

# Colour Constancy Based on Model Selection

Chuan Li, Peter M. Hall  
Department of Computer Science  
University of Bath

## Abstract

Colour constancy is the problem of mapping colours in a scene as the lighting environment changes. This problem has been studied for many years, and linear solutions that give excellent results are available for Lambertian scenes. More complex reflective environments demand non-linear transforms, but the models are harder to fit and in some cases bring worse results due to over-fitting. We introduce a fully non-linear colour constancy model and a method to select an optimal model from amongst a family of discrete models, including linear models. Once given a set of colour matches our model selection is fully automatic. The selection method is experimentally verified, and is put to use in two applications: scene re-lighting and environment mapping.

## 1 Introduction

The goal of colour constancy is to achieve consistent colour relationships as illumination changes. This requires mapping a source colour distribution to a target. Early work from Forsyth [9, 10] independently scaled each component of a source colour  $\mathbf{x} = (x_1, x_2, x_3)$  to obtain a target  $y_i = \alpha_i x_i$ ; this is the so-called "diagonal" model (also called von Kries model) because the scale coefficients occupy the diagonal of a  $3 \times 3$  matrix. These  $\alpha$  are selected by model fitting. Finlayson showed that a 2-dimensional colour gamut (division by blue) makes for easier fitting [5] [4] [6] [8]. Later, he and others also showed the accuracy for diagonal model can be improved [7] [11], especially via a sharpening transformation [12].

The diagonal model is of benefit when an automated solution is required. However, the accuracy of diagonal transforms can be improved upon using general linear transform model [15]. Even more using non-linear transforms [14], these authors restrict themselves to quadratic transforms of the forms  $y_i = \gamma_i + \sum_{j=1}^3 (\alpha_{ij} x_j^2 + \beta_{ij} x_j)$ . We refer to this as the "diagonal quadratic" model because they do not have cross items  $x_1 x_2$ ,  $x_2 x_3$  and  $x_3 x_1$ . The cost of improved accuracy is the need for a greater number of colour matches between source and target, often available only via user interaction. This strongly suggests that task plays an important role in the selection of an appropriate colour transform. Yet task is not the only governing criteria: the optimally accurate transform depends on scene content as much as on illumination change. Otherwise said, there is no single transform that is accurate for all objects in a given scene [3].

To see this consider the following example. We cannot distinguish red from white when viewed under a red light. Of course, both will be transformed to the same target

colour. This is against the fact, as we can differ these two colour in any other lights that are not pure red. The aim of colour constancy is therefore to find an optimal model for particular scenes.

Our work makes the following two contributions:

- we introduce a complete quadric transform including crossed items of the form  $x_i x_j$  to give a general quadratic model.
- we select the optimal model from all those available (diagonal, general linear, diagonal quadratic and general quadratic) rather than adhere to any single one.

The general model we use can be written as a tensor form:  $y_i = [\mathbf{x}; 1]^T \mathbf{Q}_i [\mathbf{x}; 1]$ ;  $i = 1 : 3$  are for red, green and blue channel respectively. We also bring in the homogeneous coordinate (appending 1) in order to have the linear part of the transform. Each layer of the tensor  $\mathbf{Q}$  is a  $4 \times 4$  triangular matrix, denoted  $\mathbf{Q}_i$ , that represents the mapping for a colour channel from source to target. Model fitting requires estimating all of the elements in  $\mathbf{Q}$ , given a set of sample matches.

In this paper we consider four models: diagonal linear, general linear, diagonal quadratic and general quadratic (DL, GL, DQ, GQ for short). For all these models, we constrained the entry  $\mathbf{Q}_{i44}$  to be zero. This means we don't translate all colours by the same amount (Experiments show translation plays an insignificant role in colour constancy problem). Given a coloured image and a set of colour matches, our model selection strategy selects the optimal model (see Section 2). The strategy takes the quality of the matches into account: quality being related to the distribution of the coloured matches within the distribution of the image as a whole. The rationale behind our selection mechanism is given empirical support by experiments described in Section 3. The research in this paper was motivated by the two applications we describe in Section 4. The first application is colour modification of scene for which no target exists — as if the scene were re-lit under a different illuminant. We show how matches from an example source and target pair taken from a different scene can be used to re-light the given, novel scene — this includes picking the best model. The second application is environment mapping in which we create a synthetic reflection in a coloured reflector — a coffee pot in our case.

