Neural, Parallel and Scientific Computations, no. 7, p. 103-118, July 1999

ESTIMATION OF MULTI-MODAL HISTOGRAM'S *pdf* USING A MIXTURE MODEL

Elmehdi Aitnouri[1], Shengrui Wang[1], Djemel Ziou[1],
Jean Vaillancourt[1] and Langis Gagnon[2]

[1] Department of Mathematics and Computer Science
University of Sherbrooke, Sherbrooke, Québec, J1K 2R1, Canada.
E-mail: {aitnouri, wang,ziou,vaillanc@dmi.usherb.ca}
[2] Centre de Recherche Informatique de Montréal
550 Sherbrooke Ouest, Montreal (Quebec), CANADA, H3A1B9
E-mail: langis.gagnon@crim.ca

ABSTRACT: A general model for estimating the *pdf* of a gray-level image histogram is reported. The histogram's *pdf* is approached by a mixture of Gaussian distributions. The originality of this work lies in the determination of the number of components in the mixture, which is considered as a parameter of the model and is determined using a novel algorithm. For this purpose, the model is divided into three parts. First, we use the *k-means* algorithm to set the initial values for the parameters of each component in the mixture. Our contributions are the determination of an appropriate number of clusters in the k-means algorithm and a novel algorithm for eliminating false clusters. Finally, the values of the parameters are refined using the EM algorithm. The model has been validated on both artificial and real image 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. 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. The mixture model is the appropriate one for approaching histograms with overlapping modes.

Mixture models, particularly Gaussian mixture models, have recently attracted wide-spread attention in the neural network community. They have shown better data classification capacities than many conventional neural network models such as layered networks trained with the Gradient Back-Propagation algorithm. The appeal of Gaussian distribution in the mixture is attributable to a large extent to the applicability of the *EM* (Expectation Maximization) algorithm (Dempster 77). However, two conditions must be satisfied in order to optimize the use of the EM algorithm. First, a rough estimation of the number of modes in the mixture must be available in advance, and second, accurate initial values for the parameters of each component in the mixture must be known in order to start up the EM algorithm.

Gaussian mixture models can be considered as forms of the Radial Basis Function problem (Bishop 95), where each basis function is a Gaussian. Thus, the number of basis functions (the number of components in the mixture) can be computed using methods based on the MDL (Minimum Description Length) (Tenmoto 96), combined with pruning techniques (Leonardis 98). On the other hand, determination of the number of components in the mixture can be treated as a model selection problem using cross-validation techniques (Stone 94) and Bayesian methods (MacKay 92). Moreover, Zhang (Zhang 1990) considered determination of the number of components in mixture models as a cluster validation problem and developed a formal way to estimate this number using the AIC criterion (Akaike 1973). 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 computing the number of inflection points in the histogram (Malachlan 88) (Ziou 93). 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).

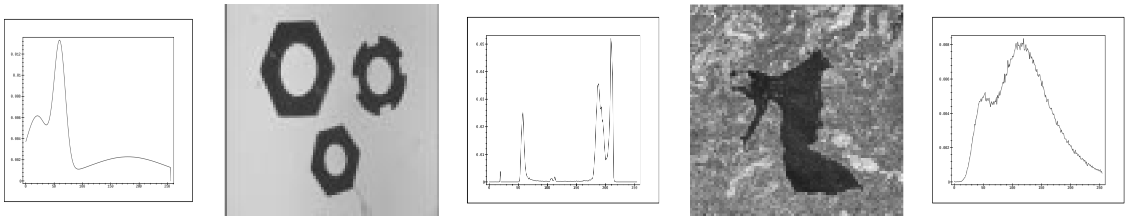
The proposed algorithm estimates the number of components in the mixture and the initial values for the parameters of each component before running the EM algorithm. We have developed a general model composed of three steps. First, the model pre-processes input data using the *k-means* algorithm (Mac Queen 67). The k-means algorithm performs unsupervised learning in order to find centers of clusters which reflect the distribution of the data points. Our contribution here 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 an algorithm, based on the Gaussian characteristics, for eliminating false clusters. The final step involves parameter refinement using the EM algorithm. The first two steps result in accurate initialization of the mixture parameters, which helps keep the EM algorithm from falling into local minima.

