BAYESIAN SEGMENTATION OF CHEST TUMORS IN PET SCANS USING A POISSON-GAMMA MIXTURE MODEL

Zacharie Irace, Marcelo Pereyra, Nicolas Dobigeon and Hadj Batatia

University of Toulouse, IRIT/INP-ENSEEIHT, 31071 Toulouse Cedex 7, France

{zacharie.irace, marcelo.pereyra, nicolas.dobigeon, hadj.batatia}@enseeiht.fr

ABSTRACT

This paper presents a Bayesian algorithm for PET image segmentation. The proposed method, which is derived from PET physics, models tissue activity using a mixture of Poisson-Gamma distributions. Moreover, a Markov field is proposed to model the spatial correlation between mixture components. Then, segmentation is performed using an Markov chain Monte Carlo algorithm that jointly estimates the mixture parameters and classifies voxels. The performance of the proposed algorithm is illustrated on synthetic and real data. Experimental results on real chest PET images suggest that the proposed method can correctly segment both small and large tumors.

Index Terms— PET imaging, Poisson-Gamma, Negative Binomial, Mixture model, Markov-Potts, Bayesian estimation, Gibbs sampler.

1. INTRODUCTION

Positron Emission Tomography (PET) is a nuclear imaging modality that quantifies biological activity, an extremely useful information for tumor assessment in clinical oncology. However, quantification of tumor properties requires precise PET segmentation algorithms, which is a challenging task mainly due to the low SNR and strong artifacts intrinsic of these images.

Numerous PET segmentation methods have been proposed in the literature, including thresholding [1], clustering [2], active contours [3], among several others. Moreover, recent work has addressed PET segmentation in a Bayesian framework. For instance, Chen *et al.* proposed a Markov Random Field [4]. Gaussian mixtures have been used in [5] and Gauss-Markov-Potts models in [6, 7]. At last, in [8], the authors proposed a Poisson mixture model for PET projections and developed a double *expectation-maximization* algorithm to solve PET reconstruction by combining anatomical information.

Starting from PET physics, this work derives a new Bayesian algorithm for PET image segmentation based on a mixture of Poisson-Gamma distributions. Moreover, a Markov field is proposed to model the spatial correlation between mixture components. Then, as in [9], segmentation is performed by jointly estimating the mixture parameters and classifying voxels.

2. PROPOSED BAYESIAN MODEL

Let $r_n \in \mathbb{N}$ denote the amount of radioactivity received from the n_{th} voxel in a 3-D PET scan $\boldsymbol{r} \in \mathbb{N}^N$. We shall assume that $\boldsymbol{r}$ is composed by multiple biological tissues $\{C_1, ..., C_K\}$, each with its own characteristic radioactivity. In addition, we explicitly introduce a hidden label vector $\boldsymbol{z} = \{z_1, \ldots, z_N\}$ associated with the observation vector $\boldsymbol{r} = \{r_1, \ldots, r_N\}$ such that $z_n = k$ if $r_n \in C_k$. Then, it is possible to state the segmentation problem as follows:

$$\hat{\boldsymbol{z}} = \underset{\boldsymbol{z}}{\operatorname{argmax}} \quad p(\boldsymbol{\theta}, \boldsymbol{z}|\boldsymbol{r}) \tag{1}$$

where $\boldsymbol{\theta}$ is a vector of parameters associated with a parametric image formation model $p(\boldsymbol{r}|\boldsymbol{z}, \boldsymbol{\theta})$. We note that $\boldsymbol{\theta}$ is considered unknown and its posterior density is estimated jointly with the posterior related to the label vector $\boldsymbol{z}$. The likelihood $p(\boldsymbol{r}|\boldsymbol{z}, \boldsymbol{\theta})$ and the prior $p(\boldsymbol{z}, \boldsymbol{\theta})$ are defined in what follows.

2.1. Likelihood

It has long been established that r_n can be correctly modeled as a Poisson random variable [10]:

$$r_n \sim \mathcal{P}(\lambda_n) \tag{2}$$

where $\mathcal{P}$ denotes the Poisson distribution and λ_n is the mean radioactivity within the n_{th} voxel. It easily follows that the likelihood could be expressed as a finite Poisson mixture if λ_n were forced to be uniform within each biological tissue. However, biological activity is not homogeneous within tissues, rendering the Poisson mixture model inaccurate. In this study we consider intra-tissue heterogeneity by assigning independent radioactivities to each voxel. More specifically, we propose to model radioactivities within a tissue as independent and identically distributed random Gamma variables:

$$\lambda_n | z_n = u \sim \Gamma(\alpha_k, \beta_k) \tag{3}$$

where $\alpha_k \in \mathbb{R}^+$ and $\beta_k \in \mathbb{R}^+$ are respectively the shape and scale parameters associated with tissue C_k.

