An Optimal Unsupervised Satellite image Segmentation Approach Based on Pearson System and k-Means Clustering Algorithm Initialization

Ahmed Rekik, Mourad Zribi, Ahmed Ben Hamida and Mohamed Benjelloun

***Abstract*—**This paper presents an optimal and unsupervised satellite image segmentation approach based on Pearson system and k-Means Clustering Algorithm Initialization. Such method could be considered as original by the fact that it utilised K-Means clustering algorithm for an optimal initialisation of image class number on one hand and it exploited Pearson system for an optimal statistical distributions' affectation of each considered class on the other hand. Satellite image exploitation requires the use of different approaches, especially those founded on the unsupervised statistical segmentation principle. Such approaches necessitate definition of several parameters like image class number, class variables' estimation and generalised mixture distributions. Use of statistical images' attributes assured convincing and promoting results under the condition of having an optimal initialisation step with appropriated statistical distributions' affectation.

Pearson system associated with a k-means clustering algorithm and Stochastic Expectation-Maximization 'SEM' algorithm could be adapted to such problem. For each image's class, Pearson system attributes one distribution type according to different parameters and especially the Skewness 'β1' and the kurtosis 'β2'.

The different adapted algorithms, K-Means clustering algorithm, SEM algorithm and Pearson system algorithm, are then applied to satellite image segmentation problem. Efficiency of those combined algorithms was firstly validated with the Mean Quadratic Error 'MQE' evaluation, and secondly with visual inspection along several comparisons of these unsupervised images' segmentation.

***Keywords*—**Unsupervised classification, Pearson system, Satellite image, Segmentation.

I. Introduction

INCREASING use of satellite images acquired periodically by satellites on different areas and for multiple purposes makes it extremely interesting for various applications. Indeed, the recent construction of multi spectral and hyper spectral images would provide detailed data with information in both the spatial and spectral domains. This data shows great promise for remote sensing applications ranging from environmental and agricultural to national security interests. Among satellite image processing, numerous algorithms using different approaches have been proposed during the past few years. These approaches include local edge detection [1, 4], deformable curves [2, 5], morphological region-based approaches [3, 6, 7], global optimization approaches on energy functions and stochastic model-based methods [8, 9].

Some intensity-based methods such as thresholding and histogram-based finite mixture models are formulated easily and quickly. However they often fail to segment objects with low contrast or noisy images with varying backgrounds. It is noted that these methods don't use the spatial morphological images information [10]. On the other hand, some other methods such as morphological segmentation, region growing and deformable curves, mainly focus on spatial information such as local structures or regions. Unfortunately, the majority of these techniques are not suitable for satellite image segmentation since such type of image presents a non homogenous texture [11].

In this context, it would be important to develop unsupervised statistical segmentation methods capable of analysing and classifying these types of images adequately. There has been recently considerable interest in stochastic model-based image segmentation techniques because of their efficiency. In such techniques, an image would be separated into a set of disjoint regions with each region associated with one of a finite numbers of classes that would be characterized by distinct parameters. In fact, there are two different statistic image segmentation approaches: the supervised approach and the unsupervised approach.

In the supervised approach, it is usually assumed that training data are available for the image classes; therefore, the parameters can be estimated from the training data before segmentation. But, this is rather unrealistic in many practical situations.

Zribi [12] says that, for unsupervised techniques, the objective is to estimate the parameters and segment the image simultaneously.

A. Rekik is with Unité de Recherche en Technologie de l'Information etElectronique Médicale 'TIEM', Ecole Nationale d'Ingénieurs de Sfax (ENIS), BP W, 3038 Sfax, Tunisia, and Laboratoire d'Analyse des Systèmes du Littoral, (LASLEA 2600), Université du Littoral Côte d'Opale, France *(*e-mail: ahmedrekik@yahoo.fr).

M.Zribi is with 2Laboratoire d'Analyse des Systèmes du Littoral, (LASLEA 2600), France (e-mail : Mourad.Zribi@lasl.univ-littoral.fr).

A. B. Hamida is the Director of the Unité de Recherche en Technologie de l'Information et Electronique Médicale 'TIEM', (ENIS), BP W, 3038 Sfax, Tunisia (e-mail : Ahmed.Benhamida@enis.rnu.tn).

Most of the proposed solutions to the unsupervised segmentation problem could be classified into two broad categories: one is a two-step procedure estimating the parameters for each class and then using a relaxation scheme to do segmentation. The other is an iterative procedure which starts with initial parameters and alternatively segments the image based on current parameters and estimates parameters based on current segmentation as if they were correct. In the first category, clustering algorithms are usually adopted to estimate the classes' parameters. In the second category, parameters are estimated in each iteration using the current segmentation as if they were correct, and the estimated parameters are used in the next segmentation as if they were true parameters.

Although these techniques have demonstrated substantial success for satellite imagery, they have some limitations. Indeed, most of the statistical image segmentation techniques need a manual initial input such as the classes' number. Meanwhile, these methods are often sensitive to these initial conditions. Moreover, all these methods cannot segment correctly the entire image if the initialization step is not optimal and if the distribution of every image classes is not known.

We propose in this paper an unsupervised satellite image segmentation approach with an optimal initialisation step based on Pearson system, in order to assure a better representation and definition of each image's classes.

