PERGAMON

Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Pattern Recognition 38 (2005) 179 – 190

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Illuminant and device invariant colour using histogram equalisation

Graham Finlayson[a], Steven Hordley[a,*], Gerald Schaefer[b], Gui Yun Tian[c]

[a] *School of Computing Sciences, University of East Anglia, Norwich NR4 7TJ, UK*
[b] *School of Computing and Technology, Faculty of Construction, Computing and Technology, The Nottingham Trent University, Burton Street, Nottingham NG1 4BU, UK*
[c] *School of Engineering, Electronics and Communications Research, University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK*

Received 19 December 2002; received in revised form 29 April 2004; accepted 29 April 2004

Abstract

Colour can potentially provide useful information for a variety of computer vision tasks such as image segmentation, image retrieval, object recognition and tracking. However, for it to be helpful in practice, colour must relate directly to the intrinsic properties of the imaged objects and be independent of imaging conditions such as scene illumination and the imaging device. To this end many *invariant* colour representations have been proposed in the literature. Unfortunately, recent work (Second Workshop on Content-based Multimedia Indexing) has shown that none of them provides good enough practical performance.

In this paper we propose a new colour invariant image representation based on an existing grey-scale image enhancement technique: histogram equalisation. We show that provided the rank ordering of sensor responses are preserved across a change in imaging conditions (lighting or device) a histogram equalisation of each channel of a colour image renders it invariant to these conditions. We set out theoretical conditions under which rank ordering of sensor responses is preserved and we present empirical evidence which demonstrates that rank ordering is maintained in practice for a wide range of illuminants and imaging devices. Finally, we apply the method to an image indexing application and show that the method out performs all previous invariant representations, giving close to perfect illumination invariance and very good performance across a change in device.

Keywords: Illuminant invariance; Device invariance; Histogram equalisation; Image indexing

1. Introduction

It has long been argued that colour (*RGB*) images provide useful information which can help in solving a wide range of computer vision problems. For example, it has been demonstrated [1–3] that characterising an image by the distribution of its colours is an effective way to identify images with similar content from amongst a diverse database of images. Or that a similar approach [4] can be used to locate objects in an image. Colour has also been found to be useful for tasks such as image segmentation [5,6] and object tracking [7,8]. Implicit in these applications is the assumption that the colours recorded by devices are an inherent property of the imaged objects and thus a reliable cue to their identity. In fact, an examination of image formation reveals that this assumption is not valid. Rather, the *RGB* that a camera records is more properly a measure of the light reflected from the surface of an object and while this does depend in part on characteristics of the object, it depends in equal measure on the composition of the light which is incident on the object in the first place. So, an object which is lit by an illuminant which is itself reddish will be recorded by a camera as more red than will the same object lit under a more bluish illuminant: image *RGB*s are illumination dependent. In addition

* Corresponding author. Tel.: +44-1603-593287; fax: +44-1603-593345.

E-mail address: steve@cmp.uea.ac.uk (S. Hordley)

0031-3203/$30.00 © 2004 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2004.04.010

image colour also depends on the properties of the recording device. Importantly, different imaging devices have different sensors which implies that an object which produces a given *RGB* response in one camera might well produce a quite different response in a different device. Moreover, a camera (or the user) may change the contrast in an image (e.g. by applying a gamma function).

In recognition of this fact many researchers have sought modified image representations such that one or more of these dependencies are removed. Research has to-date been focused on accounting for illumination dependence and can be broadly classified into one of *colour invariant* [9–13] or *colour constancy* [14,15] approaches. Colour invariant approaches seek transformations of the image data such that the transformed data are illuminant independent, whereas colour constancy approaches set out to determine an estimate of the light illuminating a scene and provide this estimate in some form to subsequent vision algorithms. Colour constancy algorithms, in contrast to invariant approaches, can deliver true object colours. Colour invariants can be calculated post-colour constancy processing but the converse is not true. Thus, in theory the colour constancy approach is a more powerful solution to the problem. This said, colour constancy has proven to be a harder problem to solve than colour invariants and for the moment at least more practical success can be achieved with invariants. Importantly, however, it has been demonstrated [15,1] that the practical performance of neither approach is good enough to facilitate colour-based object recognition or image retrieval across a change in illumination. In addition, none of the methods even attempts to account for device dependence.

In this paper we address the limitations of existing colour constancy and colour invariant approaches by defining a new image representation which we show is both illumination independent and (in many cases) also device independent. The method is based on the observation that while a change in illumination or device leads in practice to significant changes in the recorded *RGB*s, the rank orderings of the responses of a given sensor are largely preserved. So, for example, if we look at the rank order of *R* values for a set of surfaces under one illuminant and compare it to the corresponding rank order under a second light, we will find that the ordering is approximately invariant. In fact, we show in this paper that under certain simplifying assumptions, invariance of rank ordering follows directly from the image formation equation. In addition, we present an empirical study which reveals that the preservation of rank ordering holds in practice both across a wide range of illuminants and a variety of imaging devices. Thus, an image representation which is based on rank ordering of recorded *RGB*s rather than on the *RGB*s themselves offers the possibility of accounting for both illumination and device dependence at the same time.

There are a number of ways which we might exploit this rank ordering of sensor responses to derive an invariant image representation. In this paper we propose a method which borrows a tool which has long been used by the image processing community [16] for a quite different purpose. The technique is histogram equalisation and is typically applied to grey-scale images to produce a new image which is enhanced in the sense that the image has more contrast and thus conveys more information. In some cases this results in a visually more pleasing image. But in a departure from traditional image processing practice, we apply the procedure not to a grey-scale image, but rather to each of the *R*, *G*, and *B* channels of a colour image independently of one another. This departure, though subtle, is significant since if one histogram equalises *R*, *G*, and *B* independently it is possible (indeed it is common) to arrive at an image with unnatural pseudo-colours. As such the three-band histogram equalisation is usually advised against in the literature [16] since the method almost never improves the quality of the image. However, for our purposes the look of the image is not an issue. Rather, we seek only an invariant representation.

Histogram equalisation achieves the sought invariance since, as we will show, provided two images differ in such a way as to preserve the rank ordering of pixel values in each of the three channels then an application of histogram equalisation to each of the channels of the two images results in a pair of equivalent images. Thus, provided a change in illuminant or device preserves rank ordering of pixel responses, the application of histogram equalisation will provide us with an invariant representation of a scene which might subsequently be of use in a range of vision applications.

