A Bayesian Joint Mixture Framework for the Integration of Anatomical Information in Functional Image Reconstruction

Anand Rangarajan†, Ing-Tsung Hsiao‡, and Gene Gindi‡

30th December 1998

†Departments of Electrical Engineering and Diagnostic Radiology, Yale University, New Haven, CT.

‡Departments of Electrical Engineering and Radiology, SUNY at Stony Brook, Stony Brook, NY.

Abstract

We present a Bayesian joint mixture framework for integrating anatomical image intensity and region segmentation information into emission tomographic reconstruction in medical imaging. The joint mixture framework is particularly well suited for this problem and allows us to integrate additional available information such as anatomical region segmentation information into the Bayesian model. Since this information is independently available as opposed to being estimated, it acts as a good constraint on the joint mixture model. After specifying the joint mixture model, we combine it with the standard emission tomographic likelihood. The Bayesian posterior is a combination of this likelihood and the joint mixture prior. Since well known EM algorithms separately exist for both the emission tomography (ET) likelihood and the joint mixture prior, we have designed a novel EM^2 algorithm that comprises two EM algorithms—one for the likelihood and one for the prior. Despite being dove-tailed in this manner, the resulting EM^2 algorithm is an alternating descent algorithm that is guaranteed to converge to a local minimum of the negative log Bayesian posterior. Results are shown on synthetic images with bias/variance plots used to gauge performance. The EM^2 algorithm resulting from the joint mixture framework has the best bias/variance performance when compared with six other closely related algorithms that incorporate anatomical information to varying degrees.

Keywords: Mixture modeling, emission tomography, Bayesian reconstruction, expectation-maximization (EM), alternating algorithms, joint mixtures, algebraic transformations, conjugate priors, gamma distributions, bias/variance performance, region segmentation, projection data, Radon transform

1 Introduction

Maximum likelihood (ML) reconstruction in positron emission tomography (PET) and single-photon emission computed tomography (SPECT) (Shepp and Vardi, 1982) is hampered by the loss of information due to noise and other factors in the projection imaging process. One approach to obtaining improved reconstructions is to use priors (Hebert and Leahy, 1989) that reflect the nature of the underlying radionuclide distribution. Recently, in emission tomography (ET), there has been considerable interest in incorporating side information derived from highly correlated anatomical information, such as co-registered magnetic resonance (MR) images, in the form of Bayesian priors (see extensive review in Section 2.) The main attraction of this data fusion approach is that one expects to obtain improved reconstructions to the extent that function follows anatomy.

While the correlation of the spatial distribution of the intensities in functional images and their co-registered anatomical counterparts is often high, there is no relationship between the actual *values* of the two intensity patterns in common anatomical regions. Consequently, we can expect to gain more purchase by using the anatomical segmentation *label* information rather than the raw anatomical *intensity* as prior information in tomographic reconstruction. This correlation between function and anatomy is not perfect: there may be regions in the functional image that do not have an anatomical counterpart. For example, there could be a lesion which is apparent in the radionuclide distribution as a localized "hot spot" of high intensity pixels without a corresponding signature lesion in the anatomical label map. Therefore, we need a prior model which takes into account the good, but not perfect, correlation between a segmented and co-registered anatomical image and the evolving emission tomographic reconstruction.

We can capture the notion of anatomy-function correlation via the *joint histogram* formed from the raw intensity values of the anatomical and functional images. Qualitatively, we would expect the joint histogram to comprise several localized peaks, each corresponding to a *joint cluster* of anatomy and function pixels. Where the anatomy-function correlation is high, the volume of the peaks is large. For a function class with no anatomical counterpart, e.g. functional lesions, there could be multiple low volume peaks in the joint histogram. If there are K intensity classes in function and L intensity classes in anatomy, there is a cross-product of as many as $K \times L$ classes that may appear in the joint histogram. However, since function follows anatomy for the most part, the actual number of peaks seen in the joint histogram would be much less than $K \times L$.

The joint histogram is seen typically as a non-parametric estimate of the joint density between anatomy and function. In this paper, we have chosen to use a joint mixture framework to model the joint density between anatomy and function. As opposed to a non-parametric density estimator, the joint mixture framework allows us to associate a parametric density function for each joint anatomy-function class. We also have more information than is present in the joint histogram: often we have a very good estimate of the anatomical segmentation labels which act as a strong constraint on the joint mixture model. We also employ a prior on the joint mixture *occupation probability* to express the tight coupling between anatomy and function. The Bayesian maximum *a posteriori* (MAP) estimate then has the novel interpretation of including *mutual information* maximization between the anatomy and function.

We approach the ET reconstruction problem from a Bayesian perspective wherein the prior is formulated using the aforementioned joint mixture model. We use the familiar Poisson likelihood as a forward model for generating the observed projection data from the underlying radionuclide source distribution. Then, the log-posterior is proportional to the sum of the Poisson log-likelihood and the joint mixture log-prior. Our overarching goal is a MAP estimate of the functional image. The parameters of the joint mixture prior are unknown and also have to be estimated from the projection data (and the known anatomical labels). Since the Poisson log-likelihood is maximized by the well known emission tomographic (ET) expectation-maximization (EM) algorithm (Shepp and Vardi, 1982) and well known EM algorithms exist for mixture decomposition (McLachlan and Basford, 1988; Redner and Walker, 1984), it is natural to combine these two algorithms into one EM algorithm. The result is an alternating descent algorithm, which we term the EM^2 algorithm, which consists of an EM outer loop dealing with the likelihood and an EM inner loop dealing with the joint mixture model parameters. It is guaranteed to find a local minimum of the Bayesian posterior cost function, by virtue of being an alternating descent algorithm.

2 Previous Work

The integration of anatomical side information into the ET reconstruction, a problem of some practical import especially in brain imaging, has been addressed by several groups in the past decade. The strategies fall mainly into two categories: those that employ explicit boundary information from the anatomical image, and those that employ implicit boundary information derived from anatomical tissue classification labels at each pixel. In both cases, the anatomy-function correlation is premised on the physiologically plausible assumption of approximately uniform activity in distinct anatomical regions.

In the first category, loci of discontinuities from the co-registered anatomical image are used to spatially modulate the formation of discontinuities in the ET reconstruction. For one class of these methods, the mathematical machinery for imposing side information is couched within a Bayesian MAP estimation problem that uses an edge-preserving Gibbs prior. A spatially varying weight (for line-site or discontinuity formation) modulates the Gibbs potential so as to encourage the formation of discontinuities in the ET reconstruction at sites corresponding to discontinuities in the anatomical image. Implementational details differ, but such strategies have been employed in (Leahy and Yan, 1991; Gindi et al., 1993; Ouyang et al., 1994; Lipinski et al., 1997). A related strategy was employed in (Ardekani et al., 1996) in which an adaptive edge-preserving smoothing filter is calculated using multispectral segmented MR data and applied to the ET reconstruction with the aid of a cross-entropy objective function. All of these approaches are distinguished in the ways they model uncertainty in the location of anatomical discontinuities, errors in the registration, and mismatched resolution of the anatomical and functional images. In cases where the anatomy-function correlation is high, these methods work well.

