FUZZY C-MEANS FOR DEFORESTATION IDENTIFICATION BASED ON REMOTE SENSING IMAGE

Setia Darmawan Afandi1, Yeni Herdiyeni1, Lilik B Prasetyo2
1 Department of Computer Science, Fakulty of Mathematics and Natural Science, Bogor Agricultural University, West Java, Indonesia
2 Departement of Conservation of Forest Resources and Ecotourism, Faculty of Forestry, Bogor Agricultural University, West Java, Indonesia
Email: sastra.afandi@gmail.com, yeniherdiyeni@gmail.com

Abstract—This research report about Fuzzy C-Means for Deforestation Identification Based On Remote Sensing Image. Deforestation means that changes forest area into another functions. Clustering is a method of classify objects into related groups (clusters). While, Fuzzy C-Means clustering is a technique that each data is determined by the degree of membership. In this research, the data used are MODIS EVI 250 m in 2000 and 2012 to identify deforestation rate in Java island. MODIS EVI is one of kind MODIS image which is able to detect vegetation based on photosynthesis rate and vegetation density. The number of clusters used were 13 clusters. This research had succeeded to classify areas based on the value of EVI like areas who had a high EVI values (forests, plantations, grass land), moderate values (agricultural area), and low values (build up area, mining area, pond, and other land cover). But, EVI value is only influenced by photosynthesis rate and vegetation density. Thus, EVI value is not well to identify forest areas. this is because the value of EVI in forest areas are almost same with plantations, savanna, etc.

I. INTRODUCTION

Indonesia is one of many countries that has the greatest potential forest in the world. According to Indonesian forestry ministry in 2011, Indonesia has 98.8 million hectares forest area. But on the other hand, the facts show Indonesian forests are very alarming shrinkage. It is caused by changing forest areas into another functions, like industrials, residences, plantation, mining area etc. it is also called deforestation. Deforestation rate period 1985 - 1997 recorded 1.6 million hectares per year. Meanwhile, in 1997 - 2000 recorded 3.8 million hectares per year. The details of the rate of deforestation in Indonesia can be seen in Figure 1.

Under these conditions, forest monitoring is very important in forest management and determine how severe deforestation rate in Indonesia. Forest monitoring process has many methods. Forest monitoring using field measurement system spent more time and more cost. So, other methods are needed that are faster, cheaper and more efficient.

Development of a system which can detect deforestation rate using remote sensing is one method to monitoring forest that can provide information quickly and accurately. Moderate Resolution Imaging Spectroradiometer Enhanced Vegetation Index (MODIS EVI) is used in this research. Why MODIS data used in this research?. Because MODIS data has a good spectral resolution and consists of 36 channels, having 3 temporal resolution, that are 250 m, 500 m and 1000 m and MODIS data is available every day, so we can develop a real time system for the future.

Some related research such as research which conducted by Setiawan. In that research, the authors performed classifying existing vegetations in the area of the Java island in 2000-2009 based on spectral differences reflected from every object in the image by using the k-means clustering with accuracy rate of 80.16% [1]. That research was further developed in 2013 to identify rate of change in forest cover which includes deforestation and reforestation in java island in 2000-2009 with the accuracy of 50.34% [2]. While in this research, focused to analyzes deforestation rate in the java island in 2000 and 2012 by using Fuzzy C-Means clustering method.

II. REMOTE SENSING AND MODIS EVI

Remote sensing is a technique for obtaining information about an object, area, or phenomenon through the analysis of data obtained from devices that are not directly in contact with the object, area, or phenomenon being observed [3]. Moderate Imaging Spectroradiometer (MODIS) is one type of data from satellite images. MODIS data have real visibility for 2330 km and available every 1 or 2 days. Data MODIS has 36 spectral bands and has 3 spatial
resolution, those are 250 m, 500 m, and 1000 m. With these specifications, the MODIS data are able to provide much information about the state of the earth's surface, ocean, and atmosphere [4].