At this point the reader may feel surprised that we are proposing such a simple technique for image invariance. However, from our perspective the simplicity of the technique is key: we show that a hitherto neglected technique is applicable to colour imaging if the purpose is invariance (and not attractive images). We demonstrate the utility of the technique by applying the method to the problem of colour indexing: we show that the method out performs all previous approaches and in the case of a change in illumination provides close to perfect indexing.

The paper is organised as follows. In the next section we briefly review the image formation process and show formally how recorded responses depend on illuminant and device. We then consider some simplifying assumptions of the process and briefly describe a number of existing colour invariants and how they are derived. In Section 3 we set out a number of theoretical conditions under which rank ordering of pixel values are preserved and we present an empirical proof that rank orderings of sensor responses are in practice approximately invariant across a wide range of illuminants and devices. Section 4 shows how we can use histogram equalisation to exploit this rank invariance and derive an image which is illumination and device invariant We demonstrate the utility of the technique by applying it to the problem of colour-based image indexing in Section 5. Finally, we draw some conclusions from this work in Section 6.

2. Background

We adopt a simple model of image formation in which the response of an imaging device to an object depends on three factors: the light by which the object is lit, the surface reflectance properties of the object, and the properties of the device's sensors. We assume that a scene is illuminated by a single light characterised by its spectral power distribution which we denote $E(\lambda)$ and which specifies how much energy the source emits at each wavelength (λ) of the electromagnetic spectrum. The reflectance properties of a surface are characterised by a function $S(\lambda)$ which defines what proportion of light incident upon it the surface reflects on a per-wavelength basis. Finally, a sensor is characterised by $Q_k(\lambda)$, its spectral sensitivity function which specifies its sensitivity to light energy at each wavelength of the spectrum. The subscript k denotes that this is the kth sensor. Its response is defined as

$$q_k = \int_{\omega} E(\lambda)S(\lambda)Q_k(\lambda)\,\mathrm{d}\lambda, \quad k = 1, \ldots, m, \tag{1}$$

where the integral is taken over the range of wavelengths ω: the range for which the sensor has non-zero sensitivity. In what follows we assume that our devices (as most devices do) have three sensors ($m = 3$) so that the response of a device to a point in a scene is represented by a triplet of values: (q_1, q_2, q_3). It is common to denote these triplets as R, G, and B or just *RGB*s and so we use the different notations interchangeably throughout. In the context of this paper then, an image is a collection of *RGB*s representing the device's response to light from a range of positions in a scene. We note further that the ideas we present here can trivially be extended to images from devices with greater or fewer than three sensors.

Eq. (1) is an accurate model of the image formation process for Lambertian [17] surfaces for which incident light is reflected equally in all directions, and independently of the direction of the incident light. The equation makes clear the fact that a device response depends both on properties of the sensor (it depends on $Q_k(\lambda)$) and also on the prevailing illumination (on $E(\lambda)$). That is, responses are both device and illumination dependent. It follows that if no account is taken of these dependencies, an *RGB* cannot correctly considered to be an intrinsic property of an object.

An examination of the literature reveals many attempts to deal with the illumination dependence problem. One approach is to apply a correction to the responses recorded by a device to account for the colour of the prevailing scene illumination. Provided an accurate estimate of the scene illumination can be obtained, such a correction accounts well for the illumination dependence, rendering responses colour constant: that is stable across a change in illumination. The difficulty with this approach is the fact that estimating the scene illuminant is non-trivial. In 1998, Funt et al. [15] demonstrated that the existing colour constancy algorithms are not sufficiently accurate to make such an approach viable, though more recent work [14] has shown that for simple imaging conditions and given good device calibration the colour constancy approach can work.

In many situations, however, a calibrated device is not available and so a different approach is required. Alternative approaches to colour constancy set out to derive from the image data some new representation which is invariant to illumination. Such approaches are classified as colour (or illuminant) invariant approaches and a wide variety of invariant features have been proposed in the literature. One of the simplest invariants is a *chromaticity* representation of the image data. A chromaticity is derived from an *RGB* by the following transformation:

$$r = \frac{R}{R+G+B}, \quad g = \frac{R}{R+G+B}, \quad b = \frac{B}{R+G+B}. \tag{2}$$

A chromaticity vector (r, g, b) is invariant to a change in intensity of an illuminant since if (R, G, B) is the response recorded under one illuminant, then if the illuminant intensity changes the response will change to (sR, sG, sB) for some scale factor s and applying the transformation in Eq. (2) to both responses results in the same (r, g, b) triplet.

Accounting for a change in illumination colour is more difficult because, as is clear from Eq. (1), the interaction between light, surface, and sensor is complex. Researchers have attempted to reduce the complexity of the problem by adopting simple models of illumination change. One of the simplest models is the so-called *diagonal* model in which it is proposed that sensor responses under a pair of illuminants are related by a diagonal matrix transform:

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix} \begin{pmatrix} R^o \\ G^o \\ B^o \end{pmatrix}, \tag{3}$$

where the superscripts o and c characterise the pair of illuminants. The model is widely used and has been shown to be well justified under many conditions [18]. Adopting such a model one simple illuminant invariant representation of an image can be derived by applying the following transform:

$$R' = \frac{R}{R_{ave}}, \quad G' = \frac{G}{G_{ave}}, \quad B' = \frac{B}{B_{ave}}, \tag{4}$$

where the triplet $(R_{ave}, G_{ave}, B_{ave})$ denotes the mean of all *RGB*s in an image. It is easy to show that this so-called *Greyworld* representation of an image is illumination invariant provided that Eq. (2) holds.

Assuming the same conditions Finlayson et al. [11] have shown that successive and repeated application of Eqs. (2) and (4) converges to an image representation (which they call a *comprehensively normalised* image) which is both intensity and illuminant colour invariant. Many other illuminant invariant representations [9,10,12,13] have been derived, in some cases [12] by adopting different models of image formation. All derived invariants, however, share two common failings: first, it has been demonstrated that when applied to the practical problem of image retrieval [1] none of these invariants affords good enough performance across a change in illumination. Second, none of these approaches considers the issue of device invariance.

