https://doi.org/10.24867/GRID-2022-p96 Original scientific paper

IMPROVED RECONSTRUCTION OF THE REFLECTANCE SPECTRA FROM RGB READINGS USING TWO INSTEAD OF ONE DIGITAL CAMERA

Mihael Lazar, Aleš Hladnik
University of Ljubljana, Faculty of Natural Sciences and Engineering, Department of Textiles, Graphic Arts and Design, Ljubljana, Slovenia

Abstract: *The colour of an observed object can be described in many different manners, and the description by its reflectance provides the unambiguous colour representation. The reflectance description can be acquired by expensive multispectral cameras or, e.g., with time-sequential multispectral illumination. In our experiment, we propose that under the condition of constant and uniform illumination, the reflectance can be deduced from the object's RGB camera readouts, captured alongside the set of colour patches with known spectral characteristics. Translation from a colour description in RGB space into reflectance spectra, independent of illuminant and camera sensor characteristics, was performed with the help of an artificial neural network (ANN). In our study, the hypothesis was proposed that the ANN's performance of reflectance reconstruction can be enhanced by employing richer learning datasets using RGB input sets of two cameras instead of just one. Additional second camera information would be adequate only if the equivalent channels of cameras used are linearly independent. A quantitative measure of nonlinearity (QMoN), which is the metric primarily developed for use in chemistry, was employed to estimate the degree of nonlinearity. Additional attention was paid to ANN training, structure and learning set sizes. Two ANN training algorithms have been utilised, a faster GPU executed standard backpropagation and an order of magnitude slower CPU based, but with significantly better convergence Levenberg-Marquardt training algorithm. The number of neurons in the hidden ANN layer varied from the size of the input layer to a number greater than the number in the output layer. The complete set of colour samples was divided into five learning sets of different sizes, with the smaller sets being subsets of the larger ones. To assess performances of the resulting ANNs, mean squared error, the goodness of fit and colour differences calculated from original and reconstructed reflectances assuming several standard illuminations have been compared. A noticeable reflectance performance improvement has been found by using two cameras, even though the cameras' equivalent channels exerted only small degrees of nonlinearity.*

Keywords: artificial neural network, reflectance spectra reconstruction, enriched learning set, two cameras, equivalent channel's nonlinearity

1. INTRODUCTION

In many areas, the description of colours using the camera's colour space, i.e. RGB is sufficient, but for a more accurate description, we use colour spaces such as XYZ or CIELAB. The former depends on both the characteristics of the capture device and the illumination, while the latter is independent of the device but still includes illumination. If we, for some reason, e.g. archiving a medieval facsimile or an artwork, want to achieve a colour description independent of the lighting at the time of capture, colours should be described through the reflectance spectra, for which several different image capture strategies have been developed. Multispectral cameras, which describe the image with more than three frequency bands, and hyperspectral cameras, which describe the image with a large number of narrow frequency bands, are expensive, and by increasing the spatial and frequency resolution, the temporal resolution can be compromised (Cucci et al., 2011). Due to the affordability of consumer cameras, their high image resolution and capture speed, the possibility of converting an RGB image, which varies from camera to camera, into a reflectance spectrum is becoming more and more intriguing. Such methods range from purely mathematical models that incorporate knowledge of image capture conditions and limitations to learnable models that exploit the capabilities of artificial neural networks (ANNs). Some of the many methods of spectrum reconstruction from camera readings have already been mentioned in our previous work (Lazar, Javoršek & Hladnik, 2020), and some are described below.

In (Imai & Berns, 1999), an alternative approach to capturing multispectral images via a combination of RGB digital camera and either absorption filters or multiple illuminations is described. Imaging spectroscopy through faster electronically tuneable filters in front of a monochrome camera to build an image cube of spectral layers is reviewed in (Gat, 2000). Tuneable filters in front of the camera enable narrow-band filtering, where the build of the image cube requires a time sequence of consecutive photos. In (Chi, Yoo & Ben-Ezra 2010), the multispectral image is obtained with the carefully selected set of filters placed in front of the illumination source and the unknown moderate indoor ambient light is cancelled out. Besides hyper/multi-spectral imaging using special equipment, other techniques are being developed. For application in spectrally-based rendering systems, the algorithm for converting RGB labelled object and texture colours to spectrally presented reflectances has been described in (Smits, 1999). In a three-step algorithm proposed by (Jia et al., 2017), the spectrum has been accurately reconstructed from RGB values of photos taken in daylight. The first step is a nonlinear dimensionality reduction of the high dimensional spectra of natural scene images to a 3D embedding of spectra. Then, RGB values of observed spectra are calculated, considering the known lighting conditions and camera spectral response and training the ANN to connect the RGB values and 3D embedding of spectra. In the third step, conversion from 3D embedding to high dimensional spectra is done through a low to high dimensionality dictionary.

Multispectral and hyperspectral images are helpful in many fields, e.g. in geology (Ninomiya & Fu, 2019), agriculture (Hassan-Esfahani et al., 2014), food industry (Benouis et al., 2021), healthcare (Wang et al., 2022) and cultural heritage (Colantonio et al., 2018). Research in the field of multi- and hyperspectral image capture is useful and current, as well as, due to the affordability of high-resolution commercial cameras, also research that uses various techniques of converting RGB to multidimensional spectral image space.

In most studies, at least some knowledge about the image capturing conditions, e.g. lighting conditions and/or sensory response, is required. The precise measurement of the illumination spectrum is not always feasible, and the manufacturer does not necessarily specify the camera spectral response. Instead of using absorption filters or alternating bandpass illumination, which requires additional technical expertise, we propose that an array of reference colour patches with known reflectance spectra be captured alongside the object of interest with one or more cameras. Thus, the relationship between RGB values and the reflectance spectra can be established through a subsequent software approach. In our previous work, the possibility and effectiveness of an ANN-based spectral reflectance reconstruction from a single camera RGB data has been studied. An open question remained whether the efficiency of reflection reconstruction could be enhanced if the RGB data from two cameras were available.

There are many possibilities for using the data of the reflection spectrum of the observed object. Depending on the light reflection, the surface type is essential, where the reflection can be complete or diffuse. If the surface is uniformly coloured, the reflectance can be measured using a spectrophotometer. In the case of large colour variability, such as in works of art, its use is impractical or impossible. For our planned future use for archiving works of art in watercolour, pastel, chalk, crayon, or printing material on matte paper, we used non-glossy - matte samples in our experiment.