We explore also in this work two models parameters estimation, the EM model and the SEM model, and we propose a comparison between these models [13]. We focus particularly on the mixture of distribution problem. In this context, we review the available theoretical results on the convergence of these algorithms and on the behaviour of SEM. Then we show that, for some particular mixture situations, the SEM algorithm is almost always preferable than the EM algorithm.

The SEM algorithm could be used as an efficient data exploratory tool for locating significant maxima of the likelihood function. In the real data case, we show that the SEM stationary distribution provides a contrasted view of the log-likelihood by emphasizing sensible maxima.

II. Statistical Satellite Image Classification : The Unsupervised Approach

Segmentation of a satellite image into differently textured regions' 'classes' is a difficult problem. Usually, one does not know a priori what types of textures exist in a satellite image, how many textures there are, and what regions have certain textures [14].

The monitoring task can be accomplished by supervised classification techniques, which have proven to be effective categorisation tools [15].

Unfortunately, these techniques require the availability of a suitable training set (classes' numbers for example) for each new image of the considered area to be classified. However, in real applications, it is not possible to rely on suitable ground truth information for each of the available images of the analysed site. Consequently, not all the satellite images acquired on the investigated area at different times can be used for updating the related land-cover maps. In this context, it would be important to develop classification methods capable of analysing the images of the considered site for which no training data would be available, thus increasing the effectiveness of monitoring systems based on the use of remote-sensing images.

Recently, researchers faced this problem by proposing an unsupervised retraining technique for maximum-likelihood (ML) classifiers capable of producing accurate land-cover maps even for images for which ground-truth information is not available [16]. This technique allows the unsupervised parameters' updating of an already trained classifier on the basis of the distribution of the new image to be classified.

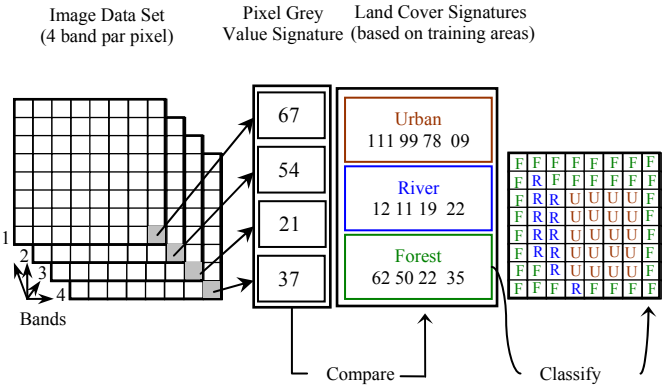

Fig.1 Important steps in supervised classification

A. Unsupervised and supervised classification principle

Classification principle could be described as follows: any individual pixel or spatially grouped sets of pixels representing some feature, class, or material is characterized by a (generally small) range of digital numbers for each band monitored by the remote sensor. The digital numbers values (determined by the radiance averaged over each spectral interval) are considered to be clustered sets of data in 2D plotting space. These are analyzed statistically to determine their degree of uniqueness in this spectral response space and some mathematical function(s) is/are chosen to discriminate the resulting clusters. Two methods for unsupervised and supervised classification are commonly used [15]:

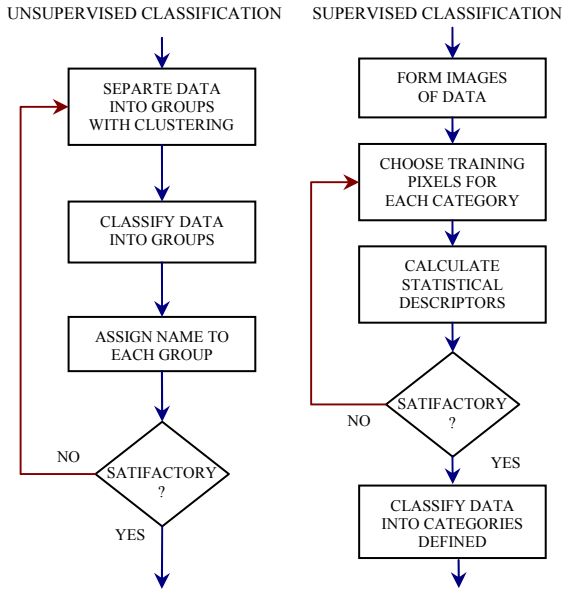

Fig. 2 Unsupervised and supervised classification principle.

In supervised classification the interpreter knows beforehand what classes are present and where each is in one or more locations within the scene. These are located on the image, areas containing examples of the class are circumscribed (making them training sites), and the statistical analysis is performed on the multiband data for each such class (see Fig.1). Instead of clusters then, one has class groupings with appropriate discriminant functions (it is possible that more than one class will have similar spectral values but unlikely when more than three bands are used because different classes/materials seldom have similar responses over a wide range of wavelengths). All pixels in the image lying outside training sites are then compared with the class discriminants, with each being assigned to the class it is closest to this makes a map of established classes.

In unsupervised classification, every individual pixel is compared to each discrete cluster to see which one it is closest to. A map of all pixels in the image, classified as to which cluster each pixel is most likely to belong, is produced (in black and white or more commonly in colors assigned to each cluster). This, then, must be interpreted by the user as to what the color patterns may mean in terms of classes, which are actually present in the real world scene; this requires some knowledge of the scene's feature/class/material content from general experience or personal familiarity with the area imaged.

