A Novel EM Algorithm for Stable Optimum

Yang Xiao

Dept. of Computer Science, Tongji University, Shanghai, China

Guorong Xuan

Dept. of Computer Science, Tongji University, Shanghai, China

Zhiqiang Yang

Dept. of Computer Science, Tongji University, Shanghai, China

Yun Q. Shi

Dept. of ECE, New Jersey Institute of Technology, Newark, New Jersey, USA

ABSTRACT: The main contribution of this paper is to propose a novel EM algorithm that utilizes ideas of EM algorithm and maximum-entropy uniform distribution to find a stable optimum in an uncomplicated way. The conventional EM algorithms may suffer from the following two problems: First, it may converge to an undetermined local maximum; second, the algorithm may suffer from singularity. The novel EM algorithm is deterministic that the solution is determined solely by the initial condition. In our novel EM algorithm, a stable and optimal solution can be obtained by using a uniform distribution instead of special initial condition. In addition, a positive perturbation scheme is adopted to avoid singularity. Experimental results have demonstrated that the novel EM is uncomplicated and effective for stable optimum compared with some prior arts.

1 [*]INTRODUCTION

The Expectation Maximization (EM) algorithm is one of the most popular methods for obtaining the maximum likelihood estimate. One of its convenient properties is that it guarantees an increase in the likelihood function in every iteration (Dempster 1977). Moreover, since EM operates on the log-scale, it is analytically simple and numerically stable for distributions that belong to the exponential family such as Gaussian. However, the conventional EM algorithm has some drawbacks. First, it may converge to a local optimum of the likelihood function. Second, the algorithm may suffer from singularity. A great deal of efforts for EM algorithm to obtain the global optimal solution has been made, including the Deterministic Annealing Algorithm (Rose 1998) and the Minimum Message Length (Figueiredo 2002). Our primary goal is to achieve stable optimal solutions in an uncomplicated way when the conventional EM attains locally optimal solutions and other efforts to obtain the globally optimal solutions are rather complicated.

In this paper, a novel EM algorithm is proposed. First, we use the distribution instead of parameter as initial condition so as to easily manipulate the EM iteration. Second, the uniform distribution instead of peak-shape distribution is used to provide the most random initialization. That is, the proposed novel EM algorithm has been devised based on the use of an uniform distribution (having maximum entropy) as initial condition that can achieve global optimality. In addition, a positive perturbation scheme is proposed in our algorithm to avoid suffering from singularity. The proposed scheme has been shown effective in the experimental works.

The rest of the paper is organized as follows. The motivation, diagram and flowchart are introduced in Section 2. We apply our method to several image data sets, experimental results of the novel EM are reported in Section 3. Conclusions are drawn in Section 4.

We first present some notations which are necessary for formulas in this paper：

C *number of components of Gaussian,* $i = 1,\cdots,C$

N *number of Samples (Pixels)*

Y_j *sample vector* $j = 1,\cdots, N$

K *number of feature (Gray levels),* $k = 1,\cdots, K$

P_i *Priori probability*

M_i *Mean vector*

Σ_i *Covariance matrix*

λ *or* λ' *Simplified form of parameters before or after estimation where* λ *or* λ' *means* $[P_i , M_i , \Sigma_i]$

w_{ik} *Weight distribution*

X_k *Feature vector of sample vector* Y_j

$h(X_k)$ *Histogram with feature* X_k

$q_i(X_k)$ *or* $q_i(Y_j)$ *Posteriori distribution for* X_k *or* Y_j *respectively*

q *Simplified form of posteriori distribution*

[*] supported by “National Fundamental Research Funds for the Central Universities” 2012KJ001.

L Likelihood function
$N(x_k | m_i, \sigma_i^2)$ or $N(X_k | M_i, \Sigma_i)$ one or mult − dimension Gaussian Distribution

2 CONVENTIONAL EM ALGORITHM

The Expectation maximization (EM) algorithm is an iterative procedure designed to produce maximum likelihood estimates in incomplete data problems. It has been mathematically proved that the likelihood will increase during each iteration but may not always produce the global optimizer of the likelihood.

