Funded by
the European Union

Hyperspectral Image Segmentation and Entropy Calculation for Qualitative Analysis of Grassland

Serban Oprisescu, Radu-Mihai Coliban, Mihai Ivanovici
Electronics and Computers Department,
Transilvania University of Brașov, România
{serban.oprisescu, coliban.radu, mihai.ivanovici}@unitbv.ro

This paper has been submitted to: ISSCS 2023

Abstract—Hyperspectral imaging provides high spectral resolution images of a scene in hundreds of narrow spectral bands. This technique proves to be very useful in Earth Observation applications such as land cover mapping, agriculture crop health assessment and other remote sensing tasks. The qualitative analysis of grassland areas is manually performed by computing the Shannon-Weaver biodiversity index. In this paper we propose a semi-automatic estimation of this biodiversity index in remotely-sensed hyperspectral images. Starting from a spectral reflectance curve chosen to be representative for grassland, we perform the image segmentation based on histogram thresholding of spectral angle mapper (SAM) values. We then compute the entropy for the pixels belonging to the segmented grassland areas. In order to apply the classic Shannon definition of entropy, we perform a clustering for data dimensionality reduction. The evaluation of the proposed method is performed considering a manually generated ground truth.

I. Introduction

Recent remote sensing technologies such as hyperspectral imaging are promising tools for precision agriculture. Their goal is to determine with much higher precision than multispectral imaging [1] (eg. SPOT optical images) the subtle time variations of different crops health and land cover properties like early detection of lack of water or nutrients or different diseases. This higher precision is due to the fact that the hyperspectral image (HSI) sensors offer usually hundreds of spectral bands - typically having a bandwidth below 10nm [1] - in the visible, near-infrared and short-wave infrared spectrum.

Biodiversity is an attribute of an area, referring to the variety within and among species within that area [2]. Biodiversity assessment is an important aspect of natural resource management [3], in particular in the case of grassland ecosystems, where biodiversity has a direct influence on productivity and sustainability [4].

The entropy of a system, in general, and of a signal in our particular case, is a statistical measure which depicts the randomness of the system, and it also represents a measure of diversity [5]. As mentioned in [6] the entropy is used to measure the biodiversity and its change. But measuring species diversity over large areas is a very time consuming and expensive task [6]. Therefore, there is a need for automatic entropy measurements techniques relying on remote sensing data.

The qualitative analysis of permanent grassland vegetation usually consists of determining a complete list of present species. The classical method is the double meter method which allows the analysis of plant cover in a grassland [7]. The data from a floristic relevé performed by the double meter method can be used to compute various vegetation parameters, like the pastoral value. One of the vegetation parameters of interest is the Shannon-Weaver biodiversity index [8]. Basically, the entropy is computed based on the estimated probabilities for each individual of the species present in a certain area. The working hypothesis of the current study is that the Shannon-Weaver biodiversity index can be estimated by or correlated by with the entropy computed in digital images acquired over a given area, in particular remotely-sensed satellite images.

In this paper we propose a novel hyperspectral image entropy estimation technique based on clustering dimensionality reduction scheme. We set the goal to determine the entropy of grassland regions from mountain areas. As stated before, the entropy is currently manually computed in situ, by counting the frequency of different species over a small area, to determine the biodiversity. This measure is directly correlated with the quality of the grasslands, hence the significance of its assessment. Our proposed algorithm first segments the hyperspectral image to extract the grassland regions. The spectral angle mapper (SAM) was used for segmentation. Then, on the segmented grassland hyperspectral pixels, a Fuzzy C-means (FCM) clustering is applied. By clustering, we are able to reduce the hyperspectral space dimensionality to a relatively small number of clusters, on which we can apply the Shannon entropy formula and compute the entropy.

As related work, the Shannon entropy of color, multispectral or hyperspectral images is usually computed for each spectral band [9]. In [10] an entropy assessment method for color images is defined. The authors of [5] propose an adaptation of the Shannon entropy for the hyperspectral domain with the use of the histogram of spectral differences (which are defined using the Kullback-Leibler pseudo-divergence).

