https://doi.org/10.2352/issn.2694-118X.2020.LIM-08

Colour Fidelity in Spectral Reconstruction from RGB Images

Yi-Tun Lin
University of East Anglia, Norwich, United Kingdom

Abstract

Spectral reconstruction (SR) aims to recover high resolution spectra from RGB images. Recent developments - leading by Convolutional Neural Networks (CNN) - can already solve this problem with low errors. However, those leading methods do not explicitly ensure the predicted spectra will re-integrate (with the underlying camera response functions) into the same RGB colours as the ones they are recovered from, namely the 'colour fidelity' problem. The purpose of this paper is to show, visually and quantitatively, how well (or bad) the existing SR models maintain colour fidelity. Three main approaches are evaluated - regression, sparse coding and CNN. Furthermore, aiming for a more realistic setting, the evaluations are done on real RGB images and the 'end-of-pipe' images (i.e. rendered images shown to the end users) are provided for visual comparisons. It is shown that the state-of-the-art CNN-based model, despite of the superior performance in spectral recovery, introduces significant colour shifts in the final images. Interestingly, the leading sparse coding and the simple linear regression model, both of which are based on linear mapping, best preserve the colour fidelity in SR.

1. Introduction

The colour signal (*i.e.* physical radiance) coming from the scene contains spectral information of the light source and object surfaces within visible range, and by human eyes or consumer RGB cameras this signal is captured in 3 separate colour channels (red, green and blue). Clearly, this 'colour image formation' process loses a significant portion of spectral information. Indeed, the continuous spectra are recorded with just 3 values per pixel. Consequently, having only RGB data on hand, our ability to predict colours in different viewing conditions (*e.g.* different lighting and/or sensors) is limited. This limitation has been reported in several colour applications including colour correction [9], colour constancy [1] and device characterisation [7].

It is well known that *hyperspectral imaging devices*, which measure the colour signal at high spectral resolution, are able to provide very accurate colour predictions. For example, camera RGBs can be accurately simulated by integrating the measured spectra with the spectral response functions of the underlying camera. However, these devices are costly and/or bulky and often suffer from long integration time, low light sensitivity and/or low spatial resolution.

Instead of building new devices for capturing hyperspectral information, *spectral reconstruction* problem (SR) studies the mapping from RGB data to its hyperspectral counterpart. It is hoped that the RGBs - which are taken by a single camera (and often under confined illumination conditions) - can recover spectra which can predict colours in other desired viewing conditions. Despite of its highly *ill-posed* nature, this SR problem has been solved with increasing accuracy over the years. Maloney and Wandell [15] first proposed a 3-dimensional linear characterisation of reflectance spectra, which suggests a simple linear transformation between the RGBs and the 3-dimensional reflectances. Followed by the development of least-squares regression [11] and Bayesian approach [6] which further improve the performance of spectral recovery. In the recent literature, highly

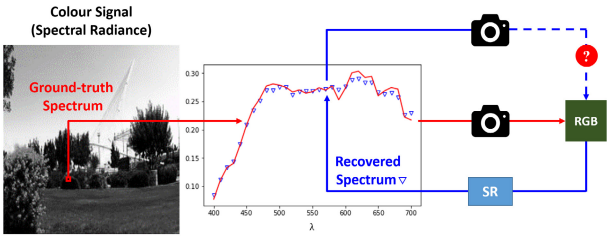

Figure 1. *Spectral reconstruction (SR) and the colour fidelity problem*

data-driven sparse coding [2, 4] and deep neural network models [5, 18] are considered the leading methods for SR.

The main concern of this paper is the *colour fidelity* of the SR algorithms. While the ongoing developments seem to blindly focus on lowering the spectral error of the recovery, they do not clearly guarantee that the recovered spectra are *colorimetrically accurate*, that is whether they can be re-integrated into the RGBs they are recovered from. In fact, the very issue has been addressed (and solved) for several primitive SR models including 3-dimensional reflectance model [3], linear least-squares regression [22] and Bayesian estimation [16]. Yet, it is still underestimated in the recent development of SR. Indeed, as a fact to be presented in this paper, the state-of-the-art SR model [5, 18] introduces significant colour shifts in the re-integrated RGB images. Ironically, this tells that the *best* SR models cannot even predict colours correctly in the **original viewing condition**.