Variation of responses across devices can occur for a number of different reasons. First, the properties of the three sensors of a device can vary significantly from one device to another. So, whilst most trichromatic devices have sensors which respond broadly to long-, medium-, and short-wavelength regions of the visible spectrum, the exact regions to which the sensors respond and the relative sensitivity of the sensors within those regions will vary from one device to another. In terms of Eq. (1), different devices have different Q_k. But even if two devices have the same sensors the colours they record will not necessarily be the same for each device. This is because device responses are often not linearly related to scene radiance as Eq. (1) suggests, but rather the pixel values which a device outputs can be subject to some non-linear transformation. Thus, more generally, the image formation equation is written as

$$q_k = f\left(\int_{\omega} E(\lambda)S(\lambda)Q_k(\lambda)\,\mathrm{d}\lambda\right), \quad k = 1, \ldots, m, \tag{5}$$

where $f()$ is some arbitrary (possibly) non-linear transform. Importantly, the nature of this transform can vary from device to device and even on an per-image basis.

The transform $f()$ is deliberately applied to *RGB* values recorded by a device for a number of reasons. First, many captured images will eventually be displayed on a monitor. Importantly, colours displayed on a screen are not a linear function of the RGBs sent to the monitor. Rather, there exists a power function relationship between the voltage driving the monitor (which is linearly related to image intensity) and the displayed intensity. This relationship is known as the gamma of the monitor, where gamma describes the exponent of the power function [19]. To compensate for this gamma function images are usually stored in a way that reverses the effect of this transformation: that is by applying a power function with exponent of $1/\gamma$, where γ describes the gamma of the monitor, to the image RGBs. Importantly, monitor gammas are not unique but can vary from system to system and so images from two different devices will not necessarily have the same gamma correction applied. In addition to gamma correction other more general non-linear "tone curve" corrections are often applied to images so as to change image contrast with the intention of creating a visually more pleasing image. Such transformations are device and, quite often, image dependent and so lead, inevitably, to device-dependent colour. In the next section we consider how we might account for the effect of a change in illumination and/or device.

3. Rank invariance of sensor responses

Suppose we adopt, like a number of previous authors [9,20], the diagonal model of illumination change defined by Eq. (3). The usual approach to deriving invariant representations is to find algebraic manipulations of image *RGB*s such that illumination dependence is factored out. In this case, the aim is to factor out the three parameters, d_1, d_2, and d_3, of the diagonal matrix in Eq. (3). But rather than taking an algebraic approach we instead begin with the observation that one implication of this model of illumination change is that the rank ordering of sensor responses is preserved under a change of illumination. To see this let us denote by R_i^o the response of a single sensor to a surface i under an illuminant o. Under a second illuminant, which we denote c, the surface will have response R_i^c and the pair of sensor responses are related by

$$R_i^c = \alpha R_i^o. \tag{6}$$

Eq. (6) is true for all surfaces (that is, $\forall i$). Now, consider a pair of surfaces, i and j, viewed under illuminant o and suppose that $R_i^o > R_j^o$, then it follows from Eq. (6) that

$$R_i^o > R_j^o \;\Rightarrow\; \alpha R_i^o > \alpha R_j^o \;\Rightarrow\; R_i^c > R_j^c \quad \forall i, j, \ \forall \alpha > 0. \tag{7}$$

That is, the rank ordering of sensor responses within a given channel is invariant to a change in illumination under the assumption of a diagonal model of illumination change.

Next, consider the more general model of image formation (Eq. (3)) in which sensor responses are allowed to undergo a possibly non-linear transformation. Rank ordering is also preserved in this case for a certain class of functions $f()$. Specifically, rank ordering is preserved provided that $f()$ is a monotonic increasing function. Importantly, many of the transformations such as gamma or tone-curve corrections which are applied to images, satisfy this condition of monotonicity and are thus rank invariant. For example power (gamma) function transformations are rank invariant since:

$$R_i > R_j \;\Rightarrow\; (R_i)^\gamma > (R_j)^\gamma \quad \forall \gamma > 0. \tag{8}$$

It makes sense that tone-curve corrections applied to images should also be monotonic (and thus rank invariant) since such corrections are essentially mappings from input pixel values to output values. If this mapping is not monotonic as, for example, in the case of the mapping shown in Fig. 1, then it can happen that two quite different input pixel values are mapped to the same output value.

Thus, we have a property of pixel values (rank ordering) which is invariant to illumination and to a range of typical transforms applied to images. For true invariance we must also consider what happens when illumination change is not diagonal and also when two devices have different spectral sensitivities. Typical imaging devices have three classes of sensor which are broadly sensitive in either the long-,

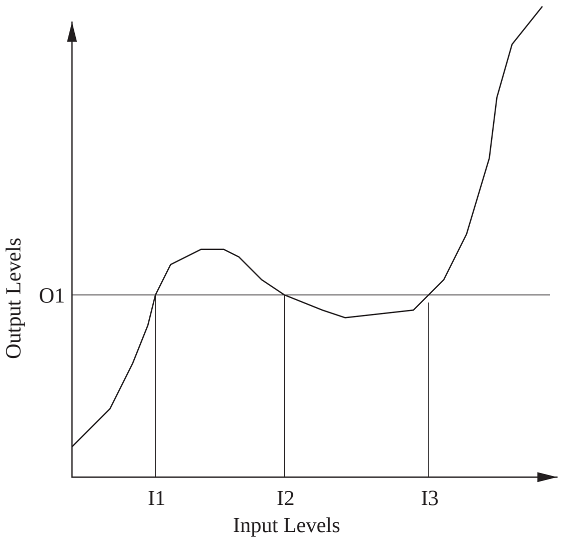

Fig. 1. A non-monotonic tone mapping results in two or more input values being mapped to the same output value.

medium-, or short-wavelength range of the visible spectrum. Fig. 2 shows the three sensor classes for a range of typical imaging devices. It is clear that there is a reasonable degree of correlation between sensors of the same class across different devices and so we might expect a good degree of rank invariance of responses across different devices. We consider this issue more carefully next.

3.1. Rank invariance in practice

To investigate further the rank invariance of sensor responses across changes in both illumination and device we conducted a similar experiment to that of Dannemiller [21] who investigated to what extent the responses of cone cells in the human eye maintain their rank ordering under a change in illumination. He found that to a very good approximation rank orderings were maintained. Here, we extend the analysis to investigate a range of different devices in addition to a range of illuminants.

