# classification of small Magellanic cloud Using Gaussian mixture model

**Ban Abed –AL-Rizak**
*University of Baghdad, College of sciences, Astronomy and space Dep.*

## Abstract

In this paper we suggesting classified the small Magellan cloud by a method of supervised classification. Supervised classification relies on the prior knowledge of the location and identity of land cover types that are in the image. This can be achieved through field our work, study of Astronomical image "Small Magellan cloud". In this paper was devoted to classify the image by using a method of supervised classification and analysis. Which is, named supervised classifier, based on Gaussian mixture model which is often used in probability classification problems to model such distributions. The experiment demonstrate that the proposed method can get a better classification result and less affected by the noise.

## الخلاصة

في هذا البحث اقترحنا طريقة للتصنيف المرشد لسحابة ماجلان الصغرى. التصنيف المرشد يعتمد على المعرفة االمسبقه للموضع والشدة التي تغطي مساحة واسعة في الصورة. وقد انجز هذا البحث ضمن مجال عملنا حقل معالجة الصور الفلكية (سحابة ماجلان الصغرى). صنفت هذه الصورة باستخدام طريقة للتصنيف المرشده ومن تحليل النتائج. والتي تسمى بالمصنف المرشدة المعتمدة على أ نمودج الكاوسي المختلط للاحتمالية. التجربة وضحت ان الطريقة المستخدمة يمكن ان تعطي نتيجة افضل واقل تأثر بالضوضاء.

## Introduction

Remote-sensing research focusing on image classification has long attracted the attention of the remote-sensing community because classification results are the basis for many environmental and socioeconomic applications. Scientists and practitioners have made great efforts in developing advanced classification approaches and techniques for improving classification accuracy. The well known and common classification methods that, usually, followed by remote sensing data users are those categorized as supervised and unsupervised methods [1] .

Small Magellan cloud ( SMC ) is a dwarf galaxy. It has diameter about 7000 light years. Contains several hundred million stars. It has total mass of approximately 7 billion times of the mass of our Sun. The Small Magellan cloud (SMC) was once a barred spiral galaxy that is once disrupted by the Milky way to be, some how shown irregular. The nearest neighbor of Milky way is at a distance of about 200000 light years, it can be seen with naked eye. If we come to analysis the visible light image of the small Magellan cloud we shall see, white and blue of x-ray image referred to distribution of the young stars . Ultra violet ray is emitte by new born (massive stars). Pinky nebula is created by ionizing Hydrogen gas. The glowing red is because of the heating up of the embedded dust grains. Grey luminosity is produced heating of the synchrotron emission resulted from supernova explosions, [2] .

In this paper, the proposed method is presented to classify Small Magellan cloud (SMC) regions which depended on Gaussian probability density functions are widely used to model the distribution of values within a dataset.

## Supervised Classification

Supervised classification always classifies pixels of unknown identify by samples of known identity located within training areas. The analyst defines training areas by identifying regions on the image that can be clearly matched to areas of known identity on the image. Such areas should typify spectral properties of the categories they represent, and, of course, must be homogeneous in respect to the informational category to be classified. Clearly, traditional supervised algorithms are based on probability distribution models for the class of interest (Richards, 1986). The selection of these training data is a key step in supervised classification. Differences in the selection of training data were more important influences on accuracy than those among other four different classification procedures (Bo and Wang, 2003). Many methods have been devised to implement the basic strategy of supervised classification as follows: parallelepiped classification; K-nearest neighbour; minimum distance classification; maximum likelihood classification; Bayes's classification and so on [3] .

## Unsupervised Classification

Unsupervised classification can be defined as the identification of natural groups, or structures, within multi-spectral data. A typical sequence for unsupervised classification might be composed of two main stages. Firstly, the analyst specifies minimum and maximum numbers of categories by classification algorithm. The second is to find new centroids for each class, then the entire scene is classified again. Again new centroids are calculated; if the new centroids differ from those found in the preceding step, then the process repeats until there is no significant change detected in locations of class centroids and the classes meet all constraints required by the operator. Apparently, distance measures are the heart of unsupervised classification. Unsupervised classification is particularly useful when training data can not be obtained to perform supervised classification. These techniques are also used in exploratory analysis to determine the number of possible spectral classes that can be considered for a supervised classification process. Many procedures for unsupervised classification are available; despite their diversity, most are based on the above general strategy, such as K-Mean cluster method and ISODATA [4] .

