**21st International Conference on Pattern Recognition (ICPR 2012)**
**November 11-15, 2012. Tsukuba, Japan**

# An Enhanced EM Algorithm Using Maximum Entropy Distribution as Initial Condition

*Guorong Xuan*[1], *Yun Q. Shi*[2], *Peiqi Chai*[1], *Patchara Sutthiwan*[2]
[1]Dept. of Computer Science, Tongji University, Shanghai, China
[2]Dept. of ECE, New Jersey Institute of Technology, Newark, New Jersey, USA
grxuan@tongji.edu.cn, shi@njit.edu, ps249@njit.edu

## Abstract

*The conventional EM algorithms may suffer from the following two problems. First, it may converge to a local maximum. Second, the algorithm may suffer from singularity. A novel Enhanced EM algorithm (EEM) using a realization of maximum-entropy uniform distribution as initial condition is proposed. A global optimal solution can be obtained. In addition, a positive perturbation scheme is adopted to avoid singularity. Experimental results have demonstrated that the EEM is simple and effective compared with some prior arts.*

## 1. Introduction

The conventional Expectation maximization (EM) algorithms may suffer from the following two problems. First, it may converge to a local maximum. 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 [1] and the Minimum Message Length [2]. The procedures are rather complicated.

In this paper, an uncomplicated novel Enhanced EM algorithm (EEM) 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. Third, a realization of probability distribution instead of probability distribution itself is utilized so as to obtain the repeatability of the maximum likelihood solution. That is, the proposed EEM algorithm has been devised based on the use of a realization of uniform distribution (having maximum entropy) as initial condition that can achieve global optimality. In addition, a positive perturbation scheme is proposed to avoid the algorithm 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. Some experimental results of EEM are reported in Section 3. Conclusions are drawn in Sections 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
$\Sigma_i$ Covariance matrix
$\lambda$ or $\lambda'$ Simplified form of parameters before or after estimation where $\lambda$ or $\lambda'$ 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
$L$ Likelihood function
$N(x_k \mid m_i, \sigma_i^2)$ or $N(X_k \mid M_i, \Sigma_i)$ one or mult-dimension Gaussian Distribution

## 2. EEM algorithm

### 2.1 Conventional EM algorithm

Expectation maximization (EM) algorithm is a well-known powerful tool for Gaussian Mixture Model (GMM) unsupervised learning [3]. It has been mathematically proved that the likelihood will increase during each iteration. Hence, the EM algorithm will be able converge (to at least a local maximum).

The block diagram of conventional EM is shown in Figure 1 .In Fig. 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 obtained by re-estimates by posteriori distributions q and samples in last iteration. The parameters will be conversed until stable, we have the output parameters.

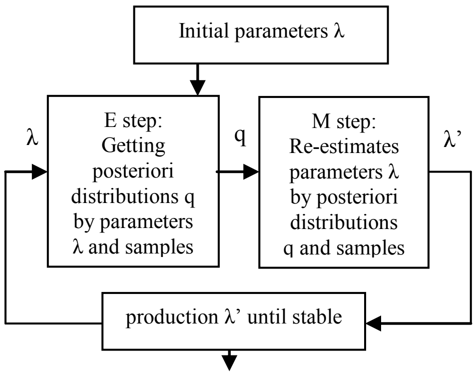


Fig. 1 Block diagram of conventional EM

The flowchart of conventional EM is shown in Fig. 2. The conventional EM algorithms, however, may suffer from local maximum and singularity. It may caused by random parameter initial condition.

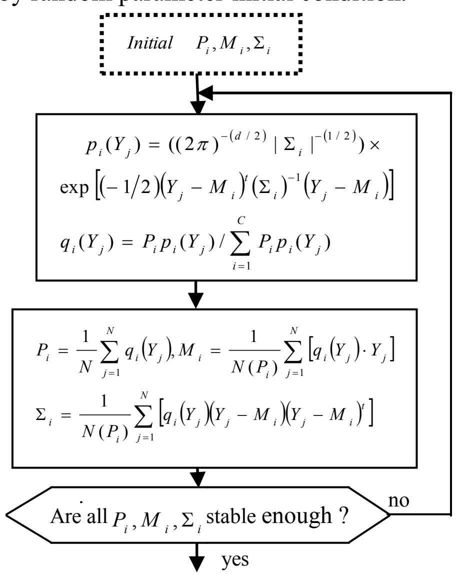


Fig. 2. Flowchart of conventional EM

## 2.2. Motivation of EEM

First, we use the distribution instead of parameter as the initial condition of EM so as to control the iterative process more easily. For example, if we are to estimate the parameters of 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, as we use the uniform distribution as 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 doing solves the issue of "local optimum."