The basic idea in the second approach is similar, but here, use is made of tissue label information at each pixel in the co-registered anatomical image (e.g. “gray”, “white” and “csf” labels derived from a segmented MR brain scan). Through a variety of mathematical strategies, local differences in the label field may be used to again modulate the strength of a smoothing prior. At loci of label discontinuities, the smoothing weight is lowered and discontinuities are allowed to form. Since a given pixel may be composed of different tissues, non-binary “partial-volume” labels are typically used in these models. In addition, some models posit a uniform concentration of radionuclide for each tissue class, thus accounting for the spatial variation of radionuclide by a weighted (by spatially varying partial-volume tissue labels) sum of activities in each class. At the interface of anatomical regions, usually corresponding to gradual transitions in ET intensity, this use of partial-volume modeling is efficacious for accurate ET reconstruction. This strategy was used in the form of a post correction to a conventionally reconstructed PET brain scan in (Meltzer et al., 1990; Videen et al., 1988; Muller-Gartner et al., 1992), with a later, more sophisticated extension of this approach reported in (Rousset et al., 1998). The incorporation of label information within a Bayesian MAP estimation procedure is reported in (Sastry and Carson, 1997). In a different Bayesian MAP approach (Hsu and Leahy, 1997), pixel labels that are not dominated by one tissue class are flagged and used to invoke a second-order “thin-plate” prior (Lee et al., 1997), corresponding in the ET reconstruction to graded intensity regions: otherwise, conventional first-order smoothing is imposed. This strategy leads to an ET reconstruction containing realistic smooth transition regions separating those areas of uniform activity corresponding to relatively pure tissue types. Again, these models work well where the anatomy-function correlation is high.

More complex models, not clearly identitified with either of the two categories, have been proposed. In (Zhang et al., 1994), a joint MAP estimation approach is proposed wherein, as with other approaches, anatomical label differences are used to modulate discontinuities in ET. Unlike other approaches, a separate provision is made to account for the possibility of localized intensity regions in the emission reconstruction that do not correspond to label regions in the anatomy image. Some user interaction with an initial ET reconstruction is required, however. An elegant but very elaborate hierarchical Bayesian model proposed in (Bowsher et al., 1996) is able to robustly account for exceptions to the presumed anatomy-function correlation, and, in fact, utilizes anatomical information in the form of an initial estimate. However, the model expressiveness is balanced by the need to estimate several hyperparameters on whose values the performance depends.

The framework proposed in this paper is motivated by a desire to capture some of the modeling flexibility afforded in (Zhang et al., 1994) and (Bowsher et al., 1996), but without the need for user interaction or difficult hyperparameter determination.

Finally, this work is similar in spirit to recent work on unsupervised category learning using mixture models in the neural net literature (Tenenbaum and Freeman, 1997; de Sa and Ballard, 1998). The themes are very similar; for example, in (Tenenbaum and Freeman, 1997), a mixture model is used to model the joint space of font styles and types and in (de Sa and Ballard, 1998), the focus is squarely on the unsupervised learning of categories from multimodality data.

3 A Bayesian Joint Mixture Framework

In this section, we begin with a brief description of the standard likelihood model for emission tomographic reconstruction. We then proceed with the development of the joint mixture model. After describing the overall joint mixture framework in some detail, we specialize the mixture parametric density to a gamma distribution. With the likelihood and priors specified, we then describe the EM^2 algorithm—an alternating descent algorithm that minimizes the negative log-posterior.

3.1 The Likelihood Model

We now write down the likelihood for our problem. The likelihood specifies the forward model for generating the observed projection data from the underlying radionuclide (PET or SPECT) source distribution. We use the conventional model of Poisson noise at the detector (Shepp and Vardi, 1982) with the standard assumption that the detector bin counts are independent of one another. Therefore, the likelihood is

$$G \sim \text{Poisson}(\mathcal{H}f) \Rightarrow \Pr(G = g | F = f) = \prod_i \frac{e^{-\sum_j \mathcal{H}_{ij} f_j} (\sum_j \mathcal{H}_{ij} f_j)^{g_i}}{g_i!} \tag{1}$$

where g is the vector of projection data, f is the underlying radionuclide source distribution vector (to be estimated) and $\mathcal{H}$ is the forward projection matrix. G and F are the associated random fields. The measurement physics is incorporated in $\mathcal{H}$. (While this linear model is appropriate for SPECT, it has to be extended to an affine model for PET. This extension does not affect our formulation.) Throughout this paper, we use $i \in \{1, \ldots, M\}$ as the index for detector measurement g_i and $j \in \{1, \ldots, N\}$ as the index for the radionuclide source distribution f_j.

The corresponding negative log-likelihood is

$$E_{\text{likl}}(f) = \sum_{ij} \mathcal{H}_{ij} f_j - \sum_i g_i \log \sum_j \mathcal{H}_{ij} f_j + (\text{terms independent of } f) \tag{2}$$

Despite the non-quadratic nature of the above log-likelihood, the well known EM algorithm that maximizes the log-likelihood has a simple, elegant form (Shepp and Vardi, 1982). The EM algorithm is conventionally derived using an incomplete/complete data parameterization (Dempster et al., 1977). Lack of space precludes us from presenting this derivation of the EM algorithm for emission tomography. Instead, we present the intuition behind this conventional derivation and then present a shorter, idiosyncratic derivation.

The complete data random field C for emission tomography is the number of photons C_{ij} that are emitted from pixel j and detected at detector bin i. The incomplete projection data random field G is derived from a straightforward many-to-one mapping $(G_i = \sum_j C_{ij})$ from the complete data C. The original incomplete data likelihood in (1) is now written in terms of the complete data likelihood. Since the actual complete data is unknown, the incomplete data log-likelihood becomes a function of expected value of the complete data. These computed expectations comprise the E-step. The M-step is a maximization step carried out after the E-step. For emission tomography, the EM algorithm turns out to be an extremely simple (but usually inefficient) discrete-time dynamical system that is guaranteed to maximize the incomplete data log-likelihood.