## Proposed Classification Method

In this section, the proposed method is presented. Is the general multi-dimensional Gaussian mixture model is described and solved by Expectation-Maximization (EM) algorithm.

For a joint-volume with N voxels, each voxel is a n dimensional vector. The voxel intensity vectors are denoted by Recall that the $x_i(i = 1,2,\ldots,N)$ goal is to estimate the class probabilities on each voxel according to the intensity vectors. The probability distribution of the kth tissue class is denoted by $p_k(x|\Theta_k)$ , which is governed by a set of parameters $\Theta_k$. Given the parameters of all the classes, the probability distribution of each voxel can be described as a mixture of probability distributions as follows [5] :

$$p(x|\Theta_k) = \sum_{k=1}^{k} a_k \, p_k(x|\Theta_k) \qquad \text{................(1)}$$

where $a_k$ denotes the mixture coefficients. The parameter set of this distribution is $\Theta = (a_1, \ldots\ldots, a_k, \Theta_1, \ldots.., \Theta_k)$ with the constraint that

$$\sum_{k=1}^{k} a_k = 1$$

Typically, $p_k(x|\Theta_k)$ is modeled by a Gaussian distribution with mean $\mu_k$ and covariance matrix $\Sigma_k$ That is :

$$p_k(x|\Theta_k) = p_k(x|\mu_k, \Sigma_k) \qquad \ldots\ldots\ldots\ldots(2)$$

$$= \frac{1}{\sqrt{\det(2\pi\Sigma_k)}} e^{-(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k)/2}$$

Maximum likelihood (ML) estimation is a common used method to find the probability distribution parameters. The log-likelihood expression for this density from the data X is given by:

$$\log(L(\Theta|X)) = \log \prod_{i=1}^{N} P(x_i|\Theta) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(3)$$

$$= \sum_{i=1}^{N} \log\left(\sum_{k=1}^{K} a_k\, p_k(x_i|\Theta_k)\right)$$

Finding the ML solution directly from Eq. (3) is difficult because it contains the log of the sum. The EM algorithm is a good way to solve this problem. The iterative solution for finding the parameters at the (t+1) th iteration step is as follows:

$$a_k^{t+1} = \frac{1}{N}\sum_{i=1}^{N} P(k|x_i, \Theta^t) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

$$\mu_k^{t+1} = \frac{\sum_{i=1}^{N} x_i, p(k|x_i, \Theta^t)}{\sum_{i=1}^{N} p(k|x_i, \Theta^t)} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(5)$$

$$\sum_k^{t+1} = \frac{\sum_{i=1}^{N} = P(k|x_i,\Theta^t) \,.\, (x_i - \mu_k^{t+1})(x_i - \mu_k^{t+1})^T}{\sum_{i=1}^{N} p(k|x_i,\Theta^t)} \qquad \text{.........................(6)}$$

$$P(k|x_i,\Theta^t) = \frac{a_k^t p_k(x_i|\Theta_k^t)}{p(x_i|\Theta^t)} \qquad \text{...........................(7)}$$

$$P(k|x_i,\Theta^t) = \frac{a_k^t p_k(x_i|\Theta_k^t)}{\sum_j^k a_j^t\, p_j(x_i|\Theta_j^t)}$$

Taking the mixing parameters $a_k$ as prior probabilities, the probability of each class can be computed using Bayes' rule.

## Experimental Result

The proposed method of classified( the general multi-dimensional Gaussian mixture model is described and solved by Expectation-Maximization (EM) algorithm). is perform by visual basic. Frist input a vectorial volume X , the number of classes K. Second, Initialization of $\Theta, p_k(x|\Theta_k)$. Any classification method could be used , we choose K-means. Third, calculate the prior probability by Eq.(7). Fourth, compute the new parameter data according to Eqs. (4), (5), and (6). Finaly, repeat steps Second- Fourth until reaching the end condition. we can notice that the final regions are not homogeneous as expected because of the noise, so this method relies on the histogram of image. The results show as following:-

Figure (1) Represent the original image and its Histogram.

Figure (2) Represent the classified image with the Gaussian mixture model with K=6, and its histogram on small Magellan cloud.

## Discuss And Conclusion

The process of classifying this image wasn't easy because of interfering of the regions with each other . So there was a difficulty to point out the area of the test, and also because of the farina of the picked image by the telescope, we found the follows:- The most of areas were for the young stars and massive stars. That makes a clue that the galaxy Small Magellan cloud (SMC) is the vessel of the small stars or at the primmer forming.

