

Improvement of Segmentation Performance for Feature Extraction on Whirlwind Cloud-based Satellite Image using DBSCAN Clustering algorithm

Nailussa'ada, Tri Harsono, Achmad Basuki

Politeknik Elektronika Negeri Surabaya
Kampus PENS, Jl. Raya ITS, Sukolilo Surabaya
E-mail: snailus96@gmail.com, {trison, basuki}@pens.ac.id

Abstract

Images contain a lot of information that can be used in a variety of areas. One of the images that have much information inside is satellite image. In order to extract the information properly, the image processing step should be performed properly. The segmentation process plays an important role in image processing, especially for feature extraction. Many ways were developed to perform the segmentation image. In this study, we apply DBSCAN clustering to segment images on whirlwind cloud feature extraction problems. DBSCAN is a density-based classifier method which means it is suitable to group a density-based data. While the image used in the segmentation process is the Himawari 8 satellite image which also contains density-based data. It contains various information about clouds condition like cloud type, cloud temperature, cloud humidity, rainfall potential based on cloud temperature, etc. This study uses Himawari 8 satellite images as input where the images taken are images several hours before a whirlwind event in an area, while the cluster method used is the DBSCAN algorithm. Clustering is done to get the extraction features of a whirlwind in the form of centroid points that characterize the movement of a cloud. Segmentation performance was observed based on the number of centroid points as a result of clustering several types of clouds in an area before a whirlwind occurred. Based on segmentation testing using the DBSCAN algorithm for cloud data in an area for several hours before a whirlwind, better segmentation performance was obtained compared to the segmentation results of the Meng hee heng k-means algorithm for the same test data specifications. DBSCAN separates a type of cloud in more detail that makes it easier to record each centroid of each cluster around the scene. It is even able to cluster small groups of clouds independently so that these small groups of clouds can also be detected as features.

Keywords: Satellite Image, Segmentation Performance, Feature extraction, Whirlwind, DBSCAN Clustering

1. INTRODUCTION

An image can contain a lot of useful information and it also can be a way of transferring that information. In order to get the information in an image, we need to understand what the image is and what information can be extracted from it. The information extracted from an image can be used to accomplish some works in various areas. One example image contains a lot of useful information is a satellite image that is mainly used in a geographic information system [1]. In order to get the necessary information from an image, the extraction process becomes one of the important task of application in digital image technology [2]. Digital image processing is a process to perform image processing into digital images using computer algorithms [3]. It also a type of signal processing with input is an image and output can be an image or features associated with that image input. The three general phases in digital image processing are preprocessing, enhancement and display or information extraction. From all of the process in image processing, segmentation is an important and challenging process [3]. The main purpose of the segmentation process is to divide an image into regions or categories, which belong to different objects or parts of an object [4]. Good segmentation comes if pixels in the same category have similar values and form a connected region. In another way, image segmentation can be explained as a process of assigning labels for every pixel in an image so that pixels in one label have certain characteristics. Each pixel in an area is related to some characteristics or properties, such as color, intensity, or texture.

In this study, we use satellite image as an input and segment it to extract the useful information from it. The satellite image was used in many different areas like astronomy, GIS, agriculture monitoring, corp census, weather forecast, disaster management and many more. In this study, we also implement the computational algorithms to interpret the information inside satellite images. Similar with common digital image processing, satellite images segmentation also contains a most significant role to play in remote sensing imaging [5]. Since 1994, there are already six classes of segmentation techniques used to accomplish image segmentation [6]. They are amplitude thresholding, component labeling, boundary-based segmentation, region-based segmentation, template matching, and texture segmentation.

In the previous research, region-based segmentation use clustering algorithms were done to Himawari 8 satellite images [7]. It shows comparison result using two clustering algorithms, meng hee heng k-means and DBSCAN algorithm. It claims that DBSCAN gives a more satisfying result than meng hee heng k-means. Himawari 8 satellite is used in this study. It contains information about clouds that classified based on their temperature. This satellite image is usually used for weather forecasting by meteorologist expert. It contains various information about clouds condition like cloud type,

cloud temperature, cloud humidity, rainfall potential based on cloud temperature, etc.

In this study, we provide an analysis and explanation of how is DBSCAN can be a better segmentation method and its relationship with feature extraction result. We analyze the segmentation result by implementing it to a whirlwind case. The purpose of implement the segmentation into whirlwind case, especially into cloud satellite image is to extract the information of cloud presence as the feature for whirlwind phenomenon. By using the clustering method, centroid can be produced, these centroid represents the cloud presence in somewhere. By knowing the cloud presence condition before whirlwind occur, analysis of the sign pre-event can be known. Here we use the segmentation result of meng hee heng k-means and DBSCAN to extract feature on a whirlwind incident data and compare the performance of them. Some preprocessing steps were performed before doing the segmentation process with Meng Hee Heng K-Means and DBSCAN algorithm. The output from the segmentation process is centroid points from some clusters. It was transformed into numeric table data for the next analysis requirement. After analyzing it, information extracted by DBSCAN is more complete, more convenience and more suitable for whirlwind incident than meng hee heng k-means.

2. RELATED WORKS

Biplab Banerjee et al propose using a novel adaptive non-parametric means-shift clustering algorithm to perform segmentation of satellite image [8]. The segmentation is a process of separating an image into its constituent object. Low-level image interpretation and high-level visual analysis are bridged by this step, it makes this step become an important step. The performance of the mean shift algorithm is affected by the size of the parzen window and the terminating criteria. In this study, a purely statistical framework has been used to take care of this issue.

Ganesan P et al, use a modified fuzzy c-means clustering algorithm to present a comparative study of satellite image segmentation in RGB and HSV color space [9]. They use GeoEye-1 satellite image as the experiment data. RGB color space is not efficient for color specifications and recognition. In addition, the chrome and luminance components are mixed that makes it is difficult to determine certain colors in RGB and space. So it is a need to transform the RGB image into another color space. HSV color space approaches human vision. The illumination variations in the image can be easily solved because the component values do not depend on the color. So we have chosen this color space for satellite image segmentation. The comparison is use number of bivariate image quality parameter on the image input and segmented image. It uses to measure the similarity between the input and the segmented image based on the corresponding pixels of both images.

Satellite images are not only used for disaster management or weather forecasting, but also for object detection. Maria Vakalopoulou et al use satellite image for building detection based on deep convolutional neural networks [10]. The core of their method is based on a supervised classification procedure with very large training dataset. The optimal label for detecting building scene was performed with an MRF model. The training model has been formed to solve a binary problem, building or not building. The training dataset contains building-centered patches and randomly generated patches for non-building classes. These patches were inserted into ImageNet framework and the feature vector extracted from FC7 layer features for training by SVM classifier. Optimized integrated MRF significantly improves final building detection maps. Quantitative validation is shown promising results with a significantly high level of completeness and correctness detection.

Packyanathan Ganesan et al proposed using hill climbing local optimization technique and modified k-means clustering to segment the satellite images [11]. Firstly, the RGB images are converted into CIELAB color space. The initial centroid is obtained by the hill climbing process when it is applied to the color histogram of CIELAB color space image. The centroids are given into k-means clustering and produce the segmented image. This proposed collaboration method takes the advantages of both algorithms. When it is tested into different satellite images, the proposed method proved the robustness of computational cost and also the number of clusters.

