The Transformation of RGB Images to Munsell Soil-Color Charts

Transformación de imágenes en RGB a tarjetas de color de suelo Munsell

Conversão de imagens em RGB em cartas de cores do solo de Munsell

Martín Solís[1*], **Erick Muñoz-Alvarado**[2], **María Carmen Pegalajar**[3]

Abstract

[Objective] The transformation from RGB to Munsell color space is a relevant issue for different tasks, such as the identification of soil taxonomy, organic materials, rock materials, skin types, among others. This research aims to develop alternatives based on feedforward networks and the convolutional neural networks to predict the hue, value, and chroma in the Munsell soil-color charts (MSCCs) from RGB images. **[Methodology]** We used images of Munsell soil-color charts from 2000 and 2009 versions taken from Millota *et al*. (2018) to train and test the models. A division of 2856 images in 10% for testing, 20% for validation, and 70% for training was used to build the models. **[Results]** The best approach was the convolutional neural networks for classification with 93% of total accuracy of hue, value, and chroma combination; it comprises three CNN, one for the hue prediction, another for value prediction, and the last one for chroma prediction. However, the three best models show closeness between the prediction and real values according to the CIEDE2000 distance. The cases classified incorrectly with this approach had a CIEDE2000 average of 0.27 and a standard deviation of 1.06. **[Conclusions]** The models demonstrated better color recognition in uncontrolled environments than the Transformation of Centore, which is the classical method to transform from RGB to HVC. The results were promising, but the model should be tested with real images at different applications, such as soil real images, to classify the soil color.

Keywords: Munsell color space; RGB color space; transformation; Munsell soil color charts; machine learning; neuronal networks.

Resumen

[Objetivo] La transformación del espacio de color RGB al de color Munsell es un tema relevante para diferentes tareas como la identificación de: la taxonomía del suelo, materiales orgánicos, materiales rocosos. tipo de piel entre otros. Esta investigación tiene como objetivo desarrollar alternativas basadas en las redes feedforward y las Redes Neuronales Convolucionales para predecir el tono, el valor y el croma en las cartas de color del suelo

Martín Solís, marsolis@itcr.ac.cr, https://orcid.org/0000-0003-4750-1198
Erick Muñoz-Alvarado, erickamacr3@gmail.com, https://orcid.org/0000-0002-6244-3398
María Carmen Pegalajar, mcarmen@go.ugr.es, https://orcid.org/0000-0001-9408-6770
* Corresponding author
1   Tecnológico de Costa Rica, Cartago, Costa Rica
2   Tecnológico de Costa Rica, Escuela de Ingeniería en Computadores, Cartago, Costa Rica
3   Department of Computer Science and Artificial Intelligence, Universidad de Granada, Granada, España

de Munsell (MSCC) a partir de imágenes RGB. **[Metodología]** Con el fin de entrenar y probar los modelos, usamos imágenes de los gráficos de colores de suelo de Munsell de las versiones 2000 y 2009 tomadas de Millota *et al.* (2018). Se utilizó una división de 2856 imágenes en 10% para pruebas, 20% para validación y 70% para entrenamiento con miras a construir los modelos. **[Resultados]** El mejor enfoque fueron las redes neuronales convolucionales para la clasificación con un 93% de precisión total de la combinación de tono, valor y croma (consta de tres CNN, uno para la predicción de tono, otra para la de valor y la última para la de croma), aunque los tres mejores modelos muestran cercanía entre la predicción y los valores reales según la distancia CIEDE2000. Los casos clasificados incorrectamente con este enfoque tuvieron un promedio CIEDE2000 de 0.27 y una desviación estándar de 1.06. **[Conclusiones]** Los modelos demostraron un mejor reconocimiento de color en entornos no controlados que la transformación de Centore, la cual es el método clásico para transformar de RGB a HVC. Los resultados fueron prometedores, pero el modelo debe evaluarse ampliamente con imágenes reales del suelo para clasificar su color.

