TELKOMNIKA, Vol.16, No.3, June 2018, pp. 1409~1415
ISSN: 1693-6930, accredited **A** by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v16i3.6876

Optimization of Parallel K-means for Java Paddy Mapping Using Time-series Satellite Imagery

Alvin Fatikhunnada[1], Kudang Boro Seminar[2], Liyantono[3], Mohamad Solahudin*[4], Agus Buono[5]
[1]Agricultural Engineering, Bogor Agricultural University, Indonesia
[2,3,4]Department of Mechanical and Biosystem Engineering, Bogor Agricultural University, Indonesia
[5]Department of Computer Science, Bogor Agricultural University, Indonesia
*Corresponding author, e-mail: mohamadso@apps.ipb.ac.id

Abstract

Spatiotemporal analysis of MODIS Vegetation Index Imagery widely used for vegetation seasonal mapping both on forest and agricultural site. In order to provide a long-terms of vegetation characteristic maps, a wide time-series images analysis is needed which require high-performance computer and also consumes a lot of energy resources. Meanwhile, for agriculture monitoring purpose in Indonesia, that analysis has to be employed gradually and endlessly to provide the latest condition of paddy field vegetation information. This research is aimed to develop a method to produce the optimized solution in classifying vegetation of paddy fields that diverse both spatial and temporal characteristics. The time-series EVI data from MODIS have been filtered using wavelet transform to reduce noise that caused by cloud. Sequential K-means and Parallel K-means unsupervised classification method were used in both CPU and GPU to find the efficient and the robust result. The developed method has been tested and implemented using the sample case of paddy fields in Java Island. The best system which can accommodate of the extend-ability, affordability, redundancy, energy-saving, maintainability indicators are ARM-based processor (Raspberry Pi), with the highest speed up of 8 and the efficiency of 60%.

Keywords*: agriculture monitoring, high-performance computer, modis, parallel k-means, spatiotemporal analysis*

1. Introduction

In Java Island Indonesia, Agricultural produces such as paddy, maize, soybean, groundnut and mung bean are cultivated in paddy field which divided into two or three cropping seasonal [1-2]. As the fourth-highest rice consumption country after Myanmar, Vietnam, and Bangladesh, Indonesia consume 1.62 kg rice per capita per week in 2014 [3]. Total rice consumption has been rising faster than production, as the growth rate of national rice area and yield has faltered. Based on U.S Department of Agriculture-Foreign Agricultural Service/USDA-FAS [4], Indonesia ranks 3rd in the world in regards to total rice production but has also been the world 7th largest rice importer over the past 5 years, on average requiring over 1.1 million tons of imports per year. Consequently, food security and the pursuit of national rice self sufficiency have become main concerns of the government in Indonesia. In the other hand, the recent modeling of climates reports that there is a potential that may threaten food security in the near future [5].

Crop phenology monitoring at a regional scale can provide useful information for agricultural management to enhance crop yield via irrigation regulation or adjustments in crop cultivation systems [6-7]. Ozdogan *et al*. [8] have reviewed the method and the result of remote sensing in irrigated agriculture by the variant of local climate. Observation satellite that provides continuous data and open access worldwide: Landsat, ASTER, MODIS, MERIS, ACHRR. Some recent studies have been proved that time-series analysis of optical type satellite describe the phenology of terrestrial vegetation which represented the condition and the land used change based on trends and seasonal change [9-16]. In order to reduce noise which caused by clouds and other aerosol problem, discrete wavelet transform (DWT) is used to denoising and extracting the agriculture vegetation information from the terrestrial surface [9],[17-18]. Decision support system for weed control based on a precision agricultural approach that has been done

Received July 25, 2017; Revised February 8, 2018; Accepted March 10, 2018

by Sampurno *et al*. [19] also used MODIS time-series data as the main information of vegetation stage.

Spatiotemporal analysis on time-series imagery in long-term data needs a lot of computational resources and a sophisticated computer to provide dynamic agriculture cropping maps. Meanwhile, monitoring the current condition can be done by that analysis gradually and continuously. Trends in computer architecture exponentially improve and multi-core machines are said to provide high-performance at low-energy cost via multi-threading [20]. In the context of parallel and distributed computing, energy efficient voltage scheduling for multi-core processors is an important issue [21]. The improvement of energy efficiency not only depending on the problem size but also the degree of multi-threading [22]. Low-power consumption computer such as Raspberry Pi (3.5 Watt) can perform as a desktop pc and used for image processing and weed fractal dimension processing [23].

