An Algorithm for Determination of the Number of Modes for *pdf* Estimation of Multi-Modal Histograms

E. M. Aitnouri [*] S. Wang [*] D. Ziou [*] J. Vaillancourt [*] L. Gagnon [†]

[*]Dépt. de Mathématiques et d'Informatique
Université de Sherbrooke
Sherbrooke, Qc, Canada J1K 2R1

[†]Centre de Recherche en Informatique de Montréal
550 Sherbrooke O., Montreal
Qc, Canada H3A1B9

Abstract

An algorithm for determination of the number of modes in a gray-level image histogram is presented in this paper. The hypothesis is that the image histogram's pdf is approached by a mixture of Gaussians. Then, the algorithm tries to estimate the number of components in the mixture, which is an important parameter when using the maximum likelihood technique to estimate the remaining of parameters of the mixture. The algorithm is divided into two parts. First, initial clustering using the k-means algorithm is performed. This allows to estimate the centers of each cluster. Second, a novel algorithm, denoted "Elimination of False Clusters" (EFC) based on the Gaussian characteristics tries to suppress clusters which have no corresponding modes in the histogram. The algorithm has been validated on both artificial and real histograms.

1 Introduction

Estimation of a histogram's probability density function (*pdf*) is one of the fundamental operations involved in image processing. It is also a difficult operation. A histogram is usually composed of several modes, each of which corresponds to a class of objects. This type of histogram is called "multi-modal". Often, segmenting an image means separating the modes in the histogram. One of the major difficulties in estimating a histogram's *pdf* is that there may be several overlapping modes in the histogram (see figure 1). The existence of such overlapping modes makes it impossible to use popular parametric forms of *pdf* to estimate each mode in an isolation. When the histogram is "uni-modal" *e.g.*, it contains one mode, it can theoretically be approached by a Gaussian distribution [1]. When the histogram is multi-modal, complex models of *pdf* estimation should be used. The Mixture model is the most appropriate model.

Mixture models, particularly Gaussian mixture models, have recently attracted wide attention in the neural network community. The major interest of the Gaussian mixture is its capacity to produce a quick and useful approximation to a multi-modal histogram. The appeal of Gaussian distribution in the mixture is attributable to a large extent to the applicability of the EM (Expectation Maximization) algorithm [2] which maximize a likelihood function. However, the use of the EM algorithm requires that the number of components in the mixture (the number of modes in the histogram) is available. Moreover, in order to avoid local minima, accurate values of the parameters must be available before running the EM algorithm.

Determination of the number of components in the mixture can be treated as model selection problem using cross validation techniques [3] and Bayesian methods [4]. Moreover, Zhang [5] considered the problem as cluster validation problem and developed a formal way to estimate the number of components using the AIC criterion [6]. The main argument against these algorithms is their computational complexity. In this paper we are interested in direct methods based on "processed data knowledge" (histograms). One such method seeks to estimate the number of components in a mixture by using the number of inflection points in the histogram [7] [8]. Estimation using such an approach, however, is subject to distortion due to noise in the histogram. In this paper, we propose a new algorithm based on a direct approach (processed data knowledge).

Our algorithm is composed of two steps. First, the algorithm pre-processes input data using the k-means algorithm. The k-means algorithm performs an unsupervised learning in order to find centers of clusters which reflect the distribution of the data points. Our contribution is the determination of an appropriate number of initial clusters in the k-means algorithm so that all the modes can be accurately located for a large class of histograms. Then, in the second step, we propose the Elimination of False Clusters (EFC) procedure which is based on the Gaussian characteristics, for eliminating false clusters. The two steps result in accurate initialization of the parameters in the mixture, which helps keep the EM algorithm to result in a more accurate final estimation of the parameters.