The rest of this paper is organized as follows. Section 2 deals with an overview of our model. Section 3 is devoted to solutions to the problems of determining the number of initial clusters and eliminating false clusters. Section 4 deals with the use of the EM algorithm. The experimental results yielded when our model is applied to artificial and real histograms are presented in Section 5. Finally, section 6 presents our conclusion.

2. GRAY-LEVEL IMAGE HISTOGRAM AS A *pdf*

A gray-level image histogram can be represented by a function, $h(x)$, $x \in Gl_N$ of the gray-level frequencies of the image, where $Gl_N = \{0, 1, ..., N-1\}$ corresponds to the gray levels of the image. When a given image contains more than one object/region, each of them will appear as a mode on the histogram of the image (see Figure 1).

This type of histogram is called “multi-modal”. However, when objects/regions in the image heve close gray-level averages, they may overlap to give a single mode. Our hypothesis is that each mode corresponds to a Gaussian distribution. This is acceptable in a large number of practical applications.



a) b) b') c) c')

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

Let us consider a multi-modal histogram $h(x)$ with M modes. Each mode is characterized by a vector of parameters denoted by θ (since the distribution is Gaussian, we have $\theta = (\mu,\ \sigma)$, where μ and σ, are respectively, the mean and the width of the Gaussian). There exists a normalized representation of $h(x)$, denoted by $h_z(x)$ such that $h_z(k) = \frac{h(k)}{\sum_{i=0}^{N-1} h(i)}$ $(k \in Gl_N)$. Since $\sum_{i=0}^{N-1} h_z(i) = 1$ and $h_z(i) \geq 0$, $h_z(x)$ can be approached by a *pdf* denoted by *p(x)*.
The reason for considering the histogram as a *pdf* is the possibility to use standard *pdf* estimation methods. Indeed, since the histogram possesses M modes, each of which corresponds by the above hypothesis to a Gaussian distribution, *p(x)* can be approached by a mixture model so that:

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

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

The literature abounds with work on the use of mixture models, particularly the Gaussian mixture which is well suited to approximate a wide class of continuous probability densities. The use of Gaussians is favored due to the applicability of the EM algorithm, which maximizes a likelihood function. The technique used to maximize the likelihood function is based on the choice of Θ most likely to give rise to the observed data. For analytical convenience, this is equivalent to minimizing the log-likelihood function. the minimization of the log-likelihood leads to the EM algorithm (Dempster 77).

There are a number of difficulties that must be overcome when using mixture models for *pdf* estimation. The two major difficulties concern the determination of M, the number of components to use in the mixture, and the initial values for the parameters. The standard model as defined in equation (1) does not include procedures to estimate M, but in contrast requires a fixed value of M. The estimation of the number of components in a mixture is an important problem which has not been completely resolved (MacLachlan 1988). The EM algorithm requires accurate initial values for the parameters, otherwise severe problems arise due to singularities and local minima in the log-likelihood function.

In this paper, we focus on the determination of both the number of modes in the histogram and accurate initial values for each component's parameters in order to optimize the use of the EM algorithm.

3. AN OVERVIEW OF OUR MODEL

Figure 2 shows a block diagram of the model comprised of three major steps. Our contribution resides in the first two steps, which constitute of a robust procedure for initializing the mixture parameters.