We now present an idiosyncratic derivation of the EM algorithm for ET. Borrowing from a similar EM algorithm for mixture decompositions (Hathaway, 1986; Gold et al., 1996; Yuille et al., 1994), we write down the following complete data energy function:

$$E_{\text{cmplikl}}(c, f) = \sum_{ij} \mathcal{H}_{ij} f_j - \sum_{ij} c_{ij} \log \mathcal{H}_{ij} f_j + \sum_{ij} c_{ij} \log c_{ij} + \sum_i \zeta_i (\sum_j c_{ij} - g_i) \tag{3}$$

In (3), c_{ij} is the complete data corresponding to the expected number of photons emanating from pixel j that are detected at detector i. The vector ζ is a Lagrange parameter vector that expresses the complete-incomplete data constraint $\sum_j c_{ij} = g_i$. It is possible to solve for the three unknowns c, f and the Lagrange parameter ζ by differentiating with respect to each variable and setting the result to zero. Rather surprisingly, this yields the emission tomography EM algorithm originally derived in (Shepp and Vardi, 1982).

$$\begin{array}{rclcrcl} \frac{\partial E_{\text{cmplikl}}}{\partial c_{ij}} & = & 0 & \Rightarrow & c_{ij} & = & \mathcal{H}_{ij} f_j \exp\left(-\zeta_i - 1\right), \\ \frac{\partial E_{\text{cmplikl}}}{\partial \zeta_i} & = & 0 & \Rightarrow & \sum_i c_{ij} & = & g_i, \\ \frac{\partial E_{\text{cmplikl}}}{\partial f_j} & = & 0 & \Rightarrow & f_j & = & \frac{\sum_i c_{ij}}{\sum_i H_{ij}} \end{array} \tag{4}$$

The final algorithm after eliminating c has a simple iterative form which at time step n is

$$f_j^{(n+1)} = \frac{\sum_i g_i \frac{\mathcal{H}_{ij} f_j^{(n)}}{\sum_k \mathcal{H}_{ik} f_k^{(n)}}}{\sum_i \mathcal{H}_{ij}} \tag{5}$$

While the above derivation is somewhat idiosyncratic, it is especially useful for our purposes. We were able to convert the original incomplete data log-likelihood energy function in (2) into a constrained energy function from which the well known EM algorithm emerges as a straightforward alternating descent algorithm. Exactly the same approach is used later in the constrained joint mixture energy function.

3.2 The joint mixture model

Assume that the anatomical image and the (evolving) functional image are perfectly registered and that the anatomical image has has been resampled so that the pixel size is the same as that of the functional image. While it is reasonable to expect that the pixel classifications of the anatomical and functional images will have some degree of overlap, there is also reason to expect that some functional image region class may overlap several anatomical region classes. This notion will be further explicated after notational definitions have been established.

The joint mixture prior is written as follows:

$$\Pr(F = f, H = h|\pi, \theta^{(1)}, \theta^{(2)}) = \prod_{j=1}^{N} \sum_{a=1}^{K} \sum_{b=1}^{L} \pi_{ab} \Pr(F_j = f_j | H_j = h_j, \theta_a^{(1)}) \Pr(H_j = h_j | \theta_b^{(2)}) \quad (6)$$

In (6), h is the anatomical image intensity with H being the associated random field. The matrix π is the *occupation probability* of the two intensity images in a $K \times L$ label cross product space. The constraints on π are $\pi_{ab} > 0$, $\forall a \in \{1, \ldots, K\}$, $\forall b \in \{1, \ldots, L\}$ and $\sum_{ab} \pi_{ab} = 1$. Throughout this paper, a is the subscript index used for the ET classes and b the subscript index used for the anatomical classes. The common index $j \in \{1, \ldots, N\}$ on f and h denotes a pixel location in either image. Typically, N is much larger than K or L (e.g. $N = 10^4$ and $K, L \approx 5$ for a 2D SPECT reconstruction problem with anatomical MR side information). The parameters $\theta^{(1)}$ and $\theta^{(2)}$ are associated with the basic modeling distributions of the joint mixture model and will be described below.

After specifying the joint mixture model (6), we turn to the prior on the occupation probability matrix π. If a perfect overlap existed between the labels in the functional and anatomical image, the matrix π would be a square matrix and without loss of generality, the diagonal elements would represent the numbers of common pixels sharing the same label. On the other hand, if the anatomical and functional labels were independent of one another, π could be factored into a product of a column and a row vector with the column (row) vector consisting of the occupancy probabilities of the functional (anatomical) image. The prior on π can be viewed as a complexity term, which, when combined with the joint mixture model has the novel interpretation of "maximizing the mutual information of π" (Wells III et al., 1996). We write the prior on π with an eye toward maximizing the mutual information:

$$p(\pi) = \frac{\Pi_{ab} \left(\frac{1}{\sum_c \pi_{cb} \sum_d \pi_{ad}} \right)^{\gamma \pi_{ab}}}{Z_{\text{part}}} \quad (7)$$

where Z_{part} is the partition function corresponding to this prior density. This partition function is a configuration space integral over all π satisfying the constraints $\pi_{ab} > 0$, $\forall a \in \{1, \ldots, K\}$, $\forall b \in \{1, \ldots, L\}$ and $\sum_{ab} \pi_{ab} = 1$. The parameter γ is a free parameter (which we later set to the number of ET pixels N). The corresponding negative log prior energy function is

$$E_{\mathrm{pi}}(\pi) = \gamma \sum_{ab} \left(\pi_{ab} \log \sum_c \pi_{cb} + \pi_{ab} \log \sum_d \pi_{ad} \right) + \text{ terms independent of } \pi. \tag{8}$$

This energy function can be viewed as a summation over the negative of the marginal entropies of π.

The only remaining piece of information left unspecified is the anatomical label information. We assume that the anatomical image has been previously segmented and registered with the ET image; for registration purposes, a filtered backprojection (FBP) or EM reconstruction will suffice. We also assume that every anatomical pixel (in the ET coordinate frame) has a region label. We represent this anatomical image segmentation information as a matrix W with each entry denoted by W_{bj}. (Recall that we use b to denote an anatomical class index throughout this paper.) Essentially, the matrix $\{W_{bj}\}$ contains the degree of membership of pixel j in the anatomical image to class $b \in \{1, \ldots, L\}$. The constraints on $\{W_{bj}\}$ are $W_{bj} > 0, \forall b \in \{1, \ldots, L\}$ and $\sum_b W_{bj} = 1$. The segmentation information $\{W_{bj}\}$ derived from a segmentation algorithm is assumed known. It plays a vital role in the computation of the pixel classifications in the functional image.

Using Bayes' rule, one could write down the *incomplete* data posterior energy function consisting of the tomographic likelihood in (1), the joint mixture model in (6) and the prior on π in (7). However, in order to derive the EM^2 algorithm, we switch to the *complete* data spaces for both the tomographic likelihood and the mixture prior and then write down the complete data posterior energy function.