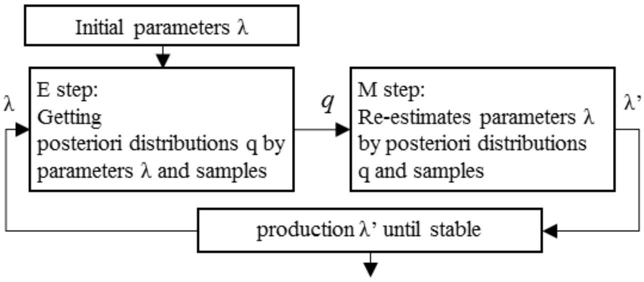

Figure 1. Block diagram of conventional EM

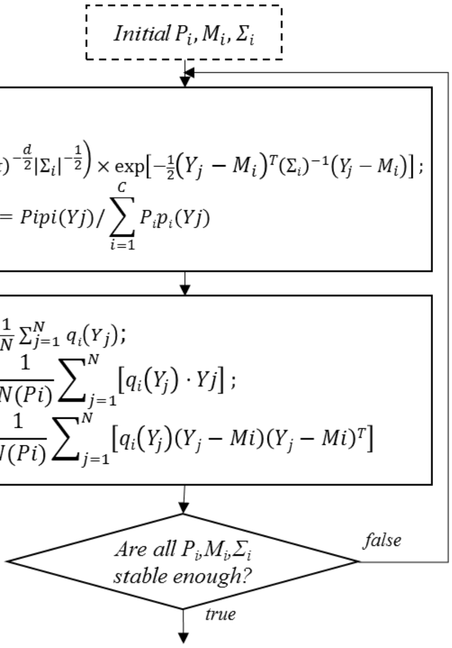

Figure 2. Flowchart of conventional EM

The block diagram of conventional EM for Gaussian Mixture Model (GMM) is shown in Figure 1. The EM iteratively produce estimates to the maximum likelihood estimate (MLE) of the parameters λ, denoted by λ^*, by maximizing the log-likelihood function. In Figure 1, the parameters λ of GMM is set as its initial condition, the iteration will begin with getting conditional distribution and posteriori distributions q and samples in E step. The parameters λ will be re-estimated by posteriori distributions q and samples in M step. It successively alternates between the E step and M step until convergence while the difference or relative difference of successive log-likelihood values fall below a specified tolerance.

The flowchart of conventional EM is shown in Figure 2. The solution generated by EM is determined solely by the initial condition. And depending on its starting value, EM is a locally converging algorithm, which may not produce the global optimizer of the likelihood function. It may suffer from local maximum and singularity.

3 MOTIVATION OF THE NOVEL EM ALGORITHM

First, we use the distribution instead of parameter as the initial condition of EM to control the iterative process more easily. In the conventional EM for GMM, we will set the initial condition for all the parameters: mean vectors, covariance matrices and prior probabilities. Differently, if we use the distribution as the initial condition, we will only need to set up an initial distribution. And, the number of distributions is equal to the number of GMM components, it will be quite simple.
Second, we use the uniform distribution as the initial condition, this will provide us with the most random initialization, and hence with the maximum entropy among various distributions that can be used in the algorithm initialization. This is most important which solves the issue of "local optimum".

4 NOVEL EM ALGORITHM

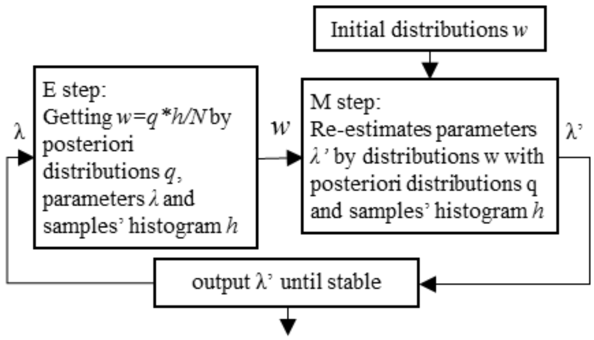

Figure 3. Block diagram of proposed novel EM

The novel EM is based on a realization of uniform distribution which is known to have maximum entropy. The block diagram of the novel EM is shown in Figure 3. The likelihood by histogram *h(Xk)* is defined as:

$L = \sum_{i=1}^{N} ln(p(Y_i|\lambda)) = \sum_{k=1}^{K} h(X_k) ln(p(X_k|\lambda))$,
Where Mixed distribution: $p(X_k|\lambda) = \sum_{i=1}^{C} P_i p_i(X_k)$
And Gaussian distribution: $p_i(X_k) = N(X_k|M_i, \Sigma_i)$

The novel EM's flowchart is shown in Figure 4.
We choose to work with the histogram instead of the samples in order to simplify the iterative computation greatly (Xuan 2001). The formulas in Figure 4 are different from those in Figure 2, i.e., the novel EM works with the histogram $h(X_k)$, while the EM works with sample Y_j. Figure 4 however still corre-

sponds to Figure 3. The proposed novel EM algorithm, on the one hand, utilizes ideas of EM algorithm and maximum-entropy uniform distribution to find a stable and globally solution in an uncomplicated and practical way. On the other hand, the features generated satisfy both uniqueness and similarity requirements for image retrieval, authentication and some forensic tasks.

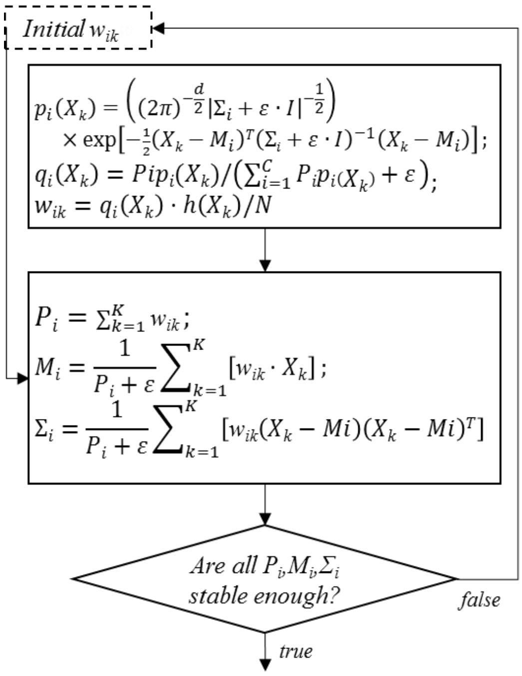

Figure 4. Flowchart of proposed novel EM

5 POSITIVE PERTURBATION SCHEME

The E step and M step of EM algorithm are repeated as necessary. Each iteration is guaranteed to increase the log-likelihood and the algorithm is guaranteed to converge to a local maximum of the likelihood function. Even though the EM algorithm will converge to a (possibly local) maximum, the algorithm may still suffer from the singularity problem, such as the boundary overflow.

Table 1. Comparison of the novel and standard EM

Novel EM proposed	Conventional standard EM
Uniform distribution used as random initialization. It converges to a global optimal solution.	Parameters λ used as random initialization. It may converge to a local maximum.
Adding a very small perturbation to avoid singularities.	Algorithm may suffer from singularity.

The proposed novel EM algorithm solves the issue of singularity by adding a small quantity (e.g., ε=10-20 or other small numbers), which is referred to as a Positive Perturbation Scheme. In this way, the reliable computation can be guaranteed. Adding a small quantity usually takes place for the following situations, e.g., division by a denominator, logarithm, the matrix inverse. Because of these measures are adopted, the summand in iteration will always not be a negative number.

The comparison of the novel EM and standard EM of GMM is shown in Table 1.

