normalisation can be used to remove lighting geometry. Another normalisation common in the colour literature is the grey-world normalisation [26]. Here the average RGB pixel in an image is calculated and the other pixels described relative to the average (if (R_i, G_i, B_i) denotes the response at a pixel and (μ_R, μ_G, μ_B) the average response over the image, then $(R_i/\mu_R, G_i/\mu_G, B_i/\mu_B)$ is calculated). Under most conditions [27, 25] this simple normalisation will remove illuminant colour. Unfortunately, neither the chromaticity nor the grey-world normalisation suffices to remove both illuminant colour and geometry dependencies.

Both dependencies are however removed in the comprehensive normalisation scheme [8]. Here the chromaticity and grey-world normalisations are applied successively and iteratively. It was shown [8] that this procedure converges to a unique fixed point. Comprehensive normalisation for the same scene viewed under most reasonable lighting conditions delivers the same intrinsic image which depends only on surface reflectance. But, is this iteration necessary? In some sense this question has been answered by research which followed the comprehensive normalisation result. The methods of Gevers and Smeulders [16] and Berens *et al* [3] remove the illumination geometry and colour in a single step. Yet to do this, these normalisations take the 3-component RGB image into a new co-ordinate frame: the idea of redness, greenness and blueness is removed and one is left with lighting independent scalars which do not have their usual colour meaning. Thus, while the comprehensive normalisation is iterative, the iterative cost of the computation delivers an RGB colour image as output (and in this image the RGBs have similar "meanings" to the input colours).

In this paper we first ask whether we can normalise an RGB image, and keep the semantic meaning of the RGBs the same but without the cost of iteration? The answer is yes [12]. We begin with the simple observation that the lighting geometry and lighting colour processes which are multiplicative in RGB space become additive in log RGB space. In a second stage we show that these additive effects can be removed using projection operators. Importantly, we show that in contrast to comprehensive normalisation, the new process is non-iterative. This is so because of the nature of projector operators in linear algebra. That is projectors are idempotent.

While experiments in this paper show that this normalisation can work well, it is based on the assumption that camera response is linear with respect to light intensity. However, in reality, responses are more typically non-linearly related to the intensity of the incident light for a number of different reasons. First, when coding image data it is usual for a power function to be applied to raw device responses. This power function (or gamma correction as it is commonly called) is applied to account for the fact that the output of CRT devices, on which images are typically displayed, is non-linearly related to the input signal. In the case of PC monitors this non-linearity is well modelled by a power function with an exponent (gamma) of 2.2. Storing images coded with a power function of 1/2.2 ensures that the images will display correctly on these devices. Unfortunately, systems can be calibrated such that their gamma is other than 2.2. For example Apple systems are typically calibrated to a gamma of 1.8. This implies that images of the same scene may differ according to the display system they are intended for. Furthermore, gamma correction is also commonly applied to change the relative contrast in an image. Applying a gamma of less than one tends to bring out detail in darker regions (at the expense of the lighter regions) and conversely a gamma larger than one is used to bring out detail in the highlights. Simply put, images of the same scene taken with different cameras can differ in terms of the gamma employed.

We address this issue in a second part of this paper in which we set out to remove the dependence of image colour on gamma whilst at the same time maintaining independence of illumination colour and geometry. To this end, we first perform a non-iterative log normalisation of the image data. We then observe that in log space x^γ maps to $\gamma \ln x$: the power function becomes multiplication by a scalar. It follows that by dividing each R, G and B by their respective standard deviations the γ will cancel and we obtain an illuminant geometry, colour, and gamma independent image representation.

The two new normalisation methods introduced in this paper are tested in the context of colour object recognition experiments. Here objects are imaged under different lighting conditions. The images are then normalised as described in the paper and we ask whether the remaining colour content suffices for recognising the object. Our experiments show that non-iterative comprehensive normalisations deliver almost perfect recognition on a moderately sized set of linear images. In addition, experiments on a set of non-linear images demonstrate that normalising for gamma in addition to lighting effects leads to a significant further improvement.

The rest of the paper is organised as follows: the basis of colour image formation is introduced and colour image normalisation methods are reviewed in Section 2. The new log-space comprehensive normalisations are presented in Section 3. In section 4 these normalisation procedures are evaluated in the context of a set of object recognition experiments. The paper concludes with a brief summary and discussion in Section 5.

