**Application of SPFCM (Superpixel Fuzzy C means) algorithm on Plant and Residue Cover Images**

**By**

**Timotius Andrean Patrick Lagaunne**

**1. INTRODUCTION**

Management of cover in agricultural fields is crucial for conservation management practices. Having residue cover especially, it helps decreasing soil erosion, increased soil organic matter and it can improved soil and water quality and reduced carbon dioxide CO2 emissions [7]. For that purpose, estimating the residue cover in a field is important. The standard technique used by USDA to estimate the residue cover is the line-point transect where a 15-30 m line with 100 evenly spaced markers intersecting plant and residue are counted. Though this technique is powerful, this technique is impractical for monitoring crop residues cover in many fields over broad areas in a timely manner. Hence, automation of residue cover estimation can greatly impact the management practice. In this paper, I investigate the capability of unsupervised machine learning algorithm specifically SPFCM (Super Pixel Fuzzy C-means) to estimate the plant and residue cover from images automatically. FCM (Fuzzy C- means clustering algorithm) by itself is a form of clustering in which each [data point](https://en.wikipedia.org/wiki/Data_point) can belong to more than one cluster. Incorporating super pixel to the clustering can incorporate spatial/semantic/ contextual information into accounts which is very intuitive since, human differentiate different class of object based on the contextual/spatial/ semantic information. At the same time, using super pixel can also significantly improve the post processing performance. [3]

**2. BACKGROUND**

2.1 Line Transect Method for Cover Residue Estimation

Line transect surveys are widely used to estimate the density and/or size of wild animal populations. As it is stated earlier in the introduction, USDA generally use 15-30 m line with 100 evenly spaced markers. Then the plant and residue that intersect with the markers are counted. As a result the percent cover would be the number of intersected plant and residue divided by 100. This method has some limitations that normally people encounter. One of the limitation is that, the method assume that the residue and plant are random and well distributed spatially, which usually not the case. Consequently, often this method can overestimate or underestimate the actual percent cover. Other limitation would be, the process of having a line and having someone count the intersecting marker and plant/residue is laborious which is why it is impractical to use this method for monitoring crop residues cover in many fields. [7]

2.2 Satellite multispectral and hyperspectral data to calculate crop and plant residue cover

Effort has been done in evaluating plant and residue cover by using remote sensing. Currently, there are several spectral indices that has been developed to calculate crop residue. There is CAI, NDTI, NDI5 and some others. [1] These spectral indices can only be acquired though multispectral and hyperspectral camera.

* 1. Rise of Drone in agriculture

Today, drones are used for a staggeringly diverse number of applications, with every indication that they’ll become more ubiquitous and capable as time goes on. Specifically, drone has been very widespread for agriculture application. Though some limitation of drone surely still exists that hinder the full utilization, it has grown so much especially as farmer start realizing the benefit of it. As an example of application, to identify non healthy plant, the use of NDVI that requires near infra-red band as addition to the visible light band (RGB bands). This need of Near-infra red can be a hindrance for farmer because having near infra-red can cost a lot more than normal bands especially with the same image resolution. Moreover, specifically as mentioned previously with the use of satellite multispectral and hyperspectral data, to discriminate residue cover with soil, hyperspectral camera are needed which is very costly. As an addition, high resolution can also be needed to get more precision, causing the cost to be higher. As a consequence, being able to utilize the power without the need of hyperspectral camera and high resolution camera can be really attractive.

* 1. Plant vs Soil and Residue

Plant generally are green because of the chlorophyll, so different with the soil. With our bare eye, we generally can tell quickly with our eyes the differences between plant vs soil and residue. Though there are diverse range of soil and residue, the residue and the soil are generally not green. Following that, in perspective of image segmentation and image processing, segmenting green vs soil and residue which are generally white, black, brown or red is very feasible. The use of special indices like exG index, which is basically extra green index, Green Red index, CIVE index or even the conversion of rgb to different color space like LAB or HSV can make it easier. [5]

* 1. Soil vs Residue

Different with plant vs Soil or Residue, differentiating soil and residue is very hard using visible light bands (RGB) only. Though residue can be, in general, brighter than soil, Moisture of the residue and soil, shadow, light exposure can tremendously impact that. As a result, texture and shape feature are needed to be used to discriminate both of them. [7]