2. REFLECTANCE RECONSTRUCTION USING ANN

It has been shown (Hornik, 1991) that multilayer feedforward networks are, under very general conditions on the hidden unit activation function (continuous, bounded and nonconstant), universal approximators provided that sufficiently many hidden units are available. In recent years, the availability of graphics cards with high computing power and data throughput has fuelled the evolution of neural network architecture. Convolutional and deep NNs are used to analyse large-scale input data through compacting the input data, like images and natural language and absorbing a huge amount of complex knowledge (Ciresan et al., 2011, Zhang & Wallace, 2015, Szegedy, Toshev & Erhan, 2013, Rolnick & Tegmark, 2017). Our research focuses on a relatively small learning set (hereafter LS) with a small number of inputs; thus, reducing input data dimensionality would not contribute to network efficiency. We built a three-layer ANN, where an input layer accepts R, G and B values of one or two cameras, a hidden layer harbours a variable number of neurons, and the output layer's exit values consist of a reconstructed reflectance vector containing spectral power distribution of visible light, with a 10 nm wavelength resolution from 380 to 730 nm.

The generic architecture of utilised ANNs is shown in Figure 1, where $x_{k,j}$ are k-th inputs, which can be RGB readings of one or two (depicted) cameras (hereinafter 1RGB or 2RGB); $b_j^{(L)}$ and $w_{j,i}^{(L)}$ stand for biases and connection weights and $y_{k,j}$ are output layer neurons' output values, representing 36 spectral components of k-th reconstructed reflectance; j is neuron index in (L)-th layer and i in $(L-1)$-th layer.

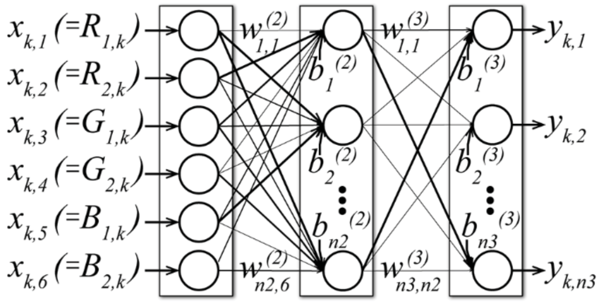

Figure 1: Architecture of an ANN with a single hidden layer and a variable number of hidden neurons, for the reconstruction of reflectances from RGB

Having a LS of (1RGB/2RGB, reflectance spectrum) pairs from the reference set of patches, the training of the ANN is performed by a learning algorithm, adjusting connection weights and biases so that the network cost function -mean squared error (Eq. 1), is minimised.

$$C = \frac{1}{N_p}\sum_{k=1}^{N_p} \left[\frac{1}{N_\lambda}\sum_{j=1}^{N_\lambda} \left(r_{k,j} - y_{k,j}\right)^2\right] . \tag{1}$$

N_p is the number of patches, N_λ number of reflectance wavelengths, k is patch index and j wavelength index, $r_{k,j}$ is a j-th spectral component of k-th patch *measured* reflectance and $y_{k,j}$ j-th spectral component of k-th patch *reconstructed* reflectance.

The software part of our experiment was programmed in Matlab, where the two ANN learning algorithms are set for the purpose of function approximation, namely "standard gradient descent based backpropagation" (hereafter BP) ANN training algorithm is high-speed due to the ability to run on GPU and another, Levenberg-Marquardt (hereafter LM) algorithm, which only supports computing on CPU, but is specially adapted to minimise the sum-of-squares error functions (Aldrich, 2002). As in our previous research, it shows again that the LM learning algorithm, with an increased number of hidden layer neurons, requires at least an order of magnitude more time but results in order of magnitude better learning convergence by means of epochs repetition and better ANN performance compared to BP.

Functions embedded in the modelled ANN translate RGB readings into components of the reflectance spectrum. Each of the 36 embedded functions in our case translates the three input readings (a point in a 3D space) into one spectral component. If we take two RGB readings that are not linearly dependent, triangulating these two points into one spectral value should give more accurate results and partially eliminate metamerism. Metamerism caused by the light source cannot be decreased or eliminated in this way. Suppose a part of the illuminant spectrum is missing. In that case, it prevents sensing the reflection of a missing part of the spectrum from the object. Whereas, if a light source with a smooth spectrum is used (e.g., sunlight or incandescent bulb) when the photo of an object and the reference array of patches is captured, their reflectance should convey its complete spectrum, which in the camera is transformed into RGB readouts. If cameras used in the experiment have a nonlinear relationship between equivalent colour channels, the observer (or camera) metamerism could be reduced. The integration defining the R, G and B values includes the product of illuminant spectrum $I(\lambda)$, object/colour patch spectral reflectance $S(\lambda)$ and camera spectral sensitivity function $\tau(\lambda)$, which is different for each channel of each camera (Eq. 2):

$$y_{cam.,ch.,i} = \int_{\lambda_{min}}^{\lambda_{max}} \tau_{cam.,ch.}(\lambda) S_i(\lambda) I(\lambda) d\lambda\,. \tag{2}$$

Here "*cam.*" indicates camera (in our case, 1 or 2), "*ch.*" represents colour channel (R, G or B) and "*i*" denotes colour index (e.g., colour 1, 2 … num. of colour patches).

If one camera gives the same RGB outputs for two different samples, a metameric pair of samples is found, meaning that the integrals of equivalent channels of two samples give the same result (R_1=R_2, G_1=G_2 and B_1=B_2). If the second camera's equivalent channels are not linearly dependent on the first camera, the integrals for outputs of the *second camera* will *likely* give different results for the previously metameric pair of samples ($R_1 \neq R_2$ and/or $G_1 \neq G_2$ and/or $B_1 \neq B_2$). Because RGB values are 8-bit integers, too

slight differences could still result in the same values. The greater the nonlinearity is between the equivalent channels, the more significant the possibility of *different values for the previously metameric pair*. We can conclude that the second camera can provide additional information. It is challenging to prove analytically; however, by expanding the ANN's LS with an algorithm that allows learning through a series of examples, an additional camera could allow for better ANN performance, which we hope to demonstrate through our experiment.
As stated in the hypotheses below, we assume that the equivalent colour channels of the cameras used are linearly independent. If so, we believe the ANN performance will improve if the LS input data is doubled.
Our hypotheses are as follows:

1. Equivalent channels (R_{C1}-R_{C2}, G_{C1}-G_{C2}, B_{C1}-B_{C2}) of different cameras are most likely linearly independent.
2. An ANN for reflectance reconstruction from RGB readings, trained with input data of two different cameras, will perform better than an ANN trained with single-camera RGB data.