Third, if the initial condition is a fixed specific realization of probability distribution, the solution will surely be deterministic.

## 2.3. EEM algorithm

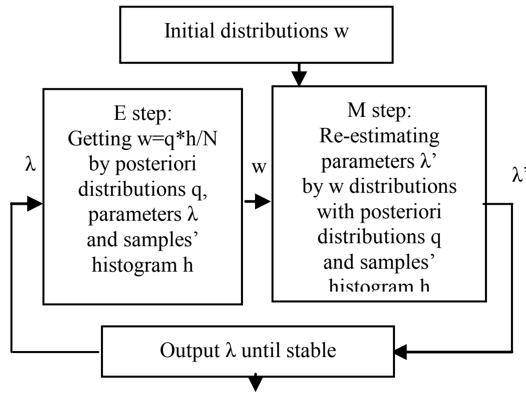


Fig. 3 Block diagram of proposed EEM

The EEM is based on a realization of uniform distribution which is known to have maximum entropy. The block diagram of the EEM is shown in Fig. 3. The EEM's flowchart is shown in Fig. 4.

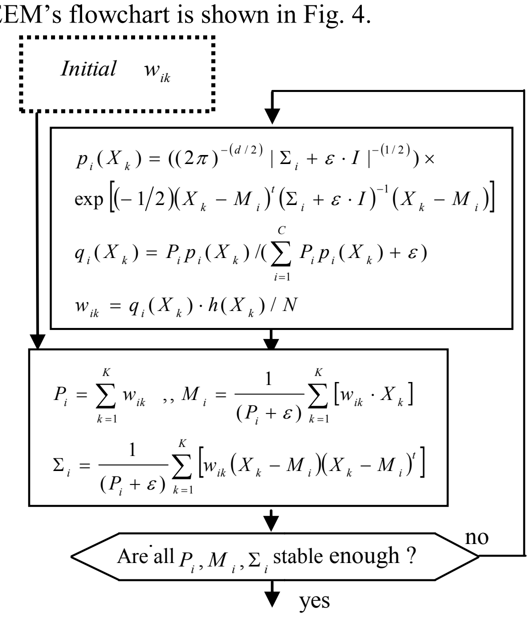


Fig. 4 Flowchart of proposed EEM

We choose to work with the histogram instead of the samples in order to simplify the iterative computation greatly [4]. The formulas in Fig. 4 are different from those in Fig. 2, i.e., in Fig. 4 the EEM works with the histogram $h(X_k)$, while in Fig. 2, the EM works with sample $Y_j$. Fig. 4 however still corresponds to Fig. 3, except using histogram. The proposed EEM algorithm, on the one hand, can solve the above mentioned problems. It is simple and practical. On the other hand, the features generated satisfy both uniqueness and similarity requirements for image retrieval, authentication and some forensic tasks.

### 2.4. An example of iteration of the parameters

We now present an example to show the sequences of iteration of the parameters $\lambda=[P_i,\mu_i,\sigma_i]$ of EEM of Lena in Fig.5

