Sentinel-2 Image Scene Classification over Alentejo Region Farmland

Kashyap Raiyani[1]
kshyp@uevora.pt
Teresa Gonçalves[1]
tcg@uevora.pt
Luis Rato[1,2]
lmr@uevora.pt
Pedro Salgueiro[1]
pds@uevora.pt
José R. Marques da Silva[3,4]
jmsilva@uevora.pt

[1] Departmento de Informática, Universidade de Évora, Portugal
[2] CIMA, Universidade de Évora, Portugal
[3] MED, Universidade de Évora, Portugal
[4] Agroinsider Lda., Évora, Portugal

Abstract

Given the wide-ranging farmland area, optical satellite images of farms are used to develop maps that reflect land dynamics and its behavior over different time frames, crops, and regions on various environmental conditions. In this regard, it is essential to identify and remove atmospheric distorted images to further prevent misleading information, since their presence severely restrict the use of optical satellite images for forecasting harvest dates, yield estimation, and manufacturing control in agriculture systems. These atmospheric distortions are frequent due to cloud, shadow, snow, and water cover over farmland. In this work, we developed a method to identify distortion covering images of corn crop farmland situated in the Alentejo Region of Portugal. The results are compared with the state-of-the-art (SOTA) Sen2Cor algorithm of the European Space Agency. Further, experimental results show that the developed image scene classifier model outperforms Sen2Cor by 10% in F1-measure.

1 Introduction

Agriculture in Europe has witnessed a substantial change after the creation of the Common Agriculture Policy (CAP)[1] in 1962. As a result, Food security [6] is ensured in most parts of Europe but the estimated global population growth 7 billion to 9 billion by 2050 [2] posses the challenge of producing more food [12]. The way to address this challenge is to rely on science and technology for possible answers.

Over the last few decades, many new technologies have been developed for or adapted to, agricultural use. Examples of these include low-cost positioning systems such as the Global Navigation Satellite System (GNSS) or the Geographic Information Systems (GIS), sensors mounted on agricultural machinery, geophysical sensors aimed at measuring soil properties, low-cost remote sensing techniques, and reliable devices to store, process and exchange/share information [3, 10]. Together, these new technologies have produced a large amount of affordable, high resolution information and have led to the development of site-specific agricultural management that is often termed Precision Agriculture.

There are many aspects related to Precision Agriculture and this work aims at investigating Sentinel-2 satellite images (or known as product) to gain information across different parcel/region and time. Resulting, a time data-series that takes land (usage) properties as input and outputs land dynamic which will provide information about environmental (such as soil, water and, weather) impact on the land and crop growth.

The existence of optical distortion such as clouds, shadows, snow, and water over land can mask true surface reflection resulting in false land information restricting the use of satellite images. To identify this distortion, sate-of-the-art (SOTA) Sen2Cor image scene classifier could be used. Sen2Cor is an algorithm whose main purpose is to correct single-date Sentinel-2 Level-1C products from the effects of the atmosphere and deliver a Level-2A surface reflectance product [7]. Level-2A (L2A) output consists of a Scene Classification (SCL) image with seven classes: Cirrus, Shadow, Snow, Water, Vegetation, Soil, and Cloud with low, mid, and high probability.

This document reports the work developed within the scope of the NIIAA (Núcleo de Investigação em Inteligência Artificial em Agricultura), project co-promoted by the company Agroinsider[1]. In this regard, we created a Sentinel-2 image scene classifier, and used the developed classifier over the corn parcel images to recognize atmospheric distortion.

[1] https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agricultural-policy/cap-glance_en

2 Developed Work

The health of plants can be determined by their biophysical parameters and can be measured by spectral information gathered using remote sensing. The physiological changes (due to crop stress) lead to a change in the spectral reflection/emission characteristics [8]. This observation of the stress factor during crop growth using, for example the Normalized Difference Vegetation Index (NDVI) [11] is a necessary stage to know the probable loss of production. NDVI values are affected by multiple factors such as available soil moisture, date of planting, air temperature, day length, and soil condition [9].

2.1 Study Area

With the help of Agroinsider, we acquired 170 (5 days apart) Sentinel-2 images from 05-01-2017 to 03-08-2019 of ten corn parcels from Alentejo region between (37°56'29.13" N , 8°22'21.95" W) and (37°55'32.44" N, 8°21'02.23" W) coordinates. Figure 1 shows the corresponding 2D image of the ten corn parcels (referred as parcel-1 to parcel-10 onwards).

Figure 1: Ten Corn Parcels from Alentejo Region.