Palabras clave: espacio de color Munsell; espacio de color RGB; transformación; cartas de color del suelo Munsell; aprendizaje automático; redes neuronales

Resumo

[Objetivo] A conversão do espaço de cor RGB para o espaço de cores Munsell é um tema relevante para diferentes tarefas como a identificação: da taxonomia do solo, dos materiais orgânicos, dos materiais rochosos, do tipo de pele, dentre outros. Esta pesquisa tem como objetivo desenvolver alternativas baseadas nas redes feed-forward e nas Redes Neurais Convolucionais (CNN) para prever o matiz, o valor e o croma nas cartas de cores do solo de Munsell (MSCC) a partir de imagens RGB. **[Metodologia]** Para treinar e testar os modelos, usamos imagens dos gráficos de cores do solo de Munsell das versões 2000 e 2009 tomadas de Millota *et al.* (2018). Foi usada uma divisão de 2856 imagens em 10% para testes, 20% para validação e 70% para treinamento com o intuito de construir os modelos. **[Resultados]** O melhor enfoque foram as redes neurais convolucionais para a classificação com 93% de precisão total da combinação de matiz, valor e croma (consta de três CNN, um para a previsão de matiz, outra para a previsão de valor e a última para a previsão de croma), embora três melhores modelos tenham mostrado proximidade entre a previsão e os valores reais dependendo da distância CIEDE2000. Os casos classificados incorretamente com este enfoque tiveram uma média CIEDE2000 de 0,27 e um desvio padrão de 1,06. **[Conclusões]** Os modelos demonstraram um melhor reconhecimento de cor em ambientes não controlados que a conversão de Centore, que é o método clássico para converter de RGB a HVC. Os resultados foram prometedores, mas o modelo deve ser amplamente avaliado com imagens reais de solo para classificar sua cor.

Palavras-chave: espaço de cor Munsell, espaço de cor RGB, conversão, cartas de cores do solo de Munsell, aprendizagem automática, redes neuronais

Introduction

The Munsell soil-color charts (MSCCs) are derived from the Munsell space color and represent a part of the Munsell hue spectrum. There are several versions, but each is composed of a subset of charts representing a different hue in the Munsell space. Each chart has multiple color chips originating from the degrees of luminance and purity of color known as value and chroma. Figure 1 shows one of the seven charts of the USA2000 version.

The transformation from RGB to Munsell color space is a relevant issue for different tasks that require applying the conversion to images taken from traditional cameras or smartphones. For example, in Edaphology, the transformation of a sample soil image to the Munsell soil-color charts (MSCCs) is required to identify the taxonomy and characteristics of the soil in geographical areas (Sánchez-Marañón *et al.*, 2005; Domínguez Soto *et al.*, 2018; Ibáñez-Asensio *et al.*, 2013). Likewise, in archaeology, Munsell soil-color charts (MSCCs) are useful for identifying organic materials, soil proles, rock materials, textiles, metals, colored glasses, paintings, and the artifacts retrieved (Milotta *et al.*, 2017; Milotta *et al.*, 2018a). In addition, the transformation from RGB color space to Munsell soil-color charts (MSCCs) could be used to identify skin, hair, and eyes in Anthropology, Criminology, Pathology, and Forensic Medicine (Munsell Soil Color Charts, 2000). Some studies have aimed to build a model to predict the value, chroma, and hue of MSCCs from RGB images (Pegalajar *et al.*, 2018; Stanco *et al.*, 2011; Milotta *et al.*, 2017, Milotta *et al.*, 2018a, Milotta *et al.*, 2020) to solve the imprecision of professionals (Edaphologists and Archaeologists) when predicting value, chroma, and hue based on MSCCs. This paper is in the same line as those investigations. The goal is to find a model to accurately predict the HVC of MSCCs from images in RGB space captured in an uncontrolled environment. It means in the field without controlling the illumination, inclination, and position of the images.