## 2 Model Selection

Our model selection picks the best fitting model for a specific scene based on  $n$  colour matches  $(\mathbf{x}_j, \mathbf{y}_j)$ , provided by user interaction. The key to our approach is to consider both the fitting error  $\lambda$  and stability  $\nu$  for each model. We find the model that minimises some linear combination of these two.

Fitting error,  $\lambda$ , is the distance between the transformed colours of source sample and their corresponding target colours in RGB space. In this case the transform is based on the full set of samples given by the user. More specifically, we define  $\lambda$  for each channel  $i = [1, 3]$ .

$$\lambda_i = \frac{1}{n} \sum_{j=1}^n (\mathbf{x}_j^T \mathbf{Q}_i \mathbf{x}_j - y_{ij}) \quad (1)$$

This allow us to select different model for different colour channels.

Stability,  $\nu$ , on the other hand, is computed from a collection of images  $g^k(\mathbf{x})$ . Each image is the transform of the full source image, but now based on the  $k$ th subset of the given matches; this is a "leave one out" strategy.

$$\nu_i = \frac{1}{|S|} \sum_{x \in S} \left( \frac{1}{N} \sum_{k=1}^N (g_i^k(\mathbf{x}) - \bar{g}(\mathbf{x})) \right) \quad (2)$$

in which  $N$  is the number of subsets chosen and  $S$  is the spatial domain of the source image,  $\bar{g}(\mathbf{x})$  is the average of all transformed images.

A standard way to select a model given accuracy and variance is to defined an "energy" as their sum:  $\xi = \lambda + \nu$  (see [2]). We found this to be insufficient, because of the quality of the samples. It is well known that a model fitted to tightly spaced points often give a poor account for distant point. This implies we will observe a high variance, if we successively select subsets of tightly points. This effect is amplified for non-linear models when compared to linear models. Our model selection strategy takes this into account: it prefers linear models when the given samples a tightly spaced, when compared to the distribution of colour in the image as a whole.

The optimal model for the  $i$ th colour channel minimises  $\xi$  given by

$$\xi_i = w\lambda_i + (1 - w)\nu_i \quad (3)$$

Here  $w$  is a weight. The role of  $w$  is to give some allowance for the "quality" of the sample points. Larger  $w$  means higher quality. This linear combination indicates when the quality of the sample is low,  $\xi$  is mainly decided by stability  $\nu$ ; when the quality of the sample is high, we credit more on the fitting error  $\lambda$ . This follows our intuition and has been experimentally proved in Section 3. (see Figure 2)

We designed two forms of weights for different applications. An easy way to choose  $w$  is via the ratio  $r$  of convex hulls: the volume contained by the samples to the volume of all colours in a scene. This definition works well when all the colours in samples belong to colours in the scene. However in the applications such as scene re-lighting, we aim to use colour samples from one scene to reproduce the illumination change for another. In general the colour distributions for the source image of scene A and the image of scene B will differ, so we are likely to have colours sampled from A that don't exist in B. In this case, we redefined  $r$  as  $r = L/\rho$ , in which  $\rho$  is the mean of the distance between all the colours in B and their nearest neighbours in the samples taken from A. The constant  $L$  is set in advance, we have found  $L = 0.01$  is a suitable value (see Figure 4 and here we assume colour components in the range  $[0, 1]$ ).

However, we found that simply setting  $w = r$  gave poor results — model switching is better served when  $w$  is a sigmoid function of  $r$ , specifically:

$$w = \frac{1}{1 + \exp(u(c - r))} \quad (4)$$

The values for the cutoff  $c = 0.7$  and the gain  $u = 20$  were decided by experiment (Figure 2 and 4).

