

# Online Estimation of B-Spline Mixture Models From TOF-PET List-Mode Data

Colas Schretter, Jianyong Sun and Leif Kobbelt

**Abstract**—In emission tomography, images are usually represented by regular grids of voxels or overlapping smooth image elements (*blobs*). Few other image models have been proposed like tetrahedral meshes or point clouds that are adapted to an anatomical image. This work proposes a practical sparse and continuous image model inspired from the field of parametric density estimation for Gaussian mixture models. The position, size, aspect ratio and orientation of each image element is optimized as well as its weight with a very fast online estimation method. Furthermore, the number of mixture components, hence the image resolution, is locally adapted according to the available data. The system model is represented in the same basis as image elements and captures time of flight and positron range effects in an exact way. Computations use approximated B-spline approximations of Gaussians and simple closed-form analytical expressions without any sampling or interpolation. In consequence, the reconstructed image never suffers from spurious aliasing artifacts. Noiseless images of the XCAT brain phantom were reconstructed from simulated data.

**Index Terms**—Time of flight (TOF), positron emission tomography (PET), Gaussian mixture model (GMM), B-spline approximation, online expectation-maximization (EM).

## I. INTRODUCTION

TIME-OF-FLIGHT (TOF) list-mode data acquisitions have recently become the new industry standard in positron emission tomography (PET). With the increasing temporal resolution of TOF-PET scanners, the position of annihilation events are better approximated but established reconstruction methods often fail to exploit fully this important information. While a wealth of publications address the problem of accurate system modeling, very few papers define suitable image models for emission tomography. Notable exceptions include the smooth *blobs* basis functions introduced by Matej and Lewitt [1] and the work of Sitek *et al.* [2], [3].

A fundamental drawback of those image models is that the number of image elements, their position and spatial extent is defined *a-priori* before image reconstruction. Therefore, some important high-intensity regions can be poorly resolved while more emission data are collected there. Simultaneously, redundant image elements cover regions where even no emission event might occur. This paper introduces an alternative continuous and sparse image model for reconstructing adaptive-resolution images from noisy TOF-PET data.

This work was supported by the Marie Curie program of the European Commission under contract PERG06-GA-2009-256519.

C. Schretter and L. Kobbelt are with the Computer Graphics Group, RWTH Aachen University, Germany (email: schretter@cs.rwth-aachen.de).

J. Sun is with the Intelligent Modelling and Analysis Research Group, University of Nottingham, UK (email: jsx@cs.nott.ac.uk).

Our new tomographic image reconstruction method builds upon a synthesis of several previous works. The core of the algorithm is essentially a simplification of a maximum-likelihood method for mixtures of Student's *t*-distributions [4], [5]. This estimation algorithm has been adapted for online estimation with the sliding windows expectation-maximization (SWEM) framework [6]. SWEM is a generalization of the event-by-event OSEM acceleration method [7] for adapting an estimate to new data by discarding progressively the contributions from older events.

## II. METHOD

This section describes a online EM algorithm for event-by-event (EBE) parametric estimation of a multivariate Gaussian mixture model (GMM) when an estimate of measurement errors is available with each list-mode event. Measurement errors are captured by a covariance matrix modeling both TOF localization and space-variant finite resolution effects that are caused by positron range, photon non-collinearity and inter-crystal scattering. Theoretical derivations of the estimation method benefits from the many elegant algebraic solutions for the Gaussian basis function. However, the computational burden inherent to its infinite support is alleviated by a substitution with the cubic B-spline kernel.

The principle of EBE image reconstruction is ideal for list-mode data streams because detected events can be processed directly during acquisition and the convergence is faster. The online parametric estimation framework requires a strategy for allocating dynamically mixture components. This problem is solved with a novel split and merge strategy, sharing similar name but different substance than the work of Ueda *et al.* [8]. Our splitting and merging operations are applied for refining or coarsening local image resolution while guaranteeing monotonous increase in terms of likelihood.

