A combined physical and statistical approach to colour constancy

Gerald Schaefer  
School of Computing and Informatics  
Nottingham Trent University, U.K.  
Gerald.Schaefer@ntu.ac.uk

Steven Hordley and Graham Finlayson  
School of Computing Sciences  
University of East Anglia, U.K.  
{sh,graham}@cmp.uea.ac.uk

Abstract

Computational colour constancy tries to recover the colour of the scene illuminant of an image. Colour constancy algorithms can, in general, be divided into two groups: statistics-based approaches that exploit statistical knowledge of common lights and surfaces, and physics-based algorithms which are based on an understanding of how physical processes such as highlights manifest themselves in images. A combined physical and statistical colour constancy algorithm that integrates the advantages of the statistics-based Colour by Correlation method with those of a physics-based technique based on the dichromatic reflectance model is introduced. In contrast to other approaches not only a single illuminant estimate is provided but a set of likelihoods for a given illumination set. Experimental results on the benchmark Simon Fraser image database show the combined method to clearly outperform purely statistical and purely physical algorithms.

1. Introduction

The sensor responses of a device such as a digital camera depend both on the surfaces in a scene and on the prevailing illumination conditions: an object viewed under two different illuminants will yield two different sets of sensor responses. For humans however, the perceived colour of an object is more or less independent of the illuminant; a white paper appears white both outdoors under bluish daylight and indoors under yellow tungsten light, though the responses of the eyes’ colour receptors, the long-, medium-, and short-wave sensitive cones, will be quite different for the two cases. Researchers in computer vision have long sought algorithms to make colour cameras equally colour constant.

Colour constancy algorithms can, in general, be divided into two main categories: physics-based and statistically-based approaches [6]. Statistical methods try to correlate the distribution of colours in the scene with statistical knowledge of common lights and surfaces. In contrast, physics-based methods are based on an understanding of how physical processes such as highlights or interreflections manifest themselves in images.

While committee based decision systems are commonly used in areas such as optical character recognition [11] there has been little research in the combination of colour constancy algorithms. The work by Cardei and Funt [2] performs a fairly simple convex combination of Grayworld, MaxRGB and neural network colour constancy algorithms. In particular, they explore a weighted average of the solutions obtained by each of the three methods as well as a non-linear combination based on a neural net. Crucially, the only information that is utilised are the illuminant estimates of each algorithm. Furthermore, a real improvement could only be noticed in the combination of Grayworld and MaxRGB algorithms. When incorporating the neural network estimates the accuracy of the combined estimation was in some cases even lower than that of the neural network approach alone.

In this paper a first approach at a combined physical and statistical approach to colour constancy is presented. Although a similar approach as in [2] could be taken to integrate a statistics-based algorithm with a physics-based method, the results reported there suggest that only little if any improvement can be gained by averaging separate estimates. The algorithm proposed in here combines the data obtained by a statistical colour constancy algorithm, a version of the Colour by Correlation algorithm [4], with that of a physics-based technique based on the dichromatic reflectance model. In particular, a vector of likelihoods that describes the probability of an image having been taken under each of a set of reference lights obtained from the Colour by Correlation framework is combined with a likelihood vector that describes similar estimation probabilities recovered by the dichromatic colour constancy algorithm. The resulting solution vector describes combined likelihoods with respect to the set of reference lights; a single solution can be obtained by choosing the maximum likelihood estimate.
Experiments on the benchmark Simon Fraser colour constancy image database [1] confirm that this novel colour constancy algorithm provides excellent recovery performance. The results achieved show significantly improved colour constancy compared to either purely statistical or purely physics-based algorithms.

2. Colour image formation

An image taken with a linear device such as a digital colour camera is composed of sensor responses that can be described by

$$p = \int_{\omega} C(\lambda) R(\lambda) d\lambda$$

(1)

where $\lambda$ is wavelength, $p$ is a 3-vector of sensor responses (RGB pixel values), $C$ is the colour signal (the light reflected from an object), and $R$ is the 3-vector of sensitivity functions of the device. Integration is performed over the visible spectrum $\omega$. The colour signal $C(\lambda)$ itself depends on both the surface reflectance $S(\lambda)$ and the spectral power distribution $E(\lambda)$ of the illumination. For pure Lambertian (matte) surfaces $C(\lambda)$ is proportional to the product $S(\lambda) E(\lambda)$ and its magnitude depends on the angle(s) between the surface normal and the light direction(s).