6 EXPERIMENTAL WORKS

The experimental works have been presented in Figures 5-8 to illustrate the proposed novel EM algorithm for comparison between using the uniform distribution (b) and peak-shape distributions (c-f) as the initial condition. The sample images are selected from each category from the test dataset (Wang 2003), which contains 1000 generic images in 10 categories, with 100 images in each category. The uniform distributions are w_{ik} = *rand(1,256)/256*.The peak-shape distributions are with different means and fixed standard deviations: $w_{ik} = N(X_k|m,\sigma_i^2)$; *i=1,2; k=(1:256);m=60,100,150,210;σ1=5, σ2=15.* The values of likelihood *L* for five images under different initial conditions are shown in Table 2.

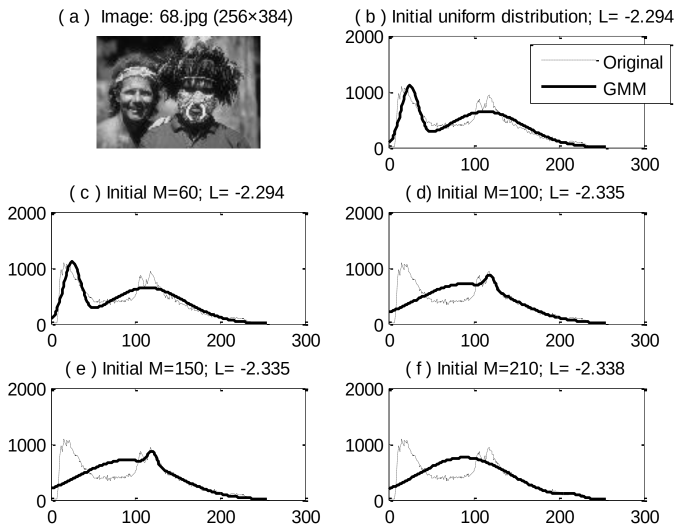

Figure 5. Novel EM of Image 68

Since the uniform distribution is a fixed specific realization of probability distribution, the solution is surely deterministic. In comparison, other solutions are determined by the different peak-shape distributions.

The likelihood L achieved with a uniform distribution as initialization are always equal to the largest one among the maximum likelihood values that can be achieved with the different peak-shape distributions as initialization. The results in Table 2 confirm that for all the images the novel EM could find the global optimal solution every single time. When the peak-shape distributions are used as starting value, it is unlikely that they all yield the exact same solution. Thus, we consider a solution as approximately equal to the global optimum if it achieves the maximum of all the likelihood values with different start values. It has been demonstrated in our experimental works that this novel EM algorithm using a realization of uniform distribution as initial condition can achieve the global maximum for natural images.