2.6 FCM

FCM algorithms is a clustering algorithm that cluster data by calculating a measure of membership, called the fuzzy membership. The membership is bounded to be between zero and one, it reflects the degree of similarity between data and each centroid. FCM is devised as the minimization of the following objective function

Where C is the predefined number of clusters, and N is the number of pixels in the given image. Uij is the membership of pixel xj to the ith clusters such that vi is the ith cluster centroid, and ||·|| is a norm metric, denoting Euclidean distance between pixels and clustering centroids. The parameter m is a weighting exponent on each fuzzy membership and determines the amount of fuzziness of the resulting classification. [3]

2.5 Modified FCM algorithm

Ahmed, proposed a modification to formula 1 by introducing a term that allows the labeling of pixel to be influenced by the labels in its immediate neighborhood. The modified objective function was given by

Where Nj stands for the set of neighbor that exists in a window around xj and NR is the cardinality of Nj. The parameter α control the effects of the neighbor item. This Modified FCM method is more insensitive to noise, but suffers from computational complexity. [4]

* 1. SLIC super pixels

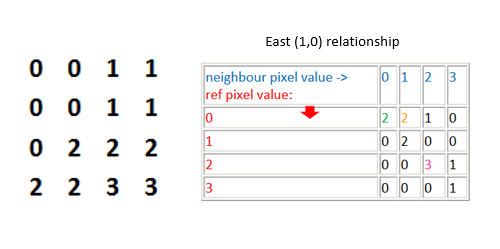
The SLIC superpixel method clusters pixels based on their color similarity and proximity in the image plane. For color images in the CIELAB color space, two feature Ci = [Li, ai, bi]T and Si=[xi,yi]T, are defined to represent the color values and 2D positions of the ith pixel. First, K initial cluster centers are sampled uniformly in the image. To avoid centering a superpixel on an edge or a noisy pixel, cluster centers are moved to the locations corresponding to the lowest gradient magnitude in a 3×3 neighborhood. Next, each pixel is labeled by the nearest cluster center based on the distance D as

where j is the index of the cluster center, and Nc and Ns are the normalization constants of the color and spatial distances. After the clustering, the cluster centers are updated as the mean of the vector so fall the pixels in the cluster as

where Gj represents the cluster centered at ϕj and N is the number of pixels in Gj. The clustering and updating processes are repeated until a predefined number of iteration is achieved. The SLIC algorithm can generate compact and nearly uniform superpixels with a low computational overhead. [4]

* 1. GLCM

GLCM is a tabulation of how often different combinations of pixel brightness values (grey levels) occur in an image. In this case, I will use GLCM to get all the combination of the superpixels in an image to get all the connection and the adjacent superpixels of each particular superpixels.



**Fig. 1** Example of GLCM calculation for East (1,0) relationship

**3. SPFCM**

Normal FCM algorithms perform clustering on image pixels, while SPFCM algorithm perform clustering on image superpixels with this propose modified objective function considering the attributes of superpixels and its immediate neighborhood.

Where Q is the number of superpixels in images, ϒr is the number of pixels in superpixels sj, and ξr is the average feature values of superpixels sj. Uij denoted the membership of superpixel sj to the ith cluster. Nj stands for the the set of the neighboring superpixels that are adjacent to sj, and NR is the cardinality of Nj.

The objective function J then can be minimized in a similar fashion to the standard FCM algorithm. Based on Lagrange multiplier method, the following equation will be constructed

Taking the derivative of F with respect to uij and vi and setting the result to zero, we can get

and

The SPFCM algorithm can be summarized in the following steps.

**Algorithm SPFCM**

1: Generate superpixels representation of original image, and collect necessary information of superpixels.

2: Find the adjacent/connection of each superpixels.

3: Initialize cluster centroid vi, i=1,…,C.