The rest of this paper is organized as follows. Section 2 is devoted to brief overview of our algorithm and solutions to the determination of the number of initial clusters and the EFC procedure. The EM algorithm for histogram's *pdf* estimation is presented in section 3. Experimental results of our algorithm applied to both artificial and real histograms as well as evaluation of our algorithm are presented in Section 4. Finally, conclusion is given in section 5.

2 An overview of the algorithm

2.1 Basic hypothesis of the algorithm

A gray-level image histogram can be represented by a function $h(x)$, $x \in G_m$ of the gray-level frequencies of the image, where $G_m = \{0, 1, ..., N-1\}$ corresponds to the gray levels of the image. There exists a normalized representation of $h(x)$, denoted by $h_z(x)$ so that $h_z(x) = \frac{h(x)}{\sum_{i=0}^{N-1} h(i)}$ $(x \in G_m)$. Since $\sum_{i=0}^{N-1} h_z(i) = 1$ and $h_z(x) \geq 0$, $h_z(x)$ can be approached by a *pdf* denoted by *p(x)*.

When a given image contains more than one object/region, the histogram of the image may represent each object/region by a mode (see figure 1).

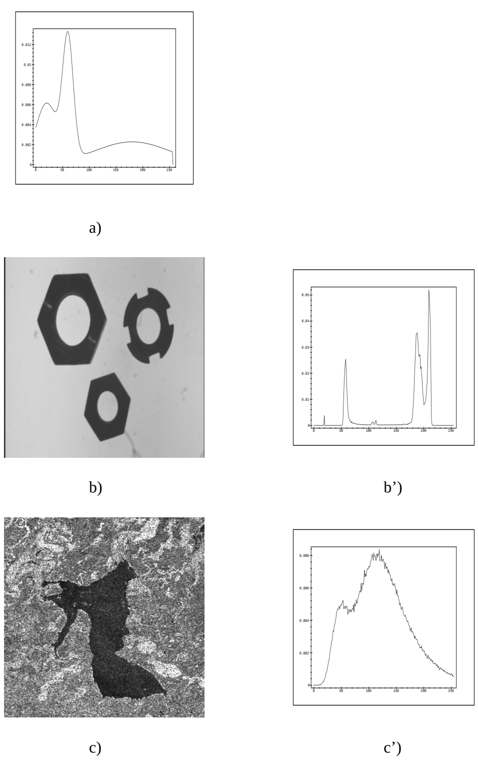


Figure 1: a) An artificial histogram generated from a mixture of three Gaussians. b) & b') An optical image and its histogram. c) & c') A typical radar image and its histogram

Our hypothesis is that each mode corresponds to a Gaussian distribution, which is acceptable in a large number of practical applications. Such a histogram can be approached by a mixture model so that:

$$p(x/\Theta) = \sum_{j=1}^{M} P_j G_j(x/\theta_{\mathbf{j}}) \quad (1)$$

with the restrictions $P_j \geq 0$ and $\sum_{j=1}^{M} P_j = 1$ $(j = 1, ..., M)$. P_j are the mixing parameters. Θ denotes the vector of parameters θ_j and θ_j the parameter of the j^{th} distribution. G_j is the j^{th} distribution of the mixture given by: $G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

We used the notation $p(x/\Theta)$ to emphasize the dependence of $p(x)$ on Θ.

2.2 Bloc diagram of the algorithm

Figure 2 shows a block diagram of the algorithm composed of two major steps. In the first step, initial estimation of the mixture parameters is done using the k-means algorithm. In order to approximate each mode by at least one Gaussian, the k-means algorithm is applied with a number K of clusters greater than the number M of modes in the image histogram. The next step mainly concerns the EFC procedure for suppressing false clusters that may result from the k-means algorithm. Basically, it takes advantage of the Gaussian *pdf* function. Before proceeding the elimination, a smoothing operation is performed on the histogram using a PNN (Probabilistic Neural Network or equivalently Parzen Window) [9]. While this operation is not essential in all cases, it greatly increases the robustness of our model against noise (especially when applied to radar images). Finding the optimal smoothing parameter for the PNN is another interesting question that we have studied [10] and will be presented in a further paper.