3. Statistical Colour by Correlation colour constancy

The correlation matrix colour constancy algorithm by Finlayson et al. [4] is based on the observation that the colours in an image provide some information about the scene illuminant. An image with a yellowish colour cast is more likely to have originated from a scene captured under yellow tungsten light than under bluish daylight. The correlation matrix colour constancy approach tries to exploit this correlation between observed colours in an image and observable colours under a given illuminant.

This correlation is performed in five stages: choice of colour space representation, characterisation of reference illuminants, characterising the input image, correlating the image information with the illuminant characteristics, and finally selecting the best illuminant estimate. Steps 1 and 2 are concerned with setting up the correlation matrix (and are typically performed only once for a given device), whereas steps 4 and 5 perform the actual scene illuminant estimation.

3.1. Colour space selection

Selecting an appropriate colour representation is the initial step for building the correlation matrix. Since - as

$$E(\lambda) S(\lambda) = \frac{E(\lambda)}{k} k S(\lambda)$$

- it is not possible to recover the brightness of the illuminant, the choice for a chromaticity space seems natural. One possibility of such as space is given by \((R/(R+G+B), G/(R+G+B))\). However, as the distribution of surface colours within this chromaticity space is highly non-uniform, this might not be the best choice. A chromaticity space with a higher entropy such as

$$\langle q_1, q_2 \rangle = \left( \sqrt[3]{\frac{R}{B}}, \sqrt[3]{\frac{G}{B}} \right)$$

(2)

is hence more desirable.

After the colour space selection follows the problem of quantising the space into $N$ cells of distinct chromaticities. This step is necessary as the correlation matrix is inherently discrete and hence a discrete representation of colour needs to be adopted. On the other hand it is not a crucial step with uniform quantisation schemes from $16 \times 16$ to $64 \times 64$ performing almost equally well [4].

3.2. Reference illuminants characterisation

The next stage is to characterise a set of reference illuminants which essentially leads to determining the entries in the correlation matrix $M$. Due to the discrete nature of the correlation matrix a discrete set of maximally $N$ possible lights has to be chosen. This feature can be automatically used to constrain the solution to only those illuminants that are likely to be encountered in real images. As pointed out in [3], placing an illuminant constraint offers vastly improved recovery performance compared to the unconstrained case. Also, if a priori information on the possible lights is available it can be directly encoded into the choice of reference illuminant set.

Each of the $N_{ill} \leq N$ selected lights is mapped to one column of $M$. Conversely, each of the $N$ chromaticity cells occupies one row of the matrix. The entries of the $N \times N_{ill}$ correlation matrix now simply encode whether it is physically possible that the given chromaticity could have come from a surface captured under the given light\(^1\). The simplest form of doing this is in a binary fashion where 1 declares that the colour could stem from a scene taken under the light and 0 states that it is physically impossible to obtain this chromaticity under a given illuminant.

While this binary representation in essence follows the strategy laid out by gamut mapping algorithms [3] with the gamut for a certain illuminant being marked by 1s and the space outside the gamut by 0s, the correlation framework offers further possibilities. In particular, rather than restricting the matrix entries to binary values, probability information are incorporated into the correlation matrix. Each entry of $M$ then records a measure of probability for each colour/illumination combination. Indeed, this extension has

\(^1\)Note that for this the device sensitivities $Q$ need to be known.
been shown to provide improved colour constancy performance compared to the binary approach which performs similar to gamut mapping algorithms [4]. Since in a later stage probabilities are to be combined, i.e. the probability values multiplied, for computational efficiency the correlation matrix is filled with \( \log \) probabilities.

### 3.3. Input image encoding

The colours of the input image are quantised according to the quantisation of the colour space in Section 3.1 (i.e. to \( N \) chromaticities). Then an \( N \times 1 \) image vector \( \vec{v} \) is built which is similar in form to one column of \( M \), that is each entry in \( \vec{v} \) corresponds to one quantised chromaticity. Since from a colour constancy point of view the relative frequencies of colours in an image are irrelevant the only information that is recorded in \( \vec{v} \) is the presence or absence of each colour by setting each entry to either 1 or 0.