In 2015, Kaliyamurthi K.P et al perform a segmentation process to a satellite image that supported color feature with unattended FCM formula [5]. The first step is to improve the separation color of satellite imagery victimization color. The second step is for the region method is sorted into a group of FCM clump formulas. FCM uses space between pixels and center clusters within the spectral domain for reasons of membership operating. The pixel in images are much related to, and this spatial info is an important characteristic that will not help their classification. Here are places that are recognized by FCM technique, Land original image pixels are seldom to form sensible and effective use of FCM. Water Body original image. Geographical region original image. The recent one, in 2018, we have been present segmentation in satellite image using two clustering algorithms [7].

DBSCAN is a density based clustering algorithm that usually implement in data mining. Among the density based mostly cluster algorithms DBSCAN is exceptionally well known due both to its low complexity and its capacity to detect clusters of any shape, which is a desired characteristics when one doesn't have any knowledge of the possible clusters' shapes, or when the objects are circulated heterogeneously, for example, along paths of a graph or a road network [15]. In their research, Jain et al [15] was analyze the performance of DBSCAN in data mining and get the similar and dissimilar style of knowledge from the foremost dense region within the dataset. DBSCAN is also implement into video processing by some

researchers. Huan yu et al [16] utilize DBSCAN to cluster and separated the corner of object moving based on their motion's vectors angles and length. The result of the experiment is accurately distinguish between corners on moving objects and background. Another good result of DBSCAN in video processing is gained by Dayang sun et al [17]. They utilize DBSCAN to recognize vehicles in order to get the number of vehicle inside the frame. It gives good performance in achieve the target of that research. Utilizing the same method and version of DBSCAN which used in those video processing research, we implement it as the segmentation method for cloud satellite images, Himawari 8.

3. ORIGINALITY

On the previous research, we have been tried using two clustering algorithms as the segmentation method into Himawari 8 satellite image [7]. Himawari 8 contains various information about clouds condition like cloud type, cloud temperature, cloud humidity, rainfall potential based on cloud temperature, etc. Bring the same data into this study, we present an analysis of the effect of segmentation result into feature extraction in whirlwind case.

In 2017, Himawari 8 satellite image is used for predict whirlwind by Rhoma Cahyanti et al [12]. Inside the whole process of build the system, segmentation was performed using meng hee heng k-means. In this study, we proposed to use DBSCAN as the segmentation process to improve the feature extraction result. The improvement given by DBSCAN can be seen on the detail grouping of cloud which more specific and more separate than previous method, using meng hee heng k-means. It proven by the number of cluster obtained and the group position of cluster. DBSCAN produces more cluster and the separation between each group of clouds is clearer. Whirlwind occurs in a local area which means it only appears in a certain location and stays there. The clearer separation of clouds makes the observation of each cloud feature easier and better.

The segmentation is performed into the whirlwind historical data in 2017. It produces some cluster of clouds groups for every incident data. From each of cluster, we collect the centroid point and transform it into numeric table data. Each column in each of row indicates the appearance of clouds in the sky. The appearance of clouds become the parameter for the whirlwind data. From that data table, we observe each row completeness and the effect of it into whirlwind feature for the purpose of prediction. DBSCAN is considered better and more suitable to perform segmentation and extracting feature in a satellite image for the whirlwind case. The main reason for that statement is because we need to cluster cloud into some groups and cloud groups are about density, so we need a method that can classify density-based data like DBSCAN.

4. SYSTEM DESIGN

The proposed system consists of 4 main steps: (1) image preprocessing, (2) segmentation with a clustering algorithm, (3) data transformation and collection, and (4) data observation. Beside of the four main steps there are several other processes such as collecting necessary data likes satellite image and historical data. The whole process is shown in Figure 1.

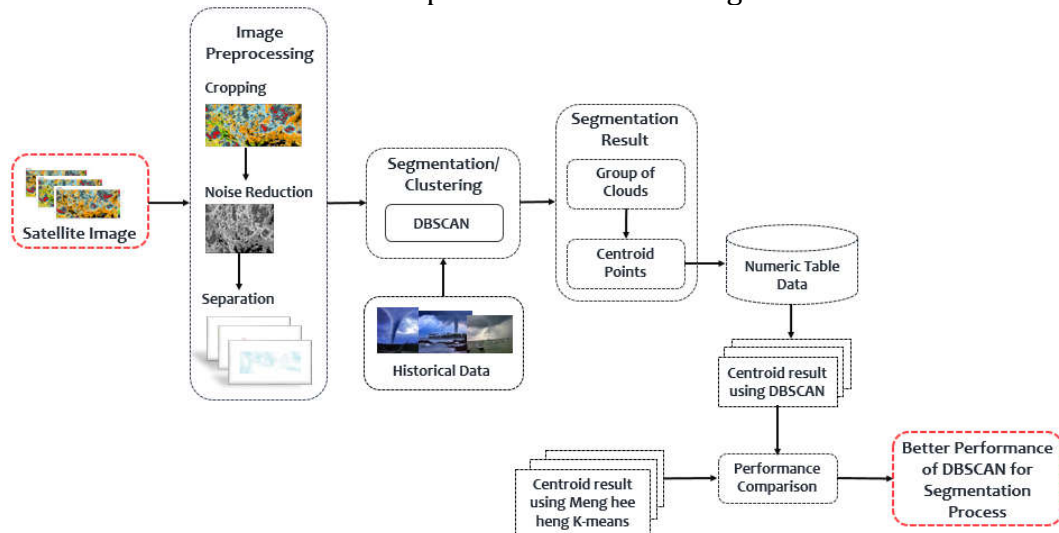


Figure 1. System Design

4.1 Satellite image collection

Himawari-8 is a Japanese weather satellite operated by the Japan Meteorological Agency. This satellite aircraft was built by Mitsubishi Electric with the help of Boeing and is one of the DS-2000-based satellites. Some types of Himawari-8 satellite images owned by BMKG include Himawari-8 Enhanced IR, Himawari 8 Natural Color, Himawari-8 WV Enhanced, Himawari-8 Rainfall Potential, Himawari-8 IR Enhanced (low res), Himawari-8 Wildfire Detection, and Himawari-8 RDCA.

In this study, the type of Himawari-8 IR Enhanced satellite image will be used. This satellite image is updated every one hour and can be downloaded from the BMKG website. Himawari 8 shows cloud peak temperatures obtained from observations of radiation at 10.4 wavelength micrometers which are then classified by staining, black or blue shows that there are not many bright clouds, whereas colder clouds peak temperatures where colors approach jungles, shows significant cloud growth and the potential for Cumulonimbus clouds to form. Low clouds generally have warmer temperatures and are relatively close to the earth's surface. Low clouds are depicted using dark blue to light green. While the cooler clouds are generally higher the peak is described as orange to bright pink. In this study to separate each type of cloud, a selection technique based on RGB is used which symbolizes each type of cloud. The following are the RGB values of each type of cloud:

- Cumulonimbus: (255, 0, 0)

- Dense: (55, 87, 87)
- High Cloud: (55, 87, 87)
- Middle Cloud: (255, 221, 85)
- Cumulus: (255, 147, 2)
- Stratocumulus: (185, 255, 55)
- Stratus / Fog: (55, 221, 8)

4.2 Pre-processing satellite image

In the preprocessing stage, separation of each cloud type is done on satellite imagery. First, to get these types of clouds, cropping has been done before to get the necessary area and also noise reduction to reduce disturbances in the image (Figure 2). Noise, in this case, is the spread of small clouds in the image.