3. MATERIALS AND METHODS

Our study aimed to explore the possibility of improving the recovery of reflectance spectra from trichromatic camera values by supervised learning of an ANN with a single hidden layer, modelled with Matlab Neural Network Toolbox. The ANN LS inputs are vectors of 3 or 6-dimensional RGB readings from one or two cameras, and outputs are higher dimensional vectors of reflectance spectra – readings of the spectrophotometer.
The source of colourimetric data for our experiment was The Munsell Book of Color Matte Collection, with 44 sheets providing 1301 colour patches, varying chroma, hue and value. Forty sheets are divided into 2,5 steps Munsell hue circle (2,5, 5, 7,5, 10 for Red, YR, Yellow, GY, Green, BG, Blue, PB, Purple and RP). Four remaining sheets contain neutral colours - neutral and subtly hued greys. Reflectance spectra of patches were measured by spectrophotometer X-Rite i1Pro 2 at five points (on both diagonals, a quarter of the distance from each corner, and at the patch centre). Reflectances - vectors with 107 components for wavelengths from 376.66 to 730 nm with a step size of 3.33 nm were calculated as an average value of the five measurements with a maximum standard deviation of less than 0.4%.
Because of the absorption characteristics of human-made and natural colourants, the sampling rate can be significantly decreased (Imai & Berns, 1999). Surface spectral reflectances of many organic and inorganic substances are characteristically smooth, low-pass functions of wavelength (Maloney, 1986). By observing reflectance spectra of all the available colour patches, which include many natural colours (soil, skin, foliage etc.), the smoothness has been confirmed without exception; therefore, the decrease in sampling rate is acceptable. Subsampling was made from 3.33 to 10 nm step, resulting in 36 reflectance spectral components in the range from 380 nm to 730 nm.
The sheets with colour patches were photographed in a photo studio under constant lighting conditions. We used three cameras (Nikon D600, D700 and Panasonic GH-4) with fixed manual settings (Figure 2). A spectrophotometer measured the light source spectral power distribution at the sheet location, and a correlated colour temperature (CCT) of 3019° K was read out. Images were captured in RAW format and required conversion to 8-bit RGB. All images were processed equally - normalised using Adobe Lightroom software to the measured CCT, the tint was balanced to 0, and chromatic aberration, even though unnoticeable, was corrected with corresponding lens profiles. This process resulted in conversion to digital photos in AdobeRGB (1998) colour space. The RGB value of each patch was calculated as the median (rather than the mean) of the inner 50% of the squared patch area to avoid the influence of possible minor colour deviations.

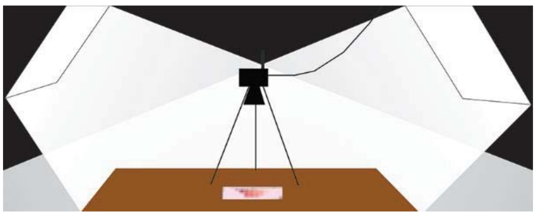

Figure 2: Photo studio setting

The complete set of RGB-reflectance pairs included data of all 1301 Munsell Matte colour patches, with three camera sets of R, G and B values and 36 values for each reflectance vector. Before training the ANN models, the complete set was split into a training set and the remaining independent samples, for which additional measures of reflectance reconstruction performance were calculated. At the beginning of each ANN model training iteration, the LS was randomly split into training, validation and testing sets in a 70:15:15 ratio.

Due to the nature of ANN learning, where the algorithm tends to find a local instead of a global minimum of the cost function, for each model with selected cameras, a training set size and number of neurons in the hidden layer (3 to 48, in steps of 1), the search for optimal ANN parameters was repeated 41 times. In preliminary experiments with 21 ANN training repetitions with selected LS sizes, it was found that increasing the number of neurons in the hidden layer (hereafter HLN) above 48 does not improve the ANN performance noticeably or even worsens, while the calculations become very time-consuming.

In this experiment, we wanted to increase the possibility of finding the best ANN models, so the number of repetitions was almost doubled. The odd number of 41 repetitions was chosen due to the calculations of some additional statistics not presented in the article. Calculations with the BP and LM learning algorithms were performed on five different sizes of learning sets and corresponding sets of independent samples (Table 1).

Table 1: Size of learning sets

descriptive size of learning set	% of learning set vs complete set of 1301 colour patches	num. of 1/2RGB - reflectance learning set pairs	num. of remaining independent samples
very large	90	1171	130
large	50	650	651
medium	30	390	911
medium	20	260	1041
smaller	15	195	1106

4. NONLINEAR RELATIONSHIPS BETWEEN CAMERAS

In our experiment, we observed RGB readings of the patches acquired with three disparate cameras. Polynomial regression has been calculated relating values of R, G and B channels of the same patches, pairwise between these three cameras. The relationship of equivalent channels for one of these pairs (Panasonic GH4 and Nikon D600), their polynomial regression functions of the second-order and the residuals as the difference between the polynomial regression of the 2nd, 3rd and 4th order and linear regression are shown in Figure 3.

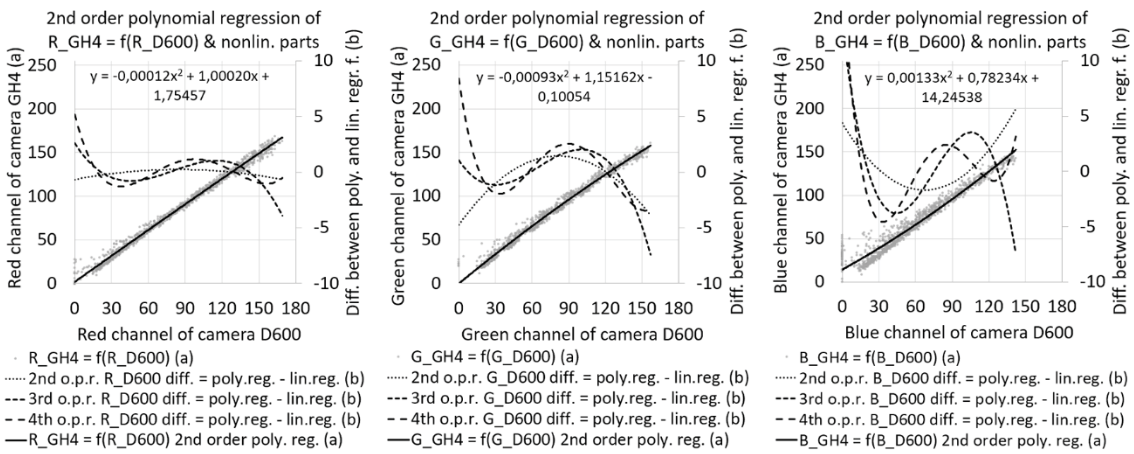

Figure 3: The relationship of equivalent channels from camera GH4 and D600, the polynomial regression functions of the second-order and the residual parts as the difference between the 2nd, 3rd and 4th order polynomial regressions and the linear regression