4: Update the membership uij

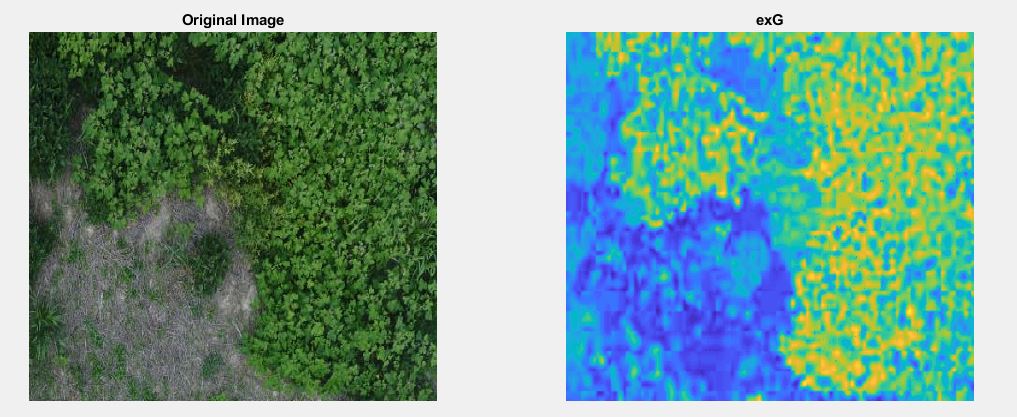
5: Update the cluster centroid vi by equation

6: Repeat Steps 3-4, until

**4. EXPERIMENT**

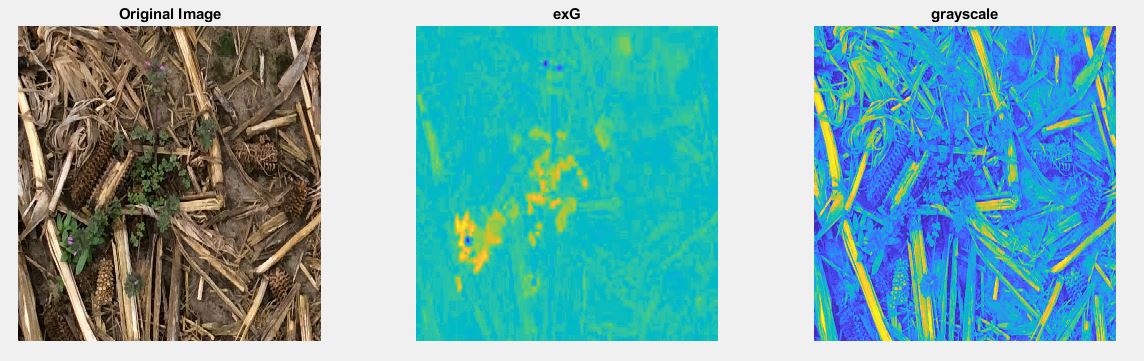
* 1. Feature Extraction

There are two images specifically used for the test of the algorithm. One image has soil and plant and the other has soil, residue and plant. In the first image, green color is extracted with exG index to be able to segment the plant from the soil [5].



**Fig. 2** Image 1 and exG index

In the second image, as seen below, it has soil, plant and residue. Similarly with the first image, exG index will be extracted and grayscale is also extracted since residue is brighter than the soil and plant.

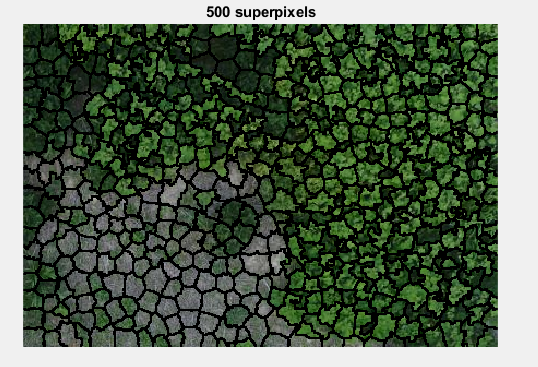


**Fig. 3** Image 2, exG index and grayscale index

After, the feature get extracted, the feature is normalized to 0 as the minimum and 1 as the maximum.

* 1. Superpixel Extraction

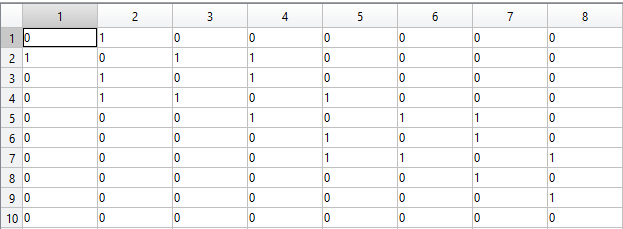
The superpixels, then, is extracted with SLIC algorithms. Compared to the explanation above, the matlab version of the SLIC has compactness and estimated superpixels number Instead of having Nc and Ns. Below are the examples of the superpixels extraction with estimated 500 superpixels.