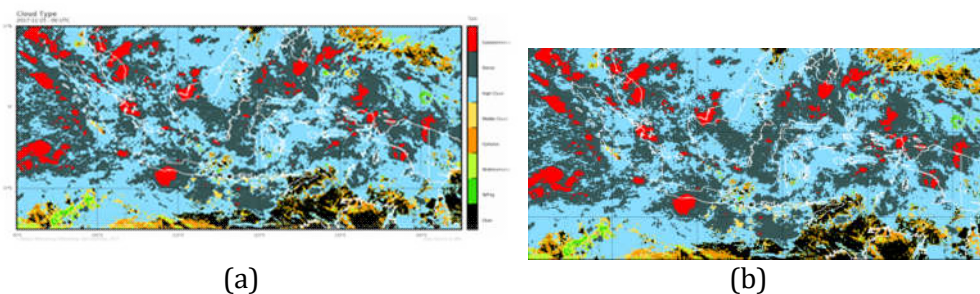


Figure 2. (a) Himawari 8 before cropped. (b) Cropping result of Himawari 8

Because of Himawari image contains some cloud type information which separated by color, the selection is done based on RGB selection for each type of cloud according to their RGB values described in section 4.1. Selection was done to get images according to each cloud type which processed separately. So that a himawari image will split into 6 different images indicates 6 different cloud type. The example result can be seen in Figure 4 in section 5.1.

4.3 Segmentation

Image segmentation is a way to split the image into multiple regions based on certain characteristics homogeneous [13]. There are two main approaches in image segmentation, which is an approach based on edge and an approach based on region. The most often used method for segmentation process is region growing, clustering (K-means clustering, ISODATA clustering, graph-theoretic clustering, etc.) or split and merge. In this study, segmentation is done to produce groupings of each type of cloud based on the type and area or position. In this study DBSCAN was proposed as the segmentation method. The output of this segmentation used in the next step is the centroid in each of cluster/group of clouds (Figure 3). The centroids points were listed as the feature of the data.

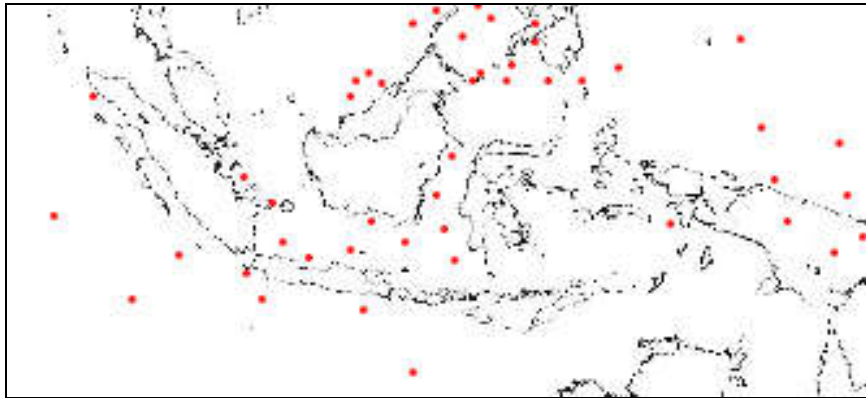


Figure 3. Centroid point produced by segmentation process

We use whirlwind case as the topic in this study, so the segmentation process was applied to the historical whirlwin incident that occur in Indonesia during 2017. The feature which is the centroid points obtained from the segmentation process in every historical data are collected so in the future process it can be used to observer the characteristic of whirlwind cases.

4.3.1 DBSCAN

DBSCAN or density-based spatial clustering applications with noise is a density-based algorithm that is commonly used in machine learning or data mining. Data points adjacent to many neighbors are collected into one cluster and points with small or far from neighboring points are considered as outliers. This algorithm is actually quite common and widely used in scientific literature even it has won the test of time award in 2014. The advantages of using this method are the ability to detect noise and outliers, does not require to specify the number of clusters and can find arbitrarily shaped cluster. In addition, to handling computing time problems, this algorithm is indeed designed for data with very large scalability. DBSCAN requires two basic parameters that are eps (ϵ) and minPoints. Eps describes the minimum distance between two points that can be considered as neighbor. MinPoints is the minimum number of points to form a group region. If the minPoints value is 3, then a group region should at least has 3 points as the member. To find a cluster with this method, DBSCAN starts with a dynamic point p and collects all adjacent points starting with p wrt. If p is the center value, wrt cluster will be generated. If p is the edge value and there are no adjacent points, DBSCAN will switch to the next point in the data store [14]. Here is the step by step description of DBSCAN algorithm in form of pseudo-code:

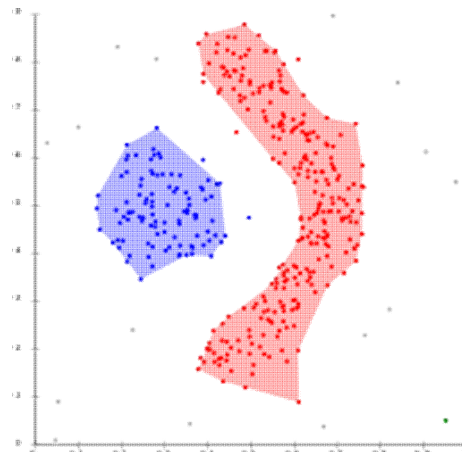

```

DBSCAN(D, epsilon, min_points):
    C = 0
    for each unvisited point P in dataset
        mark P as visited
        sphere_points = regionQuery(P, epsilon)
        if sizeof(sphere_points) < min_points
            ignore P
        else
            C = next cluster
            expandCluster(P, sphere_points, C,
epsilon, min_points)
expandCluster(P, sphere_points, C, epsilon,
min_points):
    add P to cluster C
    for each point P' in sphere_points
        if P' is not visited
            mark P' as visited
            sphere_points' = regionQuery(P',
epsilon)
            if sizeof(sphere_points') >=
min_points
                sphere_points = sphere_points
joined with sphere_points'
                if P' is not yet member of any
cluster
                    add P' to cluster C
regionQuery(P, epsilon):
    return all points within the n-dimensional sphere
centered at P with radius epsilon (including P)

```

In simpler way, the description of step by step DBSCAN algorithm are:

- Find the neighbourhood of every point with parameter ϵ (eps)
- Identify the core points that have neighbor more than minPoints
- Find connected components of core points with the neighbor graph
- Assign non-core point to the nearby cluster with nearest ϵ (eps), if there is no nearby cluster meet the ϵ (eps), assign it to noise.



Source: <https://commons.wikimedia.org/wiki/File:DBSCAN-density-data.svg#/media/File:DBSCAN-density-data.svg>

Figure 4. DBSCAN result visualization

4.4 Numerical Data Table transformation

Centroid data from the historical data that have been collected will be converted into one row of data for every event. The centroids are collected and form a data table. The column in every row becomes the feature for every incident. The number of column or feature of both algorithms indicates the completeness of the feature extracted from the original data. It can be the indication of the segmentation algorithm performance in order to extract the feature of the whirlwind incident.

5. EXPERIMENT AND ANALYSIS

We collect some Himawari 8 satellite image during 2017 and also historical data of whirlwind incident occurs in Indonesia. Figure 5 shows a Himawari 8 satellite image visualization.