This distribution choice is motivated by the fact that the marginal distribution $\int_0^\infty \mathcal{P}(r_n|\lambda_n)\Gamma(\lambda_n|\alpha_k,\beta_k)d\lambda_n$ yields the Poisson-Gamma or Negative Binomial distribution:

$$\begin{aligned} p(r_n|\alpha_k,\rho_k,z_n=k) &= \int_0^\infty \mathcal{P}(r_n|\lambda)\Gamma(\lambda|\alpha_k,\beta_k)d\lambda \\ &= \binom{r_n+\alpha_k+1}{r_n}(1-\rho_k)^{\alpha_k}\rho_k^{r_n} \end{aligned} \tag{4}$$

where $\rho_k = \frac{1}{1+\beta_k}$. To conclude, we assume the observations are mutually independent and express the likelihood as follows:

$$p(\boldsymbol{r}|\boldsymbol{\theta},\boldsymbol{z}) = \prod_{n=1}^{N} p(r_n|\alpha_k,\rho_k,z_n=k) \tag{5}$$

where $p(r_n|\alpha_k,\rho_k,z_n=k)$ is defined in (4),
$\boldsymbol{\alpha} = \{\alpha_1,\ldots,\alpha_K\}^T$, $\boldsymbol{\rho} = \{\rho_1,\ldots,\rho_K\}^T$ and $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T,\boldsymbol{\rho}^T)^T$.

2.2. Prior Distribution

The unknown parameter vector for this problem is written as $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T,\boldsymbol{\rho}^T)^T$. We note that $\rho_k \in]0,1[$ since $0 < \beta < \infty$. Moreover, the number of classes K is assumed to be known in this study. This assumption might be relaxed by using a dimension matching strategy, such as a reversible jump Markov chain Monte Carlo (MCMC) algorithm [11]. The prior distributions assigned to the unknown parameters as well as the label vector are introduced below.

2.2.1. Poisson-Gamma priors

First, the prior for the shape parameter α is elected as an inverse Gamma distribution with hyperparameters a_0 and b_0

$$\alpha_k \sim \mathcal{IG}(a_0,b_0), \quad k=1,\ldots,K \tag{6}$$

where the hyperparameters are fixed to $a_0 = 1$ and $b_0 = 1$, yielding a vague prior.
A conjugate Beta distribution is chosen as prior for ρ_k:

$$\rho_k \sim \mathcal{B}(c_0,d_0), \quad k=1,\ldots,K \tag{7}$$

where the hyperparameters are fixed to $c_0 = 1$ and $d_0 = 1$ yielding a flat prior.
At last, assuming the mixture parameters a priori independent, the joint prior distribution for the vector $\boldsymbol{\theta}$ is:

$$p(\boldsymbol{\theta}) = \prod_{k=1}^{K} p(\alpha_k)p(\rho_k) \tag{8}$$

2.2.2. Labels $\boldsymbol{z}$

The prior on the label vector $p(\boldsymbol{z})$ should promote the spatial coherence inherent to biological tissues. Since the seminal work of Geman [12], Markov Random Fields (MRF) have become a popular prior choice to express spatial correlation in images. In this study we consider a 3-D Markov-Potts field as prior distribution for $\boldsymbol{z}$ [13]:

$$p(\boldsymbol{z}) = \frac{1}{C(\gamma)}\exp\left[\sum_{n=1}^{N}\sum_{n'\in\mathcal{V}(n)}\gamma\delta(z_n - z_{n'})\right] \tag{9}$$

where γ is the granularity coefficient, $C(\gamma)$ is the normalizing constant or partition function, $\delta(\cdot)$ is the Kronecker function and $\mathcal{V}(\cdot)$ denotes a 3-D first-order neighborhood structure .

2.3. Posterior Distribution of $(\boldsymbol{\theta},\boldsymbol{z})$

Assuming the unknown parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{z}$ are a priori independent and using Bayes theorem, the posterior distribution of the parameter vector $(\boldsymbol{z},\boldsymbol{\theta})$ can be expressed as:

$$p\left(\boldsymbol{\theta},\boldsymbol{z}|\boldsymbol{r}\right) \propto p(\boldsymbol{r}|\boldsymbol{\theta},\boldsymbol{z})p(\boldsymbol{\theta})p(\boldsymbol{z}) \tag{10}$$

where the likelihood $p(\boldsymbol{r}|\boldsymbol{\theta},\boldsymbol{z})$ and the prior distributions $p(\boldsymbol{\theta})$ and $p(\boldsymbol{z})$ have been defined in (5), (8) and (9), respectively. Deriving explicit solutions to the segmentation problem (1) is not possible because of the complexity of the posterior distribution (10). Instead, we proceed as in [9] and use an MCMC algorithm to generate samples that are asymptotically distributed according to posterior (10) [14]. Then, these samples can be used to approximate the maximum a posteriori (MAP) estimator, as in [15].

