ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20210405174515
Original scientific paper

Exploiting Nonlinearity between Parallel Channels of Multiple Cameras for Accurate ANN Reconstruction of Reflectance Spectra

Mihael LAZAR*, Aleš HLADNIK

Abstract: Colour of an observed object is unambiguously described by its reflectance. Translation from a colour description in RGB space obtained with a digital camera into reflectance, independent of illuminant and camera's sensor characteristics, was performed through an artificial neural network (ANN). In the study, it was hypothesized that the ANN's performance of reflectance reconstruction could be improved by using extended learning datasets with two or three cameras RGB input sets instead of one, but only if the parallel channels of cameras used are not linearly dependent. Nonlinearity was assessed by a quantitative measure of nonlinearity (QMoN), the metric primarily developed for use in chemistry. A noticeable reflectance performance improvement has been found with two and three cameras, even though the cameras' parallel channels exerted only small degrees of nonlinearity. Close attention was paid to the impact of scattering of RGB readings around the ideal values on the reflectance reconstruction performance, and it has been found that the more pronounced scattering is inversely proportional to the performance of ANNs trained with a single-camera input learning set but shows no visible impact on the performance of ANNs trained with extended input learning sets.

Keywords: artificial neural network; extended learning set; multiple cameras; parallel channel's nonlinearity; reflectance spectra reconstruction

1 INTRODUCTION

An unambiguous description of object colours should be independent of both the illuminant and the capturing device. Unfortunately, a device-independent colour description in CIE 1931 XYZ colour space contains information on the illuminant spectral power distribution. To assure the investigated object colour match for all observers and through various illuminants, a spectral match must be achieved [1]. For this purpose, different image capturing strategies have been developed. Hyperspectral and multispectral imaging cameras can be expensive, and by increasing spatial and spectral resolution, time resolution may be compromised [2]. The other possibility is to use low-cost, ubiquitous RGB cameras, where colour readings are device-dependent and noticeably vary with different cameras. The development of methods for conversion of camera RGB triplet values into a reflectance spectrum results in the object colour description, which is independent of the image capturing conditions. Numerous methods of spectral reconstruction from RGB values were explored in earlier studies, of which several were mentioned in our previous work [3], and some are discussed below. They range from strictly mathematical, of finding a model that conforms to training data and all the constraints, to some other optimization methods, such as those based on the artificial neural networks (ANNs).

Smits in [4] described the algorithm to solve the conversion of RGB defined object and texture colours to spectrally defined reflectances for use in physically/spectrally-based rendering systems. Imai and Berns in [1] mentioned flaws of the conventional photographic process, like nonlinear photometric responses of film and possibly large variance in match equality due to metamerism, which result in a dramatic reduction of colour quality. They argued that by using conventional techniques of photography and scanning, accurate capture of the original object is impossible. Accuracy is critical for purposes such as artwork imaging for reproduction and archiving. They present an alternative approach to capture multispectral images with a combination of RGB digital camera and either absorption filters or multiple illuminations. Jia, Zheng, Gu et al. in [5] accurately reconstructed the spectrum from RGB values of photos taken in daylight and proposed a three-step algorithm. First, they reduced the dimensionality of natural scene images high dimensional spectra to a 3D embedding of spectra through nonlinear dimensionality reduction. Second, they calculated RGB values of observed spectra, with known lighting conditions and camera spectral response and learning the ANN to relate RGB to the 3D embedding of these spectra. Finally, a transformation from 3D embedding to high dimensional spectra takes place by building a low to high dimensionality dictionary. In [6], Wu proposed a novel adaptive compressive sensing-based method for recovering reflectance spectra from RGB images using dynamic principal components calculated from the dynamic eigenvector subspace formed by RGB value similarities from the training set selection. Nowadays, research is actively underway in multispectral image capture and in the area of converting RGB into spectral values or, better yet, into reflectances.

In most of the studies, the image capturing conditions - the characteristics of the camera, sensory response, and lighting conditions - must be known or assumed. However, the illumination spectrum is not always known, neither is camera spectral response, as different cameras have different sensor characteristics, which are usually unknown to us. Instead of using absorption filters or alternating bandpass illumination, an array of reference colour patches can be captured alongside the object of interest in the same conditions, with one or more cameras, and the relationship between RGB values and reflectance spectra of colour patches can be established. Our previous work [3] studied the possibility and effectiveness of an ANN-based spectral reflectance reconstruction from a single camera RGB data. Quality consumer cameras are affordable, and the question arises as to whether the efficiency of reflection reconstruction could be improved if RGB data from two or more camera sources were available.

Possible purposes for using reflectance spectra are many and vary depending on whether the surface reflection is complete or diffuse. If the surface is uniformly coloured, reflectance is obtained by a spectrophotometer. In the case where a complex scene must be spectrally described, e.g., reflectance spectrum for each point of the artistic image, however, measurement with a spectrophotometer is

impractical or impossible. For our later purpose of spectral description of works of art in the techniques of pastels, watercolours, crayons, coloured felt-tip pens or printed materials on matte paper (non-glossy), we experimented with colour patterns with diffuse reflection.

2 REFLECTANCE RECONSTRUCTION USING ANN

It was proven by Cybenko [7] that any continuous function can be uniformly approximated by a continuous ANN having a single internal, hidden layer and with an arbitrary continuous sigmoidal nonlinearity. Hornik [8] proved 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. Novel forms of NNs have been developed recently, like convolutional NNs, which are primarily used to analyse extensive input data like images and natural language [9, 10] and deep NNs with many hidden layers used to consume and store a large amount of complex knowledge [11, 12]. This research focuses on a relatively small learning set with a small number of inputs, so reducing input data dimensionality would not contribute to network efficiency. Our preliminary experiments indicated that by increasing the number of hidden layers, the reflectance reconstruction improves only to some extent. Considering this, we built a three-layer ANN consisting of an input layer that accepts *R*, *G* and *B* values of readings of one, two or three cameras, a hidden layer with a variable number of neurons and of the output layer, whose output values consist of a reconstructed reflectance vector with a 10 nm wavelength resolution ranging from 380 to 730 nm.

The structure of the ANN is shown in Fig. 1, where $x_{k,j}$ are *k*-th inputs, which can be RGB readings of one, two (shown) or three cameras (hereinafter referred to as 1RGB, 2RGB or 3RGB); $b_j^{(L)}$ 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.