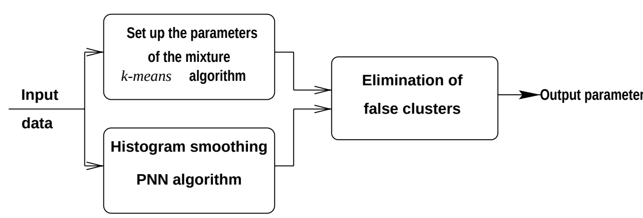


Figure 2: A block diagram of the proposed algorithm

2.3 The number of clusters in k-means algorithm

To initialize an appropriate number of clusters in the k-means algorithm, we have designed an experimental procedure as follows. First, for each of the possible number of modes $M = 1, ..., 5$, we randomly generate a set of 10000 artificial histograms containing M modes. The aim of this experiment is to find, for all the histograms with a given number of modes, the average value of the initial number of clusters, K, for which the clustering algorithm finds the true center of each mode. For this purpose, we propose an error function to measure the quality of the set of clusters computed by k-means algorithm. This is an average of the distances between each of the true centers μ_j, $(j = 1, M)$ (which are known in this experiment) and the nearest center $y(\mu_j)$ computed by the k-means algorithm:

$$err = \frac{1}{M} \sum_{j=1}^{M} \|\mu_j - y(\mu_j)\| \qquad (2)$$

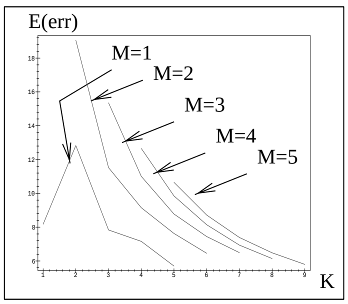


Figure 3: Average error $E(err)$, an indicator of the performance of the k-means algorithm, as a function of M and K. $E(err)$ is calculated over 10000 histograms.

Figure 3 shows a very interesting relationship between the number of clusters, K, the number of modes, M, and the precision of approximation . From the statistical point of view, the k-means algorithm accurately finds all the true centers if k is chosen at least $M + 4$. This itself is an important result concerning the k-means algorithm. We further note that the choices $M+2$ or $M+3$ are also good candidate for K. Consequently, when applied to real images, it is not necessary to impose a very strict condition on the accuracy of the estimation of M.

2.4 EFC procedure

The k-means algorithm was applied with an initial number of clusters greater than the number of modes, there are false clusters that must be eliminated. In general, EFC procedures can be performed using regulation-based methods. In this paper, we develop an EFC procedure which is based on the symmetry property of the Gaussian. Indeed, a cluster is true if its center separates a mode, on a certain interval, into two equal parts.

The EFC procedure proposed here depends on two parameters, β and γ. β is related to the relative level of the histogram at which the symmetry is measured. It specifies the percentage of the histogram height $h(y_j)$ for any cluster center y_j. In practice, β can be as large as 0.975 and as small as 0.5. The parameter β is closely related to “a concept of limit” introduced in [8] where $\beta = \frac{min_{x \in \{y_j, y_j+1, \ldots, y_{j+1}\}} P(x/\theta)}{h[y]}$, $0 \leq \beta < 1$.

The parameter γ is used as a threshold on the acceptable deviation between the true center and the closest center of the cluster computed by the k-means algorithm. If the deviation, theoretically written as $\|\mu_j - y_j\|$ is grater than γ, then y_j is rejected. In real applications, μ_j are unknown. Thanks to the fact that a true center divides a mode into two symmetric parts, an equivalent test can be performed without knowing the position of the true center. If horizontal line is plotted at the level $\beta h(y_j)$, it is sufficient to measure the absolute value of the difference between the horizontal distance from the point $(y_j, \beta h(y_j))$ to the first intersection point with the histogram at left (denoted by p_{left}), and the distance from the point $(y_j, \beta h(y_j))$ to the first intersection point with the histogram at right (denoted by p_{right}).