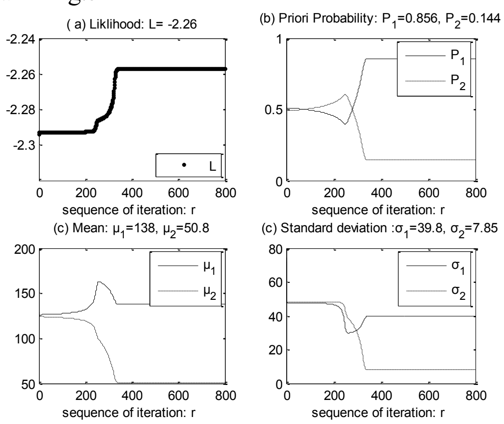


Fig. 5 Sequences of iteration of EEM of Lena

The likelihood by histogram $h(X_k)$ is defined as:

$$L = \sum_{j=1}^{N} \ln\, p(Y_j \mid \lambda) = \sum_{k=1}^{K} h(X_k) \ln\, p(X_k \mid \lambda),$$

$$\text{where Mixed distribution: } \quad p(X_k \mid \lambda) = \sum_{i=1}^{C} P_i\, p_i(X_k)$$

$$\text{and Gaussian distribution: } \quad p_i(X_k) = N(X_k \mid M_i, \Sigma_i)$$

The goodness of histogram fitting G is defined as:

$$G = 10 \log_{10}\left\{ A \times K \Big/ \left( \left( \sum_{k=1}^{K} \left( h(X_k) - N \cdot p(X_k) \right)^2 \right) + \varepsilon \right) \right\}$$

$$\text{where} \quad A = \{Max\,[Max\,(h(X_k)), Max\,(N \cdot p(X_k))]\}^2$$

### 2.5. Positive perturbation scheme

These two steps (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 proved by Dempster at 1977. 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 overflowor singular solutions [2]

The proposed EEM algorithm solves the issue of singularity by adding a small quantity (e.g., $\varepsilon=10^{(-20)}$ 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.

Table 1 Comparison of EEM and standard EM

| **EEM proposed** | **Conventional standard EM** |
|---|---|
| Uniform distribution 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 comparison of EEM and standard EM of GMM is shown in Table 1.

## 3. Experimental Works

The experimental works have been presented to illustrate the proposed EMM algorithm shown in Fig.'s 6-10 for comparison between uniform distribution (b) in Fig.' 6-10 and peak-shape distribution in (c)-(f) in Fig.'s 6-10. As shown in Fig.'s 6-10, the likelihood values, 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.

