https://doi.org/10.2352/issn.2169-2629.2021.29.19

Investigating the Upper-Bound Performance of Sparse-Coding-Based Spectral Reconstruction from RGB Images

Yi-Tun Lin and Graham D. Finlayson
University of East Anglia, Norwich, United Kingdom

Abstract

In Spectral Reconstruction (SR), we recover hyperspectral images from their RGB counterparts. Most of the recent approaches are based on Deep Neural Networks (DNN), where millions of parameters are trained mainly to extract and utilize the contextual features in large image patches as part of the SR process. On the other hand, the leading Sparse Coding method 'A+'—which is among the strongest point-based baselines against the DNNs—seeks to divide the RGB space into neighborhoods, where locally a simple linear regression (comprised by roughly 10^2 parameters) suffices for SR.

In this paper, we explore how the performance of Sparse Coding can be further advanced. We point out that in the original A+, the sparse dictionary used for neighborhood separations are optimized for the spectral data but used in the projected RGB space. In turn, we demonstrate that if the local linear mapping is trained for each spectral neighborhood instead of RGB neighborhood (and theoretically if we could recover each spectrum based on where it locates in the spectral space), the Sparse Coding algorithm can actually perform much better than the leading DNN method. In effect, our result defines one potential (and very appealing) upper-bound performance of point-based SR.

1. Introduction

The spectral signature of light signals is often regarded as a rich descriptor of object surfaces and light sources. Hence, hyperspectral cameras [14, 16] are developed to record high resolution radiance spectra at each pixel of the scene. This technology is found useful in many industrial applications such as remote sensing [15, 8], medical imaging [35], food processing [27], device color characterization [9, 29] and computer graphics [19, 33]. However, precise spectral measurements often require some form of *scanning*—either spatial, spectral and/or temporal—which results in long integration time, low light sensitivity and/or low resolutions.

Spectral Reconstruction (SR) is one of the approaches for mitigating the need for accurate physical measurements in hyperspectral imaging. In SR, high resolution spectra are recovered from spectral images of fewer spectral channels, and many consider to train SR directly from the easily accessible RGB images, e.g., [4, 5]. This problem may seem insoluble at the first glance. Indeed, only 3 coarsely averaged spectral measurements are recorded at a pixel of an RGB image—respectively in the Red, Green and Blue spectral regions, and to obtain an SR-from-RGB map would mean we are to solve a 3-to-n inverse problem [31] (where $n \gg 3$ is the number of sampled wavelengths of the hyperspectral measurement).

In spite of its highly ill-posed nature, since as early as 1986 [24] to-date, many algorithms have been proposed to solve the SR problem. Historically, the SR mapping is trained on matching point measurements of RGBs and spectra, and the solved map is 'one-to-one' from each RGB to the most probable and/or most reasonable spectral estimation. Examples include regression [18, 11, 26, 22], Bayesian inference [7, 25] and iterative optimization [6, 36]. Recently, as large hyperspectral 'image' databases are made available [3, 34, 26], highly data-driven Deep Neural Network (DNN) and Sparse Coding approaches come to the fore.

On one hand, leading DNNs [4, 5, 30, 20] seek to utilize the image content information by training a *patch-based* mapping, i.e., the RGBs are regressed as part of its proximal pixels. While the DNNs have proved to be effective (recently it is believed that DNNs provide the state-of-the-art performance [4, 5]), they are often parametrized by millions of model parameters, which results in much longer training time compared to the legacy methods, usually also with much higher hardware requirements (e.g., large-memory GPUs). And, realistically, the training data that is available is not large enough to optimally train such a large number of parameters.

On the other hand, the leading Sparse Coding method 'A+' [1] continues to train an effective point-based mapping (where each pixel of an image is regarded as an independent data entry), but instead of targeting a *single* function (i.e., a mapping that is eligible for all RGB inputs), the algorithm seeks to explicitly divide the RGB space where, in each local RGB neighborhood, a different SR mapping is used. And, each local mapping is described by a simple 'linear regression' [21]. Despite its simplicity, research shows that A+ not only provides comparable SR performance to some early DNNs [1], but also ensure much better color fidelity [21, 23] and stability under exposure change [22].