The efficiency of polynomial regression has been evaluated by observing the difference between the regression functions and actual data, averaged for each Munsell hue section. The difference for the 2nd order polynomial regression functions varies from 4,2% to -2,7%. When including neutrals and lightly hued greys, the average difference is almost 0%, with the standard deviation span between 1,49% and 2,8% through all channels and camera combinations. With the third order, the standard deviation decreases a little (1,46% to 2,68%), with a still smaller decrease with the fourth order polynomial regression (1,43% to 2,66%). Therefore, the 2nd, 3rd and optionally fourth order regression is suitable for observing the sensor functions (R, G and B) of the first in dependence on the second camera.
To assess the nonlinearity between equivalent channels of cameras, we utilised the quantitative measure of nonlinearity (QMoN) suggested in (Emancipator & Kroll, 1993). The "dimensional nonlinearity" of a method is defined as the root mean square of the deviation of the response curve from an ideal straight line, chosen to minimise the nonlinearity. The "relative nonlinearity" is the "dimensional nonlinearity" divided by the distance between the largest and smallest tested values. The definition of quantitative nonlinearity measure suggests different nonlinear regression methods, among others also polynomial, which was used in our experiment. Due to a large number of test values (1301) and the associated small number of different values along the x-axis (167), the F-distribution with the proposed degrees of freedom gives very small values of the 95th percentile of the F-distribution (1,2). Hence the optional algorithm for searching the advisable order of polynomial regression, proposed by Emancipator and Kroll, is inadequate, and the nonlinearity has been therefore calculated by relative QMoN for the polynomial regression for up to the fourth order, as shown in Table 2. The linear regression (1st order) with expectedly small values is introduced as a nonlinearity measure control group.

Table 2: Quantitative measures of nonlinearity between the equivalent channels of the cameras used, varying the order of polynomial regression

	R channel nonlinearity [%]			**G channel nonlinearity [%]**			**B channel nonlinearity [%]**			**All channel average nonlinearity [%]**		
order of polynomial regression	R_D700 = f(R_D600)	R_GH4 = f(R_D600)	R_GH4 = f(R_D700)	G_D700 = f(G_D600)	G_GH4 = f(G_D600)	G_GH4 = f(G_D700)	B_D700 = f(B_D600)	B_GH4 = f(B_D600)	B_GH4 = f(B_D700)	RGB_D700 = f(RGB_D600)	RGB_GH4 = f(RGB_D600)	RGB_GH4 = f(RGB_D700)
4th	0,929	0,707	0,498	1,160	1,448	0,599	1,520	2,953	3,258	1,203	1,703	1,452
3rd	0,962	0,646	0,399	1,124	1,268	0,666	1,656	3,116	2,659	1,247	1,676	1,241
2nd	0,729	0,160	0,251	0,738	1,083	0,720	0,603	1,446	1,746	0,690	0,897	0,906
1st = lin. r.	0,002	0,002	0,001	0,002	0,002	0,002	0,003	0,003	0,002	0,002	0,002	0,002

5. RESULTS

In our experiment, we wanted to determine the influence of using an additional camera's RGB data on the improvement of the reflectance reconstruction performance by ANN models. With three cameras, six different camera combinations and thus six sets of data have been prepared, three for modelling ANNs based on the single camera and three for ANN input RGB values from two cameras. For each of five different LS sizes (Table 1), ANNs with a varying number of neurons from 3 to 48 in their hidden layers have been trained, each one for 41 times by both BP and LM training algorithms. The average and the best results have been recorded, compared, and visualised. Unfortunately, a slower CPU-executed LM learning algorithm shows better results than a significantly faster GPU-executed BP algorithm (Lazar, Javoršek & Hladnik, 2020). The calculations were performed on two computers, partially on a 4-core 2nd gen. i7 CPU with Nvidia 550 GPU and the rest on a 6-core 9th gen. i7 with Nvidia RTX 2060 GPU. Calculations consumed more than 1000 hours of CPU and 65 hours of GPU time.
The MSE performance of each ANN model trained with BP and LM learning algorithm has been calculated for the test set of samples, and the average and the best performance of 41 model iterations for each combination of cameras, the number of HLNs and LS size has been registered. The average and best test set MSE performance depending on the number of HLNs for 2RGB ANN trained with a medium-sized LS and both learning algorithms, with D600+GH4 camera combination, is shown in Figure 4 in the form of the third order polynomial approximation.

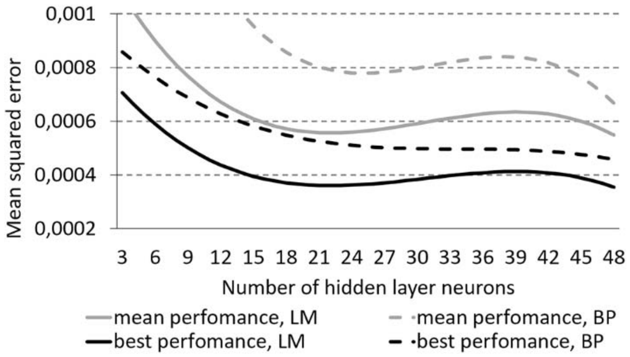

Figure 4: Mean and best test set performance of BP and LM trained ANNs, with 2RGB inputs of D600 & GH4 cameras and 41 iterations for each number of HLNs, with the third order regression trendlines for performances

Even when the training of each ANN model has been repeated many times, the mean and even more the best MSE values jump around a bit. It is possibly due to the random selection of training, validation, and test sets from the LS at the beginning of training for each ANN model. Besides, in most cases, the local minima of the cost function are found by the nature of the ANN training algorithm. In some cases, the polynomial regression curve is plotted along with the data plot, whereas for better illustration, only polynomial regression curves will be presented for the most part. The performance of ANNs trained with the LM algorithm is always better than the BP algorithm, so only the results obtained with the first one will be presented below.

