Open Source system for identification of corn leaf chlorophyll contents based on multispectral images

[1]Joe Saverio, [1]Allan Alarcón, [1]Jonathan Paillacho, [2]María Calderón, [1]Miguel Realpe

ESPOL Polytechnic University, Escuela Superior Politécnica del Litoral, ESPOL
[1] Facultad de Ingeniería en Electricidad y Computación - FIEC, CIDIS
[2] Facultad de Ciencias de la Vida - FCV
{jsaverio,asalarco,jspailla,mafercal,mrealpe}@espol.edu.ec

Abstract. It is important for farmers to know the level of chlorophyll in plants since this depends on the treatment they should give to their crops. There are two common classic methods to get chlorophyll values: from laboratory analysis and electronic devices. Both methods obtain the chlorophyll level of one sample at a time, although they can be destructive. The objective of this research is to develop a system that allows obtaining the chlorophyll level of plants using images.

Python programming language and different libraries of that language were used to develop the solution. It was decided to implement an image labeling module, a simple linear regression and a prediction module. The first module was used to create a database that links the values of the images with those of chlorophyll, which was then used to obtain linear regression in order to determine the relationship between these variables. Finally, the linear regression was used in the prediction system to obtain chlorophyll values from the images. The linear regression was trained with 92 images, obtaining a root-mean-square error of 7.27 SPAD units. While the testing was perform using 10 values getting a maximum error of 15.5%.

It is concluded that the system is appropriate for chlorophyll contents identification of corn leaves in field tests. However, it can also be adapted for other measurement and crops. The system can be downloaded at github.com/JoeSvr95/NDVI-Checking [1].

Keywords: multispectral images, image labeling, SPAD, NDVI, linear regression, open-source.

1 Introduction

One key feature for improving production quality of crops is the precise use of fertilizers, for this it is necessary to know the chlorophyll content of each crop, thus the fertilizer requirement can be estimated for each portion of the sown field. In the Experimental Agricultural Farm (Granja Experimental Agricola - GEA) of the ESPOL Polytechnic University, an institution that served as case of study for this work, the values of chlorophyll content of crops are obtained through an electronic device called SPAD, which determines the amount of chlorophyll in the plant non-destructively [2], and also through the analysis of plant tissues in a laboratory.

The processing time for obtaining the level of chlorophyll using SPAD electronic devices is immediate in addition of being not harmful to plants. On the other hand, the analysis of plant tissues is destructive to crops and requires expensive equipment and more time [3] because chlorophyll is not obtained immediately but is performed in a laboratory.

For the above, it is necessary to find a system that improves the process of obtaining chlorophyll from plants regarding data recollection time, processing time, and operating costs, as well as not being destructive to plants, in order to improve production of agricultural crops.

1.1 Analysis using SPAD devices

The SPAD (Soil Plant Analysis Development) device has two sensors where the leaf is placed and a measurement is taken in SPAD units, this measurement is given in the range of -9.9 and 199 [4]. The sensors measure the absorption of the leaf in two regions of wavelength, near infrared and red, using these two regions a numerical value of SPAD is calculated, which is proportional to the level of Chlorophyll of the plant. Using this method an immediate and acceptable response is obtained. Additionally, no harm is done to plants because it is not necessary to destroy nor to punch holes in them [5].

1.2 Geographic Information Systems (GIS) for agriculture

Geographic Information Systems (GIS) focused on agriculture refers to a non-invasive and automatic method to estimate the level of chlorophyll of plants in a crop, and therefore to determine the health status of plants. These systems capture, store, manipulate, analyze, manage and present spatial or geographic data. GIS focused on agriculture usually uses multispectral images captured by satellites or by UAVs in order to determine vegetation indexes related to the level of chlorophyll in plants [6, 7].

Vegetation indexes enhance the vegetative properties of an observation, indicating whether vegetation exists or not in an image. In order to visualize these indicators, multispectral images are used, which allow observing the radiation emitted by plants after absorbing solar energy. Multispectral images provide information that is not visible in the human's range of vision nor in images captured by conventional cameras [8]. The indexes are obtained from a mathematical calculation where different multispectral ranges are used, depending on the information to be analyzed. Due to the multispectral behavior of plants, approximately 90% of the information related to vegetation is found in the red and near infrared spectra [9]. One of the most commonly used indicators for crop monitoring applications that include image processing is the Normalized Difference Vegetation Index (NDVI).

1.3 NDVI

NDVI is an index that indicates the presence of biomass in a processed image [10, 11]. This indicator is used to estimate the quality and development of agricultural fields, making it easier for experts to take decisions in order to improve the production of the field. NDVI is commonly used to identify different areas, identify anomalies in crop seasons, give yield estimates, identify weeds, among other functions. The following equation is used to calculate the index:

$$NDVI = (NIR - Red) / (NIR + Red) \quad (1)$$

Where NIR refers to the near infrared spectral range, while Red refers to the red spectral range.

2 Methodology

A system is proposed to analyze multispectral images of crops. First, data collection is performed, which includes image capture and SPAD measurements. Then, NDVI and chlorophyll values are processed and stored. Finally, a linear regression model is trained in order to predict SPAD values using NDVI maps (see Fig. 1).