As it stands, A+ still falls short in SR accuracy compared to the top DNN models [23]. However, we found that the neighborhood assignment of A+ is actually not optimized, in that the sparse dictionary [2] used for defining the RGB neighborhoods is optimized for the spectral data but not the RGBs. Based on this finding, we investigate how A+ would have performed if the linear regressions are trained for individual spectral neighborhoods rather than the RGB neighborhoods. Surprisingly, we show that if all spectra are recovered using the *correct* regressions—that is if we pair all of them with the correct spectral-neighborhood labels—our new Sparse Coding setup is able to reach an SR performance substantially beyond the leading DNN.

While we leave the question open as of how to obtain the spectral neighborhood information when the ground-truth spectra are practically unknown (perhaps through additional physical measurements and/or by a learned estimation), our result opens up a new route for seeking a point-based SR map that can be more

accurate than the patch-based and heavily parametrized DNNs.

The rest of the paper will be organized as follows. In Section 2, we introduce the A+ Sparse Coding algorithm and the compared DNN method. In Section 3, we present the proposed Sparse Coding upgrade. Experimental procedures are introduced in Section 4. The results and future work discussion are provided in Section 5. This paper concludes in Section 6.

2. Methods

The relation between RGB and hyperspectral images at a single pixel can be written as [32]:

$$[\underline{s}_R, \underline{s}_G, \underline{s}_B]^{\mathsf{T}} \underline{r} = \underline{x} \,, \tag{1}$$

where the n-component $\underline{r}$ vector refers to the n-channel discrete hyperspectral measurement of the radiance spectrum, $\underline{s}_R, \underline{s}_G$ and $\underline{s}_B$ are the *spectral sensitivity functions* [32] of the R, G, and B sensors, respectively, and $\underline{x}$ is the resulting RGB camera response.

Hereafter, we use the matrix $\mathbf{R}$ to denote the training-set spectra (each column of $\mathbf{R}$ is one spectrum) and the matrix $\mathbf{X}$ to denote the RGB counterparts derived from $\mathbf{R}$ following $[\underline{s}_R, \underline{s}_G, \underline{s}_B]^{\mathsf{T}} \mathbf{R} = \mathbf{X}$.

In SR, we seek to solve the *inverse problem* [31] of the above RGB imaging process. That is:

$$\Psi(\underline{x}) \approx \underline{r} \,, \tag{2}$$

where Ψ denotes the SR algorithm.

2.1. A+ Sparse Coding

2.1.1. Dictionary Training

The training scheme of the A+ algorithm is illustrated in the top panel of Figure 1. The common first step of a Sparse Coding algorithm is to find an effective *dictionary* to represent the data. In A+, the K-SVD algorithm [2] is used, where a given number of K representative spectra—also called the 'atoms'—are found:

$$\mathbf{D}_r = K\text{-SVD}^{\ell}(\mathbf{R}) = [\underline{r}_1, \underline{r}_2, \cdots, \underline{r}_i, \cdots, \underline{r}_K] \,. \tag{3}$$

These atoms are optimized such that all spectra in the training dataset can be derived by their linear combinations with minimal errors. In addition, upon running the K-SVD, we often set a 'sparsity constraint' ℓ, which limits the algorithm to use only ℓ atoms to recover each spectrum ($\ell \ll K$).

Next, following Equation (1), we derive the RGB counterpart of $\mathbf{D}_r$, denoted as the RGB dictionary $\mathbf{D}_x$:

$$\mathbf{D}_x = [\underline{s}_R, \underline{s}_G, \underline{s}_B]^{\mathsf{T}} \mathbf{D}_r = [\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_i, \cdots, \underline{x}_K] \,. \tag{4}$$

Hereafter we call the columns of $\mathbf{D}_r$ and $\mathbf{D}_x$ respectively the spectral and RGB atoms for distinction.