### 3.4. Correlation

Once the correlation matrix \( M \) and the image vector \( \vec{v} \) have been generated, the correlation between the two can be calculated. Correlation is often determined as a vector dot product; if two vectors \( \vec{a} \) and \( \vec{b} \) strongly correlate then their dot product \( \vec{a} \cdot \vec{b} \) will be large. Since each column of \( M \) represents one illuminant its dot product with \( \vec{v} \) determines the correlation between this light and the input image. In terms of the whole matrix \( M \) the correlation to all illuminants is calculated by

\[
L_{C\text{BC}} = \vec{v}^T M
\]  

### 3.5. Illuminant selection

\( L_{C\text{BC}} \) from Equation 3 is a \( 1 \times N_{ill} \) vector whose entries describe the likelihoods for the image described by \( \vec{v} \) to have been captured under each of the \( N_{ill} \) lights. The remaining task is to select one of these lights as the estimated solution to the colour constancy problem. Choosing the light with the maximum value of \( L_{C\text{BC}} \) provides a straightforward way to selecting an illuminant estimate and also defines a well founded maximum likelihood solution to the colour constancy problem.

However, in contrast to other algorithms the correlation framework offers more than just a single answer. Since the entries of \( L_{C\text{BC}} \) represent likelihoods these values can be used to obtain a confidence measure for a given solution. Alternatively more than one possible illuminant estimate can be provided together with the respective likelihoods. Finally, the whole vector \( L_{C\text{BC}} \) can be returned which provides a likelihood for each chosen light source. It is this latter result that will be utilised in the algorithm described in this paper.

### 4. Physics-based robust dichromatic colour constancy

#### 4.1. Dichromatic reflectance model

The dichromatic reflection model for inhomogeneous dielectric objects (such as paints, plastics and paper) states that the colour signals for such objects are composed of two additive components, one being associated with the surface reflectance and the other describing the body (or Lambertian) reflectance part [9]. Both of these components can further be decomposed into a term describing the spectral power distribution of the reflectance and a scale factor depending on the geometry. Making the roles of light and surface explicit and incorporating that the index of refraction does not change significantly over the visible spectrum and \( S_I(\lambda) \) can hence assumed to be constant, the colour signals of a dichromatic object can be expressed as

\[
C(\theta, \lambda) = m_I(\theta)E(\lambda) + m_B(\theta)S_B(\lambda)E(\lambda)
\]  

where \( m_I \) and \( m_B \) are the corresponding weight factors depending on the geometry \( \theta \) which includes the incident angle of the light, the viewing angle and the phase angle.

Substituting Equation 4 into Equation 1, denoting \( R, G, \) and \( B \) as the red, green, and blue pixel value outputs of the digital camera, and making the observation that the interface reflectance is equal to the RGB of the illuminant explicit leads to

\[
\begin{pmatrix}
R \\
G \\
B
\end{pmatrix}
= m_I'(\theta) \begin{pmatrix}
R \\
G \\
B
\end{pmatrix}_E + m_B(\theta) \begin{pmatrix}
R \\
G \\
B
\end{pmatrix}_B
\]

The colour signals of a dichromatic object hence fall on a two-dimensional plane in RGB space with the illumination RGB being one of the vectors spanning the plane.

#### 4.2. Dichromatic colour constancy

The fact that the illuminant vector is contained in the dichromatic plane is obviously very useful when solving for colour constancy. It follows that the intersection of two or more dichromatic planes will yield the illuminant [10]. Thus, if \( V \) and \( W \) are two \( 3 \times 2 \) matrices that contain the span of two dichromatic planes an illuminant estimate based on these two planes can be identified through solving for

\[
V_{E} = W_{d} \quad \text{or} \quad [VW] \begin{pmatrix}
E \\
d
\end{pmatrix} = 0
\]

Integrating more than two surfaces is usually accomplished by finding the best intersection of all planes. A derivation of such a solution where ‘best’ is defined in a least-squares sense was derived in [5] and shown to result in an eigenvector problem.
4.3. Robust dichromatic colour constancy

While dichromatic colour constancy is theoretically able to provide an illuminant estimate based on as few as two surfaces, this nice theory rarely translates into practise. It has been demonstrated that due to noise and insufficient segmentation colour constancy based on the intersection of dichromatic planes only works well enough under lab conditions with highly saturated surfaces and under controlled lighting [6]. We therefore employ a robust approach for colour constancy based on the dichromatic model where rather than finding the best overall intersection of all planes as in [5], the result of each intersection of pairs of two dichromatic planes is utilised [8]. Proceeding this way has the advantage that it allows to discard those intersections that would lead to inaccurate illuminant estimates as is explained in the following.