Now, we want to know the reasonable values for γ. To answer this question, we have designed an experiment using the same set of data used in the figure 3. The goal is to measure the deviation between the true center and the closest center found by the k-means algorithm, for each combination of M and K. to do this, we have adopted a *tolerant* (rather than a *conservative*) approach. To be more precise, we measure the maximum deviation. The result of this approach is the largest possible value for γ. Table 1 illustrates the results obtained for each combination of K and M.

M/K	1	2	3	4	5	6	7	8	9
1	8	12	8	7	6	-	-	-	-
2	-	20	11	9	8	7	-	-	-
3	-	-	16	11	9	8	7	-	-
4	-	-	-	13	10	9	8	7	-
5	-	-	-	-	11	9	8	7	6

Table 1: Values of γ in relation with M and K

Table 1 is useful for assigning values to γ. Note that since the number of modes M is unknown, one can take the maximum value of γ corresponding to the number of clusters used. The EFC procedure is given as follow :

- Choose a cluster y_j to test.
- Find the two distances d_{left} and d_{right} on the line $\delta = \beta h(y_j)$ from $(y_j, \beta h(y_j)$ to the nearest intersection points on both sides, i.e. $(y_j - d_{left}, \beta h(y_j)$ and $(y_j + d_{right}, \beta h(y_j)$.
- if $\|d_{left} - d_{right}\| > \gamma$, eliminate the cluster.

2.5 Evaluation of the algorithm

This evaluation aims to determine the capability of our model to find the exact number of modes in the histogram. This is different of that in figure 3. Indeed, figure 3 deals only with the quality of the centers estimated by the k-means algorithm in relation with the number of initial clusters. This evaluation, however, uses the same data from that in section 2.3., namely a set of 10000 random histograms. For each histogram, we apply the algorithm with an initial number of clusters greater than the number of modes.
The results are presented in tables 2 and 3 for a number of combinations of M and K. Table 2 shows the percentage of cases in which the exact number of modes is found. Table 3 shows the percentage of cases in which the difference between the exact number of modes and the detected number of modes, plus or minus one mode.

M/K	1	2	3	4	5	6	7	8	9
1	-	100	91.3	54.7	45.2	-	-	-	-
2	-	-	49.9	55.2	57.1	46	-	-	-
3	-	-	-	42.7	46.9	51.6	39.6	-	-
4	-	-	-	-	26.7	31.6	43.5	40.5	-
5	-	-	-	-	-	21.4	29.3	32.4	19.5

Table 2: Percentage of cases in which our model finds the exact number of modes in relation with M and K with $\beta = 0.96$

A number of remarks can be made regarding the two tables. First, it appears that the performance of our model decreases with the number of modes. This is reasonable because when the number of modes increases, the chance of having overlapping modes is higher. We believe the number of bins in the histogram (or equivalently the number of gray-levels) affects the results because it influences the “separability” of the randomly generated modes. In all of our experiments, the histogram has 256 bins.

Secondly, from table 2 , we can see that $M+3$ is a good choice for K, although $M+2$ and $M+4$ are not too bad. This agrees fairly well with the results presented in section 2.3.

Thirdly, although the best performance on detection of the exact number of modes for $M = 2, 3, 4$ is around 50%, it should be noted that when an error of one mode is allowed, the percentage rises beyond 80% for many M/K combinations. Moreover, we have noted that 73% of cases of table 3 are those in which the algorithm detects an extra mode. Both 50% and 80% results are encouraging if we consider the fact that our approach is based on the k-means algorithm which is not designed to automatically determines the number of modes.