2.1.2. Linear Regression in Each RGB Neighborhood

Each RGB atom in $\mathbf{D}_x$ defines an RGB data neighborhood roughly centers at the atom. In A+, we train an SR regression for each RGB neighborhood using only the neighboring data of its centering atom.

Let us take the i^{th} neighborhood (centering at $\underline{x}_i$) as an example. We find the N nearest RGB neighbors of $\underline{x}_i$ in $\mathbf{X}$ (the training-set RGBs):

$$\mathbf{X}_i = Prox^N(\mathbf{X}, \underline{x}_i) = [\underline{x}_{i,1}, \underline{x}_{i,2}, \cdots, \underline{x}_{i,j}, \cdots, \underline{x}_{i,N}] \,. \tag{5}$$

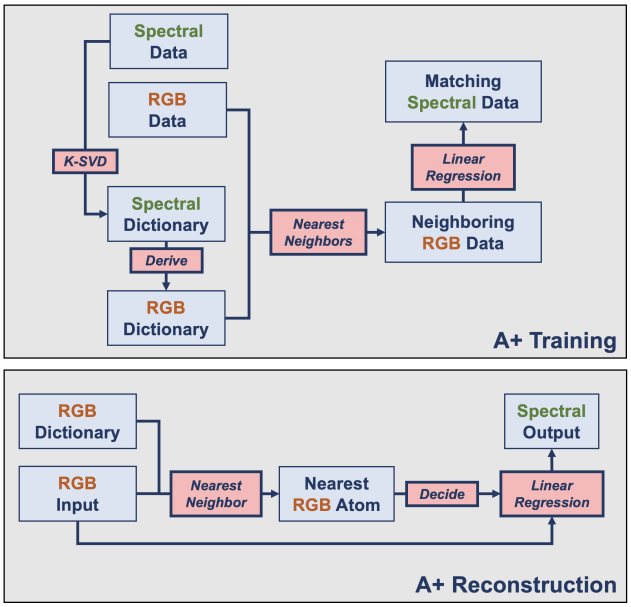

Figure 1. *The training and reconstruction schemes of A+ Sparse Coding.*

Here, the $Prox^N$ function calculates and ranks the Euclidean distances of *normalized* vectors, i.e., both $\underline{x}_i$ and the training-set RGBs (columns of $\mathbf{X}$) are normalized to unit-lengths when calculating the distances. But, in the resulting $\mathbf{X}_i$ matrix, the unnormalized ground-truth vectors are recorded.

For each nearest neighbor in $\mathbf{X}_i$, we also record the corresponding matching ground-truth spectrum in the corresponding column of $\mathbf{R}_i$:

$$\mathbf{R}_i = [\underline{r}_{i,1}, \underline{r}_{i,2}, \cdots, \underline{r}_{i,j}, \cdots, \underline{r}_{i,N}] \,. \tag{6}$$

Given both $\mathbf{X}_i$ and $\mathbf{R}_i$, we solve the i^{th}-neighborhood SR mapping, denoted as Ψ_i, following a *regularized least-squares* minimization [18]:

$$\min_{\mathbf{M}_i} ||\mathbf{R}_i - \mathbf{M}_i \mathbf{X}_i||_2^2 + \gamma_i ||\mathbf{M}_i||_2^2 \implies \Psi_i(\underline{x}) = \mathbf{M}_i \underline{x} \,, \tag{7}$$

where $\mathbf{M}_i$ is an $n \times 3$ regression matrix that constitutes $\Psi_i(\underline{x})$, and γ_i refers to the *regularization parameter* which controls the penalty for large $||\mathbf{M}_i||_2^2$ while minimizing the least-squares term (this setting stabilizes the $\mathbf{M}_i$ solution and prevents it from overfitting the training data [31]).