A note on estimating the parameters of a given model. This is easily done by factoring out the  $\mathbf{Q}$  as a vector to obtain the homogeneous linear system  $\mathbf{X}\mathbf{Q} = 0$ . The design matrix  $\mathbf{X}$  comprises rows built from matching pairs. The fully quadratic model uses rows in the

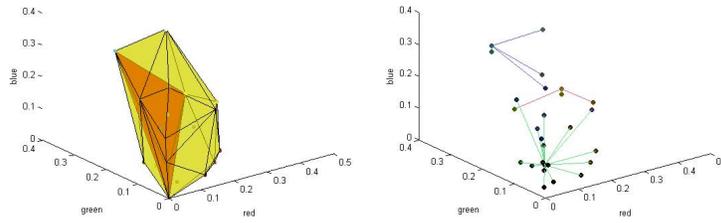


Figure 1: Two ways to estimate the sample quality. Left, the sample quality is a function of the ratio  $r$  of convex hulls: the volume contained by the samples to the volume of all colours in a scene. Right, the sample quality is a function of the mean of the distance between all the colours in the scene and their nearest neighbours in the samples. Here the sample has 3 different colour matches.

design matrix of the form

$$[x_1x_1, x_2x_2, x_3x_3, x_1x_2, x_1x_3, x_2x_3, x_1, x_2, x_3, -y]$$

for each colour channel and for each matching pair. Less general models use design matrices with appropriate terms removed. The aim is to ensure a design matrix with the smallest possible number of columns, and with at least as many rows as columns. The singular value decomposition of the design matrix then yields a solution in the least squares sense.

### 3 Experiment

This section details our experiment to show that the optimal model for a given scene depends on the quality of the samples.

Recall we constrained the candidate model into four forms: diagonal linear, general linear model, diagonal quadratic and general quadratic. These models have different degrees of freedom but as long as we have sufficient numbers of colour matches, we can find least square solutions for them. The image we use are from colour constancy test database collected by the colour group at Simon Fraser University <sup>1</sup>. In this database the scenes have been set up such that only the colour of illumination changes. Hence images from the same scene can be used as the ground truth to each other.

Given same colour matches, we estimate the transform  $\mathbf{Q}^k$  for each of the four candidate models  $M^k$  — we make no choice of model during the experiment. Note that superscripts index models so  $k \in [1, 4]$ ; subscripts  $\mathbf{Q}_i$  index colour channels. Having estimated a transform we apply it to transform the colours in a source image  $f(\mathbf{x})$  to obtain image  $g^k(\mathbf{x})$ . The rms error between  $g^k(\mathbf{x})$  and the ground truth target picture  $h(\mathbf{x})$  indicates the fitting accuracy.

To make a randomised trial the experimenter chooses a large number  $m$  of samples as the full sample set. In our experiment this full sample set contains 24 colour matches from the *GretagMacbeth ColorChecker<sup>TM</sup>* [1] chart. Then we randomly select numbers of matches from them to estimate transformations for each model. Because the

<sup>1</sup><http://www.cs.sfu.ca/colour/data>

general quadratic model has the highest degree of freedom 9, so each selected sample subset should at least contains 9 matches. Our test images cover 10 different illumination changes of the same scene. For each illumination change, we select 1600 subsets containing matches of number from 9 to 24. Using each sample subset, we compute  $\mathbf{Q}^k$  for  $M^k$ . Figure 2 shows the relationship between the sample's quality and the models' performance. We use both definitions of weight to analysis the result.

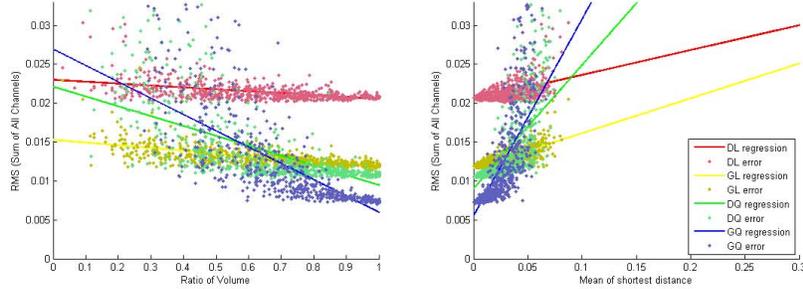


Figure 2: Error grows as sample quality falls. Left: the ratio of convex hulls measures sample quality, larger ratios means better quality. Right: using the mean of the distance between all the colours in the scene and their nearest neighbours in the samples, shorter is better.

From Figure 2 we can see there is a overall decrease of fitting error for all models when the ratio of convex hulls rises or the mean nearest neighbour distance falls. And compared to simple models such as linear ones, the advantage of complex models become much more significant when the quality is high enough (weight approaches 1). On the other hand, the complex models tend to be very unstable when weight is close to zero. This means quality of samples influences the performance of models, especially those nonlinear ones.