The challenge to obtain an accurate model with images taken in an uncontrolled environment lies in the fact that two photos captured with the same device will have different RGB values due to the lighting exposed or other conditions such as the camera inclination or the distance between the camera and the object.
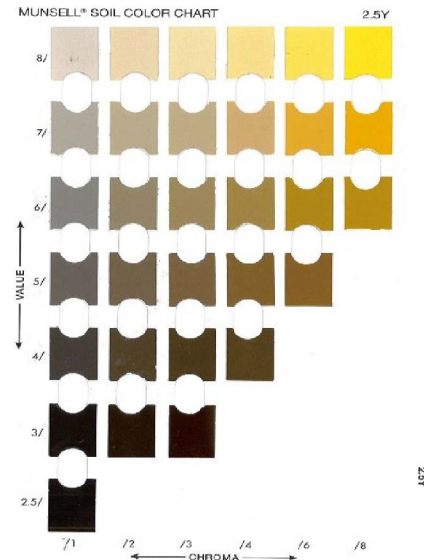

Figure 1. *2.5Y of USA2000 edition.*
Source: Munsell Soil Color Charts (2000).

Related work

ANN to Move From RGB to Another Space Color

Some studies experimented with artificial neural networks to predict a color space values from images represented in RGB. For example, León *et al.* (2006) proved different methods to move from RGB to L*a*b. These methods are the following: a) a transformation based on a linear function; b) a quadratic model, which considers the influence of the square of the variables (R, G, B) on the estimate of the values L*a*b; c) a direct model that, in the first step, carries out the RGB to XYZ transformation and, in the second step, the XYZ to L*a*b; d) a model that adds the gamma correction to the previous one; e) finally, a model based on artificial neural networks. The best prediction was obtained with an ANN of three layers.

The first layer took the RGB values of a pixel as input, the second was a hidden layer with three neurons, and the last gave the L, a, b values as output. Another example is the transformation from Munsell space to CIE XYZ through an ANN with four nodes (Viscarra-Rossel *et al.*, 2006).

The Transformation From RGB to Munsell Space

In Gómez-Robledo *et al.*, (2013), an application to transform an RGB soil sample image taken from a smartphone into Munsell notation was built. The images were taken under controlled conditions of lighting and position. Polynomial transformations using the pseudo-inversion method were implemented to convert one space into another.

Unlike Gómez-Robledo *et al.*, (2013), some researchers have created an application named ARCA (Milotta *et al.*, 2017, Milotta *et al.*, 2018a, Milotta *et al.*, 2018b) to predict hue, value, and chroma of MSCCs from images taken in an uncontrolled environment. For the prediction, they used the Centore transformation (Centore, 2011) and the application of the discretization to round the values predicted.

The study of Pegalajar *et al.*, (2018) was the first one, as far as we know, that applied machine learning to predict the HVC values from RGB images. In their system, three artificial neural networks received Red, Green, and Blue mean values of an image as input, and one of them released the hue value as output; another one, the value; and the last one, the chroma value. In a second stage, the HVC predicted for each image was used for a fuzzy system to return the exact MSCC chips.

In a recent study by Milotta *et al.*, (2020), a support vector classifier (SVC) was trained to classify images of MSCC chips taken in an uncontrolled environment to an HVC chip. The authors suggested applying deep learning techniques in future studies to improve the prediction.

Methodology

Data Generation

We used images of the MSCCs taken from Milotta *et al.* (2018b) for training and testing the model. The authors took photos of the Munsell soil-color charts from 2000 and 2009 versions in the most uncontrolled environment. We used only the hue charts that are common in both versions: 10R, 2.5YR, 5YR, 7.5YR, 10YR, 2.5Y, and 5Y. The photos were taken in Tampa-Florida (GPS coordinates 28°03' 47.9'' N 82°24' 40.9'' W), from 10:30 a.m. to 12:30 p.m. on different days, under 12 different settings (Milotta *et al.*, 2017), that come from:

1. Two devices: Canon EOS 1200D (18 megapixels) and Nexus 5X smartphone (12.2 megapixels).
2. Three automatic white balancing algorithms: automatic, sunny, and cloudy.
3. Two conditions: Only Munsell charts with a Gretag-Macbeth color checker nearby.

Of all the images, only those that had the Gretag-Macbeth color checker were used, as shown in Figure 2, because we wanted to incorporate, in the learning model, information about changes in the illumination, using the B1, D1, and E1 Gretag-Macbeth color patches. These colors work as a white reference to track the changes in the illumination.

Figure 2. *Example of Chart 10R with color patches.*
Extracted from Milotta *et al*. (2018b).

