Fuzzy statistics-based affinity propagation technique for clustering in satellite cloud image

Govindan Devika* and Sudhaman Parthasarathy

*Department of CS & IT, Kalasalingam Academy of Research and Education, Krishnankoil, India; ‡Department of Computer Applications, Thiagarajar College of Engineering, Madurai, India

ABSTRACT
Satellite-remote-sensing technologies have set off improvements in atmospheric research and developments of new tools in prospect of discovery to monitor. Classification of cloud image through satellite is well recognized as a valid approach in many climatic and environmental analyses. A multispectral cloud classifier was implemented to automate the interpretation of Kalpana-1 satellite image. In this paper, a novel image-clustering method, grounded on fuzzy statistics-based affinity propagation (FS-AP), has been proposed. It entails two steps: feature extraction and clustering. The objective is to study the volatility of the FS-AP for the classification of satellite cloud images optimally. Methods for classifying cloud type from satellite images are difficult in terms of efficiency and accuracy. Results show the effectiveness of the proposed technique for classification of satellite cloud image by comparing with fuzzy K-means and affinity propagation.

KEYWORDS
Satellite cloud image; fuzzy statistics-based affinity propagation (FS-AP); image classification; fuzzy sets; clustering

Introduction

Cloud detection is a vital issue in analyzing data related to geophysical, geomorphological and meteorological fields, which were obtained from remotely sensed images. Among various techniques, clustering techniques are found to be more suitable and widely used in classification to partition multispectral feature spaces for extracting clusters of patterns, which are associated with cloud types (Bandyopadhyay, Maulik, & Mukhopadhyay, 2007; Chi, Qian, & Benediktsson, 2008; Maulik & Saha, 2009). Classifying clouds is very much essential in order to automatically extract valuable information on occurrence of clouds and its types.

A classification problem is normally said to occur when an object needs to be assigned into a set-off predefined class or group, depending on a number of known attributes close to that object. A problem, which arises while classifying the clouds based on satellite images in a homogeneous region, is possible when we cluster its pixels as per its intensity value. For satellite images, the most prominently used channels for cloud classification are found to be in the visible and the infrared bands (Heinle, Macke, & Srivastav, 2010; Tag, Bankert, & Brody, 2000). For the visible bands, classifications of clouds depend on its property; for example, the optical depth, which could be used to differentiate the clouds.

In this paper, the issues and challenges associated with the cluster formation have been discussed. To identify the clusters for a given dataset, it is common to define similarity measure. This similarity measure does the process of placing the data close to one another so as to form a group, thereby generating different clusters. In this process, Fuzzy K-means clustering has been used to manage this issue. It should be noted that a fuzzy statistical similarity (FSS) measure is a key issue to extract cloud type information from a satellite image.

A fuzzy set consists of elements with degrees of membership. An element of a fuzzy set can be a full member (100% membership) or a partial member (between 0 and 100% membership). In other words, the membership value assigned to an element is no longer restricted to just 0 or 1, but can be any value ranging from 0 to 1. The mathematical function, which defines the degree of an element’s membership in a fuzzy set, is called as a membership function. Let “X” be a universe of discourse, whose generic elements are denoted by x. Thus, X = {x}. Membership in a classical set A of X is often viewed as a characteristic function xA from X to {0, 1} such that x A (x) = 1 if and only if x belongs to A. A fuzzy set B in X is characterized by a membership function fB, which associates with each x as a real number in [0, 1]. fk represents the “grade of membership” of x in B. The closer the value of fk (x) is to 1, the more x belongs to B.

Fuzzy set theory provides a classical representation for geographical information, thereby providing a better understanding of the images and this cannot be well defined by a single class. The fuzzy sets are
Normally represented by elements; wherein, the cloud cover classes are considered fuzzy sets and the pixels used to represent the cloud cover classes constitute the elements of the fuzzy sets. Each pixel is attached to a group of membership grading so as to indicate the extent to which the pixel belongs to a certain class. Pixels with class mixture or in intermediate conditions can be described by their membership grades. For instance, if a ground cell contains two cover-types, namely, “soil” and “vegetation”, it may have two membership grades, indicating the extents to which it is associated with the two classes (Foody, 1992, 1996; Wang, 1990).

In conventional fuzzy clustering, the clustering solely depends on the initial sequence of samples and fuzzy K-means requires the given cluster numbers. This has two main problems. They are: (i) Attraction of the centroid toward the outliers; (ii) Its inability to differentiate outliers from non-outlier. Both these two problems are referred as “noise sensitivity”. Moreover, conventional fuzzy schemes are based on low and high paradigms (Nedeljkovic, 2011). Most of the cluster analysis results get trapped in local optimizations, which increases randomness and creates difficulty in obtaining accurate results. In practice, the basic approach for classifying the clouds obtained from remote-sensing images is to form several groups by clustering their pixels in the spectral domain. In this paper, the identification of number of clusters is reported as searching some suitable number of cluster centers. In Fuzzy K-means clustering, the data points belong to one cluster having certain probability. Clustering of images is done based on similarity measures. Such clustering along with statistical techniques is used to extract the spectral and textural features from the satellite image.