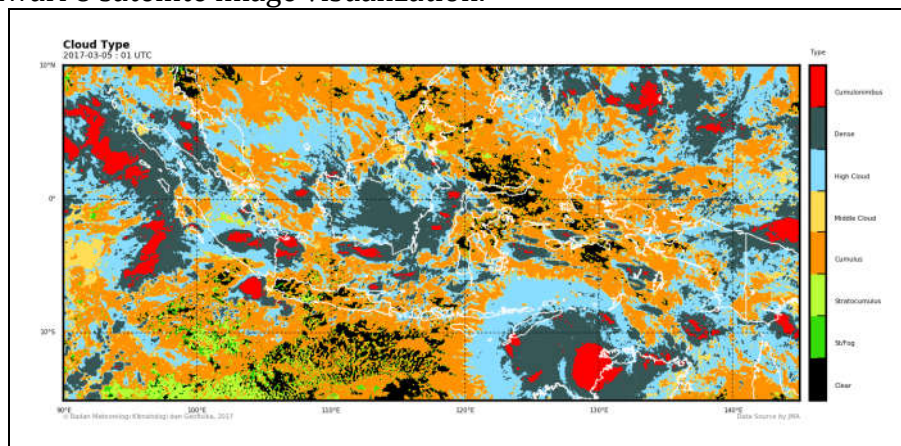


Figure 5. Himawari 8 Satellite Imagery

Himawari satellite images were collected from Indonesian agency for meteorological, climatological and geophysics. Beside of satellite images, we also collected more than 30 historical whirlwind data during 2017 from

official website of Indonesian ministry of health. The parameter that we collect from historical data are the time of the incident consist of date, month, year and time (in hour), also the location of the incident. For process a historical data we only take 7 images (7 hours data) before the incident occurs, that's why we need the time parameter. We apply step by step process into each of historical data, start from preprocessing, segmentation process, transformation data into numerical table data, and performance comparison. We implement the clustering algorithms for the segmentation process and collect the centroid from it. These centroid points used as the feature in this study.

5.1. Preprocessing

The original Himawari 8 satellite (Figure 2a) image should be prepared before begin the processing to get the correct information. We discard unnecessary area of an image by cropping the white edge area to obtain the only area contain cloud information. We also implement noise reduction method such as median filter to reduce the appearance of very small clouds in the image (Figure 2b).

In selection step, we separate every cloud type from one image into six different images. One image contains eight different colors indicate the six cloud type, one for fog and one for clear sky. Here, we only take 6 colors of cloud because we do not use stratus/fog and clear sky information for the next process. Selection was done by collecting pixel with the same RGB color. Here are the RGB values of each information in Himawari 8:

- Cumulonimbus: (255, 0, 0)
- Dense: (55, 87, 87)
- High Cloud: (55, 87, 87)
- Middle Cloud: (255, 221, 85)
- Cumulus: (255, 147, 2)
- Stratocumulus: (185, 255, 55)
- Stratus / Fog: (55, 221, 8)
- Clear: (0,0,0)

The selection result of one image produces six new images (Figure 6) that only contain one cloud type in every image, stratocumulus cloud (Figure 6a), cumulus cloud (Figure 6b), middle cloud (Figure 6c), high cloud (Figure 6d), dense cloud (Figure 6e) and cumulonimbus cloud (Figure 6f).

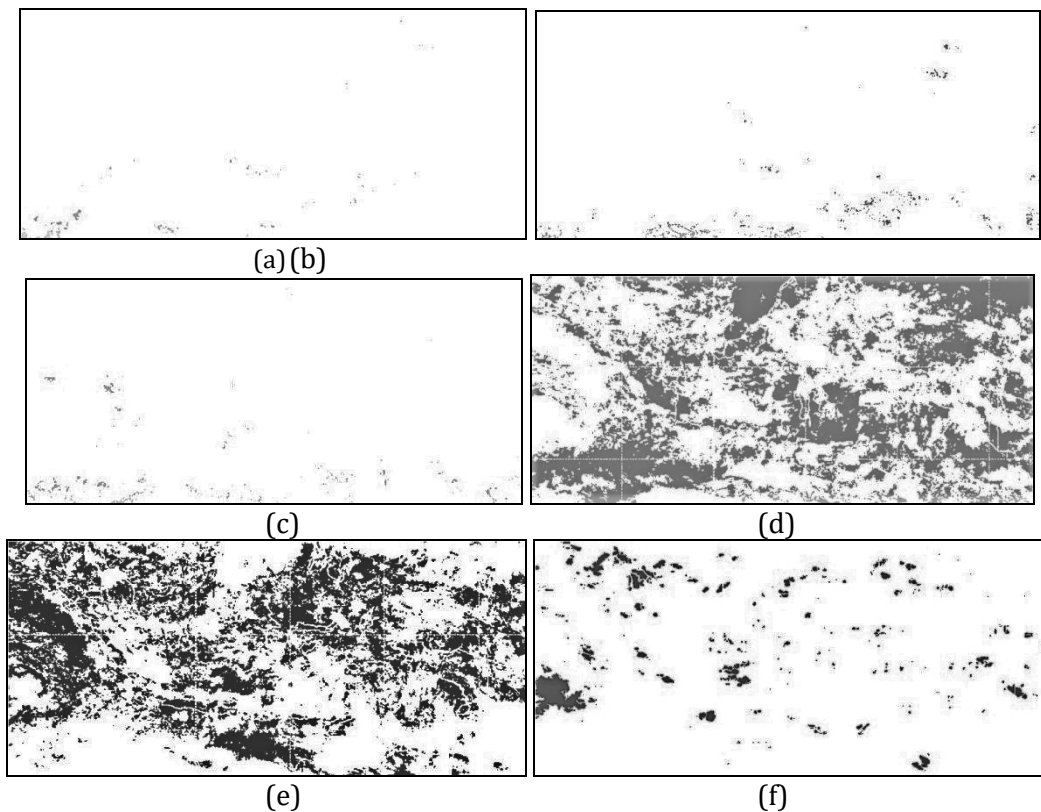


Figure 6. (a) Selection result for Stratocumulus, (b) selection result for cumulus, (c) selection result for middle cloud, (d) selection result for highcloud, (e) selection result for dense, (f) selection result for cumulonimbus

5.2. Image Segmentation

The main purpose of the segmentation process is to find a point to represent every group of clouds. First, the scattered clouds were clustered based on the region and then from every cluster we collect the centroid. In this experiment, we implement a clustering method called DBSCAN algorithm as the segmentation method.

5.2.1. DBSCAN

Data points adjacent to many neighbors are collected into one cluster, while points with small or far from neighboring points are considered as outliers. Eps and MinPts are the parameters used. Eps is the maximum distance between points to be expressed as a group. MinPts is the number of points around a point to be declared as the center point. In this experiment, we use Eps value 5 and MinPts value is 3. Figure 7 below shows one of original cumulonimbus cloud image that appear when there is whirlwind in Bojonegoro Regency on March, 14 2017 at 7 a.m.



Figure 7. Cumulonimbus cloud image before segmentation

From the original image of cumulonimbus cloud appearance (Figure 7), DBSCAN produces cloud group or cloud cluster until 74 clusters. In Figure 8, there are 74 different colors that indicate the 74 clusters form by DBSCAN. Unfortunately, the visualization can not shows much difference in color for each cluster because the RGB value of them has only one or two different values (Figure 8).