**Fig. 4** Image 1 and 2 with 500 superpixels using SLIC algorithm

* 1. Superpixel Connection

To get the adjacent superpixel for each superpixel, in this experiment I used GLCM with offset 1 for north, south, east and west. All four GLCM then summed together to get the whole superpixels connection. If the entry aij of the matrix is not zero, it means superpixel i is connected to the superpixel j, else it is not connected.

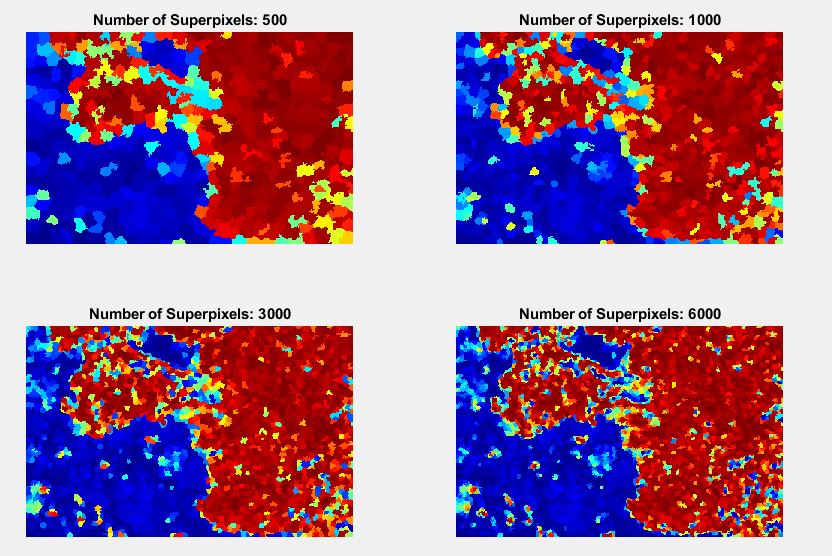


**Fig. 5** Example of final connection matrix from GLCM

* 1. Varying Parameter

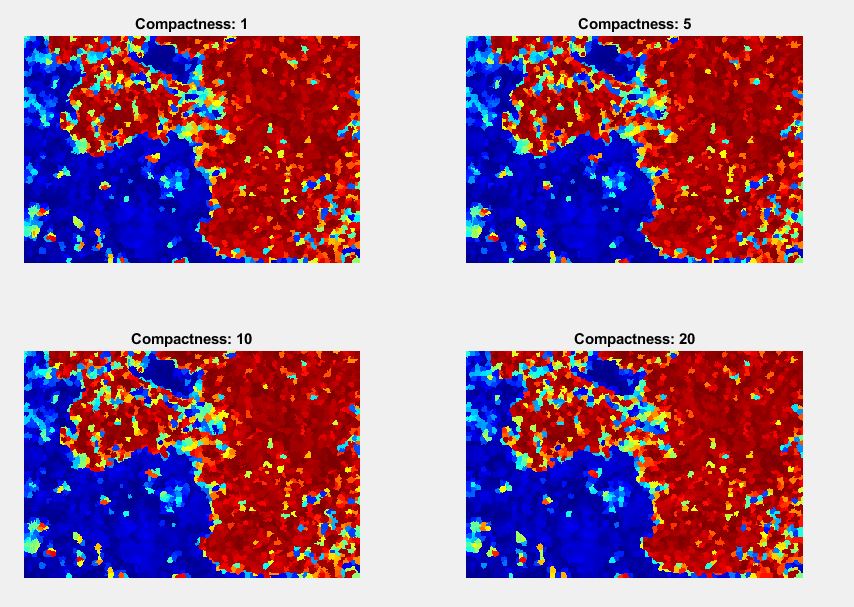
After, feature, superpixels and its connection has been extracted, all data with some parameters can be inputted to the SPFCM algorithm. In the experiment, I vary the parameters to see how the parameters affected the result. Specifically, I only vary the parameters for image1 where there is only plant and soil. As mentioned above, exG is the only feature in the data.

