

# Conditional Entropy Maximization for PET \*

**Partha Pratim Mondal**  
Department of Physics  
Indian Institute of Science  
Bangalore 560012, India  
partha@physics.iisc.ernet.in

**K. Rajan**  
Department of Physics  
Indian Institute of Science  
Bangalore 560012, India  
rajan@physics.iisc.ernet.in

**Abstract** - Maximum Likelihood (ML) estimation is extensively used for estimating emission densities from clumped and incomplete measurement data in Positron Emission Tomography (PET) modality. Reconstruction produced by ML-algorithm has been found noisy because it does not make use of available prior knowledge. Bayesian estimation provides such a platform for the inclusion of prior knowledge in the reconstruction procedure. But modeling a prior distribution is a cumbersome task and needs a lot of insight. In this work, we have proposed a conditional entropy maximization algorithm for PET modality, which is a generalization of maximum likelihood algorithm. We have imposed self-normalization condition for the determination of prior distribution. It is found that as prior distribution tends to uniform distribution, the conditional entropy maximization algorithm reduces to maximum likelihood algorithm. Simulated experimental results have shown that images reconstructed using maximum entropy algorithm are qualitatively better than those generated by ML-algorithm.

**Keywords:** Conditional Entropy, Conditional probability, Entropy Maximization, Maximum Likelihood, Poisson Process, Positron Emission Tomography.

## 1 Introduction

Maximum Likelihood estimation is a well known methodology for the estimation of unknown parameters when the outcomes i.e., measurement data are clumped and incomplete [2,3,11,13,14]. In PET [10,16], the reconstruction of the emission densities are sought on the basis of measurement data that are incomplete and clumped. Emission densities are estimated by maximizing the joint distribution function (normally termed as likelihood). The likelihood is defined as the probability of observing the measured data, given the emission parameters. It is found that due to incomplete data acquisition, the reconstruction goes noisier as the estimate approaches maximum likelihood hill. Bayesian approach [4,15] solves this problem considerably by including prior knowledge in the estimation problem via Bayes rule, which relates prior distribution and likelihood function as,

$$P(\lambda/y) = \frac{P(y/\lambda) \cdot P(\lambda)}{P(y)} \quad (1)$$

Bayes rule gives a platform for the possible inclusion of prior distribution knowledge in the estimation problem. Once the likelihood function is available, the next step is to identify a physical meaningful model for the prior distribution, which captures the structural smoothness and spatial interaction. Assuming different functional forms of prior (Gaussian [4], Gamma [7] or Gibbs distribution [15]), researchers have suggested a number of Bayesian algorithms for estimating emission densities.

Section 2.1, gives a brief introduction to the ML-estimation. In section 2.2, we introduce entropy maximization algorithm. Section 3 devotes to the application of conditional entropy maximization to PET. Determination of prior distribution for Positron Emission Tomography is explained in section 4. Simulation results are shown in section 5. Conclusions are drawn in section 6 followed by acknowledgment and references.

## 2 Formatting instructions

### 2.1 Maximum Likelihood Estimation

The most general method used for parameter estimation is the maximum likelihood algorithm [1]. The formulation of this algorithm is as follows :

Let,  $y=(y_1, y_2, \dots, y_n)$  denotes  $n$  random variables representing observables and suppose we are interested in estimating  $m$  unknown parameter  $\lambda=(\lambda_1, \lambda_2, \dots, \lambda_m)$ . The underlying assumption is that the measurement data is somehow related to the unknown parameter  $\lambda$  by a many-to-one mapping from a underlying distribution to the distribution governing the observation. Moreover, it was assumed that joint probability density function  $f(y_1, y_2, \dots, y_n | \lambda)$  depends on  $\lambda$ . Once observable  $y_1, y_2, \dots, y_n$  are given,  $f$  is a function of  $\lambda$  alone, and the value of  $\lambda$  that maximizes the joint probability density function  $f(y_1, y_2, \dots, y_n | \lambda)$  is the maximum likelihood estimate.

The ML-estimate can be obtained by maximizing the log-likelihood function

$$L(y_1, y_2, \dots, y_n | \lambda) = \log f(y_1, y_2, \dots, y_n | \lambda) \quad (2)$$

Maximum likelihood estimate may or may not exist. If  $L(y_1, y_2, \dots, y_n | \lambda)$  is differentiable, then maximum likely-hood estimate exists and it is given by,

$$\left. \frac{\partial \log f(y_1, y_2, \dots, y_n | \lambda)}{\partial \lambda} \right|_{\lambda=\lambda_{ML}} = 0 \quad (3)$$

So, the value of  $\lambda$  for which equation(3) is solvable is the ML-estimate.