Fig. 1. Proposed methodology.

The data necessary for the development of the system is composed of the corn crops multispectral images and chlorophyll values obtained with a SPAD device (Fig. 2). For multispectral images acquisition of the corn crop at GEA, a MicaSense RedEdge-M camera was used. The camera model used has 5 lenses, where each one is capable of capturing images in a specific spectral band, such as blue, red, near red, green and near infrared [12]. These images of the 5 spectral bands were then processed to obtain RGB images and images that represent NDVI values of the corn crop.

Fig. 2. Data acquisition.

Once the RGB and NDVI images were obtained, a module for associating the NDVI values of an image with the value of the chlorophyll level obtained with the SPAD device was needed. To do this, OpenCV library was used in order to implement an image labeling module that allows the user to draw a region of interest in the NDVI image, and then label it with a chlorophyll value corresponding to that region. The average of the NDVI values that are within the selected region of interest will be stored in a database in order to be used later for training a linear regression.

Finally, a linear regression that represents the correlation in the corn crop data was performed using the NDVI images obtained from the corn plants and their corresponding SPAD values. Once the linear regression function was obtained, the chlorophyll values were predicted from crops images. Different NDVI values from images of corn leaves were tested.

3 Implementation

3.1 Database

The database structure was defined in such a way that it stores relevant information regarding the image in order to facilitate identification of the images that have been labeled, as well as information on the vegetation present in the image. Since the database was created using MongoDB, there is a collection structure with the following information:

```
{
  "_id": <ObjectId>,
  "name": "filename",
  "extension": "png",
  "NDVI": [
     0: 0.34905,
                     1: 0.32853
     ...
     ...
     9: 0.39588
  ],
  "SPAD": 2,
  "LAB": 3
}
```

3.2 **Labeling module**

The labeling module is designed so that the user can open an image that represents the NDVI values of crops and to be able to select the different vegetation zones where chlorophyll data have been collected with SPAD devices. Optionally, the user can simultaneously open an RGB image of the same data in order to facilitate the identification of vegetation areas where chlorophyll data were taken (see Fig. 3).

Fig. 3. Labeling module.

3.3 **Linear regression**

Depending on the range where the NDVI value is located, approximations of its physical characteristics can be determined. Thus, NDVI values were filtered so that only values representing vegetation are used. It was decided to take NDVI values greater than 0.1 since they represent a range from a young vegetation to an abundant vegetation, as it can be appreciated below:

According to what they represent [13]

- [-1, -0.1]: Water
- (-0.1, 0.1]: Rocks, sand or snow
- (0.1, 0.4]: Young plantations, shrubs, bushes
- (0.4, 0.9): Abundant vegetation

According to density of the region [14]:

- [-1, 0]: Without vegetation
- (0, 0.15]: Very low density
- (0.15, 0.3]: Low density
- (0.3, 0.45]: Dense
- (0.45, 0.6]: High density
- (0.6, 1]: Very high density

Additionally, data structures for storing NDVI and SPAD values were implemented in order to allow compatibility with the types of data used by the Python Scikit-learn library, this library was used to obtain the simple linear regression function between the NDVI values and the theoretical SPAD values, the coefficient of linear correlation between both variables and the root of the mean square error.

3.4 Prediction module

In order to verify the linear regression, a prediction tool was implemented in which the user can upload an image of a crop (5 images from Micasense camera) and obtain its chlorophyll values. This is done by selecting any part of the image so that the system provides the chlorophyll value in that selection (see Fig. 4).

Fig. 4. Prediction module.

4 Results

To validate the system, 92 images captured in the corn crop at GEA were used. Each image contains a leaf from a plant in the crop. The Fig. 5 shows a sample image that was captured one week after seeds were planted.

Fig. 5. Sample of RGB leaf image.

4.1 Image labeling

Image labeling was done on a computer with 8GB of RAM. From the tests carried out, it was observed that in each of the labeled sheets there are around 150,000 NDVI values, then the average value was used for training. A sample of the information that was stored in the database is shown in the Table 1.

Table 1. Sample values stored in the database.

Image	NDVI min	NDVI max	NDVI avg	SPAD
NDVI0434	0.57379013	0.77697020	0.69942983	32.2
NDVI0437	0.60665011	0.8095092	0.72107878	37.3
NDVI0440	0.54207837	0.76918220	0.65276326	37.4
NDVI0443	0.78050649	0.92617207	0.87607061	36.6
NDVI0446	0.86651057	0.92999970	0.89974493	38.3

4.2 Training module test

In this experiment the database obtained from the image labeling module was used. The NDVI values belonging to a region of interest were averaged, obtaining 92 NDVI values in total. After training a linear regression (see Fig. 6 and Table 2), the following equation was obtained: $y = 43.73495324 x - 2.9578003944927787$ and the root of the mean square error was equal to 7.2728 SPAD units.

Fig. 6. Linear regression NDVI - SPAD.

Table 2. Sample values of linear regression.