2 Background

In order to develop the theory we will begin by adopting some widely used simplifying assumptions. However, as we develop the theory further some of these assumptions will be dropped. To begin let us assume that our imaging device is trichromatic and that the responses of the three sensors at a single pixel are denoted (R_i, G_i, B_i). Let us further assume that the device response is linear with respect to the incident light. That is, if we view a surface under a given light and then double the intensity of the light we expect a doubling in the recorded RGB values:

$$\begin{pmatrix} R_i \\ G_i \\ B_i \end{pmatrix} \rightarrow \begin{pmatrix} \rho_i R_i \\ \rho_i G_i \\ \rho_i B_i \end{pmatrix} \tag{1}$$

where ρ_i is a simple scalar. Note that this scalar has a subscript i indicating that all pixels can have their own individual brightness factors. That is, brightness changes, or lighting geometry, is a local phenomenon.

Changing the relative position of the light source with respect to the surface introduces shading. Assuming matte Lambertian reflectance and letting $\underline{n}$ denote surface normal and $\underline{e}$ the lighting direction then the power of the light striking a surface is proportional to the scalar $\underline{n}.\underline{e}$ (the vector dot-product). It follows then that a change in shading can also be described according to Equation(1). It is important to note that the Lambertian assumption is important here. Equation (1) cannot account for lighting geometry changes for highly specular surfaces.

Let us now consider a change in lighting colour (assuming lighting geometry is held fixed). In many circumstances, Equation (2), below, is a reasonable model of a change in illumination [27]:

$$\begin{pmatrix} R_i \\ G_i \\ B_i \end{pmatrix} \rightarrow \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \begin{pmatrix} R_i \\ G_i \\ B_i \end{pmatrix} \tag{2}$$

Furthermore, when Equation (2) is not a good model it is possible to linearly transform RGB so as to ensure that the model holds [9]. Equation (2) tells us that a change in illumination leads to a scalar change in all the responses within a single channel of the image. Responses in each colour channel change by a different scale factor a, b, and c. That is colour change is a global phenomenon affecting the whole image.

While during the image acquisition stage, device response is typically linearly related to incident light (as specified by Equation (1)), image data is typically non-linearly transformed prior to storage. Typically, this non-linearity can be well modelled by a power function transformation of the raw sensor responses. That is R_i, G_i, and B_i are raised to a fixed exponent which we refer to as *gamma* and denote γ:

$$\begin{pmatrix} R_i \\ G_i \\ B_i \end{pmatrix} \rightarrow \begin{pmatrix} R_i^\gamma \\ G_i^\gamma \\ B_i^\gamma \end{pmatrix} \tag{3}$$

Notice that the same gamma is applied to all pixels in the image. Now combining Equations (1), (2) and (3) we see that:

$$\begin{pmatrix} R_i \\ G_i \\ B_i \end{pmatrix} \rightarrow \begin{pmatrix} [a\rho_i R_i]^\gamma \\ [b\rho_i G_i]^\gamma \\ [c\rho_i B_i]^\gamma \end{pmatrix} \tag{4}$$

To simplify matters we incorporate γ into the scalars: $a' \rightarrow a^\gamma$, $b' \rightarrow b^\gamma$, $c' \rightarrow c^\gamma$ and $\rho_i' \rightarrow \rho_i^\gamma$ so that Equation (4) becomes:

$$\begin{pmatrix} R_i \\ G_i \\ B_i \end{pmatrix} \rightarrow \begin{pmatrix} a'\rho_i' R_i^\gamma \\ b'\rho_i' G_i^\gamma \\ c'\rho_i' B_i^\gamma \end{pmatrix} \tag{5}$$

Equations (1)-(5) tell us how RGB values change when one or more of lighting geometry, illuminant colour, and gamma changes. It is clear from these equations that one way to remove dependencies on these factors is to find algebraic transformations of RGBs such that the scalars ρ_i', a', b', c', and γ are canceled out.

For example, a change in light intensity as modelled by Equation (1) can be canceled out by applying a so called *chromaticity* normalisation:

$$\begin{aligned} r_i &= \frac{\rho_i R_i}{\rho_i R_i + \rho_i G_i + \rho_i B_i} \\ g_i &= \frac{\rho_i R_i}{\rho_i R_i + \rho_i G_i + \rho_i B_i}, \\ b_i &= \frac{\rho_i R_i}{\rho_i R_i + \rho_i G_i + \rho_i B_i} \end{aligned} \tag{6}$$

where the new co-ordinates r, g, and b are independent of ρ_i. We denote this chromaticity normalisation by a function C which takes as its argument an image $\mathcal{I}$ and returns a new argument such that each pixel in $\mathcal{I}$ is transformed according to Equation (6).