Figure 1 illustrates spectral reconstruction and the problem of colour fidelity. Following the red solid arrows in the graph, a ground-truth spectrum in the scene (the image on the left, and the red solid curve in the plot) integrates with the camera response functions giving the ground-truth RGB (green patch on the right). Then, the SR algorithm recovers the spectrum (following the bottom blue arrow) as the dotted blue curve in the plot. The colour fidelity of SR - which is the main arguing point of this paper - studies that, when the recovered spectrum is re-integrated with the same camera (the top blue arrow), to what extent will this re-integrated colour differ from the ground-truth RGB.

The rest of the paper is organised as follows. In section 2, the background of the tested SR models is briefly introduced. Section 3 begins with the details of quantifying colour differences and image rendering process, followed by result discussion and visual comparisons. This paper concludes in section 4.

2. Background

A simple approximation of the colour image formation process can be written as:

$$\underline{\mathbf{x}} = \mathbf{S}^T \underline{\mathbf{r}}\,, \tag{1}$$

where $\underline{\mathbf{x}}$ is the simulated RGB (a 3-dimensional vector), $\underline{\mathbf{r}}$ is the k-dimensional spectral vector referring to the *discrete hyperspectral measurement* at k points across the visible spectrum (in this paper the visible spectrum runs from 400 to 700 nm and the spectral sampling interval is 10 nm, hence $k = 31$), and $\mathbf{S}$ is a $k \times 3$ matrix with its columns to be the set of 3 camera response functions (spectrally discretised as aligned with the radiance measurements). The premise for this discrete model to work is that

the spectral measurements must be sampled at a sufficient spectral resolution ($k \gg 3$). Conversely, spectral reconstruction (SR) seeks to estimate $\underline{\mathbf{r}}$ from $\underline{\mathbf{x}}$.

2.1. Regressions

Let us consider the linear (LR) [11], polynomial (PR) [8] and root-polynomial regression (RPR) [13]. These models linearly map the 'polynomial colour features' to radiance spectra:

$$\underline{\mathbf{r}} \approx \mathbf{M}\varphi(\underline{\mathbf{x}}) \,, \tag{2}$$

where $\varphi : \mathbb{R}^3 \mapsto \mathbb{R}^p$ is the respective polynomial transformation (mainly non-linear except for LR) and $\mathbf{M}$ is a $k \times p$ regression matrix to be searched for. With $\underline{\mathbf{x}} = (R,G,B)^T$, the corresponding $\varphi(\underline{\mathbf{x}})$ for LR, PR and RPR models are summarised in Table 1. For example, for the 2^{nd} order PR model, $\varphi(\underline{\mathbf{x}}) = (R,G,B,R^2,G^2,B^2,RG,GB,BR)^T$ and $\mathbf{M}$ is $k \times 9$.

To solve for $\mathbf{M}$ in the least-squares sense, given m training pairs $\{(\underline{\mathbf{x}}_j, \underline{\mathbf{r}}_j) \mid j = 1,2,...,m\}$, we are to minimise:

$$\arg\min_{\mathbf{M}} \left(\sum_{j=1}^{m} ||\underline{\mathbf{r}}_j - \mathbf{M}\varphi(\underline{\mathbf{x}}_j)||_2^2 + \gamma||\mathbf{M}||_F^2 \right) . \tag{3}$$

Here γ is a *regularisation parameter* to be empirically selected. The purpose of this regularisation setting is to *stabilise* the regression output, that is to prevent drastic change in result at small perturbations in input RGB. For more information see [11, 20].

Then, Equation (3) is solved in closed-form:

$$\mathbf{M} = \mathbf{R}^T\mathbf{\Phi}(\mathbf{\Phi}^T\mathbf{\Phi} + \gamma\mathbf{I}_{p\times p})^{-1} \,, \tag{4}$$

where $\mathbf{R}$ and $\mathbf{\Phi}$ are respectively $m \times k$ and $m \times p$ data matrices (rows are matching spectra and polynomial features), and $\mathbf{I}_{p\times p}$ is the $p \times p$ identity matrix.