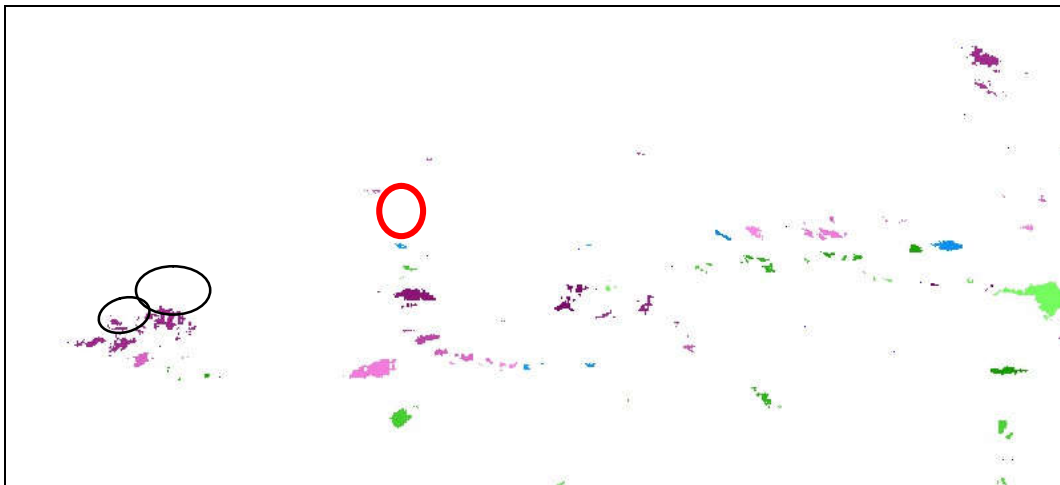


Figure 8. Segmentation result using DBSCAN clustering

Figure 8 shows the result of DBSCAN implementation which produces 74 clusters on it. The separation or grouping of group of clouds is done more detail, the separated small group of cloud also form their own cluster, look into red circle in Figure 8. Beside of that, the separated adjacent group of cloud still can be separated regardless how close they are. The two black circles in Figure 8 show two different adjacent group of clouds and they are defined as two different clusters. It is difficult to see the color difference in that figure because the RGB are generated randomly, but actually they are different. The smaller black circle cluster has RGB value (152, 50, 123),

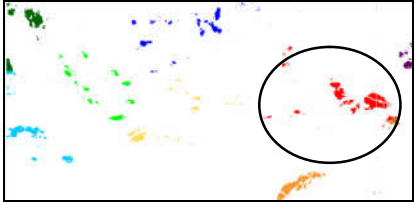

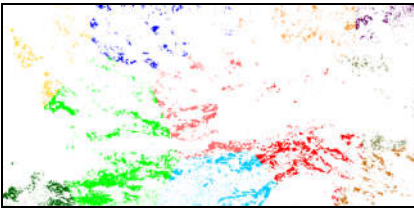
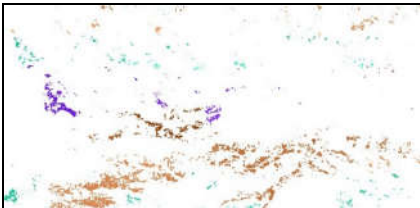
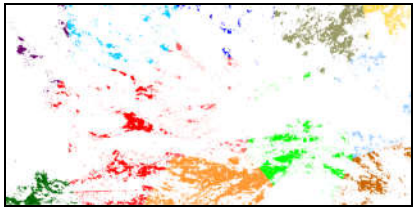
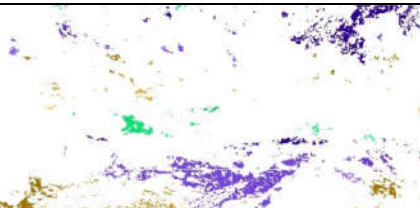
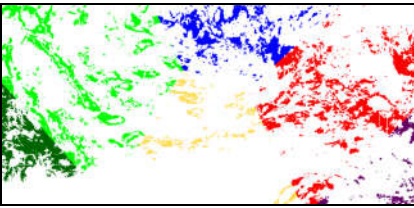
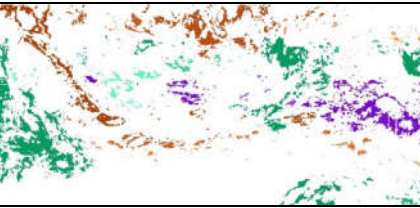
meanwhile the bigger one is (147, 48, 130). Even though it produces many clusters but the time execution needed is only 0.053 second

5.2.3. Experimental Comparison

This section explains the comparison result visualization for the previous algorithm used meng hee heng k-means and the proposed method with DBSCAN (Table 1). Meng hee heng k-means is commonly used for image segmentation in many field, just like the previous research which also implement it to segment himawari 8 satellite image [12]. Even though it is a common method for segmentation, but there is still another method that's worth to try and more suitable to the target. Meng hee heng k-means and DBSCAN have different basic concept which meng hee heng grouping the data based on the nearest centroid, meanwhile DBSCAN grouping the data based on the density. The density means the amount of data that is within the radius Eps of each data. Both of algorithms, meng hee heng k-means and DBSCAN are automatic clustering, so there is no initial k value needed, it makes they are suitable for region based image segmentation. In order to see which algorithm is more suitable for cloud satellite image segmentation, the comparison is done to obtain the same information from the same data compare to the previous one.

The comparison is shown in four types of clouds (stratocumulus, cumulus, dense and cumulonimbus cloud). Based on the visualization below, we can see that the color in DBSCAN result in every cloud type is more diverse than meng hee heng k-means. It means more cluster is formed by the DBSCAN algorithm. More clusters are formed, more centroid obtained, it means more complete feature can be obtained from DBSCAN result than meng hee heng k-means.

Table 1. Visualization of comparison result of two clustering algorithms

Cloud Type	Meng hee heng k-means	DBSCAN
Cumulonimbus		
Strato cumulus		
Cumulus		
Dense		

If we look closer into it, the difference of them can be seen from how the algorithms divide the small group of cloud into an independent cluster. Figure 9a shows that meng hee heng k-means grouping the separated small group of cloud into one cluster. Meanwhile, DBSCAN grouping them as different clusters (Figure 9b). It becomes the reason DBSCAN can produce more centroid as the feature.

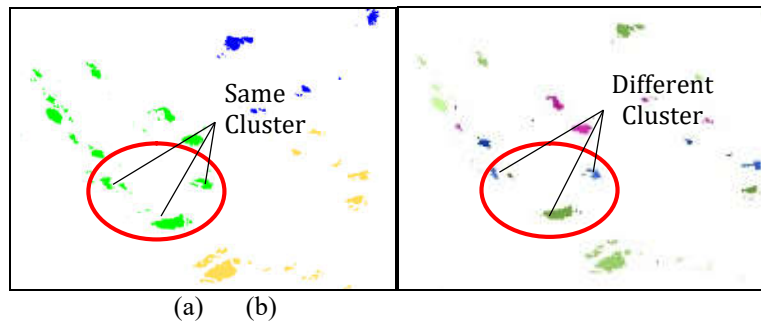


Figure 9. Comparison result on small group of clouds

Overall cluster number result of both algorithms shown in Table 2 below. There is a big difference in the number of cluster result on it. It means DBSCAN can group clouds into many more independent cluster even makes a small group of clouds has their own cluster. In order to extract feature for whirlwind in specific area, centroid of some small groups of clouds are important, because it can give more detail cloud appearance information around the scene.

Table 2. Comparison number of cluster of two algorithms

Cloud Type	Number of cluster	
	Meng Hee Heng K-Means	DBSCAN
Stratocumulus	11	153
Cumulus	12	141
Dense	6	144
Cumulonimbus	9	58

Figure 10 shows the graph that visualize the distribution value of cluster number obtained from both algorithms. Even though they have big differences but the range cluster numbers of both are stable. Meng hee heng k-means has range value 0.45 and DBSCAN has range value 0.47 (look into line inside the graph in Figure 10).