Vegetation indices is the result obtained from some combination band (wavelength) of image and has a relationship to the characteristics of pixels that contained in an image. Vegetation indices derived from wave energy emitted by vegetation in remote sensing image to show the size and amount of plant life. EVI is one of the transformations obtained from the comparison of reflectance red channel, blue channel, and near infrared sensor MODIS [3]. The formula of the EVI is

$$EVI = G \frac{\rho_{\text{nir}} - \rho_{\text{red}}^*}{\rho_{\text{nir}} + C_1 \rho_{\text{red}}^* - C_2 \rho_{\text{blue}}^* + L} (1 + L)$$

(1)

where, $\rho_{\text{nir}}, \rho_{\text{red}}^*$ and $\rho_{\text{blue}}^*$ are the remote sensing reflectances in the near-infrared, red and blue, respectively, $L$ is a soil adjustment factor and $C_1$ and $C_2$ describe the use of the blue band in correction of the red band for atmospheric aerosol scattering. The coefficients, $C_1$, $C_2$ and $L$, are empirically determined as 6.0, 7.5 and 1.0, respectively. $G$ is a gain factor set to 2.5 [6].

EVI value is strongly influenced by the rate of photosynthesis and the density of vegetation in an area. This is because, the vegetation will reflect near infrared and green waves and absorbs red and blue wave, if we see from the formula of EVI that has been described previously. If the vegetation in an area are very high, $\rho_{\text{nir}}$ will has high value. Meanwhile $\rho_{\text{red}}^*$ and $\rho_{\text{blue}}^*$ will have low value. So, it will make EVI value will be very high.

III. FUZZY C-MEANS (FCM)

The Fuzzy C-Means (FCM) algorithm is a clustering algorithm developed by Dunn, and later on improved by Bezdek. It is useful when the required number of clusters are pre-determined. Thus, the algorithm tries to put each of the data points to one of the clusters. What makes FCM different is that it does not decide the absolute membership of a data point to a given cluster. Instead, it calculates the likelihood (the degree of membership) that a data point will belong to that cluster. The algorithm of the Fuzzy C-means clustering is

1. Set the partition matrix $\mu_f (c)$ randomly
2. Set the value of $w > 1$ (ex: $w = 2$)
3. Set the value of the correction factor with a very small value (ex: Eps = 0.01)
4. Set the maximum iterations (ex: 500 iterations)
5. Set the initial value of the objective function $(P_f (c))$ randomly
6. Add iteration: $t = t + 1$
7. Calculate the center of vectors:

$$v_f = \frac{\sum_{k=1}^{N} (\mu_{ik})^w u_k}{\sum_{k=1}^{N} (\mu_{ik})^w}$$

(2)

8. Modify the membership value:

$$\mu_{ik}(y_k) = \left[ \frac{1}{\sum_{j=1}^{C} \left( \frac{|u_k - v_{fj}|}{|u_k - v_{fj}|} \right)^{1/(w-1)}} \right]$$

(3)

9. Modify the partition matrix
10. Calculate the objective function:

$$P_f (c) = \sum_{k=1}^{N} \sum_{i=1}^{C} (\mu_{ik})^w \left| Y_k - V_{fi} \right|^2$$

(4)

11. Check the stop condition

IV. METHOD

Study Area

The island of Java is located on the southern rim of the Indonesian archipelago and comprises an area of 132792 km2. Forest allocation covers about 19% of area, which consist of 1916964 ha of production forest (12.74%), 650 619 ha of protected forest (4.32%), and 442,188 ha of conservation forest (2.94%) [7]. Regarding to the statistical data of Indonesia [8], Java comprises only 7% of the total land area of Indonesia, however, Java’s population accounts for 70% of the total population of Indonesia or 1026 inhabitants per km2. The distribution of these allocations areas is shown in Figure 2.