Model	Polynomial Transformation $\varphi(\underline{\mathbf{x}})$
LR	R,G,B
PR	$R,G,B,R^2,G^2,B^2,RG,GB,BR,$ $R^3,G^3,B^3,RG^2,GB^2,BR^2,R^2G,G^2B,B^2R,RGB,$ $\ldots$ (*higher order terms if applicable*)
RPR	$R,G,B,\sqrt{RG},\sqrt{GB},\sqrt{BR},$ $\sqrt[3]{RG^2},\sqrt[3]{GB^2},\sqrt[3]{BR^2},\sqrt[3]{R^2G},\sqrt[3]{G^2B},\sqrt[3]{B^2R},\sqrt[3]{RGB},$ $\ldots$ (*higher order terms if applicable*)

Table 1. Polynomial transformations of LR, PR and RPR

2.2. Sparse Coding

The sparse coding approach [4, 2] assumes that all spectra can be approximated by linear interpolations between *neighbouring* spectral data. In addition, it is assumed that the data distribution of the real-world spectra and RGBs share similar local geometries, such that the interpolating relationship among neighbouring RGBs can be an accurate approximation of the interpolations in the corresponding spectral neighbourhood, namely the *neighbour embedding* assumption [21].

The training stage of the leading sparse coding model 'A+' [2] operates in two steps. First, the algorithm *clusters* the matching spectral and RGB data, which gives a pair of spectral and RGB *dictionaries* (*i.e.* the collection of 'centres' of all clusters), respectively denoted as $\mathfrak{D}_r$ and $\mathfrak{D}_x$:

$$\mathfrak{D}_r = \{\underline{\mathbf{r}}_1,\, \underline{\mathbf{r}}_2,\, \ldots,\, \underline{\mathbf{r}}_M\} \,,\; \mathfrak{D}_x = \{\underline{\mathbf{x}}_1,\, \underline{\mathbf{x}}_2,\, \ldots,\, \underline{\mathbf{x}}_M\} \,; \tag{5a}$$

meanwhile the corresponding *neighbourhoods* (the training data in the proximity of each dictionary member) are also saved:

$$\mathfrak{N}_r = \{\mathbf{R}_1,\, \mathbf{R}_2,\, \ldots,\, \mathbf{R}_M\} \,,\; \mathfrak{N}_x = \{\mathbf{X}_1,\, \mathbf{X}_2,\, \ldots,\, \mathbf{X}_M\} \,, \tag{5b}$$

where the rows of $\mathbf{R}_i$ and $\mathbf{X}_i$ are matching spectra and RGBs in the neighbourhood of the i^{th} dictionary member (*i.e.* $\underline{\mathbf{r}}_i$ and $\underline{\mathbf{x}}_i$). Then, for each neighbourhood the algorithm is bound to find an *interpolation vector* $\underline{\mathbf{w}}_i$ that derives $\underline{\mathbf{x}}_i$ by linearly interpolating the neighbouring data in $\mathbf{X}_i$. That is:

$$\underline{\mathbf{x}}_i = \mathbf{X}_i^T\underline{\mathbf{w}}_i \,. \tag{6}$$

This vector is optimised by linear least-squares regression:

$$\arg\min_{\underline{\mathbf{w}}_i} \left(||\underline{\mathbf{x}}_i - \mathbf{X}_i^T\underline{\mathbf{w}}_i||_2^2 + \gamma||\underline{\mathbf{w}}_i||_2^2 \right) \tag{7}$$

and solved in closed-form:

$$\underline{\mathbf{w}}_i = \mathbf{X}_i(\mathbf{X}_i^T\mathbf{X}_i + \gamma\mathbf{I}_{3\times3})^{-1}\underline{\mathbf{x}}_i \,, \tag{8}$$

where γ is the regularisation parameter as in Equation (3) and (4), and $\mathbf{I}_{3\times3}$ is the 3×3 identity matrix.