Having specified the prior on π, we return to the joint mixture model. The negative log-prior energy function corresponding to the joint mixture model in (6) is

$$E_{\mathrm{mix}}(f, h, \pi, \theta^{(1)}, \theta^{(2)}) = -\sum_j \log \sum_{ab} \pi_{ab} \Pr(F_j = f_j | H_j = h_j, \theta_a^{(1)}) \Pr(H_j = h_j | \theta_b^{(2)}) \tag{9}$$

As in the case of the tomographic likelihood, it is possible to go from the incomplete data constrained joint mixture energy function to the complete data constrained joint mixture energy function. This conversion was (to our knowledge) first pointed out in (Hathaway, 1986) for the canonical single mixture case. The extension to the joint mixture model is straightforward. The complete data energy function uses a new complete data variable $\{z_{ab,j}\}$ which is a joint classification variable. The constraints on z are $z_{ab,j} \geq 0$, $\sum_{a=1}^K z_{ab,j} = W_{bj}, \forall b \in \{1, \ldots, L\}, \forall j \in \{1, \ldots, N\}$ and $\frac{\sum_{j=1}^N z_{ab,j}}{N} = \pi_{ab}$. These constraints taken in conjunction with the constraint $\sum_b W_{bj} = 1$, imply that $\sum_{ab} z_{ab,j} = 1$. Note the appearance of the anatomical segmentation labels $\{W_{bj}\}$ in the above constraints. Since $z_{ab,j}$ is the degree of membership of ET pixel j in ET class a *and* the degree of membership of the anatomical pixel j in anatomical class b, the constraint $\sum_a z_{ab,j} = W_{bj}$ follows. An example of the complete data z is shown in Figure 1. Note the differences between the anatomical "MR" and functional image regions. Specifically, in this example, there are two "very hot" regions in the functional image that do not exist in the anatomical image. One of the "very hot" regions extends across two anatomical classes.

The complete data constrained joint mixture energy function [after adding the log-prior on π terms from (8)] is

$$\begin{aligned} E_{\mathrm{cmpmix}}(f,z,\pi,\theta^{(1)},\theta^{(2)}) &= -\sum_{abj} z_{ab,j}\log[\pi_{ab}\mathrm{Pr}(F_j=f_j|H_j=h_j,\theta_a^{(1)})\mathrm{Pr}(H_j=h_j|\theta_b^{(2)})] \\ &\quad +\sum_{abj} z_{ab,j}\log z_{ab,j}+\sum_{bj}\lambda_{bj}(\sum_a z_{ab,j}-W_{bj})+\mu(\sum_{ab}\pi_{ab}-1) \\ &\quad +\gamma\sum_{ab}\pi_{ab}\log\sum_c \pi_{cb}+\gamma\sum_{ab}\pi_{ab}\log\sum_d \pi_{ad} \qquad (10) \end{aligned}$$

In (10), λ and μ are Lagrange parameters (vector and scalar, respectively) expressing the constraints $\sum_a z_{ab,j} = W_{bj}$ and $\sum_{ab}\pi_{ab} = 1$.

From Bayes' rule, we know that the log-posterior is proportional to the sum of the log-likelihood and the log-prior. Combining the complete data mixture energy function in (10) with the complete data tomography likelihood energy function in (3), we get the generic form of the final energy function used in this paper. This posterior energy function is

$$E_{\mathrm{cmp}}(c,f,z,\pi,\theta^{(1)},\theta^{(2)}) = E_{\mathrm{cmplikl}}(c,f)+E_{\mathrm{cmpmix}}(f,z,\pi,\theta^{(1)},\theta^{(2)}) \qquad (11)$$

3.3 Mixtures of conjugate priors

We have not yet specified the modeling distributions $\mathrm{Pr}(F_j = f_j|H_j = h_j,\theta_a^{(1)})$ and $\mathrm{Pr}(H_j = h_j|\theta_b^{(2)})$. The first simplifying assumption we make is the independence of the functional image pixel conditional density function with respect to the corresponding anatomical image pixel; i.e. $\mathrm{Pr}(F_j = f_j|H_j = h_j,\theta_a^{(1)}) = \mathrm{Pr}(F_j = f_j|\theta_a^{(1)})$. This assumption is clearly limiting and forces π to mediate all of the real "coupling" between the anatomical and functional images. The main reason for this assumption is ignorance. While we could model the the joint density function $p(f_j,h_j|\theta_{ab})$ using (say) a multivariate Gaussian for each pair (a,b), this greatly increases the complexity of the estimation procedure for the mixture parameters θ. A second reason for refraining from modeling the above joint density for each pair (a,b) in the cross-product space is more technical. It turns out that using multivariate Gaussian priors (on f) complicates the estimation procedure for f.

We use a gamma prior for the prior density $p(f_j|\theta_a^{(1)})$. The use of the gamma prior is motivated by the fact that it is conjugate to the Poisson distribution used in the likelihood model. Conjugate priors usually result in highly analytically tractable posterior estimates (Bernardo and Smith, 1994). In addition, the gamma prior naturally constrains f_j to be positive unlike Gaussian densities. The gamma density with $\theta_a^{(1)} \stackrel{\mathrm{def}}{=} (\alpha_a,\beta_a)$ is defined as

$$p(f_j|\alpha_a,\beta_a) = \frac{\alpha_a^{\alpha_a}}{\beta_a^{\alpha_a}\Gamma(\alpha_a)} f_j^{\alpha_a-1} e^{-\frac{\alpha_a f_j}{\beta_a}} \qquad (12)$$

The mean of a gamma distribution parameterized in this manner is β_a and the variance is $\frac{\beta_a^2}{\alpha_a}$.

After resorting to the complete data energy function reformulation trick pioneered in (Hathaway, 1986), the complete data constrained joint mixture energy function which is a specialization of (10) to the gamma prior is written as

$$\begin{aligned} E_{\text{cmpmixgam}}(f,z,\pi,\alpha,\beta) &= -\sum_{abj} z_{ab,j}\log[\pi_{ab}\frac{\alpha_a^{\alpha_a}}{\beta_a^{\alpha_a}\Gamma(\alpha_a)}f_j^{\alpha_a-1}e^{-\frac{\alpha_a f_j}{\beta_a}}] \\ &\quad +\sum_{abj} z_{ab,j}\log z_{ab,j}+\sum_{bj}\lambda_{bj}(\sum_a z_{ab,j}-W_{bj})+\mu(\sum_{ab}\pi_{ab}-1) \\ &\quad +\gamma\sum_{ab}\pi_{ab}\log\sum_c \pi_{cb}+\gamma\sum_{ab}\pi_{ab}\log\sum_d \pi_{ad} \end{aligned} \tag{13}$$

Using the identity $\sum_j z_{ab,j} = N\pi_{ab}$, and setting $\gamma = N$, (13) can be rewritten as

$$\begin{aligned} E_{\text{cmpmixgam}}(f,z,\pi,\alpha,\beta) &= -\sum_{abj} z_{ab,j}\log[\frac{\alpha_a^{\alpha_a}}{\beta_a^{\alpha_a}\Gamma(\alpha_a)}f_j^{\alpha_a-1}e^{-\frac{\alpha_a f_j}{\beta_a}}] \\ &\quad +\sum_{abj} z_{ab,j}\log z_{ab,j}+\sum_{bj}\lambda_{bj}(\sum_a z_{ab,j}-W_{bj})+\mu(\sum_{ab}\pi_{ab}-1) \\ &\quad -N\sum_{ab}\pi_{ab}\log\left(\frac{\pi_{ab}}{\sum_c \pi_{cb}\sum_d \pi_{ad}}\right) \end{aligned} \tag{14}$$

The last term in (14) can be interpreted as the mutual information of the occupancy probability π. The prior on π was reverse-engineered in order to obtain the mutual information in the posterior. Equation (14) can be interpreted as a posterior energy function wherein the mutual information of π is maximized subject to the constraints on the complete data z and the choice of the modeling distribution (in this case, gamma).