### A. Models for measured data and image representation

The TOF-PET list-mode dataset is the sequence

$$((w_1, x_1, S_1), (w_2, x_2, S_2), \dots) \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^{d \times d}$$

of independent and identically distributed point samples with  $w_n$ ,  $x_n$  and  $S_n$  being the importance, sample's position and covariance matrix of a multivariate Gaussian distribution that model unbiased measurement errors. In TOF-PET data acquisitions, the localization of emissions is very imprecise.

Nevertheless, if a sufficient amount of heteroscedastic noisy measurements is available, it is possible to estimate the most probable emission positions given the current parameters of

an image estimate. For each measurement  $(w_n, x_n, S_n)$ , we associate a latent variable  $t_n$  for representing the corrected emission position with the following hierarchical model:

$$p(x_n|t_n) = \mathcal{N}(x_n|t_n, S_n), \quad p(t_n|k) = \mathcal{N}(t_n|\mu_k, \Sigma_k),$$

with the  $d$ -dimensional multivariate Gaussian kernel:

$$\mathcal{N}(x|\mu, \Sigma) = \frac{(2\pi)^{-\frac{d}{2}}}{\sqrt{|\Sigma|}} \exp\left[-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu)\right].$$

After diffusion of the PET tracer into tissues, the probability to observe an emission event follows an underlying continuous and smooth probability density function. Therefore, quantitative emission images are well modeled by a mixture model:

$$f(t) = \sum_{k=1}^K \pi_k p(t|k).$$

The parameters of this GMM image model is the set

$$\{(\pi_1, \mu_1, \Sigma_1), \dots, (\pi_k, \mu_k, \Sigma_k)\} \in \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^{d \times d}$$

of  $K$  triplets with  $\pi_k$ ,  $\mu_k$ , and  $\Sigma_k$  being the weight, mean vector and covariance matrix of the  $k^{\text{th}}$  freely positioned, sized and oriented smooth Gaussian image element. There an analogy between the data model and the image model.

### B. Online expectation-maximization

Parameters of each component of the mixture model are updated incrementally for each new measurement of importance  $w_n$ , approximate position  $x_n$  and error kernel  $S_n$ . First the E-step computes the posterior distribution *i.e.*, the probability that the emission event occurred in the vicinity of each mixture component. Then, the M-step increments weights, displaces means and updates covariance matrices of image elements.

According to the Bayes rule and after simplifying the product of Gaussian densities, the posterior of the latent variable  $t_n$  is

$$p(t_n|x_n, k) = \frac{p(x_n|t_n, k) p(t_n|k)}{p(x_n|k)} = \mathcal{N}(t_n|\mu_{n|k}, \Sigma_{n|k}),$$

where

$$\begin{aligned} \Sigma_{n|k} &= S_n (S_n + \Sigma_k)^{-1} \Sigma_k = (S_n^{-1} + \Sigma_k^{-1})^{-1}, \\ \mu_{n|k} &= \Sigma_{n|k} (S_n^{-1} x_n + \Sigma_k^{-1} \mu_k), \end{aligned}$$

The computation of expectations are extremely simple in our case since the convolution of two Gaussians is still Gaussian:

$$p(x_n|k) = \int p(x_n|t_n) p(t_n|k) dt_n = \mathcal{N}(x_n|\mu_k, S_n + \Sigma_k).$$

Therefore, for each of the  $K$  components, relative ownerships are computed for the current data sample  $x_n$ ,  $S_n$  by

$$p(k|x_n) = \frac{\pi_k \mathcal{N}(x_n|\mu_k, S_n + \Sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(x_n|\mu_j, S_n + \Sigma_j)},$$

and according to the *recursive* (incremental) maximization scheme of Titterton [9], the update of weight, mean and

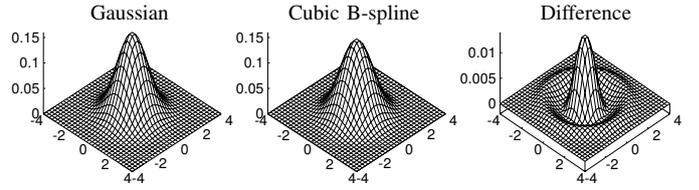


Fig. 1. The finite-supported cubic B-spline kernel approximates very well the Gaussian distribution. The difference plot emphasizes the slight dissimilarities.