On input of a query RGB $\underline{\mathbf{x}}$ in the reconstruction stage, A+ algorithm searches for its nearest neighbour in $\mathfrak{D}_x$, say it turns out to be the q^{th} dictionary member $\underline{\mathbf{x}}_q$. Based on the neighbour embedding assumption, the spectral estimate $\underline{\mathbf{r}}$ is calculated by:

$$\underline{\mathbf{r}} \;\approx\; \mathbf{R}_q^T\underline{\mathbf{w}}_q \;\approx\; \mathbf{R}_q^T\mathbf{X}_q(\mathbf{X}_q^T\mathbf{X}_q + \gamma\mathbf{I}_{3\times3})^{-1}\underline{\mathbf{x}} \,. \tag{9}$$

Interestingly, by comparing Equation (4) and Equation (9), we can clearly see that the A+ algorithm in effect operates *linear regression* (LR) locally at the discretised neighbourhoods.

2.3. Neural Networks

Neural network is a powerful approach to solving complex non-linear mappings. Recent development on deep networks such as Convolutional Neural Network (CNN) and Generative Adversarial Network (GAN) further enables *patch*-based mapping in image processing, that is to map image patches to the desired output. Especially for the SR problem, the hope for these deep-network approaches is that they can potentially learn to identify scene contents that might help spectral recovery.

According to the recent NTIRE 2018 Challenge on Spectral Reconstruction from RGB Images [5], all leading models are based on deep neural networks, among which the CNN-based HSCNN-D and HSCNN-R are the two best models [18].

3. Colour Fidelity in Spectral Reconstruction

Given the ground-truth and recovered spectra, respectively denoted as $\underline{\mathbf{r}}_{gt}$ and $\underline{\mathbf{r}}_{rec}$, the expected colour fidelity of an SR model suggests (see Equation (1)):

$$\mathbf{S}^T\underline{\mathbf{r}}_{gt} \approx \mathbf{S}^T\underline{\mathbf{r}}_{rec} \,, \tag{10}$$

where $\mathbf{S}$ is the matrix of response functions of the RGB camera used to take the training images for the SR model.

3.1. Spectral Reconstruction on Real RGB Data

Model	Approach
LR	Regression
PR6	Regression
RPR6	Regression
A+	Sparse Coding
HSCNN-R	Convolutional Neural Network

Table 2. List of tested SR models

The models listed in Table 2 are tested. The suffix '6' for PR6 and RPR6 indicates that the respective 6^{th} order polynomial expansions are used. The ground-truth hyperspectral images used for training are from the ICVL dataset [4], from which

100 images are randomly selected for training and 50 images for *validation* (in different way for each SR approach, which shall be explained later). The corresponding RGB images are simulated by Equation (1) (colour image formation) using the response functions of SONY IMX135. That is, these models are trained to map the (noiseless) SONY IMX135 raw RGBs to hyperspectral images. The spatial size of the images is around 1300×1392.

The validation process of regression-based models and A+ sparse coding involves searching for the best regularisation parameter γ in Equation (4) and (9), where a set of different γ's are chosen and the validation images are used to test each derived model. As for the CNN-based HSCNN-R model, the validation images are used to test the intermediate models out of every training epoch, in order to determine the stopping epoch of the training process.

In the reconstruction stage, the trained SR models are used to recover hyperspectral images from 139 selected raw RGB images from the INTEL-TAU dataset [12] (by far the largest open-source RGB image dataset for training and evaluating the algorithms for color constancy). All selected images were captured by SONY IMX135, with the spatial dimension of 2448×3264.

Finally, as we are bound to look at the colour fidelity of the reconstruction, the recovered spectral images are re-integrated back into raw RGB images with the camera response functions of SONY IMX135.

3.2. Quantifying Colour Differences

Now, we have on hand the ground-truth raw RGBs (*i.e.* the demosaicked raw images from INTEL-TAU dataset) and the re-integrated raw RGBs (simulated from the SR-recovered hyperspectral images). To quantify the colorimetric errors by CIE 1976 colour difference (ΔE) [17], we must consider how to transform the camera raw RGBs to their CIELAB counterparts.