Having a learning set of (1RGB/2RGB/3RGB, reflectance) pairs from the photographed set of patches, the network's supervised training is done by adjusting connection weights and biases via minimizing the network cost function. The R_n real-valued inputs to *n* input units (3, 6, or 9) in our case are limited in value range and the number of combinations, but can be viewed as random variables and thus cost function is defined as the average performance, where the average is taken with respect to the input environment measure μ, where $\mu(R_n) < \infty$. In this case, closeness is measured by the $L^P(\mu)$ distances [8]. With our choice of *p* value equal to 2, cost function corresponds to mean squared error (Eq. (1)):

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

where 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.

Matlab offers, for the purpose of function approximation, standard gradient descent based back-propagation (BP) ANN training algorithm, which is very fast due to the ability to run on GPU and also provides Levenberg-Marquardt (LM) algorithm, which only supports execution on CPU, but is particularly designed to minimize sum-of-squares error functions [13]. As in our previous research, now again we experienced that LM learning algorithm demands, with an increased number of hidden layer neurons, an order of magnitude more time, but on the other hand, offers an order of magnitude better convergence in the means of learning epochs repetition, and also gives better ANN performance compared to BP.

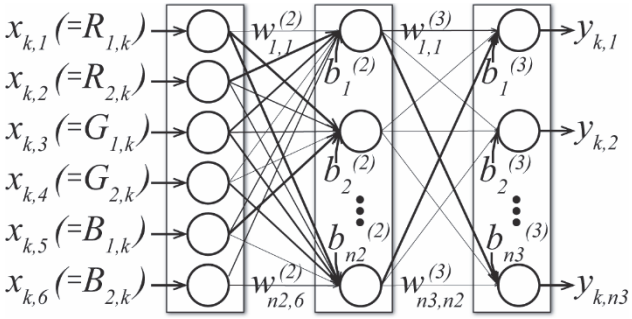

Figure 1 Architecture of an ANN with a single hidden layer and a variable number of hidden neurons, for the reconstruction of reflectances from RGB

Functions intrinsically contained in the modelled ANN translate every RGB reading into components of the reflectance spectrum. Each of these 36 functions translates the RGB readings into one spectral component; thus, one point in the 3D RGB camera space is projected into one spectral value. If we take two RGB points that are not linearly dependent, the triangulation of these into one value should give more accurate results and possibly even better if three linearly independent RGB points are used. This could be equivalent to providing a better description of the problem by giving more relevant input data to the system, which then finds a more accurate solution. The most illustrative would probably be a glance at the most doubtful look to the colours - the question of metamerism. Metamerism originating from the light source spectral power distribution cannot be decreased or eliminated this way. If, for example, one part of the illuminant spectrum is missing, it makes it impossible to capture this part of the reflected light from the object, regardless of the type or number of cameras. However, if the constant light source with smooth spectrum is used (e.g., sunlight or incandescent bulb) when the photo of an object and the photo of the reference patch array is taken, the object's and patch's reflectance should convey its complete spectrum, which is then transformed into camera RGB readouts. If cameras used in the experiment have a nonlinear relationship between parallel colour channels, the observer (or camera) metamerism could be reduced. The well-known integration defining the *R*, *G* and *B* values includes the product of illuminant spectrum $I(\lambda)$, object/sample spectral reflectance $S(\lambda)$ and camera spectral sensitivity function $\tau(\lambda)$, which is different for each channel of each camera (Eq. (2)):

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

where "*cam.*" indicates camera index (in our case, camera 1, 2 or 3), "*ch.*" represents colour channel (*R*, *G* or *B*) and "*i*" denotes colour index (e.g., colour 1, 2, … num. of colour patches).

If a metameric pair of samples is found for one camera, it means that this camera gives the same RGB outputs for both samples. The integrals of parallel 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 parallel channels are not linearly dependent on the first camera, the integrals for RGB values of the *second camera* will *possibly* give different results for the previously metameric pair of samples ($R_1 \neq R_2$ and/or $G_1 \neq G_2$ and/or $B_1 \neq B_2$). "Possibly", because RGB values can only be 8-bit integers, and too small differences result in the same values. The more nonlinearity expressed between the cameras, the greater the possibility of *different results for the previously metameric pair*. This leads to the conclusion that an additional camera provides additional information. The extraction of this information is analytically complicated to prove. Still, through the enrichment of the ANN learning set, with the learning algorithm able to learn through a series of examples, the benefit of the additional camera(s) would give a better ANN performance, which will hopefully be shown through our experiment.

As stated in the hypotheses below, we assume that the parallel colour channels of the cameras used are not linearly dependent. If so, we presume that the ANN performance will improve if the input set of readings is doubled (or tripled) so that for each input-output pair in the learning set, we present instead of one, input readings of two or three cameras, for each output reflectance.

Our hypotheses in this study are as follows:

1. Parallel channels (R_{C1}-R_{C2}, G_{C1}-G_{C2}, B_{C1}-B_{C2}) of different cameras are not necessarily linearly dependent; thus, such cameras' readings reveal some nonlinearity.
2. The scattering of the measured RGB values around the ideal curves, calculated from reflectance spectra to RGB conversion, affects the ANN model performance.
3. An ANN trained from RGB readings of two different cameras will perform better in reflectance reconstruction than an ANN trained from single-camera RGB readings.
4. An ANN trained from RGB readings of three different cameras will perform better in reflectance reconstruction than an ANN trained from RGB readings of only two cameras.

3 MATERIALS AND METHODS

The goal of our study was to explore a possibility to enhance the recovery of reflectance spectra from trichromatic camera values by supervised learning of a multilayer perceptron modelled with Matlab Neural Network Toolbox. Learning set inputs to train the ANN model are RGB values - readings from one or more cameras - and outputs are vectors of reflectance spectral components - readouts of the spectrophotometer.

Our source of colourimetric data for the experiment was The Munsell Book of Color Matte Collection, with its 44 sheets presenting 1301 patches with varying chroma, value and hue. 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. Each patch reflectance spectrum was 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 obtained with a maximum standard deviation of five measurements of less than 0.4%. For each patch, the average value of the five measurements was calculated.

Tristimulus errors are assumed to be negligible using 5 nm wavelength increment. Because of the absorption characteristics of human-made and natural colourants, the sampling rate can be greatly decreased [1]. Surface spectral reflectances of many organic and inorganic substances are characteristically smooth, low-pass functions of wavelength [14]. As our source of colour samples includes many natural colours (soil, skin, foliage etc.), the decrease of sampling rate from 5 nm to 10 nm is acceptable. The smoothness of observed reflectances has also been confirmed by spectrophotometric readings of all available colour patches in 3,33 nm step. Therefore subsampling was made from 107 to 36 reflectance spectral components, with a step of 10 nm in the range from 380 nm to 730 nm.