In practice, $\mathbf{M}_i$ can be solved in closed form [18], while the proper value for γ_i is typically searched for using a cross-validation methodology [13]—a separate *validation*-set data is tested with a wide range of different γ_i values (we search between 10^{-20} and 10^{20}), and the γ_i that minimizes the recovery error is selected (here we use the MRAE error metric [4, 5], which will be introduced in the experimental section).

2.1.3. Reconstruction

The reconstruction procedure of A+ is illustrated in the bottom panel of Figure 1. Straightforwardly, because the local linear regressions are optimized for the data around each RGB atom, for each testing RGB in reconstruction we are to apply the regression that is attached to its closest RGB atom in $\mathbf{D}_x$.

Mathematically, we write:

$$Prox^1(\mathbf{D}_x, \underline{x}) = \underline{x}_q \implies \Psi_q(\underline{x}) \approx \underline{r}\,, \tag{8}$$

which means if among $\mathbf{D}_x$ the closest atom to the testing RGB $\underline{x}$ is $\underline{x}_q$, we say $\underline{x}$ locates in the q^{th} neighborhood and apply the q^{th} local regression $\Psi_q(\underline{x})$ (trained using $\mathbf{X}_q$ and $\mathbf{R}_q$) to recover spectrum.

Note that the $Prox^1$ function—similarly to the $Prox^N$ function—also calculates the distances of normalized vectors.

2.2. Compared Method: HSCNN-R Network

There are many DNN-based SR algorithms proposed in the recent literature, most of which are based on the Convolutional Neural Network (CNN) or Generative Adversarial Network (GAN) architectures. In this paper, we benchmark the HSCNN-R method [30], which is among the top models trained on our concerned hyperspectral image dataset 'ICVL' [3], according to the NTIRE 2018 Spectral Reconstruction Challenge [4].

HSCNN-R is a CNN-based method, which is trained to map 50×50 RGB image patches to their corresponding hyperspectral image patches. The network adopts a deep residual learning framework [17]. Each residual block consists of 2 convolutional layers and 1 ReLU layer, and all convolutional kernels are set to 3×3. We set the depth of the network to be 34 and the filter numbers in each layer to be 64, based on one of the suggestions in the original work [30].

3. Proposed Sparse Coding Upgrade

3.1. Linear Regression in Each Spectral Neighborhood

In our proposed framework (Figure 2), we do not operate the neighborhood regressions in the RGB space. Instead, we seek to train a bespoke regression in each spectral neighborhood centering at the spectral atom.

Returning to Equation (5), and let us continue to use the i^{th} neighborhood as an example, here we group the N nearest spectral neighbors of $\underline{r}_i$ in $\mathbf{R}$ (training-set spectra) instead:

$$\mathbf{R}'_i = Prox^N(\mathbf{R}, \underline{r}_i)\,. \tag{9}$$

Correspondingly (and similarly to Equation (6)), we record the matching RGBs of these spectral neighbors in the corresponding columns of $\mathbf{X}'_i$.

Clearly, $\mathbf{R}'_i$ and $\mathbf{X}'_i$ represent a different set of data compared to $\mathbf{R}_i$ and $\mathbf{X}_i$ (Equation (5) and (6)). Indeed, for example, it is well known in color science that two very different spectra can correspond to the same RGB response under a given viewing condition—i.e., these two spectra are *metameric* [12], which means it is possible for two spectra to be neighbors in the RGB space but not simultaneously in the spectral space.

Then, following Equation (7) while replacing $\mathbf{R}_i$ and $\mathbf{X}_i$ by $\mathbf{R}'_i$ and $\mathbf{X}'_i$, respectively, the i^{th} local regression $\Psi_i(\underline{x})$ is calculated under our new proposed setting.

3.2. The 'Oracle' Reconstruction Scenario

Since the local regressions are optimized for each spectral neighborhood, ideally, we should reconstruct each spectrum $\underline{r}$ based on where it locates in the spectral space:

$$Prox^1(\mathbf{D}_r, \underline{r}) = \underline{r}_q \implies \Psi_q(\underline{x}) \approx \underline{r}\,. \tag{10}$$

However, in the practical situation where the ground-truth spectrum $\underline{r}$ is unknown, we shall not be able to calculate $Prox^1(\mathbf{D}_r, \underline{r})$.