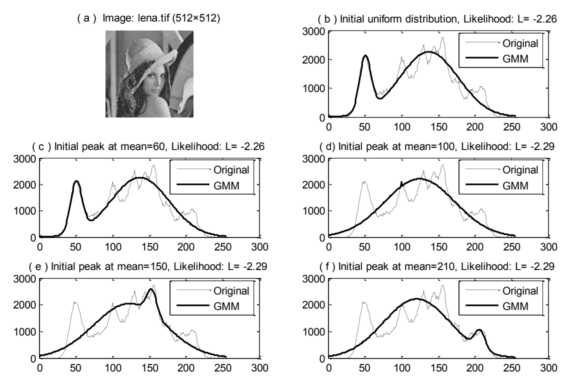


Fig. 6 EEM of Lena

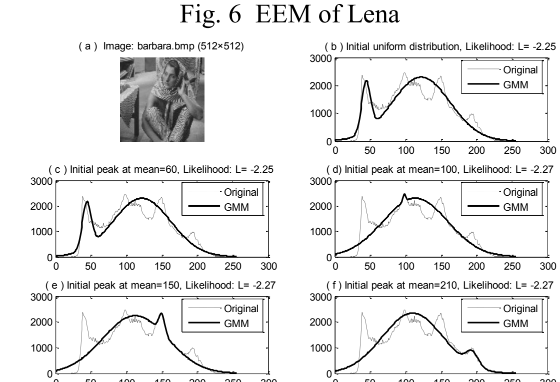


Fig. 7 EEM of Barbara

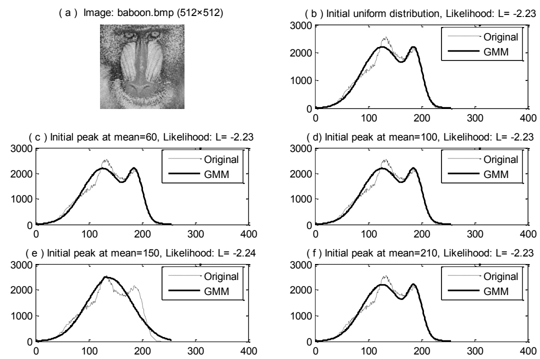


Fig. 8 EEM of Baboon

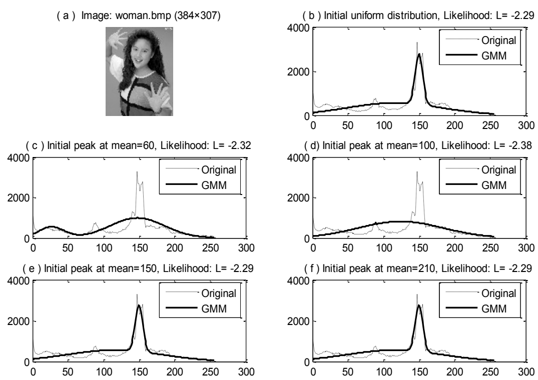


Fig. 9 EEM of Woman

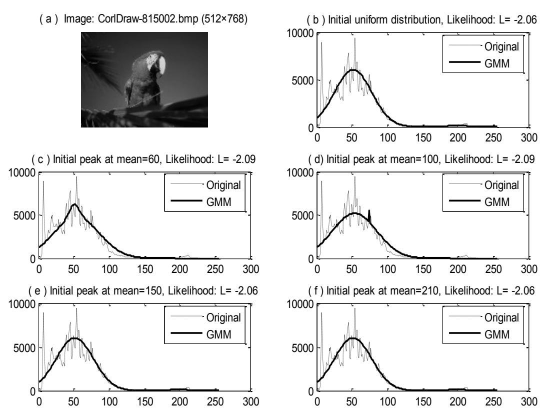


Fig. 10 EEM of CorlDraw-815002

Table 2 Likelihood L and goodness of fitting G

| L and G of images | | Uniform distribution | Mean m of peak distribution | | | |
|---|---|---|---|---|---|---|
| | | | 60 | 100 | 150 | 210 |
| Lena | L | -2.26 | -2.26 | -2.29 | -2.29 | -2.29 |
| | G | 20.72 | 20.72 | 15.62 | 16.50 | 15.49 |
| Barbara | L | -2.25 | -2.25 | -2.27 | -2.27 | -2.27 |
| | G | 18.65 | 18.65 | 16.12 | 16.66 | 16.05 |
| Baboon | L | -2.23 | -2.23 | -2.23 | -2.24 | -2.23 |
| | G | 23.72 | 23.72 | 23.72 | 18.66 | 23.72 |
| Woman | L | -2.29 | -2.32 | -2.38 | -2.29 | -2.29 |
| | G | 23.70 | 18.34 | 16.93 | 23.70 | 23.70 |
| 815002 | L | -2.06 | -2.09 | -2.09 | -2.06 | -2.06 |
| | G | 21.00 | 21.08 | 20.91 | 21.00 | 21.00 |

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), $\sigma_1$=5, $\sigma_2$=15.

The likelihood L and the goodness of histogram fitting G of five images are shown in Table 2. Since the L is in accordance with G all the times; we only use one of them to show performances.

It has been demonstrated in our experimental works that this novel Enhanced EM algorithm (EEM) using a realization of uniform distribution as initial condition can achieve the global maximum for natural images.

## 4. Conclusion

A novel Enhanced EM algorithm (EEM) has been reported in this paper. First, using the uniform distribution (known to have maximum entropy) as the initial condition, the EEM converges iteratively to a global optimality. Furthermore if a realization of the uniform distribution is used as the initial condition, then the process is repeatable. The discussion is made with respect to Gaussian Mixture Model (GMM); the EEM is used for digital images.

The histogram has been used in order to speed up computation involved in the EEM.

A positive perturbation scheme is proposed to avoid boundary overflow, often occurring with the conventional EM algorithms.

Experimental results have demonstrated that the EEM is simple and effective compared with some prior arts, including the Deterministic Annealing Algorithm and the Minimum Message Length.

## References

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

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

[3] J. A. Bilmes ,"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*, April 1998.

[4] G. Xuan , W. Zhang, P. Chai , "EM algorithm of Gaussian mixture model and hidden Markov model," *IEEE International Conference on Image Processing (ICIP2001),* vol.1, pp.145-148, Oct.7-10, 2001, Thessaloniki, Greece