The sheets with colour patches were photographed in a photo studio, under constant lighting conditions with Nikon D600, D700 and Panasonic GH-4 cameras and fixed manual camera settings (Fig. 2). The spectral power distribution of the light source was measured with a spectrophotometer at the sheet position to obtain a correlated colour temperature (CCT) of 3019 K. To ensure constant conditions and to process all the RAW images equally, images were normalized using Adobe Lightroom software to the measured CCT, the tint was balanced to 0, and chromatic aberration, even though unnoticeable, was corrected for the profile "Nikon AF-S NIKKOR 50 mm f/1.4G" for photos taken with Nikon D600 and D700 and for built-in profile "Lumix G Vario 12-35/F2.8" for photos taken with Panasonic GH-4. In this process, images in RAW format were converted to RGB values of digital photo in Adobe RGB (1998) colour space. To avoid the influence of possibly divergent dots, the median of RGB values (instead of averaging) at the inner 50% square surface of each patch was acquired for all colour patches.

Figure 2 Photo studio setting

The complete set consisting of (RGB, reflectance) pairs for all 1301 Munsell Matte colour patches, with three camera's sets of *R*, *G* and *B* values and 36 values for each

reflectance vector, was assembled to feed into an ANN supervised learning algorithm. Before training, the entire set was divided into a learning set and the remaining independent samples for which additional reflectance reconstruction performance measures were calculated. The learning set was randomly split into training, validation and testing sets in a 70:15:15 ratio at the beginning of each repetition of the ANN model training.

Due to the nature of the ANN training algorithm, which usually finds a local instead of a global minimum of the cost function, the search for optimal ANN parameters was repeated 41 times for each model with selected cameras and selected learning set size, with the size of hidden layer varying between 3 and 48 neurons in steps of 1. The preliminary experiments with 21 learning repetitions revealed that by increasing the number of hidden layer neurons (hereafter HLNs) beyond 48, the probability of finding better ANN parameters does not improve noticeably, whereas the calculations become very time-consuming.

Table 1 Size of learning sets

descriptive size of learning set	% of learning set vs complete set of 1301 colour patches	num. of 1/2/3RGB - 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

In the presented study, however, we wanted to find the best ANN models with greater certainty, so the number of repetitions was doubled, i.e. to 42. The odd number of 41 repetitions was chosen due to the calculations of some additional statistics not presented in the article. The calculations with both learning algorithms (LM and BP) were performed for five different sizes of learning sets and sets of independent samples (Tab. 1).

4 NONLINEAR RELATIONSHIP BETWEEN CAMERAS AND SCATTERING OF CAMERA READINGS

In our experiment, we obtained *R*, *G* and *B* values for the patches with three different cameras. Polynomial regression has been performed in two ways: relating values of *R*, *G* and *B* channels of the readouts for the same patches between the three cameras and relating channel-wise each camera readouts (for the same patches) to the RGB values calculated from patch reflectance spectra measured by the spectrophotometer. For the calculation from spectrum to XYZ, the D50 illuminant and 2° standard observer was taken into account, and from XYZ to RGB we used the transformation matrix for D50 illuminant and conversion into Adobe RGB (1998). The relationship of parallel channels for GH4 and D600 cameras, their polynomial regression functions of the second-order and the residuals as the difference between the polynomial regression of the 2nd to 4th order and linear regression are shown in Fig. 3.

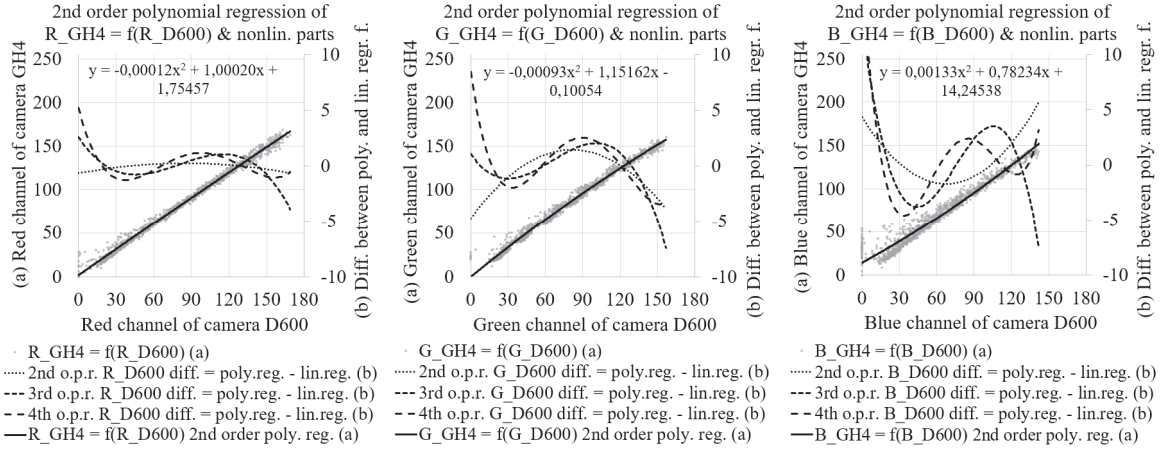

Figure 3 The relationship of parallel channels from camera GH4 and D600, the polynomial regression functions of the second-order and the residual parts as the difference between the 2nd to 4th order polynomial regression and the linear regression

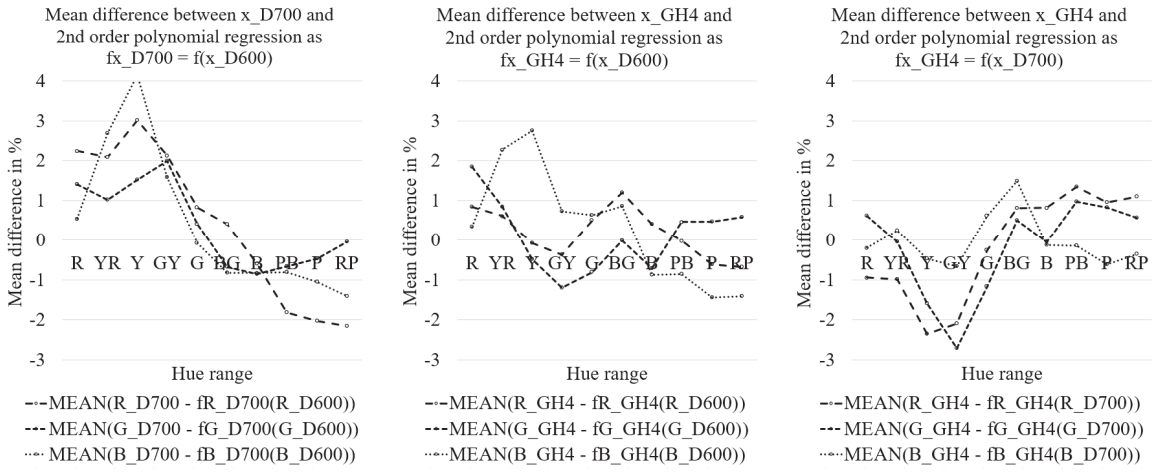

Figure 4 Average difference between RGB readings and the 2nd order polynomial regressions of the three cameras in the Munsell hue sections (x_camera denotes measured (*R*/*G*/*B*) channel values; fx_camera denotes second-order linear regression channel values as a function of another camera's channel values