Considering invariance of rank orderings of sensor responses for a single device under changing illumination we proceed as follows. Let R_k represent the spectral sensitivity of the kth sensor of the device we wish to investigate. Now suppose we calculate (according to Eq. (1)) the responses of this sensor to a set of surface reflectance functions under a fixed illuminant $E^1(\lambda)$. We denote those responses by the vector P_k^1. Similarly, we denote by P_k^2 the responses of the same sensor to the same surfaces viewed under a second illuminant $E^2(\lambda)$. Next, we define a function $rank()$ which takes a vector argument and returns a vector whose elements contain the rank of the corresponding element in the argument. Then, if sensor responses are invariant to the illuminants E^1 and E^2, the following relationship must hold:

$$rank(P_k^1) = rank(P_k^2). \tag{9}$$

In practice, the relationship in Eq. (9) will hold only approximately and we can assess how well the relationship holds using Spearman's rank correlation coefficient [22] which is defined as

$$\rho = 1 - 6 \sum_{j=1}^{N} \frac{d_j^2}{N_s(N_s^2 - 1)}, \tag{10}$$

where d_j is the difference between the jth elements of $rank(P_k^1)$ and $rank(P_k^2)$ and N_s is the number of surfaces. This coefficient takes a value between −1 and 1: a coefficient of zero implies that Eq. (9) holds not at all, while a value of one will be obtained when the relationship is exact. Invariance of rank ordering across devices can be assessed in a similar way by defining two vectors: P_k^1 defined as above and Q_k^1 representing sensor responses of the kth class of sensor of a second device under the illuminant E^1. By substituting these vectors in Eq. (10) we can measure the degree of rank correlation. Finally, we can investigate rank order invariance across device and illumination together by comparing, for example, the vectors P_k^2 and Q_k^1.

We conducted such an analysis for a variety of imaging devices and illuminants, taking a set of 462 Munsell chips [23], which represent a wide range of reflectances that might occur in the world as our surfaces. For illuminants we chose 16 different lights, including a range of daylight illuminants, Planckian blackbody radiators, and fluorescent lights, again representing a range of lights which we will meet in the world. We investigated performance for all the sensors shown in Fig. 2 which include the spectral sensitivities of the human colour matching functions [24] as well as those of four digital still cameras and a flatbed scanner.

Fig. 3 is typical of the results we obtained with our analysis. The figure is a plot of rank orderings of long-wavelength sensor responses to all surfaces under a fixed illuminant, against rank orderings of the corresponding responses for the same sensor but under a different light. If rank ordering was completely invariant all points in this plot would lie along the line $y = x$, which is also shown on the figure. It is clear that deviations from this line are small and thus rank ordering is approximately preserved for these two imaging conditions. This fact is reinforced by the correlation coefficient which in this case is 0.99.

Table 1 summarises the rank correlation coefficients for a number of different situations. The first six rows of this table show mean correlation for the three different classes of sensor of a range of devices across a change in illumination. In all cases correlation is very high implying that rank ordering is well maintained across a change in device. The next three rows show the results when illumination is kept fixed and device is allowed to change. Once again mean correlation is very high for the three illuminants shown. Similar results

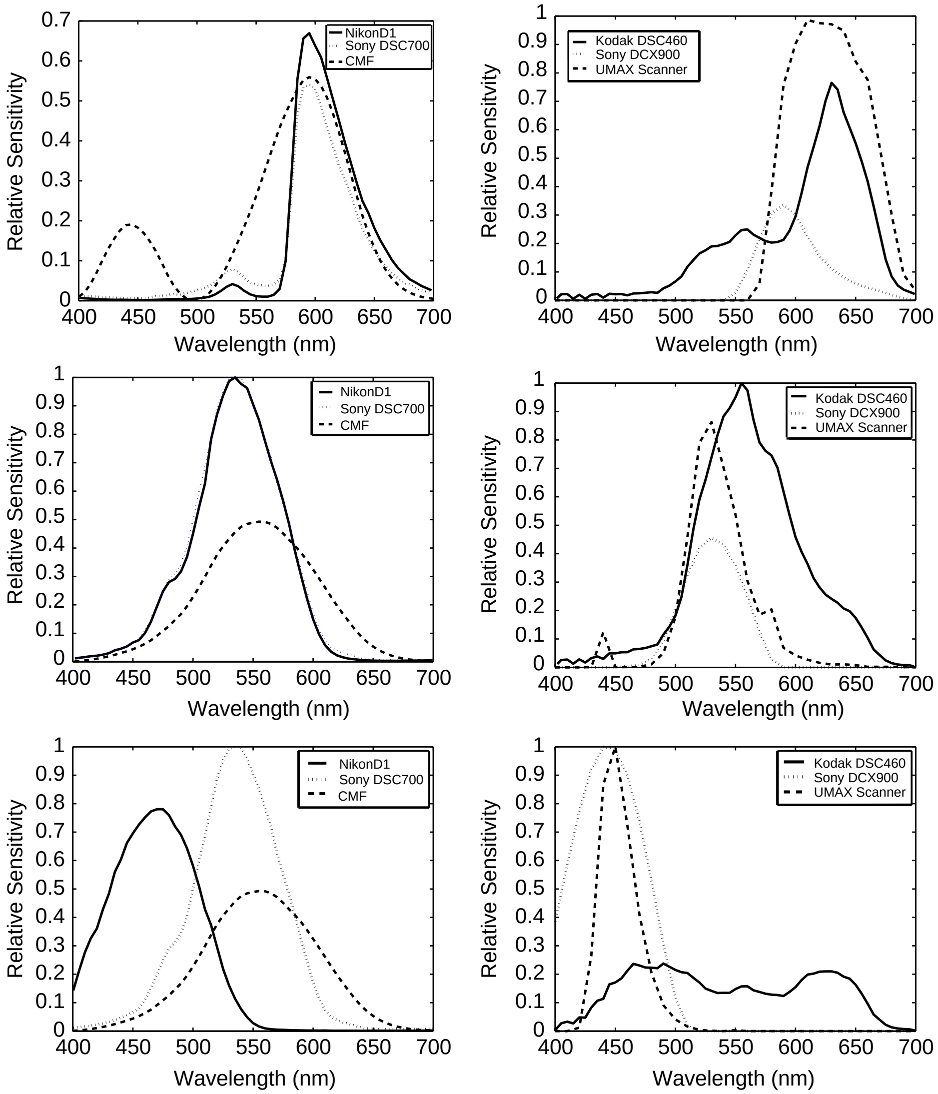

Fig. 2. Relative spectral sensitivity functions for the long-wavelength (top), medium-wavelength (middle), and short-wavelength (bottom) sensors for all the devices tested. First column shows sensors for a Nikon D1 digital camera (solid line), a Sony DSC700 (dotted line) and the colour matching functions (dashed line). The second column shows a Kodak DSC460 (solid line), a Sony DXC900 digital video camera (dotted line), and a UMAX flatbed scanner (dashed line).

were obtained for all other illuminants. Finally, correlation results are shown when both illumination and device are allowed to change. In this case correlation is slightly lower on average but is still sufficiently high to conclude that rank orderings are maintained in practice across a change in both device and illuminant.