For the dataset preparation, we developed a tool to manually extract and label a piece of a sample of every chip along with the reference patches of the Gretag-Macbeth color checker. The sample images generated by the tool had a size of 8x8 pixels. In total, 78 pictures were used to create the complete dataset; this included the Munsell charts images from both available versions.

After the chip and reference extraction, two approaches were implemented. The first one appended the reference samples to the chip image to train convolutional neural networks. Since most of the information came from the chip, the images were resized until the output image had approximately a 3:1 size relationship between chip and references, as shown in Figure 3. In total, we obtained 2856 images to train, validate and test the models.

Figure 3. *Examples of the CNN Gretag-Macbeth input images.*
Note: derived from research.

The second approach permits training a feedforward neural network. In this case, a CSV file was generated for each chip of the dataset; each row in this file contained the ground truth HVC values of the chip, the average RGB values of the chip and the average RGB values of each B1, D1, and E1 reference.

In both cases, a division of 10% for testing, 20% for validation, and 70% for training was used. The partitions had the same image data for both the convolutional neural network and the feedforward neural network; this was done to have a common ground for comparison between the two approaches.

Proposed Models

The feedforward neural network architecture for regression and classification was as follows:

1. The input consisted of twelve values, three of which correspond to the average of the R, G, and B values of each chip. Nine values were the average of the R, G, B of the reference blanks. The input and output values were normalized from 0 to 1.
2. A hidden layer with 60 nodes and a sigmoid activation function.
3. The output layers were three for the regression nets: hue, value, and chroma. In the case of classification, we developed three nets with different outputs. The hue net had seven outputs, one for each possible hue. The Value net had seven outputs. The chroma net had six outputs.

The optimization function is the combined mean square error of hue, value, and chroma in the regression nets and the cross-entropy error in the classification

nets. The optimization algorithm is Levenberg-Marquardt, which combines the steepest gradient descent and the Gauss-Newton algorithms (Afifi & Brown, 2019), with a maximum of 1000 epochs and a stop criterion, which consists of stopping after 50 epochs without improvement in the validation sample's loss function. The regression model gives numerical predictions of hue, value, and chroma. However, we need to classify each output in one of the 248 possible value combinations of hue, value, and chroma (chips) of our Munsell Charts. Therefore, the Euclidean distance was used with the regression net predictions to find the most similar combination (chip) from all HVC possible values.
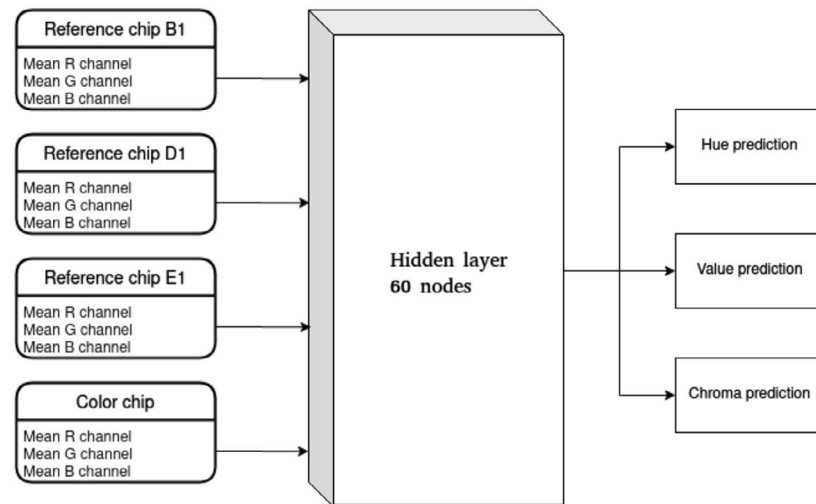

Figure 4. *Feedforward Neural Network. Note. For classification, we developed three nets with different outputs. The Hue and Value net had seven outputs, one for each possible value. The chroma net had six output.*
Note: derived from research.