There are previous research works (Yang et al., 2010), which address the classification of multispectral remote-sensing images; however, our present research takes a specific analysis on clustering of satellite cloud images. Such a research was motivated due to the growing number of spectral channels and requirement for analyzing a large quantum of information so as to classify satellite images with reasonable. Traditional classification methods appear to be less efficient when a large dataset needs to be analyzed and unsupervised clustering techniques are required to be considered for data analysis.

In remote-sensing images, based on the spatial resolution of the images, a pixel can be used to represent a mixture of land covers (Wang & Mo, 2004). Such images cannot be normally well described by a single class. Previous research works on remote-sensing images have extensively used the fuzzy approach for the classification problems. These problems would use the one-to-many relation of a pattern with the related information classes. As all these images comprise different patterns; a single class cannot be used to better describe these patterns, and thus fuzzy clustering was developed and applied for classification of satellite cloud images in this research.

**Literature survey**

For remote-sensing images, clustering implies a grouping of pixels in a multidimensional space (Wang & Mo, 2004). For a given image, normally, pixels from a particular cluster would be spectrally similar. A similarity measurement should be carried out to quantify the relationship between the pixels. Many similarity measurement techniques have been proposed in the literature. However, all these techniques use the clustering procedures as they are found to be simpler ways of measuring distance for images captured from multidimensional spaces. There are also similarity measures used in the literature for image analysis namely – Euclidian distance, correlation coefficient, spectral angle, encoding and matching, and spectral information divergence. – Each of these measures has its own strengths and limitations.

Classification of clouds based on satellite remote-sensing data has gained momentum as it is found useful for analyzing the climatic conditions and generates weather reports (Key, Maslanik, & Barry, 1989). Classification of images is basically carried out to extract meaningful information from these images. Such a classification is not a simple task (Li, Zang, Zhang, Li, & Wu, 2014; Mountrakis, Im, & Ogole, 2011). Classification of clouds is considered important and useful for many real-time applications namely weather forecast, surface identification, and climate-change detection (Elhag & Bahrawi, 2014; Surya & Simon, 2013).

Surya and Simon (2013) have designed a cloud-detection algorithm, which would perform color transformation on any given input image to generate a ratio image using spectral-image-rationing technique, and finally clusters the ratio image using Fuzzy C-means clustering for detecting clouds automatically. Banerjee et al. (2015) have addressed the problem of unsupervised land-cover classification of remotely sensed multispectral satellite images from the perspective of cluster ensembles and self-learning. Previous research works have considered remote-sensing technique to be extremely important for classification of cloud satellite images (Gómez-Chova, Camps-Valls, Calpe-Maravilla, Guanter, & Moreno, 2007; Mondal, Khare, & Kundu, 2013). An exemplar-based clustering algorithm, i.e. affinity propagation (AP), was proposed by Guan, Shi, Marchese, Yang, and Liang (2010). Though there are numerous clustering techniques in the literature (Klančar & Škrjanc, 2015; Tian, Azimi-Sadjadi, Haar, & Reinke, 2000),
there is no way to find a suitable method/technique for the purpose of clustering of images for a given dataset. Moreover, a single clustering technique cannot handle all the types of images with varying sizes and shapes (Fred & Jain, 2005).

A large number of cloud images is extracted from the satellite positioned on the earth on day-to-day basis; however, the cloud covers and its associated clarity primarily depend on the climatic conditions and the geographical location from where the data have been gathered (Gómez-Chova, Camps-Valls, Bruzzone, & Calpe-Maravilla, 2010). Extracting as many features as possible would be of great use to design forecasting models. In such a scenario, prediction could be improved using its properties namely cloud motion, cloud formation and cloud texture (Rathnayake, Premaratne, & Sonnadara, 2011).

Previous research works have used the following radiance threshold, radiative transfer model or statistical techniques, and in some cases, the spectral and the textural features in the images for identifying clouds from satellite images (Alonso, Sanz, & Malpica, 2007; Bandyopadhyay, Maulik, & Mukhopadhyay, 2007). A previous research suggests that the usage of remote-sensing dataset for cloud satellite image classification triggers a reliability issue during validation of the collected samples, if the sample size is very less (Su & Du, 2014). Based on the features associated with an image, the classifiers are expected to categorize the images to different types of clouds [Gurve and Sarup, 2012]. Previous research works in the literature have used clustering method to classify the cloud covers obtained from remote-sensing images (Kuril, Saini, & Saini, 2013).

These days, the concept of fuzzy logic is widely used in pattern recognition and related research. More specifically, it is mostly used in remote-sensing image classification. The significance of fuzzy logic is that it makes no assumption about statistical distribution of the data; however, it provides an in-depth analysis of images by means of fuzzy classification (Choodarathnakara, Ashok Kumar, & Shivaprakash Koliwad, 2012). Image classification is the process of categorizing all the pixels in an image automatically into a finite number of cloud cover classes (Rokach & Maimon, 2005). It is very often found used in quantitative data analysis of remote-sensing images to describe ground cover types or material classes.