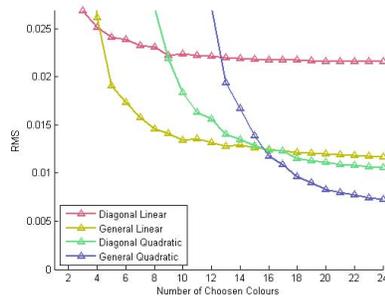


Figure 3: Error as a function of the number of samples. When the number of samples rises, more complex models can be chosen.

Based on this, we design the combination of  $\lambda$  and  $\nu$  based as an interpolation by weight  $w$ . Though this combination is as simple as linear, the effect of  $w$  has been pre-

adjusted by the sigmoid function (see Equation 4). And through *cutoff* and *gain*, we can control the manner of model selection, for example, large cutoff is conservative and small cutoff is greedy.

Though in Figure 2 the GL model (yellow) seems always better than the DL model, this is because we at least have 9 colour matches to estimate the transform. This situation will change if the number of matches falls: the degrees of freedom of a model gives a lower bound on the number of matches. As shown in Figure 3, the DL model (red line) is best when the number of matches is less than 4.

Figure 4 shows the result of our model selection. With increasing sample quality, our model selection result switches from DL to GQ at a volume ratio of about 0.7, which is the cutoff in Equation 4. This value is suggested by Figure 3 and been confirmed by extensive testing across many scenes.

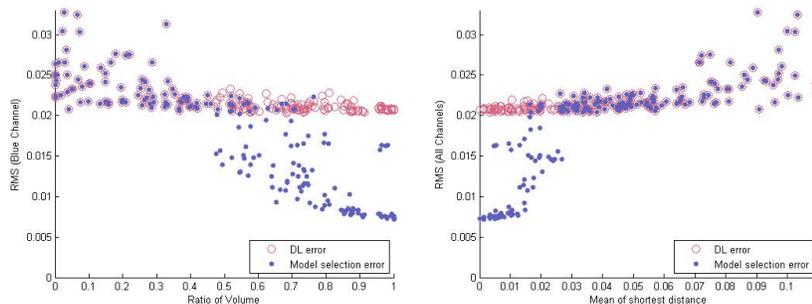


Figure 4: The results of model selection. Left, the result given by weighting the sample quality via ratio of convex hulls. Right, the result given by weighting the sample quality by the mean of the distance between all the colours in the scene and their nearest neighbours in the samples.

Note that although the results presented above are typical, they come from a single scene. We have already commented that model selection is scene depended. Therefore, we cannot conclude that any model is redundant. Our philosophy is that the DL model is the most stable but can be improved using other models if conditions support a switch. Hence conservative choice of parameters in Equation 4 keeps us save.

## 4 Applications

We used our colour model selection method in two applications. The first one is scene re-lighting and the second one is environment mapping into coloured reflectors.

### 4.1 Scene Re-lighting

Scene re-lighting as we intend involves transforming the colours in an image to give the appearance of an illuminance change in the real world scene. We limited changes of illuminant to colour only—we do not consider changes in a light’s spatial location, for example. To re-light a scene B we need to know the colour changes for a subset of its

colours, but we will have no matches — the target picture is to be synthesised. So we use set of matches taken from from other scene, scene A, for which both a source and target set exist.

The simplest approach is to use the matching samples between the source and target for scene A to compute a transform for scene B. However, we know that an optimal transform is scene dependent. Our application attempts to do chose an optimal transform for the scene B, given samples taken from the source and target of scene A. This is possible because accuracy  $\lambda$  is computed from the samples alone, but variance  $\nu$  is computed from the set of transformed images. Since we can always transform scene B, we can compute a variance, and the existence of the sample means we can compute accuracy.

As previously mentioned (in Section 2), in general the colour distributions for the source image of scene A and the image of scene B will differ. So here we use the second definition of the weight: a function of the mean of the distance between all the colours in B and their nearest neighbours in the samples taken from A. Figure 5 shows the best form of transform can vary from scene to scene, and our model selection can help us to choose the best one.

