Using Colour for Image Indexing

Graham D. Finlayson
The Colour & Imaging Institute
The University of Derby
Britannia Mills
Derby DE22 3BL
United Kingdom

Bernt Schiele
Media Lab
MIT
20 Ames Street
Cambridge MA 02139
USA

James L. Crowley
Laboratoire GRAVIR
I.N.P.G
46 avenue Felix Viallet
Grenoble 38000 Cedex
France

Abstract

Image colour is often thought to be an intrinsic correlate of surface reflectance and so is a common feature for image indexing. In this paper we point out that image colour is actually a function of surface reflectance **and** *imaging geometry* **and** *the colour of the viewing illuminant. Fortunately methods exist for normalizing away these dependencies. Pixel based and colour channel-based normalizations remove dependency on geometry and light colour respectively. Unfortunately, neither method removes both dependencies simulataneously and so a single normalization must be chosen. Common practice dictates that pixel-based normalization is the most useful. In contrast experiments that we carried out, on a variety of image databases cited in the computer vision literature, favour colour channel normalization.*

1 Introduction

At first glance, image colour would appear to be an ideal feature for image indexing by image content. For example, because images of ‘beach scenes’ typically comprise sea (deep blue), sky (light blue) and sand (beige) a tripartite colour query of *Find*(light blue, sky blue and beige) should suffice in finding, at least some, images of beach scenes. Indeed IBM’s QBIC[1] (Query by image content) system supports exactly this sort of query. Another advantage of colour queries is that they can be applied to non-annotated images. In contrast in order to query by shape (e.g. find ‘fish’, in images) all images in a given database must be pre-segmented (in order to make shapes explicit). Unfortunately, this segmentation must carried out by a human operator and so is a costly exercise for all but the smallest of image databases.

In this paper we sound a cautionary note regarding the use of image colour as a querying mechanism. We show that the colours that are recorded in an image depend on two confounding factors: the relative pose of surface and light and the colour of the light source. This dependence is sufficiently strong to confound indexing[2] (a beach viewed under a red dusk sky will result in an image where all the colours are redder than they ought to be).

Fortunately, however, various authors have shown that images can be normalized in order to remove dependence on individual viewing conditions[3, 4, 5]. Scaling every image (r, g, b) pixel triple to sum to one, $(\frac{r}{r+g+b}, \frac{g}{r+g+b}, \frac{b}{r+g+b})$, removes ambiguity due to illumination/surface pose[5]. Similarly scaling each colour channel, that is all values in the $\underline{R}$, $\underline{G}$ and $\underline{B}$ pixel planes, to sum to one, $(\frac{\underline{R}}{\sum_{i=1}^{n} R_i}, \frac{\underline{G}}{\sum_{i=1}^{n} G_i}, \frac{\underline{B}}{\sum_{i=1}^{n} B_i})$ removes dependence due to the colour of the light[3]. While, neither normalization removes dependence due to both the relative pose of the light source and the colour of the light the former normalization is more prevalent in the literature[6, 5, 4] than the latter[3]. This suggests that pixel based normalization is thought to be the most useful normalization procedure.

In in our research we quantitatively address the question “which image normalization, pixel- or colour channel-based, should we use?”. To answer this question we carried out image indexing experiments for a variety of image datasets. In each case the database comprised images of colourful objects and the query images were images of the same objects viewed under different capture conditions. These ‘object’ databases have the advantage that there is a notional correct answer to image indexing; when indexing works well both the query and indexed database image should contain same object. In line with previous colour indexing studies[7, 8, 1], indexing proceeds by finding the database image colour distribution which is most similar to the distribution of colours in a query image. Experiments were carried out for distributions calculated post pixel- and colour channel-based normalizations. Contrary to current database practice we found colour channel based normalization to be much more useful than pixel-based normalization.

In section 2 of this paper we review colour image formation and illustrate how image colours depend on the relative pose of light and surface and on the colour of the light. Pixel- and colour channel-based normalizations are introduced as mechanisms for removing these dependencies. These normalizations are applied as preprocessing steps for image indexing experiments reported in section 3. For the image data sets tested, colour channel normalization proved to be the most useful compared with pixel based normalization. We conclude the paper with a short discussion in section 4.