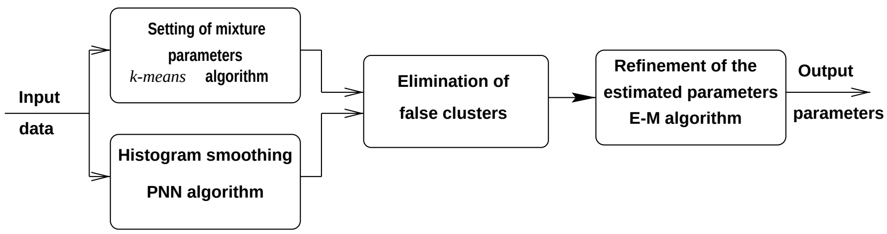


Figure 2: A block diagram of the proposed model

In the first step, an initial estimation of the mixture parameters is done using the k-means algorithm. This algorithm computes the center of each Gaussian in the mixture. In order to approximate each mode by at least one Gaussian, the clustering algorithm is applied with a number of clusters K greater than M the number of modes in the image histogram. Determination of an appropriate number of clusters is dealt with in the following subsections. The next step mainly concerns the elimination of false clusters that might result from the k-means algorithm. We propose a procedure utilising the Gaussian *pdf* function. Before proceeding with the elimination, a smoothing operation is performed on the histogram using a PNN (Probabilistic Neural Network or equivalently Parzen Window) (Parzen 62). 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 (Jiang 1998); we will summarize our basic approach in this paper. Finally, the EM algorithm is applied to the estimated parameters.

3.1. The number of clusters in k-means algorithm

Unsupervised learning and clustering methods offer useful pre-processing techniques for solving classification problems. These techniques can also be used to estimate the means of components in mixture models. Indeed, clustering methods make it possible to find the set of centers which most accurately reflect the distribution of the data points. The k-means algorithm is one of suitable algorithm for this purpose (Moody 89). However, use of the k-means algorithm requires setting of K, the initial number of clusters. K should be set according to some *a priori* knowledge about the data to be processed. In the case of image histograms, this knowledge basically concerns a rough estimate of M, the number of modes. The question is: how to choose the number of initial clusters such that the k-means algorithm accurately finds the center of each mode.

To our knowledge, a formal answer to this question is not available in the literature. Thus, to initialize an appropriate number of clusters, we have designed the following experimental procedure: First, for each of the possible numbers 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 k-means 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. This is an average of the distances between each of the true centers μ_j (which are known in this experiment) and the nearest center $y(\mu_j)$ computed by the k-means algorithm:

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

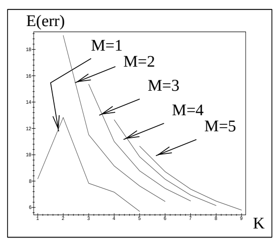


Figure 3: Average error $E(err)$, an indicator of the performance of the k-means algorithm, as a function of M and K. $E(err)$ is calculated over 10000 histograms. A subset of 15000 points has been generated according to each histogram with different values of K, the number of clusters.

Figure 3 shows a very interesting relationship between the number of clusters, K, the number of modes, M, and the precision of the approximation. The average error $E(err)$ in this figure is calculated over the 10000 random histograms for each combination of M and K. We note that it is very small when K is chosen to be 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 candidates 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. For example, if according to some prior knowledge we know that $M \leq 4$, then $K = 6$ is a very good choice in most cases.

3.2. Elimination of false clusters

The clustering algorithm was initialized *a priori* with a number of clusters greater than the number of modes in the histogram to ensure that the center of each mode would be located. However, the price paid for this insurance is the presence of false clusters along with the true ones. These false clusters must be eliminated, keeping only true clusters.

In general, false cluster elimination could be approached using regulation-based methods. For example, we can introduce a penalty term in the likelihood function to be optimized (Ormoneit 93). In this paper, false cluster elimination has been approached in a more direct way. The procedure 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 elimination 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. A large value of β such as 0.975 means that the symmetry is measured almost at the top, while a small value such as 0.5 means that the symmetry is measured in the middle. The parameter β is closely related to the concept of "limit" introduced in (Ziou 93).

The parameter γ also plays an important role in our procedure. 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 greater or equal to 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 $\beta h(y_j)$ (see Figure 4), 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}).

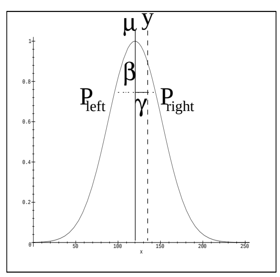


Figure 4: The principle of the deviation γ

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 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 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 result obtained for each combination of K and M. Table 1 is useful for assigning values to γ. Note that if the number of modes M is unknown, one can take the maximum value of γ corresponding to the number of clusters used. The algorithm for false cluster elimination is given as follows:

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

- 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. P_{left} and P_{right}.
- If $\|d_{left} - d_{right}\| < \gamma$, keep the cluster, else eliminate the cluster.