The interpretation of results opens various aspects due to the collection of the ANN models' mean and best performance, variations in modelling ANNs with two different RGB input sets, six camera combinations and a varying number of HLNs:

- comparison of mean vs best MSEs, and varying LSs, for each camera combination
- comparison of all camera combinations best MSEs depending on the number of HLNs, for each fixed LS size
- comparison of the mean and best one- and two-camera MSEs, depending on the number of HLNs, for each fixed LS size
- comparison of all six pairs of 2RGB vs 1RGB ANN MSE performance improvements as a function of LS sizes

In the following, the procedures of these four aspects are described in more detail.

Observing the mean and best MSE in parallel as a function of the number of HLNs, by varying LS sizes for each of the six camera combinations shows that the best MSE values for all five LS sizes reside significantly below the mean values. Only in the case of one-camera data does the best MSE curve with the smallest LS at a higher number of HLNs partially surpass the mean MSE curve of the largest LS. For the D700+GH4 camera combination, the mean vs best trend observation is shown in Figure 5.

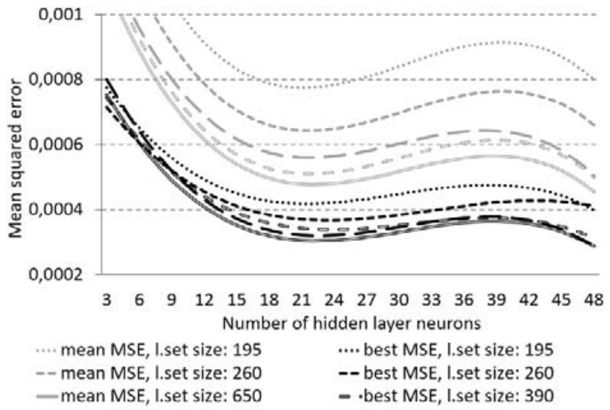

Figure 5: Mean and best MSE of 2RGB LM trained ANNs with D700 & GH4 camera combination and varying learning set size in dependence of the number of HLNs, shown by the trend lines of the 3rd order polynomial approximation.

Larger learning sets give better results in the search for the best performance. Here the ANN performances of LS sizes from 1171 to 390 colour patches almost overlap, and the trends for smaller LS sizes of 260 and 195 are still very close to the best results.

If we focus on ANN MSE performance as a function of the number of HLNs, observing in parallel all six camera combinations at each LS size separately, the ANNs trained with 2RGB perform appreciably better than with 1RGB input data. Figure 6 depicts the best performance for all six camera combinations and the smallest LS. By increasing the LS size from the smallest to the largest, the area between the trend curve of the 1RGB ANNs largest MSE and the curve of the 2RGB smallest MSE is almost halved. The range of MSE minima for the six camera combinations at the largest LS spans from 0.00030 to 0.00047, while at the smallest LS, from 0.00040 to 0.00068.

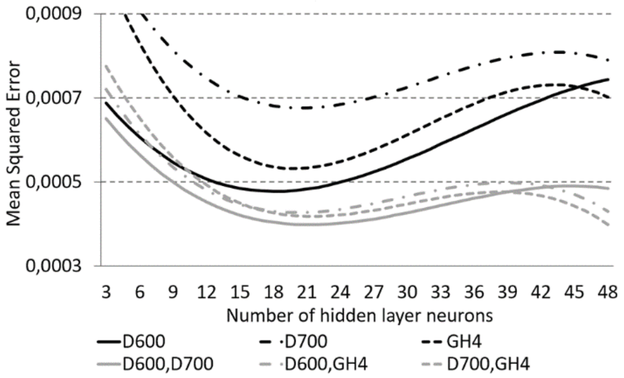

Figure 6: 3rd order polynomial regression trends of the best MSE performance ANNs trained with LM algorithm and smallest learning set (195), plotted for all six camera RGB input combinations

For each combination of two-camera models, the comparison of MSE performance has been made with 1RGB models of two involved cameras, and mean, and best performances have been visualised. In Figure 7, only one such combination is shown: the mean and best performance of ANNs trained with inputs from 2RGB D600+GH4 combination and 1RGB GH4, alongside the *MSE performance improvement with two versus one camera input, as a difference in [%]* = 100*(MSE_{1RGB}-MSE_{2RGB})/MSE_{1RGB}). For clarity, the third-order polynomial regression curves are displayed, with the actual values in only two cases, for illustration. The circle at 1 and 2RGB best MSE regression curves indicates the first minimum. In the nearby area, the best performing ANN models could be found within a moderate ANN modelling time (Lazar, Javoršek & Hladnik, 2020).

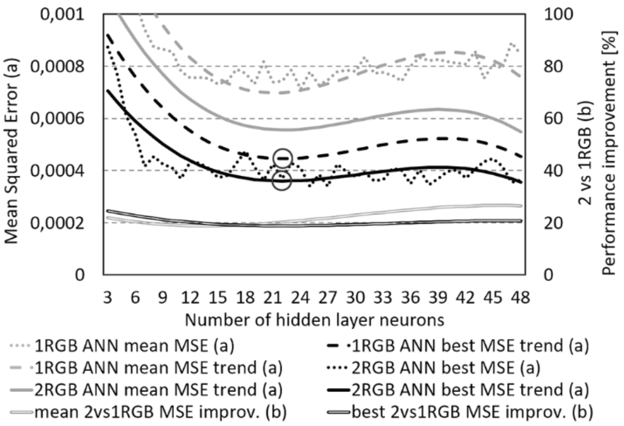

Figure 7: Mean and best performance of ANNs trained with LM algorithm and 1RGB (GH4) and 2RGB (D600+GH4) medium size learning set

To compare 2RGB versus 1RGB ANN MSE performance improvements as a function of LS sizes, the MSE performance improvement at the points of the first minimum of MSE 3rd order polynomial regression has been calculated for six two versus one-camera combinations and five LS sizes (Figure 8). Comparing the mean and the best MSE performance of 2RGB ANNs, the best is, on average, between 33% and 47%

better than the mean MSE, but, in terms of 2 vs 1RGB ANN MSE performance improvement, for the best ANNs, it is only about 5% better than the mean.