There are various HSI segmentation methods presented in the literature [11] for instance deep learning approaches are widely used, support vector machines, Markov random fields etc. Some methods reduce the dimensionality and extract features by principal component analysis (PCA) or linear discriminant analysis (LDA).

The rest of the paper is structured as follows: Section II presents the proposed method, Section III the experimental results and Section IV the conclusions and further work.

II. The Proposed Method

We chose to use a hyperspectral image provided by the PRISMA mission of the Italian Space Agency [12]. Fig. 1 shows an RGB band selection [13] from a PRISMA image.

A block diagram of our approach is presented in Fig. 2. The first step in our image analysis chain is to manually choose a grassland pixel and keep its spectral reflectance curve (SRC), denoted r. Then we build a SAM value pseudo-image by scanning the hyperspectral image and computing for each pixel (having the SRC t), the value of $SAM(t,r)$ using (1). The image containing the SAM values is denoted I_{SAM}. In Fig. 3 the chosen reference grassland pixel SRC is shown with continuous line; other three similar grassland pixels are plotted.

In Fig. 4 the histogram of the I_{SAM} image is shown. One can see that the histogram can be modelled by a Gaussian mixture. The first Gaussian within the mixture (the one with the lowest mean) corresponds to the grassland class because it holds the minimum values of SAM. To automatize the segmentation algorithm, we applied the Expectation-Maximization (EM) algorithm [14] on the I_{SAM} image by imposing a number of four components. Then we used the mean value μ_1 of the first Gaussian mixture component as threshold for the segmentation of the I_{SAM} image.

The proposed segmentation algorithm is summarized using the pseudocode in Algorithm 1.

Algorithm 1 The proposed image segmentation method

Input: the hyperspectral image I
Output: S (the segmented image)
manually choose a known grassland pixel
denote r the SRC of this pixel
for each pixel (*i*,*j*) of *I*
    denote t the SRC of the (i,j) pixel
    compute $I_{SAM}(i,j)$ = the SAM value between r and t
apply the EM algorithm on I_{SAM} to extract the first Gaussian mixture mean μ_1
$S \leftarrow$ thresholding segmentation of I_{SAM} using μ_1 threshold
compute border error (BE) rand index (RI) and Dice coef.

Figure 1. A full-size RGB band selection from a PRISMA image.

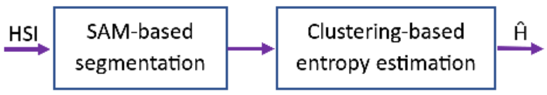

Figure 2. Block diagram of our approach.

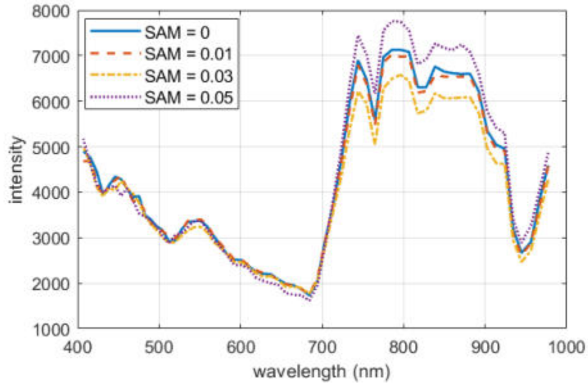

Figure 3. The reference grassland pixel SRC and other three similar SRCs.

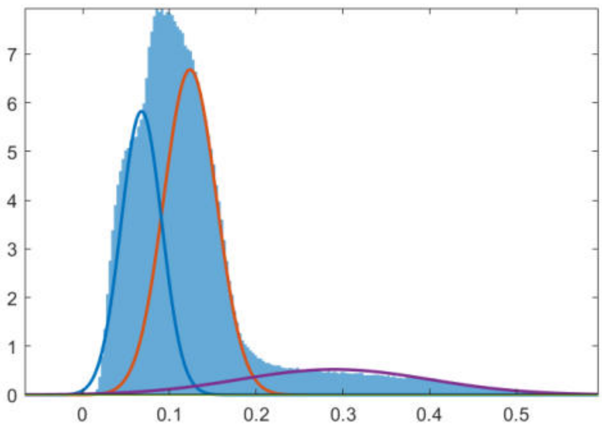

Figure 4. Histogram of the I_{SAM} image, with a Gaussian mixture model overlayed.