To conclude the initialization, we use the following known formula to set the values of the variance σ_j^2 and the mixing parameter P_j such that:

$$\sigma_j^2 = \frac{1}{N_j} \sum_{x \in S_j} \|x - \mu_j\|^2 \qquad P_j = \frac{N_j}{N}$$

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

The first two steps of our model, namely the k-means algorithm and false cluster elimination, allow us to initialize all the parameters of the mixture. Following this initialization, the EM algorithm can be applied to refine the values estimated for the parameters of each component.

4. EXPERIMENTAL RESULTS

In this section, we show the results of our model for the three histograms illustrated in Figure 1. For each histogram, we show the results of the k-means algorithm and false cluster elimination, the initial parameter values used as input to the EM algorithm, and the final parameter values output by it. The k-means algorithm was applied over 10000 iterations. Results are presented in Tables 3, 4 and 5.

4.1. Artificial histograms

The artificial histogram in Figure 1.a is generated from a true mixture of Gaussians. Table 2 illustrates its parameters.

This histogram represents a case that is typically difficult for all algorithms. Indeed, it contains a large mode next to a small one. The large mode will tend to monopolize more than one cluster of the k-means algorithm. Moreover, the histogram contains two overlapping modes, which introduces a difficulty for the elimination of false clusters. For this example, the value of β should be closer to one in order to detect the

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

Table 2: Parameters of the artificial histogram

small mode. Indeed, since there is an important overlap between the first two modes, the test for symmetry should be done above the separation level of the two modes. We have tested our model with $K = 5, 6, 7$ for two different values of β, namely $\beta = 0.97$ and $\beta = 0.975$

For the case of $K = 5$, our model accurately finds each parameter of the mixture for both values of β. Table 3 summarizes the estimated parameter values. First of all, we can see that the k-means algorithm ("k-means" column in Table 3) finds the centers of the modes. The false cluster elimination procedure ("before EM" column in Table 3) suppresses all false clusters (clusters represented by a dash). Finally, using the initial values of the parameters obtained from the first two steps, the EM algorithm can easily find the final values of the parameters.

	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 3: Result of our model on the artificial histogram, $K = 5$ and $\beta = 0.97$.

For the case of $K = 6$, our model accurately finds each parameter of the mixture for $\beta = 0.97$. The final estimated parameters resemble those of the case with $K = 5$ and $\beta = 0.97$ (see Table 4). However, when $\beta = 0.975$, one real cluster is suppressed by the elimination procedure. Table 4 summarizes the estimated parameter values. The k-means algorithm finds the centers of the modes, unfortunately, the elimination procedure suppresses the first real mode $\mu = 17.13$ (first dash in the "before EM" column of Table 4).

For the case of $K = 7$, our model accurately finds each parameter of the mixture for $\beta = 0.975$ (the table is not given). The final estimated parameters resemble those of the case with $K = 5$ and $\beta = 0.97$ (see Table 3). However, when $\beta = 0.97$, two false clusters, namely $\mu = 165.12$ and $\mu = 210.77$ (clusters 5 and 7), cannot be suppressed by the elimination procedure (see Table 5).

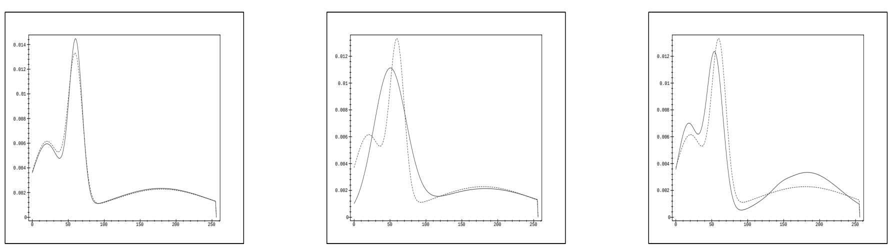
Figure 5 gives a graphic comparison of the original and reconstructed histograms for the cases illustrated in Tables 3, 4, 5.