3.2. Rank invariance of images

The analysis so far has demonstrated that under certain theoretical conditions two images which differ only in certain aspects of their capture conditions have rank invariant sensor responses. We have also shown that the sensor responses of images which differ not in content but only in capture conditions are in practice rank invariant under a wide range of capture conditions. We are thus in a position to define the equivalence of two or more images with respect to their rank orderings.

Let us begin by defining an image I^1, a set of *RGB* pixel values which represent the response of an imaging device to a number of surfaces viewed under a certain set of capture conditions. Suppose further that I^2 represents a

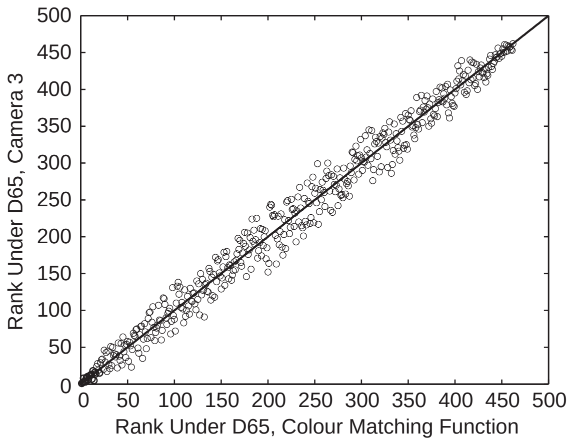

Fig. 3. Correlation plot of long-wave sensor responses to a set of surfaces viewed under two different lights.

Table 1
Spearman's rank correlation coefficient

	Long-wave sensor	Medium-wave sensor	Short-wave sensor
Across illumination			
Colour matching functions	0.9957	0.9922	0.9992
Camera 1	0.9983	0.9984	0.9974
Camera 2	0.9978	0.9938	0.9933
Camera 3	0.9979	0.9984	0.9972
Camera 4	0.9981	0.9991	0.9994
Scanner	0.9975	0.9989	0.9995
Across devices			
Daylight (D65)	0.9877	0.9934	0.9831
Fluorescent (cwf)	0.9931	0.9900	0.9710
Tungsten (A)	0.9936	0.9814	0.9640
Across device and illuminant			
	0.9901	0.9886	0.9774

Rows 1–6 show results for each sensor (*R*, *G*, and *B*) of a range of devices. Results are averaged over all pairs of a set of 16 illuminants. Rows 7–9 show results averaged over all devices for three different illuminants. Row 10 shows results averaged over six devices and 16 illuminants.

second image. Further let P_k^1 and P_k^2 ($k=1,2,3$) be vectors representing the kth sensor responses of each image. We say that I^1 is equivalent to I^2 (written $I^1 \equiv I^2$) if the following is true:

$$rank(P_k^1) = rank(P_k^2), \quad k = 1, 2, 3, \tag{11}$$

where $rank()$ is a function which takes a vector valued argument $\underline{v}$ and returns a vector whose ith element is the rank of v_i when the elements of $\underline{v}$ are ordered from smallest to largest.

In addition, given an image I we can define an equivalence class of images with respect to I which we denote $\mathscr{I}$. This equivalence class is defined as follows:

$$\mathscr{I} = \{I^j | rank(P_k^j) = rank(P_k), \ k = 1, 2, 3\}. \tag{12}$$

Any image which is in the equivalence class of an image I is equivalent to that image in the sense that its rank ordering is the same. As we have seen, this implies that the images are equivalent modulo a change in capture conditions (such as illumination or imaging sensors). Thus, comparing images in a manner which is invariant to capture conditions can be achieved by determining whether or not two images belong to the same equivalence class as defined above.

Alternatively, we can define a new image representation which exploits the rank invariant properties discussed above and leads to a representation which is invariant to capture conditions. In this case, all images within a single equivalence class (as defined above) will have the same invariant representation. There are many ways in which we might exploit rank invariance to define a new invariant image representation. We define one possible representation below which is simple to implement and which, we will demonstrate, has a number of useful properties.

4. Histogram equalisation for colour invariance

To understand our method consider a single channel of an *RGB* image recorded under an illuminant o where without loss of generality we restrict the range of R^o to be on some finite interval: $R^o \in [0 \ldots R_{max}]$. Now, consider further a value $R_i^o \in [0 \ldots R_{max}]$, where R_i^o is not necessarily the value of any pixel in the image. Let us define by $P(R^o < R_i^o)$, the number of pixels in an image with a value less than or equal to R_i^o. Under a second illuminant, c, a pixel value R^o under illuminant o is mapped to a corresponding value R^c. We denote by $P(R^c < R_i^c)$ the number of pixel values in the second image whose value is less than R_i^c. Assuming that the illumination change preserves rank ordering of pixels we have the following relation:

$$P(R^c < R_i^c) = P(R^o < R_i^o). \tag{13}$$

That is, the number of pixels in our image under illuminant o which have a value less than R_i^o is equal to the number of pixels in the image under illuminant c which have a value less than the transformed pixel value R_i^c: a change in illumination preserves cumulative proportions. So, if we represent our image data by these cumulative proportions we obtain a new image representation which is illuminant invariant. That is, we define the mapping from original image to invariant image thus:

$$R_i^{inv} = \frac{R_{max}}{N_{pix}} P(R^o \leqslant R_i^0), \tag{14}$$

Fig. 4. Top row shows the same scene captured with the same camera under three different lights. The second row shows the corresponding images post histogram equalisation.