Jabari et al., (2013) have observed in their research that object-based image analysis is found suitable for image classification. However, they also indicate that the problems underlying such methods are that, in the case of classification of high-resolution images, one has to take into account the uncertainties in the position of object borders in satellite images, and the multiplex resemblance of the segments to different classes. To address this issue, Jabari & Zhang (2013) used fuzzy logic for image classification, as it does image analysis without including thresholds in the class-assignment process. It is observed that the geostationary satellites are capable of providing cloud information covering wide areas, thus enabling weather forecasting using clouds at the earliest possible time than actual schedule.

Fuzzy K-means clustering has been broadly used in satellite images, and previous research works have proposed several algorithms, which were grounded on Fuzzy K-means. However, these algorithms are found to have limitations during satellite-image analysis. They are: (i) Specifically, the use of Euclidean distance measure is not applicable for satellite-image clustering. This is due to the fact that the scatter diagram of multispectral satellite data tends to hyper-ellipsoid distributions in the feature space, due to uncertainty and existence of mixed pixels. This can be moderated by considering kernelized versions of fuzzy K-means, known as the kernel fuzzy K-means (Zhang, Wu, & Pu, 2007). (ii) Initial centers of K-means and fuzzy K-means are defined randomly, thus leading to uneven clustering results and requiring multiple trials for obtaining reasonable results.

**Hence, in this research work, we have used the Fuzzy K-means clustering based on AP technique for classification of satellite cloud images.**

In recent years, artificial intelligence-based methods, such as artificial neural networks (Rathnayake et al., 2011) and support vector machine (Mountrakis et al., 2011), have been applied for cloud classification. These methods have improved the accuracy of cloud classification by using machine learning to extract the inherent features of training samples.

In AP, the clusters data points are based on similarity measures and here, it is considered that all data points can be equally viewed as exemplars. This algorithm aims to find several exemplars such that the sum of the similarities between the data points and the corresponding exemplars is maximized. It works well for pattern recognition and image classification. However, it has not gained significance for satellite images. Here, cluster analysis is not possible unless a meaningful measure of similarity between pairs of points has been established. To address this problem, in this paper, an improved similarity measure integrating fuzzy statistics with AP has been proposed, and this approach is referred to in this paper as fuzzy-statistics-based AP (FS-AP). The proposed method considers that all data points can be equally treated as initial exemplars. First, according to the characteristics of multispectral images, we propose fuzzy mean deviation and then develop a FSS measure while evaluating the similarity between two pixel vectors. We iteratively merge cluster centers to extract cloud cover information by FSS. The present work aims at imposing a new method for cloud detecting and cloud type classifying of multispectral images acquired using Kalpana-1 satellite.
Description of dataset

In the surface-meteorological-observation method, the cloud forms are classified into 10 types based on their base height and texture. Low-level clouds consist of cumulus (Cu), stratocumulus (Sc), stratus (St) and fractostratus (Fs). Mid-level clouds consist of altocumulus (Ac), altostratus (As) and towering cumulus. High-level clouds are cirrus (Ci), cirrostratus (Cs), cirrocumulus (Cc) and cumulonimbus (Cb).

Cloud-type identification in the satellite image is based on retrieving cloud properties from satellite data. The most useful properties are cloud height (low, mid or high), cloud fraction (broken or uniform) and visual opacity (thin or thick).

A few algorithms, which deal with extraction of properties of a cloud depend primarily on the label attached to a cloud type. For example, a low cloud can be stratocumulus or cumulus. However, the effective cloud amount and the cloud height contribute to the effectiveness of a cloud’s property. For high-level clouds, a separate class is defined as “uniform and thick”, because optically thick, extremely cold clouds, such as cumulonimbus, have different spectral and textural properties than other high-level clouds such as cirrus. Cloud classes are separated by cloud fraction and cloud height. The cloud types are classified into six groups by the identification of satellites, Table 1 shows the different types of clouds identified by satellite images.

METSAT (renamed as Kalpana-1) is the first in the series of exclusive meteorological satellites built by the Indian Space Research Organization. It is a multipurpose satellite for meteorological services including disaster-warning services. This satellite features a very high resolution radiometer, with 2-km resolution in visible band, and 8-km resolution in infrared and water-vapor band. The weather report generation and management using satellite images play important role in national development. Indian National Satellite Systems provide continuous monitoring of climatic changes through a series of satellites located over the Indian Ocean region.

Kalpana-1 is a geostationary satellite positioned at 72°E, observing earth with an imaging radiometer in three channels namely VIS (Visible), IR (Infrared) and UV (Ultraviolet) with central frequency of 0.7, 10.5 and 6.3 μm. The images are at a spatial resolution of 2-km × 2-km to obtain atmospheric cloud cover, water vapor and temperature. The temporal resolution of each image is 30 min, and the spatial resolution is 8 km for IR and WV channels, and 2 km for the VIS channel. It carries also a data relay transponder to provide data from fixed/mobile ground-level weather platforms. It was plotted from the transfer orbit to a geostationary at 37°E longitude, on 16 September, and then to the final parking at 74°E longitude on 24 September. The VIS channel of Kalpana-1 has not been used in the past for wind vector retrieval. It provides much better spatial resolution than the IR channel. The Kalpana-1 VIS channel has a sub-satellite point resolution of 2 km × 2 km with an image frequency of 30 min. Due to recent navigational and registration issues with the Kalpana-1 satellite, the main emphasis is given on retrieval from IR equivalent VIS image.