The aim of the unsupervised classification methods is to find partitions of individuals set according to proximity criteria's of their attribute vectors in the representation space. The objective is in fact to group multiband spectral response patterns into clusters that are statistically separable.

Thus, a small range of digital numbers (DNs) say three bands, can establish one cluster that is set apart from a specified range combination for another cluster (and so forth).

Separation depends on the parameters we choose to differentiate. However, the unsupervised classification methods are used to do blind classification and so to achieve segmentation without priori knowledge of the image, but some important parameters like the class numbers must be fixed. Indeed, the majority of unsupervised classification algorithms need an initialization step on which parameters like the class numbers must be known. It exists different techniques for the estimation of this parameter.

We are going to present in this paper our initialization and estimation techniques used for this aim, and we try to describe the different steps of our statistical satellite image segmentation algorithm.

B. Unsupervised Satellite Image Segmentation Approach

The approach that we are going to evoke, thereafter, is in fact an optimal unsupervised segmentation method based on the adoption of the Pearson system [17], which is going to guarantee an optimal adjustment of the mixture densities. Our approach involves four steps described as follows:

a) A first step of initialisation and definition of image classes by the use of the k-means algorithm in order to provide an optimal classes' number, and to assure a better convergence of these different images attributes.
b) A second step permits the set of distribution families for every classes of our image, based on the Pearson system.
c) A third step of estimation and optimization of the different classes distribution parameters.
d) And finally, a bayesien segmentation step in order to define the different regions of interest in the image.

III. Modelling and Classification Algorithm

A. Initialisation step: K-means clustering algorithm

The K-means is an unsupervised classification algorithm based on a clustering technique [18] through which one set of data (observations) is divided into different groups ('clusters'), while introducing a similarity criteria, elements of a same group are most similar possible.

The objective of this algorithm is to regroup pixels in K distinct regions (classes); K being fixed by the user.

The K-means technique's takes pixels intensities as a basis. One randomly assigns each pixel to a class and one reiterates as follows: Centers of various groups (class) would be recalculated and each pixel would be again affected to the group according to its nearest center. Convergence would be reached when all centers would be fixed.

Principle: For one space referred to as 'V', observations consists of M vectors v_i $(i=1,\ldots, M)$. K-means algorithm consists in finding a partition of V noted $G=\{G_1, G_2,\ldots, G_K\}$, whose subsets would be called clusters, and each group would be represented by a vector C_k called centroïd. Let $C=\{C_1,\ldots, C_K\}$ are the whole of the centroïds, also called alphabet or partition dictionary. One introduces a distance measurement $d(v_i,C_k)$ between the i^{th} sample and the k^{th} centroïd.

The K-means algorithm proceeds in an iterative way starting from an initial alphabet $C^{[0]}$. With every iteration *[q]*, every sample, represented by a vector v_i, is an associated closest centroïde C_k , and a new partition $G^{[q]}$ is thus found. Total distance, relatively to every iteration, could be formulated as follows:

$$D^{[q]} = \frac{1}{M}\sum_{j=1}^{M}\min(d(v_j, c_k^{[q-1]})) \text{ with } k=1,\ldots,K \qquad (1)$$

For each group, (noting Q_k the number of vectors associated with the k^{th} group), the new centroïds are found by minimizing the distance:

$$D(G_k^{[q]}) = \frac{1}{Q_k^{[q]}}\sum_{v_j \in G_k^{[q]}} d(v_j, c_k^{[q]}) \qquad (2)$$

i.e.:
$$\frac{\partial(G_k^{[q]})}{\partial c_k^{[q]}} = 0 \qquad (3)$$

In our application, we will consider the Euclidean distance 'd' as a metric distance.

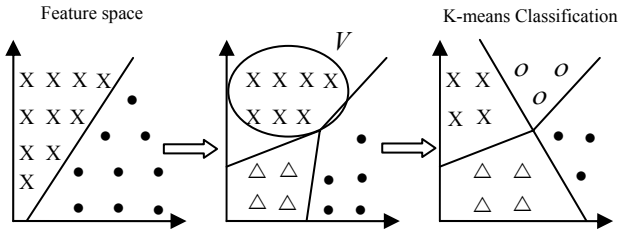

Fig. 3 K-means classification principle

B. Modelling and classification step

In our statistical image segmentation algorithm, we suppose the existence of two random fields: for the '*S*' set of pixels, we consider two sets of random variables $X=(X_s)_{s\in S}$, and $Y=(Y_s)_{s\in S}$ called "random fields". Each X_s takes its value in a finite set of classes $\Omega = \{\omega_1, \ldots, \omega_m\}$, and each Y_s takes its value in *IR*. The problem of segmentation (classification) is then that of estimating the unobserved realisation $X=x$ of the field X from the observed realisation $Y = y$ of the field Y, where $y = (y_s)_{s\square S}$ is the digital image to be segmented. The problem is then solved by the use of Gaussian mixture model and a Bayesian strategy which is "the best" in the sense of some criterion.