We also applied a convolutional architecture for regression and classification. The objective was to develop a lightweight model that could complete the task successfully. The architecture of the model is as follows:

1. Input: The image generated from the chip and the white of reference, as in Figure 2, but amplified to 40x40 pixels in order to have a larger and wider image for the convolution process (Input).
2. 2D convolution, out_channels=32, kernel_size=5, stride=1, ReLU activation function, batch normalization and average pooling (conv_1+pool_1 in Figure 5).
3. 2D convolution, out_channels=64, kernel_size=3, stride=1, ReLU activation function, batch normalization and average pooling (conv_2+pool_2, in Figure 5).
4. Fully connected Layer, out_channels=4096, ReLU activation function (FCL_1, in Figure 5).
5. Fully connected Layer, out_channels=2048, ReLU activation function (FCL_2, in Figure 5).
6. Fully connected Layer, out_channels=1024, Sigmoid for regression and softmax for classification. (FCL_3, in Figure 5)
7. Output: Three outputs for the regression nets: hue, value, and chroma. In the case of classification, we developed three nets with different outputs. The Hue net had seven outputs, one for each possible hue. The Value net had seven outputs. The chroma net had six outputs.

The mean absolute error was used as the loss function for regression and cross-entropy for the classification nets. Adam was
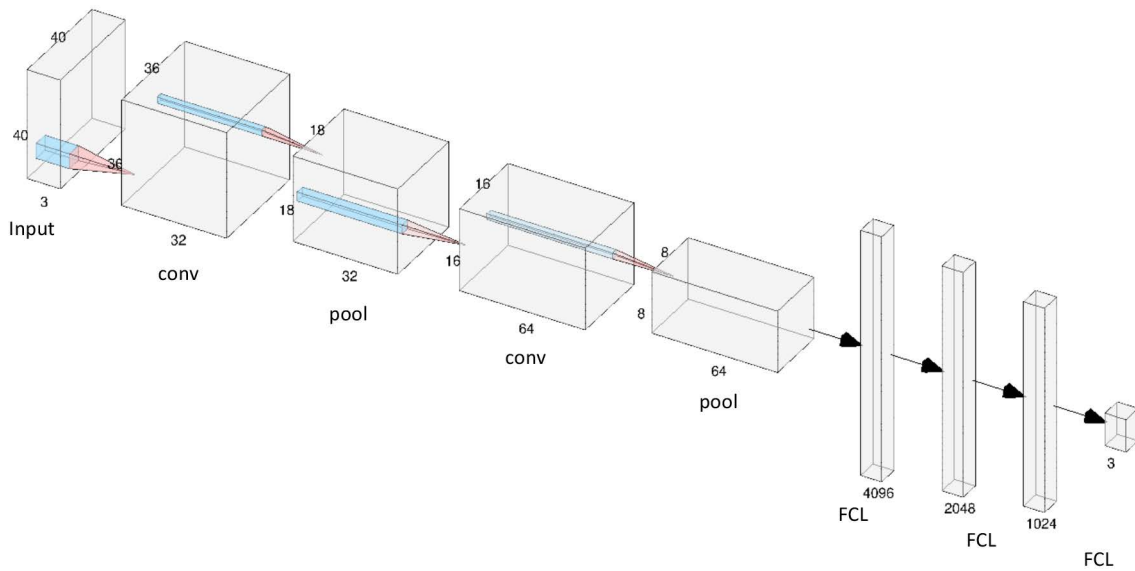
Figure 5. *Convolutional Neural Network for regression.*
Note. For classification, we developed three nets with different outputs. The Hue and Value net had seven outputs, one for each possible value. The chroma net had six output.
Note: derived from research.

the optimizer, with a learning rate and weight decay value of 0.0001. The weights were updated only when the validation accuracy of a specific epoch surpassed the best validation accuracy reported to the moment or a base accuracy value of 50%. The Euclidean distance was used with the regression net predictions to find the most similar chip from all HVC possible values. Thus, the numerical values were converted to classification values. The models were implemented in Python using Keras and executed in Google Colaboratory.

Analysis and Results

For the testing sample, Table 1 shows the percentage of chips classified correctly, the correct percentage classification of hue, value, and chroma, and the CIEDE2000 distance between prediction and ground truth for the different models. This distance was proposed to determine the color differences between the two images.