Figures 1 and 2 show the Kalpana-1 satellite image on two different days. Due to registration and navigational issues in the full disk Kalpana-1 images, a resampled image popularly known as sector generated image has been made available for the retrieval community, for all types of clouds.

Proposed method and implementation

Cloud classification is required to meet the objective of accomplishing an automatic extraction of data on cloud types. Mostly, clouds are described by higher reflectance and lower temperature, than the fundamental earth surface. A classification issue occurs when a pixel needs to be assigned into a predefined group or class, based on the features to categorize the image pixels to different cloud types (Gómez-Chova,
Camps-Valls, & Amorós, 2005). The process is described below:

1. Feature Extraction: Physically inspired features are extracted from the image.
2. FS-AP: All data points in the feature space to be initial clustering exemplars and iteratively refined.
3. Image Classification: Resulting clusters are subsequently labeled into different classes according to their spectral signatures.

**Feature Extraction**

Statistical methods are used to extract a large volume of information on clouds from satellite images. This facilitates a better understanding of the properties of the clouds. To improve the interpretability of a cloud image, the distinguishing features of the image have to be studied.

Level slicing is an enhancement technique whereby the digital number (DNs) distributed along the X-axis of an image histogram are divided into a series of analyst specified intervals or “slices”. The DN of the output image is generated by calculating the center pixel of the kernel with its coefficients, each time the kernel is moved. The influence of the convolution depends on the kernel size and the values of the coefficients. All of the DNs falling within the given interval of the input image are then displayed at a single DN in the output image. In this research, for image classification, we use level slicing to process cloud images. Figures 3 and 4 show the level slicing of the satellite cloud image.

Features of a given image can be divided into two categories. They are: (i) Spectral Features or Thematic map and (ii) Textural Features. The spectral and textural features extracted in the cloud image are based on the following statistical measures proposed by Christodoulou, Michaelides, and Pattichis (2003):

1. The spectral features used are the following:

   - Mean (ME):

   \[
   ME = \sum_{i=0}^{L-1} z_i p(z_i) \tag{1}
   \]

Figure 2. Input image-II.

Figure 3. Level slicing of input image-I.

Figure 4. Level slicing of input image-II.
where \( z \) is the variable, which indicates the values in the image (intensity in this case), \( p(z) \) is the frequency distribution of these values in the image and \( L \) is the number of possible levels of \( z \).

- **Standard deviation** (SD):
  \[
  SD = \sqrt{\frac{1}{L-1} \sum_{i=0}^{L-1} (z_i - ME)^2 p(z_i)}
  \]  
  (2)

  The standard deviation is a measure of contrast in the image.

- **Smoothness** (SM):
  \[
  SM = 1 - \frac{1}{1 + \sigma^2}
  \]  
  (3)

  where the variance \( \sigma^2 \) is defined here as:
  \[
  \sigma^2 = SD^2/(L-1)^2
  \]  
  (4)

  Values of SM result are in the range \([0, 1]\):

  - SM is 0 for an image of constant values, and 1 for an image with large variability.

- **Third moment** (TM):
  \[
  TM = \sum_{i=0}^{L-1} z_i p(z_i - ME^3) p(z_i)
  \]  
  (5)

  which measures the skewness of the histogram.

- **Uniformity** (UF):
  \[
  UF = \sum_{i=0}^{L-1} p^2(z_i)
  \]  
  (6)

  This feature is maximum when there is only one gray level present in the whole image, and minimum, when a large number of levels are present in the same amount of pixels.

- **Entropy** (EY):
  \[
  EY = \sum_{i=0}^{L-1} p \log_2 p(z_i)
  \]  
  (7)

  which is a measure of the randomness in the level values of the image.

(2) In a cloud image, texture is an important feature as the cloud image is based on the texture structure (Li, Dong, Xiao, & Xu, 2015). The image processing based on texture analysis would provide ways for analysis of all types of cloud images. From the satellite images used in this research work, nine feature sets have been extracted. They are: Auto Correlation, Cross Correlation, Cluster Prominence, Cluster shade, Dissimilarity, Area, Perimeter, Circularity and Max probability.