Next let us consider how we might cancel out dependence on illumination colour, as modelled by Equation (2). First, let $\mu(R)$ denote the mean red pixel value for an image:

$$\mu(R) = \frac{\sum_{i=1}^{N} R_i}{N} \tag{7}$$

where N is the number of pixels in the image: Under a change in illuminant colour (see Equation (2)) the mean becomes:

$$\mu(R)' = \frac{\sum_{i=1}^{N} aR_i}{N} = a\mu(R) \tag{8}$$

That is, the mean changes by the same scale factor a. Thus to cancel the effect of light colour on RGBs we can apply the following transformation:

$$R_i' = \frac{aR_i}{a\mu(R)}, \quad G_i' = \frac{bG_i}{b\mu(G)}, \quad B_i' = \frac{cB_i}{c\mu(B)} \tag{9}$$

Equation (9) is commonly referred to as a *grey-world* normalisation and we denote this normalisation by a function $G()$ such that the image $\mathcal{I}$ post gray world normalisation is denoted $G(\mathcal{I})$.

Equations (6) and (9) remove the effects of lighting geometry and illumination colour respectively but neither, by itself, suffices to remove the effect of both factors. Finlayson *et al* [8] defined a third normalisation which they called *Comprehensive normalisation*, which can remove both dependencies. It is defined as:

$$\begin{array}{lllll} 1. & \mathcal{I}_0 & = & \mathcal{I} & \text{Initialisation} \\ 2. & \mathcal{I}_{i+1} & = & G(C(\mathcal{I}_i)) & \text{Iteration step} \\ 3. & \mathcal{I}_{i+1} & = & \mathcal{I}_i & \text{Termination condition} \end{array} \tag{10}$$

That is, chromaticity normalisation and gray world normalisation are applied successively and repeatedly to an image until the resulting image converges to a fixed point.

Finlayson *et al* [8] proved that this iterative process always converges to a unique fixed point. Thus if $\mathcal{I}$ and $\mathcal{I}'$ denote images of the same scene where only the lighting conditions have changed then $CN(\mathcal{I}) = CN(\mathcal{I}')$ where $CN()$ denotes the comprehensive normalisation procedure as defined by Equation (10). As might be expected this method has been shown to outperform the individual normalisation functions $C()$ and $G()$ in object recognition experiments.

Yet despite the good performance of Comprehensive Normalisation procedure it suffers from a number of drawbacks which need to be addressed. First, is the fact that Comprehensive Normalisation is an iterative procedure. While it is proven that the method will always converge to a fixed point the number of iterations required to reach this fixed point is not defined, and further, from an implementation point of view it would be preferable to have a closed form, non-iterative procedure. In addition, underlying the procedure is the assumption that the image formation process is linear, whereas in practice, as we have discussed, the process is often non-linear. In particular, it is important that we account for the power function transform given by Equation (3). We address these shortcomings in the next section.

3 Non-iterative Comprehensive normalisation in Log-space

Equations (1) and (2) make it clear that in RGB space, as the lighting conditions change, the effect on a pixel is multiplicative. But suppose that rather than working in RGB space we work with the *log* of RGBs. Taking logarithms turns multiplications into additions so that our model of image formation becomes:

$$\begin{array}{lcl} \log(R_i) & & a'' + \rho_i'' + \gamma\log(R_i) \\ \log(G_i) & \rightarrow & b'' + \rho_i'' + \gamma\log(G_i) \\ \log(B_i) & & c'' + \rho_i'' + \gamma\log(B_i) \end{array} \tag{11}$$

where $a'' = \log a'$, $b'' = \log b'$, $c'' = \log c'$, and $\rho_i'' = \log \rho_i'$. Representing log RGBs by R', G' and B' we have:

$$\begin{pmatrix} R_i' \\ G_i' \\ B_i' \end{pmatrix} \rightarrow \begin{pmatrix} a'' \\ b'' \\ c'' \end{pmatrix} + \rho_i'' \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \gamma R_i' \\ \gamma G_i' \\ \gamma B_i' \end{pmatrix} \tag{12}$$