In order to have a baseline, the Transformation of Centore (Centore, 2011) was applied to the chips' RGB mean values for the prediction of the HVC values, similar to Stanco *et al*. (2011). After that, we used a discretization to round the values obtained from the transformation. This methodology achieved 61.2% accuracy, and the mean CIEDE2000 was 8.16, with a standard deviation of 3.64. Almost all the models generated surpassed the Accuracy of the Centore's Transformation; the only exception was the feedforward for regression without the whites of references (FFN_WWF). This result shows the importance of including information in the models about the illumination changes.

The best approach was the convolutional neural networks for classification (CNN_cla) with approximately 93% of total accuracy (it consisted of three CNN). The cases classified incorrectly with this approach had a CIEDE2000 average of 0.27 and a standard deviation of 1.06, which

were lesser than the method of Centore (Centore, 2011). According to Yang, Ming, and Yu (2012), a distance between 0 and 0.5 is hardly perceived. We developed three models in this approach instead of one because there were not enough images to build a classification model with 7*7*6 outputs.

Conclusions

The transformation from RGB to Munsell color space is a relevant issue for different tasks that require applying the conversion to images taken from traditional cameras or smartphones. In this paper, we proposed different learning models to predict the hue, value, and chroma in the Munsell soil-color charts (MSCCs) from RGB images. The best approach surpasses the Transformation of Centore (Centore, 2011) and shows 93.0% of accuracy. This approach is based on three convolutional neural networks that used images to combine the chip and whites of references as input to track the changes in the illumination. Although this percentage is promising, the model should be tested with real images at different applications, such as soil real images, to classify the soil color.

Furthermore, future research should emphasize improving the accuracy and the construction of a device-independent model to apply in images taken from other devices. An alternative to accomplish this task is the application of recent DNN that estimates the illuminant of an image independent of the sensor device RGB response to the scene's illumination (Hernandez-Juarez et al., 2020; Marquardt, 1963). These DNN could be applied, in the first stage, to the Munsell images to neutralize the illuminant effect, independent of the sensor's device. In the second stage, the neutralized images could be used to train an independent device model for HVC prediction.

Conflict of Interest

The authors declare no competing interests.

Author Contribution Statement

The total contribution percentage for the conceptualization, preparation, and correction of this paper was as follows: M.S. 35%., E.M. 35% and M.P. 30%.

Table 1
Accuracy and CIEDE200 distance of the models.

Model	Total	Hue	Value	Chroma	Mean_ CIEDE	Sd_CIEDE
Centore	61.2%	71.3%	94.1%	91.6%	8.16	3.64
FFN_WWF	23.4%	55.6%	81.8%	49.7%	4.15	3.40
FFN_reg	87.1%	95.5%	99.0%	90.2%	0.52	1.40
FFN12_cla	76.2%	94.1%	96.2%	83.9%	1.08	2.06
CNN_reg	82.2%	87.0%	98.1%	93.7%	0.57	1.41
CNN_cla	**92.9%**	**99.3%**	**98.9%**	**94.4%**	**0.27**	**1.06**

FFN_WWF= feedforward without white of reference, FFN_reg= feedforward for regression, FFN_cla= feedforward for classification, CNN_reg Convolutional Neural Network for regression, CNN_cla= Convolutional Neural Networks for classification. The best result is in bold.
Note: derived from research.

Data Availability Statement

The data supporting the results of this study is only available at http://iplab.dmi.unict.it/ARCA328/.

References

Afifi, M. & Brown, M. S. (2019). *Sensor-independent illumination estimation for DNN models.* arXiv preprint arXiv:1912.06888

Centore, P. (2011). An open-source inversion algorithm for the Munsell renotation. *Color Research & Application*, *37*(6), 455-464. https://doi.org/10.1002/col.20715

Domínguez Soto, J. M., Román Gutiérrez, A. D., Prieto García, F. & Acevedo Sandoval, O. (2018). Sistema de Notación Munsell y CIELab como herramienta para evaluación de color en suelos. *Revista Mexicana de Ciencias Agrícolas, 3*(1), 141–155. https://doi.org/10.29312/remexca.v3i1.1489