This study investigated the computational method both parallel and sequential of k-means unsupervised classification for long-term spatiotemporal data using both CPU and GPU as a processing unit. Both CPU and GPU have a potential for an efficient outcome, but both processing units have limitation. To optimize those problems, the performance of sequential and parallel k-means using Multi-CPU with raspberry pi 2, Multi-core with Intel Core-i7 3th gen and Nvidia GTX 1060 6Gb GPU were compared. Energy consumption and processing time are considered with minimization goal and maximum processing time constraint is the period of satellite images.

2. Research Method

2.1. Study area

The study area on this research involves paddy field in Java Island which defines by Geospatial Information Agency (BIG) of Indonesia. The method was validated and tested for paddy field in Banten, West Java, Central Java, and East Java Province in Java Island. Ground check location selected by class distribution and the percentage area of paddy field inside the MODIS pixel. Paddy field map as shown in Figure 1 was used as the mask for pattern classification analysis.

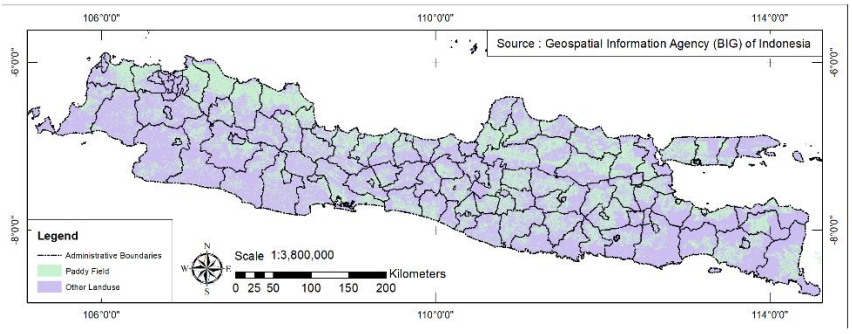


Figure 1. Paddy field maps of Java Island define by Geospatial Information Agency of Indonesia

2.2. Satellite Imagery

The MODIS product used in this study is the Vegetation Indices (VI) Composite 16-day Global 250 m SIN grid V006 or MOD13Q1 product [24], which provided vegetation coverage in Java Island. The Enhanced Vegetation Index (EVI) as in equation (1) is embedded in the MOD13Q1 product. The EVI developed to optimize the vegetation signal with improved sensitivity in high biomass regions and improved vegetation monitoring through a decoupling of the canopy background signal and a reduction in atmosphere influences [11].

$$\mathrm{EVI} = G.\frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{red}}}{\rho_{\mathrm{NIR}} + C_1.\rho_{\mathrm{red}} - \mathrm{C}_2.\rho_{\mathrm{blue}} + L} \quad (1)$$

2.3. Spatiotemporal Analysis

The collection and preprocessing geo-spatial data from MODIS imagery using MODIStsp library which provide by Busetton and Renghetti [25] was conducted. MODIS image from LP DAAC [24] collected from 2010 until 2016 for every 16-days. Image tiles for Java Island located in h28v09 and h29v09. Both of those images were merged to provide single raster image of Java Island. Research procedure involve RAW data acquisition, pre-processing, denoising, pattern classification, and spatial analysis as shown in Figure 2.

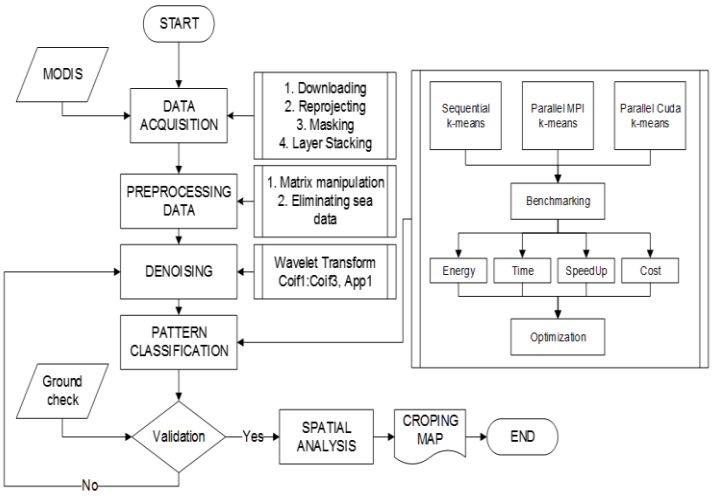


Figure 2. Research framework in spatiotemporal analysis to provide annual cropping maps