Figure 2 shows the mean NDVI Value from 05-01-2017 to 03-08-2019 for parcel-1[2]. In it, the presence of atmospheric disturbance can be observed as sudden dips in the NDVI values, supported by the fact that it is not possible to lose crop growth and regain it within a range of 5 days (the observation cycle time). To calculated mean NDVI, for each point in the parcel, NDVI was calculated using equation 1, and the overall sum value was divided by the total number of points. Here, NIR means Near Infra-Red (Band 8) and RED is Band 4.

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

2.2 Scene Classification and Results

Holstein [4] created a database of manually labeled Sentinel-2 spectra. The database consists of images acquired over the entire globe and comprises 6.6 million points from 60 different products classified into six classes as clear-sky, cloud, cirrus, shadow, snow, and water. The database is described by 4 attributes: *product_id*, *latitude*, *longitude* and *class*. To build a classifier, we extended this database adding corresponding Sentinel-2 13 bands values and, for comparison purposes, Sen2Cor scene classification. The final structure of the database is detailed in Table 1.

Instead of using standard train_test_split from Scikit-Learn library [5], we selected 59 products for training, and 1 for testing. The main reason to split the dataset in this way was to make sure that the knowledge about

[2] The same can be replicated to rest of parcels.

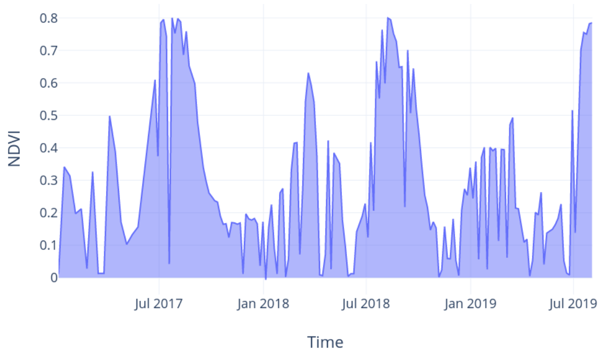

Figure 2: Mean NDVI Value for parcel-1 from 05-01-2017 to 03-08-2019.

Header	Column Value
Product ID	1 Column (78 character string)
Coordinates	4 Columns (latitude, longitude, east and, north)
Bands	13 Columns (Band 1 to 12 and 8A)
Tagged Class	1 Column (Manual tagged class value)
Sen2cor - SCL	1 Column (Scene classification class value)

Table 1: Structure of Final Dataset.

a region is not essential to classify that region. This reasoning enables us to pose the following question: will the system be able to classify it with good performance a new, non seen product? To evaluate this, it would be interesting to pick a complete region as a test set while all the rest of the points compose the training set. We replicated this procedure for each of the 60 products (use 1 for test and the rest 59 for train). We present the average F1 results. Equation 2 calculates the average $F1$ value (over 60 products) for each class where $F1_p$ is the $F1$ value of the particular class within the product p. N_p is the number of points of the class within the product p, T is the total number of points of the class for all products and $p \in (1,60)$ is the number of products.

$$F1 = \sum_{p=1}^{60} (F1_p \times N_p) \div T \;\; with \;\; T = \sum_{p=1}^{60} N_p \tag{2}$$

We used the Scikit-Learn library implementation of Decision Tree (DT), Random Forest (RF) and Extreme Trees (ET) algorithms. The obtained results were compared with the Sen2Cor algorithm. Table 2 details the results. These results show an F1 average value of 76.77% over all classes (using Extreme Trees), an improvement over 10% when compared to Sen2Cor F1 average value of 66.40%.

Class	DT	RF	ET	Sen2Cor	Support
Clear-sky	63.29	72.3	**74.16**	64.96	1694454 (25.56%)
Water	63.81	73.4	76.69	**80.73**	1071426 (16.16%)
Shadow	53.98	**63.96**	61.45	50.57	991393 (14.96%)
Cirrus	47.58	**56.63**	42.97	24.08	956623 (14.43%)
Cloud	65.25	75.08	**75.33**	75.04	1031819 (15.57%)
Snow	74.67	84.90	**87.00**	61.40	882763 (13.32%)
$F1_{avg}$	67.95	76.43	**76.77**	66.40	6628478 (100%)

Table 2: F1 values of ML algorithms and Sen2Cor.

Using the developed Extreme Tree model, the new, unseen optical images (with 13 bands) of the ten parcels mentioned in Subsection 2.1 were classified as no atmospheric disturbance image (clear-sky) or image with disturbance (cloud, shadow, snow, and water coverage). Here, each point within the parcel was classified using the model ET model built, resulting in a value between 0 if all points were classified as clear sky and 1 when all points were classified as atmospheric disturbance. Figure 3 presents the calculated disturbance over dates 14-06-2017 to 01-12-2017, with red line for the ET model and blue line mean NDVI. These results sync with sudden dips of the NDVI values supporting the claim of the presence of atmospheric disturbance in the optical image.