4.3.1 Robust dichromatic estimates

Eliminating unlikely or inaccurate illuminant estimates proceeds in several stages. Let denote \( A_i = [a_i b_i] \), \( i = 1, ..., N_{pl} \) the definition of a dichromatic plane which is found through the application of SVD on the colour signal matrix\(^2\). Since, due to noise or other reasons such as incorrect segmentation illuminant estimates may fall outside the gamut of common light sources, in a first step those dichromatic planes are discarded that do not intersect the convex hull \( C_{ill} \) of common lights. Alternatively, relaxing this scheme, those planes are discarded that do not pass close enough to \( C_{ill} \) where ‘close’ is defined as the minimal angle between the plane \( A_i \) and any vector in \( C_{ill} \)

\[ \angle(A_i, C_{ill}) = \min \{ \angle(A_i, c_j), c_j \in C_{ill} \} \] (7)

which is compared to a (small) threshold \( \tau_p \). Applying this constraint leaves \( N_{sel} \) dichromatic planes \( (N_{sel} \leq N_{pl}) \).

From the \( N_{sel} \) planes \( N_{allints} = \binom{N_{sel}}{2} \) intersections can be calculated. However, due to noise not all intersection will provide good estimates. As has been shown in [6] the intersection of dichromatic planes with similar orientations is far less stable than an estimate based on planes that are close to orthogonal to each other. Estimates of planes with similar orientations will therefore provide little accuracy and should not be considered. Hence, a threshold \( \tau_a \) is set which defines the minimal angle between two dichromatic planes so that the resulting intersection will be considered in later stages of the algorithm.

Since an intersection \( e_k \) of two dichromatic planes might still produce a physically impossible or unlikely estimate a further constraint is applied which ensures that \( e_k \) falls inside the convex hull \( C_{ill} \) of common lights or close to it. A threshold \( \tau_i \) is defined for the distance between \( e_k \) and \( C_{ill} \) where distance is defined similar to Equation 7 as the minimal angle between \( e_k \) and any vector in \( C_{ill} \)

\[ \angle(e_k, C_{ill}) = \min \{ \angle(e_k, c_j), c_j \in C_{ill} \} \] (8)

Applying \( \tau_a \) and \( \tau_i \) leaves \( N_{ints} \) \( (N_{ints} < N_{allints}) \) intersection for the illuminant estimation.

To summarise only those intersections \( \{e_k, k = 1, ..., N_{ints}\} \) that satisfy \( \{A_i A_j \} \left[ \begin{array}{c} p_i \\ p_j \end{array} \right] = 0 \) subject to

\[
\begin{align*}
1. & \quad \angle(A_i, C_{ill}) < \tau_p \\
2. & \quad \angle(A_j, C_{ill}) < \tau_p \\
3. & \quad \angle(A_i, A_j) > \tau_a \\
4. & \quad \angle(e_k, C_{ill}) < \tau_i
\end{align*}
\] (9)

are considered\(^3\).

4.3.2 Integration of intersection estimates and selection of solution

Following the procedure above produces \( N_{ints} \) intersections of dichromatic planes many of which should provide a good estimate to the colour constancy problem. A simple way of integrating these estimates would be to take the average vector of all intersection as a solution for the algorithm. However, a different approach is taken in here.

Rather than allowing every possible colour (within the illuminant hull) to serve as a potential solution, a discrete method is employed. A set of \( N_{ill} \) illuminants \( E_i \) is selected beforehand. As has been pointed out in Section 3.2 choosing a discrete set of possible lights allows the implicit integration of any type of constraint on the illumination.