Sometimes the approach we are taking is called an 'oracle method'. To solve Equation (10) we need to know the radiance spectrum that we are trying to recover. Or, we need an all seeing oracle to make the assignment for us. Oracle methods are often used to make perfect decisions that are—at the time of writing a paper—difficult to make algorithmically. In effect, in this paper, we have defined the 'Oracle A+' method. And, this method usefully bounds how well A+ can work, in general.

We remark that the spectral neighborhood information might be obtained via physical measurements (from the sound of it, measuring the spectral neighborhood should only need to be a coarser measurement than hyperspectral imaging); or, we may seek to estimate the spectral neighborhood labels from the RGBs; or a combination of both.

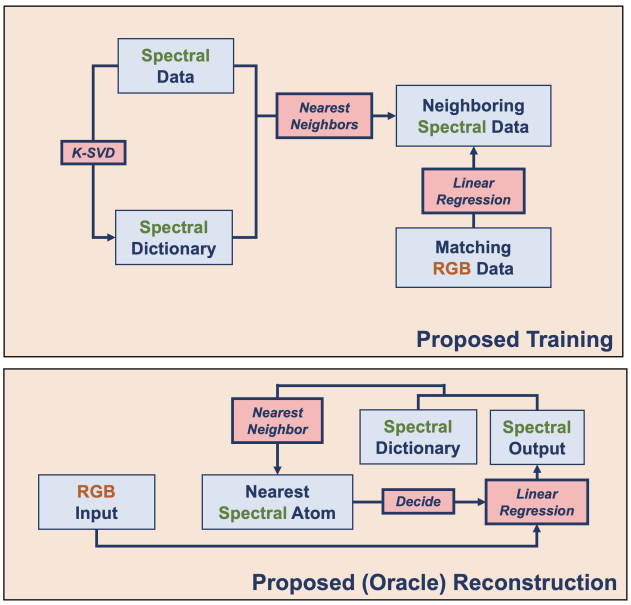

Figure 2. *The training and (oracle) reconstruction schemes of our new Sparse Coding setup.*

4. Experiment

4.1. Database

In this paper, we benchmark on the ICVL hyperspectral image database [3]. In ICVL, there are 200 scenes of size 1300×1392 and 31 spectral channels, referring to the spectral measurements between 400 to 700 nanometers (the 'visible range') with 10-nanometer intervals.

The ground-truth RGB images were generated from the hyperspectral images using Equation (1) with CIE 1964 color matching functions [10] as the camera's spectral sensitivity functions. We note that this setting is standard for the bi-annually NTIRE Spectral Reconstruction Challenges [4, 5].

4.2. Performance Metric

We use the following metrics to evaluate the methods:

- Mean-Relative-Absolute Error

$$\text{MRAE} = \frac{1}{n}\left|\left|\frac{\underline{r}-\Psi(\underline{x})}{\underline{r}}\right|\right|_1, \tag{11}$$

where $\underline{r}$ and $\Psi(\underline{x})$ are respectively the ground-truth and SR-recovered spectra, n is the number of spectral channels ($n = 31$ for ICVL database), and the division is element-wise to the vectors. In essence, MRAE calculates the averaged percentage L1 deviation with respect to the ground-truth spectral intensities over all spectral channels. This metric is the standard protocol used to benchmark the most recent models, particularly those based on DNNs [4, 5].

- Root-Mean-Square Error

$$\text{RMSE} = \sqrt{\frac{1}{n}||\underline{r}-\Psi(\underline{x})||_2^2}. \tag{12}$$

Here, unlike MRAE, the RMSE is a 'scale-dependent' metric, which means the brightness level and/or the bit-depth of the spectral data will influence the scale of RMSE. Hence, it is essential to note that the ICVL database uses 12-bit encoding (with the maximal value of 4095) [3] when interpreting the RMSE results.