where N_{pix} is the number of pixels and the constant R_{max}/N_{pix} ensures that the invariant image has the same range of values as the input image. Repeating the procedure for each channel of a colour image results in the required invariant image:

$$R_i^{inv} = \frac{R_{max}}{N_{pix}} P(R^o \leqslant R_i^0), \tag{15}$$

$$G_i^{inv} = \frac{G_{max}}{N_{pix}} P(G^o \leqslant G_i^0), \tag{16}$$

$$B_i^{inv} = \frac{B_{max}}{N_{pix}} P(B^o \leqslant B_i^0). \tag{17}$$

The reader familiar with the image processing literature might recognise Eqs. (15)–(17). Indeed, this transformation of image data is one of the simplest and most widely used methods for image enhancement and is commonly known as histogram equalisation. Histogram equalisation is an image enhancement technique originally developed for a single channel, or grey-scale, image. The aim is to increase the overall contrast in the image since doing so typical brightens dark areas of an image, increasing the detail in those regions which in turn can sometimes result in a more pleasing image. Histogram equalisation achieves this aim by transforming an image such that the histogram of the transformed image is as close as possible to a uniform histogram. The approach is justified on the grounds that amongst all possible histograms, a uniformly distributed histogram has maximum entropy [25]. Maximising the entropy of a distribution maximises its information and thus histogram equalising an image maximises the information content of the output image. Accepting the theory, to histogram equalise an image we must transform the image such that the resulting image histogram is uniform. Now, suppose that x_i represents a pixel value in the original image and x_i^t its corresponding value in the transformed image. Let us further assume that x_i and x_i^t are continuous variables and let us denote by $p(x)$ and $p_t(x^t)$ the probability density functions of the original and transformed image. We would like to transform the original image such that the proportion of pixels less than x_i^t in the transformed image is equal to the proportion of image pixels less than x_i in the original image and the histogram of the output image is uniform. This implies:

$$\int_0^{x_i} p(x)\,\mathrm{d}x = \int_0^{x_i^t} p_t(x^t)\,\mathrm{d}x^t = \frac{N_{pix}}{x_{max}} \int_0^{x_i^t} \mathrm{d}x^t. \tag{18}$$

Evaluating the right-hand integral we obtain and rearranging terms we have:

$$x_i^t = \frac{x_{max}}{N_{pix}} \int_0^{x_i} p(x)\,\mathrm{d}x. \tag{19}$$

Eq. (19) tells us that to histogram equalise an image we transform pixel values such that a value x_i in the original image is replaced by the proportion of pixels in the original image which are less than or equal to x_i. A comparison of Eqs. (15)–(17) and Eq. (19) reveals that, disregarding notation, they are the same. So, the invariant image is obtained by simply histogram equalising each of the channels of our original image. In practice, applying the histogram equalisation procedure to an image results in a transformed image whose resulting histogram is only approximately uniform. This is because the range of values a pixel can take is discrete and not continuous as we assumed in the analysis above.

In the context of image enhancement it is argued [16] that applying an equalisation to the channels of a colour image separately is inappropriate since this can produce significant colour shifts in the transformed image. However, in the current context, we are interested not in the visual quality of the image but in obtaining a representation which is illuminant and/or device invariant. Histogram equalisation achieves just this provided that the rank ordering of sensor responses is itself invariant to such changes. In addition, by applying histogram equalisation to each of the colour channels we maximise the entropy in each of those channels. This in itself seems desirable since our intent in computer vision is to use the representation to extract information about the scene and thus maximising the information content of our scene representation ought to be helpful in itself.

Fig. 4 illustrates the effect of applying the histogram equalisation technique to three images of the same scene, captured by the same camera under three different illuminants: a simulated daylight (D65), a fluorescent tube (TL84), and a tungsten filament light (A). The first row shows the

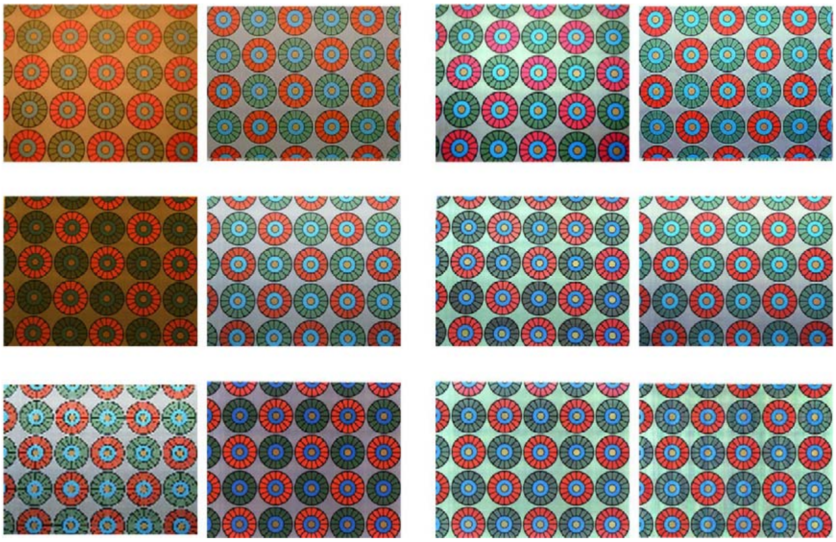

Fig. 5. The first two columns show images of the same scene captured with six different devices (four cameras and two scanners), while columns 3 and four show the corresponding images histogram equalised.

three images as they are captured by the camera and highlights the fact that a change in illumination leads to a significant change in the colours captured by the camera. To obtain the images in the second row of the figure we histogram equalised each of the channels in the three images. It is clear that the resulting images are much more similar than are the three original images, illustrating the illuminant invariant properties of the images. A second example is shown in Fig. 5, only this time the original images (first two columns) are captured with different devices. Once again, the histogram equalised images (columns three and four) exhibit a high degree of similarity, and the representation can be said to be device independent. While these two examples illustrate that the technique can work, we are interested in characterising more carefully, the degree to which the technique renders images invariant and its appropriateness for practical application. We consider this issue in the rest of the paper.

5. An application to colour indexing

To test the invariance properties of histogram equalisation further we applied the method to an image retrieval task. Finlayson et al. [1] recently investigated whether existing invariant approaches were able to facilitate good enough image indexing across a change in either or both illumination and device. Their results suggested that the answer to this question was no. Here, we repeat their experiment but using histogram equalised images as our basis for indexing to investigate what improvement, if any, the method brings.

The experiment is based on a database of images of coloured textures captured under a range of illuminants and devices. This database was used for two main reasons. First, the database is to our knowledge the only one which contains images of the same scene captured under a variety of both device and illumination. Second, using this database allows us to easily compare the performance of histogram equalisation with a number of other invariant methods for which performance on the same database has been published in [26]. Set against these advantages is the fact that scene content in these databases is quite restricted consisting only of simple colour/texture patterns and it is not immediately clear how the results would generalise to scenes with more diverse content. However, since we are indexing purely on colour, actual scene content is less important than colour diversity. Most of the texture images contain only a relatively small number of colours, so in these terms the indexing task is probably more difficult on this database than it would be on a more general database of natural images or objects.