We have now specified in great detail the likelihood, the joint mixture with the gamma distribution as the basic modeling distribution, the prior on π (motivated by mutual information) and the constraints due to the available anatomical segmentation information $\{W_{bj}\}$. In addition, we have written down constrained complete data joint mixture energy functions as well as the complete data likelihood. Next, we develop the EM^2 algorithm to minimize the complete data log-posterior energy function.

3.4 The EM^2 algorithm

We now write the full complete data energy function which comprises the complete data tomographic likelihood as well as the complete data joint mixture prior (which includes the prior on π). The final energy function used in this paper is a summation of the energy functions in (3) and in (13).

$$\begin{aligned} E_{\text{cmpgam}}(c,f,z,\alpha,\beta,\pi,\zeta,\lambda,\mu) &= \sum_{ij}\mathcal{H}_{ij}f_j-\sum_{ij}c_{ij}\log\,\mathcal{H}_{ij}f_j+\sum_{ij}c_{ij}\log\,c_{ij}+\sum_i\zeta_i(\sum_j c_{ij}-g_i) \\ &\quad +\sum_{abj} z_{ab,j}\left[\frac{\alpha_a f_j}{\beta_a}-(\alpha_a-1)\log f_j-\alpha_a\log\alpha_a+\alpha_a\log\beta_a+\log\Gamma(\alpha_a)\right] \end{aligned}$$

$$
+\sum_{abj} z_{ab,j} \log z_{ab,j} + \sum_{bj} \lambda_{bj} (\sum_{a} z_{ab,j} - W_{bj}) + \mu(\sum_{ab} \pi_{ab} - 1)
$$
$$
-\sum_{abj} z_{ab,j} \log \pi_{ab} + N \sum_{ab} \pi_{ab} \log \sum_{c} \pi_{cb} + N \sum_{ab} \pi_{ab} \log \sum_{d} \pi_{ad} \tag{15}
$$

In Appendix A, we derive the final set of update equations from the above energy function. The final set of update equations is derived by differentiating w.r.t. each variable in the transformed energy function and setting the result to zero. The result is an alternating descent algorithm that consists of two EM algorithms—one for the likelihood with a feedback contribution from the prior and one for the joint mixture prior. The mixture EM algorithm is nested inside the outer likelihood EM. The pseudocode is presented below in the box entitled "The EM^2 Algorithm: Mutual Information".

The EM^2 Algorithm: Mutual Information

Initialize $(z, \pi, \alpha, \beta, f)$

Begin A: EM outer loop. Do A until f converges

$c_{ij} \leftarrow g_i \frac{\mathcal{H}_{ij} f_j}{\sum_k \mathcal{H}_{ik} f_k}$

$f_j \leftarrow \frac{\sum_i c_{ij} + \sum_{ab} z_{ab,j}(\alpha_a - 1)}{\sum_i \mathcal{H}_{ij} + \sum_{ab} z_{ab,j} \frac{\alpha_a}{\beta_a}}$

Begin B: EM inner loop. Do B until (z, π, α, β) converge

$z_{ab,j} \leftarrow W_{bj} \frac{\frac{\pi_{ab}}{\sum_d \pi_{ad}} \frac{\alpha_a^{\alpha_a}}{\beta_a^{\alpha_a} \Gamma(\alpha_a)} f_j^{\alpha_a - 1} e^{-\frac{\alpha_a f_j}{\beta_a}}}{\sum_c \frac{\pi_{cd}}{\sum_d \pi_{cd}} \frac{\alpha_c^{\alpha_c}}{\beta_c^{\alpha_c} \Gamma(\alpha_c)} f_j^{\alpha_c - 1} e^{-\frac{\alpha_c f_j}{\beta_c}}}$

$\pi_{ab} \leftarrow \frac{\sum_j z_{ab,j}}{N}$

$\beta_a \leftarrow \frac{\sum_{bj} z_{ab,j} f_j}{\sum_{bj} z_{ab,j}}$

$\alpha_a \leftarrow \arg\min_{\alpha_a} \sum_{bj} z_{ab,j} \left[\frac{\alpha_a f_j}{\beta_a} - (\alpha_a - 1) \log f_j - \alpha_a \log \alpha_a + \alpha_a \log \beta_a + \log \Gamma(\alpha_a) \right]$

End B

End A

4 Experimental Results

In this section, we compare the performance of the EM^2 algorithm with six other closely related variants. The different algorithms model to varying degrees the inclusion of the various forms of the anatomical information present in the the EM^2 algorithm. The seven algorithms are briefly summarized in Appendix B; they are designated by EM-1, EM-2, MIX, MR, JENT, MI and MIz. The EM-1 and EM-2 are standard EM emission-tomography maximum-likelihood algorithms as in (5) with the only difference between them being the stopping criterion, i.e. the number of iterations.

(The EM maximum likelihood algorithms for ET have a over-fitting problem and hence never actually converge (Shepp and Vardi, 1982).) The algorithm MIX uses a mixture of gamma prior to stabilize the reconstruction and does not include any anatomical information. It is a standard Bayesian MAP algorithm with a mixture of gamma prior. The fourth algorithm, MR, is derived based on the assumption that the anatomical and functional classes have a perfect overlap. In MR, the complete data classification variable z_{aj} (note the absence of the second subscript b) is directly constrained by the anatomical label information W_{aj} and is inhibited from forming "hot spot" lesions in ET that do not appear in anatomy. The fifth and sixth algorithms, JENT and MI, respectively are EM^2 algorithms with the only difference between them being the prior on π. MI, the EM^2 algorithm derived in this paper, is the only algorithm that has the interpretation of mutual information maximization. JENT can be derived from the joint mixture formulation by dropping the prior on π. MR can be derived by forcing the classification spaces in anatomy and function to be identical ($K = L$) and so on. Finally, the last algorithm MIz, is used for validation purposes only. It assumes that the complete data variable $z_{ab,j}$ is known. Whereas z has to be estimated by the preceding algorithms MR, JENT and MI, MIz assumes it from the outset, leading to a simple EM algorithm with the various mixture parameters "frozen" at their true values.