4.3. Training, Validation, and Testing

We follow a 4-fold cross validation process [28]: the total of 200 scenes are separated into four sets of 50 scenes, and while each set is in turn reserved for evaluation, the rests are used for training (two sets; 100 scenes) and validation (one set; 50 scenes).

The training process of Sparse Coding includes training the spectral dictionary and the local linear regressions. Following the original work of A+ [1], we randomly sample 3,000 pixels per training image for dictionary training and 30,000 pixels per image for training the local linear regressions. Also, for fair comparison, we use the same settings for number of atoms ($K = 1024$), sparsity ($\ell = 8$) and number of nearest neighbors ($N = 8192$) in our new method, as what was originally suggested for A+ [1].

Then, the validation process of Sparse Coding refers to tuning the γ_i value in Equation (7) (for each and every local regression). Since each local regression is trained specifically for that neighborhood, we also need to determine which data points among the validation set (also downsampled to 30,000 pixels per image) are in the concerned neighborhood, and subsequently use only these neighborhood data to tune the γ_i value. We do this by using the $Prox^N$ function in Equation (5) for A+ (or Equation (9) for the proposed method), while setting $N = 8192/2 = 4096$—this is because the validation set is half the size of the training set.

Finally, in testing, we test on all pixels and images of the testing set, and the averaged result of the 4 cross-validation trials is presented.

Note that for the compared method, HSCNN-R, we do not downsample the pixels in training and validation, and the validation step refers to determining the stopping epoch of its iterative training process. Interested readers are pointed to [30] for more details.

Table 1. The mean (± standard deviation) hyperspectral image reconstruction accuracy measured in MRAE and RMSE.

Method	MRAE	RMSE
A+	0.0381 (±0.0199)	23.26 (±11.85)
HSCNN-R	0.0176 (±0.0111)	16.38 (±11.48)
Proposed	0.0149 (±0.0095)	10.27 (± 6.01)

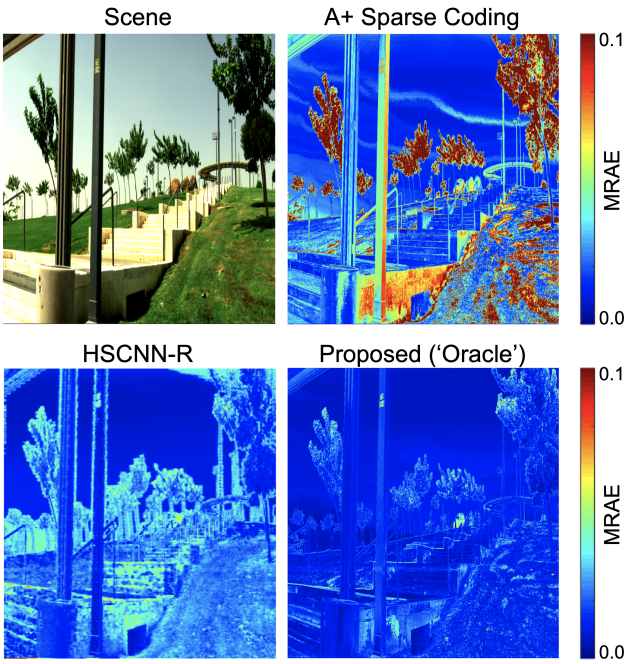

Figure 3. *Reconstruction error maps of an example scene in MRAE.*

5. Results and Future Work

In Table 1, we show the mean and standard deviation of the *per-image mean errors* in MRAE and RMSE. Also, in Figure 3, we present the MRAE error maps of one example scene recovered by the tested methods.

Evidently, our proposed modification to A+ considerably improves the upper-bound performance of Sparse-Coding-based SR. This advance indicates that using our new way of defining neighborhoods, on average, the local regressions are trained to recover spectra much more accurately. It would be interesting to study the cause of this result in the future work.