M/K	1	2	3	4	5	6	7	8	9
1	-	0	8.7	35.3	45.8	-	-	-	-
2	-	-	50.1	43.5	39.4	31.8	-	-	-
3	-	-	-	45.1	40.9	41.5	43.1	-	-
4	-	-	-	-	34.3	36.2	38	41.1	-
5	-	-	-	-	-	21.1	15.7	20.3	27.2

Table 3: Percentage of cases where our model finds the real number of modes plus or minus one mode with $\beta = 0.95$.

3 Histogram’s *pdf* and the EM algorithm

The algorithm described in the previous section is designed to estimate the number of components in the mixture (the number of modes in the histogram). The number of components is not sufficient to represent the histogram’s *pdf*. Indeed, the mixture model requires the estimation of the parameters of each component [11]. Since by hypothesis we deal with a mixture of Gaussians, the parameters to be estimated are respectively the means, the widths and the mixing parameters of each component. The classical method for estimating these parameters is the maximization of the likelihood (ML) procedure.

The literature is rich of works on the use of the ML procedure for estimating the parameters of mixture models, particularly Gaussian mixtures. The technique of maximizing the likelihood function is based on the choice of the parameters most likely to give rise to the observed data. The use of Gaussians is favored due to the applicability of the EM algorithm which maximize a likelihood function. It is an iterative algorithm for searching the optimal values of the parameters. The EM algorithm consists of the iterative application of the following two steps:
1. The E-step based on the current parameter estimates: the posterior probability that the unit j is responsible of the generation of pattern x_k is estimated as:

$$h_j^k = \frac{P_j^{old} G_j(x_k/j, \mu_j^{old} \sigma_j^{old})}{\sum_{j=1}^{M} P_j^{old} G_j(x_k/j, \mu_j^{old} \sigma_j^{old})} \quad (3)$$

2. The M-step in which we obtain the new parameter estimates by:

$$P_j^{new} = \frac{1}{N} \sum_{k=1}^{N} h_j^k \quad (4)$$

$$\mu_j^{new} = \frac{\sum_{k=1}^{N} h_j^k x_k}{\sum_{k=1}^{N} h_j^k} \quad (5)$$

$$(\sigma_j^{new})^2 = \frac{\sum_{k=1}^{N} h_j^k (x_k - \mu_j^{new})^2}{\sum_{k=1}^{N} h_j^k} \quad (6)$$

The use of the EM algorithm to estimate the mixture parameters supposes that the number of components is available, which is determined by our algorithm. However, the number of components itself is not sufficient. Indeed, since the EM algorithm is an iterative algorithm, it requires accurate initial values of the parameters. Each detected component by the EFC procedure corresponds to a cluster that is represented by its center. From the experience of section 2.2., the centers of detected clusters are very good estimation of the means of the modes in the histogram. This gives an accurate initialisation of the means. Concerning the estimation of initial values of the width σ and the mixing parameters P_j, we use the *nearest-neighbors* classifier in relation with the estimated centers, and the following known formula:

$$\sigma_j^2 = \frac{1}{N_j} \sum_{x \in S_j} \|x - \mu_j\|^2 \quad (7)$$

$$P_j = \frac{N_j}{N} \quad (8)$$

where S_j is the j^{th} data set obtained by the application of the nearest-neighbors classifier, N_j is the total number of points classified in S_j, and N is the total number of points.

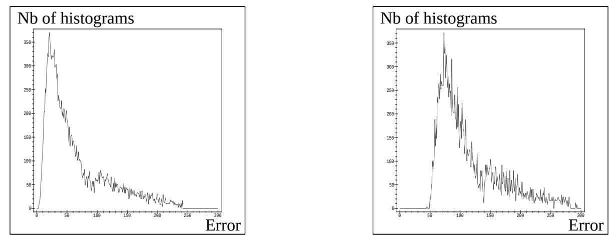
Eq. 7 and 8 allow our algorithm to be used as initialisation procedure for the EM algorithm for *pdf* estimation.