Equation (12) tells us that in log RGB space, lighting geometry changing only affects the length of the log RGB vector in the direction of $U = (1,1,1)$. That is, the directions orthogonal to $(1,1,1)$ are unaffected by brightness change. It follows that we can normalise a log RGB to remove brightness by projecting it onto the 2-dimensional space which is orthogonal to the line that is spanned by U. We can do this by applying some simple results from linear algebra. We define a 3×3 projection matrix Pr for the space spanned by U and a complementary projection matrix $Pr^{\perp} = [I - Pr]$ for the space which is orthogonal to the space spanned by U:

$$Pr = U^t(UU^t)^{-1}U = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \tag{13}$$

and so

$$Pr^{\perp} = I - Pr = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix} \tag{14}$$

where t denotes the matrix transpose operator and I denotes the 3×3 identity matrix. By definition these matrices have the property [17] that $Pr * (1,1,1)^t = (1,1,1)^t$ and $[I - Pr] * (1,1,1)^{=}(0,0,0)^t$: To project a log RGB onto the space orthogonal to U we simply multiply by the projection matrix $Pr^{\perp}$:

$$\begin{pmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} R_i' \\ G_i' \\ B_i' \end{pmatrix} = \begin{pmatrix} \frac{2R_i'}{3} - \frac{G_i'}{3} - \frac{B_i'}{3} \\ \frac{2G_i'}{3} - \frac{R_i'}{3} - \frac{B_i'}{3} \\ \frac{2B_i'}{3} - \frac{R_i'}{3} - \frac{G_i'}{3} \end{pmatrix} = \begin{pmatrix} R_i - \frac{R_i'+G_i'+B_i'}{3} \\ G_i' - \frac{R_i'+G_i'+B_i'}{3} \\ B_i' - \frac{R_i'+G_i'+B_i'}{3} \end{pmatrix} \tag{15}$$

That is, we can remove dependency on lighting geometry by subtracting the mean log response at a pixel (a "brightness" correlate) from each pixel.

The effect of illuminant colour can be removed in a similar way. However, rather than dealing with log RGB vectors we must operate on the vector of all log red (or green, or blue) responses. Considering all red responses we can write:

$$\begin{pmatrix} a'' + \gamma R_1' \\ a'' + \gamma R_2' \\ \vdots \\ a'' + \gamma R_N' \end{pmatrix} \rightarrow \begin{pmatrix} \gamma R_1' \\ \gamma R_2' \\ \vdots \\ \gamma R_N' \end{pmatrix} + a'' \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tag{16}$$

It follows that to remove dependence due to illumination colour we need to project responses onto the space orthogonal to the vector $(1,1,\cdots,1)^t$. This can be achieved similarly to the intensity normalisation, by defining a projection matrix Pc which projects onto the space spanned by $(1,1,\cdots,1)^t$ and its complement $Pc^{\perp}$ which projects onto the orthogonal space:

$$Pc = \begin{pmatrix} \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\ \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\ \vdots & \vdots & & \vdots \\ \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \end{pmatrix} \tag{17}$$

$$Pc^{\perp} = I - Pc = \begin{pmatrix} \frac{N-1}{N} & -\frac{1}{N} & \cdots & -\frac{1}{N} \\ -\frac{1}{N} & \frac{N-1}{N} & \cdots & -\frac{1}{N} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{N} & -\frac{1}{N} & \cdots & \frac{N-1}{N} \end{pmatrix} \tag{18}$$

Again, by inspecting the structure of the projection matrix, we can see that to implement this normalisation we only need to subtract the mean log red value from all log red pixel values and subtract the mean log green and log blue values from the log green and log blue pixel values.

To remove lighting geometry and illuminant colour both at the same time we just apply the projectors (14) and (18) consecutively to the log image. This operation is easy to write down mathematically if we think of an N pixel image as an $N \times 3$ matrix of log RGBs. Denoting this image $Y\gamma$, we can write down an explicit equation for the application of the lighting geometry and light colour normalisations:

$$\mathcal{Y} = (I - Pc)Y\gamma(I - Pr) \tag{19}$$

here $\mathcal{Y}$ represents the normalised image. Though, it is important to remember that to implement (19) we simply subtract row means from rows and then column means from columns.

From projection theory, we know that matrix $[I - Pr]$ and $[I - Pc]$ are both idempotent. That is to say that $[I - Pr][I - Pr] = [I - Pr]$ and $[I - Pc][I - Pc] = [I - Pc]$. It follows then that removing shading or light colour once, removes it completely:

$$\mathcal{Y} = (I - Pc)(I - Pc)Y\gamma(I - Pr)(I - Pr) = (I - Pc)Y(I - Pr) \tag{20}$$

Thus far we have not attempted to remove the dependence of γ nor have we considered its effect on our approach to removing lighting geometry and colour. In fact, because γ term is a multiplicative factor in log space, it is easy to see that it does not affect the operations of the projectors involved in the procedure. Therefore the dependencies due to lighting geometry and illuminant colour are removed regardless of γ.