The procedure is summarised in Figure 2. The INTEL-TAU database provides with each raw image the colour correction matrix (CCM) that transforms the image to sRGB colours and the ground-truth white point (WP) that is crucial for the transformation between sRGB and CIELAB [19]. Then, the ΔE colour differences are calculated between the derived CIELAB images from the ground-truth and re-integrated images.

In each image, the mean and 99 percentile (the *worst-case*) ΔE errors across all pixels are calculated. Then, the mean and standard deviation of the above two measures across all images are calculated, which are provided in Table 3.

Model	Mean ΔE	99 pt. ΔE
LR	0.69 ($\pm$ 0.18)	3.76 ($\pm$ 0.73)
PR6	1.21 ($\pm$ 2.14)	13.96 ($\pm$ 23.91)
RPR6	0.83 ($\pm$ 0.59)	4.35 ($\pm$ 3.71)
A+	0.00 ($\pm$ 0.00)	0.00 ($\pm$ 0.04)
HSCNN-R	1.73 ($\pm$ 0.69)	7.83 ($\pm$ 4.23)

Table 3. Mean ($\pm$ standard deviation) of the mean and 99 percentile ΔE errors of individual images. Best results are shown in green and the worst results are in red.

3.3. Visual Comparisons

Before showing to the end users, a camera raw image might undergo, but not limited to: black level and saturation correction, white balancing, colour correction and gamma correction. As we are already given the expected end-of-pipe image with each raw image in the INTEL-TAU database, we can alternatively build a 3D Look-up-table (LUT) which approximates the actual image processing pipeline.

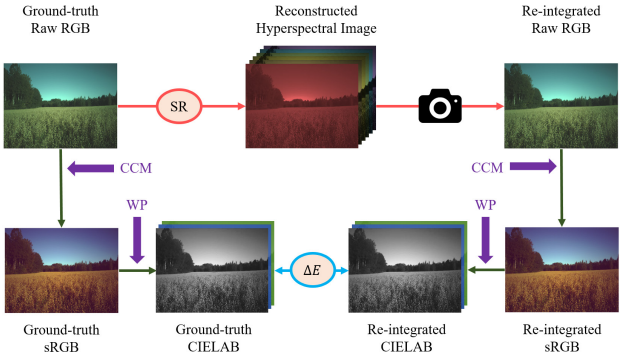

Figure 2. *The process of calculating CIE 1976 colour difference ΔE between ground-truth and re-integrated raw images.*

For each image, the LUT is built to relate the colours in the ground-truth raw RGB image to the colours in the supplied (expected) end-of-pipe image. This LUT can be optimised - in a least-squares sense - by *lattice regression* [10, 14]. To speed up the optimisation process, the LUT are trained on *thumbnail images* (the images are downsampled from the original 2448×3264 to 108×144), and the colours are binned by $24 \times 24 \times 24$ in the three colour channels. Then, the full resolution ground-truth and re-integrated raw RGB images are mapped to their respective end-of-pipe renditions by applying this pre-built 3D LUT.

Visual comparisons are provided in the top rows of Figure 3 and 4, where four regions of interest are selected as the marked areas (by white squares) in the bottom left-most image. The corresponding ΔE error maps between ground-truth and the SR-predicted images are given in the bottom rows of the figures. It is evident that the extent of colour shifts in the images are well consistent with the quantitative ΔE measure.

3.4. Discussion

First, the result in Table 3 teaches that **accurate spectral recovery does not imply colorimetric accuracy**. Indeed, the LR model is expected to deliver the *worst* spectral recovery [13], but it shows good colour fidelity. In contrary, with the state-of-the-art spectral accuracy, HSCNN-R performs poorly in colour fidelity. Second, it is worthwhile to remark that for the A+ model, the number of clusters and the size of each neighbourhood should be properly selected to ensure the colorimetric accuracy of the algorithm. As an extreme case, we may assume a single cluster with all the data in its neighbourhood, then A+ becomes LR.

Also, it appears that linear-mapping based LR and A+ models provide better colour fidelity - those non-linear models seem to create drastic worst-case performance. Lastly, whether the SR model is invariant to exposure change [13] may also lead to colorimetric inaccuracy, which can be interesting for future study.