Having selected a set of reference lights, the intersection estimates need to be combined to provide a single solution. Two possibilities of such a combination have been explored. The first one employs a voting strategy where each intersection \( e_k \) is voting for the nearest of the \( N_{ill} \) lights. The vote counts are then normalised by \( N_{ints} \) to produce likelihoods for each light. To put it formally an \( 1 \times N_{ill} \) solution vector \( l_{DCC} \) is formed as

\[ l_{DCC}(i) = \Pr(\angle(e_k, E_i) < \angle(e_k, E_j) \quad \forall j, j \neq i) \] (10)

An alternative way of defining a solution vector is to increment the likelihood for each illuminant by \( \frac{1}{\angle(e_k, E_j)} \). This method might be preferable as is less dependent on the spatial distribution of the reference lights. If for example two

\[^2\]Images are divided into \( 30 \times 30 \) subblocks; subblocks whose first two eigenvectors capture at least 98.5% of the variance are deemed dichromatic.

\[^3\]It is apparent that the 4-th constraint implicitly encodes the first two conditions, too. However, these two constraint are explicitly expressed to make the algorithm computationally less expensive.
lights are close to each other (and close to the real scene illuminant) it is likely that each of them will get about half the votes according to Equation 10 which could lead to neither of them ending up with the highest likelihood. In contrast, adding a measure of distance for each $e_l$ will generate likelihoods independent from each other which in the above scenario will still ensure that either of the two lights will be the most likely one. Formally the solution vector $l_{DCC}$ is thus defined as

$$l_{DCC}(i) = \sum_k \frac{1}{l(e_k, E_i)}$$

(11)

Based on $l_{DCC}$ a single solution has to be identified. Similar to Section 3.5 a maximum likelihood selection can be performed, i.e. the illuminant with the highest value in $l_{DCC}$ is returned

$$e = E_s \quad \text{with} \quad s = \arg \max_{s} (l_{DCC}(s) > l_{DCC}(t), \quad \forall t, t \neq s)$$

(12)

However, similar to the correlation framework, more information than just the single illuminant $e$ can be provided. The likelihoods of the solution vector $l_{DCC}$ can be used to quantify the confidence in the solution or to produce error bars of the recovery. Rather than just a single estimate a few most likely illuminants can be returned together with their respective likelihoods. Finally, the whole vector $l_{DCC}$ could be provided for further processing.

5. Combined physical and statistical colour constancy

Looking at the structure of the solution vector $l_{DCC}$ from the dichromatic algorithm in Section 4 it is immediately apparent that its output is very similar to that of $l_{CbC}$ of the Colour by Correlation framework from Section 3. Of course, this is by no means coincidental. Since the aim of the work presented in this paper is an attempt at a colour constancy algorithm that exploits both physical and statistical evidence, both sides of the algorithm have been designed to match each other. Due to this design integrating the two algorithms into a single unified physical-statistical colour constancy algorithm is fairly straightforward.

The correlation framework returns a likelihood vector $l_{CbC}$ whereas the dichromatic part provides the solution vector $l_{DCC}$. It is apparent that the same reference lights are to be used for both algorithms so that each entry of $l_{CbC}$ corresponds to one entry in $l_{DCC}$. $l_{CbC}$ contains log-likelihoods for each reference light whereas $l_{DCC}$ contains likelihoods according to either Equation 10 or Equation 11. In order to combine the two results first the exponential of $l_{CbC}$ has to be taken to transform log-likelihoods into likelihoods. Then both $l_{CbC}$ and $l_{DCC}$ are normalised by their respective maximum and cube rooted to attenuate spikes in the distribution

$$l_{XXX} = \sqrt[3]{\frac{l_{XXX}}{\max(l_{XXX})}}$$

(13)

Once $l_{DCC}$ and $l_{CbC}$ have been obtained from an image the only step left is to combine the two to form a solution to the colour constancy problem based on both physical and statistical evidence. Obviously the simplest would be to add $l_{DCC}$ and $l_{CbC}$ to yield $l_{comb}$ the combined solution vector. Extending this to allow different weights for the statistical and physical contributions gives

$$l_{comb} = \omega l_{CbC} + (1 - \omega) l_{DCC}$$

(14)

with $\omega \in [0; 1]$ where $\omega$ denotes the weight for the correlation algorithm and $(1 - \omega)$ that of the physics-based approach. Setting $\omega > 0.5$ will hence favour the statistical side while $\omega < 0.5$ puts more emphasis on the physical algorithm.