The images are available from the University of East Anglia (UEA)[1] and are described in [26]. In summary, there are 28 different coloured textures each captured under six different devices (four cameras and two scanners). In addition, each camera was used to capture each of the textures under three different lights so that in total there are $(3 \times 4+2) \times 28 = 392$ images. Image resolution varies with capture device: for four out of the six devices it is approximately 500×400, while for the other two devices it is either

[1] http://www2.cmp.uea.ac.uk/Research/groups/compvis/CGmainData.htm

670 × 508 or 745 × 603. In image indexing terms this is a relatively small database and it is chosen for the reasons above. In our experiments we tested indexing performance across three different conditions: (1) across illumination, (2) across a change in device (for fixed illumination), and (3) across a change of both device and illumination.

In each case the experimental procedure was as follows. First, we choose a set of 28 images all captured under the same conditions (same device and illuminant) to be our image database. Next, we select from the remaining set of images a subset of appropriate query images. So, if we are testing performance across illumination, we select as our query images the 56 images captured by the device corresponding to the database images, under the two non-database illuminants. Then, for all database and query images we derive an invariant image using either the histogram equalisation method set forth above, or one of a range of previously published [9–13] invariant methods. Finally, we represent the invariant image by its colour distribution: that is, by a histogram of the pixel values in the invariant image. All results reported here are based on three-dimensional histograms of dimension 16 × 16 × 16. This allows us to directly compare performance of our algorithm with a number of other invariant methods which were compared in an identical experiment reported in [1]. It is worth noting that the performance of any individual invariant method might be improved by varying the number of histogram bins so that these results are not optimal for any method. However, by experimenting with different numbers of bins we have found that the overall trend in the results is constant so that the results we report here are a good indicator of relative performance between invariant methods.

Indexing is performed for a query image by comparing its histogram to each of the histograms of the database images. The database image whose histogram is most similar to the query histogram is retrieved as a match to the query image. We compare histograms using the intersection method described by Swain et al. [4] which we found to give the best results on average. Indexing performance is measured using average match percentile (AMP) [4] which gives a value between 0% and 100%. A value of 99% implies that the correct image is ranked amongst the top 1% of images whilst a value of 50% corresponds to the performance we would achieve using random matching. In addition to results for histogram equalisation we also show results based on histograms of the original images (*RGB*), and on Greyworld normalised images, that is on images calculated according to Eq. (4). Results for a variety of other invariant representations can be found in [1]: all perform significantly worse than Greyworld.

Tables 2 and 3 summarise the results for the experiment in which only illumination changes. The fourth column of Table 2 gives an overall summary of the results of this experiment. These results confirm the argument that without compensation for a change in illumination colour-based indexing is poor (indexing on *RGB* histograms gives very poor performance). In addition, the results show that on average histogram equalisation outperforms Greyworld: an AMP of close to 97% is achieved with histogram equalisation as compared to 94% for Greyworld. The first three columns of Table 2 give further insight into the relative performance of the two methods. In these columns AMP performance is broken down by illumination. For example, column one shows the average results over four different cameras when illuminant A is chosen as the database illuminant. It is clear that this choice of database illuminant has a significant effect on performance with results using illuminant A being considerably less good than those achieved with either of the other two lights. In fact, if both results for this light are ignored performance for both histogram equalisation and Greyworld is almost perfect.

Table 4 summarises indexing results across a change in device. In this experiment we used images taken with one camera under a fixed light as the database images and corresponding images under the same light but taken with three different cameras as query images. Each column of Table 4 corresponds to results obtained using images from one of the four cameras as the database. It is clear that performance is once again dependent on the choice of database images but in all cases performance using histogram equalisation is as good as or better than that obtained using a Greyworld normalisation. Averaging results over all four cameras gives an average match percentile of approximately 97% for histogram equalisation as compared to 94% for Greyworld which represents a significant performance gain using the new method.

Table 2
Results (AMP) of indexing experiment over a change in illuminant (by illuminant)

Colour model	Ill A	D65	TL84	All lights
Greyworld	90.08	95.28	96.53	93.96
Hist. eq.	93.21	98.23	98.73	96.72

Table 3
Results (AMP) of indexing experiment over a change in illuminant (by camera)

Colour model	Camera 1	Camera 2	Camera 3	Camera 4
Greyworld	96.23	81.59	99.12	98.90
Hist. eq.	99.25	92.35	96.91	98.37

Table 4
Results (AMP) of indexing experiment over a change of device, with illuminant fixed (by camera)

Colour model	Camera 1	Camera 2	Camera 3	Camera 4
Greyworld	95.81	89.92	93.67	97.50
Hist. eq.	98.16	92.34	93.62	98.99

Table 5
Results (AMP) of indexing experiment over a change of device and illuminant (by device)

Colour model	Cameras	Scanners	All devices
Greyworld	92.77	89.36	92.28
Hist.	94.99	88.48	94.54

In a final experiment we tested indexing performance allowing both device and illuminant to change. In this case we selected images taken with one of the four cameras or the two scanners as database images and used all remaining images as query images. In common with the other experiments performance in this case is sensitive to the choice of database images. Table 5 shows that performance is significantly reduced when using scanner rather than camera images as the database images. When scanners are used Greyworld and histogram equalisation perform similarly, while results averaged over all conditions reveal a performance improvement of approximately 3% for histogram equalisation as compared to Greyworld.

6. Discussion

Taken as a whole the results of the three experiments detailed above demonstrate a modest but significant advantage for histogram equalisation over the previous best performing method: Greyworld. While the results are quite good the experiments do raise a number of issues. First, it is surprising that one of the simplest invariants—Greyworld—performs as well as it does. This performance indicates that for this data set a diagonal scaling of sensor responses accounts for most of the change that occurs when illuminant or device is changed. It also suggests that any non-linear transform applied to the device responses post-capture (the function $f()$ in Eq. (5)) must be very similar for all devices: most likely a simple power function is applied. In the context of the current paper a second and more important issue concerns the performance of histogram equalisation. The analysis in Section 3.1 suggests that sensor responses are almost perfectly rank invariant under both a change in illumination and a change in device so that we might expect histogram equalisation to deliver perfect indexing; however, it does not. In particular, while illumination invariance is very good, device invariance is somewhat less than we might have hoped for given the analysis in Section 3. Possible reasons for this can be found by an examination of the images which make up the UEA database. We found that in addition to differences due to device and illumination, images in the database also differ spatially: i.e. the illumination varies spatially across the extent of an image and this spatial variation differs from image to image. Images of the same scene under a constant and spatially varying illuminant do not look the same after histogram equalisation. We are currently investigating how this spatial aspect of illumination can be dealt with.