<u>*Gaussian mixtures:*</u> The Gaussian mixtures are usually used in classification because they correspond often with the variable distribution law. In the Gaussian case, the ϕ_k represents the averages of k^{th} class, noted μ_k, and Σ_k is the matrix of covariance. The density of multivariate probability in an X conditionally to $-\phi_k$ is written as :

$$f(x_i|\phi_k) = \frac{1}{(2\pi)^{\frac{p}{2}}\sqrt{\left|\sum_k\right|}} e^{-\frac{1}{2}(x-\mu_k)^T\sum_k^{-1}(x-\mu_k)} \qquad (4)$$

The Gaussian mixture model [19] is an omnipresent statistical model for density estimation, pattern recognition, and function approximation thanks to its benefits from analytical tractability, asymptotic properties, and universal approximate capability for continuous density functions.

The associated parameters can be estimated in an approximate maximum a posteriori (MAP) estimation or the maximum likelihood estimation. The MAP or ML estimation can be obtained by the expectation maximization (EM) algorithm or its more recent stochastic version (SEM) [20].

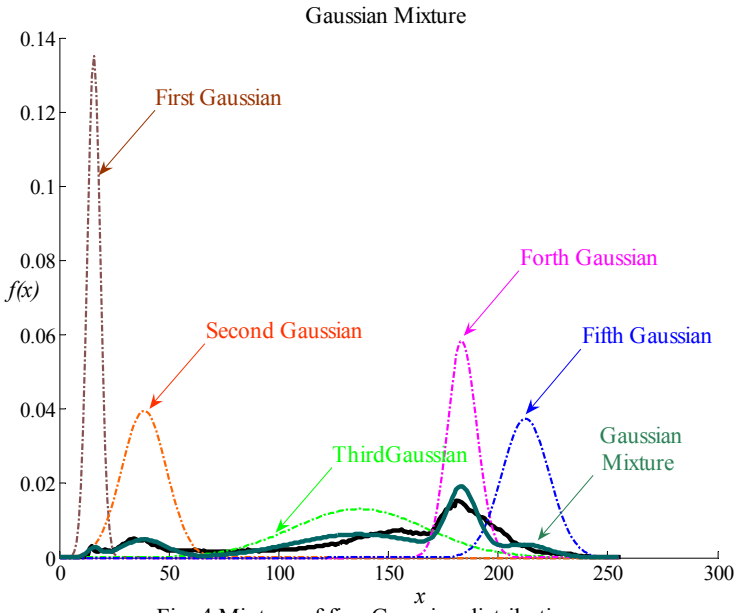

Fig. 4 Mixture of five Gaussian distribution.

Through this example, we can notice that the Gaussian mixture density permits to come as close as possible to the histogram of the image test. However, a few errors of approximations exist again in different classes. To minimize these errors, it is indispensable to estimate and to optimize the different parameters of the Gaussian mixture density. To remedy this problem, we are going to use the maximum likelihood estimator thereafter associated to the EM and SEM algorithm, and we are going to justify and to approve this choice through the calculation of the mean quadratic error measured, between the mixture density and the one of the image test.

C. Gaussian mixtures estimation

Consider a mixture model with $M>1$ components in $\Re^n$ for $n \geq 1$:

$$p(x) = \sum_{m=1}^{M} \pi_m p(x|m) \quad \forall x \in \Re^n, \qquad (5)$$

where $0 \le \pi_m \le 1, (\forall m = 1,2,\ldots,M)$ are the mixing proportions subject to $\sum_{m=1}^{M} \pi_m = 1$. For the Gaussian mixture, each component density $p(x|m)$ is a normal probability distribution.

$$p(x|\theta_m) = \frac{1}{(2\pi)^{n/2} \det(\sum_m)^{1/2}} . e^{\left(-\frac{1}{2}(x-\mu_m)^T \sum_m^{-1}(x-\mu_m)\right)} \quad (6)$$

where T denotes the transpose operation. Here we encapsulate these parameters into a parameter vector, writing the parameters of each component as $\theta_m = (\mu_m, \sum_m)$, to get $\Theta = (\pi_1, \pi_2, \ldots, \pi_M, \theta_1 \theta_2 \ldots \theta_M)$. Then, Eq. (5) can be rewritten as

$$p(x|\Theta) = \sum_{m=1}^{M} \pi_m p(x|\theta_m) \quad (7)$$

If we knew the component from which x came, then it would be simple to determine the parameters Θ. Similarly, if we knew the parameters Θ, we could determine the component that would be most likely to have produced x. The difficulty is that we know neither. However, algorithm like 'EM', or 'SEM' could be introduced to deal with this difficulty through the concept of missing data.

EM algorithm: The EM algorithm is a highly successful tool especially in statistics, but it has also found an array of different applications. One of the more common applications within the statistics literature is for the fitting of linear mixed models or generalized linear mixed models [21]. Another very common application is for the estimation of mixture models. We will show thereafter, the principle of this algorithm.

Given a set of samples $X=\{x_1, x_2, \ldots, x_K\}$, the complete data set $S=(X;Y)$ consists of the sample set X and a set Y of variables indicating from which component of the mixtures the sample came. We describe, below how to estimate the parameters of the Gaussian mixtures with the EM algorithm.