2.3. Pre-processing and denoising

Preparation of MODIS image data begins with image mosaicking on tile h28v09 and h29v09. Clipping the image of the merging is done with the help of Java vector map obtained from Geospatial Information Agency (BIG) of Indonesia. The imagery at each time frame is arranged in order to produce raster data with z axis matrix as time domain. The matrix manipulation of raster data in the form of three dimensional matrices begins with the separation of wetland data with the help of wetland rice maps from BIG as Mask. Matrix with X and Y dimensions as location information, while Z as data in time domain is modified into two dimensional matrix. The modified matrix is a data table with information on the dynamics of the vegetation index in each column for each location.

Wavelet transforms are used to reduce the interference caused by clouds and other weather disturbance. Wavelets are employed at each data location to correct the dynamics of the vegetation value in the time domain. Mother wavelet used is a coiflet with expected results in the form of approximation function that describes the data at low frequencies.

$$Wf(a,b) = \int_{-\infty}^{+\infty} f(x) \frac{1}{\sqrt{a}} \Psi\left(\frac{x-b}{a}\right) dx \quad (2)$$

$$f(x) = f_j(x) + g_j(x) + ... + g_1(x) \quad (3)$$

where Ψ is mother wavelet function, $f_j(x)$ is approximation and $g_j(x)$ is detail.

3. Results and Analysis

3.1. Parallel k-means

K-means clustering is used to categorize a data set based on its average value (ISODATA). Categorization of larger data sets takes a lot of time and intensive calculations.

Modification of serial processing into parallel processing by employing MPI library can speed up processing time by optimizing the application of multithreading [26]. The process starts from the initial centroid determination. Then the serialized determination of the data sketch against the centroid of the K groups and the calculation of the new centroid. The process will be done until the number of cluster changes per amount of data smaller than the threshold. Pseudo-code from K-Means serial shows in Algorithm 1.

Algorithm 1 K-Means clustering
1: Select initial centroid from K class
2: **while** d/n data < threshold **do**
3: Find nearest cluster for K class to the centroid
4: Recompute new cluster center
5: Count cluster change as d
6: **end while**

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k}\min(\text{Distance}(m_i, c_j)) \quad (4)$$

From equation 4, k-means algorithm finds k data points on the instance space such that the mean square error (that is, the total distance of all instances to the nearest cluster center) is minimized [27]. The sequential version of the k-means algorithm is depicted in Algorithm 1. Parallel modifications to k-means are performed on data sharing according to the number of processors involved [26]. Calculations are performed on each processor, then collected for results. The proposed method for processing of k-means using MPI is shown in the Figure 3.

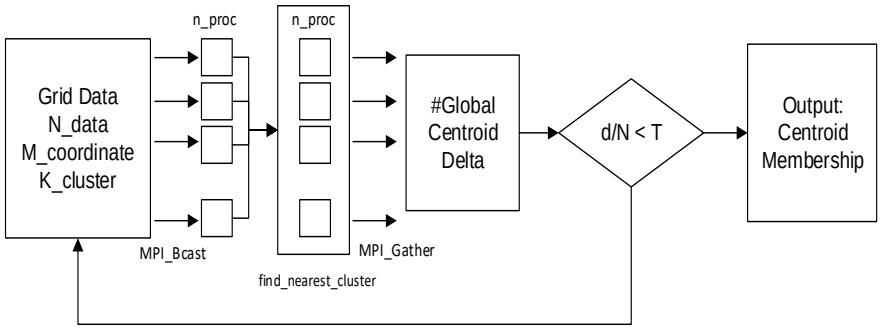


Figure 3. The proposed method for parallel processing of k-means clustering using MPI