Next, given that the upper-bound performance of the point-based and shallow-learned Sparse Coding is now better than the leading DNN method, we may further ask how we can *approach* this upper-bound performance (i.e., how we can correctly predict the spectral neighborhood of an RGB). The answer to this question may potentially lead us to an entirely point-based SR solution that performs on par with or even better than the leading DNNs.

On the other hand, we may also explore the possibility of a 'Sparse-Coding-guided' hyperspectral imagery, where the spectral neighborhood information is physically measured apart from the RGB imaging—essentially, our new Sparse Coding can serve as a guide to what to measure in order to make SR more accurate.

6. Conclusion

Spectral Reconstruction (SR) algorithms are trained to recover hyperspectral information from RGB camera responses. Recently, Deep Neural Networks (DNN) are considered the state-of-the-art approaches to SR, however, they are often excessively parametrized, which leads to long training time and high hardware requirements. Alternatively, the much simpler leading Sparse Coding 'A+' seeks to define local neighborhoods in the RGB space, where in each RGB neighborhood only a simple linear regression is trained and used. Yet its performance do not compare to the leading DNN methods.

In this paper, we show that the upper-bound performance of the leading Sparse Coding method can be further advanced if we localize the regressions in spectral neighborhoods rather than the RGB neighborhoods (the upper-bound is met when we could locate the neighborhoods of all target spectra). Moreover, we show a cross-validated experimental result that this upper-bound performance is clearly better than the leading DNN method. For future work, we shall seek the physical and/or learning means of acquiring the spectral neighborhood information of the incoming spectral signal, as such to approach the upper-bound performance of the proposed method.

References

[1] J. Aeschbacher, J. Wu, and R. Timofte. In defense of shallow learned spectral reconstruction from RGB images. In *Proceedings of the International Conference on Computer Vision*, pages 471–479. IEEE, 2017.

[2] M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322, 2006.

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

[4] B. Arad, O. Ben-Shahar, R. Timofte, et al. NTIRE 2018 challenge on spectral reconstruction from RGB images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops*, pages 929–938. IEEE, 2018.

[5] B. Arad, R. Timofte, O. Ben-Shahar, Y. Lin, G.D. Finlayson, et al. NTIRE 2020 challenge on spectral reconstruction from an RGB image. In *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, June 2020.

[6] S. Bianco. Reflectance spectra recovery from tristimulus values by adaptive estimation with metameric shape correction. *Journal of the Optical Society of America A*, 27(8):1868–1877, 2010.

[7] D.H. Brainard and W.T. Freeman. Bayesian color constancy. *Journal of the Optical Society of America A*, 14(7):1393–1411, 1997.

[8] C. Chen, W. Li, H. Su, and K. Liu. Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine. *Remote Sensing*, 6(6):5795–5814, 2014.

[9] V. Cheung, S. Westland, C. Li, J. Hardeberg, and D. Connah. 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.

[10] Commission Internationale de L'eclairage. *CIE Proceedings (1964) Vienna Session, Committee Report E-1.4. 1*, 1964.

[11] D.R. Connah and J.Y. Hardeberg. Spectral recovery using polynomial models. In *Proceedings of the Color Imaging X: Processing, Hardcopy, and Applications*, volume 5667, pages 65–75. International Society for Optics and Photonics, 2005.

[12] G.D. Finlayson and P. Morovic. Metamer sets. *Journal of the Optical Society of America A*, 22(5):810–819, 2005.

[13] N.P. Galatsanos and A.K. Katsaggelos. Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. *IEEE Transactions on Image Processing*, 1(3):322–336, 1992.

[14] N. Gat. Imaging spectroscopy using tunable filters: a review. In *Proceedings of the Wavelet Applications VII*, volume 4056, pages 50–64. International Society for Optics and Photonics, 2000.

[15] P. Ghamisi, M. Dalla Mura, and J.A. Benediktsson. A survey on spectral–spatial classification techniques based on attribute profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5):2335–2353, 2014.

[16] R.O. Green, M.L. Eastwood, C.M. Sarture, T.G. Chrien, M. Aronsson, B.J. Chippendale, J.A. Faust, B.E. Pavri, C.J. Chovit, M. Solis, et al. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sensing of Environment*, 65(3):227–248, 1998.