The usual EM algorithm consists of an *E-step* and an *M-step*. Suppose that $\Theta^{(t)}$ denotes the estimation of Θ obtained after the tth iteration of the algorithm. Then at the *(t + 1)th* iteration, the *E-step* computes the expected complete data log-likelihood function :

$$Q(\Theta|\Theta^{(t)}) = \sum_{k=1}^{K}\sum_{m=1}^{M} \{\log \pi_m p(x_k|\theta_m)\} P(m|x_k;\Theta^{(t)}), \quad (8)$$

where $P(m|x_k;\Theta^{(t)})$ is a posterior probability and is computed as

$$P(m|x_k;\Theta^{(t)}) = \frac{\pi_m^{(t)} p(x_k|\theta_m^{(t)})}{\sum_{l=1}^{M} \pi_l^{(t)} p(x_k|\theta_l^{(t)})} \quad (9)$$

And the *M-step* finds the *(t + 1)th* estimation of $\Theta^{(t+1)}$ by maximizing $Q(\Theta|\Theta^{(t)})$

$$\pi_m^{(t+1)} = \frac{1}{k}\sum_{k=1}^{K} P(m|x_k;\Theta^{(t)}), \quad (10)$$

$$\mu_m^{(t+1)} = \frac{\sum_{k=1}^{k} x_k P(m|x_k;\Theta^{(t)})}{\sum_{k=1}^{k} P(m|x_k;\Theta^{(t)})} \quad (11)$$

$$\sum_m^{(t+1)} = \frac{\sum_{k=1}^{k} P(m|x_k;\Theta^{(t)})\{(x_k - \mu_m^{(t+1)})(x_k - \mu_m^{(t+1)})^T\}}{\sum_{k=1}^{k} P(m|x_k;\Theta^{(t)})} \quad (12)$$

SEM algorithm: The SEM algorithm is a stochastic alternative and an improvement of the EM algorithm, obtained by the addition of a stochastic component. Indeed, in order to prevent a possible stabilization of EM algorithm in the neighbourhood of the maximum likelihood, a simulation phase (stochastic, *step S*) is added between the estimation phase (*step E*) and the maximization phase (*step M*) with an aim of making it less sensitive to the local minima.

Indeed, the SEM algorithm contains primarily three stages (Estimation, Simulation, and Maximisation), thus allowing to optimize the problem of the classes' number in the classification procedures while decreasing this one starting from a raised value K_{max}. The classes having few pixels are eliminated before a restarting.

In fact the principal idea of this algorithm is to insert a stochastic step between the *step E* and the *step M*. To every iteration one draws in each point x_i the multinomial random variable $e^n(x_i) = (e_k^n(x_i); k=1,K)$ with Multk $(1; \hat{z}_{i1},\ldots,\hat{z}_{iK})$ parameters, were:

$$\hat{z}_{ik}^t = \frac{\hat{\pi}_k^t f(x_i|\hat{\phi}_k^t)}{\sum_{k=1}^{K} \hat{\pi}_k^t f(x_i|\hat{\phi}_k^t)} \quad (13)$$

The multinomial distribution (*MultK*) characterizes the phenomena where, (a random pulling), K distinct and complementary events can occur, each one with a probability p_i subjected to the condition: $\sum_{i=1}^{k} p_i = 1$.

The probability that event i occurs $xi \ge 0$ times during N independent tests is given by this function: $f(x|p,n) = \frac{n!}{x_1!x_2!\ldots x_k!}\prod_{i=1}^{k} p_i^{x_i}$

The $e^n(x_i)$ define a $P^n = (P_1^n,\ldots,P_K^n)$, were $P_k^n = \{x_i / \hat{z}_k^n(x_i) = 1\}$

The *step M* is based then on these sub-samples for the estimation of the maximum likelihood: $\theta_k^{t+1} = ((\pi,\mu,\Sigma)_1^{t+1},\ldots,(\pi,\mu,\Sigma)_K^{t+1})$.

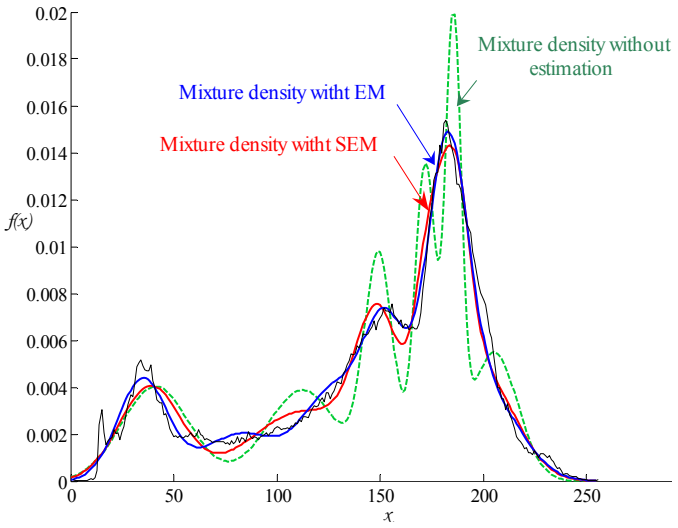

Fig. 5 The histogram of an image test, associated to the Gaussian mixture density before and after the use of EM and SEM algorithm.

The MQE measured respectively between the image density (histogram) and the one of Gaussian mixture without estimation is equal to 3.5467E-009, and the one of Gaussian mixture after the use of EM is equal to 7.4668E-010, and after the use SEM is equal to 3.8664E-010.