But our normalised image still depends on gamma, and so, to make the procedure fully comprehensive we would like to remove the effect of gamma. By definition the mean of all the elements in $\mathcal{Y}$ must be zero (if the mean of the rows and columns are individually 0 then the overall mean must also be 0). It follows then that γ cannot be removed by dividing by the mean. Rather we must use a second order statistic. It is easy to show that the variance of all the elements in $\mathcal{Y}$ can be calculated as:

$$\sigma^2(\mathcal{Y}) = \frac{\text{trace}(\mathcal{Y}^t\mathcal{Y})}{3N} \tag{21}$$

where $\mathcal{Y}$ has N rows and 3 columns (there are N pixels in the image) and trace() is the sum of the diagonal elements of a matrix. It follows that gamma can be removed by dividing by the standard deviation.

$$\frac{\mathcal{Y}}{\sigma(\mathcal{Y})} = \frac{\gamma Y}{\gamma\sigma(Y)} = \frac{Y}{\sigma(Y)} \tag{22}$$

In summary we take the log of the RGB image. At each pixel we subtract the pixel mean (of the log R, G and B responses). We then subtract the mean of all the resulting red responses from each pixel and the mean green and blue channel responses for the green and blue colour channels. The result is an image independent of the light colour and lighting geometry. Further dividing by the standard deviation of the resultant image renders the representation independent of gamma.

4 Object Recognition Experiments

To test the efficacy of the normalisations proposed above we conducted object recognition experiments. The experiment consists of first building an image database such that each image represents an object. Then we select query images and try to match the query to the database images. That is we compare the query image to each of the images in the database and the object corresponding to the database image which is closest to the query image is taken to be the query object. We do this for many different query images and the number of times we obtain a correct match (that is, we correctly identify the object in the query image) is a measure of how well we do in the experiment.

Following Swain et al [23] we represent images by their colour histograms: each bin of the histogram corresponds to a particular RGB value and the bin's entry is the proportion of pixels in an image which have this RGB. Clearly, images which differ in terms of lighting geometry, illuminant colour, and/or gamma will have different colour histograms. So if no account is taken of these factors we might expect that poor matching performance will result. To overcome this problem we apply one or other of the normalisations set out above to the images prior to constructing the colour histogram. If the normalisations successfully account for illumination and gamma we would expect the resulting histograms to be invariant to these factors and thus we should achieve accurate matching.

In a first series of experiments we used images which are linear with respect to incident light so it is unnecessary in this case to factor out gamma dependencies. We conducted colour indexing experiments on these each images applying each of the normalisations proposed in this paper: non-iterative log comprehensive normalisation and log-gamma normalisation. Since the images are linear there is in no need to apply a gamma normalisation but we did so because we were interested in whether the log gamma normalisation might degrade matching performance (since the more dependencies we cancel out the less information is left for matching). For comparison purposes we also tested comprehensive normalisation, no normalisation (that is, histograms are constructed from raw RGB values) and the grey-world normalisation.

This first experiment used images from three different datasets: Swain and Ballard's [23] image set (66 database images, 31 queries), the *legacy* Simon Fraser image set [13] (13 database images, 26 queries) and the Berwick and Lee image set[4] (8 database images and 9 queries). The Simon Fraser images and the Berwick and Lee images include images of the same object under different illuminants, whereas the Swain images are all taken under a single, fixed light. We report results for each method applied to each dataset separately and also for each method applied to a composite dataset formed by grouping together the images from all three datasets. To more thoroughly test the methods under changing illumination we conducted a second experiment using a dataset of 20 objects imaged under 11 different lights [1] [20].

Tables 1, 2, 3, 4, and 5 summarise the indexing performance of the various methods for the first two experiments. Performance is measured using the average match percentile match [23]. If the closest database histogram to the query is the correct answer (both corresponding images are of the same object) then the correct answer has rank 1. If the correct answer is the kth closest then the correct answer has rank k. The corresponding percentile is calculated as $\frac{N-k}{N-1}$, where N is the number of images in the model dataset. We also tabulate the % of matches of rank 1, 2, and above 2 together with the worst case rank statistic.

The results for the datasets containing objects under different illuminants demonstrate the need for some kind of normalisation: when images are indexed on raw RGB, poor performance results. When the three datasets are considered separately all of the tested normalisations perform well with the grey-world normalisation only slightly behind the comprehensive normalisations discussed in this paper. This suggests that normalising for a change of illuminant colour accounts for most of the differences in these images. The results for Swain's dataset are also important since they show that even when illumination doesn't change the normalised images still provide good indexing performance: that is sufficient information is retained post-normalisation to achieve good recognition.