The functional source "ground truth" image for all the reconstruction experiments is shown in Figure 2. It comprises four regions including a uniform background. The anatomical source image is chosen to be almost the same as the functional source image except that the two central bright regions seen in the functional "ground truth" image are absent. Note that the actual intensity values in the anatomical source image are irrelevant. For this "ground truth" image, we have $K = 4$ functional classes and $L = 3$ anatomical classes.

We generated 300K simulated projection data counts using the phantom in Figure 2. Noiseless projections were generated via line integrals (Radon transform) followed by the addition of Poisson noise. Anecdotal reconstructions using all seven algorithms are shown in Figure 3. The EM-1 reconstruction was obtained with a stopping criterion of 20 iterations. EM-2 used a much larger iteration cap of about 100 iterations. MIX—the mixture of gamma prior with no MR information—was executed using an EM-1 initial condition. The inner EM loop in MIX used a reasonable convergence criterion based on the total change in (z, π, α, β). We noticed that α (regardless of region label) would usually increase without approaching a finite limiting value. To circumvent this problem, we imposed a ceiling on α of 500. We do not yet understand why α does not approach a finite fixed point. JENT and MI used the same initial conditions as MIX for f. For π, we used an initial condition which somewhat favored a good overlap between the first three MR and ET classes. This initial condition was not necessary for the algorithm MR since the anatomical pixel-region segmentation labels exert a much stronger constraint in this case. Finally, when we assume perfect knowledge about the joint complete data classification matrix z in the algorithm MIz, the values of z are frozen and never updated. Since π depends on z, it is never updated as well.

The results shown in Figure 3 show the qualitative behaviors associated with each of the seven algorithms. Since EM-1 uses an early stopping criterion, the reconstructed image looks smoother than its late stopping rule EM-2 counterpart. The EM maximum likelihood reconstruction progressively becomes less biased with increasing number of iterations at the cost of higher variance. Addition of the mixture of gamma prior regularizes the reconstruction. This can be seen in the MIX reconstruction which has a piecewise flat appearance (in contrast to EM-1 and EM-2). Inclusion of the anatomical segmentation labels appears from the second row on in Figure 3. In the MR reconstruction, the high bias incurred from the misclassification of the two central bright regions (in ET) is apparent. The central regions have been classified in the same class as the outer non-background class. Consequently, the intensities in this region are completely wrong, as anticipated. This anecdotal reconstruction clearly illustrates the bias produced when the ET reconstruction is overly constrained by the anatomical segmentation labels. The JENT and MI reconstructions are quite similar and both avoid the problems created by the aforementioned over commitment. Finally, the MIz reconstruction, when compared with the original phantom, demonstrates the improvement that can be obtained when assuming perfect knowledge about the joint classification matrix z.

The JENT and MI classification images shown in Figures 4 and 5, respectively, depict the classification matrix z in image form. The classification matrices are not binary despite appearances to the contrary. The differences between JENT and MI are somewhat subtle. Note the big difference in the "outlier" class—the two ET central regions which have no homology in the anatomical image. In JENT, this class is under-represented and the opposite is true for MI.

The bias/variance images shown in Figures 6 demonstrate performance differences amongst the seven algorithms. The bias and variance images were computed using $P = 40$ noise trials. We define the bias image, $b_j, j \in \{1, \ldots N\}$ as

$$b_j \stackrel{\text{def}}{=} \frac{1}{P} \sum_{k=1}^{P} \left(\hat{f}^{(k)} - f_j^{\text{true}} \right)$$

where $\hat{f}$ is the estimated reconstruction and f^{true} is the "ground truth" functional source image. We define the variance image , $\sigma_j^2, j \in \{1, \ldots, N\}$ as

$$\sigma_j^2 \stackrel{\text{def}}{=} \frac{1}{(P-1)} \sum_{k=1}^{P} \left(f_j^{(k)} - \bar{f}_j \right)^2$$

where

$$\bar{f}_j \stackrel{\text{def}}{=} \frac{1}{P} \sum_{k=1}^{P} \hat{f}_j^{(k)}$$

Note that the bias image being bipolar is depicted so that gray corresponds to zero bias and lighter (darker) regions correspond to positive (negative) bias. Also, for reasons of dynamic range, we display the standard deviation image as opposed to the variance image. This image is unipolar. The EM-2 reconstruction has a lower bias at the expense of increased variance in comparison to the EM-1 reconstruction. The MR reconstruction shows a high negative bias in the "hot spot" central regions and has a uniformly very low variance. MI has a lower bias *and* variance than JENT for the central "hot spot"regions. The perfect information algorithm MIz has a uniformly low bias and almost zero variance.

To illustrate more quantitatively the performance characteristics of the seven algorithms, we compute scalar summary metrics of bias and variance according to the following prescription.

$$\mathrm{BIAS_G} \stackrel{\mathrm{def}}{=} \sqrt{\frac{1}{N}\sum_{j=1}^{N} b_j^2}, \ \mathrm{STD_G} \stackrel{\mathrm{def}}{=} \sqrt{\frac{1}{N}\sum_{j=1}^{N} \sigma_j^2}$$

and these are displayed in Figure 7. Note that MIz has the best overall performance in the sense of being closest to the origin. Inclusion of the prior on π in the MI algorithm leads to an overall performance improvement over JENT. As expected, the relative positions of EM-1 and EM-2 illustrate well known bias/variance tradeoffs with iteration number. In this case, MIX performs about as well as EM-1. Note the poor bias behavior manifest in MR. This is is to be expected due to the inordinate influence of the anatomical segmentation label matrix W on the complete data variable z. Overall, the inclusion of the various forms of prior information in MI leads to the best bias/variance performance.

5 Discussion and Conclusion

We have introduced a novel Bayesian joint mixture framework for incorporating anatomical information into emission tomographic reconstruction. Unlike many other previous approaches, this approach has few free parameters. In fact, the principal free parameter is K—the number of ET classes. Note that L, the number of anatomical classes is presumed known from an independent anatomical segmentation. This approach does not require that the anatomical and functional regions are exactly homologous; the joint mixture model is designed specifically to handle a mismatch of anatomical and functional regions. As illustrated in Section 4, the statistical trials show a progressive improvement in bias/variance performance as more information is included in the joint mixture model. The EM^2 algorithm using a prior on the occupation probabilities π which we term MI, has the best bias/variance performance. This indicates that careful modeling of the joint classification space of anatomy and function is an important factor in obtaining improved functional reconstructions.