To assess the polynomial regression's efficiency, the difference between the real data and the regression functions has been investigated. The difference between the 2nd order polynomial regression functions and the actual data averaged for each Munsell hue section varies from a maximum of 4,2% to –2,7% (Fig. 4). The average difference is practically 0% when including neutrals and lightly hued greys, with the standard deviation ranging from 1,49% to 2,8% through all channels and combinations of cameras. Standard deviation decreases with the 3rd order regression function, but the decrement is small, (1,46% to 2,68%). The decrease is smaller still with the 4th order (1,43% to 2,66%). Thus, the 2nd, the 3rd and optionally the 4th order regression is appropriate to observe one camera's sensor functions (*R*, *G* and *B*) in dependence of the other camera's sensor functions.

As a quantitative measure of nonlinearity (QMoN), the metric proposed by Emancipator and Kroll [15] was adopted. The "(dimensional) nonlinearity" of a method is defined as the square root of the mean of the square of the deviation of the response curve from a straight line, where the straight line is chosen to minimize the nonlinearity and the "relative nonlinearity" is defined as the dimensional nonlinearity divided by the difference between the maximum and minimum assayed values. The definition of quantitative nonlinearity measure accommodates different nonlinear regression procedures, among others also polynomial, which was used in our experiment. Due to a large number of assays, or in our case, patches (1301), and a relatively small number of different samples in terms of 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, around 1,2 and consequently the optional algorithm to search for suitable order of the polynomial regression, proposed by Emancipator and Kroll, fails. Nonlinearity has been thus calculated by their proposed relative QMoN for the polynomial regression for up to the 4th order, as shown in Tab. 2. The linear regression (1st order) was used as a nonlinearity measure control group, with indeed small values.

Table 2 Quantitative measures of nonlinearity between the same 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

To later assess the possible impact of camera readouts scattering on the performance of 1/2/3RGB trained ANNs, standard deviations were calculated to estimate the scatter for the measured values around the ideal ones for each channel of all three cameras (Tab. 3). The channel-wise ideal values are polynomial regression functions of each camera patch colour readout in dependence to the RGB value calculated from the patch reflectance spectrum (measured by the spectrophotometer) for all colour patches.

The standard deviation in *R* and *B* channels of values obtained by D700 camera is substantially higher than those of D600 and GH4. In *G* channel GH4 camera surpasses the other two cameras. The Blue channel of all three camera readouts in relation to the ideal values is shown in Fig. 5, where the difference in readout scattering around the 3rd order polynomial regression for each camera is visible.

Table 3 Standard deviations of measured from ideal values calculated by formulas for the 4th, 3rd and 2nd order polynomial regression

	R channel st. dev / %			*G* channel st. dev / %			*B* channel st. dev / %		
order of poly.	D600	D700	GH4	D600	D700	GH4	D600	D700	GH4
4th	3,33	**4,75**	3,81	2,74	4,54	**4,95**	4,66	**7,35**	6,84
3rd	3,53	**4,86**	3,94	2,85	4,64	**5,02**	4,79	**7,52**	6,86
2nd	4,18	**5,59**	4,53	3,35	4,93	**5,34**	5,34	**8,06**	7,03

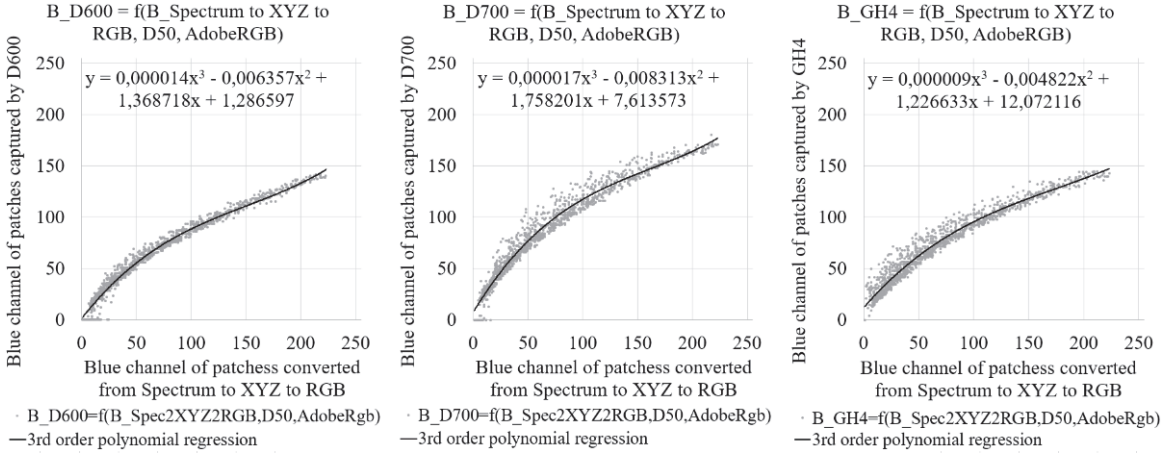

Figure 5 The blue channel of all three camera readouts in relation to the ideal values, calculated from reflectance spectra of all colour samples, with noticeably different scattering for different cameras around the 3rd order polynomial regression

Here the positions of D600/D700/GH4 *R*, *G* and *B* measured values are shown according to the ideal values - their regression functions (polynomials of 4th, third and second order) as a function of *R*, *G* and *B* values calculated from spectrums of observed patches (spectrum → XYZ → RGB (space: Adobe RGB, light source: D50)).

The scattering of D700 measurements is the largest of all three sets of measurements, while that of D600 measurements is the smallest.