3.2. Optimization

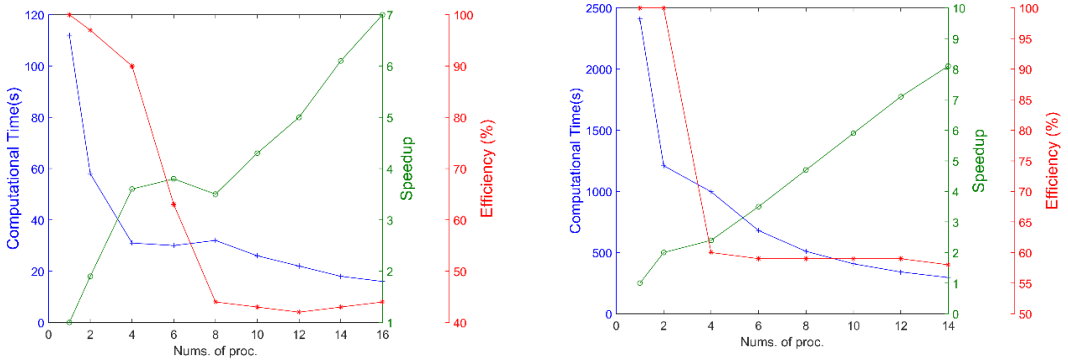
The selection of the best system for image data classification computing is performed based on the performance of each parallel method. Parameters to be considered for optimization calculations include: processing times, speedup, efficiency, durability, extend-ability, affordability, redundancy, power consumption, maintainability, and simplicity. Increased speedup (Sp) on the utilization of parallel processing is approximated by the Amdahl's Law equation. In the equation 5, speedup can be computed from the serial fraction. So that parallel efficiency (η) can be obtained from the equation 6. In this study, the Amdahl's law calculation is performed on RPi and x86 systems because the number of processors on both systems can be determined based on the number of available devices. While on CUDA/GPU, the number of processors is determined by the type of device and performed with single card configuration not dual GPU or SLI.

$$Sp = \frac{1}{x_s + \frac{1 - x_s}{n}} \quad (5)$$

$$\eta = \frac{Sp}{n} \times 100\% \quad (6)$$

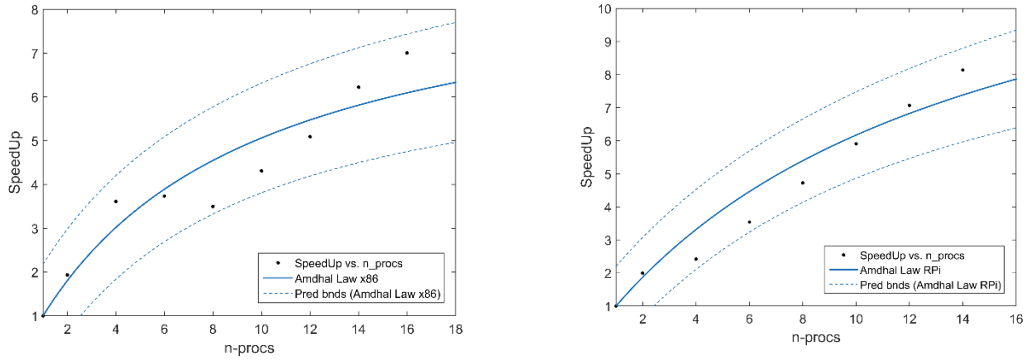
3.3. Parallel benchmarking

Performance test results from parallel computing using raspberry pi (RPi) and x86 computer desktop are shown in Figure 4. The speedup improves by adding more processors/cores both on RPi and x86 system. These results show that the ability of the RPi is still far below x86, but the increase in speed of the RPi is higher. The speedup reaches 8 times with 14 processors working in parallel. While the speedup on x86 reached 7 times with 16 processors working in parallel. The result of the Amdahl's law approach used to assess the performance of parallel processing shows that RPi has higher efficiency than x86 as shown in Figure 5. The result of the performance projection based on the Amdahl's equation as shown in Figure 6 shows the parallel utilization opportunity in RPi higher than in x86.



(a) x86/PC (b) ARM/RPi

Figure 4. Performance of parallel k-means clustering on (a) Intel core i7 and (b) Raspberry Pi 2 using MPI



(a) x86/PC (b) ARM/RPi

Figure 5. Parallel speedup based on Amdahl's Law in (a) Intel core i7, and (b) Raspberry Pi 2 using MPI

The parallel processing test on CUDA/GPU is done with 3 types of devices representing mobile GPU (Quadro 1000M), low end desktop GPU (GT 740), and mid end desktop GPU (GTX 1060 6G). In this study, GPU utilization is limited to single GPU card and does not use SLI or

dual GPU configuration. Then, the comparison is done with the three types of devices used. In addition, the memory capabilities of the GPU are limited, so it needs to be done by comparing the ability of each GPU in processing small and large capacity data. The results of this test are presented in Table 1.