After analyzing Figure 3 closely, the authors would like to state that 'NDVI value is not the sole parameter to find disturbance'. This claim is supported by Figure 3 as on 08, 13, and 18 Aug'17, the mean NDVI ranges from 0.78 to 0.68 (a drop) to 0.76 but the value of atmospheric disturbance remains 0.0.

3 Conclusion

From our experiment results (Table 2), RF and ET are comparatively providing equivalent results and outperforming Sen2Cor by 10% F1 measure for image scene classification over a specific dataset composed by 6.6M entries acquired from 60 different products. Further, the results in Figure 3 support our claim: the ML model presented in this work is applicable as a base tool to identify the existence of clouds, shadows, snow, and water coverage over agriculture farmland images acquired using Sentinel 2 optical satellite. As a result, classified parcel images will help to prevent false surface reflectance information and allow the use of selected optical images for forecasting harvest dates, yield estimation, and manufacturing control. Given that the train ML model is over 60 different products acquired over the entire globe comprises 6.6 million points, the author expects similar results of identifying atmospheric disturbances over different crops.

As future work, we would like to: (1) manually label individual data points for each parcel as (0 or 1) atmospheric disturbance and, (2) compare the performance of the ML method to Sen2Cor over each parcel.

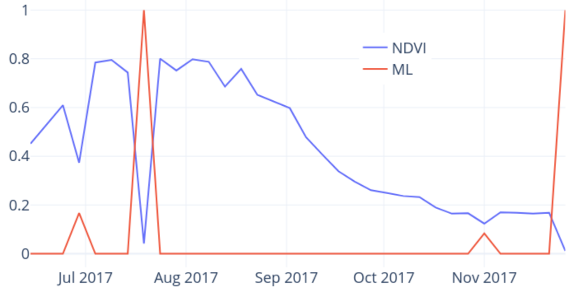

Figure 3: Parcel-1: Mean NDVI and Atmospheric Disturbance Identification by ML (over dates 14-06-2017 to 01-12-2017).

Funding

This work was supported by the NIIAA (Núcleo de Investigação em Inteligência Artificial em Agricultura) project, Alentejo 2020 program (reference ALT20-03-0247-FEDER-036981).

References

[1] Agroinsider an agricultural consulting company. `https://www.agroinsider.com/`. Accessed: 04 09 2020.

[2] Nikos Alexandratos and Jelle Bruinsma. World agriculture towards 2030/2050: the 2012 revision. 2012.

[3] Gibbons. Turning a farm art into science-an overview of precision farming. *URL: http://www. precisionfarming. com*, 2000.

[4] André Hollstein, Karl Segl, Luis Guanter, Maximilian Brell, and Marta Enesco. Ready-to-use methods for the detection of clouds, cirrus, snow, shadow, water and clear sky pixels in sentinel-2 msi images. *Remote Sensing*, 8(8):666, 2016.

[5] Oliver Kramer. Scikit-learn. In *Machine learning for evolution strategies*, pages 45–53. Springer, 2016.

[6] Deepak Kumar and Prasanta Kalita. Reducing postharvest losses during storage of grain crops to strengthen food security in developing countries. *Foods*, 6(1):8, 2017.

[7] Jérôme Louis, Vincent Debaecker, Bringfried Pflug, Magdalena Main-Knorn, Jakub Bieniarz, Uwe Mueller-Wilm, Enrico Cadau, and Ferran Gascon. Sentinel-2 sen2cor: L2a processor for users. In *Proceedings Living Planet Symposium 2016*, pages 1–8. Spacebooks Online, 2016.

[8] ARR Menon. Remote sensing application in agriculture and forestry. *T Sabu, T., Vinod. T R., M Subramonia, I., C. Bhaskaran., and B. Ambat*, pages 222–235, 2012.

[9] M Duane Nellis, Kevin P Price, and Donald Rundquist. Remote sensing of cropland agriculture. *The SAGE handbook of remote sensing*, 1:368–380, 2009.

[10] Francis J Pierce and Peter Nowak. Aspects of precision agriculture. In *Advances in agronomy*, volume 67, pages 1–85. Elsevier, 1999.

[11] JW Rouse, RH Haas, JA Schell, and DW Deering. Monitoring vegetation systems in the great plains with erts. *NASA special publication*, 351:309, 1974.

[12] Pablo J Zarco-Tejada, N Hubbard, and P Loudjani. Precision agriculture: An opportunity for eu farmers—potential support with the cap 2014-2020. *Joint Research Centre (JRC) of the European Commission*, 2014.