- **Auto Correlation**:
  \[
  f_1 = AC = \sum_{pq} (pq) GI(p, q)
  \]  
  (8)

- **Cross Correlation**:
  \[
  f_2 = CR = \sum p \sum q(pq) GI(p, q) - \mu_p \mu_q / \sigma_p \sigma_q
  \]  
  (9)

  where
  \[
  \mu_p = \sum_{p,q=0}^{N-1} p^* GI(p,q)
  \]
  \[
  \sigma_p = \sqrt{\sum_{p,q} (p - \mu_p)^2 GI(p,q)}
  \]
  \[
  \sigma_q = \sqrt{\sum_{p,q} (q - \mu_q)^2 GI(p,q)}
  \]

- **Cluster Prominence**:
  \[
  f_3 = CP = \sum_{p,q} (p + q - M_p, M_q)^4 GI(p,q)
  \]  
  (10)

- **Cluster Shade**:
  \[
  f_4 = CS = \sum_{p,q} (p + q - M_p, M_q)^3 GI(p,q)
  \]  
  (11)

- **Dissimilarity**:
  \[
  f_5 = DS = \sum_{p,q} (p + q GI(p,q))
  \]  
  (12)

- **Area**:
  \[
  f_6 = A = \sum_{p,q} [GI(p,q)]^2
  \]  
  (13)

- **Perimeter**:
  \[
  f_7 = P = 4A
  \]  
  (14)

- **Circularity**:
  \[
  f_8 = C = \frac{p^2}{4\pi A}
  \]  
  (15)

- **Max probability**:
  \[
  f_9 = MP = \text{MAX} GI(p,q)
  \]  
  (16)

  where \( p, q \) - Spatial positions

  \( GI \) - co-occurrence matrix

  \( M_p \) and \( M_q \) - intensity values

  The features of the images outlined in this study were computed by considering all pixels in the respective images rather than considering their types. The previous features were computed by using the fuzzy toolbox, which was executed in the MATLAB R12. This enabled us to label the pixels of cloud images easily.
**FS-AP**

The FSS measure is a key parameter to extract the type of the cloud from a satellite image. The characteristics of satellite images were extracted using fuzzy mean deviation. These were then used to develop a FSS measure, which would be useful in evaluating the similarity between two pixel vectors. We iteratively merge cluster centers to extract cloud-cover information by FSS. Previous research works have used the AP as a technique for pattern learning, which helps to identify patterns among data points and forms clusters of data points around these patterns (Frey & Dueck, 2007).

Each pattern is a data point that represents itself, and the related cluster of the other data points. The AP takes input measures as similarity between pairs of data points. The AP works by simultaneously considering all data points as likely patterns and iteratively exchanges values between data points, until a good set of patterns and clusters arrives. In the AP, a common choice for similarity is the negative Euclidean distance, even if more general notions of similarity can be used (Yang, Liu, Bruzzone, Guan, & Du, 2013). The algorithm proposed in this research work is based on FS-AP technique (Yang et al., 2010).

Figure 5 shows the schematic diagram of the FS-AP. When the FS-AP is compared with the conventional fuzzy-clustering methods, it simultaneously considers all data points in the feature space to be initial clustering patterns, and iteratively refines according to the mean distance deviation, until an optimal FSS is reached. A clustering algorithm is used to classify the cloud-patch features into a number of groups.

---

**Figure 5.** Flowchart of the FS-AP (Yang et al., 2010).
According to Rosenberg and Hirschberg (2007), the clustering problem is treated as a mapping from each data point to its cluster assignments. The target partitions are referred to as classes and only the hypothesized clusters are referred to as clusters. A commonly used clustering algorithm is the K-means, which is described in detail in a previous research work by Gordo, Martinez, Gonzalo, and Arquero (2011).

K-means clustering is a simple algorithm in which, for K clusters \( \{C_1, C_2, C_3, \ldots, C_k\} \) each of them having \( k \) patterns aims to find cluster centers \( m_k \) to minimize the cost function. Where,

\[
m_k = \frac{1}{n_k} \sum_{x \in C_k} X
\]

(17)

\[
E_k^2 = \sum_{k=1}^{k} \sum_{x \in C_k} ||X - m_k||^2
\]

(18)

where “X” is the feature set.

The initial cluster centers are chosen randomly and the algorithm is applied repeatedly until a steady state is reached. The procedure associated with the FS-AP as described in Yang et al. (2010) is as follows:

\textbf{Step 1: Initialize cluster centers randomly.}

\textbf{Step 2:} Set the initial value of the patterns and parameters.

\textbf{Step 3:} For all the pixels in the image do the following:

\textit{(i) Compute the Euclidean distance between the sample vector the clustering exemplar.}

\textit{(ii) Assign the pixel to that cluster whose center yields the minimum distance vector.}

\textbf{Step 4:} Identify the fuzzy cluster centers and the number of clusters

\textbf{Step 5:} Update the cluster centers by computing the mean of the feature vectors of the pixels belonging to that cluster.

\textbf{Step 6:} Between two consecutive updates, if the changes in the cluster centers are less than a specified value, then stop else go to Step 3.