Gómez-Robledo, L., López-Ruiz, N., Melgosa, M., Palma, A. J., Capitán-Vallvey, L. F. & Sánchez-Marañón, M. (2013). Using the mobile phone as Munsell soil-colour sensor: An experiment under controlled illumination conditions. *Computers and Electronics in Agriculture, 99,* 200–208. https://doi.org/10.1016/j.compag.2013.10.002

Hernandez-Juarez, D., Parisot, S., Busam, B., Leonardis, A., Slabaugh, G. & McDonagh, S. (2020). *A Multi-Hypothesis Approach to Color Constancy.* 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/cvpr42600.2020.00234

Ibáñez-Asensio, S., Marqués-Mateu, A., Moreno-Ramón, H. & Balasch, S. (2013). Statistical relationships between soil colour and soil attributes in semiarid areas. *Biosystems Engineering, 116*(2), 120–129. https://doi.org/10.1016/j.biosystemseng.2013.07.013

León, K., Mery, D., Pedreschi, F. & León, J. (2006). Color measurement in L*a*b* units from RGB digital images. *Food Research International, 39*(10), 1084–1091. https://doi.org/10.1016/j.foodres.2006.03.006

Marquardt, D. W. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics, 11*(2), 431–441. https://doi.org/10.1137/0111030

Milotta, F. L. M., Stanco, F. & Tanasi, D. (2017). *ARCA (Automatic Recognition of Color for Archaeology): A Desktop Application for Munsell Estimation.* Lecture Notes in Computer Science, 661–671. https://doi.org/10.1007/978-3-319-68548-9_60

Milotta, F. L. M., Stanco, F., Tanasi, D. & Gueli, A. M. (2018a). Munsell Color Specification using ARCA (Automatic Recognition of Color for Archaeology). *Journal on Computing and Cultural Heritage, 11*(4), 1–15. https://doi.org/10.1145/3216463

Milotta, F. L. M., Quattrocchi, C., Stanco, F., Tanasi, D., Pasquale, S. & Gueli, A. M. (2018b). ARCA 2.0: Automatic Recognition of Color for Archaeology through a Web-Application. 2018 Metrology for Archaeology and Cultural Heritage (MetroArchaeo). https://doi.org/10.1109/metroarchaeo43810.2018.9089781

Milotta, F. L. M., Furnari, G., Quattrocchi, C., Pasquale, S., Allegra, D., Gueli, A. M., … Tanasi, D. (2020). Challenges in automatic Munsell color profiling for cultural heritage. *Pattern Recognition Letters, 131,* 135–141. https://doi.org/10.1016/j.patrec.2019.12.008

Munsell Soil Color Charts. (2000). *The Year 2000 revised washable edition.* Michigan, USA: Munsell Color 4300 44th Street SE, GrandRapids, MI 49512, USA; 2000.

Pegalajar, M. C., Sánchez-Marañón, M., Baca Ruíz, L. G., Mansilla, L. & Delgado, M. (2018). Artificial Neural Networks and Fuzzy Logic for Specifying the Color of an Image Using Munsell Soil-Color Charts. Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Foundations, 699–709. https://doi.org/10.1007/978-3-319-91473-2_59

Sánchez-Marañón, M., Huertas, R. & Melgosa, M. (2005). Colour variation in standard soil-colour charts. *Soil Research, 43(*7), 827. https://doi.org/10.1071/sr04169

Stanco, F., Tanasi, D., Bruna, A. & Maugeri, V. (2011). *Automatic Color Detection of Archaeological Pottery with Munsell System.* Lecture Notes in Computer Science, 337–346. https://doi.org/10.1007/978-3-642-24085-0_35

Viscarra-Rossel, R. A., Minasny, B., Roudier, P. & McBratney, A. B. (2006). Colour space models for soil science. *Geoderma, 133*(3-4), 320–337. https://doi.org/10.1016/j.geoderma.2005.07.017

Yang, Y., Ming, J., & Yu, N. (2012). Color Image Quality Assessment Based on CIEDE2000. Advances in Multimedia, 2012, 1–6. https://doi.org/10.1155/2012/273723