covariance matrix for each component follows:

$$\begin{aligned} \pi_k &\leftarrow \pi_k + p(k|x_n) w_n, \\ \mu_k &\leftarrow \mu_k + p(k|x_n) \frac{w_n}{\pi_k} [\mu_{n|k} - \mu_k], \\ \Sigma_k &\leftarrow \Sigma_k + p(k|x_n) \frac{w_n}{\pi_k} [\tilde{\Sigma}_{n|k} - \Sigma_k], \end{aligned}$$

with

$$\tilde{\Sigma}_{n|k} = [(\mu_{n|k} - \mu_k)(\mu_{n|k} - \mu_k)^\top + \Sigma_{n|k}].$$

### C. Active learning with sliding window

In the maximization step, the mean and covariance matrix of each component is updated with an influence that is inversely proportional to its weight. In consequence, the importance of new data is decreasing according to a geometrical decay law to ensure convergence. However, it is often wanted to use a bounded amount of statistical information for estimation. Learning from limited data is often used for the sake of acceleration as in the case of OSEM. In dynamic image reconstruction for example, only the most recent events should contribute to the estimate.

A common solution for active learning is using a linear *forgetting rate*  $\alpha = 1/T$  for a chosen constant  $T > 0$ . The weight of each component decays exponentially by multiplication with  $\alpha$  after each processed data element [10]. Unfortunately, decaying weights never reach zero and this scheme often suffers from numerical inaccuracies. In alternative, we use a more accurate sliding window strategy, dubbed SWEM. With SWEM, a history of the most recent weight increments is recorded by a simple bookkeeping mechanism until their integral reaches  $T$ . Then, older contributions can be subtracted seamlessly and replaced by the newest expected weights.

In practice, a certain storage granularity is required for limiting storage or memory requirements. The most recent history is accumulated in 64 memory pages and older contributions are removed sporadically by small batches. The granularity should be chosen according to available memory resources but virtually no difference in terms of speed and accuracy was observed when using 32 pages instead of 64. A formal description of SWEM with an additional window expansion mechanism can be found in [6].

### D. Model selection with split and merge

The image model defined above assumed that the number of image elements  $K$  is known beforehand. Instead, we want to select locally the image resolution and decide both the value

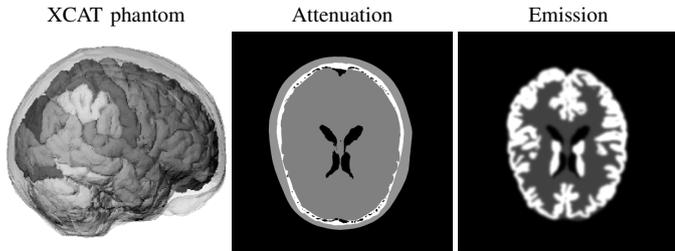


Fig. 2. High-resolution attenuation and emission images are rasterized from the NURBS surface representation of the segmented XCAT brain phantom.

of  $K$  and the initialization of parameters in a principled way. Starting from a unique component whose weight is initialized to zero, the image model is locally refined according to the amount of emission events, *i.e.*, the weight of components.