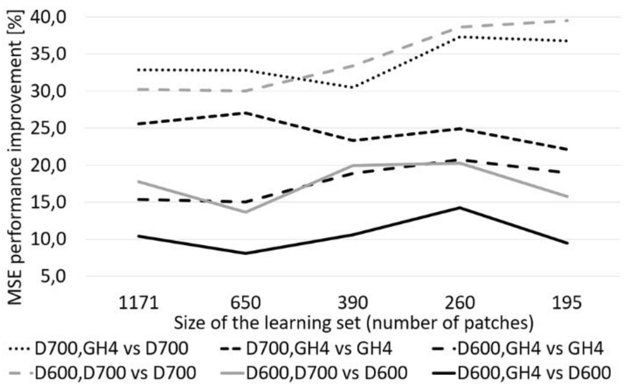

Figure 8: MSE performance improvement of 2RGB over 1RGB ANNs at peak values of the best 2RGB ANN performance for all two-camera combinations as a function of the learning set size

Reflectance reconstruction efficacy with additional quality measures has been evaluated for the best 1/2RGB ANN LM trained models, employing the five sets of leftover independent samples (Table 1). Two quality measures and one error measure were calculated to compare the measured (original) and the reconstructed spectra of independent samples.

The goodness of Fit Coefficient (GoFC) and the Peak Signal to Noise Ratio (PSNR) were used as the quality measures to compare both spectra directly. GoFC results were divided into four classes (Poor, Accurate, Good, Excellent) as proposed in (Hernández-Andrés, Romero & Lee, 2001), while PSNR into three (Poor, Accurate, Good) as proposed in (Lehtonen et al., 2009). Figure 9 compares the quality estimations of the best 1RGB and 2RGB LM trained ANNs with LS of 390 samples and, consequently, the 911 independent samples classified into the proposed classes. Only the sums of percentages for the two best classes are shown.

CIE 2000 was calculated as the error measure nearing the human eye's colour difference perception. Combining different illuminations (incandescent A, daylights D50, D65, and fluorescent F2) with the measured and reconstructed reflectance spectra, the pairs of L*a*b* colour values were calculated. Then their ΔE_{00} colour differences were sorted into seven quality classes (Hardly, Slight, Noticeable, Appreciable, Much, Very much and Strongly Perceptible Colour Difference) proposed in (Yang, Ming & Yu, 2012).

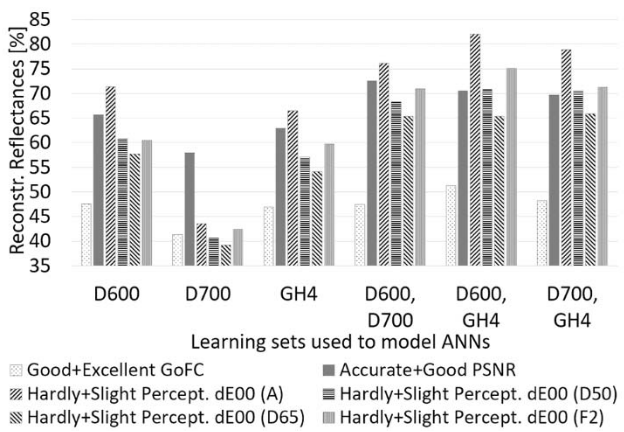

Figure 9: Reflectance reconstruction quality assessment of the best 1 and 2RGB ANNs trained with the LM algorithm and medium learning set size of 390 patches, with results that fall into the two best classes

The composite in Figure 10 presents some examples of bad and good reflectance spectra reconstruction. Falu Red, Sahara Yellow, Casal Blue and Deep Saphire Blue have "noticeable" or "appreciable" perceptible colour differences, "poor" PSNR and "poor" or "accurate" GoFC. Basket Ball Orange, Pastel Olive Green, Fountain Blue and Opera Mauve Purple have "hardly" or "slight" perceptible colour difference, "good" or "Excellent" GoFC and either "accurate" or "good" PSNR.

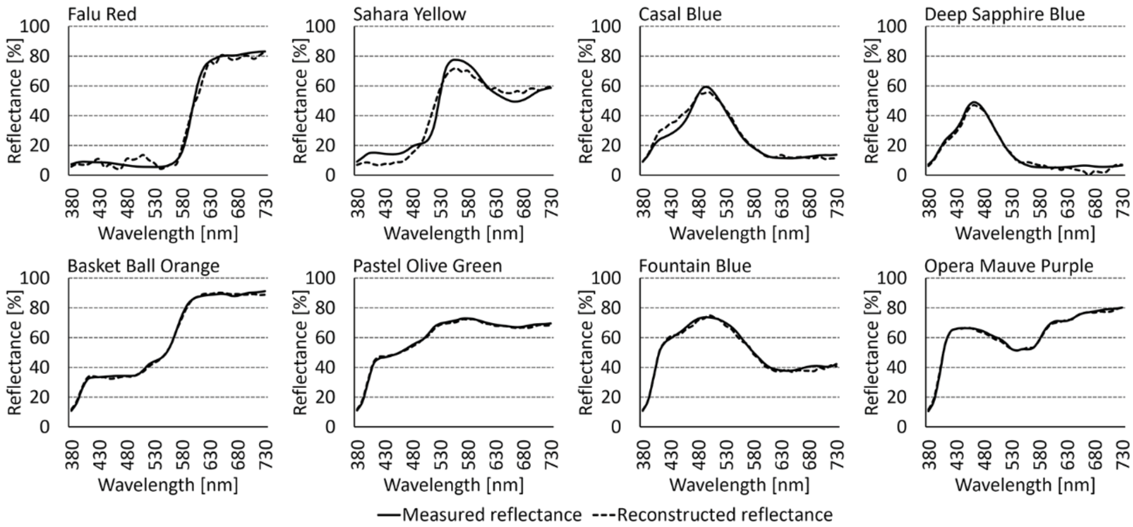

Figure 10: Examples of some bad and good reflectance reconstruction by 2RGB ANN trained with LM algorithm and medium learning set size of 260 patches

Shown reconstructions are made by 2RGB ANN trained upon LS of 260 patches (20% of the complete set) captured by D600 and GH4 cameras. Reconstructed samples belong to the separate independent set of 1041 patches not included in LS. For a better representation of the colour differences, in Figure 11, the above colours are also presented with colour fields, where the left field represents the measured (original) colour and the reconstructed colour on the right. The colours of the example patches have been calculated from the original and reconstructed spectra into AdobeRGB colour space considering D50 illuminant and a 2° observer.