The results for the composite dataset and for the larger Simon Fraser dataset containing a wide range of illuminants more clearly demonstrate the power of the comprehensive normalisations proposed in this paper. Once again it is clear that accounting for illumination conditions is important (as evidenced by the fact that indexing results for raw RGB are poor) but now it is also clear that the grey-world normalisation is not sufficient. However applying any of the more comprehensive normalisation leads to almost perfect indexing performance. Furthermore discounting gamma does not result in a significant fall in performance.

In a third set of experiments we set out to test the gamma normalisation procedure. To this end we used a dataset [11] comprising 28 designs captured under a variety of devices and illuminants. The colour response of these devices is highly non-linear and it has been shown [11] that a large variety of existing "linear" normalisations do not support image indexing. Here we test whether accounting for gamma in addition to light colour and light geometry will lead to better indexing.

Our evaluation of performance in this experiment differs slightly from the first two sets of experiments. First we choose images of the 28 designs captured with a camera under a fixed illuminant as our database. Next we take all images in the dataset as our query (test) set: this set consists of 28 designs captured with four different cameras under three different illuminants and also with two different scanners: a total of $4*3*28+2*28 = 392$ images. Prior to indexing we introduced an ordering into this dataset. Specifically we asked how well a diagonal model of illuminant change together with a gamma correction accounted for the differences in colour distributions for corresponding designs in the database and test set. So, given

[1] This dataset is available from Simon Fraser University: http://www.www.cs.sfu.ca/~colour/image_db/

an image in the test set and its corresponding image in the database we find, by an optimisation process, the parameters in Equation 5 which best transform the query image to the database image. The error in this transformation then gives us a measure of how well the model holds for this test image. We repeat this process for all test images and order the images according to this error. This ordering allowed us to choose the images in the test set for which a diagonal gamma colour model (Equation (5)) worked with a certain degree of accuracy. For instance we found that the diagonal gamma model of colour change accounted for about 20% of the test set with an error of less than 10%.

With this ordering defined we next ran a colour indexing experiment as described previously: each image in the test set is matched to the database using a colour histogram comparison where histograms are formed from normalised images. Now, rather than looking at matching results for the whole set of query images we investigate match performance as a function of how well the test images correspond to the diagonal-gamma model in Equation 5. So, for example we can investigate match performance for the 10% of images for which the model works best. These results are summarised in Figure 1. The x-axis of this figure denotes the proportion of test images for which matching performance is investigated and the corresponding match performance (average match percentile) is shown on the y-axis. So, for example we see that if we look at 0.5 on the x-axis (corresponding to the 50% of the test set for which a diagonal gamma model works best) the log gamma normalisation delivers a percentile match of about 0.93 compared with 0.88 and 0.82 respectively for the comprehensive and log-normalisations. From Figure 1 it is clear that the non-iterative plus gamma normalisation works significantly better than either the log normalisation or comprehensive normalisation on non-linear dataset. By incorporating gamma into the invariant model we have improved matching performance. However, the overall match percentiles can be quite low. If we have a database where the average match percentile is 0.95 then this means that the correct answer is in the top 5% of matches. This might be a tolerable number for some applications. However, percentiles of 0.9 or even 0.8 are indicative of rather poor matching performance. From the figure we see that at the 0.95 level that comprehensive normalisation delivers adequate matching only for about 10% of the test set. This increases to about 18% for the log normalisation and about 33% for the log-gamma procedure.

The import of this is that many of the test images cannot be normalised using the non-iterative plus gamma normalisation (in that the normalised images do not form a stable cue for indexing). The reason for this is readily understood at an average percentile of 0.95 the diagonal gamma model of image formation models the data with an error of 20%. That is, if we find by optimisation the best diagonal matrix and best gamma that maps the database image colours as close as possible to the test images there is a residual error of 20%. Empirically, an error of above 20% is too high to afford good matching.

5 Discussion

We have presented two new normalisation methods. The first cancels image dependence on lighting geometry and illuminant colour together at the same time. In addition, and in contrast to previous methods, this normalisation is a simple, non-iterative method. The second normalisation procedure extends the first to additionally cancel the effect of power functions which are typically applied to captured images. Again the method is a simple non-iterative procedure. Experimental results showed that both normalisation schemes performed similarly when images are captured with linear response cameras but that significant performance improvement is obtained with the second normalisation when images have non-linearities.

While the experimental results are encouraging there are a number of possible limitations which the methods presented here share with other normalisation techniques. First, there is a degenerate case in which the method will be unable to distinguish between objects. Specifically, if two or more objects consist of a single colour then removing their mean colour will render them grey and thus in terms of the colour histogram, identical. In practice this situation will arise infrequently though it draws attention to the fact that indexing on colour information alone is unlikely to ever result in perfect recognition.