3. HYBRID GIBBS SAMPLER

This section presents a Metropolis-within-Gibbs sampler that solves the segmentation problem (1). The principle of this algorithm is to generate samples that are asymptotically distributed according the posterior density (10) and use those samples to compute the MAP estimate. The generation of these samples is achieved by iteratively drawing from the conditional distributions of the posterior distribution (10) provided below. The interested reader is referred to [14] for more details about MCMC methods.

3.1. Conditional probability $p(\boldsymbol{z}|\boldsymbol{\theta},\boldsymbol{r})$

The conditional distribution of the discrete class label z_n is fully characterized by the probabilities

$$\begin{aligned} &p(z_n = k|r_n,\alpha_k,\rho_k,\boldsymbol{z_{-n}}) \\ &\qquad \propto p(r_n|\alpha_k,\rho_k,z_n=k)p(z_n|\boldsymbol{z_{-n}}) \end{aligned} \tag{11}$$

where $k = 1, \ldots, K$ and $\boldsymbol{z_{-n}}$ denotes the vector $\boldsymbol{z}$ whose n_{th} element has been removed. In view of equations (4) and (9) it is possible to express these probabilities as follows:

$$\begin{aligned} p(z_n = k | r_n, \alpha_k, \rho_k, \boldsymbol{z_{-n}}) & \\ \propto \binom{r_n + \alpha_k + 1}{r_n} (1-\rho_k)^{\alpha_k} \rho_k^{r_n} & \\ \times \exp\left[\sum_{n=1}^{N} \sum_{n' \in \mathcal{V}(n)} \gamma \delta(z_n - z_{n'}) \right] & \end{aligned} \tag{12}$$

Note that the posterior probabilities of the label vector $\boldsymbol{z}$ in (12) define an MRF. Consequently, sampling from this conditional distribution can be achieved by drawing a discrete value in the finite set $1, \ldots, K$ with probabilities (12).

3.2. Conditional distribution of $p(\boldsymbol{\alpha}|\boldsymbol{\rho}, \boldsymbol{z}, \boldsymbol{r})$

Moreover, a Metropolis-Hasting (MH) algorithm can be used to generate samples that are asymptotically distributed according to $p(\boldsymbol{\alpha}|\boldsymbol{\rho}, \boldsymbol{z}, \boldsymbol{r})$, leading to a Metropolis-within-Gibbs algorithm [14, p. 317]. More specifically, $\boldsymbol{\alpha}$ is updated coordinate-by-coordinate using a random walk MH algorithm [14, p. 245] with the following proposal distribution:

$$\alpha_k^* \sim \mathcal{N}(\alpha_k^{(t-1)}, \sigma^2). \tag{13}$$

where $\alpha_k^{(t-1)}$ is the previous value of the chain and the variance σ^2 is chosen to ensure an acceptance ratio close to $\frac{1}{2}$, as recommended in [16]. Moreover, given that the proposal distribution is symmetric, the acceptance ratio is reduced to the likelihood and the prior ratios

$$\mathrm{a} = \min\left\{1, \prod_{n:z_n=k}^{N} \frac{p(r_n|\alpha_k^*, \rho_k, z_n = k) p(\alpha_k^*|a_0, b_0)}{p(r_n|\alpha_k^{(t-1)}, \rho_k, z_n = k) p(\alpha_k^{(t-1)}|a_0, b_0)}\right\}$$

where the prior distribution $p(\alpha_k|a_0, b_0)$ is defined in (6).

3.3. Conditional distribution of $p(\boldsymbol{\rho}|\boldsymbol{\alpha}, \boldsymbol{z}, \boldsymbol{r})$

At last, we draw samples directly from the conditional distribution $p(\boldsymbol{\rho}|\boldsymbol{\alpha}, \boldsymbol{z}, \boldsymbol{r})$, which has the following expression ($k = 1, \ldots, K$):

$$\rho_k \sim \mathcal{B}\left(c_0 + N_k \alpha_k, d_0 + \sum_{n:z_n=k} r_n\right). \tag{14}$$

4. EXPERIMENTAL RESULTS

The proposed method has been validated using synthetic data and evaluated on real patient PET images. This section presents the results of these simulations and experimentation.