4. Conclusion

Spectral reconstruction (SR) seeks the estimation of hyperspectral measurements from RGB camera responses. It is well known that accurate 'spectral measurement' guarantees accurate colour predictions, but can accurate 'spectral estimation' give the same promise? This paper reveals the apparent paradox that **the state-of-the-art spectral reconstruction is actually colorimetrically inaccurate**.

In this paper, a thorough benchmark on colour fidelity for three main SR approaches - regression, sparse coding and deep neural network - is conducted. The results are demonstrated both quantitatively by CIE 1976 colour difference (ΔE) and qualitatively as visual comparisons on rendered camera outputs. While the state-of-the-art deep-network model fails to recover spectra

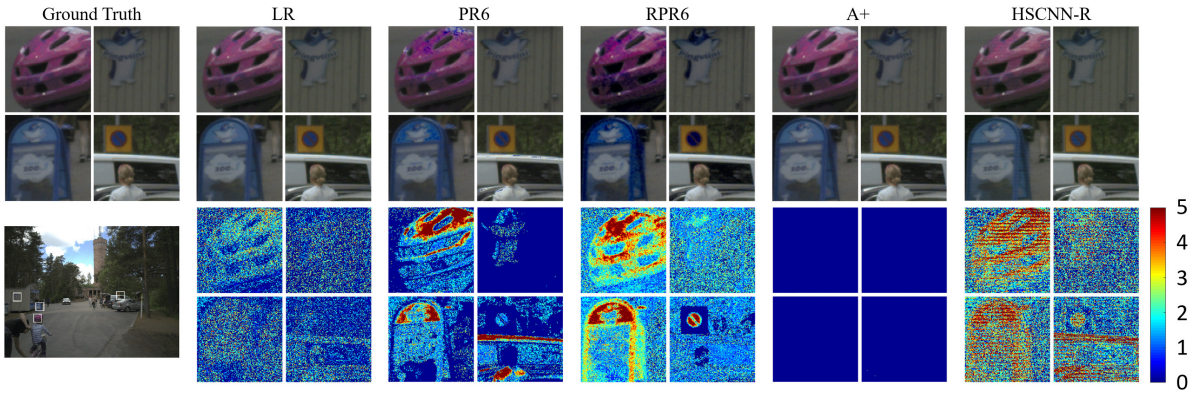

Figure 3. *Colour fidelity test on an example outdoor scene. Top row: the rendered images. Bottom row: the corresponding ΔE error maps.*

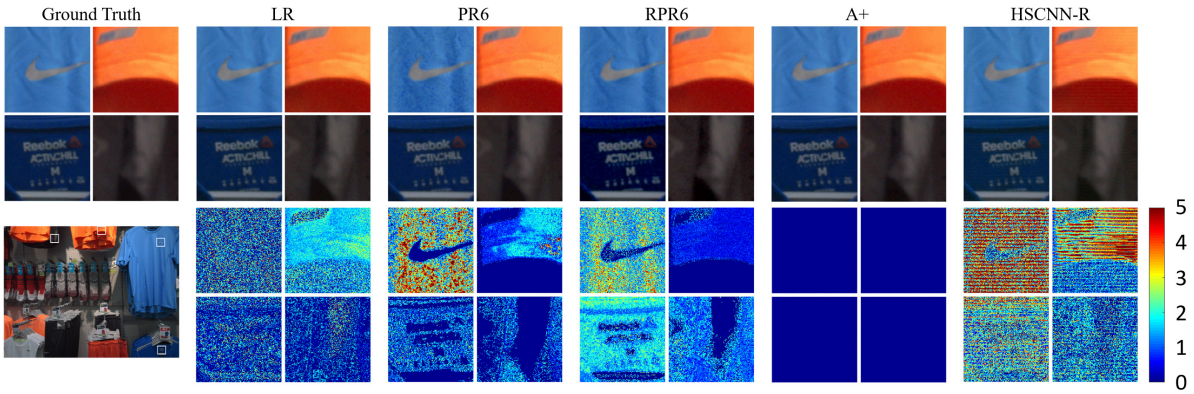

Figure 4. *Colour fidelity test on an example indoor scene. Top row: the rendered images. Bottom row: the corresponding ΔE error maps.*

with good colorimatric accuracy, the primitive linear regression and the leading sparse coding model best preserve the colour fidelity in spectral reconstruction.