While in the present framework, there is room for modeling the raw anatomical intensity information h, the EM^2 algorithm does not, in fact, do so. This is a consequence of using a gamma prior. Since there is no multivariate extension of the gamma prior, it was difficult to jointly model the anatomical intensity h along with the functional intensity f. In future work, we plan to overcome this limitation by modeling the pair (f_j, h_j) as a multivariate Gaussian. Note that the additional anatomical segmentation information W would act as a constraint in this extension. The joint mixture model using a multivariate Gaussian would be forced to respect the segmentation constraints imposed by W which is independently obtained. Unfortunately, this would increase the number of parameters to be estimated. This requires further investigation.

Currently, a perfect registration between the anatomical and functional images is assumed prior to functional image reconstruction. While it is certainly possible to register an anatomical scan and an EM reconstructed ET image of the same patient, registration errors are likely to be present. We have not specifically taken into account the mis-registration error between anatomy and ET in this paper. However, we think the joint mixture model is robust enough to accommodate small overlap errors. In future work, we intend to model these and other factors in order to exploit this framework to obtain practical reconstruction algorithms for the clinic.

Acknowledgments

We wish to thank Anthony Quinn, Petar Djuric, Michael Leventon and Thomas Hofmann for helpful discussions and Doug Dougherty for help with the figures. This work was partially supported by NIH grant RO1 NS32879.

A Algebraic transformations of the constrained complete data energy functions

An algebraic transformation—essentially a Legendre transformation—transforms an energy function by introducing auxiliary variables (Mjolsness and Garrett, 1990). Examples of algebraic transformations are:

$$\begin{aligned} -\frac{x^2}{2} &= \min_{\sigma}\left(-\sigma x + \frac{\sigma^2}{2}\right) \\ \log\, x &= \min_{\sigma}\left(x\sigma - \log\, \sigma - 1\right) \end{aligned}$$

Algebraic transformations are employed when alternating descent algorithms are being designed to minimize complex nonlinear objective functions. Each auxiliary variable results in a separate phase of optimization.

We now transform the complete data constrained mixture energy function in (10). The final objective function is:

$$\begin{aligned} E_{\mathrm{cmpmixalg}}(f,h,z,\pi,\sigma,\tau,\lambda,\mu,\nu,\kappa,\theta^{(1)},\theta^{(2)}) \;=\; & -\sum_{abj} z_{ab,j} \log[\Pr(F_j = f_j | H_j = h_j, \theta_a^{(1)}) \Pr(H_j = h_j | \theta_b^{(2)})] \\ & + \sum_{abj} z_{ab,j} \log z_{ab,j} + \sum_{bj} \lambda_{bj} (\sum_a z_{ab,j} - W_{bj}) \\ & - \sum_{abj} z_{ab,j} \log \pi_{ab} + \mu(\sum_{ab} \pi_{ab} - 1) \\ & + \sum_{abj} z_{ab,j} \log \sigma_a - \nu(\sum_a \sigma_a - 1) \\ & + \sum_{abj} z_{ab,j} \log \tau_b - \kappa(\sum_b \tau_b - 1) \end{aligned} \tag{16}$$

The above algebraically transformed energy function involves the mixture prior only. Since the variables in the prior alone are involved, this is justified. The alternating update equation can be derived by differentiating w.r.t. $z, \pi, \sigma, \tau, \lambda, \mu, \nu, \kappa, \theta^{(1)}$ and setting the results to zero:

$$\begin{array}{rcccll}
\frac{\partial E}{\partial \pi_{ab}} & = & 0 & \Rightarrow & \pi_{ab} & = \frac{\sum_j z_{ab,j}}{N} \\
\frac{\partial E}{\partial \mu} & = & 0 & \Rightarrow & \mu & = N \\
\frac{\partial E}{\partial \sigma_a} & = & 0 & \Rightarrow & \sigma_a & = \frac{\sum_{bj} z_{ab,j}}{N} \\
\frac{\partial E}{\partial \nu} & = & 0 & \Rightarrow & \nu & = N \\
\frac{\partial E}{\partial \tau_b} & = & 0 & \Rightarrow & \tau_b & = \frac{\sum_{aj} z_{ab,j}}{N} \\
\frac{\partial E}{\partial \kappa} & = & 0 & \Rightarrow & \kappa & = N \\
\frac{\partial E}{\partial z_{ab,j}} & = & 0 & \Rightarrow & z_{ab,j} & = W_{bj} \frac{\frac{\pi_{ab}}{\sigma_a} \Pr(F_j = f_j | H_j = h_j, \theta_a^{(1)})}{\sum_c \frac{\pi_{cb}}{\sigma_c} \Pr(F_j = f_j | H_j = h_j, \theta_c^{(1)})} \\
\frac{\partial E}{\partial \lambda_{bj}} & = & 0 & \Rightarrow & \lambda_{bj} & = \log[\sum_a \frac{\pi_{ab}}{\sigma_a \tau_b} \Pr(F_j = f_j | H_j = h_j, \theta_a^{(1)}) \Pr(H_j = h_j | \theta_b^{(2)})] - 1 \\
\frac{\partial E}{\partial \theta^{(1)}} & = & 0 & \Rightarrow & \theta^{(1)} & = \arg\min_{\theta^{(1)}} \left(-\sum_{bj} z_{ab,j} \log \left[\Pr(F_j = f_j | H_j = h_j, \theta_a^{(1)}) \right] \right)
\end{array} \tag{17}$$

Since the solutions w.r.t. all the variables are inter-dependent, we have merely written down the final closed-form solutions on the right hand side of all the above implications. For example, the actual partial derivative w.r.t. $z_{ab,j}$ is:

$$z_{ab,j} = \pi_{ab} \Pr(F_j = f_j | H_j = h_j, \theta_a^{(1)}) \Pr(H_j = h_j | \theta_b^{(2)}) \exp(-\lambda_{bj} - 1) \tag{18}$$

The solution for z depends on the solution for λ. Solving for the Lagrange parameter λ using the constraint $\sum_a z_{ab,j} = W_{bj}$ and using the solution $\sigma_a = \sum_b \pi_{ab}$, we get

$$z_{ab,j} = W_{bj} \frac{\frac{\pi_{ab}}{\sum_d \pi_{ad}} \Pr(F_j = f_j | H_j = h_j, \theta_a^{(1)})}{\sum_c \frac{\pi_{cb}}{\sum_d \pi_{cd}} \Pr(F_j = f_j | H_j = h_j, \theta_c^{(1)})} \tag{19}$$

The EM^2 algorithm can be derived by adding in the complete data likelihood term from (3) and differentiating w.r.t. f and c. Since, we've already derived the original EM maximum likelihood algorithm for ET in this manner, this derivation will not be repeated here. In the next section, we summarize the seven different algorithms used in this paper.

B Description of the different algorithms

The seven different algorithms used (for comparison purposes) are described here. The names given to the seven algorithms are EM-1, EM-2, MIX, MR, JENT, MI and MIz.