## 2.2 Entropy Maximization Estimation

This approach is based on maximizing the conditional entropy (which includes both the likelihood function and prior distribution function) rather than maximizing the likelihood function (i.e, the maximum likelihood of observing the data given the parameter  $\lambda$ ). The point to be noted is that likelihood doesn't utilize the knowledge of prior distribution. On the other hand, entropy includes the prior distribution along with the knowledge of likelihood. Once entropy is maximized, we impose certain constraints to get the functional form of prior distribution based on the available data  $y=(y_1, y_2, \dots, y_n)$ .

The entropy associated with a data set  $y=(y_1, y_2, \dots, y_n)$  assuming the parameter to be estimated is  $\lambda=(\lambda_1, \lambda_2, \dots, \lambda_n)$  is given by,

$$H(y/\lambda) = \sum_i P(\lambda_i) \cdot \sum_j \left( f(y_j/\lambda_i) \log \left( \frac{1}{f(y_j/\lambda_i)} \right) \right) \quad (4)$$

where,  $P(\lambda)$  is the prior distribution and  $f(y/\lambda)$  is the likelihood function.

Now, differentiating with respect to the parameter  $\lambda_i$  we get,

$$\left. \frac{\partial H(y_1, y_2, \dots, y_n | \lambda_1, \lambda_2, \dots, \lambda_n)}{\partial \lambda_i} \right|_{\lambda_i=\lambda_{iME}} = 0 \quad (5)$$

The value of  $\lambda$  for which equ.(5) is true is the maximum entropy estimate ( $\lambda_{ME}$ ). Due to the incompleteness of the problem often the solution comes in the form of a recursion relation.

The determination of the prior distribution  $P(\lambda)$  is accomplished by imposing available normalization or boundary conditions. For positron emission tomography (PET), we have utilized self-normalization property of the problem which preserves total activity of the estimate to be a constant. These constraint are problem specific,

hence proper selection of constraint plays a vital role. In the next section, we will study the entropy maximization approach for PET

## 3 Positron Emission Tomography

The measurements in PET,  $y_j, j=1, \dots, M$  are modeled as independent Poisson random variables with mean parameters given by,

$$\sum_{i=1}^N \lambda_i \cdot p_{i,j}, \quad j = 1, \dots, M$$

where,  $\lambda_i, i=1, \dots, N$  are the mean parameters of the emission process,  $N$  is the total number of pixels and  $p_{ij}$  is the probability of detection of a ray emitted in pixel  $i$  and detected in tube  $j$ .

Since the emission process can be modelled as Poisson process, the likelihood of observed data i.e, the probability of observing the measurement data  $y$  given emission parameter  $\lambda$  is expressed as,

$$f(y|\lambda) = \prod_{j=1}^M \left[ \frac{e^{-\sum_{i=1}^N \lambda_i \cdot p_{i,j}} \left( \sum_{i=1}^N \lambda_i \cdot p_{i,j} \right)^{y_j}}}{y_j!} \right] \quad (6)$$

Substituting equation(6) in equation(4) we get,

$$H(y/\lambda) = - \sum_{i=1}^N P(\lambda_i) \cdot \sum_{j=1}^M \left( \left[ \prod_{j=1}^M \frac{e^{-\sum_{i=1}^N \lambda_i \cdot p_{i,j}} \left( \sum_{i=1}^N \lambda_i \cdot p_{i,j} \right)^{y_j}}}{y_j!} \right] \right) \cdot \left( \sum_{j=1}^M \left( - \sum_{i=1}^N \lambda_i \cdot p_{i,j} + y_j \cdot \log \left( \sum_{i=1}^N \lambda_i \cdot p_{i,j} \right) - \log(y_j!) \right) \right) \quad (7)$$

Taking the partial derivative of  $H(y/\lambda)$  with respect to  $\lambda_i$  and simplification produces,

$$\frac{\partial(H(y/\lambda))}{\partial \lambda_i} = - \sum_y \left[ \frac{\partial P(\lambda_i)}{\partial \lambda_i} P(y/\lambda) \log(P(y/\lambda)) + P(\lambda_i) P(y/\lambda) \log(P(y/\lambda) + 1) - \frac{\partial}{\partial \lambda_i} (\log P(y/\lambda)) \right] \quad (8)$$