Figure 2. Java island and forest allocation [2]

Data that used in this research was the EVI MODIS satellite data in 2000 and 2012 that had been in the 16-day composite. This data was obtained from one of the NASA web which can be downloaded for free.[9]

Methods

This research consists of several steps: Data Acquisition, Preprocessing Data, Image Feature
Reduction, Clustering, and Evaluation. The flow of this research can be seen in Figure 3.

**Data Acquisition**

The data used is MODIS EVI (MOD13Q1) \[9\]. File extension is HDF. MOD13Q1 is one of the data generated by NASA for the vegetation index. This data is provided every 16 days with a spatial resolution of 250m. In this research, the analyzed area is Java island with data period in 2000 and 2012. To get an overall image of Java island requires 2 data HDF. So, it takes 46 data in a year.

**Preprocessing data**

At this step, it does not need to change the pixel values into EVI values, because MOD13Q1 already automatically have the value of EVI. The data have been downloaded from the NASA website with HDF extension. It will be converted into GEOTIFF using MODIS Reprojection Tools (MRT). The next step, the image is cropped in order to produce the image area of Java island only. Then, combine 23 images into one image in a year using average composite. Average composite is the process of taking the average value of each pixel from some images data. So, from 46 data used will be generated into 2 images data.

**Image Feature Reduction**

The next step, reducing image feature based on wavelet method. In addition to reducing image feature, wavelet can also reduce noises in the image and reduce execution time in clustering. The type of wavelet used in this research is Haar wavelet level 1 and level 2.

**Clustering**

Clustering process in this research using Fuzzy C-Means (FCM). Clustering process will classify forest and non-forest areas based on the value of the EVI. To determine forest and non-forest areas of the clustering results based on data from visual classification of Land cover. After the specified area of forest and non-forest, then the data will be compared with data in 2012 and 2000 to get the area deforested.

**Evaluation**

The evaluation process in this result is to compare the results of this research with data from visual classification of Land cover.

**V. RESULT**

**Data acquisition**

Data had been taken by downloading MODIS EVI that were obtained from NASA's official website which can be downloaded for free \[9\]. The code of data used in this research is MOD13Q1 H28-V29 and H29-V29 with coverage of the Java island. There were 86 data had been downloaded in this research which were consist of 40 data in 2000 and 46 data in 2012. Total of data size were 5.56 GB with HDF extension file.

**Preprocessing data**

At this step, data were converted from HDF into GEOTIFF using MODIS Reprojection Tools (MRT). Conversion process was used UTM projection type and scale of 1 pixel: 250 meters. The conversion process produced images that were 20 images in 2000 and 23 images in 2012 with the image dimensions of 9711x4501. This image covered an area of the Java island and some parts of the Borneo and Sumatra island. The next process was cropped image in order to get the image of the Java island only. Cropping process had been done by using OpenCV in order to produce data accurately and equally in every image.

Cropping process produced an image that had 4229x1385 dimensions. Next process, image pixel value in a year were averaged to produce 1 image in 2000 and 1 image in 2012. This process was due to the EVI value in agricultural areas is very volatile. Thus, in the growing season, the value of EVI in agricultural areas have a very high value, almost like EVI values in the forest area. Meanwhile, when the harvest season, agricultural area has a fairly low EVI values. So, average composite was done to anticipate the detection of agricultural areas as forest area.

**Image Feature Reduction**

Reduction process was done to reduce the computational time on the clustering process. Image reduction process in this research using the Haar wavelet method. Wavelet can also reduce noise in the image. Wavelet had been processed until level 2. Thus, the resulting image with dimensions 2114 x 692 at level 1 and 1057 x 346 at level 2.
Clustering

Implementation of FCM clustering was done using C++ programming language and openCV2.2 by using codeblock 13.12. Clustering would categorize EVI value of image which had 2114 x 692 in wavelet level 1 and 1054 x 346 in wavelet level 2. Clustering process was implemented using global-based approach, that’s means clustering was done simultaneously without image partition. The parameters used in this research were 500 iterations, 0,001 tolerance value and 13 clusters.

