Comprehensive Colour Image Normalization

Graham D. Finlayson[1], Bernt Schiele[2], and James L. Crowley[3]

[1] The Colour & Imaging Institute
The University of Derby
United Kingdom

[2] MIT Media Lab
Cambridge MA
USA

[3] INRIA Rhônes Alpes
38330 Montbonnot
France

Abstract. The same scene viewed under two different illuminants induces two different colour images. If the two illuminants are the same colour but are placed at different positions then corresponding rgb pixels are related by simple scale factors. In contrast if the lighting geometry is held fixed but the colour of the light changes then it is the individual colour channels (e.g. all the red pixel values or all the green pixels) that are a scaling apart. It is well known that the image dependencies due to lighting geometry and illuminant colour can be respectively removed by normalizing the magnitude of the rgb pixel triplets (e.g. by calculating chromaticities) and by normalizing the lengths of each colour channel (by running the 'grey-world' colour constancy algorithm). However, neither normalization suffices to account for changes in both the lighting geometry and illuminant colour.

In this paper we present a new **comprehensive image normalization** which removes image dependency on lighting geometry and illumination colour. Our approach is disarmingly simple. We take the colour image and normalize the rgb pixels (to remove dependence on lighting geometry) and then normalize the r, g and b colour channels (to remove dependence on illuminant colour). We then repeat this process, normalize rgb pixels then r, g and b colour channels, and then repeat again. Indeed we repeat this process until we reach a stable state; that is reach a position where each normalization is idempotent. Crucially this iterative normalization procedure always converges to the same answer. Moreover, convergence is very rapid, typically taking just 4 or 5 iterations.

To illustrate the value of our "comprehensive normalization" procedure we considered the object recognition problem for three image databases that appear in the literature: Swain's database, the Simon Fraser database, Sang Wok Lee's database. In all cases, for recognition by colour distribution comparison, the comprehensive normalization improves recognition rates (the results are near perfect and in all cases improve on results reported in the literature). Also recognition for the composite database (comprising almost 100 objects) is also near perfect.

1 Introduction

The light reaching our eye is a function of surface reflectance, illuminant colour and lighting geometry. Yet, the colours that we perceive depend almost exclusively on surface reflectance; the dependencies due to lighting geometry **and** illuminant colour are removed through some sort of image normalization procedure. As an example, the white page of a book looks white whether viewed under blue sky or under artificial light and remains white, independent of the position of the light source. While analogous normalizations exist in computer vision for discounting lighting geometry **or** illuminant colour there does not exist a normalization which can do both, together at the same time. Yet, such a **comprehensive normalization** is clearly needed since both lighting geometry and illuminant colour can be expected to change from image to image. A comprehensive normalization is developed in this paper.

Image normalization research in computer vision generally proceeds in two stages and we will adopt the same strategy here. First, the physics of image formation are characterized and the dependency due to a given physical variable is made explicit. In a second stage, methods for removing this dependency (that is, canceling dependent variables) are developed. As an example of this kind of reasoning, it is well known that, assuming a linear camera response, if light intensity is scaled by a factor s then the image scales by the same factor: each captured (r, g, b) pixel becomes (sr, sg, sb). Relative to this simple physical model it is easy to derive a normalization procedure which is independent of the intensity of the viewing illuminant:

$$\frac{r}{r+g+b}, \frac{g}{r+g+b}, \frac{b}{r+g+b} \tag{1}$$

The normalized image colours are sometimes represented using only the *chromaticities* $\frac{r}{r+g+b}$ and $\frac{g}{r+g+b}$ (since $\frac{b}{r+g+b} = \frac{1-r-g}{r+g+b}$).

The normalization shown in (1) is well used, and well accepted, in the computer vision literature (e.g. [SW95,CB97,MMK95,FDB91,Hea89]) and does an admirable job of rendering image colours independent of the power of the viewing illuminant. As we shall see later, lighting geometry in general (this includes the notions of light source direction and light source power) affects only the magnitude of a captured rgb and so the normalization shown in (1) performs well in diverse circumstances.

Dependency due to illumination colour is also very simple to model (subject to certain caveats which are explored later). If (r_1, g_1, b_1) and (r_2, g_2, b_2) denote camera responses corresponding to two scene points viewed under one colour of light then $(\alpha r_1, \beta g_1, \gamma b_1)$ and $(\alpha r_2, \beta g_2, \gamma b_2)$ denote the responses induced by the same points viewed under a different colour of light[WB86] (the red, green and blue colour channels scale by the factors α, β and γ). Clearly, it is easy to derive algebraic expressions where α, β and γ cancel:

$$\left(\frac{2r_1}{r_1+r_2}, \frac{2g_1}{g_1+g_2}, \frac{2b_1}{b_1+b_2}\right), \left(\frac{2r_2}{r_1+r_2}, \frac{2g_2}{g_1+g_2}, \frac{2b_2}{b_1+b_2}\right) \tag{2}$$

The two pixel case summarized in (2) naturally extends to N-pixels: the denominator term becomes the sum of all pixels and numerators are scaled by N. Notice that after normalization, the mean image colour maps to (1,1,1); that is, to 'grey'. For this reason, Equation (2) is sometimes called 'grey-world' normalization[Hun95,GJT88,Buc80].