References

[1] A. Abrardo *et al.* Color constancy from multispectral images. In *International Conference on Image Processing*, volume 3, pages 570–574, 1999.

[2] J. Aeschbacher *et al.* In defense of shallow learned spectral reconstruction from rgb images. In *International Conference on Computer Vision*, pages 471–479, 2017.

[3] F. Agahian *et al.* Reconstruction of reflectance spectra using weighted principal component analysis. *Color Research & Application*, 33(5):360–371, 2008.

[4] B. Arad and O. Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *European Conference on Computer Vision*, pages 19–34, 2016.

[5] B. Arad *et al.* Ntire 2018 challenge on spectral reconstruction from rgb images. In *Conference on Computer Vision and Pattern Recognition Workshops*, pages 929–938, 2018.

[6] D. H. Brainard and W. T. Freeman. Bayesian method for recovering surface and illuminant properties from photosensor responses. In *Human Vision, Visual Processing, and Digital Display V*, volume 2179, pages 364–376, 1994.

[7] V. Cheung *et al.* Characterization of trichromatic color cameras by using a new multispectral imaging technique. *Journal of the Optical Society of America A*, 22(7):1231–1240, 2005.

[8] D. R. Connah and J. Y. Hardeberg. Spectral recovery using polynomial models. In *Color Imaging X: Processing, Hardcopy, and Applications*, volume 5667, pages 65–75, 2005.

[9] M. F. Deering. Multi-spectral color correction, September 27 2005. US Patent 6,950,109.

[10] E. Garcia and M. Gupta. Lattice regression. In *Advances in Neural Information Processing Systems*, pages 594–602, 2009.

[11] V. Heikkinen *et al.* Evaluation and unification of some methods for estimating reflectance spectra from rgb images. *Journal of the Optical Society of America A*, 25(10):2444–2458, 2008.

[12] F. Laakom *et al.* Intel-tau: A color constancy dataset. *arXiv preprint arXiv:1910.10404*, 2019.

[13] Y. Lin and G. D. Finlayson. Exposure invariance in spectral reconstruction from rgb images. In *Color and Imaging Conference*, volume 2019, pages 284–289, 2019.

[14] H. Lin *et al.* Nonuniform lattice regression for modeling the camera imaging pipeline. In *European Conference on Computer Vision*, pages 556–568, 2012.

[15] L. T. Maloney and B. A. Wandell. Color constancy: a method for recovering surface spectral reflectance. *Journal of the Optical Society of America A*, 3(1):29–33, 1986.

[16] P. Morovic and G. D. Finlayson. Metamer-set-based approach to estimating surface reflectance from camera rgb. *Journal of the Optical Society of America A*, 23(8):1814–1822, 2006.

[17] A. R. Robertson. The cie 1976 color-difference formulae. *Color Research & Application*, 2(1):7–11, 1977.

[18] Z. Shi *et al.* Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In *Conference on Computer Vision and Pattern Recognition Workshops*, pages 939–947, 2018.

[19] S. Süsstrunk *et al.* Standard rgb color spaces. In *Color and Imaging Conference*, volume 1999, pages 127–134, 1999.

[20] A. N. Tikhonov *et al.* *Numerical methods for the solution of ill-posed problems*, volume 328. Springer Science & Business Media, 2013.

[21] R. Timofte *et al.* A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian Conference on Computer Vision*, pages 111–126, 2014.

[22] Y. Zhao and R. S. Berns. Image-based spectral reflectance reconstruction using the matrix r method. *Color Research & Application*, 32(5):343–351, 2007.

Author Biography

Yi-Tun (Ethan) Lin is a Ph.D. student in the Colour & Imaging Lab, School of Computing Sciences, University of East Anglia, UK. He received a joint M.Sc. degree in Colour Science in 2018 from University Jean Monnet (France), University of Granada (Spain) and University of Eastern Finland (Finland), and B.Sc. in Physics in 2016 from National Taiwan University, Taiwan. His research interest is physics and machine learning based spectral reconstruction.