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 (these are the only kind 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. Illumination, surface reflection and sensor function, combine together in forming a sensor response:

$$\underline{\rho}^{\hat{x},E} = \underline{e}.\underline{n}^x \int_w S^x(\lambda)E(\lambda)\underline{F}(\lambda)d\lambda \tag{1}$$

where λ is wavelength, $\underline{\rho}$ is the 3-vector of sensor responses (*rgb* pixel value) $\underline{F}$ is the 3-vector of response functions (red-, green and blue- sensitive), E (assumed constant across the scene) is the incident illumination and S^x is the surface reflectance function at location x on the surface which is projected onto location $\hat{x}$ on the sensor array. The relative orientation of surface and light is taken in account by the dot-product of the surface normal vector $\underline{n}^x$ with the light source direction $\underline{e}$ (both these vectors have unit length).

Let us denote $\int_w S^x(\lambda)E(\lambda)\underline{F}(\lambda)d\lambda$ as $\underline{q}^{x,E}$. It follows that (1) can be rewritten as:

$$\underline{\rho}^{\hat{x},E} = \underline{q}^{x,E}\underline{e}.\underline{n}^x \tag{2}$$

where . is the vector dot-product[1]. Equation (2) informs us that the pixel recorded in an image, for a fixed illuminant E, is in a fixed direction in colour space, $\underline{q}^{x,E}$ but but has variable magnitude proportional to the scalar $\underline{e}.\underline{n}^x$

Pixel based normalization removes this variation by dividing each camera response vector by the sum of all responses (at that pixel):

$$\frac{\underline{\rho}^{\hat{x},E}}{\sum_{i=1}^{3}\rho_i^{\hat{x},E}} = \frac{\underline{q}^{x,E}\underline{e}.\underline{n}^x}{\underline{e}.\underline{n}^x\sum_{i=1}^{3}q_i^{x,E}} = \frac{\underline{q}^{x,E}}{\sum_{i=1}^{3}q_i^{x,E}} \tag{3}$$

The dependent scale term $\underline{e}.\underline{n}^x$ cancels in Equation (3). When $\underline{\rho}^{\hat{x},E} = (r,g,b)$ then pixel normalization returns $(\frac{r}{r+g+b}, \frac{g}{r+g+b}, \frac{b}{r+g+b})$. Notice, however, that throughout the development summarized in (1), (2) and (3) the superscript E still appears: the pixel normalized colours in an image still depend (that is, vary with) the capture illuminant.

To understand the effect of the viewing illuminant on captured colours it is useful to think of the camera sensors as delta functions: $F_i(\lambda) = \delta(\lambda - \lambda_i)$ $(i = 1,2,3)$. Delta functions are sensors which are sensitive to a single wavelength of light. Under this assumption:

$$p_i^{\hat{x},E} = \underline{e}.\underline{n}^x \int_\omega S^x(\lambda)E(\lambda)\delta(\lambda-\lambda_i)d\lambda = \underline{e}.\underline{n}^x S^x(\lambda_i)E(\lambda_i) \tag{4}$$

The same surface viewed under a different coloured light, $E_1(\lambda)$, but under the same viewing geometry induces the following response:

---

[1] Equation (2) is an incarnation of Lambert's Law

$$p_i^{\hat{x},E_1} = \underline{e}.\underline{n}^x \int_\omega S^x(\lambda)E_1(\lambda)\delta(\lambda-\lambda_i)d\lambda = \underline{e}.\underline{n}^x S^x(\lambda_i)E_1(\lambda_i) \tag{5}$$

Combining (4) and (5) together we see that:

$$p_i^{\hat{x},E_1} = \frac{E_1(\lambda_i)}{E(\lambda_i)} p_i^{\hat{x},E} \tag{6}$$

As the colour of light changes the values recorded in each colour channel scale by the same factor (independent of surface reflectance and viewing geometry). It follows that 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. Under a change of viewing illuminant the captured image becomes $\alpha\underline{R}$, $\beta\underline{G}$ and $\gamma\underline{B}$ (where α, β and γ are scalars).

Colour Channel based normalization removes dependence on illumination colour by dividing each colour channel by the sum of all the values in the colour channel:

$$\begin{array}{ccc} \frac{\alpha\underline{R}}{\alpha\sum_{i=1}^{n} R_i} & = & \frac{\underline{R}}{\sum_{i=1}^{n} R_i} \\ \frac{\beta\underline{G}}{\beta\sum_{i=1}^{n} G_i} & = & \frac{\underline{G}}{\sum_{i=1}^{n} G_i} \\ \frac{\gamma\underline{B}}{\gamma\sum_{i=1}^{n} B_i} & = & \frac{\underline{B}}{\sum_{i=1}^{n} B_i} \end{array} \tag{7}$$

The light colour dependent scale terms α, β and γ cancel in Equation (7). The normalization in (7) is sometimes referred to *grey-world* colour constancy[9, 10, 11].