Unfortunately, neither normalization (1) or (2) suffices to remove dependency on both lighting geometry and illuminant colour. To see that this is so, it is useful to step through a worked example. Let $(s_1\alpha r_1, s_1\beta g_2, s_1\gamma b_1)$ and $(s_2\alpha r_2, s_2\beta g_2, s_2\gamma b_2)$ denote image colours corresponding to two scene points where (s_1, s_2) and (α, β, γ) are scalars that model lighting geometry and illuminant colour respectively. Under lighting geometry normalization, Equation (1), the pixels become:

$$\begin{array}{l} (\frac{\alpha r_1}{\alpha r_1+\beta g_1+\gamma b_1}, \frac{\beta g_1}{\alpha r_1+\beta g_1+\gamma b_1}, \frac{\gamma b_1}{\alpha r_1+\beta g_1+\gamma b_1}) \ , \\ (\frac{\alpha r_2}{\alpha r_2+\beta g_2+\gamma b_2}, \frac{\beta g_2}{\alpha r_2+\beta g_2+\gamma b_2}, \frac{\gamma b_2}{\alpha r_2+\beta g_2+\gamma b_2}) \end{array} \tag{3}$$

and under illuminant colour normalization, Equation (2):

$$(\frac{2s_1r_1}{s_1r_1+s_2r_2}, \frac{2s_1g_1}{s_1g_1+s_2g_2}, \frac{2s_1b_1}{s_1b_1+s_2b_2}) \ , \ (\frac{2s_2r_2}{s_1r_1+s_2r_2}, \frac{2s_2g_2}{s_1g_1+s_2g_2}, \frac{2s_2b_2}{s_1b_1+s_2b_2}) \tag{4}$$

In both cases only some of the dependent variables cancel. This is unsatisfactory since both lighting geometry and illuminant colour will change from image to image.

Lin an Lee[LL97] proposed that this problem (cancelation failure) could be solved using normalized colour distribution manifolds. In their method images are normalized for lighting geometry and the variation due to illuminant colour is modelled explicitly. They show that the lighting geometry normalized distribution of colours in a scene viewed under all illuminant colours occupies a continuous manifold in distribution space. In later work by Berwick and Lee[BL98] this manifold is represented implicitly. Here a pair of lighting geometry normalized image colour distributions are defined to be the same if they can be 'matched' by a shift in illuminant colour; this matching effectively reconstructs the manifold 'on the fly'. Unfortunately both these solutions incur a substantial computational overhead. A high dimensional manifold is, at the best of times, unwieldy and implies costly indexing (i.e. to discover if a distribution belongs to a given manifold). Similarly the distribution matching solution, which operates by exhaustive distribution correlation, is very expensive.

In this paper we develop a new comprehensive normalization which can normalize away variation due to illuminant colour **and** lighting geometry together at the same time. Our approach is simplicity itself. We take an input image and normalize for lighting geometry using Equation (1). We then normalize for illuminant colour using Equation (2). We then iterate on this theme, successively normalizing away lighting geometry and light colour until each normalization step is idempotent.

We prove two very important results. First, that this process always converges. Second, that the convergent image is unique: the same scene viewed under any lighting geometry and under any illuminant colour has the same comprehensively normalized image. We also found that convergence is very rapid, typically taking just 4 or 5 iterations.

To illustrate the power of the comprehensive normalization procedure we generated synthetic images of a yellow/grey wedge viewed under white, blue and orange coloured lights which were placed at angles of 45^o (close to the surface normal for grey), 90^o (halfway between both surfaces) and 135^o (close to the normal of yellow). The image capture conditions are illustrated at the top of Figure 1 (a blue light at 80^o is shown). The 9 generated synthetic images are shown at the bottom of the figure together with corresponding normalized images. Lighting geometry normalization (Equation (1)) suffices to remove variation due to lighting position but not illuminant colour. Conversely, illuminant colour normalized images (Equation 2) are independent of light colour but depend on lighting position. Only the comprehensive normalization suffices to remove variation due to lighting geometry and illuminant colour.

Examples of the comprehensive normalization acting on real images are shown in Figure 2. The top two images are of the same object viewed under a pair of lighting geometries and illuminant colours. Notice how different the images appear. After comprehensive normalization the images, shown underneath, are almost the same. This experiments is repeated on a second image pair with similar results[1].

As a yet more rigorous test of comprehensive normalization we carried out several object recognition experiments using real images. We adopted the recognition paradigm suggested by Swain and Ballard[SB91] (which is widely used e.g. [SO95], [NB93] and [LL97]) where objects are represented by image colour distributions (or in the case of our experiments by the distribution of colours in comprehensively normalized images). Recognition proceeds by distribution comparison: query distributions are compared to object distributions stored in a database and the closest match identifies the query.

For the image databases of Swain (66 database objects, 31 queries), Chatterjee (13 database, 26 queries), Brewster and Lee (8 objects and 9 queries) and a composite set 87 objects and 67 queries, comprehensive normalization facilitated almost perfect recognition. For the composite database all but 6 of the objects are correctly identified and those that are not are identified in second place. This performance is quite startling in its own right (it is a large database compiled by a variety of research groups). Moreover, recognition performance surpasses, by far, that supported by the lighting geometry or illuminant colour normalizations applied individually.

In section 2 of this paper we discuss colour image formation and derive the normalizations shown above in equations (1) and (2). The comprehensive normalization is presented in section 3 together with proofs of uniqueness and

[1] All four input images shown in Figure 2 were taken by Berwick and Lee[BL98]