Clustering process was done by classifying the EVI values were normalized. normalization used to reduce the computation time. In this research, the process of clustering with data that has been normalized able to reduce computation time by 20%. The results of clustering on the level 1 wavelet image can be seen in figure 4, while the results of clustering on level 2 wavelet image can be seen in figure 5. The normalization process was done by using the formula:

\[
EV_{\text{normalized}} = \frac{EV_{\text{max}} - EVI}{EV_{\text{max}} - EV_{\text{min}}}
\]  

(5)

Figure 4 and Figure 5 shows the membership of each cluster. The horizontal axis shows the cluster, while the vertical axis shows the number of pixels in the cluster.

Table 1 and Table 2 described the label of each cluster. Labeling was done by compared with reference map from planologi department, ministry of forestry. Based on the reference map there are 12 areas, those are dryland forest, mangrove forest, swamp forest, bush and grass land, industrial forest plantations, plantations, agricultural land and bush, pond, build up area, mining area, swamps, and other land cover. The technique of determining the area on the map references using the knowledge of experts.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Build up area, mining area, pond, and other land cover</td>
</tr>
<tr>
<td>2</td>
<td>Industrial forest, Agricultural land, Build up area</td>
</tr>
<tr>
<td>3</td>
<td>Agricultural land</td>
</tr>
<tr>
<td>4</td>
<td>Hutan industri, perkebunan, hutan, daerah pertanian</td>
</tr>
<tr>
<td>5</td>
<td>Dryland forest, Mangrove forest, Swamp forest, bush and grass land, Plantation, industrial forest, Agricultural land</td>
</tr>
<tr>
<td>6, 7, 8, 9, 10, 11, 12</td>
<td>Boundary of island (Noise)</td>
</tr>
<tr>
<td>13</td>
<td>Outlier</td>
</tr>
</tbody>
</table>

Based on Figure 4 and Figure 5, and an explanation of each clustering on Table 1 and Table 2 can be seen that the wavelet level 2 was better than wavelet level 1. It was because the wavelet level 2 was able to define the image with more detail and can reduce noise by doing a filter on pixel values that are too high or too low. In this research, wavelet will reduce outlier regions and boundary regions of the island.
Another advantage of wavelet in this research were able to reduce the computational time with very significant in clustering process. In the original image, the computational time on clustering process more or less for 360 minutes. Meanwhile the image of the wavelet level 1 and level 2 were for 120 minutes and 30 minutes. It can be concluded that the wavelet level 1 and level 2 can reduce the computational time for 240 minutes (66%) and 330 minutes (91%).

If we look at Table 2, there were some clusters that had the same label such as the agricultural and industrial forest areas. It was because the type of agricultural area that had many variations such as dryland farming, wetlands farming, and others. Besides that, differences in period of each plant can made differences in EVI values significantly. Meanwhile, the industrial forest was heavily influenced during growth, harvest and post-harvest period. When the post-harvest period EVI value was very low. It was because there were no photosynthetic that occurs because vegetation was not there anymore. Meanwhile, when the growth period, the value of EVI was very high because the plant had a very high rate of photosynthesis.

On the other hand, there was a cluster defined in several different areas such as cluster 6 was defined as a dryland forest, mangrove forest, swamp forest, industrial forest, plantations, agricultural land, bush and grass land. It was because the areas had same rate of photosynthesis value and density of vegetation value. So the areas were defined in the same cluster.

Table 3 described range of EVI value for each cluster. It was seen that, the outliers had the highest value. It was because the representation of value in the MODIS image used was 16 bits with a value of 65 536 as outlier. While EVI values was defined from 0 to 10 000. The boundary area Noise) had EVI value from 120 - 253. Noise occured because when the average composite, EVI value of the boundary island were averaged with outlier values. So, boundary area had a very high value.