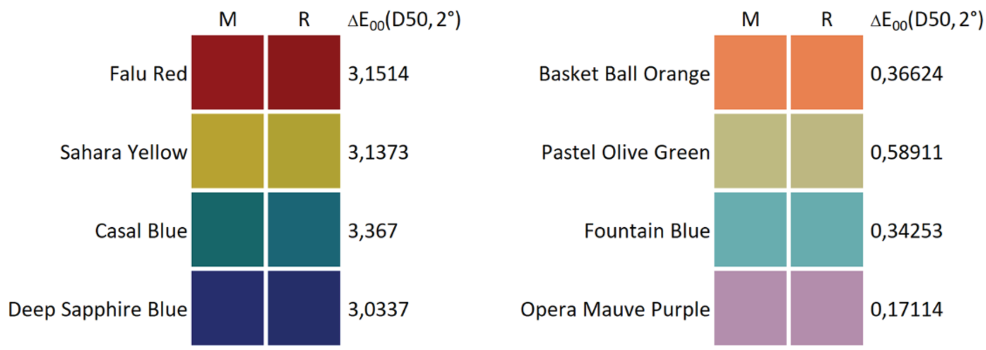

Figure 11: Colour examples of some measured (M) and reconstructed (R) colours, where the colours on the left have considerable and the colours on the right have small error estimates.

6. DISCUSSION

The described method has some limitations regarding camera settings and illumination, but they are not challenging to brace:

- The camera capturing mode must be set to manual and should not be changed during the time interval from capturing the object to capturing the table of reference colour patches.
- The illumination should be invariant in the area of the object and reference patches. Uneven illumination, shadows and glares are unwelcome.
- If possible, the reference colour table of suitable size should be captured simultaneously with the object of interest (e.g. drawing or painting). Otherwise, we should strive for constant

illumination during the time interval of capturing one after another, which is especially important when shooting in daylight.

- For capturing small flat objects, a flatbed scanner could be used. In this case, scanner options for image auto-correct options must be switched off during the time interval of sequential scanning of objects and reference samples.

When using two cameras to capture an object and reference patches, it is also necessary to ensure the registration of both captured images, which would be essential in practical implementation. Image registering is a broad topic that goes beyond the purpose of our experiment, in which capturing the same colour "pixels" with different cameras was provided through sequential capturing of the same homogenous colour patch surface under invariant illumination.

The expansion of the ANN training set from the data of one camera makes sense as long as the new input data contains additional information, which would not happen in the case of linear dependence of the equivalent channels of the cameras used. Finding nonlinearities to enhance reflectance reconstruction is thus crucial. In (Emancipator & Kroll, 1993), the lower limit of 2,5% as a criterion of nonlinearity assessment of curves for use in chemistry is proposed with the remark that in other fields of use, the lower limit may differ. In our experiment, only the blue channel, in some cases, with the 4th and third-order regression function, exceeds the proposed lower limit (Table 2). But, considering the improvement of the reflectance reconstruction, despite the otherwise small nonlinearities, we can assume that nonlinearities close to 1% are sufficient to improve the performance of 2RGB trained ANNs. In all considered cases, a nonlinear connection was detected between the equivalent channels of the experimental camera pairs, thus confirming our first hypothesis.

It is not straightforward to see a connection between the nonlinearities of paired cameras' equivalent channels and reflectance reconstruction improvement. In the R channel, QMoN values of D700 vs D600 exceed the other two camera combinations, in the G channel, this is the case with GH4+D600, and in the B channel, the GH4+D700 beats the other two camera pairs. Even though the D700+GH4 camera pair's QMoN for the R channel is lower than 0,5%, the 2RGB ANN performance still surpasses both the 1RGB ANNs performances, most likely due to some degree higher nonlinearities in the G and B channels.

Observing MSE performance in dependence on camera combinations (Figure 6), when training ANN with the 1RGB LS, the ANNs trained with the D600 camera performs better than with GH4 and even better than with D700. With these results, it could be expected that 2RGB ANNs trained with data from the combination of better-performing single camera learning sets will thus perform better. However, on the contrary, ANNs trained with the D700+GH4 perform better than those trained with the D600+D700 or D600+GH4 learning sets. It suggests that when training ANNs for reflectance reconstruction from RGB data of two or more cameras, the impact of individual cameras' somewhat better performance is not crucial.

Set side by side, the MSE performance of double and single-camera readings trained ANN models at peak values of the best ANN performance as a function of LS sizes (Figure 8), the most pronounced improvement (> 35%) appears at D600+D700 compared to D700 at smaller LS, and the smallest at D600+GH4 compared to D600 (10%) with all LS sizes. It can be attributed to a good performance of the ANNs trained with the D600 single camera readings. But despite this, and only slight nonlinearity of the GH4 R channel in relation to the R channel of D600 (Table 2 - middle R column, Figure 3 - left graph), the improvement of these two-camera ANNs' performance, in comparison to D600 single-camera trained ANNs, is still evident. The most noticeable improvement has been accomplished with D700+GH4 compared to both single-camera ANN performances, where two-camera ANN performance vs D700 is for all learning sets well above 30%, and vs GH4 above 20%. Comparing the performance of ANNs trained with the learning set of a pair and then with two individual cameras, the performance improvement is evident in all two-camera combinations. Therefore, our second hypothesis was experimentally confirmed.

To verify our findings, the reflectance reconstruction efficacy has been tested on 911 independent samples, wholly separated from the medium-sized training set of 390 samples (Figure 9). Various quality measures were used: MSE clearly confirms our findings, ΔE_{00} for four different illuminants also shows noticeable improvement with 2RGB compared to 1RGB trained ANNs, whereas GoFC does not manifest an unequivocal distinction between 1 and 2RGB trained ANNs.

7. CONCLUSION

With the presented study, we wanted to improve the spectral reflectance reconstruction from camera RGB values while capturing object colours with two instead of one camera. As detected, the relationship between the equivalent channels of cameras employed proved to be nonlinear. Therefore, each camera gives a bit different aspect to colour readings and adds some enrichment to the learning set. The ANN training benefits from this information, resulting in a better reflectance reconstruction, as presented through the attending experiment and confirmation of our hypotheses.