Given a target weight  $W > 0$ , components are split in two if their weights exceed  $2W$ . The new components have identical parameters but their means are displaced in opposite direction. The weights of split components are halved, their means are displaced and their matrix are squashed such that

$$\pi_k \leftarrow \frac{1}{2}\pi_k, \quad \mu_k \leftarrow \mu_k \pm e_k \sqrt{\frac{1}{2}\lambda_k}, \quad \Sigma_k \leftarrow Q_k \Sigma_k,$$

where  $\lambda_k$  and  $e_k = (u_k, v_k)$  are respectively the dominant eigenvalue and eigenvector of the covariance matrix in *e.g.*, two dimensions and  $Q_k$  is an anisotropic squashing operator:

$$Q_k = \begin{bmatrix} u_k & -v_k \\ v_k & u_k \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_k & v_k \\ -v_k & u_k \end{bmatrix} \\ = \begin{bmatrix} 1 - \frac{1}{2}u_k^2 & -\frac{1}{2}u_k v_k \\ -\frac{1}{2}u_k v_k & 1 - \frac{1}{2}v_k^2 \end{bmatrix}.$$

The displacement and the squashing deformation are chosen such that the likelihood of the mixture model do not decrease and such that the sum of the two split components remains unimodal. The determinant of the covariance matrix is also exactly divided by two after splitting. Moreover, a merge operation removes non-essential components whenever their weights fall below  $W/2$ . The weight and the history of removed components are distributed among remaining neighbors according to the sharing proportions derived by Vlassis and Verbeek [11]. Those proportions are similar to the expectations  $p(k|x_n)$  but the trace of the covariance matrix of the removed component is involved in calculations.

### E. Cubic B-spline approximation

Presented as it is, the estimation method would require an exceeding amount of computational resources since every mixture component is modified for each measurement. In our implementation, evaluations of the normal distribution are substituted by the following cubic B-spline kernel:

$$\mathcal{N}(x|\mu, \Sigma) \approx \frac{(2\pi)^{-\frac{d}{2}}}{\sqrt{|\Sigma|}} \begin{cases} \frac{2}{3} - t^2 + \frac{1}{2}t^3 & t < 1 \\ \frac{4}{3} - 2t + t^2 - \frac{1}{6}t^3 & 1 \leq t < 2 \\ 0 & \text{otherwise} \end{cases}$$

with

$$t = \left[ \frac{c_3}{3} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right]^{1/2}.$$

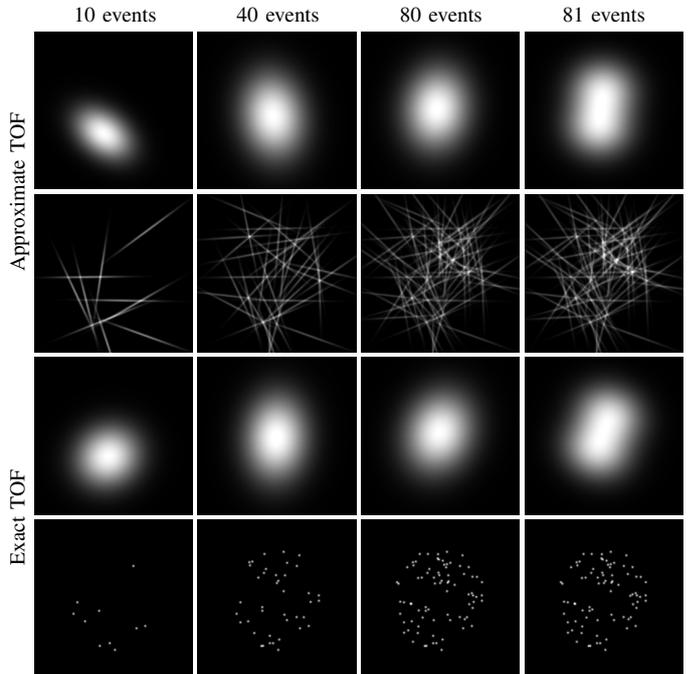


Fig. 3. Visualization of the early estimation process until the first component splitting. The mixture models (above) and the error distributions of emission data (below) are compared with the case of exact TOF measurements.

Note that the support of this approximation is apodized after four standard deviations. In two and three dimensions, the factor  $c_3$  is respectively equal to the inverse of the area and volume integral of the non-corrected ( $c_3 = 1$ ) radially-symmetric cubic B-spline kernel. Hence,  $c_3 \simeq 0.948965$  in two dimensions and  $c_3 \simeq 0.900025$  in three dimensions. Such correction is not required for substitutions of one dimensional mixture components. This approximation is very accurate, as illustrated in Fig. 1 and shown theoretically in [12] and [13].