	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 4: 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 5: Results of our model on the artificial histogram , $K = 7$ and $\beta = 0.97$.



a) b) c)

Figure 5: Reconstruction of the estimated histograms. a) $K = 5$, $\beta = 0.97$, $MSE = 0.0025$, b) $K = 6$, $\beta = 0.975$, $MSE = 0.013$, c) $K = 7$, $\beta = 0.97$, $MSE = 0.0096$

4.2. Estimation of the *pdf* of real image histograms

A major difference between artificial and real histograms is the presence of noise in real histograms. The false cluster elimination step in our model is sensitive to noise. Thus, in order to increase its robustness against noise, a smoothing operation is performed. Note that the k-means algorithm is not sensitive to noise.

There are many classical algorithms that can be used to perform a smoothing operation. Indeed, since noise corresponds to high frequency signals, Fourier transform analysis of the histogram can be used to suppress high frequencies (noise). Another simple method is to use low-pass filtering. We have chosen the PNN to perform the smoothing operation because it is easy to tune it for the trade-off between noise suppression and mode conservation. Moreover, the PNN is an algorithm that can be used to compute non-parametric *pdf*. Indeed, the PNN is now used as a *pdf* estimator in an application involving ship detection (Jiang 98).

The PNN is an algorithm which computes the Parzen window *pdf* (Parzen 1962). It is a kernel-based non-parametric estimation method and is particularly appropriate when the data are intensive, as in our case. The resulting *pdf* is a mixture of kernels superposed on each input point, weighted by the normalized original histogram $h_z(x)$ $(x = 0, \cdots, N-1)$. When Gaussians are used as kernel windows, the resulting PNN-pdf will depend on the width σ of the Gaussians. σ in this case controls the smoothness of the resulting PNN-pdf, denoted by $h_s(x)$ (smoothed histogram) which is given by:

$$h_s(x) = \sum_{i=0}^{N-1} h_z(i) * G(i,\sigma)(x) \qquad (3)$$

where $h_z(x)$ $(x = 0, ..., N-1)$ is the normalized histogram, and $G(i,\sigma)$ is a Gaussian.

4.2.1. Estimation of σ

There are no formal ways for choosing σ: the choice depends on the application in question. In our case, h_s should satisfy two major conditions. First, h_s should be an accurate estimate of h_z. Second, h_s should be smooth. We have developed a selection criterion for setting the value of σ in order to satisfy these two conditions, namely the value which leads to the minimum value of the error E_p given by:

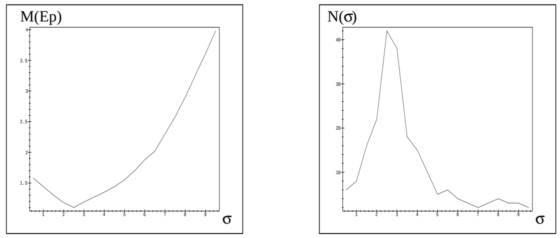
$$E_p = E_q + \eta \Delta h_s.$$

Here E_q is the quadratic error $E_q = \sqrt{\sum_{i=0}^{N-1}(h_z(i) - h_s(i))^2}$. h_z and h_s respectively, are the original and estimated histograms. The second term of E_p introduces a penalty term that punishes any estimated histogram which tends to oscillate. Δh_s is the sum of absolute value of the differences between two consecutive values of h_s and is given by $\Delta h_s = \sum_{n=1}^{N-1} \|h_s(n) - h_s(n-1)\|$. η expresses the penalty rate.
For a specific database of SAR images, we can estimate a statistically optimal value of σ. The experiment was performed on a database of 200 radar images. For each image, the best value of σ which minimizes E_p was computed.

Figure 6.a illustrates the mean error $M(E_p)$ calculated over the 200 images for each value of σ. The minimum mean error was obtained for $\sigma = 2.5$. Figure 6.b shows, for any given value of σ, $N(\sigma)$ the number of images that give rise to the minimum error. Figure 6.b provides information about the distribution of the number of images *versus* the best choice of σ. In our experiments, the concentration of this distribution was around $\sigma = 2.5$.

4.3. Radar image histograms

As an application to real histograms, our model was used to estimate the *pdf* of the radar image histogram illustrated in Figure 1.c. Such an application has a direct interest for radar image segmentation (El Zaart 1998). Indeed, mixture parameters are used in algorithms which directly give the thresholds required for segmentation.



a) b)

Figure 6: Results of the experiment for estimating the best value of the width σ to perform a smoothing. The experiment was performed for $\sigma = 0.5$ to 9.5 with a step of 0.5 and $\eta = 1$. a) Mean error. b) Number of images that give rise to the minimum error.