While cluster 6 was a cluster that had a very high EVI values, followed by cluster 5, cluster 4, cluster 3, cluster 2, and cluster 1. As described previously, the EVI value was determined by photosynthesis rate and vegetation density. It can be concluded that cluster 6 had a high photosynthesis rate and vegetation density very much. Whereas in cluster 1 are defined as Build up area, mining area, pond, and other land cover that had a little vegetation.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Label</th>
<th>EVI value (degree of membership)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Build up area, mining area, pond, and other land cover</td>
<td>1 (1) – 34 (0.43)</td>
</tr>
<tr>
<td>2</td>
<td>Agricultural land, Build up area</td>
<td>35 (0.45) – 47 (0.6)</td>
</tr>
<tr>
<td>3</td>
<td>Agricultural land</td>
<td>48 (0.53) – 56 (0.66)</td>
</tr>
<tr>
<td>4</td>
<td>Industrial forest, Agriculture land</td>
<td>57 (0.83) – 63 (0.85)</td>
</tr>
<tr>
<td>5</td>
<td>Industrial forest, Plantation, Dryland forest</td>
<td>64 (0.9) – 71 (0.9)</td>
</tr>
<tr>
<td>6</td>
<td>Dryland forest, Mangrove forest, Swamp forest, bush and grass land, Plantation, industrial forest, Agricultural land</td>
<td>72 (0.9) – 89 (0.9)</td>
</tr>
<tr>
<td>7,8,9, 10,11, 12</td>
<td>Boundary of island (Noise)</td>
<td>120 (0.9) – 253 (0.9)</td>
</tr>
<tr>
<td>13</td>
<td>Outlier</td>
<td>254 (1) – 255 (1)</td>
</tr>
</tbody>
</table>

If we look at table 3, cluster 1 was a cluster that had a very high EVI values and most of the members of the cluster was forested areas, so that the cluster was defined as forest area. The forest area in this research can be seen in Figure 6. Forest area on image wavelet level 2 can be calculated using the formula: 

$$\text{Area} = \text{total pixel} \times 6.25 \text{ ha} \times 16$$

This formula based on scale of MODIS image used is 1: 250 m so, to calculate the area of a pixel can be multiplied with 6.25 ha. While multiplication with 16 based on wavelet level 2 will reduce 1/16 image size than the actual size.

From Figure 6, we can see that the experiment result showed the increase of forest areas or reforestation of 214 700 ha. Meanwhile, reference showed the decrease of forest area or deforestation of 165392.60
ha. This difference was happened because the EVI value just depends on the photosynthesis rate and vegetation density. Thus, the forest area which was defined in this research covers dryland forest, mangrove forest, industrial forest, plantations, agricultural land, bush and grass land that had EVI values are almost same. In the reference, swamp forest, grass land and plantations areas had expanded. On the other hand, industrial forest area had been detected as a forest in this research had been also increased.

Evaluation
This research had successed to classify areas based on the value of EVI such as the area that had high EVI values (forests, plantations, grass land), medium (agricultural land), and low EVI value (Build up area, mining area, pond, and other land cover). But, EVI value was only influenced by the rate of photosynthesis and vegetation density. Thus, it was difficult to detect areas in more detail. For example, forest areas were detected in cluster 1 with other areas such as mangrove forest, swamp forest, industrial forest, plantations, agricultural land, bush and grass land. In addition, the forest area was also detected in cluster 2 with industrial forest and plantation area.

VI. CONCLUSION
This research had successed to classify areas based on the value of EVI such as the area that had high EVI values (forests, plantations, grass land), medium (agricultural land), and low EVI value (Build up area, mining area, pond, and other land cover). But, EVI value was only influenced by the rate of photosynthesis and vegetation density. Thus, it was difficult to detect areas in more detail.

VII. REFERENCES
[1] Setiawan et al, Characterizing Temporal Vegetation Dynamics of Land Use in Regional Scale of Java Island, Indonesia, J.Land Use Sci,iFirst, 1-30, 2011