Kuhn-Tucker condition for an estimate  $\hat{\lambda}$  to be a maximizer is given by,

$$\lambda_i \frac{\partial(H(y/\lambda))}{\partial \lambda_i} \Big|_{\lambda_i=\hat{\lambda}_i} = 0, \text{ if } \lambda_i > 0 \quad (9a)$$

$$\frac{\partial(H(y/\lambda))}{\partial \lambda_i} \Big|_{\lambda_i=\hat{\lambda}_i} \leq 0, \text{ if } \lambda_i = 0 \quad (9b)$$

for all pixels  $i=1, \dots, N$

Using the Kuhn-Tucker conditions (9a) and (9b) along with the approximation  $\log(P(y/\lambda)) \gg 1$ , we get the following recursive solution for the determination of  $\lambda_i$ 's,

$$\hat{\lambda}_i^{k+1} = \frac{\hat{\lambda}_i^k}{P(\hat{\lambda}_i^k) \cdot \sum_{j=1}^M p_{ij}} \left( \frac{\partial P(\hat{\lambda}_i^k)}{\partial \lambda_i} + P(\hat{\lambda}_i^k) \cdot \sum_{j=1}^M \frac{y_j p_{ij}}{\sum_{i'=1}^N \hat{\lambda}_{i'}^k p_{i'j}} \right) \quad (\text{where, } i = 1, \dots, N) \quad (10)$$

Above approximation is valid for PET system, since  $P(y/\lambda) \ll 1$ .

It can be seen that as,  $p(\lambda_i)$  tends to uniform distribution the recursion relation (10) approaches the ML-algorithm i.e.,

$$\hat{\lambda}_i^{k+1} = \frac{\hat{\lambda}_i^k}{\sum_{j=1}^M p_{ij}} \sum_{j=1}^M \frac{y_j p_{ij}}{\sum_{i'=1}^N \hat{\lambda}_{i'}^k p_{i'j}} \quad (\text{where, } i = 1, \dots, N) \quad (11)$$

Bayesian algorithm also produces ML-algorithm as the prior distribution tends to uniform [4,15]. It is clear that the proposed algorithm (10) is a generalized version of ML-algorithm. We call this as the Maximum Entropy (ME) estimate.

## 4 Extraction of Prior Knowledge

Self-normalization property of preserving the total activity of the estimate a constant gives,

$$\sum_{i=1}^N \hat{\lambda}_i^{k+1} = c, \quad (12)$$

where,  $c$  is a constant representing the total activity of the estimate.

Expanding (12) by substituting equ.(10) we get,

$$\sum_{i=1}^N \frac{\hat{\lambda}_i^k}{P(\hat{\lambda}_i^k) \cdot \sum_{j=1}^M p_{ij}} \left( \frac{\partial P(\hat{\lambda}_i^k)}{\partial \lambda_i} + P(\hat{\lambda}_i^k) \cdot \sum_{j=1}^M \frac{y_j p_{ij}}{\sum_{i'=1}^N \hat{\lambda}_{i'}^k p_{i'j}} \right) = c \quad (13)$$

Simplifying and rearrangement of eqn.(13) produces,

$$\sum_{i=1}^N \frac{\hat{\lambda}_i^k}{P(\hat{\lambda}_i^k)} \frac{\partial}{\partial \lambda_i^k} P(\hat{\lambda}_i^k) = c_1 \quad (14)$$

$$\text{where, } c_1 = c \cdot \sum_{j=1}^M p_{ij} - \sum_{j=1}^M y_j$$

Solving (14) and simplification gives the following form of the prior,

$$P(\hat{\lambda}_i^k) = e^{\alpha \hat{\lambda}_i^k} \quad (15)$$

where,  $\alpha$  is an unknown constant.

It is clear that the prior given by equ.(15) belongs to improper priors. This prior does not integrate to a finite quantity. The problem of improper prior can be meaningful if the posterior probability density function is proper [16]. This is true because the likelihood always dominates.