1. The most of areas were for the young stars and massive stars. That makes a clue that the galaxy Small Magellan cloud (SMC) is the vessel of the small stars or at the primmer forming.
2. The results were much for concerning the green color because of the area of dust cloud.
3. We show from original image and Classified Images There are many Pieces of evidence indications between the magellanic clouds(the small Magellan cloud and large Magellan cloud) and our Galaxy .The existence of the eastern wing of Small Magellan cloud (SMC), extending out in the direction of the Large Magellan cloud (LMC).

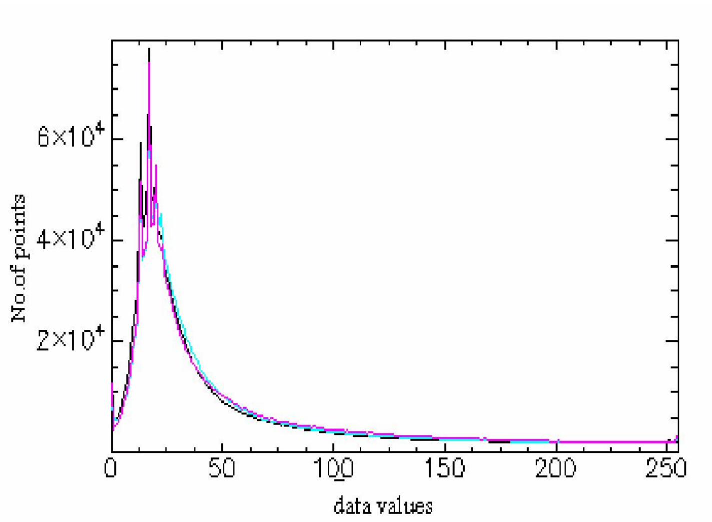
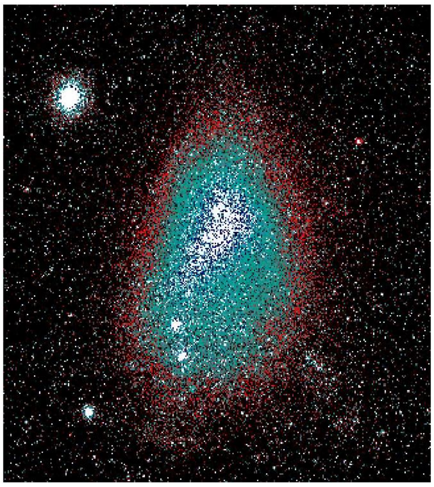



Fig.(1) Original image and its Histogram



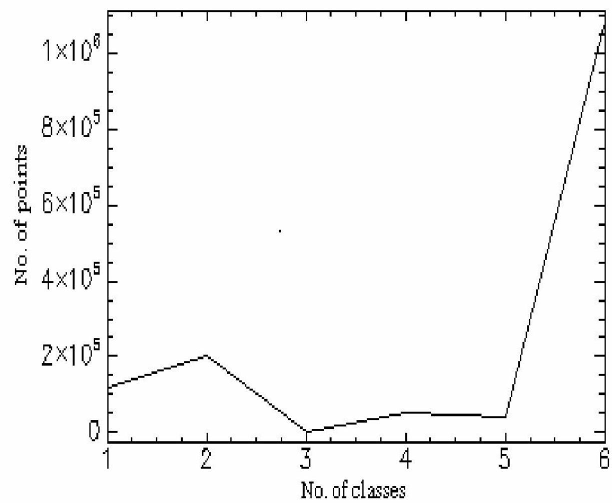



Fig.(2) Classified image and its histogram

## References


Xiangwei Liu "Several Different Remote Sensing Image Classification Technology Analysis" 2011, Journal of Sustainable Development Vol. 4, No. 5, October.

Fraser Cain "Small Magellanic Cloud", 2009, 7 May.

ALZUKY A.A.D. "Quantitative Analysis of Synthetic Aperture Radar (SAR) Images " 1998, ph.D. thesis, college of Science, Baghdad Univ.

D.Lu and Q.Weng "A Survey of Image Classification Methods And Techniques For Improving Classification Performance" 2007, International Journal of Remote Sensing Vol. 28, No. 5, 10 March, 823-870.

Hu.i Tang, Jean-Louis Dillenseger, and Li Min Luo "A Vectorial Image Classification Method Based On Neighborhood Weighted Gaussian Mixture Model" 2008, Author manuscript, published in Conference proceedings:Annuual International Conference of the IEEE Engineering in Medicine and Biology Society.