At the end of this phase, we get a class label for each of the pixels and the centroids for each of the classes for the \( k^{th} \) cluster. The mean is given by:

\[
\mu_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i
\]

(19)

where \( \mu_k \) is the mean vector and \( n_k \) is the number of vectors in the \( k^{th} \) cluster. For the \( k^{th} \) cluster, the covariance matrix is given by:

\[
C_k = \frac{1}{n_k} \sum_{i=1}^{n_k} (x_i - \mu_k)^2
\]

(20)

where \( n_k \) is the number of vectors in the \( k^{th} \) cluster, \( x_i \) is the vector in the cluster \( k \) and \( \mu_k \) is the mean vector of cluster \( k \). Different random initializations of the cluster centers results in creating different clusters with less similarity. The convergence criteria is taken as,

\[
|\mu_k^{(n+1)} - \mu_k^{(n)}| < \text{Threshold}
\]

(21)

\section*{Image classification}

The efficiency of the proposed approach has been tested by Kalpana-1 satellite images, which were captured on two different days. Comparisons between FS-AP and traditional unsupervised algorithms (K-means and fuzzy K-means) and the standard AP method, based on the Euclidean distance (ED-AP), were carried out systematically. Accuracy assessment of obtained results is an important step in the classification process. The performance of the classification process is evaluated using confusion matrix (Demir, Bovolo, & Bruzzone, 2013). Report about the results of the image classification contains: confusion matrix, accuracy comparison and training time comparison. The confusion matrix gives the information about how much of original training areas pixels was actually classified as being in the class that the training areas were meant to represent. An error matrix, or often referred to as a confusion matrix, summarizes the relationship between the two sources of information. From an error matrix the images overall accuracy, producer’s accuracy, omission errors, user’s accuracy and commission errors can be determined (Jensen 2005). The confusion matrix is shown in Table 2.

Considering the level of classification accuracy, fuzzy logic can be satisfactory used for image classification (Table 3). Table 4 tabulates training time needed to compute the features for a given image.

\begin{table}[h]
\centering
\caption{Confusion matrix of the classification results.}
\begin{tabular}{|c|c|c|c|c|c|c|}
\hline
\textbf{Cloud Type} & \textbf{Ci} & \textbf{Cm} & \textbf{St} & \textbf{Sc} & \textbf{Cu} & \textbf{Cb} \\
\hline
\textbf{Ci} & \textbf{89.6} & 2.8 & 2.8 & 0 & 3.4 & 3.2 \\
\hline
\textbf{Cm} & 1.2 & 93 & 2.3 & 2 & 1.7 & 3.1 \\
\hline
\textbf{St} & 3.4 & 3.6 & \textbf{87.4} & 0 & 3.2 & 2.8 \\
\hline
\textbf{Sc} & 1.6 & 4.8 & 3.2 & \textbf{84.7} & 2.9 & 3.5 \\
\hline
\textbf{Cu} & 1.8 & 2.8 & 0 & 2.9 & \textbf{92} & 2.7 \\
\hline
\textbf{Cb} & 1.5 & 0 & 0 & 1.2 & 2.8 & \textbf{94.5} \\
\hline
\end{tabular}
\end{table}

\begin{table}[h]
\centering
\caption{Accuracy comparison of the classifiers.}
\begin{tabular}{|c|c|}
\hline
\textbf{Classifier} & \textbf{Accuracy} (\%) \\
\hline
K-means & 81 \\
Fuzzy K-means & 83.4\% \\
AP & 86 \\
FS-AP & 93.6 \\
\hline
\end{tabular}
\end{table}

\begin{table}[h]
\centering
\caption{Training time comparison.}
\begin{tabular}{|c|c|}
\hline
\textbf{Classifier} & \textbf{Training Time (sec)} \\
\hline
K-means & 6.7 \\
Fuzzy K-means & 5.7 \\
AP & 4.5 \\
FS-AP & 3.4 \\
\hline
\end{tabular}
\end{table}
Results and discussion

Cloud classification through satellite image is an approach which is essential in numerous atmospheric and environment studies; for example, weather analysis and forecasting. The classification of satellite clouds image is the best technique for classifying the cloud type as cumulonimbus, cumulus, stratocumulus and stratus/fog middle level and high level.

Each data point in the fuzzy clustering has some degree of belongingness to the cluster. Hence, the data points, which are presented on the edge of the cluster are different from the data points, which are shown in the centroid. In this research work, assignment of points to different clusters has been done based on the newly developed point. A cloud-texture classification implementing FS-AP is used to process cloud images in this aspect.

Figures 6 and 7 show the classification of the images. In these two figures, one could see the different colors indicating the different types of cloud formation in the images. The feature-extraction method is based on kernels, and can deal with non-linear relations between samples at different dates. To provide a confidence level for the assessment of whether a pixel contains more than one cloud layer, the method is applied on the pixel array, or tile. The...
validation of a cloud-classification method would consider the dataset of cloud types to match against the classifier output. Accuracy assessment is an important step in the classification process. The performance of the classification process is analyzed using confusion matrix. The confusion matrix is shown in Table 2. This table gives the summary of cloud-type classification results of the different cloud types. Based on the training data, this table shows relative accuracies, for each of the cloud groupings: C, Cm, St, Sc, Cu and Cb. The diagonal value represents the accuracy of image classification. As an illustration, consider classification result for cloud type “Cb”, which is shown as 94.5% in this table.

This result stems primarily from the fact that the training database is far from being complete in terms of cloud types for all regions. For this reason, region was eliminated as a feature. We understand that this limitation gives us an observation that the classifier must be retested for any new region in which the classifier is applied. This research work proves that the FS-AP method can be a potential classification technique for high-performance satellite-image classification. The performance of the proposed FS-AP is solely based on the accuracy of the classifier. Table 3 depicts the performance of proposed FS-AP when compared with the K-means, the fuzzy K-means algorithms and the AP method of two images.