1. *Varying Number of superpixels*

**

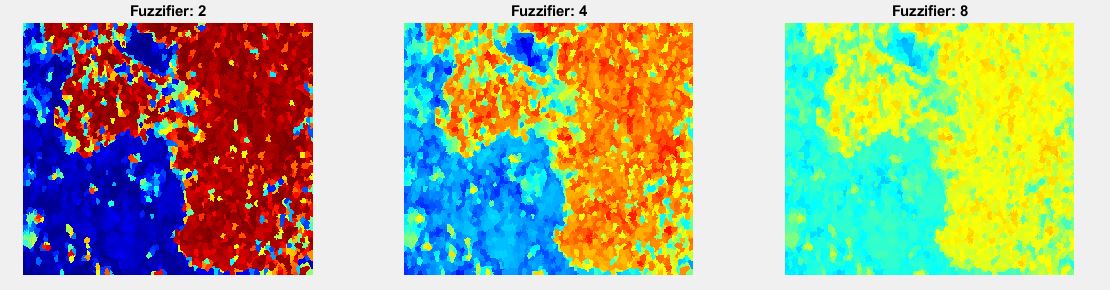
**Fig. 6** Superpixels Variations (500, 1000, 3000, 6000)

*b. Varying compactness*

**

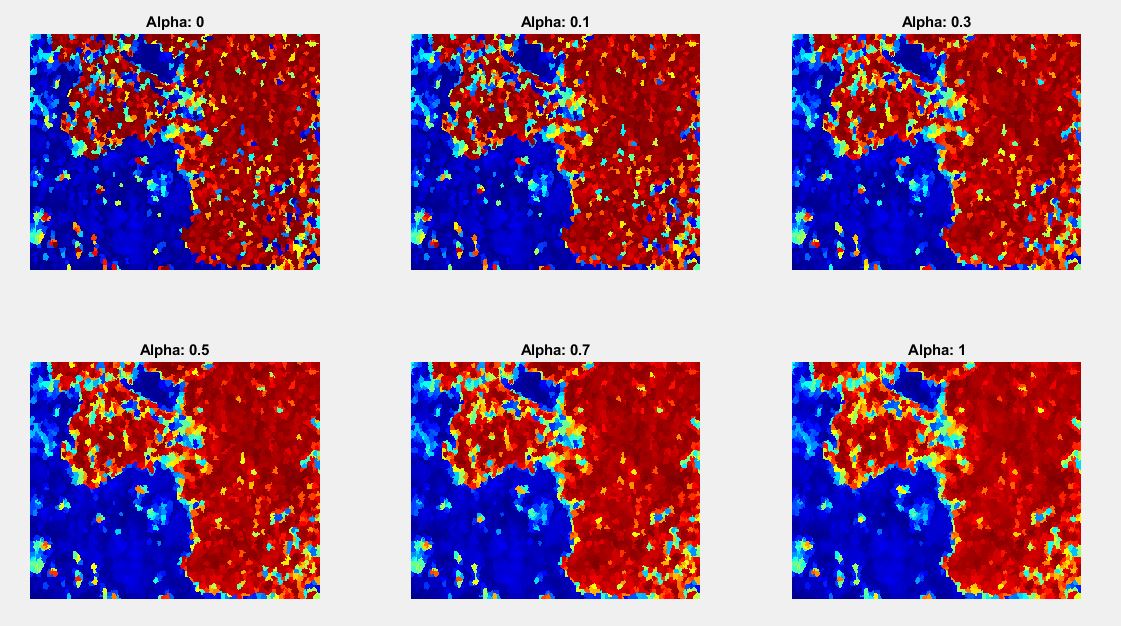
**Fig. 7** Compactness Variations (1, 5, 10, 20)