$$SAM(t,r) = cos^{-1}\left(\frac{\sum_{i=1}^{N} t_i r_i}{\sqrt{\sum_{i=1}^{N} t_i^2}\sqrt{\sum_{i=1}^{N} r_i^2}}\right) \tag{1}$$

As mentioned in the introduction, the quality of grassland areas can be measured by the entropy. Usually, agricultural research institutes manually measure the entropy by counting the number of different plant species on a land area. In this paper we propose a method to assess the entropy using remotely-sensed hyperspectral data. The Shannon entropy is given by (2).

$$\text{H} = - \Sigma p(i)\log(p(i)) \tag{2}$$

We cannot apply directly (2) on the hyperspectral values due to the high dimensionality of the corresponding hyperspace. Thus, we need a dimensionality reduction technique. We propose here to reduce the dimensionality by applying a clustering technique. For this, we selected the Fuzzy C-means (FCM) clustering technique. We applied FCM on all the previously segmented grassland hyperspectral pixels, by imposing a number of *N* clusters. Then, for each pixel we selected the most probable class (the maximum value). The probability $p(i)$ from (2) was computed for each cluster (i = 1,..,N) as a frequency, i.e. the number of pixels within the class i divided by the total number of grassland pixels. See Algorithm 2.

Algorithm 2 The calculation of entropy

Input: the hyperspectral image I and the segmented image S
Output: the entropy H
for each pixel (*i,j*) of *S*
    if S(i,j) == 1
    add the SRC of $I(i,j)$ to a matrix SD
apply FCM clustering on SD, with N classes
for *i* = 1 to *N*
    $P(i)$ = (no. of pixels from class i) / (total no. of pixels)
    $H = H + P(i)$*log($P(i)$)
$H \leftarrow - H$

III. Experimental Results

For this study we used hyperspectral images provided by the PRISMA satellite, containing 66 VNIR (Visible and Near-InfraRed) bands (from 400 to 1010 nm), 173 SWIR (Short-Wave InfraRed) bands (from 920 to 2500 nm) and a medium resolution panchromatic image. More precisely we worked on the VNIR bands, each band having a resolution of 1000 x 1000 pixels. The area on the ground covered by an image is around 30 x 30 km. To reduce the computational cost and because we had to use a precise ground truth image, we selected a crop of 300 x 300 pixels; in Fig. 5 we show an example of such crop in its RGB colored version obtained using the visualization method from [15] which uses a neural network trained on the CAVE dataset [16]. In Fig. 6 the corresponding manually generated ground truth (GT) image is presented. The black regions correspond to grasslands and the white ones to other land cover types.

After manually choosing a grassland pixel and then applying $SAM(t,r)$ on all image pixels (1), the SAM value image from Fig. 7 is obtained. Darker areas within this image correspond to a higher similarity degree with the reference grassland hyperspectral pixel, and lighter areas (having greater SAM values) correspond to other land coverings (forest, different crops, water, roads, houses etc.). After running the EM algorithm and choosing the threshold value (0.05 in this case), the final segmented image was obtained, like in Fig. 8.

To assess the accuracy of the segmentation we computed three quality measures: border error (BE), rand index (RI or accuracy) and Dice coefficient. The BE [17] was calculated using (3), where A represents grassland regions obtained after the segmentation and M is the manually segmented region (from the ground truth).

$$BE = Area((A \cup M) - Area(A \cap M))/Area(M) \tag{3}$$

For the image in Fig. 8, these values are: BE = 0.78, RI = 0.95 and Dice = 0.64.

Afterwards, we applied the proposed hyperspectral entropy estimation algorithm on the segmented grassland pixels, as described in the previous section. By clustering or vector quantization techniques, the computation of entropy leads to an underestimation of the complexity. We varied the number of clusters, in order to investigate the variation of the resulting entropy value as a function of the number of clusters. Table 1 presents the obtained entropy values and in Fig. 9 a graphic representation is shown. As expected, the entropy increases asymptotically with the number of clusters, to a maximum value of approximately 5.9.