## III. RESULTS

An emission image of the XCAT brain phantom shown in Fig. 2 was reconstructed from 40 millions simulated detection events. The Monte Carlo simulation modeled Gaussian random walks between emission and annihilation and photon non-collinearity with a FWHM of 2.8mm. A TOF precision of 600ps corresponding to a FWHM of about 9cm was modeled as well. The attenuation probability was derived by conventional raytracing in the anatomical image. We observed that about 24.36 emission events were rejected before registering a coincidence event. Modeling more advanced physical phenomena such as photon scattering will be added in future.

The effect of component splitting is illustrated by comparing the third and last columns in Fig. 3. No attenuation was simulated for this example such that every event have the same unit importance. We remark that very similar component's parameters are estimated when an exact TOF information is given. For the image reconstruction, the importance of each event is set to the inverse of its attenuation probability, compensating for the more important sampling of line of responses corresponding to low attenuation. The distributions

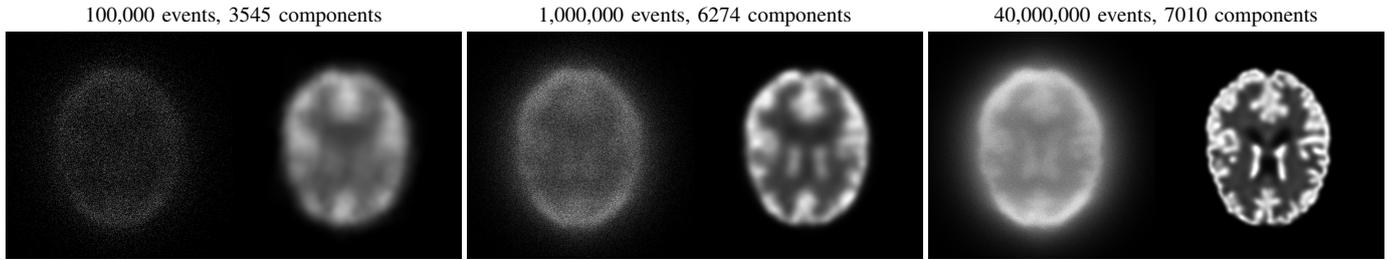


Fig. 4. The mixture model (right) is estimated from the stream of noisy TOF-PET measurements (left) and matches progressively the ground truth in Fig. 2.

of measurements errors are elongated multivariate Gaussians positioned and oriented along line or responses.

The emission image was reconstructed with various values of the parameters  $T$  and  $W$ . Some intermediate steps of the image formation process are shown in Fig. 4 for  $T = 4, 000, 000$  and  $W = 500$ , hence  $T/W = 8000$ . After processing 40 millions events, the mixture size reached 7010 components with min. and max. weight of 258.9 and 982.5. The mean component weight was 568.0 with std. deviation of 143.8.

Fig. 5 shows quantitative analyzes of the progressive convergence towards the ground truth emission image. The convergence is faster when using a smaller sliding window of a total weight  $T$  of two millions (approx. 82,000 events). The convergence speed is two times slower with  $T$  equals to four millions. However, after convergence, the accuracy of parametric estimation and the image similarity are always better with larger values of  $T$ .

The image dissimilarity is measured in terms of Kullback-Liebler (KL) divergence with a rasterized image of the mixture model, matching the high-definition reference emission image (0.5 mm pixels). The KL divergence is an appropriate figure of merit for assessing images that are proportional to probability density functions [7]. Since the GMM image model is continuous, rasterizing close-up images for regions of interest does not introduce any interpolation artifacts.

#### IV. CONCLUSION AND FUTURE WORK

This work describes and evaluates a new approach for image reconstruction from time of flight list-mode PET data. Parameters of a continuous and sparse image representation are estimated by an online algorithm. The system model captures exactly positron's random walks and imprecise TOF measurements. The image model is a mixture of multivariate Gaussian basis functions and yields closed-form expressions for both the E and M steps of maximum-likelihood expectation-maximization methods. Results show that estimating large mixtures is tractable when substituting Gaussians with finite-support cubic B-spline polynomial kernels.