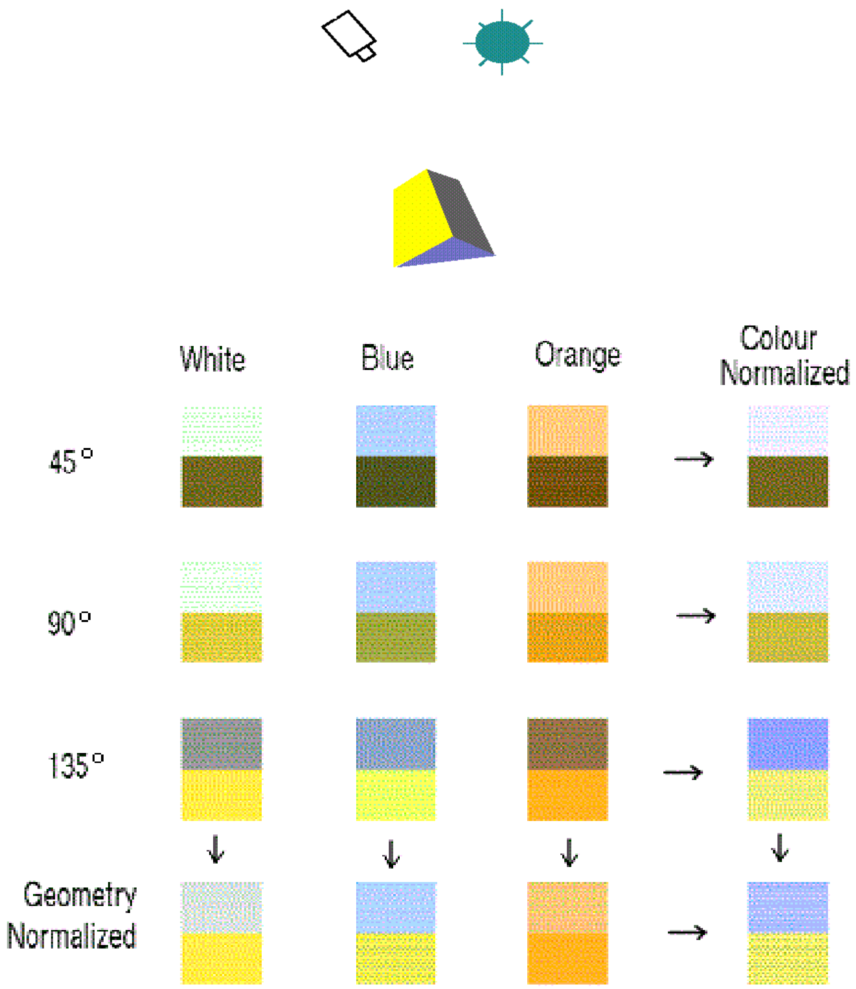

Fig. 1. A yellow/grey wedge, shown top of figure, is imaged under 3 lighting geometries and 3 light colours. The resulting 9 images comprise the 3 × 3 grid of images shown top left above. When illuminant colour is held fixed, lighting geometry normalization suffices (first 3 images in the last row). For fixed lighting geometry, illuminant colour normalization suffices (top 3 images in the last column). The comprehensive normalization removes the effects of both illuminant colour and lighting geometry (the single image shown bottom right)

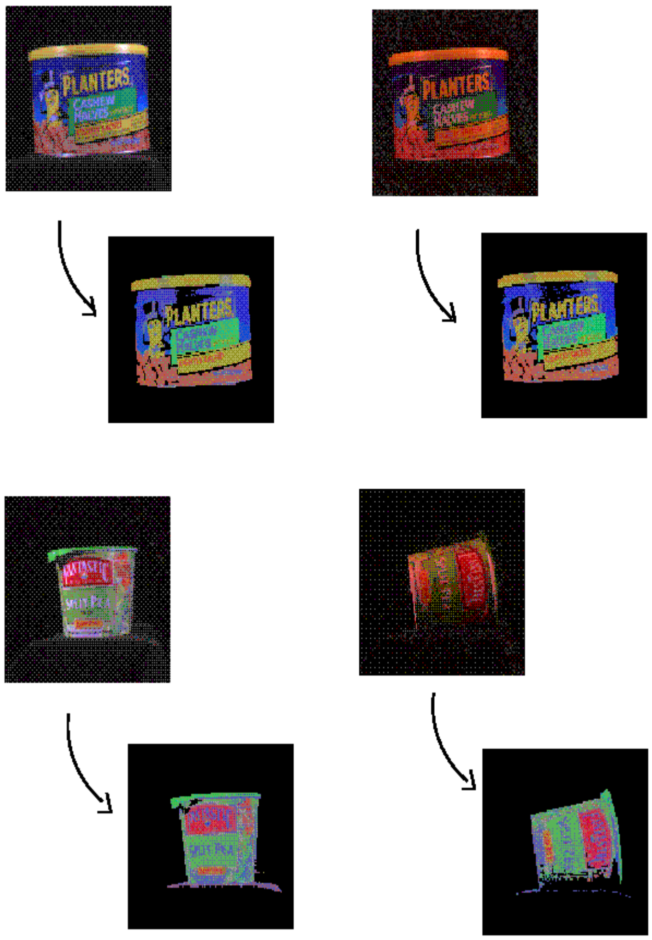

Fig. 2. A peanut container is imaged under two different lighting geometries and illuminant colours (top of figure). After comprehensive normalization the images appear the same (2nd pair of images). A pair of 'split-pea' images (third row) are comprehensively normalized (bottom pair of images). Notice how effectively comprehensive normalization removes dependence illuminant colour and lighting geometry

convergence. In section 4 the object recognition experiments are described and results presented. The paper finishes with some conclusions in section 5.

