

Cloud Classification Using Ground Based Images Using CBIR and K-Means Clustering

Gujanatti Rudrappa¹, Nataraj Vijapur², Sushant Jadhav³ and Prabhakar Manage⁴

^{1,3,4}Department of Electronics and Communication Engineering, KLE Dr. M.S.

Sheshgiri College of Engineering and Technology, Belagavi, India

²Department of Electronics and Communication Engineering, RV College of Engineering and Management, Bangalore, India

ABSTRACT

Artificial Intelligence (AI) and especially Machine learning (ML) is finding to be useful in many tasks that are simple to carryout to complex tasks that are found to be challenging in nature. One such application of ML is in classification of images. In this paper an attempt to blend the application of unsupervised ML (k-means clustering) approach along with content based image retrieval (CBIR) approach is presented to classify clouds. K-means is a simple approach which can be applied for image classification, also k-means easily adapts to new examples of classification. An attempt is made to combine the features of k-means and CBIR to classify the cloud images. It is performing a double check on the cloud image being classified. Clustering is included with CBIR to obtain an easy retrieval of cloud image. Three categories are chosen for classification - low level clouds, high level clouds and medium level clouds. The classification of clouds is achieved with the help of ground based images (or whole sky images). High resolution of ground based images can be obtained with the help of new high resolution cameras. These ground based images are processed to classify the clouds present in the images into the three categories as mentioned above. Ground based images captured by ground based cameras provide better ground truth. The results find its application in various domains such as agriculture, aviation, military, and various meteorological applications.

KEY WORDS: ARTIFICIAL INTELLIGENCE, CBIR, CLOUD CLASSIFICATION, K-MEANS CLUSTERING, MACHINE LEARNING.

INTRODUCTION

In various domains and application such as aviation, agriculture, etc., (Rajendrakumar et al 2019) the usage of ground based images is getting popular day by day and are turning out to be useful in those domains and application. We can obtain high quality images with the help of ground based image acquisition systems. The

low level clouds can be classified in a much better way as the ground based imagers face the sky to capture the images of the sky and clouds. The images captured can be used for cloud detection and classification. But this also has some challenges as the cloud shape, size and intensity varies widely (Liu et al., 2014) also lighting conditions add to the challenges of cloud classification (Dev et al., 2016).

ARTICLE INFORMATION

*Corresponding Author: rudraguj@gmail.com

Received 9th Oct 2020 Accepted after revision 28th Dec 2020

Print ISSN: 0974-6455 Online ISSN: 2321-4007 CODEN: BBRBCA

Thomson Reuters ISI Web of Science Clarivate Analytics USA and Crossref Indexed Journal



NAAS Journal Score 2020 (4.31)

A Society of Science and Nature Publication, Bhopal India 2020. All rights reserved.

Online Contents Available at: <http://www.bbrc.in/>

Doi: <http://dx.doi.org/10.21786/bbrc/13.13/13>

proposed that the intensity information in the image can be used for image restoration and classification to conduct 2-D analysis on images (Parameshchhari, B. D et al 2019). Spectral features, textural features and contextual features can be used for classification of images (Haralick et al., 1973), which can also be applied for classification of clouds. We may achieve the cloud classification using satellite images or ground based images with the aid of image processing and ML methods (Haralick et al., 1973; Mahrooghy et al., 2012; Rudrappa & Vijapur, 2019). In, the present paper we propose the cloud classification using the ground based images by blending the Content Based Image Retrieval (CBIR) method along with k-means clustering method.

The classification of clouds in the present proposal is based in classifying the clouds into Low Level Clouds (LLC), Medium Level Clouds (MLC) and High Level Clouds (HLC). HLC basically fall at a distance of 6000m above ground level. Cirrocumulus, Cirrus and Cirrostratus fall under HLC, MLC are within 2000m to 6000m above ground level. Altopcumulus and Altostratus fall under MLC. LLC fall in between 0m to 2000m from the ground level. Stratocumulus, Stratus and Nimbostratus fall under LLC.

Literature Review: In (Zhang et al., 2018) used deep CNN for cloud classification using ground based images and attained accuracy up to 90.00%. In (Dev et al., 2016) authors proposed the use of supervised segmentation framework for ground-based sky/cloud images based on systematic analysis of different colour spaces and components using partial least-squares regression (PLS). In (Shi et al., 2017) the authors presented the use of Deep Convolutional Activations-Based Features (DCAFs) for classification of clouds using ground based images; they used dataset from Keil and SWIMCAT DBs. Ground based images were used in (Moughyt et al., 2015) for identifying the cloud cover with the help of Otsu's method and Multi-objective optimization algorithms.