A second and more important issue is the fact that the theory developed in the paper relies (like grey-world normalisation and others) on a diagonal model of illumination change. It is possible that significant deviations from this model would cause the method to fail. However, the experimental results are obtained on images captured under a wide range of common illuminants and the good performance achieved suggests that the model is valid in most practical situations.

Finally, extending previous normalisation approaches to also include gamma normalisation generalises previous work in the sense in which it is founded on a model which more accurately models image formation in a camera. However, the results of the third experiment suggest that this model is still not sufficiently general to fully account for the image formation process. It might be expected that a further performance improvement will be obtained by more accurately modelling and accounting for the factors of image formation.

References

[1] Donald A.Adjeroh and M. C.Lee. *On Ratio-Based Color Indexing.* IEEE transaction on image processing, vol(10), 2001.

[2] J Bach, C. Fuller, A. Gupta, and A. Hampapur. *The virage image search engine: An open framework for image management.* In SPIE Conf. on Storage and Retrieval for Image and Video Databases, volume 2670, pages 76-87, 1996.

[3] J. Berens, G.D. Finlayson, and G. Qiu. Image indexing using compressed colour histograms. *Vision, Image and Signal Processing*, 147(4):349–355, 2000.

[4] D. Berwick and S. W.Lee. *A chromaticity space for specularity for specularity-, illumination color- and illumination pose-invariant 3-d object recognition.* ICCV98, Session2.3, 1998.

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

[6] G. D. Finlayson. *Colour Object Recognition.* Simon Fraser Univ. School of Computing Science, 1992.

[7] G. D. Finlayson. *Color in perspective.* IEEE Trans. PSMI, 18(10):1034-1038, 1996.

[8] G. D.Finlayson, Bernt Schiele, and James L.Crowley. *Comprehensive Colour Image Normalisation.* In 5th European Conference on Computer Vision,page 475-490, 1998.

[9] G. D.Finlayson, M S.Drew, and B. V.Funt. *Spectral sharpening: Sensor transformations for improved color constancy.* J.Optical Society of America A, 11(5):1553-1563, 1994.

[10] M. D.Zmura and G. Iverson. *Color constancy, basic theory of two-stage linear recovery of spectral description for lights and surfaces.* JOSA, 10:2148-2165, 1993.

[11] G.D Finlayson and G. Schaefer. Colour indexing across devices and viewing conditions. *2nd Int. Workshop on Content-Based Multimedia Indexing,Brescia*, pages 215–221, 2001.

[12] G.D. Finlayson and R. Xu. *Non-Iterative Comprehensive Normalisation.* First European Conference on Color in Graphics, Image and Vision(CGIV), 2002.

[13] B.V. Funt, K. Barnard, and L. Martin. Is machine colour constancy good enough. In *The Fifth European Conference on Computer Vision (Vol II)*, pages 445–459. European Vision Society, 1998.

[14] B.V. Funt and G.D. Finlayson. Color constant color indexing. *PAMI*, 17(5):522–529, May 1995.

[15] Ron Gershon, Allan D.Jepson, and John K.Tsotsos. *From [R,G,B] to Surface Reflectance:Computing Color Constant Descriptors in Images.* In International Joint Conference on Artificial Intelligence, pages 755-758, 1987.

[16] G. Gevers and A. W.M.Smeulders. *Color based object recognition.* Pattern Recognition,32:453-464, 1999.

[17] G.H. Golub and C.F. van Loan. *Matrix Computations.* John Hopkins U. Press, 1983.

[18] E H.Land. *The retinex theory of color constancy.* Scientific American, pages 108-129, 1977.

[19] M. J.Swain and D. H.Ballard. *Color indexing.* International Journal of Computer Vision,7(11):11-32, 1991.

[20] Brian Funt Kobus Barnard, Lindsay Martin and Adam Coath. *A Data Set for Colour Research and Application, Volume 27, Number 3, pp. 147-151,2002.*

[21] S. Lin and S. W. Lee. *Using chromaticity distributions and eigenspace analysis for pose-, illumination- and specularity-invariant recognition of 3d object.* In CVPR97, pages 426-431, 1997.

[22] W. Niblack and R. Barber. *The QBIC project:Querying images by content using color, texture and shape.* In Storage and Retrieval for Image and Video Databases I, volume 1908 of SPIE Proceedings Series, 1993.