2 Colour Image Formation

The light reflected from a surface depends on the spectral properties of the surface reflectance and of the illumination incident on the surface. In the case of Lambertian surfaces (the only ones we consider here), this light is simply the product of the spectral power distribution of the light source with the percent spectral reflectance of the surface. Assuming a single point source light, illumination, surface reflection and sensor function, combine together in forming a sensor response:

$$\underline{p}^{\hat{x},E} = \underline{e}^x.\underline{n}^x \int_\omega S^x(\lambda)E(\lambda)\underline{F}(\lambda)d\lambda \tag{5}$$

where λ is wavelength, $\underline{p}$ is a 3-vector of sensor responses (*rgb* pixel value), $\underline{F}$ is the 3-vector of response functions (red-, green and blue- sensitive), E is the illumination striking surface reflectance S^x at location x. Integration is over the visible spectrum ω. Here, and throughout this paper, underscoring denotes vector quantities. The light reflected at x, is proportional to $E(\lambda)S^x(\lambda)$ and is projected onto $\hat{x}$ on the sensor array. The precise power of the reflected light is governed by the dot-product term $\underline{e}^x.\underline{n}^x$. Here, $\underline{n}^x$ is the unit vector corresponding to the surface normal at x and $\underline{e}^x$ is in the *direction* of the light source. The length of $\underline{e}^x$ models the power of the incident light at x. Note that this implies that the function $E(\lambda)$ is actually constant across the scene. Substituting $\underline{q}^{x,E}$ for $\int_\omega S^x(\lambda)E(\lambda)\underline{F}(\lambda)$ allows us to simplify (5):

$$\underline{p}^{\hat{x},E} = \underline{q}^{x,E}\underline{e}^x.\underline{n}^x \tag{6}$$

It is now understood that $\underline{q}^{x,E}$ is that part of a scene that does not vary with lighting geometry (but does change with illuminant colour). Equation (6), which deals only with point-source lights is easily generalized to more complex lighting geometries. Suppose the light incident at x is a combination of m point source lights with lighting direction vectors equal to $\underline{e}^{x,i}$ $(i = 1, 2, \cdots, m)$. In this case the the camera response is equal to:

$$\underline{p}^{\hat{x},E} = \underline{q}^{x,E}\sum_{i=1}^{m}\underline{e}^{x,i}.\underline{n}^x \tag{7}$$

Of course, all the lighting vectors can be combined into a single *effective* direction vector (and this takes us back to Equation (6)):

$$\underline{e}^x = \sum_{i=1}^{m}\underline{e}^{x,i} \Rightarrow \underline{p}^{\hat{x},E} = \underline{q}^{x,E}\underline{e}^x.\underline{n}^x \tag{7a}$$

Equation (7) conveys the intuitive idea that the camera response to m light equals the sum of the responses to each individual light. Simple though (7) is, it suffices to model extended light sources such as fluorescent lights[Pet93].

Since we now understand the dependency between camera response and lighting geometry, it is a scalar relationship dependent on $\underline{e}^x.\underline{n}^x$, it is a straightforward matter to normalize it away:

$$\frac{\underline{p}^{\hat{x},E}}{\sum_{i=1}^{3} p_i^{\hat{x},E}} \;=\; \frac{\underline{q}^{x,E}\underline{e}^x.\underline{n}^x}{\underline{e}^x.\underline{n}^x\sum_{i=1}^{3} q_i^{x,E}} \;=\; \frac{\underline{q}^x}{\sum_{i=1}^{3} q_i^{x,E}} \tag{8}$$

When $\underline{p}^{\hat{x},E} = (r,g,b)$ then the normalization returns: $(\frac{r}{r+g+b}, \frac{g}{r+g+b}, \frac{b}{r+g+b})$.

It is useful to view the dynamics of this normalization in terms of a complete image. Let us place the N image rgb pixels in rows of an $N \times 3$ image matrix $\mathcal{I}$. It is clear that (8) scales the rows of $\mathcal{I}$ to sum to one. The function $R()$ row-normalizes an image matrix according to (8):

$$R(\mathcal{I})_{i,j} \;=\; \frac{\mathcal{I}_{i,j}}{\sum_{k=1}^{3}\mathcal{I}_{i,k}} \tag{8a}$$

Here, and henceforth, a double subscript $_{i,j}$ indexes the ijth element of a matrix.

Let us now consider the effect of illuminant colour on the rgbs recorded by a camera. Here we hold lighting geometry, the vectors $\underline{e}^x$, fixed. To simplify matters still further camera sensors are delta functions: $F(\lambda) = \delta(\lambda - \lambda_i)$ $(i = 1,2,3)$. Under $E(\lambda)$ the camera response is equal to:

$$p_i^{\hat{x},E} \;=\; \underline{e}^x.\underline{n}^x\int_\omega S^x(\lambda)E(\lambda)\delta(\lambda-\lambda_i)d\lambda \;=\; \underline{e}^x.\underline{n}^x S^x(\lambda_i)E(\lambda_i) \tag{9}$$

and under $E_1(\lambda)$:

$$p_i^{\hat{x},E_1} \;=\; \underline{e}^x.\underline{n}^x\int_\omega S^x(\lambda)E_1(\lambda)\delta(\lambda-\lambda_i)d\lambda \;=\; \underline{e}^x.\underline{n}^x S^x(\lambda_i)E_1(\lambda_i) \tag{10}$$

Combining (9) and (10) together we see that:

$$p_i^{\hat{x},E_1} \;=\; \frac{E_1(\lambda_i)}{E(\lambda_i)}p^{\hat{x},E} \tag{11}$$

Equation (11) informs us that, as the colour of the light changes, the values recorded in each colour channel scale by a multiplicative factor (one factor per colour channel). If $\underline{R}$, $\underline{G}$ and $\underline{B}$ denote the N values recorded in an image (for each of the red, green and blue colour channels) then under a change in light colour the captured image becomes $\alpha\underline{R}$, $\beta\underline{G}$ and $\gamma\underline{B}$ (where α, β and γ are scalars). It is a simple matter to remove dependence on illumination colour:

$$
\begin{aligned}
\frac{\alpha N/3\underline{R}}{\alpha\sum_{i=1}^{N} R_i} &= \frac{N/3\underline{R}}{\sum_{i=1}^{N} R_i} \\
\frac{\beta N/3\underline{G}}{\beta\sum_{i=1}^{N} G_i} &= \frac{N/3\underline{G}}{\sum_{i=1}^{N} G_i} \qquad (12)\\
\frac{\gamma N/3\underline{B}}{\gamma\sum_{i=1}^{N} B_i} &= \frac{N/3\underline{B}}{\sum_{i=1}^{N} B_i}
\end{aligned}
$$