For detection of clouds Super Pixel Segmentation (SPS) method was used in (Liu et al., 2014). In (Mahrooghy et al., 2012) authors propose to use Link-based Cluster Ensemble for the purpose of cloud classification and precipitation estimation. In (Liu et al., 2012) the authors propose a novel descriptor - Illumination-Invariant Completed Local Ternary Pattern (ICLTP); which tries to overcome the challenges posed in cloud classification using ground based images because of illumination presented in the images.

For automatic classification of clouds in (Christodoulou et al., 2003) the authors presented to use multifeature texture analysis and modular neural networks. Satellite images from geostationary operational environmental satellite (GOES) 8 data were used to build and test the cloud classification method in (Tian et al., 1999). In (Tian et al., 1999) the authors proposed to use Singular Value Decomposition (SVD) and Wavelet Packet (WP) image transforms to gain the special features related to spectral and textural aspects. In (Buch et al., 1995) the

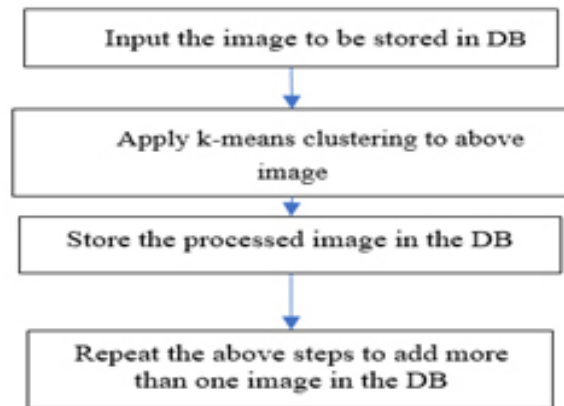
author proposed to use binary decision trees to identify different types of clouds based on cloud vector features and using the whole sky imagers data.

Dataset: The data for the proposed work is taken from Singapore High-dynamic-range Whole sky IMagingSEGmentation Database (Dev et al., 2018). The dataset consists of three types of cloud images for high, medium and low level clouds. 52 images are present for each of the cloud types. Each image is of 500 X 500 pixels in size. The ground truth of these images was done with the help of experts from Singapore Meteorological Services. This available dataset is used for testing the classification results.

METHODOLOGY

Two different methods can be used for retrieving an image Text Based Image Retrieval (TBIR) and Content Based Image Retrieval (CBIR). In the present system we propose to use CBIR for retrieving and image from the DB (DB) created for the purpose of cloud classification. Prior to creation of the DB for CBIR we need to create a DB. The process of DB creation is given in (Figure 1). We can store the images without any processing or apply some preprocessing and then store the image(s) in the DB. We choose the second option of preprocessing the images and storing it in the DB.

Figure 1: Flowchart for storing image(s) in DB which will be later used for CBIR for classification of clouds



The images stored in the DB will help us to classify the cloud present in the ground based images. In order to preprocess the image, we use k-means clustering. Once the image is passed through the k-means process the cloud image chosen is stored in the DB. The number of images to be stored in the DB can vary. (Figure 1) shows the flowchart for storing image(s) in the DB which will be later used for classification of clouds. (Figure 2) shows a sample image that is stored in the DB for reference, (Figure 3) shows the processed image after k-means is applied to image in (Figure 2). (Figure 2) shows a sample image that is stored in the DB for reference, (Figure 3) shows the processed image after k-means is applied to image in (Figure 2).

For classifying cloud present in a given ground based image we provide the image as input to the classification system. When the input is applied to the system, k-means clustering is applied to the input image. After the input image is classified, we estimate the Euclidian distance for each of the image stored in the DB. Euclidian distance equation is given in (1). The smaller the value of Euclidian distance the closer is the input image to the image in the DB. Euclidian distance is utilized as it suits for distance calculation in majority if images. (Figure 5) shows the flow chart for classification of the cloud present in the input image provided to the system.

Figure 2: Sample image that is stored in the DB for reference



Figure 3: Processed image after k-means is applied to image in (Figure 2).



The statistical parameters used for cloud classification is below.