The limitation to a maximum number of 85 due to Matlab capabilities clearly shows that a more efficient algorithm should be designed and, in the same time, to ensure the correct estimation of the entropy value. The previous segmentation quality metrics and the entropy value are given for the image crop in Fig. 5, but one can apply the proposed algorithm on the entire image (Fig. 1) provided the existence of the corresponding GT image.

Figure 5. A PRISMA image RGB crop.

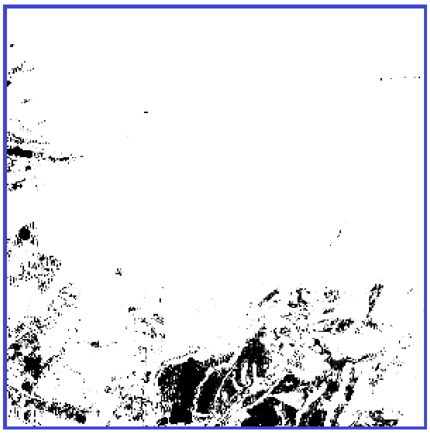

Figure 6. Ground truth image for the crop from Fig. 5

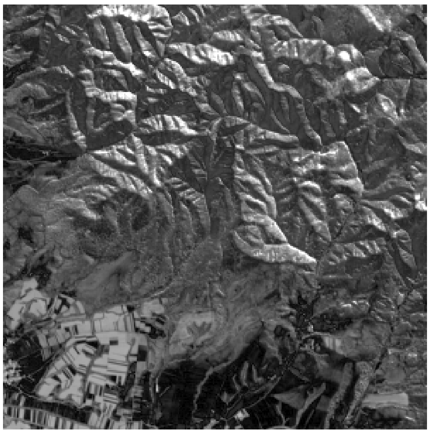

Figure 7. The computed SAM values pseudo-image.

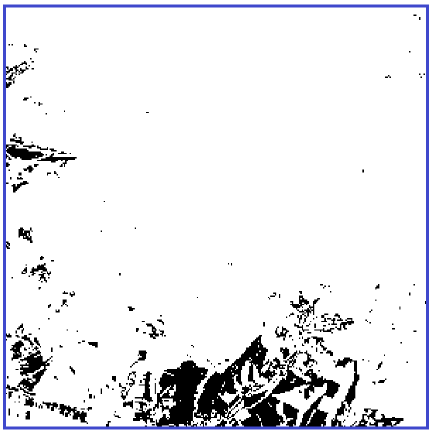

Figure 8. The final image segmentation result

TABLE I. THE ENTROPY FUNCTION OF NO. OF CLUSTERS

No. of clusters	25	35	45	50	55	65	75	85
Entropy	4.55	4.94	5.23	5.33	5.43	5.72	5.87	5.95

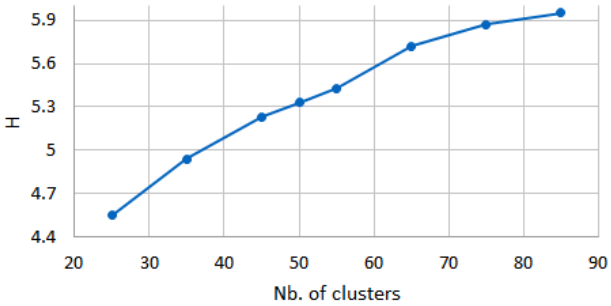

Figure 9. The entropy function of no. of clusters.

IV. Conclusions and Further Work

We proposed a hyperspectral image entropy assessment algorithm which was used to compute the entropy of grassland regions. This statistical measure is very useful in practice to assess the quality of grasslands. The first step was to segment the grassland areas from the HSI, and for this the spectral angle mapper was used in a semi-supervised manner, by manually choosing a representative grassland pixel. The quality of the segmentation was assessed using the border error, rand index and Dice coefficient. To reduce the HSI space dimensionality we applied Fuzzy C-means clustering by imposing a relatively small number of clusters. This allowed us to directly apply the Shannon entropy formula on the obtained clusters.