5 RESULTS

In our experiment, we wanted to determine the influence of using RGB data from more than one camera on the modelled ANN's reflectance reconstruction performance. With three cameras, seven different camera combinations and thus seven learning sets have been prepared: three for modelling ANNs based on the single-camera RGB input values, three with ANN input RGB values of two cameras and one with inputs from all three cameras, each time with five different sizes of learning sets (Tab. 1). For each learning set, ANN with a varying number of HLNs from 3 to 48 has been trained 41 times with both BP and LM training algorithms, and the average and the best results were recorded, compared, and visualized. Unfortunately, CPU executed time-consuming LM algorithm gives noticeably better results than significantly faster GPU supported BP learning algorithm [3]. The calculations were performed on two computers, partially on 4-core i7 (2nd gen.) CPU with Nvidia 550 GPU and the rest on 6-core i7 (9th gen.) with Nvidia RTX 2060 GPU and in total demanded almost 1230 hours of CPU and 72 hours of GPU time.

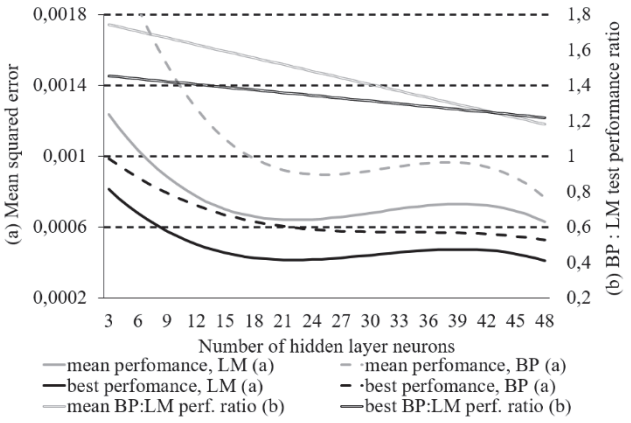

Figure 6 Mean, and best test set performance of BP and LM trained ANNs, with 2RGB inputs of D600 & GH4 cameras and 41 iterations for each num of HLN, with the 3rd order regression trendlines for performances and linear regression trendlines for BP:LM performance ratios

The 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 MSE performance of 41 models with the same number of HLNs and camera settings was recorded. The average and best test set performance depending on the number of HLNs for 2RGB ANN trained with medium size learning set and both learning algorithms, with D600 + GH4 camera combination are shown in Fig. 6, alongside with the BP/LM performance ratios, in the form of the 3rd order polynomial trends for performance and with linear trends for performance ratios. The BP/LM performance ratio is flatter (between 1,4 and 1,3) at larger learning sets. The slope increases as the learning set size decreases and is always steeper for the mean than for the best performance ratios. The largest slope is for smaller learning set at D600, D700 camera combination, where it spans for the mean performance ratio from 1,9 to 1,0, and for the best from 1,75 at small to 1,0 for a large number of HLNs. The BP/LM ratio remains inside these boundaries for all camera combinations, with the mean value between 1,3 and 1,5.

Although the training of each ANN model was repeated many times, the mean (and even more the best) MSE slightly varies, because the ANN training algorithm in most cases finds one of the local minima of the cost function. Variation of performance is also influenced by the random selection of training, validation and test sets from the learning set at the beginning of training for each ANN model. The polynomial regression curve is plotted along with the data plot. For better illustration, in most cases, only polynomial regression curves will be shown below.

Since the LM trained ANNs performance is always better than that of the BP trained, only the results obtained with the LM learning algorithm will be given below.

Variants of modelled ANNs with a different number of RGB input sets, inputs from different camera combinations, different number of HLNs, and collection of the mean (or average) and best performance lead the interpretation of results in various directions.

Monitoring the mean and best MSE in parallel for each of seven camera combinations and varying learning set size as a function of the number of HLNs, we found that *the best MSE values for all five learning set sizes lie far below mean values for all seven camera combinations.* Only in the case of 1RGB ANNs, it happens that the best MSE curve with the smallest learning set at a higher number of HLNs partially exceeds the mean MSE curve of the largest learning set. The mean vs best trend observation of D700 + GH4 camera combination is shown in Fig. 7. *Larger learning sets give better result in search of best MSE. Here the best MSE trends of learning set sizes from 90% to 30% (of the complete set) almost overlap, and the MSE performance trends for learning set sizes of 20% and 15% are still very close to best trends*.

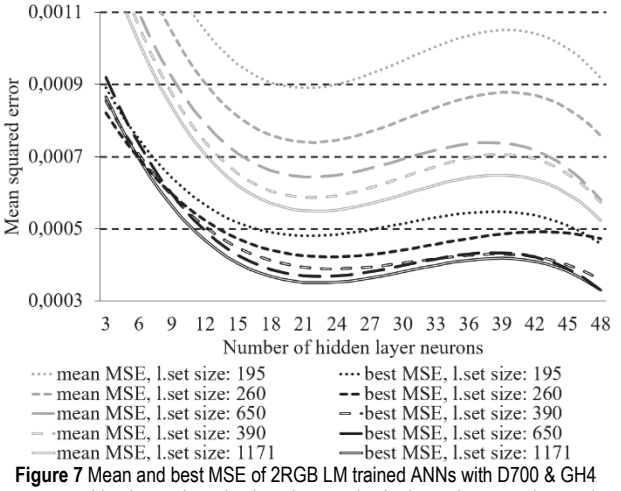

Figure 7 Mean and best MSE of 2RGB LM trained ANNs with D700 & GH4 camera combination and varying learning set size in dependence on the number of HLNs, shown by the trend lines of the 3rd order polynomial approximation.

If the learning set size is fixed and we focus on the MSE performance of ANNs, varying RGB input sets (alias camera combinations) as a function of the number of HLNs, the 2RGB trained ANNs perform visibly better than 1RGB ANNs and 3RGB ANNs perform better than 2RGB ANNs. The best performance for all camera combinations and the smallest learning set is shown in Fig. 8. As the learning set size increases from the smallest to the largest, the band (area) between the trend curve of the 1RGB ANN worst MSE and the curve of the best MSE for the 3RGB ANN narrows to almost half. The range of MSE minima for the seven camera combinations at the largest learning set spans from 0,00033 to 0,00053, while at the smallest learning set from 0,00043 to 0,00078.

For every ANN model with the combination of two-camera inputs, the comparison of MSE performance has been made to the performance of ANN models with single-camera RGB inputs, and for the ANN modelled with three-camera inputs the comparison has been made to the performance of all three ANN model with the combinations of two-camera RGB inputs.