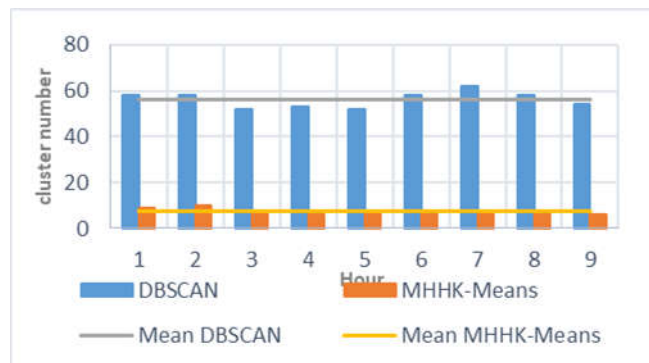


Figure 10. Data range average of both algorithms

More groups of clouds are formed means the feature obtained is more complete and better. Especially in whirlwind case, the observation area is small, so the appearance of small groups of clouds are important. More explanation about comparison of both algorithms implementation in whirlwind case is explained in the section 5.3.

5.3. Comparison of Observations on Whirlwind Incidents

In this section, segmentation is performed into real whirlwind historical data. The segmentation step is produces clusters which also bring centroid point in every cluster. In this experiment we apply both algorithms into whirlwind incident data and listing the centroid of them.

One of whirlwind incident occurs in Indonesia at March 14, 2017, used as the historical data sample. Figure 11 shows result of segmentation using DBSCAN in one of the cumulonimbus image on March 14, 2017, that occurs in Bojonegoro. It



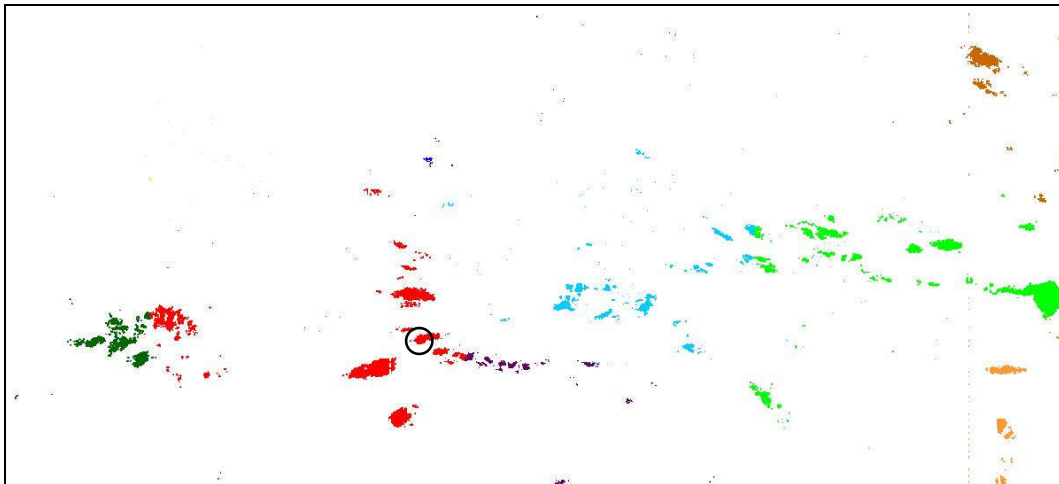
Figure 11. Segmented image result of DBSCAN

In Figure 11 there are 74 different colors indicates the 74 clusters formed by DBSCAN algorithm. The centroids are collected and transformed into numerical data table. Table 3 contains 74 rows consist of 74 clusters produces in one image for cumulonimbus cloud.

Table 3. Centroid points table

Centroid X	Centroid Y
735.8278	39.01667
767.5	50
736.0455	60.31818
751	65.5
477.1667	110.5
796	110.5
...	...
308.6667	300.3333
582	300.3333
296.781	308.8759
750.5294	316.6176
749.8333	356.8333
417.3333	358.75

Implement the previous clustering algorithm [12] using meng hee heng k-means into the same data, we obtain only 10 clusters. Figure 12 shows there are 10 different groups/clusters that are indicate by 10 different colors produced by meng hee heng k-means.

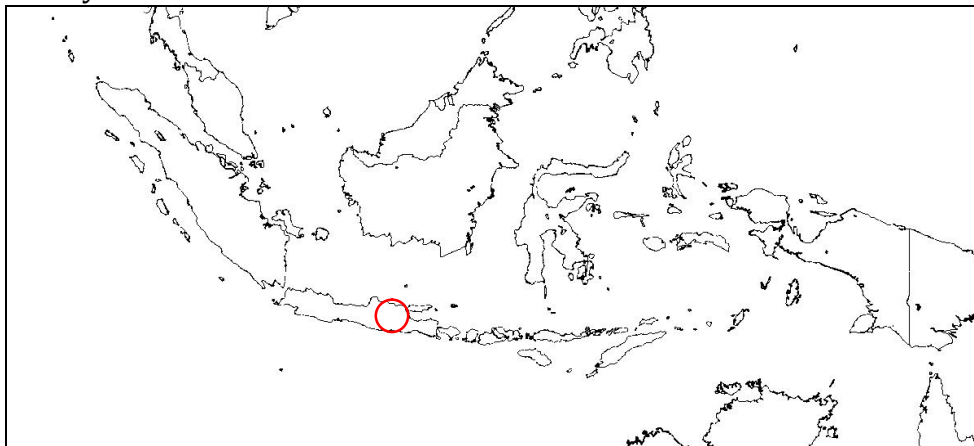
**Figure 12.** Segmented image result of meng hee heng k-means

The centroid points are transformed into numerical table data looks like Table 4. There are only 10 rows which means there are 10 clusters produced from an image. The number of clusters produced from both algorithms has big differences value. DBSCAN produces 74 clusters and meng hee heng k-means only 10 clusters.

Table 4. Centroid points of meng hee heng k-means

Centroid X	Centroid Y
226.6	248.6
661.3	194.6
400	7.3
148.67	40.6
424	357.3
7.3	295.3
793.3	23.3
754	360.67
464	180
627.33	20.66

We point out that whirlwind occurs in a local area which means its only appear in a certain location and stay there. The centroids result above (Table 3 and Table 4) are the result for all over Indonesia area in one image. In order to extract the centroid as a parameter for whirlwind we need to look closer into the area of incidents only. So the next step is to select the centroid that appears in the incident area. For the incident data sample above that occur in Bojonegoro, we select only for specific pixel area which indicates the incident boundary.

**Figure 13.** Incident area of whirlwind in Bojonegoro

The red circle in Figure 13 shows the specific location of Bojonegoro. That area is specified in pixel range area between (330,242) and (352,264). This range area obtained by dividing the original Himawari 8 image into grids with scale 2.5 degree. The centroid produced from both algorithms is filtered so only centroid that located inside the specific area will be stored into new table data. Segmentation with both algorithms was applied into seven sequence images which are images for seven hours before the incident, then the centroids are filtered. Table 5 shows the filtered centroids from DBSCAN and Meng hee heng k-means in cumulonimbus cloud case during 7

hours. In DBSCAN result, centroids didn't appear at the 4th image and 7th image. Meanwhile for Meng hee heng k-means result, none centroids appear during 7 hours before the incident.

Table 5. Filtered centroid points

	1		2		3		4		5		6		7	
	x	y	x	y	x	y	x	y	x	y	x	y	X	y
DBSCAN	346	263.1	347.7	261.5	344.7	261.2	-	-	346.5	253.1	334.5	251.0	-	-
MMHK-Means	-	-	-	-	-	-	-	-	-	-	-	-	-	-

When Meng hee heng k-means result can not produce centroid points around the scene area, especially inside the nearby area, it becomes the worst case that can happen on segmentation process. If there is no centroid obtain from it then there is no feature characteristics can be seen.