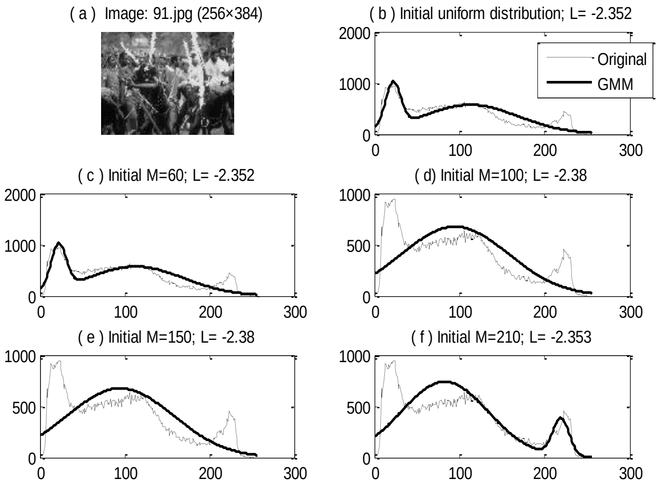

Figure 6. Novel EM of Image 91

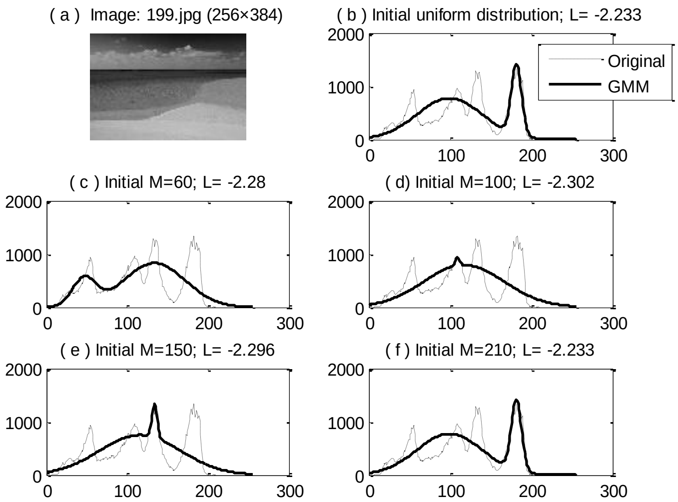

Figure 7. Novel EM of Image 199

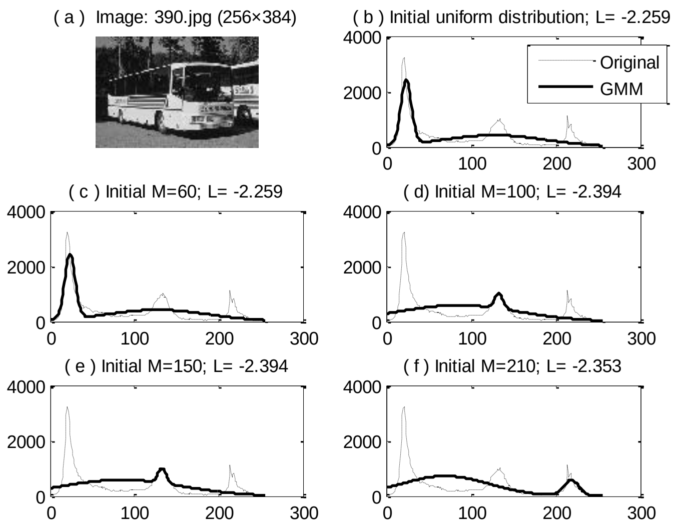

Figure 8. Novel EM of Image 390

Table 2. Likelihood L for Images with different initialization

Images No.	Uniform Distribution	Mean M of peak distribution			
		60	100	150	210
68	-2.294	-2.294	-2.335	-2.335	-2.338
91	-2.352	-2.352	-2.38	-2.38	-2.353
199	-2.233	-2.28	-2.302	-2.296	-2.233
390	-2.259	-2.259	-2.394	-2.394	-2.353

7 CONCLUSION

In this paper we introduce a new a novel EM algorithm to find stable optimal solution. Using a random uniform distribution (known to have maximum entropy) as the initial condition, the novel EM converges iteratively to an optimal solution. Furthermore, the estimation process is repeatable, stable and uncomplicated. For every time, the novel EM algorithm produces stably the optimal solution which could achieve the maximum of all the likelihood values obtained by different initial conditions. A positive perturbation scheme is also proposed to avoid boundary overflow, often occurring with the conventional EM algorithms.

In our Experiments, the novel EM algorithm is used for digital images and the discussion is made with respect to Gaussian Mixture Model (GMM). The histogram has been used in order to speed up computation involved in the novel EM algorithm. Numerical experiments indicate the ability for obtaining the stable optimal solution in uncomplicated way over solutions obtained with some prior arts, including the Deterministic Annealing Algorithm and the Minimum Message Length.

REFERENCES

Bilmes J. A. 1998. A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixed and Hidden Markov Models. U.C. Berkeley. TR-97-021.

Dempster A., Laird N. & Rubin.D.1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society B* 39:1-22,

Figueiredo M. A.T. & Jain A. K. 2002. Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24:381-396.

Li J. and Wang J. Z. 2003.Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1075-1088.

Rose K. 1998. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE* 86 (11): 2210–2239.

Xuan G., Zhang W.& Chai P. 2001. EM algorithm of Gaussian mixture model and hidden Markov model. *IEEE International Conference on Image Processing (ICIP2001)*. vol.1:145-148.Thessaloniki, Greece.