There are certainly possibilities to improve the results of our experiment. The selection of samples in five learning sets is fixed, does not vary, and each smaller LS is a subset of a previous larger one. Learning sets have been selected visually and are most likely not entirely optimal. A better selection could give better results. Another possibility is upgrading the structure of the artificial neural network, e.g., by adding additional hidden layers. Nevertheless, these suggestions are out of the scope of the current experiment and may be explored in the future.

8. REFERENCES

Aldrich, C. (2002) *Exploratory analysis of metallurgical process data with neural networks and related methods*. Elsevier. Available from: doi: 10.1016/s1572-4409(02)x8001-8

Benouis, M., Medus, L. D., Saban, M., Ghemougui, A. & Rosado-Muñoz, A. (2021) Food Tray Sealing Fault Detection in Multi-Spectral Images Using Data Fusion and Deep Learning Techniques. *Journal of Imaging*, 7 (9), 186. Available from: doi: 10.3390/jimaging7090186

Chi, C., Yoo, H. & Ben-Ezra, M. (2010) Multi-spectral imaging by optimised wide band illumination. *International Journal of Computer Vision*. 86 (2), 140-151. Available from: doi:10.1007/s11263-008-0176-y

Ciresan, D. C., Meier, U., Masci, J., Gambardella, L. M. & Schmidhuber, J. (2011) Flexible, high performance convolutional neural networks for image classification. *Twenty-second international joint conference on artificial intelligence.* Available from: doi:10.5591/978-1-57735-516-8/IJCAI11-210

Colantonio, C., Pelosi, C., D'Alessandro, L., Sottile, S., Calabrò, G. & Melis, M. (2018) Hypercolorimetric multispectral imaging system for cultural heritage diagnostics: an innovative study for copper painting examination. *The European Physical Journal Plus*. 133 (12), 1-12. Available from: doi: 10.1140/epjp/i2018-12370-9

Cucci, C., Casini, A., Picollo, M., Poggesi, M. & Stefani, L. (2011) Open issues in hyperspectral imaging for diagnostics on paintings: when high-spectral and spatial resolution turns into data redundancy. In *O3A: Optics for Arts, Architecture, and Archaeology III* (Vol. 8084, p. 808408). International Society for Optics and Photonics. Available from: doi: 10.1117/12.889460

Emancipator, K. & Kroll, M. H. (1993) A quantitative measure of nonlinearity. *Clinical chemistry*. 39 (5), 766-772. Available from: doi: 10.1093/clinchem/39.5.766

Gat, N. (2000) Imaging spectroscopy using tunable filters: a review. *Wavelet Applications VII*, *4056*, 50-64. Available from: doi: 10.1117/12.381686

Hassan-Esfahani, L., Torres-Rua, A., Ticlavilca, A. M., Jensen, A. & McKee, M. (2014) Topsoil moisture estimation for precision agriculture using unmmaned aerial vehicle multispectral imagery. In *2014 IEEE geoscience and remote sensing symposium*, pp. 3263-3266. Available from: doi: 10.1109/IGARSS.2014.6947175

Hernández-Andrés, J., Romero, J. & Lee, R. L. (2001) Colorimetric and spectroradiometric characteristics of narrow-field-of-view clear skylight in Granada, Spain. *JOSA A*, 18 (2), 412-420. Available from: doi: 10.1364/josaa.18.000412

Hornik, K. (1991) Approximation capabilities of multilayer feedforward networks. *Neural networks*. 4 (2), 251-257. Available from: doi: 10.1016/0893-6080(91)90009-t

Imai, F. H. & Berns, R. S. (1999) Spectral estimation using trichromatic digital cameras. In *Proceedings of the International Symposium on Multispectral Imaging and Color Reproduction for Digital Archives.* Chiba University Chiba, Japan. pp. 1-8.

Jia, Y., Zheng, Y., Gu, L., Subpa-Asa, A., Lam, A., Sato, Y. & Sato, I. (2017) From RGB to spectrum for natural scenes via manifold-based mapping. In *Proceedings of the IEEE International Conference on Computer Vision*. pp. 4705-4713. Available from: doi: 10.1109/iccv.2017.504

Lazar, M., Javoršek, D. & Hladnik, A. (2020) Study of Camera Spectral Reflectance Reconstruction Performance using CPU and GPU Artificial Neural Network Modelling. *Tehnički vjesnik*. 27 (4), 1204-1212. Available from: doi: 10.17559/tv-20190526202030

Lehtonen, J., Parkkinen, J., Jaaskelainen, T. & Kamshilin, A. (2009) Principal component and sampling analysis of color spectra. *Optical review*. 16 (2), 81-90. Available from: doi: 10.1007/s10043-009-0015-6

Maloney, L. T. (1986) Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *JOSA A*. 3 (10), 1673-1683. Available from: doi: 10.1364/josaa.3.001673

Ninomiya, Y. & Fu, B. (2019) Thermal infrared multispectral remote sensing of lithology and mineralogy based on spectral properties of materials. *Ore Geology Reviews*. 108, 54-72. Available from: doi: 10.1016/j.oregeorev.2018.03.012

Rolnick, D. & Tegmark, M. (2017) The power of deeper networks for expressing natural functions. *arXiv preprint arXiv:1705.05502*.

Smits, B. (1999) An RGB-to-spectrum conversion for reflectances. *Journal of Graphics Tools*. 4 (4), 11-22. Available from: doi: 10.1080/10867651.1999.10487511

Szegedy, C., Toshev, A. & Erhan, D. (2013) Deep Neural Networks for Object Detection. *Advances in Neural Information Processing Systems.* 26, 2553-2561.

Wang, P., Wang, S., Zhang, Y. & Duan, X. (2022) Multispectral Image under Tissue Classification Algorithm in Screening of Cervical Cancer. *Journal of Healthcare Engineering.* Available from: doi: 10.1155/2022/9048123

Yang, Y., Ming, J., & Yu, N. (2012) Color image quality assessment based on CIEDE2000. *Advances in Multimedia*. Available from: doi: 10.1155/2012/273723

Zhang, Y. & Wallace, B. (2015) A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *arXiv preprint arXiv:1510.03820*. Available from: doi: 10.48550/arXiv.1510.03820