Additional investigation of images for which indexing performance was particularly poor reveals a number of artifacts of the capture process which might also account for the performance. First, a number of images captured under tungsten illumination have values of zero in the blue channel for many pixels. Second, a number of the textures have uniform backgrounds but the scanning process introduces significant non-uniformities in these regions. For both cases the resulting histogram equalised images are far from invariant. Excluding these images leads to a significant improvement in indexing performance. However, for an invariant image representation to be of practical use in an uncalibrated environment it must be robust to the limitations of the imaging process. Thus we have reported results including all images. We stress again in summary, that the simple technique of histogram equalisation, posited only on the invariance of rank ordering across illumination and/or device outperforms all previous invariant methods and in particular gives excellent performance across changes in illumination. Further testing on more diverse and larger images databases is required to properly determine the power of this method as compared to other invariant approaches.

References

[1] G. Finlayson, G. Schaefer, Colour indexing across devices and viewing conditions, in: Second International Workshop on Content-Based Multimedia Indexing, 2001.
[2] W. Niblack, R. Barber, The QBIC project: querying images by content using color, texture, and shape, in: Storage and Retrieval for Image and Video Databases I, SPIE Proceedings Series, Vol. 1908, 1993.
[3] J. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. Jain, The virage image search engine: an open framework for image management, in: SPIE Conference on Storage and Retrieval for Image and Video Databases, Vol. 2670, 1996, pp. 76–87.
[4] Michael J. Swain, Dana H. Ballard, Color indexing, Int. J. Comput. Vision 7 (1) (1991) 11–32.
[5] G. Healey, Using color for geometry-insensitive segmentation, J. Opt. Soc. Am. A 6 (1989) 920–937.
[6] B.A. Maxwell, S.A. Shafer, A framework for segmentation using physical models of image formation, in: Computer Vision and Pattern Recognition, IEEE, 1994, pp. 361–368.
[7] B. Schiele, A. Waibel, Gaze tracking based on face-color, in: International Workshop on Automatic Face- and Gesture-Recognition, June 1995.
[8] Yogesh Raja, Stephen J. McKenna, Shaogang Gong, Colour model selection and adaptation in dynamic scenes, in: Fifth European Conference on Computer Vision, Springer, Berlin, June 1998, pp. 460–474.
[9] B.V. Funt, G.D. Finlayson, Color constant color indexing, IEEE Trans. Pattern Anal. Mach. Intell. 17 (5) (1995) 522–529.
[10] G.D. Finlayson, S.S. Chatterjee, B.V. Funt, Color angular indexing, in: The Fourth European Conference on Computer Vision (Vol. II), European Vision Society, 1996, pp. 16–27.
[11] G.D. Finlayson, B. Schiele, J. Crowley, Comprehensive colour image normalization, in: eccv98, 1998, pp. 475–490.

[12] T. Gevers, A.W.M. Smeulders, Color based object recognition, Pattern Recogn. 32 (1999) 453–464.
[13] M. Stricker, M. Orengo, Similarity of color images, in: SPIE Conference on Storage and Retrieval for Image and Video Databases III, Vol. 2420, 1995, pp. 381–392.
[14] G.D. Finlayson, S. Hordley, P.M. Hubel, Illuminant estimation for object recognition, Color Res. Appl. 27 (4) (2002) 260–270.
[15] B. Funt, K. Barnard, L. Martin, Is machine colour constancy good enough?, in: Fifth European Conference on Computer Vision, Springer, Berlin, June 1998, pp. 455–459.
[16] A. Watt, F. Policarpo, The Computer Image, Addison-Wesley, Reading, MA, 1997.
[17] B.K.P. Horn, Robot Vision, MIT Press, Cambridge, MA, 1986.
[18] G. Finlayson, Coefficient colour constancy, Ph.D. Thesis, Simon Fraser University, 1995.
[19] C. Poynton, The rehabilitation of gamma, in: SPIE Conference on Human Vision and Electronic Imaging III, Vol. 3299, 1998, pp. 232–249.
[20] D.A. Forsyth, A novel algorithm for colour constancy, Int. J. Comput. Vision 5 (1) (1990) 5–36.
[21] J.L. Dannemiller, Rank orderings of photoreceptor photon catches from natural objects are nearly illuminant-invariant, Vision Res. 33 (1) (1993) 131–140.
[22] R.J. Freund, W.J. Wilson, Statistical Methods, Academic Press, New York, 1993.
[23] S.M. Newhall, D. Nickerson, D.B. Judd, Final report of the OSA subcommittee on the spacing of the Munsell colours, J. Opt. Soc. Am. A 33 (1943) 385–418.
[24] G. Wyszecki, W.S. Stiles, Color Science: Concepts and Methods, Quantitative Data and Formulas, second ed., Wiley, New York, 1982.
[25] R.C. Gonzalez, R.E. Woods, Digital Image Processing, Addison-Wesley, Reading, MA, 1992.
[26] G. Finlayson, G. Schaefer, G.Y. Tian, The UEA uncalibrated colour image database, Technical Report SYS-C00-07, School of Information Systems, University of East Anglia, Norwich, UK, 2000.

About the Author—PROFESSOR FINALYSON leads the Colour Group in the School of Computing Sciences at the University of East Anglia, Norwich, UK. In the past 5 years Professor Finalyson has published over 100 scientific papers in colour. His research interests include colour imaging, computer vision, artificial intelligence, and computational geometry.

About the Author—DR. HORDLEY is a Lecturer in the School of Computing Sciences at the University of East Anglia, Norwich, UK. His research interests include colour imaging computer vision, and image processing.

About the Author—GERALD SCHAEFER is a Senior Lecturer in the Department of Computing and Mathematics at The Nottingham Trent University and is also enrolled as a part-time Ph.D. student at the University of East Anglia, Norwich. His research interests include colour image analysis, physics-based vision, image retrieval, and image coding.

About the Author—DR. TIAN is a lecturer in Multimedia and Electronics at the University of Huddersfield. His research interests include sensors and instrumentation, colour imaging, digital signal processing, and Internet technology. He is a member of the BCS and the American Precision Engineering Society and has published over 40 articles in the past 5 years.