One notices therefore that the evaluation of Gaussian mixture parameters by EM and SEM, minimize the mean quadratic error, permitting a better representation of the image test density. However, one sees on Figure 5 that some classes don't have the shape (probability density) of a Gaussian, and that it would be discriminating to change the type of distribution. Indeed, instead of representing every class by a Gaussian law, we are going to try to insert new distributions (beta law, gamma law,…) in the mixture, and we are going to validate and to approve our choice by the calculation of the MQE measured between the Gaussian mixture and the mixture of several Pearson distributions.

IV. Classes' Distribution Optimization and Bayesian Segmentation

A. Pearson's System

A distribution density f on IR belongs to Pearson's system if it satisfies:

$$\frac{df(x)}{dx} = -\frac{(x+a)}{c_0 + c_1 x + c_2 x^2} f(x) \tag{14}$$

14

The variation of the parameters a, c_0, c_1, c_2 provides distributions of different shape and, for each shape, defines the parameters fixing a given distribution. Let Y be a real random variable whose distribution belongs to Pearson's system. For $q = 1,2,3,4$ let us consider the moments of Y defined by:

$$\mu_1 = \mathrm{E}[Y] \tag{15}$$

$$\mu_q = \mathrm{E}\left[(Y - E(Y))^q\right] \text{ for } q \geq 2 \tag{16}$$

and two parameters (β_1, β_2) defined by:

$$\beta_1 = \frac{(\mu_3)^2}{(\mu_2)^3} \quad \beta_2 = \frac{\mu_4}{(\mu_2)^2} \tag{17}$$

$\sqrt{\beta_1}$ is called 'Skewness' and $\beta 2$ 'Kurtosis'.

On the one hand, the coefficients a, c_0, c_1, c_2 are related to μ_1, μ_q, β_1, β_2 by the following formula :

$$c_0 = \frac{\hat{m}_2(4\hat{\beta}_2 - 3\hat{\beta}_1)}{10\hat{\beta}_2 - 10\hat{\beta}_1 - 18} \quad c_1 = \frac{\sqrt{\hat{\beta}_1 \hat{m}_2}(\hat{\beta}_2 + 3)}{10\hat{\beta}_2 - 10\hat{\beta}_1 - 18}$$

$$c_2 = \frac{2\hat{\beta}_2 - 3\hat{\beta}_1 - 6}{10\hat{\beta}_2 - 10\hat{\beta}_1 - 18} \tag{18}$$

$$a = \frac{(\beta_1 - \beta_2 + 1)\mu_1 + (\beta_2 + 3)\sqrt{\beta_1 \mu_2}}{10\hat{\beta}_2 - 10\hat{\beta}_1 - 18}$$

On the other hand, given:

$$\lambda = \frac{\beta_1(\beta_2 + 3)^2}{4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)}, \tag{19}$$

The eight families are illustrated in the Pearson's graph given in Figure 6.

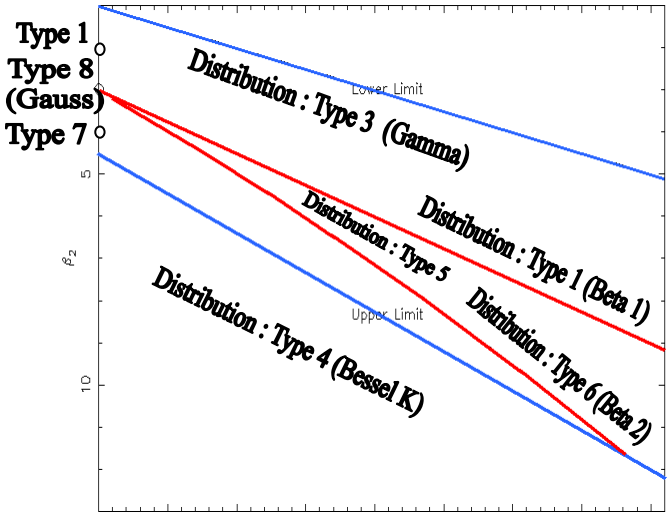

Fig. 6 the eight families of Pearson's system function of (β1, β2).

The aim of this system is to estimate all parameters defining the components of the distribution mixture as the shape of the marginal distributions of classes.

We most therefore calculate the moments: $\mu_1, \mu_2, \mu_3, \mu_4$, which can be estimated from empirical moments, from which we deduce the estimated values of β_1, β_2 by (17). Then, we estimate the family using (19). Once the family is estimated, values of *a, c_0, c_1, c_2*, given by (18) can be used to solve the parameters defining the corresponding densities. And finally, one must estimate the different parameters of each family by the SEM algorithm.

B. Bayesian segmentation

The aim of the Bayesian segmentation approach is to calculate the distribution of *Y*, and the model parameters if unknown, given *X*, that is:

$$P(\omega_i \mid X) = \frac{P(X/\omega_i).P(\omega_i)}{P(X)} \quad (20)$$

Knowing that:

$$P(X) = \sum_i P(\omega_i)P(X/\omega_i) \quad (21)$$

With:

- $P(\omega_i/X)$: a posterior probability.
- $P(X/\omega_i)$: adherence probability to a class.
- $P(\omega_i)$: a prior probability.