The value of β was set to 0.97, since the presence of overlapping modes meant the test of symmetry should be done at a high level. It is clear that for small values of β, the small mode would be suppressed. The histogram in Figure 1.c was smoothed with a width value $\sigma = 2.5$. The k-means algorithm was applied with number of clusters $K = 5$. The Results of each step of the model are illustrated in Table 6. Figure 7 shows the reconstructed histogram.

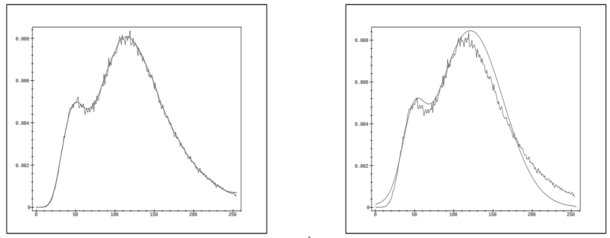
	k-means	before EM			after EM		
Cluster	μ_j	μ_j	σ_j	P_j	μ_j	σ_j	P_j
1	53.73	53.73	17.40	0.21	50.51	19.72	0.2
2	96.48	-	-	-	-	-	-
3	124.35	124.35	40.12	0.79	130.29	43.06	0.72
4	165.63	-	-	-	-	-	-
5	208.39	-	-	-	-	-	-

Table 6: Result of our model on the histogram in Figure 1.b, $K = 5$

$K = 5$ or $K = 6$ might be good choices for most radar images. Our experiments on quite a number of radar images actually confirm these choices. Figure 7 also shows that there are significant differences between the form of a mode in a radar image histogram and the form of a Gaussian *pdf*. Indeed, a mode in a radar image histogram is not symmetric, particularly if the number of looks is small. Such a mode is best approximated by an asymmetric *pdf* function such as the Gamma function or the K-function. Thus, the Gaussian mixture can only provide an approximation of the radar image histogram. Nevertheless, this may be sufficient for many segmentation problems (El Zaart 1998). Extension of the present model to improve the mixture of Gamma distribution is a direction for future investigation.

4.4. Optical image histograms

This section deals with the application of our model to the histogram of the optical image illustrated in Figure 1.b. Such an application serves to segment the image in order to separate objects from the background. The modes in the histogram of Figure 1.b. are relatively well separated. Thus, in this case β can be set to both 0.6 and 0.9. The k-means algorithm is applied with number of clusters $K = 9$. Table



a) b)

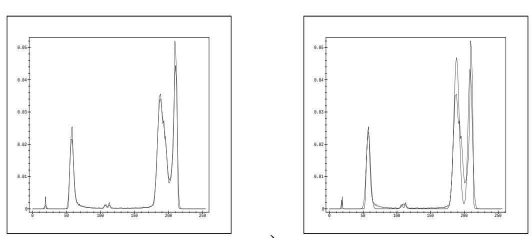
Figure 7: Radar image histogram estimation. a) Smoothing the histogram using the PNN approach with $\sigma = 2.5$, $MSE = 0.00026$. b) Estimation of parameters by our model with $K = 5$ and $\beta = 0.97$, $MSE = 0.0011$.

7 summarizes the results obtained in each step, with $\beta = 0.9$. The histogram was smoothed with $\sigma = 1.5$. Figure 8 shows the reconstructed histogram.

	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 7: Result of our model on the histogram in Figure 1.b, $K = 9$ and $\beta = 0.9$

Our model could not detect the small modes located in the middle of the histogram. However, it was perfectly able to separate the large modes in order to facilitate the extraction of parameters required for segmentation.



a) b)

Figure 8: On optical image histogram estimation. a) Smoothing the histogram using the PNN approach with $\sigma = 1.5$, $MSE = 0.00011$. b) Estimation of parameters using our model with $k = 9$, $\beta = 0.9$, $MSE = 0.00089$.