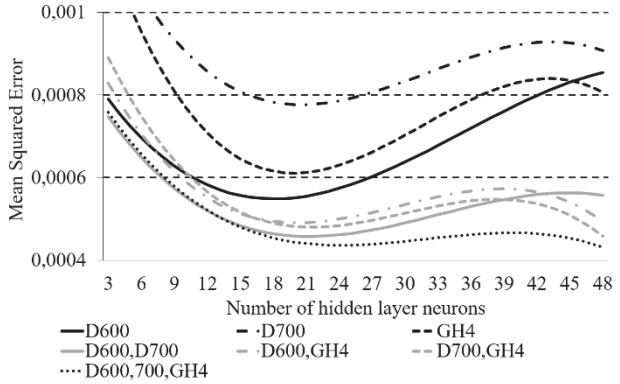

Figure 8 3rd order polynomial regression trends of the best MSE performance ANNs trained with LM algorithm and smallest learning set (195), plotted for all seven camera RGB input combinations

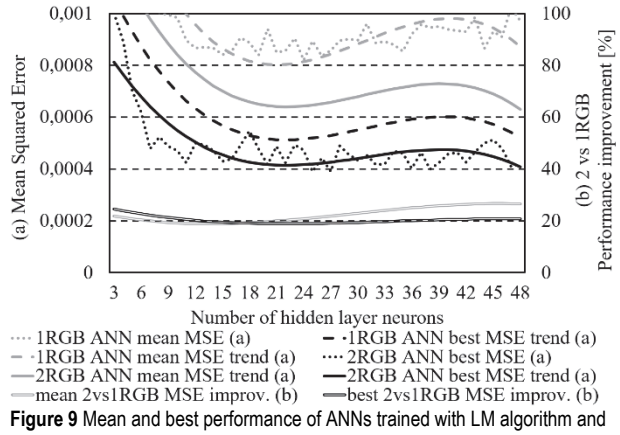

Figure 9 Mean and best performance of ANNs trained with LM algorithm and 1RGB (GH4) and 2RGB (D600 + GH4) medium size learning set

Mean, and best performances have been visualized. Here, only one such combination is shown: the mean and best ANN performances of ANNs trained with 2RGB learning set of D600 + GH4 inputs combination and 1RGB GH4 inputs along with *MSE performance improvement of two over one camera input* (their difference in [%] = $(100(MSE_{1RGB} - MSE_{2RGB})/MSE_{1RGB})$. For better comprehensibility, Fig. 9 shows the third-order polynomial regression curves, and the actual values are shown in only two cases for illustration. The first minimum of the regression function (marked in the illustration with a circle) suggests the number of HLNs in the vicinity of which the ANN model with good performance could be found with reasonable time investment [3].

To compare MSE performance improvements of 2RGB ANNs over 1RGB ANNs as a function of learning set sizes, the MSE performance improvement at the point of the first minimum of 2RGB MSE 3rd order polynomial regression has been found for all two-camera combinations and learning set sizes (Fig. 10). The same procedure has been followed for 3RGB ANNs over 2RGB ANNs.

Compring the mean (over 41 models for each number of HLNs) and the best MSE *performance* of 2RGB ANNs, the best is on average between 33% and 47% lower (better) than the mean MSE. But, in terms of *performance improvement*, for the best ANNs, it is only about 5% better than for the mean.

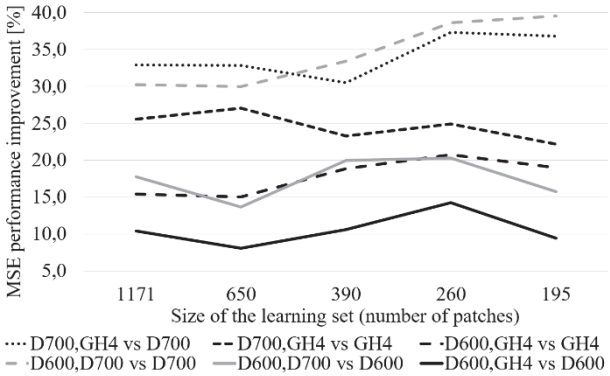

Figure 10 MSE performance improvement of 2RGB over 1RGB ANNs at peak values of the best 2RGB ANN performance for all two-camera combinations as a function of the learning set size

Evaluation of reflectance reconstruction efficacy of the best 1/2/3RGB ANN models trained with the LM algorithm was performed over independent samples that were excluded from the five learning sets (Tab. 1). One error measure and two quality measures were applied to compare the original and the reconstructed spectra.

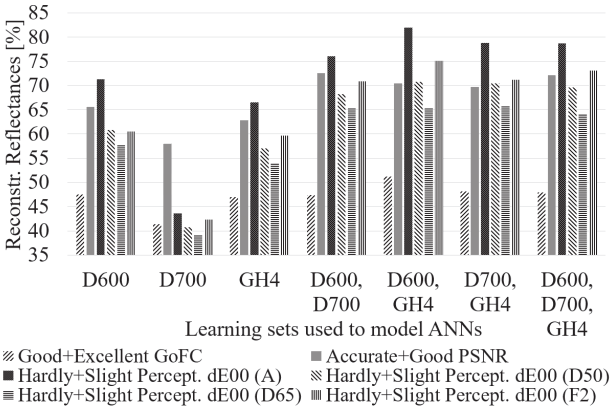

Figure 11 Reflectance reconstruction quality assessment of the best 1/2/3RGB ANNs trained with the LM algorithm and medium learning set size of 390 patches, with results that fall into the two best classes

As the error measure based on the human eye's characteristics, the CIE 2000 colour difference was used. Pairs of *L*a*b** colour values were calculated from the measured and reconstructed reflectances by considering different illuminants (incandescent A, daylight D50 and

D65, and fluorescent F2). For the obtained colour pairs, ΔE_{00} colour differences were calculated and classified into seven quality classes according to the groups (Hardly, Slight, Noticeable, Appreciable, Much, Verymuch and Strongly Perceptible Colour Difference) proposed by Yang, Ming and Yu [16].

As the quality measures, based on a direct comparison of the original and the reconstructed reflectance spectra of independent samples, *Goodness of Fit Coefficient* (GoFC) and *Peak Signal to Noise Ratio* (PSNR) were used. GoFC results were classified into four classes (Poor, Accurate, Good, Excellent) as proposed by Hernandez and Romero [17], and PSNR into three classes (Poor, Accurate, Good) as proposed by Lehtonen et al. [18]. Comparison of the quality assessment of best 1/2/3RGB ANNs trained with medium learning set of 390 samples, and thus 911 independent samples classified into two best classes, is shown graphically in Fig. 11.