Selecting the maximum likelihood entry from $l_{comb}$ provides a combined illuminant estimate. Alternatively, the few top ranking lights can be returned together with their likelihood values which can be used as a confidence measure.

6. Experimental results

An extensive set of experimental results was carried out in order to evaluate how well the approach introduced works on a set of real images. The dataset that was used for this purpose is provided by the Computer Vision Lab at Simon Fraser University (SFU). These images were captured with the intent of providing a benchmark set for colour constancy algorithms [1]. The ‘Mondrian’ subset which consists of images of 22 objects taken under up to 11 different lights was chosen for the experiments reported here.

For the physics-based algorithm the thresholds $\tau_p = 5^\circ$, $\tau_a = 10^\circ$, and $\tau_t = 5^\circ$ were used to eliminate unlikely solution estimates. From the remaining intersections the solution vector $l_{DCC}$ was built using Equation 11 i.e. by accumulating the inverse angular distances to each of the reference illuminants\footnote{The approach using Equation 11 gave slightly worse results which are hence not listed here.}. Two sets of reference lights were used. The first one comprised 87 lights used in [1] which were chosen so as to fill the gamut of possible illuminants. For the second light set a priori knowledge was assumed and the possible illumination restricted to the 11 SFU lights.

For the statistical side of the algorithm the $(q_1, q_2)$ chromaticity space from Equation 2 was used due to its superior entropy characteristics. A uniform quantisation of
The results for the SFU dataset are given in Table 1 and are provided in terms of median and mean angular error between the estimate and the real scene illuminant. The results for the combined physical/statistical algorithm are based on a combination of the solution vectors \( l_{CbCn} \) and \( l_{DCCn} \) according to Equation 14 with \( \omega = 0.5 \) i.e. equal weightings for the algorithms. The median error of the combined method for all 223 images based on 87 reference illuminants is 2.20 degrees and the mean error 4.50°. This compares favourably to the 2.72/5.59 (median/mean) and 2.52/5.68 degrees achieved by the statistical and physical approaches on their own. When applying prior knowledge through the use of the 11 SFU lights the resulting median and mean errors are 0.96 and 3.70 respectively with the parameters specified above. The two vectors were integrated following the procedure outlined in Section 5 and the maximum likelihood candidate chosen as the final solution estimate.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Median Error</th>
<th>Mean Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Col by Corr 87 lights</td>
<td>2.72</td>
<td>5.59</td>
</tr>
<tr>
<td>Col by Corr SFU lights</td>
<td>1.08</td>
<td>4.97</td>
</tr>
<tr>
<td>Rob Dichr CC - 87 lights</td>
<td>2.52</td>
<td>5.68</td>
</tr>
<tr>
<td>Rob Dichr CC - SFU lights</td>
<td>1.06</td>
<td>4.39</td>
</tr>
<tr>
<td>Combined CC - 87 lights</td>
<td>2.20</td>
<td>4.50</td>
</tr>
<tr>
<td>Combined CC - SFU lights</td>
<td>0.96</td>
<td>3.70</td>
</tr>
</tbody>
</table>

Table 1. Performance of algorithms on SFU dataset.

In [7] the Colour By Correlation algorithm was shown - based on the same dataset - to perform better than gamut mapping colour constancy which in turn was shown to outperform many other approaches [1].

3In [7] the Colour By Correlation algorithm was shown - based on the same dataset - to perform better than gamut mapping colour constancy which in turn was shown to outperform many other approaches [1].

7. Conclusions

In this paper a first approach at deriving a truly combined physical and statistical colour constancy algorithm was presented. Statistical knowledge on surfaces and lights is correlated with the colour information in an image by a version of the Colour by Correlation algorithm. Physical knowledge is exploited in an algorithm based on the dichromatic reflectance model. Both techniques are inherently discrete and return likelihoods for a set pre-defined reference illuminants. By integrating the two types of likelihoods a combined physical/statistical solution is extracted.

Results obtained from the Simon Fraser image dataset confirm that the combined approach provides excellent colour constancy and significantly outperforms both statistics-based and physics-based algorithms.

References