*c. Varying Fuzzifier*

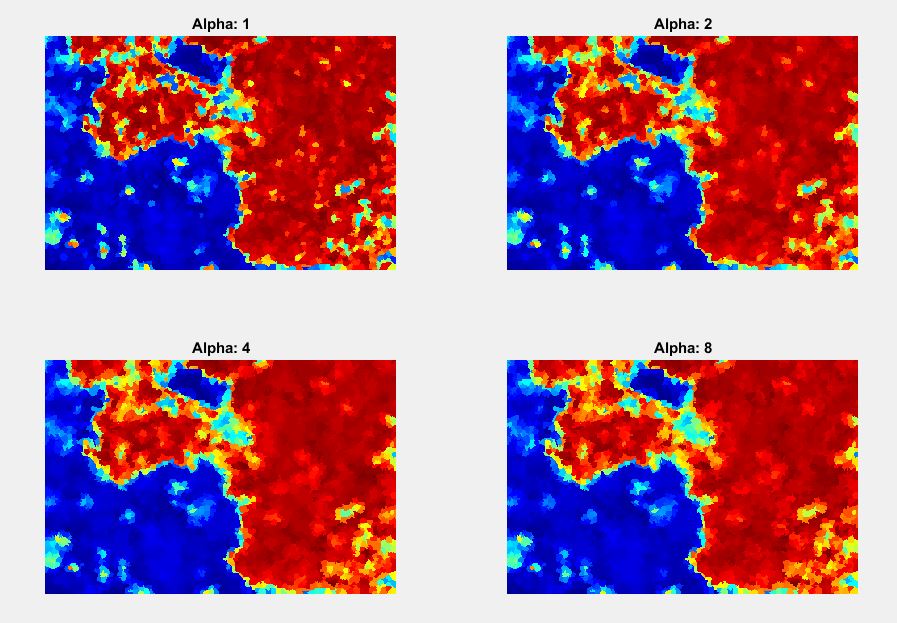
**

**Fig. 8** Fuzzifier Variation (2,4,8)

*d. Varying Alpha*

**

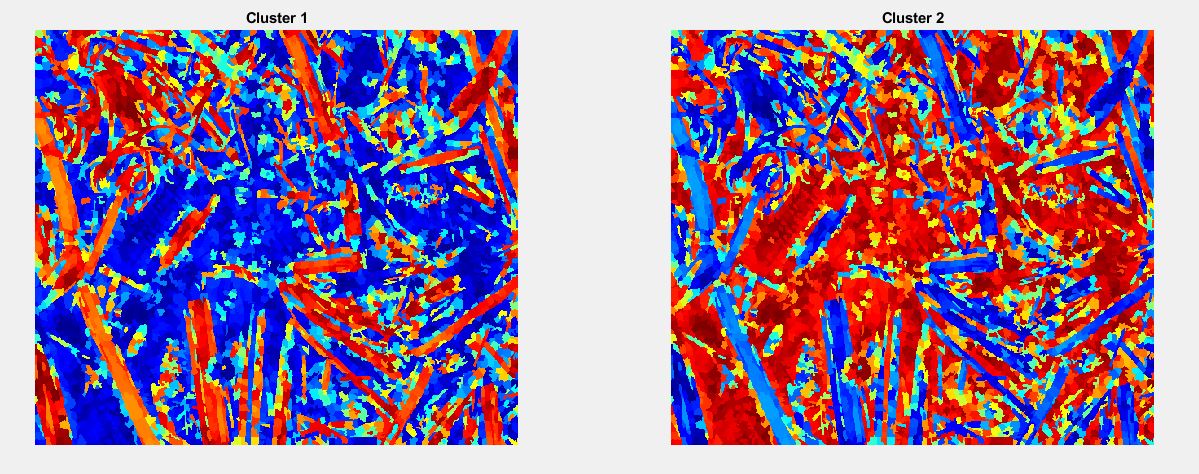
**Fig. 9** Alpha Variation 1 (0, 0.1, 0.3, 0.5, 0.7, 1)

**

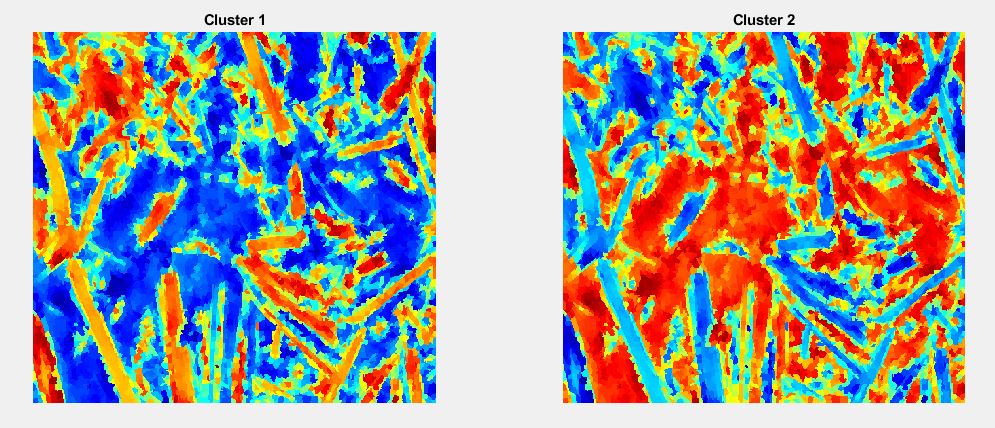
**Fig. 10** Alpha Variation 2 (1, 2, 4, 8)