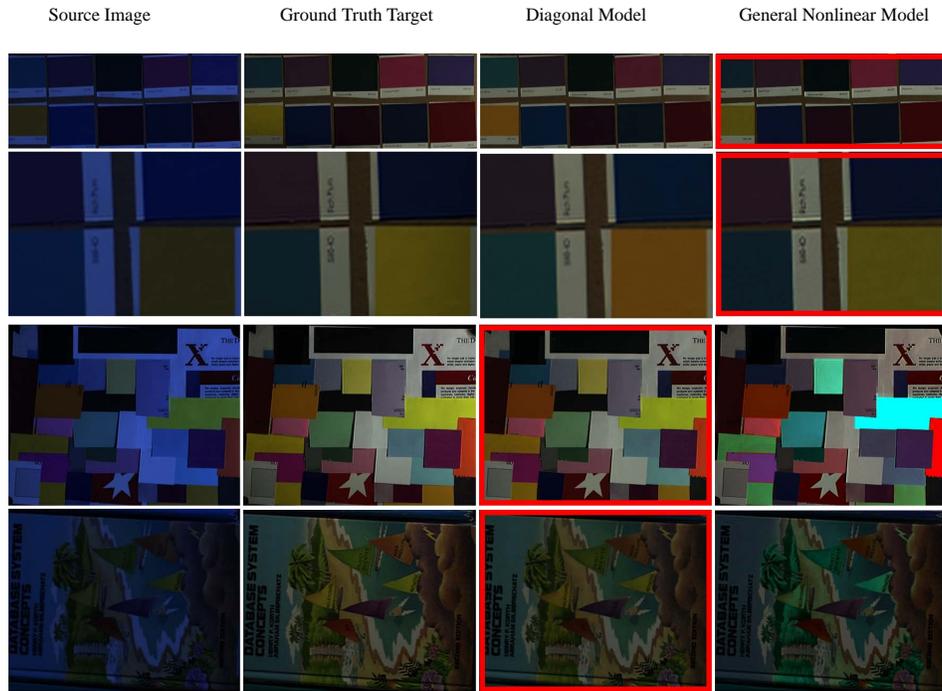


Figure 5: Result of different models for scene re-lighting. Model selection results are indicated by red rectangles. We sampled colours from the first scene (the top row) and use them to produce the same illumination change for different scenes. When the scene can be well described by samples (the first row and second rows), the complex model gives better results. When sample quality is low according to the scene (the third and fourth rows), the simple model tends to be stable. In both cases, our model selection strategy chose the best models.

## 4.2 Environment Mapping

Inserting graphical objects into real world imagery is a contemporary aim that arises in many applications. There are many interesting problems, here we focus on reflecting synthetic scenes in mirrors, which is one kind of environment mapping. The mirror in our example happens to be both coloured and curved, which make the problem even more interesting. A full discussion of reflecting into curved visual objects (without reconstruction in three dimensions) would take us beyond the bounds of this paper, but see [13]. Here we focus on the colour transform problem.



Figure 6: Non-linearity in specular reflection. Bottom row shows the re-lit coffee pot. The left-most pair were created using low quality samples, The right-most pair using high-quality samples. The left of each pair is the result of a diagonal linear model, the right a general quadratic model. The diagonal model is selected for low quality samples, the general quadratic for high quality samples.

The need for general non-linearity is suggested in Figure 6. It shows a silver coffee pot and a green Christmas bauble, captured under identical conditions. The reflections of these two reflectors are both specular. We aim to accurately transform the coffee pot's silver surface to the bauble's green surface (including the reflection). To achieve this, sufficient colour samples and a non-linear transform are needed. This is because the highlights remain approximately a constant colour regardless of the colour of the reflector, but coloured regions away from highlights change to a considerable degree. Figure 6 shows the results given by transforms estimated from different samples and different models. The best result is from general quadratic transform when samples are sufficient, which has been successfully picked up by our model selection strategy. However, when samples are not sufficient, our model selection can also choose the less complex transforms.