4.1. Synthetic Data

A synthetic 2-component Poisson-Gamma mixture ($50 \times 50 \times 5$) 3D-map has been generated with $\boldsymbol{\rho} = [0.1; 0.2]$ and $\boldsymbol{\alpha} = [5; 6]$. A single MCMC chain of $30,000$ iterations (including 250 burn-in iterations) has been applied to this data. Figure 1 shows histograms of the estimated posterior density for the unknown parameters. This illustrates the accuracy of the estimations despite the fact that parameters are very close.

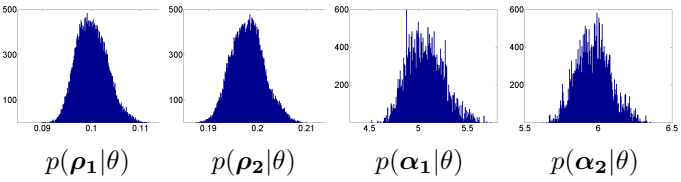

Fig. 1. Histograms of parameters resulting from the MCMC chain.

Moreover, in order to assess the performance of the algorithm, 100 independent Monte Carlo runs (each of 500 iterations after 250 burn-in iterations) have been performed with parameters $\boldsymbol{\rho} = [0.2; 0.5]$ and $\boldsymbol{\alpha} = [7; 12]$. MMSE estimates of each parameter and corresponding standard deviations were $\hat{\boldsymbol{\rho}} = [0.199 \pm 0.008; 0.490 \pm 0.014]$ and $\hat{\boldsymbol{\alpha}} = [6.96 \pm 0.36; 11.58 \pm 0.64]$. As it can be observed, the estimates are in agreement with the theoretical parameters.

Finally, labeling has been illustrated on a 3-component spatially coherent simulated dataset with $\boldsymbol{\rho} = [0.2; 0.5; 0.8]$ and $\boldsymbol{\alpha} = [7; 12; 14]$. Figure 2 shows the true labels $\mathbf{z}$, the generated data, and the MAP estimates. One can observe the good retrieval of the classes (0.09% of misclassified pixels).

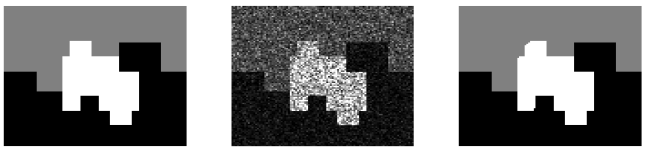

Fig. 2. True labels, generated values and MAP estimates for a 3-class mixture.

4.2. Application to real data

The proposed segmentation method has been applied to real chest PET images acquired using the GE-DST system. Figure 3 shows three slices of a region of interest ($116 \times 151 \times 3$) and the corresponding MAP class labels estimated using a 5-class MCMC chain with 250 iterations and 250 additional burn-in iterations. We observe that two tumors have been segmented (white regions). These are in agreement with the clinician's evaluation. Delineation of these regions using the widely used thresholding method would have required a threshold value of 13% as opposed to 40% used by clinicians. These results

illustrate the robustness of the proposed method to heterogeneous activity. In addition, we observe that the different biological tissues have also been labeled distinctly.

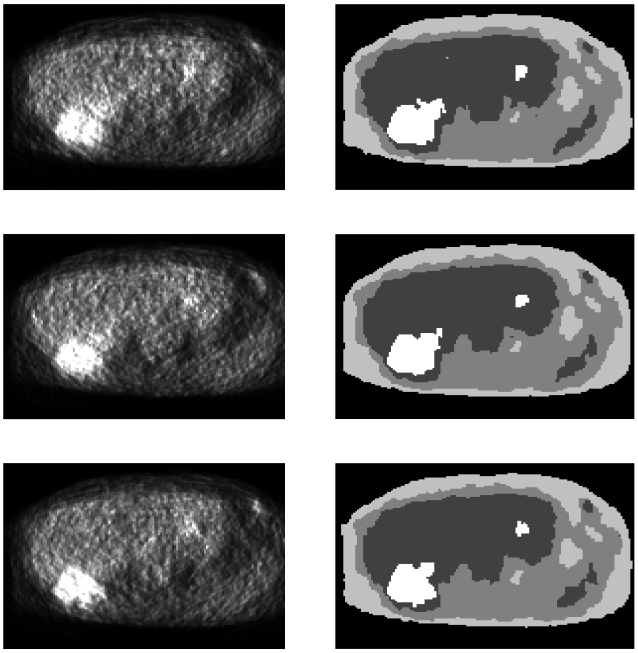

Fig. 3. Original images and the corresponding estimated labels.