5. EVALUATION OF THE ALGORITHM

In Section 4, we tested our algorithm on both artificial and real histograms. The results are encouraging, however they cannot reflect a general evaluation of the algorithm itself. Thus, it is necessary to evaluate our model on several histograms. In this section, we will perform two of evaluations. The first evaluation concerns the capability of our model to find the exact number of modes in the histogram, independent of the EM algorithm. This will give us an idea of the capacity of our model to serve as an estimator of the number of modes in a multi-modal histogram. The second evaluation, highlights particularly the effect of false cluster elimination on the performance of our model.

5.1. Experimental data

For both evaluations, we use a procedure similar to the experiment in Section 3.1. However, the generated data (artificial histograms) have to satisfy certain conditions. The aim of these conditions is to ensure that the basic hypothesis that the distribution of each mode is a Gaussian is satisfied. Thus, the generated modes should not totally overlap, in order to maintain the Gaussian distribution behavior. The conditions are given in (Aitnouri 99).

5.2. Estimation of the number of modes

This evaluation aims to determine the preprocessing capability of our model to find the exact number of modes in the histogram. For this purpose, it is sufficient to use only the k-means algorithm followed by the false cluster elimination procedure, without computing the final values of the parameters. This experiment is different from 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. The results of this evaluation are presented in Tables 8 and 9 for a number of combinations of M and K. Table 8 shows the percentage of cases in which the exact number of modes is found. We are also interested in cases where our model makes an error of one mode. Table 9 shows the percentage of cases in which our model finds an number of modes that is different of one from the exact number of modes.

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

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 is large, 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 8, 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 3.1.
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, note that the situation in which our model detects one extra mode, constitutes 73% of cases in Table 9.

Both the 50% for exact detection and the 80% for detection with a possible error of one mode are encouraging if we consider the fact that our approach is based on the k-means algorithm which is not designed to automatically determine the number of modes. Note that the detected clusters correspond to the true modes of the histograms, since their parameters, particularly the centers, are well estimated and their quality was dealt with in section 3.1.

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 9: Percentage of cases where our model finds the real number of modes plus or minus one mode with $\beta = 0.95$ in relation with M and K

5.3. Evaluation of the model

Our model improves the EM algorithm, however it would be imfair to compare directly our model with the EM algorithm (with any pre-processing) since the latter requires initial values for the parameters. For this reason, we make use of the k-means algorithm to set the initial values before the application of the EM algorithm. Consequently, the comparison will reflect basiclly the effect of false cluster elimination. Figure 9.a and Figure 9.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 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 and the accuracy of the false cluster elimination procedure.

6. CONCLUSION

The proposed model significantly reduces the need for human intervention in initializing the parameters in mixture models. Strictly speaking, it requires only that the

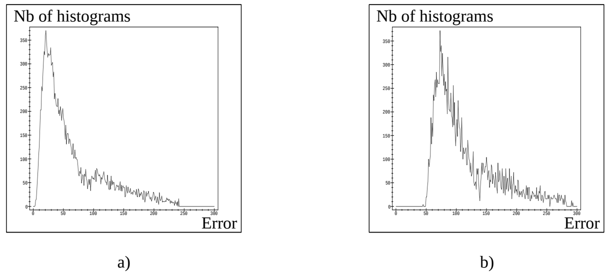


Figure 9: Frequencies of estimated histograms resulting with a given error x 4.10^{-4}. a) Histograms estimated with our model (Average Mean Square Error: $AMSE = 0.0132$), b) Histograms estimated with classical method (Average Mean Square Error: $AMSE = 0.0308$).

parameter β be specified in order to adjust the model's sensitivity to *small* modes luing beside a large mode. This is clearly a parameter that depends on the application in question. This new model offers a promising alternative to systems based on histogram processing, since it is an automatic procedure that represents a whole image signature (histogram) with only a few parameters. The model is robust w.r.t. the estimate of the number of modes M and the presence of noise. A wrong estimate has only a slight effect on the performance of the k-means algorithm, mainly in terms of execution time. The elimination step suppresses most false centers, which greatly facilitates the task of the EM algorithm in the last step.