The 1st order histogram in any image is given (1).

$$\text{Probability of pixel } k = nk/N \tag{1}$$

Where,

- k – is pixel intensity value
- nk – number of pixels with intensity value k
- N – number of pixels in the image

Mean value of brightness pertaining to an image is provided by equation (2).

$$\bar{k} = \sum_{k=0}^{L-1} [k (\text{Probability of } k)] \tag{2}$$

Standard Deviation is used for contrast details of the image and is given in equation (3).

$$\sigma_k = \sqrt{\sum_{k=0}^{L-1} (k - \bar{k}) (\text{Probability of } k)} \tag{3}$$

Intensity level distribution is obtained from energy (E) information present in the image which is given in equation (4)

$$E = \sqrt{\sum_{k=0}^{L-1} (\text{Probability of } k)^2} \tag{4}$$

Similarity in the image under test is obtained based on equation (5).

$$ED = \sqrt{\sum_{k=0}^{M-1} [x_{in} - x_{db}]^2} \tag{5}$$

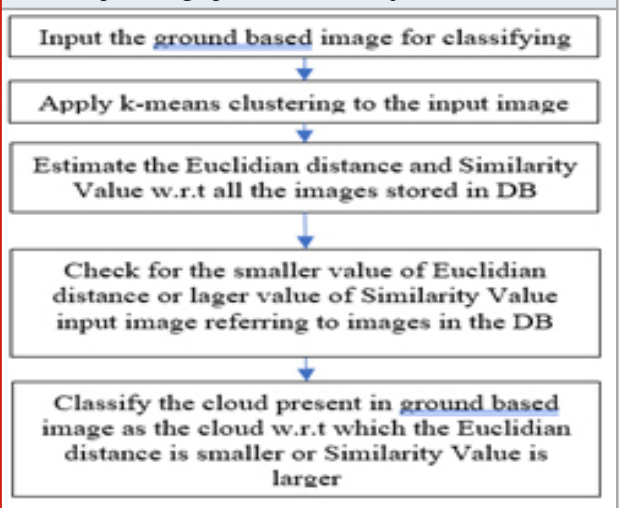
- ED – Euclidian distance
- xin – coordinate of input image
- xdb – coordinate of DB image

Based on the value of similarity we can classify the cloud present in the ground based image. The cloud is classified by considering the largest value of similarity measured based on equation (6).

$$SV = \sum_{k=0}^{M-1} [x_{in} * x_{db}] \tag{6}$$

For implementation of the proposed work Matlab 2017 Ra was used. It was used on system with 8GB RAM, Windows Operating System, with 8th generation i3 processor.

Figure 4: Flow chart for classification of the cloud present in the input image provided to the system



RESULTS

Here we present few results of applying the machine learning approach for classifying the clouds present in

the ground based images. (Figure 6) shows a sample ground based image related to test image that is given as input for classifying the cloud. For simplicity only one image for each cloud is stored in the DB. In (Figure 7) we can observe the estimated values of Euclidian Distance

Figure 5: Input test image in which cloud classification has to be obtained (Cloud present in the image is High)



and Similarity Values for the test image which is given as input to the system. In (Figure 7) the highlighted values show the values corresponding to Euclidian Distance and Similarity Value which is close to High Level Cloud in this case considered.

The ground based images used in the system are taken from the dataset Singapore Whole Sky Imaging Segmentation Singapore High-dynamic-range Whole sky IMagingSEGmentation DB (SHWIMSEG) (Dev et al., 2018) and in weatherscapes website (www.weatherscapes.com, 2019). The dataset SHWIMSEG consists of fifty-two images present for each of the clouds mentioned above. (Table 1) shows the overall accuracy of the system for different conditions considered with respect to the process mentioned in (Figure 1). Here three conditions are considered and presented in (Table 1). (Table 2) provides the comparison between different classification approaches used.

Table 1. Overall accuracy of cloud classification for different number of cloud images stored in DB.

No: of images stored in DB for reference (Figure 1)	No: of correctly classification of clouds in the ground based images referring to SHWIMSEG dataset			
	High Level Clouds	Middle Level Clouds	Low Level Clouds	Overall accuracy of the system
1	49	47	41	87.82%
2	49	47	48	92.31%
3	49	47	10	67.31%

Table 2. Comparison of cloud classification approaches with respect to their accuracy of classification

Author(s)	Method used for cloud classification	Performance measure	Result (%)
J. Zhang et al. (2018)	CNN	Accuracy	83.00 - 90.00
A. Kazantzidis et al. (2017)	K-Nearest Neighbour	Accuracy	87.9
Our method	Using CBIR and k-means	Accuracy	67.31- 92.31

CBIR and k-means are simple approaches that can be used for image retrieval and classification, also both methods can adapt to new examples for image retrieval and classification. The accuracy of 67.31% which is less as compared to others is observed. This variation in accuracy of classification even when more images are present in the database is due to the variation in the intensities of the images which are stored in the database. If the images stored in databased have less intensity variations then it results in the reduced accuracy in image classification. This can be avoided in our system by carefully choosing the images to be stored in database.