[17] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 770–778. IEEE, 2016.

[18] V. Heikkinen, R. Lenz, T. Jetsu, J. Parkkinen, M. Hauta-Kasari, and T. Jääskeläinen. 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.

[19] A. Lam and I. Sato. Spectral modeling and relighting of reflective-fluorescent scenes. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 1452–1459. IEEE, 2013.

[20] J. Li, C. Wu, R. Song, Y. Li, and F. Liu. Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from RGB images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops*, pages 462–463. IEEE, 2020.

[21] Y. Lin. Colour fidelity in spectral reconstruction from rgb images. In *Proceedings of the London Imaging Meeting, London, UK*, volume 2020, pages 144–148. Society for Imaging Science and Technology, 2020.

[22] Y. Lin and G.D. Finlayson. Exposure invariance in spectral reconstruction from RGB images. In *Proceedings of the Color and Imaging Conference*, volume 2019, pages 284–289. Society for Imaging Science and Technology, 2019.

[23] Y. Lin and G.D. Finlayson. Physically plausible spectral reconstruction. *Sensors*, 20(21):6399, 2020.

[24] 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.

[25] 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.

[26] R.M.H. Nguyen, D.K. Prasad, and M.S. Brown. Training-based spectral reconstruction from a single RGB image. In *Proceedings of the European Conference on Computer Vision*, pages 186–201. Springer, 2014.

[27] J. Qin, K. Chao, M.S. Kim, R. Lu, and T.F. Burks. Hyperspectral and multispectral imaging for evaluating food safety and quality. *Journal of Food Engineering*, 118(2):157–171, 2013.

[28] P. Refaeilzadeh, L. Tang, and H. Liu. *Encyclopedia of Database Systems – Cross Validation*, pages 532–538. Springer US, Boston,

MA, 2009.

[29] H. Shen and J.H. Xin. Spectral characterization of a color scanner by adaptive estimation. *Journal of the Optical Society of America A*, 21(7):1125–1130, 2004.

[30] Z. Shi, C. Chen, Z. Xiong, D. Liu, and F. Wu. Hscnn+: Advanced cnn-based hyperspectral recovery from RGB images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops*, pages 939–947. IEEE, 2018.

[31] A.N. Tikhonov, A.V. Goncharsky, V.V. Stepanov, and A.G. Yagola. *Numerical Methods for the Solution of Ill-posed Problems*, volume 328. Springer Science & Business Media, 2013.

[32] B.A. Wandell. The synthesis and analysis of color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1):2–13, 1987.

[33] P. Xu, H. Xu, C. Diao, and Z. Ye. Self-training-based spectral image reconstruction for art paintings with multispectral imaging. *Applied Optics*, 56(30):8461–8470, 2017.

[34] F. Yasuma, T. Mitsunaga, D. Iso, and S.K. Nayar. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE Transactions on Image Processing*, 19(9):2241–2253, 2010.

[35] Y. Zhang, X. Mou, G. Wang, and H. Yu. Tensor-based dictionary learning for spectral ct reconstruction. *IEEE Transactions on Medical Imaging*, 36(1):142–154, 2016.

[36] S. Zuffi, S. Santini, and R. Schettini. From color sensor space to feasible reflectance spectra. *IEEE Transactions on Signal Processing*, 56(2):518–531, 2008.

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 a B.Sc. degree in Physics in 2016, from National Taiwan University, Taiwan. His research interest is physics and machine-learning-based spectral reconstruction.

Graham Finlayson is a Professor of Computer Science at the University of East Anglia (UEA) where he leads the Colour & Imaging Lab. Professor Finlayson is interested in 'computing how we see' and his research spans computer science (algorithms), engineering (embedded systems) and psychophysics (visual perception). Significantly, some of Graham's research is implemented and used in commercial products including photo processing software, dedicated image processing hardware and in embedded camera software.