Examples of some bad and good reflectance spectra reconstruction are shown in Fig. 12. Barn Red, Ecru Yellow, Army Green and Trout Gray have "noticeable" or "appreciable" perceptible colour difference, "poor" PSNR and "poor" or "accurate" GoFC. Clay Orange, Sea Green, Helio Violet, Walnut Brown have "hardly" or "slight" perceptible colour difference, good GoFC and either "good" or "accurate" PSNR.

Shown reconstructions are made by 2RGB ANN trained upon learning set containing values of 390 patches (30% of the complete set) captured by D700 and GH4 cameras. Reconstructed samples belong to the separate independent test set.

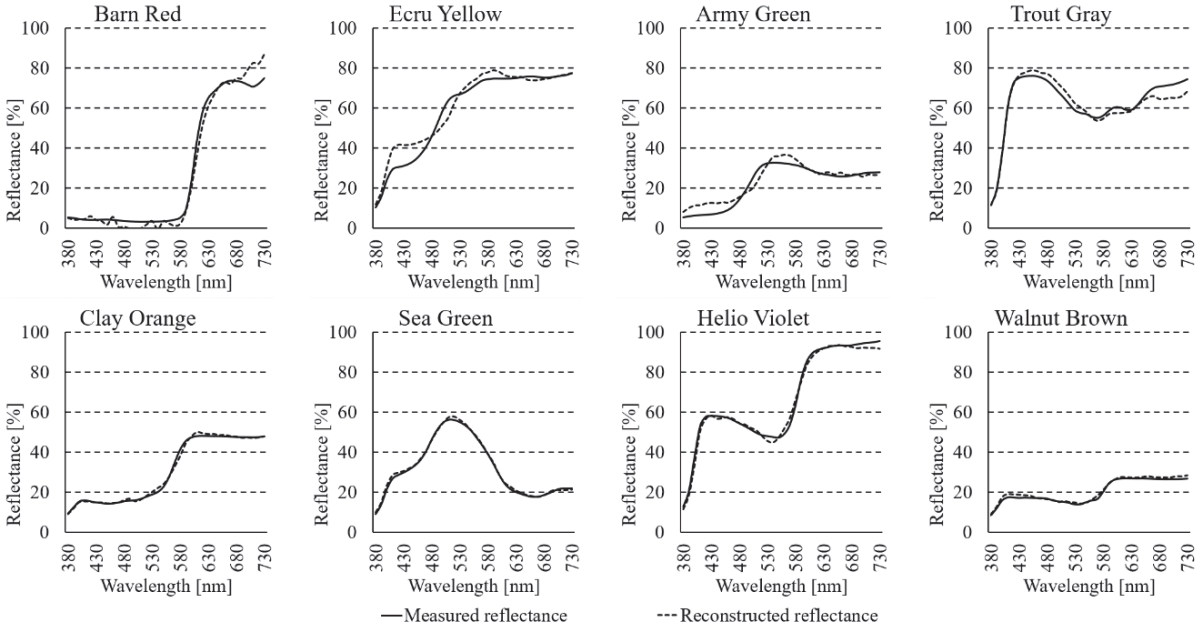

Figure 12 Examples of some bad and good reflectance reconstruction by 2RGB ANN trained with LM algorithm and medium learning set size of 390 patches

6 DISCUSSION

It should be mentioned that the described method has several limitations, which are not difficult to overcome:

- The illumination should be homogenous in the area of interest (in the area of photographed object and reference samples). Shadows or spotlights are undesirable.
- If reference samples are photographed before or after the capture of the object, time invariance of illumination during this time interval must be achieved. This is important if daylight is used for illumination. With the appropriate size of reference samples table, it could be captured simultaneously with the object of interest, e.g., work of art.
- Camera capturing mode must be put to manual settings and should not be changed during object capture - reference samples table capture interval.
- If scanners are used instead of cameras, which is possible for completely flat small objects (e.g., small pastel/aquarelle paintings), image auto correct options of the scanner software must be disabled to ensure consistency of sequentially scanned objects and reference samples.

Expansion of the ANN learning set input side, from 1RGB set to 2 or 3RGB sets, is beneficial only if some additional information could be drawn from this modification, which would not make the case if the parallel channels of the cameras used are linearly dependent. However, a nonlinear connection would, with an additional camera, add a look at the input data from an additional, slightly different perspective.

Only slight nonlinearities were found when analysing the nonlinear relationship between the parallel channels of the three cameras used in the experiment. The lower limit of 2,5% as a criterion of nonlinearity is proposed by Emancipator and Kroll, as these curves appear to be acceptably linear by visual interpretation, with the explanation that the limit is meant for use in chemistry. In other areas, this limit could be different [15]. In our case, only the blue channel in some cases, with the 4th and third-order regression function, suits the criterion (Tab. 2). Nevertheless, even so, we presume the nonlinearity near 1% could be sufficient to enhance the performance of 2RGB and 3RGB trained ANNs. In all cases, we detected some nonlinear connection between the parallel channels of the cameras used, thus confirming our first hypothesis.

In terms of reflectance reconstruction efficacy in connection to QMoN it is challenging to find an obvious correlation, as QMoN values of D700 vs D600 in *R* channel exceed the other two camera combinations, in *G* channel, this is the case with QMoN of GH4 vs D600, and in *B* channel QMoN of GH4 vs D700 surpasses other camera combinations. Although the *R* channel's nonlinearity for the D700 + GH4 combination is very small with QMoN under 0,5%, the ANN performance is still better with the 2RGB input combination than the 1RGB performance for individual cameras, possibly helped by slightly higher nonlinearity in the *G* channel and significantly higher nonlinearity in *B* channel.

Considering MSE performance in dependence on learning set camera combinations (Fig. 8), when training ANN with the 1RGB learning set of GH4 camera, it performs better than when training with the learning set of D700, which has the most considerable scattering of its RGB readings. Training of ANN with 1RGB learning set of D600 with the smallest scattering results in even better MSE performance. The performance of 1RGB trained ANNs is inverse proportional to the scattering of their camera RGB readings' input values. With these results, it could be expected that a 2RGB learning set containing data from the combination of cameras with smaller scattering will result in better ANN performance. However, on the contrary, ANNs trained with the D700 + GH4 learning set, containing higher scattering cameras, perform better than ANNs trained with the D600 + D700 or D600 + GH4 learning sets, with lower camera scattering readings. This suggests that when training ANNs for reflectance reconstruction from RGB data of two or more cameras, the impact of individual camera scattering is not significant. The 2nd hypothesis can be valid for one- but not for two-camera input RGB learning sets.