The rule used for the probabilistic and Bayesian decision, for the choice of the class to which a pixel belongs, is the maximum a posteriori:

$$x \in C_i \text{ if } p(x \in C_i \mid I_1,\ldots,I_P) = \max\left\{p(x \in C_k \mid I_1,\ldots,I_p)\right\} \quad (22)$$

V. Experimental Results

In this section, we apply the proposed above approach to some test images. We also make a comparison between the Mean Quadratic Error (*MQE*) found with the use of k-means initiation step, and *MQE* found with the use of histogram thresholding step.

The first test image is a natural image representing a boat. The initialisation parameters for this image are: K_{max}=10, iteration N^{br} =50.

The *MQE* measured respectively between image density (histogram) and the one of Gaussian mixture with K-means initialization is equal to *3.3725E-010*, and the one with histogram thresholding is equal to *4.2737E-009*

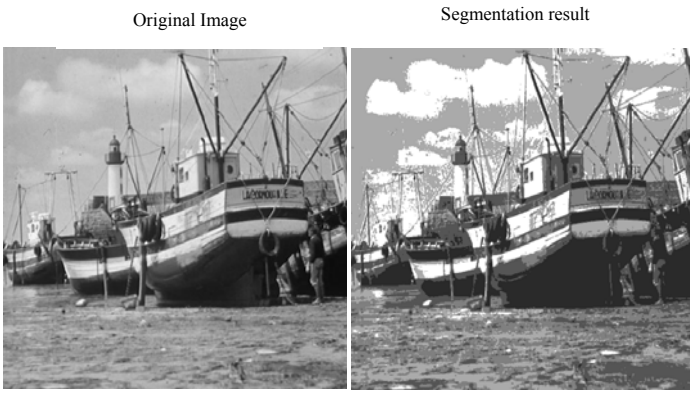

Fig.7 Unsupervised segmentation of 'the boat' image.

The second test image is a spot satellite image representing the city of Cairo with a 2,5m resolution. The initialisation parameters for this image are: K_{max}=15, iteration N^{br} =50.

The *MQE* measured respectively between image density (histogram) and the one of Gaussian mixture with K-means initialization is equal to *4.6015E-010*, and the one with histogram thresholding is equal to *1.04569E-009.*

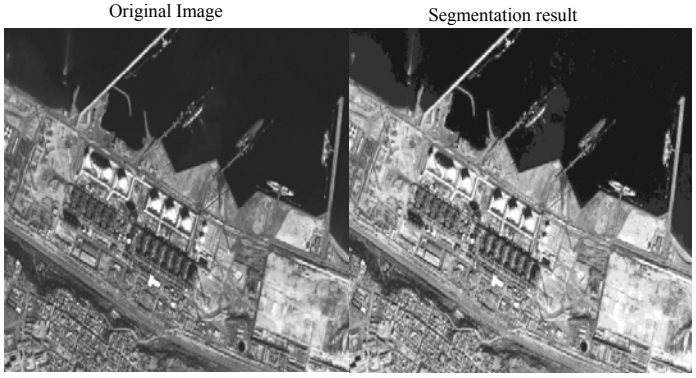

Fig. 8 Unsupervised segmentation of a Spot5 – 2,5 m image, representing the city of Cairo – EGYPT.

The third image test is a spot image representing the city of Rotterdam with a 20 m resolution. The initialisation parameters for this image are: K_{max}=20, iteration N^{br} =50.

The *MQE* measured respectively between this image density (histogram) and the one of Gaussian mixture with K-means initialization is equal to *3.6243E-010*, and the one with histogram thresholding is equal to *0.32679E-009*

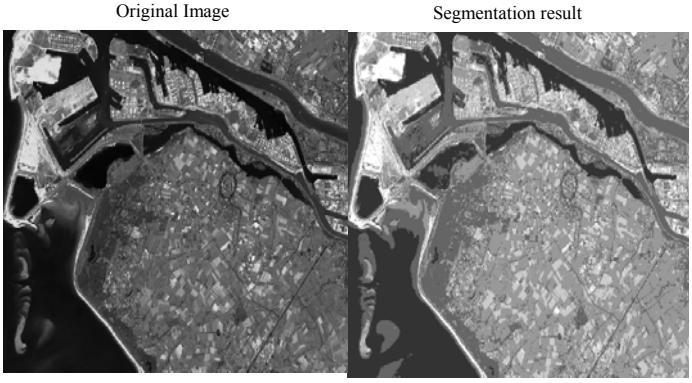

Fig. 9 Unsupervised segmentation of a Spot4- 20 m image, representing the city of Rotterdam –NETHERLAND.

The forth image test is a spot image representing the city of El Keela with a resolution equal to 5 m. The initialisation parameters for this image are:

K_{max}=20, iteration numbers =50.

The *MQE* measured respectively between this image density (histogram) and the one of Gaussian mixture with K-means initialization is equal to *3.4623E-010*, and the one with histogram thresholding is equal to *2.23369E-009*.