* 1. Experiment with image 2 (Plant Cover, Residue Cover, Soil)

1. Residue vs Soil and Plant

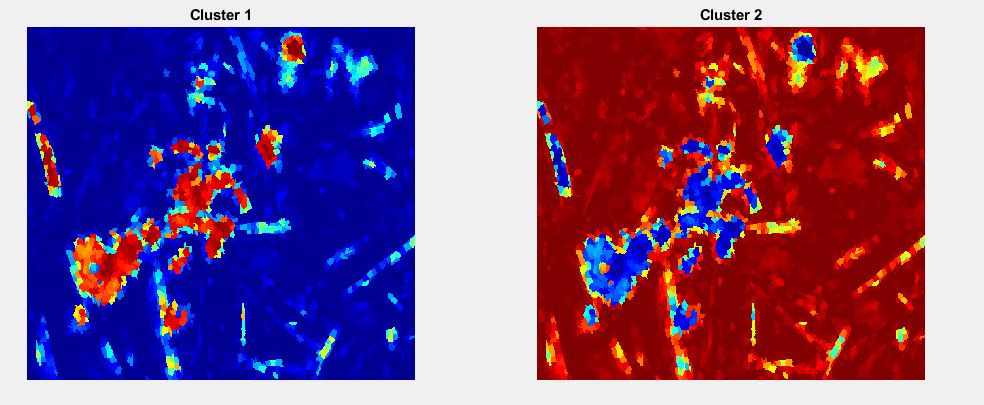


**Fig. 11** Residue Cover vs Not Residue Cover (Plant Cover and Soil), *Notes: Feature used was grayscale index only*



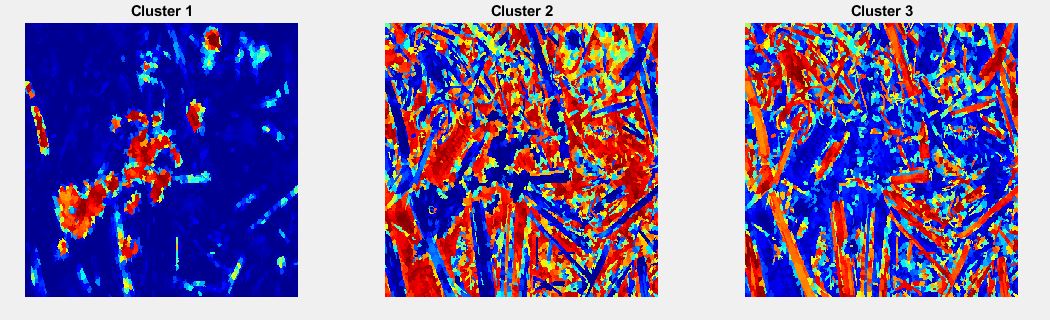
**Fig. 12** Residue Cover vs Not Residue Cover (Plant Cover and Soil), *Notes: Feature used was grayscale index only with α=2*

1. Plant vs Soil and Residue



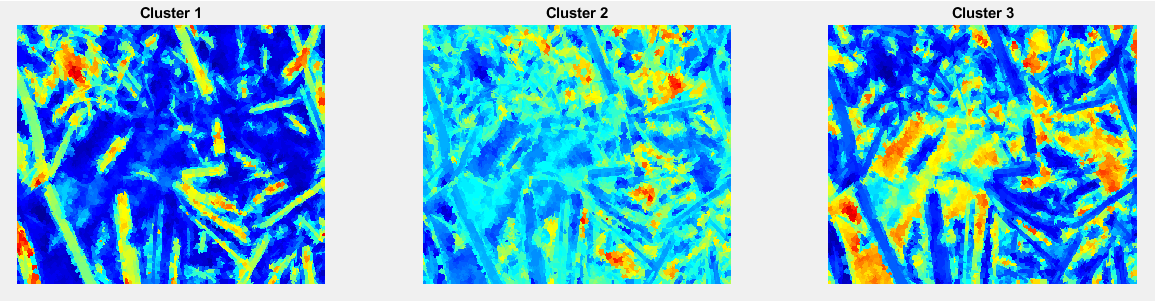
**Fig. 13** Plant Cover vs Not Plant Cover (Residue Cover and Soil), *Notes: Feature used was exG index only with α=0.3*