Regarding the MSE performance improvement of two-camera over single-camera ANN models at peak values of the best 2RGB ANN performance as a function of learning set sizes (Fig. 10), the highest improvement (more than 35%) is shown with D600 + D700 vs D700 with smaller learning set, and the lowest with D600 + GH4 vs D600 (around 10%) for all learning set sizes. This is due to good single-camera D600 ANN performance. But despite this and the small nonlinearity of the GH4 *R* channel in relation to the D600 *R* channel (second *R* column in Tab. 2 and left graph in Fig. 3), the improvement of MSE performance of ANNs, trained with these two-camera RGB sets, in comparison to 1RGB D600 trained ANNs, is still noticeable. The most significant average improvement has been achieved with D700+GH4 over both single-camera ANN performances, where two-camera performance vs D700 is for all learning sets well above 30%, and vs GH4 fairly above 20%. Given the noticeable improvement in the efficiency of ANN when using two cameras compared to one camera in all combinations of two-camera input sets, our 3rd hypothesis was experimentally confirmed.

The MSE performance improvement of three-camera over two-camera inputs trained ANN models is not pronounced but recognizable. Compared to D600 + GH4 trained ANNs, it ranges from 19% for large to 11% for medium and smaller learning sets. Compared to the other two 2RGB combinations, performance improvement with 3RGB is smaller but still between 14% and 4%. A slight improvement in the efficiency of the ANNs when using three cameras compared to two cameras in all combinations of input sets does not experimentally refute our 4th hypothesis. However, it would be worth considering whether a relatively small improvement justifies using three instead of two cameras.

To verify our findings, the reflectance reconstruction efficacy has been tested on 911 independent samples, completely separated from the medium-size training set of 390 samples (Fig. 11). The appraise of the reconstructed reflectances has been made through diverse quality measures, where MSE clearly confirms our findings, whereas GoFC does not show major differences comparing 1, 2 or 3RGB trained ANNs. ΔE_{00}, taking into account the possibility of different illuminations, also shows noticeable improvement with 2RGB compared to 1RGB trained ANNs. The improvement of 3RGB compared to 2RGB of trained ANNs is not explicitly expressed in any of the quality criteria used.

7 CONCLUSION

With the presented study, we wanted to improve the spectral reflectance reconstruction from camera RGB values while capturing object colours with more than one camera. As shown, the relationship between the parallel channels of cameras used in the experiment is not linear. Thus each camera gives a slightly different view of each RGB colour reading. Through such enrichment of the ANN learning set, the artificial neural network's learning algorithm benefited from the additional camera(s) information resulting in a better ANN performance, which has been shown through testing and discussion of our hypotheses in the experiment.

There is still some space to improve the results of our experiment of ANN supported reflectance reconstruction from the RGB sets of camera readings. The selection of our five learning sets has been fixed. Each smaller learning set is a subset of the larger one. Learning sets have been selected visually, and the selection most probably was not optimal. A better selection could give better results. Another possibility is a different ANN architecture, e.g., using two or more hidden layers of neurons instead of only one. However, these questions are out of the scope of the current experiment and remain to be explored in the future.

8 REFERENCES

[1] Imai, F. H. & Berns, R. S. (1999). Spectral estimation using trichromatic digital cameras. *Proceedings of the International Symposium on Multispectral Imaging and Color Reproduction for Digital Archives*, *42*, 1-8.

[2] 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. *O3A: Optics for Arts, Architecture, and Archaeology III*, *8084*, 808408. https://doi.org/10.1117/12.889460

[3] 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. https://doi.org/10.17559/tv-20190526202030

[4] Smits, B. (1999). An RGB-to-spectrum conversion for reflectances. *Journal of Graphics Tools*, *4*(4), 11-22.

https://doi.org/10.1080/10867651.1999.10487511
[5] 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. *Proceedings of the IEEE International Conference on Computer Vision*, 4705-4713. https://doi.org/10.1109/iccv.2017.504
[6] Wu, G. (2019). Reflectance spectra recovery from a single RGB image by adaptive compressive sensing. *Laser Physics Letters*, *16*(8), 085208. https://doi.org/10.1088/1612-202x/ab2b36
[7] Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, *2*(4), 303-314.https://doi.org/10.1007/bf02551274
[8] Hornik, K. (1991). Approximation capabilities of multilayerfeedforward networks. *Neural networks*, *4*(2), 251-257. https://doi.org/10.1016/0893-6080(91)90009-t
[9] 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.*
[10] Zhang, Y. & Wallace, B. (2015). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification.
[11] Rolnick, D. & Tegmark, M. (2017). The power of deeper networks for expressing natural functions.
[12] Szegedy, C., Toshev, A., & Erhan, D. (2013). Deep Neural Networks for Object Detection. *Advances in Neural Information Processing Systems, 26*, 2553-2561.
[13] Aldrich, C. (2002). *Exploratory analysis of metallurgical process data with neural networks and related methods*. Elsevier. https://doi.org/10.1016/s1572-4409(02)x8001-8
[14] Maloney, L. T. (1986). Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *JOSA A*, *3*(10), 1673-1683. https://doi.org/10.1364/josaa.3.001673
[15] Emancipator, K. & Kroll, M. H. (1993). A quantitative measure of nonlinearity. *Clinical chemistry*, *39*(5), 766-772. https://doi.org/10.1093/clinchem/39.5.766
[16] Yang, Y., Ming, J., & Yu, N. (2012). Color image quality assessment based on CIEDE2000. *Advances in Multimedia*, *2012*. https://doi.org/10.1155/2012/273723
[17] 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. https://doi.org/10.1364/josaa.18.000412
[18] Lehtonen, J., Parkkinen, J., Jaaskelainen, T., & Kamshilin, A. (2009). Principal component and sampling analysis of color spectra. *Optical review*, *16*(2), 81-90. https://doi.org/10.1007/s10043-009-0015-6

Contact information:

Mihael LAZAR, B.Sc.E.E.
(Corresponding author)
University of Ljubljana, Slovenia,
Faculty of Natural Sciences and Engineering,
Department of Textiles, Graphic Arts and Design,
Snežniška ulica 5, SI-1000 Ljubljana, Slovenia
E-mail: miha.lazar@ntf.uni-lj.si

Aleš HLADNIK, Associate Professor, PhD
University of Ljubljana, Slovenia,
Faculty of Natural Sciences and Engineering,
Department of Textiles, Graphic Arts and Design,
Snežniška ulica 5, SI-1000 Ljubljana, Slovenia
E-mail: ales.hladnik@ntf.uni-lj.si