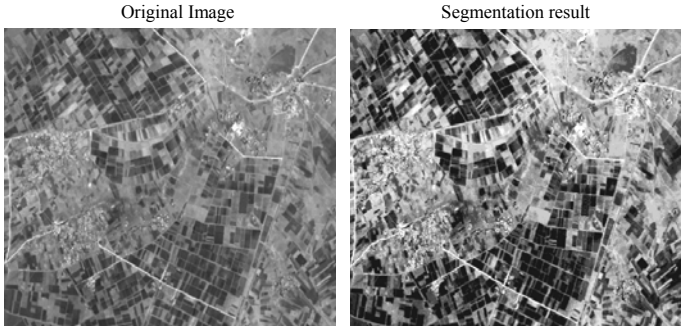

Fig. 10 Unsupervised segmentation of a Spot 5 – 5 m image, representing the city of El Keela – MOROCCO.

VI. Conclusion

Classifying satellite images is a challenging problem as these images often contain different textured regions or varying backgrounds, and are often subjected to illumination changes or environmental effects.

We proposed in this article one convivial solution to the optimization of unsupervised satellite image segmentation, in which a method based on K-means technique and Pearson system is used in order to optimize segmentation results. A mixture Gaussian model and SEM classification algorithms are also developed for Bayesien image satellite segmentation.

Our proposed method showed a high performance in satellite image classification since other distributions could be used for an optimal modelling. Such method is not sensitive to the selection of parameter values, and does not require any prior knowledge about the number of class (regions) in the image. Finally, it provides significantly better results than the existing unsupervised segmentation approaches.

References

[1] S. Li, T. Fevens, A. Krzyżak and S. Li, "Automatic clinical image segmentation using pathological modelling PCA and SVM", Engineering Applications of Artificial Intelligence, Volume 19, pp. 403–410, June 2006.

[2] A. Schwaighofer, V. Tresp, P. Mayer, A.K. Scheel and G. Muller, "The RA scanner: prediction of rheumatoid joint inflammation based on laser imaging", IEEE Trans. Biomed. Imaging 50, pp. 375–382. 2003.

[3] Y. Zheng, H. Li and D. Doermann, "Machine printed text and handwriting identification in noisy document images", IEEE Trans. Pattern Anal. Mach. Intell. 26, pp. 337–345, 2004.

[4] W.Y. Manjunath, "A framework of boundary detection and image segmentation", IEEE Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, pp. 744–749, 2005.

[5] C.Y. and L.P. Jerry, "Snakes, shapes, and gradient vector flow", IEEE Trans. Image Process. 73, pp. 359–369, 1998.

[6] S.C. and Y. Alan, "Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation", IEEE Trans. Pattern Anal. Mach. Intell. 189, pp. 884–900, 1996.

[7] G.-P. and G. Chuang, "Extensive partition operators, gray-level connected operators, and region merging/classification segmentation algorithms: Theoretical links", IEEE Trans. Image Process. 109, pp. 1332–1345, 2001.

[8] K.S. and K.U. Jayaram, "Optimum image thresholding via class uncertainty and region homogeneity", IEEE Trans. Pattern Recog. Mach. Intell. 237, pp. 689–706, 2001.

[9] J.P, "Stochastic relaxation on partitions with connected components and its application to image segmentation", IEEE Trans. Pattern Anal. Mach. Intell. 206, pp. 619–636, 1998.

[10] J. Xie, and H. T. Tsui, "Image segmentation based on maximum-likelihood estimation and optimum entropy-distribution(MLE–OED)", Pattern Recognition Letters, Volume 25, Issue 10 , pp. 1133–1141, July 2004.

[11] Y. Deng, and B.S. Manjunath, "Unsupervised segmentation of color–texture regions in images and video", IEEE Trans. Pattern Anal. Mach. Intell. 238, pp. 800–810, 2001.

[12] M. Zribi and F. Ghorbel, "An unsupervised and non-parametric Bayesian classifier", Pattern Recognition Letters 24, pp. 97–112, 2003.

[13] A. Rekik, M. Zribi, A. Ben Hamida and, M. Benjelloun, "Unsupervised Bayesian Image Segmentation Using Adaptive EM Algorithm based on Pearson ssystem", Vol. V, WMSCI 2006, Florida, USA, pp. 165–169, 2006.

[14] T.R. Reed, J.M.H. Du Buf, "A review of recent texture segmentation, feature extraction techniques", CVGIP Image Understanding 57, pp. 359–372, 1993.

[15] J.A. Richards, "Remote Sensing Digital Image Analysis", second ed., Springer-Verlag, New York, 1993.

[16] L.Bruzzone, D.Fernandez Prieto, "Unsupervised retraining of a maximum-likehood classifier for the analysis of multitemporal remote-sensing images", IEEE Transactions on Geoscience and Remote Sensing 39 (2001) pp.456-460.

[17] A. Rekik, M. Zribi, M. Benjelloun and A. ben Hamida, "A k-Means Clustering Algorithm Initialization for Unsupervised Statistical Satellite Image Segmentation", IEEE-International Conference on E-Learning in Industrial Electronics, Hammamet – Tunisia, 2006.

[18] H. Frigui and R. Krishnapuram; "Clustering by competitive agglomeration", Pattern Recogntion, Vol. 30, No.7, pp.1109–1119, 1997.

[19] G. McLachlan, D. Peel, "Finite Mixture Models", Wiley, New York, 2000.

[20] G. McLachlan, T. Krishnan, "The EM Algorithm and Extensions", Wiley, New York, 1997.

[21] McCulloch, C. E, "Maximum likelihood algorithms for generalized linear mixed models", Journal of the American Statistical Association, pp. 62–170, 1997.