Table 4 shows the training time comparison of the proposed approach, the K-means, the fuzzy K-means algorithms and the AP method. These results are also validated by other quality measures such as confusion matrix and accuracy considered, including those related to the computational time. Thus, we can conclude that, for the chosen dataset, the FS-AP technique outperforms the other classifiers.

The results show that this method can improve not only the classification accuracy, but also in the aspects of training time, compared with the standard K-means, the fuzzy K-means algorithms and the AP method. The overall conclusion is that FS-AP algorithm appears to be superior to other existing methods.

## Conclusion

Previous research works have used various approaches to automatically classify clouds obtained from satellite images. From the satellite images presented in this research work, different spectral and texture feature sets were extracted. By calculating the spectral ratio of the image as input, it is very effortless to identify the spectral properties of cloud, which enhances the quality of the images in terms of their pixels. The extracted features were used effectively for the classification of cloud types with reasonable accuracy.

In the process of classification of clouds from satellite images, fleeting change in the images is one of the key factors, which cause degradation of the classifier performance. The developed system was able to correctly classify cloud for the six classes. This research work has succeeded in developing a method that can classify pixels in the Kalpana-1 satellite data according to cloud types, through the use of fuzzy clustering with AP. The results suggest several ways of improving the overall accuracy of the classification. In this research work, multifeature analysis was utilized for the classification of satellite cloud images received from the Kalpana-1 satellite. It is observed that relatively first-rate clustering of different classes is provided by the proposed method outlined in this work.

Primary characteristic features of the clouds are its higher reflectance and lower temperature than the underlying earth surface. It should be noted that some cloud types such as thin cirrus, low stratus at night and small cumulus are difficult to be determined because of its insufficient contrast characteristics with the surface radiance. Future research should explore the possibility of handling such clouds.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References

Alonso, M.C., Sanz, A., & Malpica, J.A. (2007). Classification of high resolution satellite images using texture from the panchromatic band. *International symposium on visual computing* (pp. 499–508). Berlin, Heidelberg: Springer.

Bandyopadhyay, S., Maulik, U., & Mukhopadhyay, A. (2007). Multiobjective genetic clustering for pixel classification in remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 45(5), 1506–1511.

Banerjee, B., Bovolo, F., Bhattacharya, A., Bruzzone, L., Chaudhuri, S., & Mohan, B.K. (2015). A new self-training-based unsupervised satellite image classification technique using cluster ensemble strategy. *IEEE Transactions on Geoscience and Remote Sensing Letters*, 12(4), 741–745.

Chi, M.M., Qian, Q., & Benediktsson, J.A. (2008). Cluster-based ensemble classification for hyperspectral remote sensing. *IEEE IGARSS proceedings*, Boston, MA. (pp. 209–212).

Choodarathnakara, A.L., Ashok Kumar, T.A., & Shivaraprakash Koliwad, D.C. (2012). Satellite image classification with fuzzy logic: From hard to soft computing situation. *International Journal of Computer Science & Applications*, 1(9).

Christodoulou, C.I., Michaelides, S.C., & 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.
Demir, B., Bovolo, F., & Bruzzone, L. (2013). Classification of time series of multispectral images with limited training data. *IEEE Transactions on Image Processing*, 22(8), 3219–3233.

Elbhag, M., & Bahrawi, J. (2014). Cloud coverage disruption for groundwater recharge improvement using remote sensing techniques in *Asir Region Saudi Arabia*. *Life Science Journal*, 11(1), 192–200.

Foody, G.M. (1992). A fuzzy sets approach to the representation of vegetation continua from remotely sensed data: An example from lowland heath. *Photogrammetric Engineering Remote Sensing*, 58(2), 443–451.

Foody, G.M. (1996). Approaches for the production and evaluation of fuzzy land cover classification from remotely-sensed data. *International Journal of Remote Sensing*, 17(7), 1317–1340.

Fred, A.L.N., & Jain, A.K. (2005). Combining multiple clustering using evidence accumulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(6), 835–850.

Frey, B.J., & Dueck, D. (2007). Clustering by passing messages between data points. *Science*, 315, 972–976.

Gómez-Chova, L., Camps-Valls, G., & Amorós, J. (2005). Cloud detection for MERIS multispectral images. *Proceedings of the MERIS (A)ATSR Workshop 2005* (ESA SP-597). 26 - 30 September 2005 ESRIN, Frascati, Italy. Editor: H. Lacoste. Published on CDROM., id.5.3.1.

Gómez-Chova, L., Camps-Valls, G., Bruzzone, L., & Calpe-Maravilla, J. (2010). Mean map kernel methods for semisupervised cloud classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(1), 207–220.

Gómez-Chova, L., Camps-Valls, G., Calpe-Maravilla, J., Guanter, L., & Moreno, J. (2007). Cloud-screening algorithm for ENVISAT/MERIS multispectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12), 4105–4118.