CONCLUSION

Cloud classification could be achieved by using unsupervised ML approach along with image retrieval approach. The classification of clouds was done as high-level clouds, middle-level clouds and low-level clouds. One of the challenging aspects observed was that shape and size of the clouds in the ground based images cannot be used for classification of clouds. Different intensity present in the images also make the classification a more challenging aspect. Also, to improve the overall accuracy of cloud classification we can consider two major aspects;

first one being proper selection of cloud images to be stored in DB and second one is to put more images in DB prior to testing the classification of clouds present in the image. We can test the system to classify the clouds into more categories such as classifying the clouds as cirrus clouds, stratus clouds, etc., and also estimate the cloud cover and use the result in precipitation estimation or prediction along with the other meteorological information.

REFERENCES

- Andrews, H.C., and Enloe, L.H. (1974) Special issue on digital picture processing. *Proc. IEEE*, 60 : 768-894.
- Andrews, H.C., Tescher, A.G. and Kruger, R.P. (1972) Image processing by digital computer. *IEEE spectrum*, 9(7) : 20-32.
- Buch Jr, K.A. and Sun, C.H. (1995) Cloud classification using whole-sky imager data. Sandia National Labs, Albuquerque, NM, United States.
- Christodoulou, C.I., Michaelides, S.C. and Pattichis, C.S. (2003) Multifeature texture analysis for the classification of clouds in satellite imagery. *IEEE transactions on geoscience and remote sensing*, 41(11) : 2662-2668.
- Dev, S., Lee, Y.H. and Winkler, S. (2016) Color-based segmentation of sky/cloud images from ground-based cameras. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(1) : 231-242.
- Dev, S., Savoy, F.M., Lee, Y.H. and Winkler, S. (2018) High-dynamic-range imaging for cloud segmentation. *arXiv preprint arXiv:1803.01071*.
- Haralick, R.M., Shanmugam, K. and Dinstein, I.H. (1973) Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6) : 610-621.
- Liu, S., Wang, C., Xiao, B., Zhang, Z. and Shao, Y. (2012) Illumination-invariant completed LTP descriptor for cloud classification. *5th International Congress on Image and Signal Processing* : 449-453.
- Liu, S., Zhang, L., Zhang, Z., Wang, C. and Xiao, B. (2014) Automatic cloud detection for all-sky images using superpixel segmentation. *IEEE Geoscience and Remote Sensing Letters*, 12(2) : 354-358.
- Mahrooghi, M., Younan, N.H., Anantharaj, V.G., Aanstoos, J. and Yarahmadian, S. (2012) On the use of a cluster ensemble cloud classification technique in satellite precipitation estimation. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(5) : 1356-1363.
- Parameshachari, B. D., Rashmi P. Kiran, P. Rashmi, M. C. Supriya, Rajashekarappa, and H. T. Panduranga. "Controlled partial image encryption based on LSIC and chaotic map." In *ICCS*, pp. 60-63. 2019.
- Rajendrakumar, Shiny, and V. K. Parvati. "Automation of irrigation system through embedded computing technology." In *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, pp. 289-293. 2019.
- Rudrappa, G. and Vijapur, N. (2019) Intelligent Methods Used for Obtaining Weather Derivatives: A Review. *Sciences*, 4(6) : 144-148.
- Shi, C., Wang, C., Wang, Y. and Xiao, B. (2017) Deep convolutional activations-based features for ground-based cloud classification. *IEEE Geoscience and Remote Sensing Letters*, 14(6) : 816-820.
- Tian, B., Shaikh, M.A., Azimi-Sadjadi, M.R., Haar, T.H.V. and Reinke, D.L. (1999) A study of cloud classification with neural networks using spectral and textural features. *IEEE transactions on neural networks*, 10(1) : 138-151.
- Wan, X. and Du, J. (2020) Cloud Classification for Ground-Based Sky Image Using Random Forest. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43 : 835-842.
- Zhang, J., Liu, P., Zhang, F. and Song, Q. (2018) CloudNet: Ground-based cloud classification with deep convolutional neural network. *Geophysical Research Letters*, 45(16) : 8665-8672.