Environment mapping is not confined to changing the colour of reflectors, as in Figure 6. It is possible to insert purely synthetic imagery into the reflector and change the

reflector's colour. In this case, we also need to consider the stability of the transform for the insert. This can be easily achieved via considering colour of the insert during computing variance  $v$ . Figure 7 shows how we insert new object into reflectors, and re-lighting the synthetic reflections. In this case a diagonal model was selected. This is because the insertion changes the colour gamut from the colour calibration pictures in Figure 6. Hence, in this case, samples are of lower quality and a stable model is selected



Figure 7: Environment mapping into a coloured reflector. The first picture is the calibration scene. The second picture is a insert object. The third picture is the mapping result for the insert – its colour was transformed to match the silver surface of the coffee pot and its shape was non-linear warped. The fourth picture is the synthetic reflection. The fifth picture is the result of re-lighting the synthetic reflection to the green bauble in Figure 6 .

## 5 Conclusion

We have introduced a fully non-linear colour constancy model and a method to automatically select an optimal model for a given scene. The non-linear model and selection method have been used in two applications. Each application forced novelty of its own: in scene re-lighting we changed the colours of a second picture based on samples taken from an example pair thereby forcing a redefinition of sample quality; in environment mapping we had to account for the non-linearity brought about by specular reflections which motivated the general non-linear model.

Our work has lead us to the following conclusions. First, that although colour constancy has no generally optimal solution, specific solutions can be found. Second, that the optimal model for a specific solution can be selected automatically. The particular model chosen critically depends on the quality of the samples used to calibrate the mapping. Third, non-linear mappings are of value in practical applications such as environment mapping.

We do rely on interaction in that as users must input initial matches. This gives us the space for future work – to produce an fully automatic system. This is challenging because scene content can change dramatically, hence colour matches between them are difficult to be found. But as to applications such as environment mapping, the users will benefit from a fully automatic system as it can reduce their workload.

## 6 Acknowledgements

This work was supported by EPSRC grant. The authors also gratefully acknowledge the colour group at Simon Fraser University for opening their collection of test database.

## References

- [1] *GretagMacbeth Color Management Solutions*. <http://www.gretagmacbeth.com>.
- [2] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley Interscience, 2001.
- [3] M. D. Fairchild. *Color Appearance Models*. Addison Wesley, 1997.
- [4] G. D. Finlayson. Color in perspective. In *IEEE Trans. Pattern Analysis and Machine Intelligence*, volume 18, pages 1034–1038, Washington, DC, USA, 1996. IEEE Computer Society.
- [5] G. Finlayson. *Coefficient color constancy*. PhD thesis, Simon Fraser Univ., School of Computing, Burnaby, BC, Canada., 1995.
- [6] G. Finlayson and S. Hordley. A theory of selection for gamut mapping color constancy. In *CVPR*, volume 00, page 60, Los Alamitos, CA, USA, 1998. IEEE Computer Society.
- [7] G.D. Finlayson, M.S. Drew, and B.V. Funt. Diagonal transforms suffice for color constancy. In *ICCV*, pages 164–171, 1993.
- [8] G.D. Finlayson, B.V. Funt, and K. Barnard. Color constancy under varying illumination. In *ICCV*, volume 00, page 720, 1995.
- [9] D. A. Forsyth. Colour constancy and its applications in machine vision. *PhD thesis, Oxford University, Department of Engineering Science*, 1988.
- [10] D. A. Forsyth. A novel algorithm for color constancy. In *IJCV*, volume 5, pages 5–36, Hingham, MA, USA, 1990. Kluwer Academic Publishers.
- [11] B. V. Funt and Hao Jiang. Nondiagonal color correction. In *ICIP*, pages 481–484, 2003.
- [12] M. Drew G.D. Finlayson and B. Funt. Spectral sharpening: Sensor transformations for improved color constancy. In *Journal of Optical Society of America*, volume 11, pages 1553–1563, 1994.
- [13] P.M. Hall, J.P. Collomosse, Y.Z. Song, P.Y. Shen, and C. Li. Rtcams: a new perspective on non-photorealistic rendering from photographs. to appear in *IEEE Transactions on Visualization & Computer Graphics*.
- [14] A. Ilie and G. Welch. Ensuring color consistency across multiple cameras. In *ICCV*, pages 1268–1275, Washington, DC, USA, 2005. IEEE Computer Society.
- [15] N. Joshi. Color calibration for arrays of inexpensive image sensors. Master’s thesis, Stanford University Department of Computer Science, 2004.