In terms of the $N \times 3$ image matrix $\mathcal{I}$, the normalization acts to scale each column to sum to $N/3$. This $N/3$ tally is far from arbitrary, but rather ensures that the total sum of all pixels post- column normalization is N which is the same as the total image sum calculated post row normalization. Thus, in principle, an image can be in both row- and column-normal form (and the goal of comprehensive normalization, discussed in the next section, is feasible). The function $C()$ column normalizes $\mathcal{I}$ according to (12):

$$
C(\mathcal{I})_{i,j} = \frac{N/3\mathcal{I}_{i,j}}{\sum_{k=1}^{N} \mathcal{I}_{k,j}} \qquad (12a)
$$

It is prudent to remind the reader that in order to arrive at the simple normalization presented in (12) delta function sensitivities were selected for our camera. While such a selection is not generally applicable, studies[FF96,FDF94b,FDF94a] have shown that most camera sensors behave, or can be made to behave, like delta functions.

3 The Comprehensive Normalization

The **comprehensive normalization** procedure is defined below:

$$
\begin{array}{llll}
1. & \mathcal{I}_0 = \mathcal{I} & \text{Initialization} & \\
2. & \mathcal{I}_{i+1} = C(R(\mathcal{I}_i)) & \text{Iteration step} & (13)\\
3. & \mathcal{I}_{i+1} = \mathcal{I}_i & \text{Termination condition} &
\end{array}
$$

The comprehensive procedure iteratively performs row and column normalizations until the termination condition is met. In practice the process will terminate when a normalization step induces a change less than a criterion amount.

Obviously this iterative procedure is useful if and only if we can show convergence and uniqueness. The procedure is said to converge, if for all images termination is guaranteed. If the convergent image is always the same (for any fixed scene) then uniqueness follows.

As a step towards proving uniqueness it is useful to examine the effects of lighting geometry and illuminant colour using the tools of matrix algebra. From the discussion in section 2, we know that viewing the same scene under a different lighting geometry results in an image where pixels, that is **r**ows of $\mathcal{I}$, are scaled. This idea can be expressed as an $N \times N$ diagonal matrix $\mathcal{D}^r$ premultiplying $\mathcal{I}$:

$$\mathcal{D}^r \mathcal{I} \tag{14}$$

Similarly, a change in illuminant colour results in a scaling of individual colour channels; that is, a scaling of the **c**olumns of $\mathcal{I}$. This may be written as $\mathcal{I}$ post multiplied by a 3×3 matrix $\mathcal{D}^c$:

$$\mathcal{I}\mathcal{D}^c \tag{15}$$

Equations (14) and (15) taken together inform us that the same scene viewed under a pair of lighting geometries and illuminant colours induces the image matrices: $\mathcal{I}$ and $\mathcal{D}^r\mathcal{I}\mathcal{D}^c$.

By definition, each iteration of the comprehensive normalization procedure scales the rows and then the columns of $\mathcal{I}$ by pre- and post-multiplying with the appropriate diagonal matrix. The operations of successive normalizations can be cascaded together and so we find that an image and its comprehensively normalized counterpart are related:

$$comprehensive(\mathcal{I}) \;=\; \mathcal{D}^{r*}\mathcal{I}\mathcal{D}^{c*} \tag{16}$$

where *comprehensive*() is a function implementing the iterative procedure shown in (13) and the symbol $*$ conveys the idea of a sequence of diagonal matrices.

Theorem 1. *Assuming the iterative procedure converges, if* $\mathcal{C}^1 = comprehensive(\mathcal{I})$ *and* $\mathcal{C}^2 = comprehensive(\mathcal{D}^r\mathcal{I}\mathcal{D}^c)$ *then* $\mathcal{C}^1 = \mathcal{C}^2$ *(proof of uniqueness).*

Proof. Let us assume that $\mathcal{C}^1 \neq \mathcal{C}^2$. By (16), $\mathcal{C}^1 = \mathcal{D}^r_1\mathcal{I}\mathcal{D}^c_1$ and $\mathcal{C}^2 = \mathcal{D}^r_2\mathcal{D}^r\mathcal{I}\mathcal{D}^c\mathcal{D}^c_2$ for some diagonal matrices $\mathcal{D}^r_1$, $\mathcal{D}^r_2$, $\mathcal{D}^c_1$ and $\mathcal{D}^c_2$. It follows that:

$$\mathcal{C}^2 = \mathcal{D}^a\mathcal{C}^1\mathcal{D}^b \tag{17}$$

where $\mathcal{D}^a \;=\; \mathcal{D}^r_2\mathcal{D}^r[\mathcal{D}^r_1]^{-1}$ and $\mathcal{D}^b \;=\; [\mathcal{D}^c_1]^{-1}\mathcal{D}^c\mathcal{D}^c_2$. By the assumption that $\mathcal{C}^1 \neq \mathcal{C}^2$, $\mathcal{D}^a$ and $\mathcal{D}^b$ are not equal to identity (or scaled) identity matrices. Clearly, for any $\mathcal{D}^a$ and $\mathcal{D}^b$ satisfying (17) so do $k\mathcal{D}^a$ and $\frac{1}{k}\mathcal{D}^b$ so, without loss of generality we assume that $\mathcal{D}^b_{i,i} > 1$. We also assume that $\mathcal{D}^b_{1,1} > \mathcal{D}^b_{2,2} > \mathcal{D}^b_{3,3}$ (since if this is not the case it can be made true by interchanging the columns of $\mathcal{C}^1$). Since $\mathcal{C}^2$ is comprehensively normalized we can express the components of $\mathcal{D}^a$ in terms of $\mathcal{D}^b$ and $\mathcal{C}^1$. In particular the ith diagonal term of $\mathcal{D}^a$ is, and must be, the reciprocal of the sum of the ith row of $\mathcal{C}^1\mathcal{D}^b$:

$$\mathcal{D}^a_{i,i} \;=\; \frac{1}{\mathcal{D}^b_{1,1}\mathcal{C}^1_{i,1} + \mathcal{D}^b_{2,2}\mathcal{C}^1_{i,2} + \mathcal{D}^b_{3,3}\mathcal{C}^1_{i,3}} \tag{18}$$

From which it follows that:

$$\mathcal{C}^2_{i,1} \;=\; \frac{\mathcal{D}^b_{1,1}\mathcal{C}^1_{i,1}}{\mathcal{D}^b_{1,1}\mathcal{C}^1_{i,1} + \mathcal{D}^b_{2,2}\mathcal{C}^1_{i,2} + \mathcal{D}^b_{3,3}\mathcal{C}^1_{i,3}} \tag{19}$$

Since have assumed that $\mathcal{D}^b_{1,1} > \mathcal{D}^b_{2,2} > \mathcal{D}^b_{3,3}$, it follows that

$$\frac{\mathcal{D}^b_{1,1}\mathcal{C}^1_{i,1}}{\mathcal{D}^b_{1,1}\mathcal{C}^1_{i,1} + \mathcal{D}^b_{1,1}\mathcal{C}^1_{i,2} + \mathcal{D}^b_{1,1}\mathcal{C}^1_{i,3}} < \frac{\mathcal{D}^b_{1,1}\mathcal{C}^1_{i,1}}{\mathcal{D}^b_{1,1}\mathcal{C}^1_{i,1} + \mathcal{D}^b_{2,2}\mathcal{C}^1_{i,2} + \mathcal{D}^b_{3,3}\mathcal{C}^1_{i,3}} \tag{20}$$

which implies that

$$\frac{\mathcal{C}^1_{i,1}}{\mathcal{C}^1_{i,1} + \mathcal{C}^1_{i,2} + \mathcal{C}^1_{i,3}} < \frac{\mathcal{D}^b_{1,1}\mathcal{C}^1_{i,1}}{\mathcal{D}^b_{1,1}\mathcal{C}^1_{i,1} + \mathcal{D}^b_{2,2}\mathcal{C}^1_{i,2} + \mathcal{D}^b_{3,3}\mathcal{C}^1_{i,3}} \equiv \mathcal{C}^1_{i,1} < \mathcal{C}^2_{i,1} \tag{21}$$

Equation (21) informs us that every element in the first column of $\mathcal{C}^1$ is strictly less than the corresponding element in in $\mathcal{C}^2$. However, this cannot be the case since both $\mathcal{C}_1$ and $\mathcal{C}_2$ are comprehensively normalized which implies that the sum of their respected first columns must be the same (and this cannot be the case if the inequality in (21) holds).
We have a contradiction and so, $\mathcal{C}^1 = \mathcal{C}^2$ and uniqueness is proven. □

Theorem 2. *The comprehensive normalization, (13), procedure always converges.*

Proof. Our proof follows directly from **Sinkhorn's** thoerem[Sin64] which we invoke here as a Lemma.
Lemma. Let $\mathcal{B}$ denote an arbitrary $n \times n$ all positive matrix. Sinkhorn showed that the process where the rows of $\mathcal{B}$ are iteratively scaled to sum to $n/3$ and then the columns are scaled to sum to $n/3$ (in an analogous process to (13)) is guaranteed to converge[2].

First, note that images, under any imaging conditions, are all positive. Now, let matrix $\mathcal{B}$ be a $3N \times 3N$ where the $N \times 3$ image matrix $\mathcal{I}$ is copied N times in the horizontal and 3 times in the vertical direction:

$$\mathcal{B} = \begin{bmatrix} \mathcal{I}\,\mathcal{I} \cdots \mathcal{I} \\ \mathcal{I}\,\mathcal{I} \cdots \mathcal{I} \\ \mathcal{I}\,\mathcal{I} \cdots \mathcal{I} \end{bmatrix} \tag{22}$$

Suppose that $\mathcal{D}^r$ and $\mathcal{D}^c$ are diagonal matrices such that the rows of $\mathcal{D}^r\mathcal{B}$ sum to N and the columns of $\mathcal{B}\mathcal{D}^c$ sum to N (note $N = 3N/3$). From the block structure of $\mathcal{B}$, it follows that $\mathcal{D}^r_{i,i} = \mathcal{D}^r_{i+kN,i+kN}$ $(i = 1, 2, \cdots, N)$ $(k = 2, 3)$. Similarly because columns sum to N, $\mathcal{D}^c_{i,i} = \mathcal{D}^c_{i+kN,i+kN}$ $(i = 1, 2, 3)$ and $k = (2, \cdots, N)$. Setting $\mathcal{D}^a = \mathcal{D}^r_{i,i}$ $(i = 1, 2, \cdots, N)$ and $\mathcal{D}^b = \mathcal{D}^c_{i,i}$ $(i = 1, 2, 3)$, we can write $\mathcal{D}^r\mathcal{B}$ and $\mathcal{B}\mathcal{D}^c$ as:

$$\mathcal{D}^r\mathcal{B} = \begin{bmatrix} \mathcal{D}^a\mathcal{I}\;\mathcal{D}^a\mathcal{I} \cdots \mathcal{D}^a\mathcal{I} \\ \mathcal{D}^a\mathcal{I}\;\mathcal{D}^a\mathcal{I} \cdots \mathcal{D}^a\mathcal{I} \\ \mathcal{D}^a\mathcal{I}\;\mathcal{D}^a\mathcal{I} \cdots \mathcal{D}^a\mathcal{I} \end{bmatrix} \;,\; \mathcal{B}\mathcal{D}^c = \begin{bmatrix} \mathcal{I}\mathcal{D}^b\;\mathcal{I}\mathcal{D}^b \cdots \mathcal{I}\mathcal{D}^b \\ \mathcal{I}\mathcal{D}^b\;\mathcal{I}\mathcal{D}^b \cdots \mathcal{I}\mathcal{D}^b \\ \mathcal{I}\mathcal{D}^b\;\mathcal{I}\mathcal{D}^b \cdots \mathcal{I}\mathcal{D}^b \end{bmatrix} \tag{23}$$

Each row in $\mathcal{D}^a\mathcal{I}$ sums to $\frac{3}{3N} * N = 1$ and each column in $\mathcal{I}\mathcal{D}^b$ sums to $\frac{N}{3N} * N = N/3$. That is each $\mathcal{I}$ in $\mathcal{B}$ is normalized according to the functions $R()$

[2] In fact we could choose any positive number here; $n/3$ will work well for our purposes.

and $C()$ and so after sufficient iterations, Sinkhorn's iterative process converges to:

$$Sinkhorn(\mathcal{B}) = \begin{bmatrix} comprehensive(\mathcal{I}) & comprehensive(\mathcal{I}) & \cdots & comprehensive(\mathcal{I}) \\ comprehensive(\mathcal{I}) & comprehensive(\mathcal{I}) & \cdots & comprehensive(\mathcal{I}) \\ comprehensive(\mathcal{I}) & comprehensive(\mathcal{I}) & \cdots & comprehensive(\mathcal{I}) \end{bmatrix} \tag{24}$$

Clearly, Sinkhorn's theorem implies the convergence of the comprehensive normalization procedure and our proof is complete. □

Experimentally, we found that the comprehensive normalization converges very rapidily: 4 or 5 iterations generally suffices.

4 Object Recognition Experiments

We carried out image indexing experiments for the Swain and Ballard[SB91], Simon Fraser[Cha95,GFF95], Berwick and Lee[BL98] image sets and a set of all images combined. Swain and Ballard's image set comprises 66 database and 31 query images. All images are taken under a fixed colour light source and there are only small changes in lighting geometry. Because there are, effectively, no confounding factors in Swain's images, we expect good indexing performance for lighting geometry, illuminant colour and comprehensive normalizations. The Simon Fraser dataset comprises a small database of 13 object images and 26 query images. Query images contain the same objects but viewed under large changes in lighting geometry and illuminant colour. In Lee and Berwick's image set there are 8 object images and 9 query images. Again queries images are captured under different conditions (viewing geometry and light colour change). The composite set comprises 87 database images and 67 queries.

To test the efficacy of each normalization we proceed as follows. For all images, database and query, we separately carried out lighting geometry, illuminant colour and comprehensive normalizations. At a second stage colour histograms, representing the colour distributions, of the variously normalized images are constructed. This involves histogramming only the (g, b) tuples in the normalized images. The pixel value r is discarded because $r + g + b = 1$ after lighting geometry and comprehensive normalizations, and so is a dependent variable. After illuminant colour normalization *on average* $r + g + b = 1$. A 16×16 partion of (G, B) colour space (which have values between 0 and 1) define the bins for the colour histograms. If H_i and Q denotes the histograms for the ith database and query images then the similarity of the ith image to the query image is defined as:

$$||H_i - Q||_1 \tag{25}$$

where $||.||_1$ denotes the L1 (or city-block distance) between the colour distributions. This distance is equal to the sum of absolute differences of corresponding histogram bins. Reassuringly, if $H_i = Q$ then $||H_i - Q||_1 = 0$; closeness corresponds to small distances.

For each query colour distribution, we calculate the distance to all distributions in the database. These distances are sorted into ascending order and the rank of the correct answer is recorded (ideally the image ranked in first place should contain the same object as the query image). Tables 1, 2 and 3 summarize indexing performance for all three normalizations operating on all four data sets. Two performance measures are shown: the % of queries that were correctly matched (% in 1st place) and the rank of the worst case match.

Image Set	% correct	worst ranking
Swain's	96.7	2nd out of 66
Simon Fraser	42.3	13th out of 13
Lee and Berwick	33.33	6th out of 8
Composite	58.2	86th out of 87

Table 1. Indexing performance of lighting geometry normalization

Image Set	% correct	worst ranking
Swain's	87.1	5th out of 66
Simon Fraser	80.8	6th out of 13
Lee and Berwick	67.7	4th out of 9
Composite	79.1	16th out of 87

Table 2. Indexing performance of illuminant normalization

Image Set	% correct	worst ranking
Swain's	80.6	2nd out of 66
Simon Fraser	100	1st out of 13
Lee and Berwick	100	1st out of 9
Composite	93.1	2nd out of 87

Table 3. Indexing performance of comprehensive normalization

A cursory look at the matching performance for Swain's images appears to suggest that lighting geometry normalization works best and the comprehensive normalization worst. This is, in fact, not the case: all three normalizations work very well. Notice that only the comprehensive normalization and lighting geometry normalizations place the correct answers in the top two ranks and this is an admirable level of performance given such a large database. For the Simon Fraser Database the comprehensive normalization procedure is vastly superior, 100% recognition is supported compared with 42.3% and 80.8% for lighting geometry and illuminant colour normalizations. The latter normalizations also perform very poorly in terms of the worst case rankings which are 13th and 6th respectively. This is quite unacceptable given the very small size (just 13 objects) of the Simon Fraser database. It is worth noting that no other colour distribution comparison method has come close to delivering 100% recognition on this dataset[FCF96] (these methods include, colour-angular indexing[FCF96], affine-invariants of colour histograms[HS94] and Colour constant colour indexing[FF95]). The same recognition story is repeated for the Berwick and Lee database. Comprehensive normalization supports 100% recognition and the other normalization perform very poorly.