Future work will investigate a scale-space image representation with hierarchical mixtures for real-time reconstructions. A hierarchy of mixture components allows for accessing very quickly mixture components, given a query point and its associated error distribution represented by a covariance matrix. This multiresolution image model will be used for interactive visualization as well.

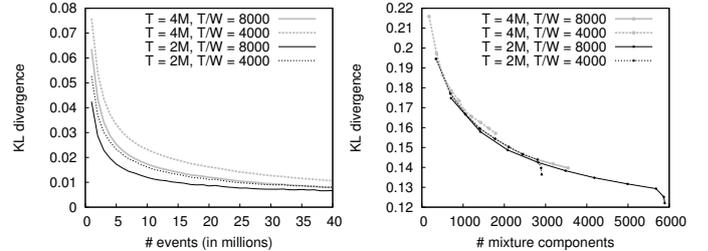


Fig. 5. The convergence towards the ground truth emission image is measured in terms of Kullback-Liebler divergence for various values of  $T$  and  $W$  (left). The simultaneous improvement of image similarity with the increasing number of mixture components is also shown for the 100,000 first events (right).

#### ACKNOWLEDGMENT

We thank Arman Rahmim for sharing the XCAT human brain phantom and providing experimental configuration files.

#### REFERENCES

- [1] S. Matej and R. M. Lewitt, "Efficient 3D grids for image reconstruction using spherically-symmetric volume elements," *IEEE Transactions on Nuclear Science*, vol. 42, no. 4, pp. 1361–1370, 1995.
- [2] A. Sitek, R. H. Huesman, and G. T. Gullberg, "Tomographic reconstruction using an adaptive tetrahedral mesh defined by a point cloud," *IEEE Transactions on Medical Imaging*, vol. 25, no. 9, pp. 1172–1179, 2006.
- [3] N. F. Pereira and A. Sitek, "Evaluation of a 3D point cloud tetrahedral tomographic reconstruction method," *Physics in Medicine and Biology*, vol. 55, no. 9, pp. 5341–5361, 2010.
- [4] J. Sun, A. Kabán, and S. Raychaudhury, "Robust mixtures in the presence of measurement errors," in *Proc. International Conference on Machine Learning*, pp. 847–854, 2007.
- [5] J. Sun and A. Kabán, "A fast algorithm for robust mixtures in the presence of measurement errors," *IEEE Transactions on Neural Networks*, vol. 21, no. 8, pp. 1206–1220, 2010.
- [6] C. Schretter, "Online PET reconstruction from list-mode data," in *Proc. of the Fully3D*, pp. 195–198, 2009.
- [7] C. Schretter, "Event-by-event image reconstruction from list-mode PET data," *IEEE Transactions on Image Processing*, vol. 8, no. 1, pp. 117–124, 2009.
- [8] N. Ueda, R. Nakano, Z. Ghahramani, and G. E. Hinton, "SMEM algorithm for mixture models," *Neural Computation*, vol. 12, no. 9, pp. 2109–2128, 2000.
- [9] D. M. Titterton, "Recursive parameter estimation using incomplete data," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 46, no. 2, pp. 257–267, 1984.
- [10] Z. Zivkovic and F. van der Heijden, "Recursive unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 651–656, 2004.
- [11] N. Vlassis and J. J. Verbeek, "Gaussian mixture learning from noisy data," Tech. Rep. IAS-UVA-04-01, University of Amsterdam, 2004.
- [12] M. Unser, "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Magazine*, vol. 16, no. 6, pp. 22–38, 1999.
- [13] R. Brinks, "On the convergence of derivatives of B-splines to derivatives of the Gaussian function," *Computational and Applied Mathematics*, vol. 27, no. 1, pp. 79–92, 2008.