4 Experimental results

In section 2.5, we were interested in evaluating our algorithm, namely its capability to estimate the number of modes in a multi-modal histogram. The algorithm must necessarily be followed by the EM algorithm for estimating the histogram's *pdf*. In this section, we are interested in evaluating our algorithm when it is associated with the EM algorithm for *pdf* estimation. For this purpose, we apply both algorithms (our algorithm and the EM algorithm) to estimate *pdf* of artificial as well as real histograms.

4.1 Artificial histograms

Before estimating the *pdf* of the histogram in figure 1.a, we are interested in determination in general the effect that our algorithm brings to the EM algorithm for estimating of the mixture parameters. Thus, we have designed an experiment to study a comparison between results obtained using the EM algorithm initialized by our algorithm, and results obtained by the EM algorithm initialized by classical methods. As classical initialisation of the EM algorithm, we use the k-means algorithm to set the initial values of the parameters. This comparison will emphasize the EFC procedure since it is the only part which differs one procedure from the other. Figure 4.a and 4.b show the results of the comparison for a sample of 10000 histograms. For each histogram, we applied both methods, and for each method we computed the Mean Square Error (MSE) between original and estimated histograms. Then we graphed the number of histograms having a given error as a frequency plot. This allows us to identify the error around which each model is concentrated. We remark that the error around which our model is concentrated is significantly less than for the classical method. This shows the robustness of the EFC procedure.



a) b)

Figure 4: Frequencies of estimated histograms resulting with a given error x 4.10^{-4}. a) Histograms estimated with our model (Average Mean Square Error: $AMSE = 0.0132$), b) Histograms estimated with classical method (Average Mean Square Error: $AMSE = 0.0308$).

Concerning the histogram of figure 1.a, we tested $K = 5, 6, 7$. Figure 5 gives a graphic comparison between original and reconstructed histograms for the three values of K. Table 4 gives the real parameters of the histogram of figure 1.a. Tables 5,6 and 7 gives detailed results of the application of our algorithm respectively for $K = 5$, $K = 6$ and $K = 7$.

Modes	Means	variances	Mixing parameters
1	20	20	0.3
2	60	10	0.3
3	180	70	0.4

Table 4: Parameters of the artificial histogram

	k-means	before EM			after EM		
Cluster	μ_j	μ_j	σ_j	P_j	μ_j	σ_j	P_j
1	19.5	19.5	10.7	0.23	19.9	20.0	0.29
2	64.4	64.4	18.4	0.44	60.4	9.1	0.30
3	115.9	-	-	-	-	-	-
4	174.1	174.1	39.0	0.33	179.1	69.8	0.41
5	222.9	-	-	-	-	-	-

Table 5: Result of our model on the artificial histogram, $K = 5$

Note that since the artificial histogram is noise free, the algorithm was applied without the smoothing operation. In contrast, for the two real images of figure 1, we make use of the smoothing operation.

	k-means	before EM			after EM		
Cluster	μ_j	μ_j	σ_j	P_j	μ_j	σ_j	P_j
1	17.13	-	-	-	-	-	-
2	55.32	55.32	20.53	0.41	49.96	22.56	0.61
3	94.79	-	-	-	-	-	-
4	147.81	-	-	-	-	-	-
5	186.97	186.97	90.41	0.59	182.76	73.63	0.39
6	215.97	-	-	-	-	-	-

Table 6: Results of our model on the artificial histogram, $K = 6$ and $\beta = 0.975$.