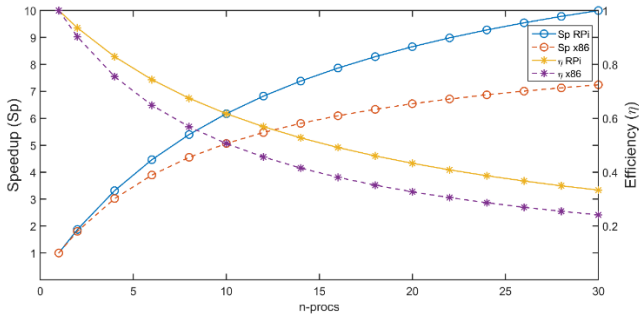


Figure 6. Performance projection based on the Amdahl's equation both in Intel core i7 and Raspberry Pi 2 using MPI

Table 1. Performance of parallel k-means clustering using CUDA accelerator

Devices Name	CUDA Core	Numbers of Data Point	Processing Time (s)
Quadro 1000M	96	2309262	682.83
Quadro 1000M	96	8082417	2396.44
GT 740	386	2309262	563.10
GT 740	386	8082417	1935.25
GTX 1060	1280	2309262	88.88
GTX 1060	1280	8082417	279.70

4. Conclusion

The method for parallel processing of k-means clustering using MPI has been developed, tested and implemented. The utilization of parallel processing in k-means clustering for Java paddy mapping gives some advantages such as faster computational time, more robust system by redundant configuration, and higher efficiency by using an ARM based system. All of those architectures are can be use to perform fast computation for paddy mapping. The performance of parallel k-means can be improved by adding more processors or processing cores. From this research, the best system architecture which can satisfy almost all of the performance indicator is ARM-based processor (Raspberry Pi), which achieved the highest speed up and the efficiency. It is important to highlight that both CPU-based and GPU-based (CUDA) systems can be employed for parallel processing to classify the paddy fields patterns.

Acknowledgement

The authors would like to thank the NASA LP DAAC for making MODIS data available, Geospatial Information Agency of Indonesia for providing paddy field maps in Java Island available for this research. This research was financially supported by Directorate General of Higher Education, Ministry of Research, Technology, and Higher Education Republic Indonesia.

References

[1] FAO. Fertilizer use by crop in Indonesia 2005: 78.
[2] Myers N. Tree-crop based agroecosystems in Java. For Ecol Manage 1986; 17: 1-11. doi:10.1016/0378-1127(86)90071-X.
[3] BPS. Statistical Yearbook of Indonesia 2014. Jakarta: BPS - Statistics Indonesia; 2014.
[4] USDA-FAS. INDONESIA: Stagnating Rice Production Ensures Continued Need for Imports. Foreign Agricultural Service, U.S. Department of Agriculture. USGS; 2012.
[5] Schmidhuber J, Tubiello FN. *Global food security under climate change*. Proc Natl Acad Sci U S A 2007; 104: 19703-8. doi:10.1073/pnas.0701976104.