Figure 14 shows why there is no centroid appear at the incident area. Incident area is drawn by black circle, inside the circle there is cumulonimbus cloud appear but it clustered with others separated group of clouds, indicate by red color. It makes the centroid located in the center of them (a red circle with centroid caption in Figure 14). Even, it is located in a position where there is no cloud appearance. Meanwhile, the value that extracted as the feature is the centroid point value around the incident area with specific boundary (black circle in Figure 14). Even though there is cloud appearance inside the scene area, but the centroid didn't appear inside the boundary area that makes no feature listed from the image.

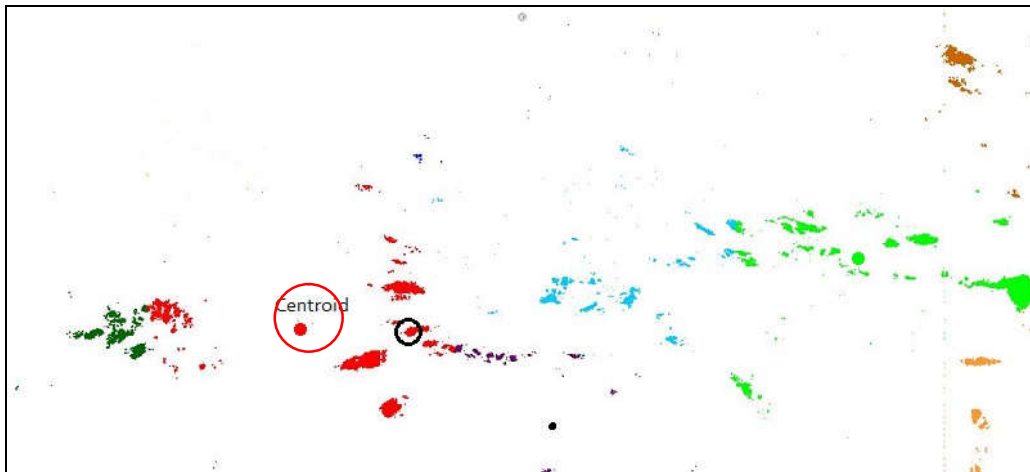


Figure 14. Centroid position condition using meng hee heng k-means

The result from meng hee heng k-means is different with DBSCAN result. DBSCAN produces independent cluster for the group of clouds inside scene boundary area which also makes a centroid (indicates by red point) appear in that area (black circle inside Figure 15). Even though it is only a small group of clouds, it was formed as an independent cluster.

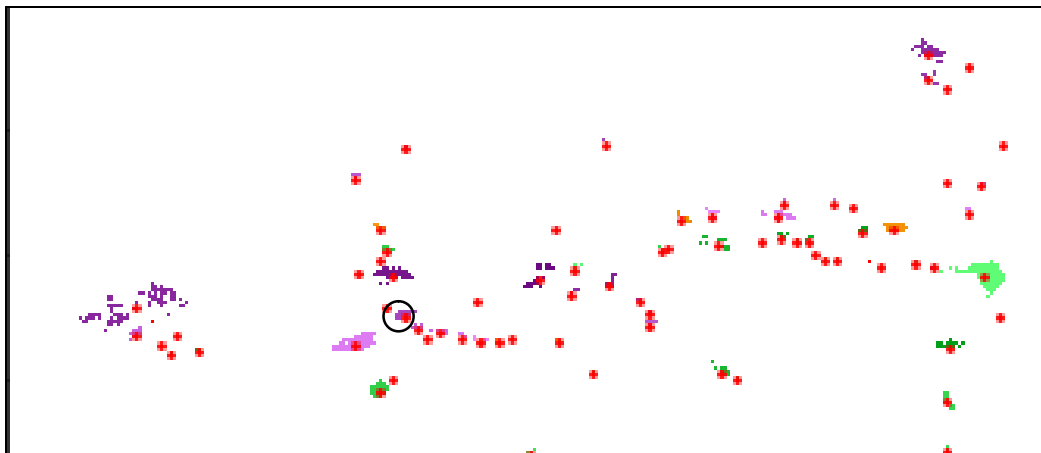


Figure 15. Centroid position condition using DBSCAN

After doing 30 experiments which were also performed into the other three cloud types such as dense, high cloud and also cumulus cloud, DBSCAN obtains at least one centroid point appears. Even when there is no cloud appear at seven sequence images on three cloud types, there will be at least one centroid appear on one of the cloud type. Here are some sample implementations of both algorithms in whirlwind incidents. Table 6 shows the centroid result for DBSCAN algorithm and table 7 shows the centroid result for Meng hee heng k-means.

Table 6. Centroid Result with DBSCAN

Incident	1		2		3		4		5		6		7	
	x	y	x	y	x	y	x	y	X	y	x	y	x	y
Bojonegoro	346.0	263.2	347.7	261.6	344.7	261.3			346.5	253.1	334.5	251.0		
Gresik	395	258.4	391.2	262.2	392.5	259.8	376	263.3						
Madun			348	252	351.3	251.4	346.3	252.7	350.5	249.8	348.7	249.2		
Nganjuk					349.5	257	332.7	251.1	334	255.4	339.3	251.7	340.1	259.7
Nganjuk							351.9	261.1			332.2	259	330.4	258.4
Bali			371.8	268	371.3	268.5	368.7	270.1	362.3	267.5				
Pasuruan							347.3	250.9	338.9	255.3			336.3	258.3
Banjar							291	253.8	298.6	252.6	298.5	252.5	297.4	253.9
Tasik					278.9	253.4	268.9	253.4					271.6	261
Aceh	76.9	76.2	78	84.5	86.2	70.6								

Table 6 shows how many centroid collected and how many missing centroid data in every incident data produced by DBSCAN algorithm. Some of

incident data have more than two missing data. In another hand, table 7 shows the centroid result produced by Meng hee heng k-means. It shows there are many missing data in every incident, even in some incident there is no centroid that can be collected.

Table 7. Centroid Result with Meng hee heng K-means

Incident	1		2		3		4		5		6		7	
	X	y	x	y	x	y	x	y	x	y	x	y	x	y
Bojonegoro														
Gresik			389	252.8										
Madiun											333.01	247.1		
Nganjuk														
Nganjuk														
Bali														
Pasuruan					332.6	251.1								
Banjarnegara														
Tasikmalaya														
Aceh														

Comparing the result shown in table 3 and table 4, we can see that DBSCAN can collect more complete feature than Meng hee heng k-means. DBSCAN data result has 44% missing value, meanwhile Meng hee heng k-means has 96% missing value inside the data result. For every whirlwind historical data, DBSCAN can obtain more feature than meng hee heng k-means, with missing feature value difference 67%. More centroids extracted from every incident, more complete features obtained. More features obtained makes the characteristics or the pattern of the data can be seen. The feature recorded is visualized sequentially. It will make a shape that can be the pattern of the data based on the feature that have been listed. Figure 16 shows the visualtization of both algorithms on whirlwind data occur in Madiun (Table 6 and Table 7 3rd row).