Gordo, O., Martinez, E., Gonzalo, C., & Arquero, A. (2011). Classification of satellite images by means of fuzzy rules generated by a genetic algorithm. *IEEE Latin America Transactions*, 9(1), 743–748.

Guo, R., Shi, X., Marchese, M., Yang, C., & Liang, Y. (2010). Text clustering with seeds affinity propagation. *IEEE Transactions on Knowledge and Data Engineering*, 23(4), 627–637.

Gurve, M.K., & Sarup, J. (2012). Satellite cloud image processing and information retrieval system. 2012 *World Congress on Information and Communication Technologies*, Trivandrum, (pp. 292–296).

Heinle, A., Macke, A., & Srivastav, A. (2010). Automatic cloud classification of whole skymages. *Journal of Atmospheric Measurement Technology*, 3, 557–567.

Jabari, S., & Zhang, Y. (2013). Very high resolution satellite image classification using fuzzy rule-based systems. *Algorithms*, 6(4), 762–781.

Jensen, J.R. (2005). Introductory digital image processing: a remote sensing perspective - 3rd ed. Prentice hall. 526 p.

Key, J.R., Maslak, J.A., & Barry, R.G. (1989). Cloud classification from satellite data using a fuzzy sets algorithm: A polar example. *International Journal of Remote Sensing*, 10(12), 1823–1842.

Kláncár, G., & Škrjanc, I. (2015). Evolving principal component clustering with a low run-time complexity for LRF data mapping. *Applied Soft Computing*, 35(1), 349–358.

Kuril, S., Saini, I., & Saini, B.S. (2013). B-S cloud classification for weather information by artificial neural network. *International Journal of Applied Physics and Mathematics*, 3(1), 28–30.

Li, M., Zhang, S., Zhang, B., Li, S., & Wu, C. (2014). A review of remote sensing image classification techniques: The role of spatio-contextual information. *European Journal of Remote Sensing*, 47, 389–411.

Li, P., Dong, L., Xiao, H., & Xu, M. (2015). A cloud image detection method based on SVM vector machine. *Neurocomputing*, 169, 34–42.

Maulik, U., & Saha, I. (2009). Modified differential evolution based fuzzy clustering for pixel classification in remote sensing imagery. *Pattern Recognition*, 42(9), 2135–2149.

Mondal, A., Khare, D., & Kundu, S. (2013). Comparison of MLC and FCMT techniques with satellite imagery in a part of Narmada River Basin of Madhya Pradesh, India. *International Journal of Advanced Remote Sensing and GIS*, 2(1), 130–137.

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), 247–259.

Nedeljkovic, I. (2011). Image classification based on fuzzy logic. *Remote Sensing*. *Spatial Information Science*, 34, 1–6.

Rathnayake, V.S., Premaratne, H.L., & Sonnadara, D.U.J. (2011). Performance of neural networks in forecasting short range occurrence of rainfall. *Journal of the National Science Foundation of Sri Lanka*, 39(3), 251–260.

Rokach, L., & Maimon, O. (2005). Top-down induction of decision trees classifiers a survey. *IEEE Transaction on System, Man,Cybermetics. Part C: Applications and Reviews*, 35(4), 476–487.

Rosenberg, A., & Hirschberg, J. (2007). V-Measure: A conditional entropy-based external cluster evaluation measure. *EMNLP-CoNLL*, 7, 410–420.

Su, H., & Du, P. (2014). Multiple classifier ensembles with band clustering for hyperspectral image classification. *European Journal of Remote Sensing*, 47, 217–227.

Surya, S.R., & Simon, P. (2013). Automatic cloud detection using spectral rationing and fuzzy clustering. *Advanced computing, networking and security (ADCONS), 2013 2nd international conference on IEEE, Mangalore. (pp. 90–95).

Tag, P.M., Bankert, R.L., & Brody, L.R. (2000). An AVHRR multiple cloud type classification package. *Journal of Applied Meteorology*, 39(2), 125–134.

Tian, B., Azimi-Sadjadi, M.R., Haar, T.H., & Reinke, D. (2000). Temporal updating scheme for probabilistic neural network with application to satellite cloud classification. *IEEE Transactions On Neural Networks*, 11(4), 903–920.

Wang, F. (1990). Improving remote sensing image analysis through fuzzy information representation. *Photogrammetric Engineering and Remote Sensing*, 56(8), 1163–1169.

Wang, Y., & Mo, J. (2004). Fuzzy logic applied in remote sensing image classification. *Proceeding of international conference on systems,man, and cybernetics*, The Hague (pp. 6378–6382).

Yang, C., Bruzzone, L., Sun, F., Lu, L., Guan, R., & Liang, Y. (2010). A fuzzy-statistics-based affinity propagation technique for clustering in multispectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 48(6), 2647–2659.

Yang, C., Liu, S., Bruzzone, L., Guan, R., & Du, P. (2013). A feature-metric-based affinity propagation technique for feature selection in hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 10(5), 1152–1156.

Zhang, H.Y., Wu, Q.T., & Pu, J.X. (2007). A novel fuzzy kernel clustering algorithm for outlier detection. *Proceedings of IEEE ICMA*, Harbin (pp. 2378–2382).