[23] M.J. Swain and D.H.. Ballard. Color indexing. *International Journal of Computer Vision*, 7(11):11–32, 1991.

[24] B. V.Funt, K. Barnard, and L. Martin. *Is colour constancy good enough?* In 5th European Conference on Computer Vision,pages 445-459, 1998.

[25] B. V.Funt and G D.Finlayson. *Color constant color indexing.* IEEE Trans. Pattern Recognition and Machine Intelligence,17(5):522-529, 1991.

[26] R. W.G. *The Reproduction of Color.* Fountain Press, 5th edition, 1995.

[27] J.A. Worthey and M.H. Brill. Heuristic analysis of von Kries color constancy. *Journal of The Optical Society of America A*, 3(10):1708–1712, 1986.

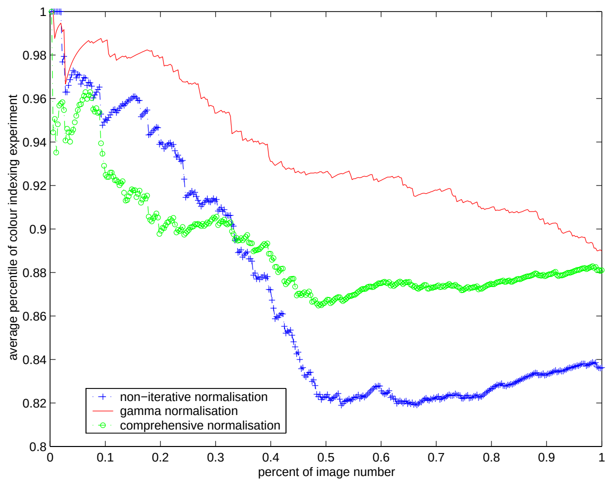

Figure 1: This figure shows performance of each of the three normalisation methods(log gamma, non-iterative, comprehensive normalisation) on the design dataset. The x-axis corresponds to the proportion of query images for which average match percentile (y-axis) is investigated. The query set is sorted according to how well the images conform to a diagonal-gamma model. See text for further details.

Methods	Av.percentile	rank 1	rank 2	rank>2	worst rank
nothing	69.84	33.33	33.33	33.33	8 out of 8
gray-world	98.41	88.89	11.11	0	2 out of 8
non-iterative	100	100	0	0	1 out of 8
comprehensive	100	100	0	0	1 out of 8
log gamma	100	100	0	0	1 out of 8

Table 1: Indexing Performance of Lee and Berwick Dataset (ranks are % of the dataset)

Methods	Av.percentile	rank 1	rank 2	rank>2	worst rank
nothing	98.54	93.33	3.33	3.33	9 out of 66
gray-world	98.77	83.33	13.33	3.34	21 out of 66
non-iterative	99.59	83.33	13.33	3.34	5 out of 66
comprehensive	98.46	86.67	3.33	10	9 out of 66
log gamma	98.05	56.67	26.67	16.66	17 out of 66

Table 2: Indexing Performance of Swain's Dataset (ranks are % of the dataset)

Methods	Av.percentile	rank 1	rank 2	rank>2	worst rank
nothing	73.08	30.77	15.38	53.85	13 out of 13
gray-world	98.08	84.62	11.54	3.84	4 out of 13
non-iterative	98.40	84.62	11.54	3.84	2 out of 13
comprehensive	98.08	84.62	11.54	3.84	4 out of 13
log gamma	97.76	84.62	7.69	7.69	4 out of 13

Table 3: Indexing Performance of Simon Fraser Dataset (ranks are % of the dataset)

Methods	Av.percentile	rank 1	rank 2	rank>2	worst rank
nothing	93.92	58.46	6.15	35.38	65 out of 87
gray-world	92.11	56.92	9.23	33.85	72 out of 87
non-iterative	99.71	95.38	1.15	3.47	4 out of 87
comprehensive	99.71	92.31	2.3	5.39	3 out of 87
log gamma	98.60	84.62	4.62	10.76	19 out of 87

Table 4: Indexing Performance of Composite Dataset (ranks are % of the dataset)

Methods	Av.percentile	rank 1	rank 2	rank>2	worst rank
nothing	73.50	28	12	60	20 out of 20
gray-world	94.21	72.50	10.5	17	20 out of 20
non-iterative	99.13	91.00	3.5	5.5	5 out of 20
comprehensive	98.84	91.00	3.5	5.5	16 out of 20
log gamma	97.61	84.50	6	9.5	10 out of 20

Table 5: Indexing Performance on Large Simon Fraser Dataset (ranks are % of the dataset)