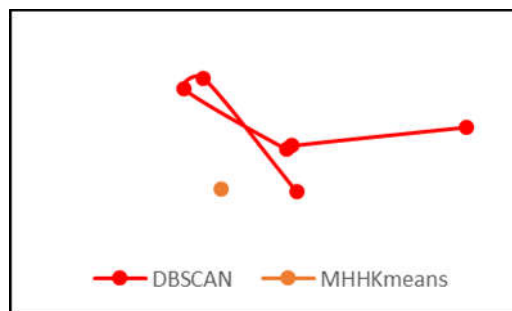


Figure 16. Feature Visialization

Red line shows the visualization of feature pattern produced by DBSCAN, and orange is for meng hee heng k-means. If there is less feature

listed then it is difficult to find the pattern or characteristic of the feature in the data. The completeness of feature result obtained by using DBSCAN indicates that using this method can be affected in characteristics or pattern recognition of the data. More complete result given by DBSCAN can be indication of better performance on segmentation step as the feature extraction process than previous method which is using meng hee heng k-means. Another thing to pay attention is the time execution of the algorithm, meng hee heng k-means need more than 1 second to obtain only 10 cluster, meanwhile DBSCAN only need time 0,05 second to obtain 74 cluster in the same data with the same resolution. Even though 1 second is not a long time, but if we compare it with 0,05 second, the gap is very visible, especially if we run more than 1 data continuously.

6. CONCLUSION

Segmentation plays important role in images processing regardless of what kind of image used. Every image has different information inside that can be useful for various works. The satellite image is one of the images that contain much useful information. It should be properly extracted so the information can be used according to the target. There is so many methods of segmentation images developed since 1994 like amplitude thresholding, component labeling, boundary-based segmentation, region-based segmentation, template matching, and texture segmentation. Recently many researchers were developed and implemented a computational algorithm such as clustering as the segmentation method.

Satellite images also being used so many times not only for a common weather forecast but also used in GIS or disaster management. A lot of researcher interest in processed the satellite image. Satellite images itself has various type and function. Himawari 8 is one of image satellite that has many information inside it i.e. cloud type, cloud temperature, cloud humidity, rainfall potential based on cloud temperature, etc. The appearance and location of clouds in Himawari 8 can be useful information to be extracted. Clouds are forming some groups and located separately. In order to find the location of each group of clouds region-based segmentation was performed in previous research, Himawari 8 was used for predicting whirlwind [12]. In the segmentation process, they perform meng hee heng k-means algorithms. In this study, another clustering algorithm was proposed to improve the segmentation result using the DBSCAN clustering algorithm. DBSCAN provide a better result in classifying groups of clouds. The classifying process is suitable for classify density-based data like groups of clouds. For a whirlwind case which occurs in a specific area and mostly stay (not moving), more detail feature is better. The clearer separation of clouds makes easier to observe each cloud feature and see the pattern or characteristics of the data. We implement DBSCAN and meng hee heng k-means into whirlwind historical data still to observe the feature extraction performance. DBSCAN can record centroid appearance although the observation area is minimized.

The minimization area was done because whirlwind needs more specific area observation. The extracted features are not from one whole image but from a specific range where the incident occurred. Unfortunately, meng hee heng k-means cannot satisfy this requirement. Even the centroid produced by it appears in position with no cloud appear. It makes there are no clouds recorded in the incident area by meng hee heng k-means, although there should be. DBSCAN can give more complete centroid points than meng hee heng k-means that makes DBSCAN performance in order to extract centroid points as the feature is better than meng hee heng k-means. Moreover, DBSCAN need only 0,05 second to finish the clustering, yet meng hee heng k-means need more than 1 second to finish. The exploration of others method in as the segmentation process as the feature extraction should be continue in the future, notably in many more field and more various data case. The utilization of feature obtained from this research also must continue to be developed for the needs of community, for example for the development of whirlwind early warning system.

Acknowledgements

The authors express their appreciation to Indonesian agency for meteorological, climatological and geophysics for providing the Himawari 8 satellite images for this study.

REFERENCES

- [1] Shalini Bhatia, Kumkum Saxena, **Satellite Image Segmentation using Watershed based Algorithms**, *International Conference on Soft computing and Intelligent Systems*, Bali, 2007.
- [2] Song Yuheng, Yan Hao, **Image Segmentation Algorithms Overview**, *arXiv preprint arXiv: 1707.02051*, 2017.
- [3] Dilpreet Kaur, Yadwinder Kaur, **Various Image Segmentation Techniques: A Review**, *International Journal of Computer Science and Mobile Computing*, Vol. 3, No. 5, pp.809 – 814, 2014.
- [4] Chris A Glasbey, Graham W Horgan, **Image analysis for the biological sciences**, *John Wiley & Sons, Inc (New York)*, pp. 93-84, 1995.
- [5] Kaliyamurthie K.P, Parameswari D, **Remote Sensing Imaging for Satellite Image**, *Indian Journal of Science and Technology*, Vol. 8, 2015.
- [6] Peak, James E., and Paul M. Tag, **Segmentation of Satellite Imagery Using Hierarchical Thresholding and Neural Networks**, *Journal of Applied Meteorology*, Vol. 33, No. 5, pp. 605-616, 1994.
- [7] Nailussa'ada, Tri Harson, Achmad Basuki, **Cloud Satellite Image Segmentation using Meng Hee Heng K-Means and DBSCAN Clustering**, *2018 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC)*, Bali, pp. 367-371, 2018.
- [8] Biplab Banerjee, Surender Varma G, Krishna Mohan Buddhiraju, **Satellite Image Segmentation: A Novel Adaptive Mean-Shift Clustering Based**

- Approach**, *2012 IEEE International Geoscience and Remote Sensing Symposium*, Munich, pp. 4319 – 4322, 2012.
- [9] Packyanathan Ganesan, Rajini V, **Assessment of satellite image segmentation in RGB and HSV color space using image quality measures**, *2014 International Conference on Advances in Electrical Engineering (ICAEE)*, Tamilnadu, pp. 1-5, 2014.
- [10] Maria Vakalopoulou, Konstantinos Karantzalos, Nikos Komodakis, Nikos Paragios, **Building detection in very high resolution multispectral data with deep learning features**, *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Milan, pp. 1873-1876, 2015
- [11] Packyanathan Ganesan, V Kalist, B. S. Sathish, **Histogram based hill climbing optimization for the segmentation of region of interest in satellite images**, *2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave)*, Coimbatore, pp. 1-5, 2016.
- [12] Rhoma Cahyanti, Rendra Budi Hutama, Rafi Haidar Ramdlon, Windasari Dwiastuti, Fadilah Fahrul Hardiansyah, Achmad Basuki, **Whirlwind Prediction using Cloud Movement Patterns on Satellite Image**, *2017 International Electronics Symposium on Knowledge Creation and Intelligent Computing*, Surabaya, pp. 252-257, 2017.
- [13] Robert M. Haralick, Linda G. Shapiro, **Image Segmentation Techniques**, *Computer Vision Graphics and Image Processing*, pp. 100-132. 1985.
- [14] Rafael C. Gonzales, Paul Wintz, **Digital Image Processing**, *Addison-Wesley (United States of America)*, Ed. 2, 1987.
- [15] Deepak Jain, Manoj Singh, Dr. Arvind K. Sharma, **Performance Enhancement of DBSCAN Density based Clustering Algorithm in Data Mining**, *International Conference on Energy, Communication, Data Analytics and Soft Computing*, 2017.
- [16] Huan Yu, Wenhui Zhang, **DBSCAN Data Clustering Algorithm for Video Stabilizing System**, *2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC)*, Shenyang, 2013.
- [17] Dayang Sun, Binbin Li, Zhihong Qian, **Research of Vehicle Counting Based on DBSCAN in Video Analysis**, *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, pp. 1523-1527.