Substituting  $P(\lambda)$  from equ.(15) in equ.(10) gives,

$$\hat{\lambda}_i^{k+1} = \frac{\alpha \hat{\lambda}_i^k}{\sum_{j=1}^M p_{ij}} + \frac{\hat{\lambda}_i^k}{\sum_{j=1}^M p_{ij}} \sum_{j=1}^M \frac{y_j p_{ij}}{\sum_{i'=1}^N \hat{\lambda}_{i'}^k p_{i'j}} \quad (16)$$

Writing in an additive form for iterative implementation of the algorithm produces,

$$\hat{\lambda}_i^{k+1} = \hat{\lambda}_i^k + \frac{\alpha \hat{\lambda}_i^k}{\sum_{j=1}^M p_{ij}} + \frac{\hat{\lambda}_i^k}{\sum_{j=1}^M p_{ij}} \left( \sum_{j=1}^M \frac{y_j - \sum_{i'=1}^N \hat{\lambda}_{i'}^k p_{i'j}}{\sum_{i'=1}^N \hat{\lambda}_{i'}^k p_{i'j}} \cdot p_{ij} \right) \quad (17)$$

This is the proposed maximum entropy algorithm for PET modality. Similar algorithms can be developed for other forms of emission tomography like single photon emission computed tomography (SPECT).

## 5 Computer Simulation Studies

We have performed computer simulation on a 64x64 test phantom. The simulated PET system consists of 64 detectors in a ring geometry and the object space is decomposed into 64 x 64 square pixels. The object was assumed to be confined within the circle which is confined in the square region of the object, to exploit the eight fold symmetry available in the circular region of the system [10]. Emission from the test image is taken as 100,000. The convolution back projection (CBP) performs much better than ML if the emission count is more than 100,000. For small emission count, ML is found to give better signal to noise ratio compared to that of CBP [8,5,10].

An annihilation event and subsequent emission occurring inside a pixel is governed by a number of physical phenomena that comes into play during the interaction of  $\gamma$ -ray with the matter like scattering, attenuation etc. In this study, we assume that the probability of an emission in a pixel  $i$  and its detection in tube  $j$ ,  $P_{ij}$ , depends only on the geometry of the measurement system. In such a case,  $P_{ij}$  is proportional to

the angle of view,  $\theta_{ij}$ , from the centre of the pixel into the detector tube  $j$ . Shepp [8,10] has shown that the choice of  $P_{ij}$  based on the geometry of the measurement system is reasonable. Due to the multiplicative nature of the recursion relation (eqn.10) we have started the iterative process by considering the initial distribution as uniform distribution.

Determination of correct  $p_{ij}$  is difficult. Equation (17) holds good if correct  $p_{ij}$  values are available. In order to compensate for the error that creeps in the values of  $p_{ij}$ , eqn.(17) has been modified to the following form:

$$\hat{\lambda}_i^{k+1} = \hat{\lambda}_i^k + \frac{\alpha \cdot (\hat{\lambda}_i^k)^\alpha}{\sum_{j=1}^M P_{ij}} + \frac{(\hat{\lambda}_i^k)^\alpha}{\sum_{j=1}^M P_{ij}} \left( \frac{\sum_{j=1}^M y_j - \sum_{i'=1}^N \hat{\lambda}_{i'}^k P_{i'j}}{\sum_{i'=1}^N \hat{\lambda}_{i'}^k P_{i'j}} \cdot P_{ij} \right) \quad (18)$$

Shepp and Logan phantom used for simulation studies is shown in Fig.1(a), and the corresponding histogram is also shown in Fig.1(b). Reconstructed images using ML-algorithm and the proposed maximum entropy (ME) algorithm, along with the histogram plots are shown in Fig.2 and Fig.3 respectively.

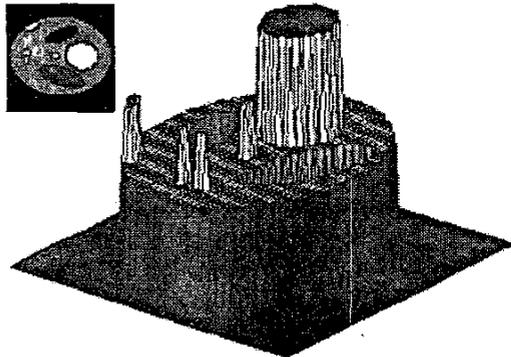


Fig.1. (a) Original Phantom, (b) Histogram plot of Shepp and Logan Phantom

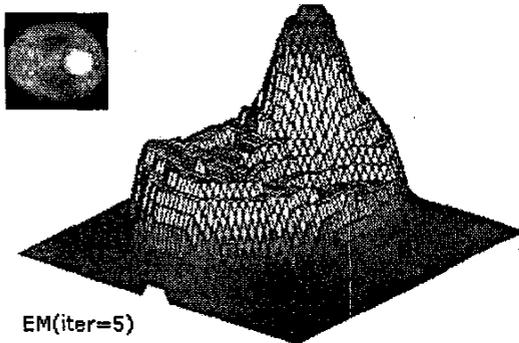


Fig.2. (a) Reconstructed phantom image after 5 EM-iterations, (b) Corresponding histogram