Perhaps the recognition results for the composite data set are the most interesting. Over 93% (of 67) queries are correctly identified using comprehensive normalization and the worst case match is in second place. Such recognition performance is quite startling. The database is large comprising 87 objects and these were compiled by three different research groups. Also the means of recognition is a simple colour distribution comparison which is bound to fail when images, or objects, have the same mixture of colours. Indeed, most of the, 2nd place matches that are recorded have colour distributions which are similar to the overall best match. For example an image of 'Campbell's chicken soup' is confused with an image of 'Campbell's special chicken soup'. Both images are predominantly red and white (as we expect with Campbell's soup).

In comparison the lighting geometry and illuminant colour normalizations, used individually, perform very poorly. The former succeeds just 58% of the time and the worst case ranking is an incredibly poor 86th (out of 87). Illuminant colour normalization performs better, a recognition rate of 79% but again the worst case match is unacceptable: 16th placed out of 87.

5 Conclusion

The colours recorded in an image depend on both the lighting geometry and the colour of the illuminant. Unless these confounding factors can be discounted, colours cannot be meaningfully compared across images (and so object recognition by colour distribution comparison cannot work). In this paper we developed a new comprehensive normalization procedure which can remove dependency due to lighting geometry **and** illuminant colour and so can facilitate cross-image colour comparison.

Our approach is simplicity itself. We simply invoke normalization procedures which discount either lighting geometry **or** illuminant colour and apply them together and iteratively. We prove that this iterative process always converges to a unique *comprehensively normalized* image.

The power of our comprehensive normalization procedure was illustrated in a set the object recognition experiments. For four image databases: Swain's database, the Simon Fraser database, Sang Wok Lee's database, and a composite set (with almost 100 objects), recognition by colour distribution comparison, post-comprehensive normalization was found to be near perfect. Importantly, performance greatly surpassed that achievable using the lighting geometry or illuminant colour normalizations individually.

References

[BL98] D. Berwick and S.W. Lee. A chromaticity space for specularity-, illumination color- and illumination pose-invariant 3-d object recognition. In *ICCV98*, page Session 2.3, 1998.

[Buc80] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310:1–26, 1980.

[CB97] J.L. Crowley and F. Berard. Multi-modal tracking of faces for video communications. In *CVPR 97*, pages 640–645, 1997.

[Cha95] S.S. Chatterjee. Color invariant object and texture recognition, 1995. MSc thesis, Simon Fraser University, School of Computing Science.

[FCF96] 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.

[FDB91] B.V. Funt, M.S. Drew, and M. Brockington. Recovering shading in color images, 1991. Submitted for publication.

[FDF94a] G.D. Finlayson, M.S. Drew, and B.V. Funt. Color constancy: Generalized diagonal transforms suffice. *J. Opt. Soc. Am. A*, 11:3011–3020, 1994.

[FDF94b] G.D. Finlayson, M.S. Drew, and B.V. Funt. Spectral sharpening: Sensor transformations for improved color constancy. *J. Opt. Soc. Am. A*, 11(5):1553–1563, May 1994.

[FF95] B.V. Funt and G.D. Finlayson. Color constant color indexing. *IEEE transactions on Pattern analysis and Machine Intelligence*, 1995.

[FF96] G.D. Finlayson and B.V. Funt. Coefficient channels: Derivation and relationship to other theoretical studies. *COLOR research and application*, 21(2):87–96, 1996.

[GFF95] S.S. Chatterjee G.D. Finlayson and B.V. Funt. Color angle invariants for object recognition. In *3rd IS&T and SID Color Imaging Conference*, pages 44–47. 1995.

[GJT88] R. Gershon, A.D. Jepson, and J.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.

[Hea89] G. Healey. Using color for geometry-insensitive segmentation. *J. Opt. Soc. Am. A*, 6:920–937, 1989.

[HS94] G. Healey and D. Slater. "Global color constancy: recognition of objects by use of illumination invariant properties of color distributions". *Journal of the Optical Society of America, A*, 11(11):3003–3010, November 1994.

[Hun95] R.W.G. Hunt. *The Reproduction of Color.* Fountain Press, 5th edition, 1995.

[LL97] 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.

[MMK95] J. Matas, R. Marik, and J. Kittler. On representation and matching of multi-coloured objects. In *Proceedings of the fifth International Conference on Computer Vision*, pages 726–732. IEEE Computer Society, June 1995.

[NB93] 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.

[Pet93] A.P. Petrov. On obtaining shape from color shading. *COLOR research and application*, 18(4):236–240, 1993.

[SB91] M.J. Swain and D.H.. Ballard. Color indexing. *International Journal of Computer Vision*, 7(11):11–32, 1991.

[Sin64] R. Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *Annals of Mathematical Statistics*, 35:876–879, 1964.

[SO95] M. A. Stricker and M. Orengo. Similarity of color images. In *Storage and Retrieval for Image and Video Databases III*, volume 2420 of *SPIE Proceedings Series*, pages 381–392. Feb. 1995.

[SW95] B. Schiele and A. Waibel. Gaze tracking based on face-color. In *International Workshop on Automatic Face- and Gesture-Recognition*, June 1995.

[WB86] 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.