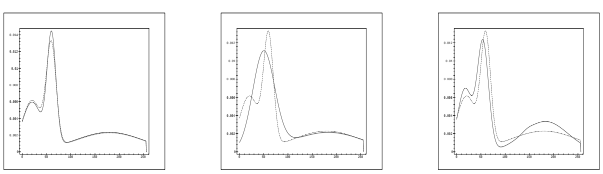
	k-means	before EM			after EM		
Cluster	μ_j	μ_j	σ_j	P_j	μ_j	σ_j	P_j
1	18.71	18.71	9.43	0.21	17.33	15.04	0.26
2	45.25	-	-	-	-	-	-
3	63.61	63.61	18.84	0.33	54.18	11.54	0.34
4	112.18	-	-	-	-	-	-
5	165.12	165.12	11.54	0.11	145.76	10.03	0.004
6	185.92	185.92	33.87	0.24	182.82	46.06	0.38
7	210.77	210.77	15.12	0.12	221.30	17.70	0.00028

Table 7: Results of our model on the artificial histogram , $K = 7$ and $\beta = 0.97$.

4.2 Real histograms

As application on real image histograms, we have tested our algorithm followed by the EM algorithm to estimate the *pdf* of histogram of figures 1. b and 1.c.

Concerning image b of figure 1, the k-means algorithm was applied with $K = 9$. Table 8 illustrates the result details, and figure 6 shows the reconstructed histogram using the estimated parameters.



a) b) c)

Figure 5: Reconstruction of the estimated histograms. a) $K = 5,\ \beta = 0.97$, b) $K = 6,\ \beta = 0.975$, c) $K = 7,\ \beta = 0.97$

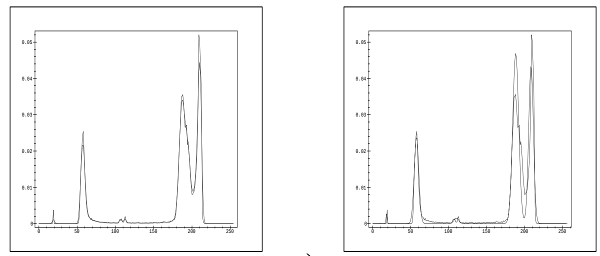
Concerning image c of figure 1, the k-means algorithm was applied with $K = 5$. Table 9 illustrates the result details, and figure 7 shows the reconstructed histogram using the estimated parameters.

5 Conclusion

The proposed algorithm significantly helps histogram's *pdf* estimation using mixture model. Indeed, it allows to obtain a good estimation of the number of components in the mixture, which is an essential parameter. With the determi-

	k-means	before EM			after EM		
Cluster	μ_j	μ_j	σ_j	P_j	μ_j	σ_j	P_j
1	15.7	15.7	6.02	0.0051	18.25	0.71	0.005
2	43.2	-	-	-	-	-	-
3	58.1	58.1	14.19	0.1811	57.64	3.03	0.181
4	94.4	-	-	-	-	-	-
5	114.5	114.5	15.13	0.0032	110.13	3.16	0.0119
6	168.9	-	-	-	-	-	-
7	189.9	189.9	6.98	0.4697	188.22	3.93	0.4620
8	209.0	209.0	3.58	0.3409	208.41	3.11	0.3401
9	238.5	-	-	-	-	-	-

Table 8: Result of our model on the histogram figure 1.b, $K = 9$

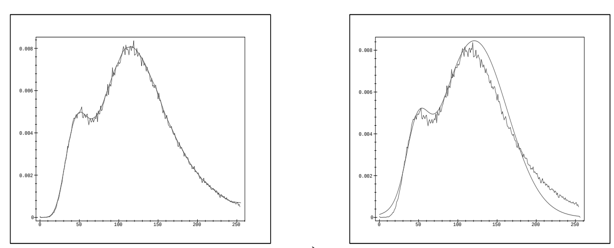


a) b)

Figure 6: an optical image histogram estimation. a) Smoothing the histogram using the PNN Approach. b) estimation of the parameters using the Gaussian mixture.

	k-means	before EM			after EM		
Cluster	μ_j	μ_j	σ_j	P_j	μ_j	σ_j	P_j
1	53.73	53.73	17.40	0.21	50.51	19.72	0.2
2	96.48	-	-	-	-	-	-
3	124.35	124.35	40.12	0.79	130.29	43.06	0.72
4	165.63	-	-	-	-	-	-
5	208.39	-	-	-	-	-	-