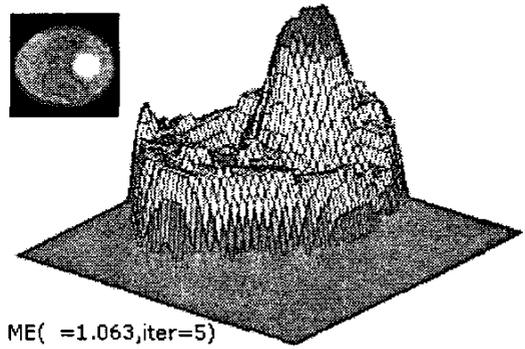


Fig.3. (a) Reconstructed phantom image after 5 ME-iterations ( $\alpha=1.063$ ), (b) corresponding Histogram

Due to the iterative nature, an appropriate measure for the evaluation of the performance of proposed algorithm is a plot of entropy versus iterations rather than likelihood versus iterations. Entropy is given by,

$$H(\hat{\lambda}) = \log \sum_{i=1}^N P(\hat{\lambda}_i) \cdot \sum_{j=1}^M \left\{ \left[ \hat{\lambda}_i \cdot P_{i,j} + y_j \cdot \hat{\lambda}_i \cdot P_{i,j} - y_j \right] \frac{e^{-\sum_{i=1}^N \hat{\lambda}_i \cdot P_{i,j}} \left( \sum_{i=1}^N \hat{\lambda}_i \cdot P_{i,j} \right)^{y_j}}}{y_j!} \right\} \quad (19)$$

The entropy plot (Fig.4) shows that entropy increases with the number of iterations, similar to the log-likelihood plot of EM-algorithm, conforming the convergence of the algorithm. For comparison, the log-likelihood plot is also shown in Fig. 4.

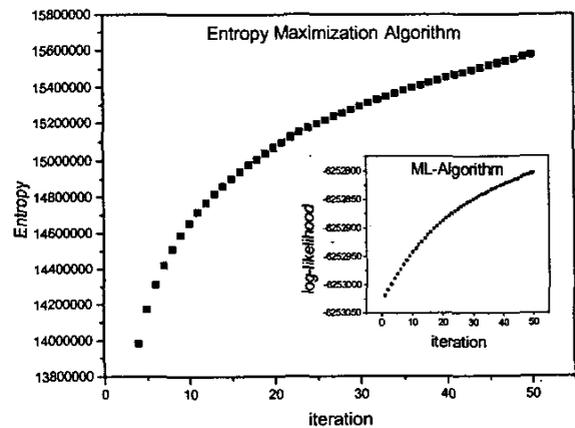


Fig.4. Plot of entropy versus the number of iterations of ME-algorithm showing a definite convergence for  $\alpha=1.063$ .

The quality of images reconstructed using proposed algorithm increases much faster than ML-algorithm with increasing iterations. The image quality starts degrading when the number of iteration is approximately 10 (see fig.5). In the case of ML-estimation, deterioration of the image starts much later at around 20<sup>th</sup> iteration [12]. The quality of the image reconstructed using the proposed algorithm is found to be much better in the early iterations as compared to those generated by ML-algorithm.

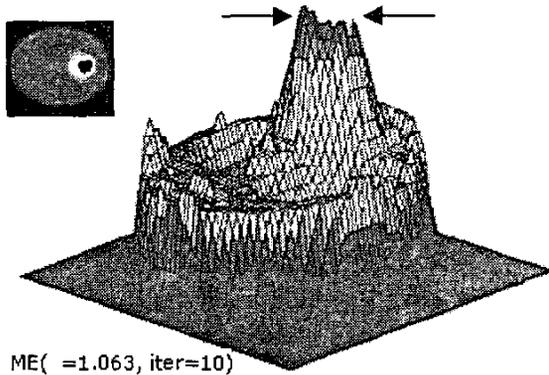


Fig.5. Histogram plot of the reconstructed Shepp and Logan phantom, showing degradation at 10<sup>th</sup> iteration

## 6 Conclusions

We have presented a new algorithm for image reconstruction via conditional entropy maximization. The main contributions of this paper is its generalization and the derivation of prior distribution. From the surface plots of the original phantom (Fig.1), EM reconstructed phantoms (Fig.2), ME reconstructed phantom with  $a=1.063$  (Fig.3) and the entropy vs. iteration plot (Fig.4), the following conclusions have been drawn :

(1) The minute features are missing in the reconstructed images using ML-algorithm, whereas the proposed algorithm retains those features.