At last, figure 4 shows that the probability density function estimated using the proposed mixture model (in red) closely fits the empirical density (in blue), which has been estimated using a Gaussian kernel method.

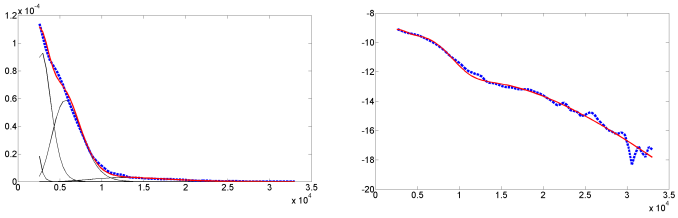

Fig. 4. Voxel-intensity distribution in linear (left) and logarithmic (right) scales. [Red]: proposed Bayesian estimation. [Dashed blue]: non-parametric kernel estimation

5. CONCLUSION

Starting from PET physics, a Poisson-Gamma mixture model has been proposed to represent the activity in PET images. In addition, a Markov field was used to describe the spatial correlation between mixture components. Then, segmentation was performed using an MCMC algorithm that jointly estimates the mixture parameters and classifies voxels. Experimental results on real chest PET images suggested that the proposed method outperforms the widely used 40% thresholding technique. At last, future work will focus on designing characterization indicators based on the posterior density of activity within tumors.

6. ACKNOWLEDGMENTS

The authors would like to thank Prof. J.-Y. Tourneret (University of Toulouse, IRIT) for interesting discussion regarding this work. They are also very grateful to Dr. F. Courbon and his team at ICR for their help with the images.

7. REFERENCES

[1] L. Drever, W. Roa, A. McEwan, and D. Robinson, "Iterative threshold segmentation for PET target volume delineation," *Med. Phys.*, vol. 34, pp. 1253–1265, 2007.

[2] M. E. Kamasak and B. Bayraktar, "Clustering dynamic PET images on the projection domain," *IEEE Trans. Nucl. Sci.*, vol. 53, no. 3, pp. 496–503, 2007.

[3] N. Joshi and M. Brady, "Non-parametric mixture model based evolution of level sets and application to medical images," *Int. J. Comp. Vis.*, vol. 88, pp. 52–68, 2010.

[4] J. L. Chen, S. R. Gunn, M. S. Nixon, and R. N. Gunn, "Markov random field models for segmentation of PET images," in *Proc. Inf. Proc. Med. Im.*, vol. 2082, 2001, pp. 468–474.

[5] M. Aristophanous, B. Penney, M. Martel, and C. A. Pelizzari, "A gaussian mixture model for definition of lung tumor volumes in positron emission tomography," *Med. Phys.*, vol. 34, no. 11, pp. 4223–4235, 2007.

[6] D. Montgomery, A. Amira, and F. Murtagh, "An automated volumetric segmentation system combining multiscale and statistical reasoning," in *Proc. IEEE Int. Symp. Circ. and Syst.*, vol. 4, 2005, pp. 3789–3792.

[7] D. W. G. Montgomery, A. Amira, and H. Zaidi, "Fully automated segmentation of oncological PET volumes using a combined multiscale and statistical model," *Med. Phys.*, vol. 34, no. 2, pp. 722–736, 2007.

[8] A. Rangarajan, I. Hsiao, and G. Gindi, "A Bayesian joint mixture framework for the integration of anatomical information in functional image reconstruction," *J. Math. Im. Vis.*, vol. 12, pp. 199–217, 2000.

[9] M. A. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tourneret, "Labeling skin tissues in ultrasound images using a generalized Rayleigh mixture model," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Proc. (ICASSP)*, Prague, Czech Republic, May 2011.

[10] L. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. 1, no. 2, pp. 113–122, May 1982.

[11] P. J. Green, "Reversible jump Markov chain Monte Carlo methods computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, Dec. 1995.

[12] S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov 1984.

[13] F. Y. Wu, "The potts model," *Rev. Mod. Phys.*, vol. 54, no. 1, pp. 235–268, Jan 1982.

[14] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 1999.

[15] N. Dobigeon and J. Y. Tourneret, "Bayesian orthogonal component analysis for sparse representation," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2675–2685, May 2010.

[16] G. O. Roberts, "Markov chain concepts related to sampling algorithms," in *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Eds. London: Chapman & Hall, 1996, pp. 259–273.