1. Plant vs Soil vs Residue (Combining Together)



**Fig. 14** Plant Cover vs Soil vs Residue Cover, *Notes: no SPFCM used, Result is formulated by combining the previous two result (Fig 11 and Fig 13), Soil (Cluster 2) is generated by subtracting Soil and Plant (Cluster 2) membership from Fig 11 with Plant (Cluster 1) from Fig 13. Any negative value from the subtraction is set to zero*

1. Plant vs Soil vs Residue (exG and grayscale inputted to the SPFCM)

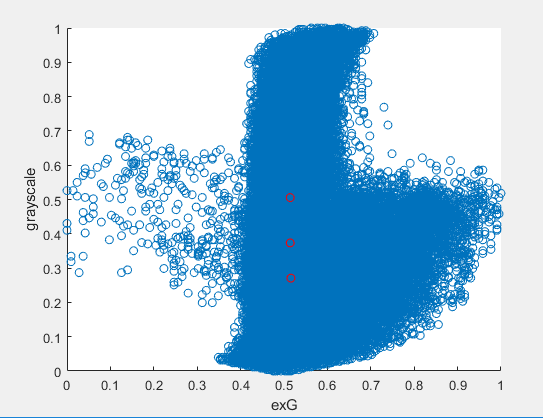


**Fig. 15** Residue Cover vs Soil vs Plant Cover, *Notes: Bad result is explained in the limitation*

**5. LIMITATION**

As I go through all the parameters and debugging the code, these are some limitation of the code I have found.

1. I haven’t optimize it fully and the code can be really slow as I have more features, clusters, superpixels.
2. Some results, especially the result of clustering/ segmenting with 3 clusters using 2 feature (grayscale and exG) does not give good result because the features are not good enough (not clustered together)



**Fig. 16** normalized exG and Normalized grayscale plotted in scatter plotwith blue color,*Notes: Red color represent the cluster center from clustering exG and grayscale together with SPCFM*

1. GLCM requires a lot of memory with a lot of superpixels.

**6. CONCLUSION**

This experiment though is done in pursue of getting good way to segment residue cover, plant cover and soil in rgb spaces, has a lot of limitations and problems. Several shape or texture feature could have been part of the experiment, but cannot be done because of the time limitation. As a result, this experiment is only done in order to just implement the algorithm and to explore the capability of the algorithm. Though my implementation in matlab is kind of slow but it can be optimized and has the potential to be faster than normal fcm because of the use of superpixels. It’s also incorporate the spatial information by incorporating neighbor superpixel information that is better compare to fcm when context and spatial information matter in the image.

**7. REFERENCES**

[1] C. S. T. Daughtry, P. C. Doraiswamy, E. R. Hunt Jr., A. J. Stern, J. E. McMurtrey, and J. H. Prueger, “Remote sensing of crop residue cover and soil tillage intensity,” Soil and Tillage Research, vol. 91, pp. 101–108, 2006.

[2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, “Slic superpixels compared to state-of-the-art superpixel methods,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 34, no. 11, pp. 2274–2282, 2012

[3] Jia, S., Zhang, C.: Fast and robust image segmentation using an superpixel based FCM algorithm. In: 2014 IEEE International Conference on Image Processing (ICIP). IEEE (2014)

[4] M. N. Ahmed, S. M. Yamany, N. Mohamed, A. A. Farag, and T. Moriarty, “A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data”, IEEE Transactions on Medical Imaging, 21(3), pp. 193-199, 2002.

[5] Meyer, G. E. and Neto, J. C., Verification of color vegetation indices for automated crop imaging applications, Computers and Electronics in Agriculture 63(2):282– 293 (2008).

[6] T. Adao, J. Hruška, L. Pádua , J Bessa, E Peres, R. Morais, J.J Sousa. Hyperspectral Imaging: A Review on UAV-Based Sensors, Data Processing and Applications for Agriculture and Forestry. Remote Sens. Agric. Veg. 2017, 9, 1110

[7] Beeson P. C,Daughtry C. S. T,Hunt E R, etal．Multispectral Satellite Mapping of Crop Residue Cover and Tillage Intensity in Iowa[J]．Journal of Soil and Water Conservation, 2016, 7 1 (5)