(2) Though the entropy associated with the image increases slowly with increasing iterations, the image improvement is very fast using the entropy maximization method.

## Acknowledgement

The author thank T. Hemanth for the help he provided during the course of this work.

Thanks are due to Council of Scientific and Industrial Research for Junior Research Fellowship to the first author. First author declares that this work is partly

supported by Council of Scientific and Industrial Research, New Delhi, India.

## References

[1] A. P. Dempster, N. M. Liard, and D. B. Rubin, "Maximum Likelihood from incomplete data via the EM algorithm," *Jl. Royal Statistical Soc., Ser.B, Vol. 39, NO.1, pp.1-38, 1977.*

[2] A. Katsaggelos and K. Lay, "Maximum likelihood blur identification and image restoration using the algorithm", *IEEE trans. Signal Proc., Vol. 39, No.3, pp.729-733, 1991.*

[3] E. S. Chernoboy, C. J. Chen, M. I. Miller, T. R. Miller and D. L. Snyder, " An evaluation of maximum likelihood reconstruction for SPECT ", *IEEE Trans. Med. Imaging, vol. 9, pp.99-110, 1990.*

[4] E. Levitan and G.T. Herman. A maximum *a posteriori* probability expectation maximization algorithm for image reconstruction in emission tomography. *IEEE Trans. on Medical Imaging, MI-6(3) : pp. 185-192, 1987.*

[5] E. Veklerov and J.Llacer, " Stopping rule for MLE algorithm based on statistical hypothesis testing ", *IEEE Trans. on Med. Img., MI-6 : pp. 313-319, 1987*

[6] Jeff Kershaw, Babak A. Ardekani, and Iwao Kanno. "Application of baysian inference to fMRI data analysis", *IEEE Trans. on Medical Imaging, Vol. 18, No.12, Dec. 1999*

[7] K. Lange, M. Bahn, and R. Little. A theoretical Study of some maximum likelihood algorithm for emission and transmission tomography. *IEEE Trans.on Medical Imaging, MI-6(2) : pp. 106-114, 1987*

[8] K. Lange and R. Carson, " EM reconstruction algorithms for emission and transmission tomography, " *J. Comput. Assist. Tomog. , vol. 8, pp. 306-316, 1984.*

[9] L. A .Shepp and Y. Vardi. Maximum likelihood estimation for emission tomography. *IEEE Trans. on Medical Imaging, MI-1: pp. 113-121, 1982*

[10] L. Kaufman, " Implementing and accelerating the EM algorithm for Positron Emission tomography ", *IEEE Trans. On Med. Img., Vol. MI-6, No.1, pp. 37-51, March , 1987*

[11] M. Segal and E. Weinstein, "Parameter estimation of continuous dynamical linear systems using discrete time observations", *Proc. IEEE, Vol. 7, NO.5, pp.727-729, 1987.*

[12] N. Rajeevan, "*PhD Thesis: A Stochastic estimation approach to emission tomography*", Department of Electrical Engineering, Indian Institute of Science, Bangalore, India, 1991.

[13] P.J. Green, "Bayesian reconstruction from emission tomography data using a modified EM algorithm", *IEEE Trans. on Med. Img.*, vol.9, No.1, March, 1990.

[14] R. Legendijk, J. Biemond, and D. Boeke, "Identification and restoration of noisy blurred images

using the expectation maximization algorithm", *IEEE Trans. ASSP*, Vol. 38, NO. 7, pp.1180-1191, 1990.

[15] T. Hebert and R. Leahy. A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Trans. on Medical Imaging*, MI-8(2) : pp. 194-202, 1989.

[16] Y. Vardi, L. A. Shepp and L. Kaufman, "A Statistical model for Positron Emission Tomography", *Jl. Amer. Stat. Asso.*, Vol.80, pp. 8-37, Mar. 1985.