Of course to arrive at the simple normalization presented in (7) we had to assume that camera sensors were delta functions. While this need not be true, camera sensors generally behave like delta functions[12, 13, 14]

3 Colour Indexing Experiments

We carried out image indexing experiments for the Swain and Ballard[15], Simon Fraser[16], Berwick and Lee[6] image sets and a set of all images combined. Swain and Ballard's image set comprises 66 database and 31 queries images. All images are taken under a fixed colour light source and there are only small changes in light/surface pose. Because there are, effectively, no confounding factors in Swain's images, we expect good indexing performance for both pixel-based and colour-channel based normalizations. 24 of Swain and Ballard's query images are shown in Figure 1. 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 relative pose and light 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 pixel based normalization we proceed as follows. For all images, database and query, we carried out pixel based normalization. At a second stage colour histograms, representing the colour distributions, of the normalized images are constructed. However, because $R + G + B = 1$ (by definition), R is not independent of G and B, so only the distribution of (G, B) tuples is actually recorded. A 16×16 partion of (G, B) colour space (which have values between 0 and 1) define the bins for the colour

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 pixel-based 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 colour channel-based normalization

histograms. If D_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:

$$||D_i - Q||_1 \tag{8}$$

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 $D_i = Q$ then $||D_i - Q||_1 = 0$. 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 (ideally the query and closest indexed image (rank 1) should contain the same object) is recorded. Table 1 summarizes indexing performance for 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.

We repeated the same experiment for the colour channel based normalization. However, after channel normalization the sum of R, G and B at each pixel will **not** equal 1 but we would like the sum to equal 1 so that we can compare the channel and pixel based normalizations directly. To achieve this level playing field we apply pixel based normalization at a second stage to the channel normalized images and histogram (B, G) tuples as before. Indexing performance is summarized in Table 2.

Clearly, both normalizations support good indexing performance for Swain's dataset with the pixel-based normalization giving slightly better results. However, for all other data sets the colour channel normalization provides superior indexing: the percentage correctly matched is always high and the worst case ranking is always low. In contrast, pixel based normalization performs quite poorly. For the Simon Fraser dataset less than half of the queries are correctly identified and the worst case match is 13th out of 13 (which is the worst that is possible!). Similar results are reported for Lee and Berwick's images. The composite data set statistics are perhaps the most interesting. Almost 80% of queries are correctly matched after colour channel normaliztion compared to less than 60% for pixel-based normalization. Moreover the worst case matches are respectively 16th and 86th (out of 87). While 16th is almost acceptable 86th is completely unacceptable.

So why does channel normalization work better? Partly, we believe this is due to a sleight of hand that we performed. Specifically, in order to compare the pixel based normalization to the colour channel normalization we were forced to pixel normalize all colour channel normalized images (since this second normalization step rendered all image colours in the same coordinate frame). Thus, our experiments tell us that pixel based normalization is not as good as colour channel & pixel based normalizations combined. Of course what is really needed is an image normalization procedure that is sufficiently general that it can remove both dependencies due to viewing geometry and the colour of the light source. Perhaps the combined colour channel & pixel based normalization is a first step towards achieving this goal.

4 Conclusions

Colour is an appealing feature for indexing by image content. We can naturally describe scenes in terms of their relative colour composition and this is readily formulated as a database query and this query can be executed quickly. Unfortunately, this approach is only reasonable if the colours captured in an image correspond to the colours that we ourselves see. They do not. Image colours depend on viewing geometry (the relative orientation of surface and light source) and on the colour of the light.

It follows that effective image indexing is only possible if we can account for and discount these confounding factors. Fortunately, techniques exist for doing just this. Pixel based normalization removes dependence on viewing geometry and colour channel based normalization removes dependence on light source colour. While neither normalisation suffices to remove both dependencies, we present experiments that indicate that normalizing for light source colour is more important for image indexing based on colour compared with the pixel based normalization. This observation is at odds with common practice in the image database community where pixel based normalization is more prevalent.

References

[1] 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.

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

[3] 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.

[4] M.J. Swain. *Color Indexing*. PhD thesis, University of Rochester, Department of Computer Science, 1990.

[5] 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.

[6] 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.

[7] M.J. Swain. Interactive indexing into image databases. In *Storage and Retrieval for Image and Video Databases I*, SPIE Proceedings Series, pages 95–103. Feb 1993.

[8] 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.

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

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

[11] R. Gershon, A.D. Jepson, and J.K. Tsotsos. From $[r, g, b]$ to surface reflectance: Computing color constant descriptors in images. *Perception*, pages 755–758, 1988.

[12] 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.

[13] 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.

[14] 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.

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

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

Figure 2:

Figure 1. 24 of Swain's object images