[6] Wang H, Lin H, Munroe DK, Zhang X, Liu P. Reconstructing rice phenology curves with frequency-based analysis and multi-temporal NDVI in double-cropping area in Jiangsu, China. Front Earth Sci 2016; 10: 292-302. doi:10.1007/s11707-016-0552-9.
[7] Liyantono, Kato T, Yoshida K, Kuroda H. The Influence of El Niño Southern Oscillation on Agricultural Production Sustainability in a Tropical Monsoon Region : Case Study in Nganjuk District, East Java, Indonesia 2012; 74: 65-74.
[8] Ozdogan M, Yang Y, Allez G, Cervantes C. Remote sensing of irrigated agriculture: Opportunities and challenges. Remote Sens 2010; 2: 2274-304. doi:10.3390/rs2092274.
[9] Sakamoto T, Yokozawa M, Toritani H, Shibayama M, Ishitsuka N, Ohno H. A crop phenology detection method using time-series MODIS data. Remote Sens Environ 2005; 96: 366-74. doi:10.1016/j.rse.2005.03.008.
[10] Xiao X, Boles S, Frolking S, Li C, Babu JY, Salas W, et al. Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images. Remote Sens Environ 2006; 100: 95-113. doi:10.1016/j.rse.2005.10.004.
[11] Huete a., Didan K, Miura T, Rodriguez EP, Gao X, Ferreira LG. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens Environ 2002; 83: 195-213. doi:10.1016/S0034-4257(02)00096-2.
[12] Setiawan Y, Liyantono, Fatikhunnada A, Permatasari PA, Aulia MR. Dynamics Pattern Analysis of Paddy Fields in Indonesia for Developing a Near Real-time Monitoring System Using MODIS Satellite Images. *Procedia Environ Sci* 2016; 33: 108-16. doi:10.1016/j.proenv.2016.03.062.
[13] Fatikhunnada A, Permatasari PA, Setiawan Y, Liyantono. *Bfast Change Detection for Agricultural Land Use Changes Analysis in Sidoarjo Regency, East Java, Indonesia*. ACRS 2015-36th Asian Conf Remote Sens Foster Resilient Growth Asia, Proc 2015.
[14] Kim S-R, Prasad AK, El-Askary H, Lee W-K, Kwak D-A, Lee S-H, et al. Application of the Savitzky-Golay Filter to Land Cover Classification Using Temporal MODIS Vegetation Indices. Photogramm Eng {&} Remote Sens 2014; 80: 675-85. doi:10.14358/PERS.80.7.675.
[15] Nugroho AP, Okayasu T, Hoshi T, Inoue E, Hirai Y, Mitsuoka M, et al. Development of a remote environmental monitoring and control framework for tropical horticulture and verification of its validity under unstable network connection in rural area. *Comput Electron Agric* 2016; 124: 325-39. doi:10.1016/j.compag.2016.04.025.
[16] Martinez B, Gilabert MA. Vegetation dynamics from NDVI time series analysis using the wavelet transform. Remote Sens Environ 2009; 113: 1823-42. doi:10.1016/j.rse.2009.04.016.
[17] Setiawan Y, Rustiadi E, Yoshino K, Liyantono E, Effendi H. Assessing the Seasonal Dynamics of the Java's Paddy Field Using MODIS Satellite Images. *ISPRS Int J Geo-Information* 2014; 3: 110-29. doi:10.3390/ijgi3010110.
[18] Shao Y, Lunetta RS, Wheeler B, liames JS, Campbell JB. An evaluation of time-series smoothing algorithms for land-cover classifications using MODIS-NDVI multi-temporal data. Remote Sens Environ 2016; 174: 258-65. doi:10.1016/j.rse.2015.12.023.
[19] Sampurno RM, Boro Seminar K, Suharnoto Y. Weed Control Decision Support System Based on Precision Agriculture Approach. *TELKOMNIKA (Telecommunication Comput Electron Control* 2014; 12: 475. doi:10.12928/TELKOMNIKA.v12i2.1982.
[20] Tseng C, Figueira S. *An analysis of the energy efficiency of multi-threading on multi-core machines*. 2010 Int Conf Green Comput Green Comp 2010 2010:283-90. doi:10.1109/GREENCOMP.2010.5598301.
[21] Mishra A, Tripathi AK. Energy efficient voltage scheduling for multi-core processors with software controlled dynamic voltage scaling. Appl Math Model 2014; 38: 3456-66. doi:http://dx.doi.org/10.1016/j.apm.2013.12.009.
[22] Lien H, Natvig L, Al Hasib A, Meyer JC. Case Studies of Multi-core Energy Efficiency in Task Based Programs. In: Auweter A, Kranzlmüller D, Tahamtan A, Tjoa AM, editors. ICT as Key Technol. against Glob. Warm. Second Int. Conf. ICT-GLOW 2012, Vienna, Austria, Sept. 6, 2012. Proc., Berlin, Heidelberg: Springer Berlin Heidelberg; 2012, p. 44-54. doi:10.1007/978-3-642-32606-6_4.
[23] Suriansyah MI, Sukoco H, Solahudin M. Weed detection using fractal-based low cost commodity hardware Raspberry Pi. Indones J Electr Eng Comput Sci 2016; 2: 426-30. doi:10.11591/ijeecs.v2.i2.pp426-430.
[24] Didan K. MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006 2015. doi:10.5067/MODIS/MOD13Q1.006.
[25] Busetto L, Ranghetti L. MODIStsp: An R package for automatic preprocessing of {MODIS} Land Products time series. Comput Geosci 2016; 97: 40-8. doi:http://dx.doi.org/10.1016/j.cageo.2016.08.020.
[26] Liao Q, Yang F, Zhao J. An improved parallel K-means clustering algorithm with MapReduce. Int Conf Commun Technol Proceedings, ICCT 2013:764-8. doi:10.1109/ICCT.2013.6820477.
[27] Kucukyilmaz T. Parallel K-Means Algorithm for Shared Memory Multiprocessors. J Comput Commun 2014:15-23.