NDVI	SPAD	y
0.7012959	34.8	36.78035163
0.8696189	43	40.48576784
0.8162994	41.1	40.35309443
0.69903266	36.3	35.53141767

4.3 Prediction module test

Once the linear regression was obtained, the prediction module was tested using crop images (see Fig. 4). After loading an image to the program, any region can be selected, and then the NDVI value of that particular position and the chlorophyll value are shown using the linear regression calculated above.

The following table shows 10 chlorophyll level values obtained using the simple linear regressions mentioned previously and their respective relationship with the NDVI and the actual SPAD chlorophyll value.

Table 3. Prediction module test.

NDVI	SPAD	y
0.6682494	38.1	38.44229234
0.5511283	38.1	37.69348461
0.82342315	42.6	40.89074824
0.864155	47.9	41.14669231
0.6238252	35.7	41.24439288
0.5811929	35	35.64828601
0.5918739	35.4	35.20947708
0.61275303	35.2	36.07730117
0.6876595	35.7	35.95191117
0.9160053	45.1	42.59750256

5 Conclusions and future works

It is important that labeling module allows users to freely draw regions of interest in irregular shapes in order to avoid the acquisition of data of NDVI values that do not correspond to chlorophyll, such as areas that represent water, earth, stones, among others. Thus, the implemented module is not restricted to a specific shape (e.g. rectangle), this proved to be useful for quickly getting sample values of pixels inside leaves.

Regarding the results obtained with the simple linear regression, it is concluded that, because the leaves are in a similar state of health, most of the data were distributed in the same NDVI ranges where SPAD values are very close to each other, as seen in Figure 7. This makes the training phase difficult since a diverse range of NDVI values is not obtained as well as SPAD values.

Once the regressions were obtained, they were compared using GEA corn crops. Table 3 shows 10 tests realized where a maximum error of 15.5% was obtained, with an average of 4.67%.

Future experiments are planed where diverse values of chlorophyll are obtained by applying different levels of fertilizers to the plants. Additionally, a segmentation preprocess is recommended in the labeling module in order to reduce labeling time.

References

1. J. Saverio, Allan Alarcón. “NDVI-Checking” [Online]. Available: https://github.com/JoeSvr95/NDVI-Checking
2. H. Sainz Rozas and H. E. Echeverría, "Sainz Rozas, Hernán, and Hernán E. Echeverría. "Relación entre las lecturas del medidor de clorofila (Minolta SPAD 502) en distintos estadios del ciclo del cultivo de maíz y el rendimiento en grano.," Revista de la Facultad de Agronomía , p. 103 , 1998.
3. R. F. Muñoz-Huerta, R. G. Guevara-Gonzalez, L. M. Contreras-Medina, I. Torres-Pacheco, J. Prado-Olivarez and R. V. Ocampo-Velazquez, "A Review of Methods for Sensing the Nitrogen Status in Plants: Advantages, Disadvantages and Recent Advances,," Sensors, vol. 8, no. 13, 2013.
4. "The SPAD," [Online]. Available: http://www.gisiberica.com/MEDIDOR%20DE%20CLOROFILA/MCL502.htm.
5. R. Novoa S.-A. and N. Villagrán A., "Evaluación de un instrumento medidor de clorofila en la determinación de niveles de nitrógeno foliar en maíz," Agricultura Técnica, vol. 62, no. 1, pp. 166-171, 2002.
6. José Reyes 1 , Axel Godoy 1 , Miguel Realpe. "Uso de software de código abierto para fusión de imágenes agrícolas multiespectrales adquiridas con drones", 17th LACCEI International Multi-Conference for Engineering, Education, and Technology: Industry, Innovation, And Infrastructure for Sustainable Cities and Communities, 24-26 July 2019, Jamaica.

7. "Solución GIS" [Online]. Available: http://site.geoagro.com/es/producto/soluci%C3%B3n-gis.
8. "Soluciones inteligentes para el manejo de cultivos," [Online]. Available: https://www.micasense.com/es/altum.
9. M. Gilabert, J. Gonzalez-Piqueras and J. García-Haro, "Acerca de los índices de vegetación," Revista de teledetección: Revista de la Asociación Española de Teledetección, no. 8, pp. 1133-0953, 1997.
10. "An overview of the available layers and indices in Atlas," 16 Agosto 2019. [Online]. Available: https://support.micasense.com/hc/en-us/articles/22783730
11. E. Taipale, "NDVI vs. NDRE: What's the Difference?," 1 Mayo 2018. [Online]. Available: https://sentera.com/ndvi-vs-ndre-whats-difference.
12. "Micasense RedEdge-M" [Online]. Available: https://www.micasense.com/rededge-m.
13. "Why does NDVI, NDBI, NDWI Ranges From -1 to 1?," 6 Marzo 2018. [Online]. Available: www.gisresources.com/ndvi-ndbi-ndwi-ranges-1-1/.
14. A. López-Pérez, M. Ménez and D. Fernández-Reynoso, "Prioritization of Intervention Areas using a Morphometric Analysis and Vegetation Index," Water Technology and Sciences, no. 6, pp. 121-137, 2015.