Table 9: Result of our model on the histogram figure 1.b, $K = 5$



a) b)

Figure 7: radar image histogram estimation. a) Smoothing the histogram using the PNN Approach. b) estimation of the parameters using the Gaussian mixture.

nation of the number of components, the algorithm reduces the necessity of human intervention for initializing the other parameters of the mixture. Strictly speaking, it requires only that the parameter β be specified in order adjust the algorithm's sensitivity to *small* modes lying near a large mode. This is clearly a parameter depending on the application in question. The algorithm is robust *w.r.t.*, the estimate of the number M of modes and the presence of noise (this point is not detailed in this paper). A wrong estimate slightly affects only the performance of the k-means algorithm, mainly in terms of execution time. The EFC procedure suppresses most false centers, which greatly facilitates the job of EM algorithm at the last step when used for *pdf* estimation.

There are a number of interesting alternatives to the choices made in this algorithm. For example, we could use a more advanced data clustering algorithm such as fuzzy k-means rather than the k-means algorithm. The k-means algorithm has been chosen mainly for its simplicity. Another example is the *tolerant* choice of the parameter γ for testing the symmetry. Other choices (*conservative*, or *statistical average*) are equally valid. Finally, the smoothing operation could also use many other existing techniques. Although all these alternatives offer interesting directions for further application of the algorithm, for example to extend it to a mixture of Gammas to deal with many images whose histograms are not mixture of Gaussians.

References

[1] D. B. Lawrence and J. T. Gene Hwang, How to Approximate a Histogram by a Normal Density, *The American Statistician*, Vol.47, $N^o 4$, 1993.

[2] A. P. Dempster, N. M. Lairol and D. B. Rubin, *Maximum Likelihood form Incomplete Data, Via the EM Algorithm* , Journal of the Royal Statistical Society, B 39 [1], pp. 1-38, 1977.

[3] M. Stone, Cross-Validatory choice and assessment of Statistical Predictions, *Journal of the Royal Statistical Society* B:, 36(1), 11-147, 1974.

[4] J. Mac Queen, *Some Methods for Classification and Analysis of Multivariate Observation*, in *Proc.* of the 5^{th} Berkeley Symp. on Math. Stat. and Prob., pp. 281-297, Berkeley, CA, Uni. Cali. Pres., 1967.

[5] J. Zhang and J. W. Modestino, (1990), *A model-fitting approach to cluster validation with application to stockastic model-based image segmentation*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 12, $N^o 10$, 1990.

[6] H. Akaike, Information and an extension of the maximum likelihood principle, In B.N. Petrov & F. Csaki (Eds.), 2^{nd} *International Symposium on Information Theory*, pp. 267-281, 1973.

[7] G. J. MaLachlan and K. E. Basford, *Mixture Models, Inference and applications to clustering*, Statistics, textbooks and monographs; V.84, 1988.

[8] D. Ziou, *Optimal Thresholding for Image Segmentation*, in *Proc.* 2^{nd} African Conference on Research in Computer Science, Burkina-Fasso, 1994.

[9] E. Parzen, *On the Estimation of a Probability Density Function and Mode*, Annals of Math. Stat. 33, pp. 1065-1076, 1962.

[10] Q. Jiang, E. M. Aitnouri, S. Wang, and D. Ziou, *Ship Detection in RADARSAT SAR Imagery Using the PNN Model*, in *Proc.* (CD) of the ADRO Symposium, Montreal, QC, 1998.

[11] E. M. Aitnouri, S. Wang, D. Ziou, J. Vaillancourt and L. Gagnon, *Estimation of a Multi-modal Histogram's pdf by a Mixture Model*, to appear in *Neural, Parallel & Scientific Computation*, 1999.