For the practical computation of the entropy, a compromise should be made regarding the choice of the number of clusters: higher the number, higher the accuracy in estimation. However, a large number of clusters results in high computing demands and lower running time. As further work we will propose a new definition and computation of entropy for hyperspectral images and refinement of the segmentation approach by using artificial intelligence paradigms.

Acknowledgment

This work was funded by the European Union. The AI4AGRI project entitled “Romanian Excellence Center on Artificial Intelligence on Earth Observation Data for Agriculture” received funding from the European Union’s Horizon Europe research and innovation program under grant agreement no. 101079136.

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

References

[1] Lu, B.; Dao, P.D.; Liu, J.; He, Y.; Shang, J. Recent Advances of Hyperspectral Imaging Technology and Applications in Agriculture. Remote Sens. 2020, 12, 2659.

[2] DeLong Jr, D. C. (1996). Defining biodiversity. Wildlife society bulletin, 738-749

[3] Vermeulen, S., & Koziell, I. (2002). Integrating global and local values: a review of biodiversity assessment. International Institute for Environment and Development, Biodiversity and Livelihoods Group

[4] Tilman, D., Wedin, D., & Knops, J. (1996). Productivity and sustainability influenced by biodiversity in grassland ecosystems. Nature, 379, 718-720, doi: 10.1038/379718a0

[5] Y. -J. Chen, N. Richard, H. Deborah, A. Tournié, A. Michelin and C. Andraud, "Statistics of hyperspectral data/image analysis: Entropy," 2018 Colour and Visual Comp. Symp. (CVCS), Gjovik, Norway, 2018, pp. 1-6, doi: 10.1109/CVCS.2018.8496573.

[6] Thouverai, E., Marcantonio, M., Bacaro, G. et al. Measuring diversity from space: a global view of the free and open source rasterdiv R package under a coding perspective. COMMUNITY ECOLOGY 22, 1–11 (2021). https://doi.org/10.1007/s42974-021-00042-x

[7] I. Samfira, C. A. Popescu, Sustainable Management of Grasslands: A Guide for the Maintenance of Ecosystem Services, Mirton Publishing House, 2021 Bean.

[8] Ortiz-Burgos, S. (2016). Shannon-Weaver Diversity Index. In: Kennish, M.J. Encyclopedia of Estuaries. Encyclopedia of Earth Sciences Series. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-8801-4_233

[9] D. Long and V. P. Singh, "An Entropy-Based Multispectral Image Classification Algorithm," in IEEE Trans. on Geos. and Rem. Sen., vol. 51, no. 12, pp. 5225-5238, Dec. 2013, doi: 10.1109/TGRS.2013.2272560.

[10] M. Ivanovici, N. Richard, “Entropy versus fractal complexity for computer-generated colour fractal images”, 4th CIE Expert Symp. on Colour and Vis. Appearance, 6-7 Sept. 2016, Prague, Czech Republic

[11] Liu L, Awwad EM, Ali YA, Al-Razgan M, Maarouf A, Abualigah L, Hoshyar AN. Multi-Dataset Hyper-CNN for Hyperspectral Image Segmentation of Remote Sensing Images. Processes. 2023; 11(2):435. https://doi.org/10.3390/pr11020435

[12] www.asi.it/en/earth-science/prisma/

[13] Su, H., Du, Q., & Du, P. Hyperspectral imagery visualization using band selection. In 2012 4th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2012, pp. 1-4, doi: 10.1109/WHISPERS.2012.6874236

[14] Dempster, A.P.; Laird, N.M.; Rubin, D.B. (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm". Journal of the Royal Statistical Society, Series B. 39 (1): 1–38

[15] I. Plajer, A. Baicoianu, L. Majercsik, "AI-based Visualization of Remotely-Sensed Spectral Images", submitted to ISSCS 2023, Iasi, Romania

[16] F. Yasuma, T. Mitsunaga, D. Iso, and S.K. Nayar, Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum, Tech. Rep., Dep. of Comp. Sci., Columbia Univ., 2008

[17] Zambanini, S.; Sablatnig, R.; Maier, H.; Langs, G. Automatic image-based assessment of lesion development during heman-gioma follow-up examinations. Artif. Intell. Med. 2010, 50, 83–94, doi: 10.1016/j.artmed.2010.06.003.