There are a number of interesting alternatives to the choices made in this model. For example, we could use a more advanced data clustering algorithm such as fuzzy k-means rather 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 of a model (Section 3.2). Other choices (*conservative*, or *statistical average*) are equally valid. The smoothing operation could also be performed using many other techniques. Although all these alternatives offer interesting directions for further works, the most urgent step is to extend the model to a mixture of Gammas, in order to deal with many images whose histograms are not mixtures of Gaussians.

The results obtained suggest that it is possible to make an efficient use of the k-means algorithm to initialize the number of modes in a mixture model. Finally we mention that further investigation is needed to make the model easier to use. In particular, there should be some way to set the value of β in order to avoid the manual choice now required. The relation between M and K also needs to be further clarified. In fact, although $K = M + 3$ seems to be a very good choice, the formula is not necessarily applicable since M may be unknown in practice. How does the model behave when only an estimate of M is available?

REFERENCES

1. Aitnouri, E. M., Wang, S., & Ziou, D., (1998). Estimation of a Multi-modal Histogram's *pdf* by a Mixture Model, *Technical Report* N^o213, Dept. Math. and Comp. Scien., Uni. of Sherbrooke.

2. Aitnouri, E. M., Wang, S., & Ziou, D., (1999). On a Standard Data-Base for Clustering Algorithms Testing, *Technical Report* (in preparation), Dept. Math. and Comp. Scien., Uni. of Sherbrooke.
3. Akaike, H., (1973). 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.
4. C. M. Bishop (1996). *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford University Press.
5. El Zaart, A. , Ziou, D., B. Bénié, G., Wang, S., & Jiang, Q., (1998). Oil Spills Detection in SAR Imagery, *Technical Report* N^o211, Dept. Math. and Comp. Scien., Uni. of Sherbrooke.
6. Dempster, A. P., Lairol, N. M., & Rubin, D. B., (1977). Maximum Likelihood from Incomplete Data, Via the EM Algorithm, *Journal of the Royal Statistical Society*, B 39 [1], pp. 1-38.
7. Duda, R. O., & Hart, P. E., (1973). *Pattern classification and scene analysis*, New York: John Wiley.
8. Guo, R., & Pandit, S. M., (1998). *Automatic threshold selection based on histogram modes and discriminant criterion*, Machine Vision and Applications, Berlin, Springler-Verlag.
9. Jiang, Q., Wang, S., Ziou, D., El Zaart, A., & Bénié, G. B. (1998). Ship Detection in RADARSAT SAR Imagery, In *Proc.* of CMS'98, San Diego.
10. Kelly, P. M., Cannon, M., & Barros, J. E., (1996). Efficiency issues to probability density function comparison, SPIE vol.2670.
11. Kohonen, T., (1995). *Self-Organizing Map*, Berlin: Springler.
12. Kull-back, S., (1959). *Information Theory and Statistics*, New York: Doven publications.
13. Leonardis, A., & Bichof, H., (1998). An efficient MDL-based construction of RBF networks, *Neural Networks* 11 963-973.
14. MacKay, D. J. C., (1992). A Practical Bayesian Interpolation. 4(3), 415-447.
15. Mac Queen, J., (1967). 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.
16. MaLachlan, G. J., & Basford, K. E., (1988). Mixture Models, Inference and applications to clustering, Statistics, textbooks and monographs; V.84.
17. Ormoneit, D., (1993). Estimation of Probability Densities Using Neural Networks, Master Thesis, Technical University of Munich.
18. Parzen, E., (1962). On the Estimation of a Probability Density Function and Mode, Annals of Math. Stat. 33, pp. 1065-1076.
19. Sprecht, D. F., (1990). Probabilistic Neural Network, *Neural Network* 3 (1), 109-118.
20. Stone, M., (1974). Cross-Validatory choice and assessment of Statistical Predictions, *Journal of the Royal Statistical Society* B:, 36(1), 11-147.
21. Tenmoto, H., Kudo, M., & Shimbo, M., (1998). MDL-Based Selection of the Number of Components in Mixture Models for Pattern Classification, *Lecture Notes in Computer Science*, Vol.1451, p.831, 6p.
22. Zhang, J. & Modestino, J. W., (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^o10.
23. Ziou, D., (1993). Optimal Thresholding for Image Segmentation, In *Proc.* 2^{nd} African Conference on Research in Computer Science, Burkina-Fasso.