The first and second algorithms are standard EM algorithms operating on the original likelihood energy function in (2). Since the EM algorithm (for tomographic reconstruction) typically has convergence problems and quite different intensity images can be obtained for different stopping conditions, we use two different EM algorithms with the only difference between them being the stopping rule. EM-1 uses an early stopping rule (and hence incurs higher bias) whereas EM-2 uses a late stopping rule (and incurs higher variance).

EM-1 and EM-2: Maximum Likelihood

Initialize f

Begin A: Do A for a fixed number of iterations

$$c_{ij} \leftarrow g_i \frac{\mathcal{H}_{ij} f_j}{\sum_k \mathcal{H}_{ik} f_k}$$

$$f_j \leftarrow \frac{\sum_i c_{ij}}{\sum_i \mathcal{H}_{ij}}$$

End A

The third algorithm is the mixture of gamma algorithm, termed MIX. A mixture of gamma prior is added to the emission tomography likelihood energy function. Note that the anatomical segmentation information does not appear in this algorithm. We merely use this algorithm for the sake of comparison. An EM algorithm is used for the mixtures prior. This results in an EM algorithm for the likelihood and an EM algorithm for the prior. Hence, this is an EM^2 algorithm which has been specialized for a mixtures of gamma prior.

MIX: Single Mixture of Gamma Prior

Initialize $(z, \pi, \alpha, \beta, f)$

Begin A: EM outer loop. Do A until f converges

$$c_{ij} \leftarrow g_i \frac{\mathcal{H}_{ij} f_j}{\sum_k \mathcal{H}_{ik} f_k}$$

$$f_j \leftarrow \frac{\sum_i c_{ij} + \sum_a z_{aj}(\alpha_a - 1)}{\sum_i \mathcal{H}_{ij} + \sum_a z_{aj} \frac{\alpha_a}{\beta_a}}$$

Begin B: EM inner loop. Do B until (z, π, α, β) converge

$$z_{aj} \leftarrow \frac{\pi_a \frac{\alpha_a^{\alpha_a}}{\beta_a^{\alpha_a} \Gamma(\alpha_a)} f_j^{\alpha_a - 1} e^{-\frac{\alpha_a f_j}{\beta_a}}}{\sum_c \pi_c \frac{\alpha_c^{\alpha_c}}{\beta_c^{\alpha_c} \Gamma(\alpha_c)} f_j^{\alpha_c - 1} e^{-\frac{\alpha_c f_j}{\beta_c}}}$$

$$\pi_a \leftarrow \frac{\sum_j z_{aj}}{N}$$

$$\beta_a \leftarrow \frac{\sum_j z_{aj} f_j}{\sum_j z_{aj}}$$

$$\alpha_a \leftarrow \arg\min_{\alpha_a} \sum_j z_{aj} \left[\frac{\alpha_a f_j}{\beta_a} - (\alpha_a - 1) \log f_j - \alpha_a \log \alpha_a + \alpha_a \log \beta_a + \log \Gamma(\alpha_a) \right]$$

End B

End A

The fourth algorithm directly constrains the ET pixel classification variable z_{aj} using the *a priori* anatomical segmentation W. However, the set of ET classification labels are not allowed to be different from that of anatomy. This has the effect, for example, of misclassification of functional lesions which do not appear in anatomy. In the algorithm description, note the direct modulation of ET pixel classification labels z_{aj} by the anatomical label W_{aj}. The algorithm is termed MR.

MR: Anatomical Segmentation Constrained Mixture Prior

Initialize $(z, \pi, \alpha, \beta, f)$

Begin A: EM outer loop. Do A until f converges

$c_{ij} \leftarrow g_i \frac{\mathcal{H}_{ij} f_j}{\sum_k \mathcal{H}_{ik} f_k}$

$f_j \leftarrow \frac{\sum_i c_{ij} + \sum_a z_{aj}(\alpha_a - 1)}{\sum_i \mathcal{H}_{ij} + \sum_a z_{aj} \frac{\alpha_a}{\beta_a}}$

Begin B: EM inner loop. Do B until (z, π, α, β) converge

$z_{aj} \leftarrow \frac{W_{aj} \pi_a \frac{\alpha_a^{\alpha_a}}{\beta_a^{\alpha_a} \Gamma(\alpha_a)} f_j^{\alpha_a - 1} e^{-\frac{\alpha_a f_j}{\beta_a}}}{\sum_c W_{cj} \pi_c \frac{\alpha_c^{\alpha_c}}{\beta_c^{\alpha_c} \Gamma(\alpha_c)} f_j^{\alpha_c - 1} e^{-\frac{\alpha_c f_j}{\beta_c}}}$

$\pi_a \leftarrow \frac{\sum_j z_{aj}}{N}$

$\beta_a \leftarrow \frac{\sum_j z_{aj} f_j}{\sum_j z_{aj}}$

$\alpha_a \leftarrow \arg\min_{\alpha_a} \sum_j z_{aj} \left[\frac{\alpha_a f_j}{\beta_a} - (\alpha_a - 1) \log f_j - \alpha_a \log \alpha_a + \alpha_a \log \beta_a + \log \Gamma(\alpha_a) \right]$

End B

End A

The first clue that the algorithm MR introduces a significant, unwarranted bias in the reconstruction is the absence of a second, differentiating segmentation index b. Since the anatomical and the ET classes are not allowed to separately vary, it follows that the ET pixel classes are very tightly coupled with the anatomical segmentation labels. Unfortunately, such a coupling prohibits, by its very nature, the formation of a functional lesion that is a separate ET class not found in anatomy. An even more egregious error occurs at the algorithmic level. The update equation for z depends multiplicatively on the anatomical segmentation labels $\{W_{aj}\}$. Consequently, if any W_{aj} is zero, the *corresponding ET pixel classification variable* z_{aj} *can never be non-zero.*

The fifth algorithm uses the joint mixture model. Since no prior is added on the occupation probabilities π, the algorithm can be interpreted as minimizing the joint entropy between the anatomical and functional images. Hence, the algorithm is termed JENT.

JENT: Joint Entropy

Initialize $(z, \pi, \alpha, \beta, f)$

Begin A: EM outer loop. Do A until f converges

$$c_{ij} \leftarrow g_i \frac{\mathcal{H}_{ij} f_j}{\sum_k \mathcal{H}_{ik} f_k}$$

$$f_j \leftarrow \frac{\sum_i c_{ij} + \sum_{ab} z_{ab,j}(\alpha_a - 1)}{\sum_i \mathcal{H}_{ij} + \sum_{ab} z_{ab,j} \frac{\alpha_a}{\beta_a}}$$

Begin B: EM inner loop. Do B until (z, π, α, β) converge

$$z_{ab,j} \leftarrow W_{bj} \frac{\pi_{ab} \frac{\alpha_a^{\alpha_a}}{\beta_a^{\alpha_a} \Gamma(\alpha_a)} f_j^{\alpha_a - 1} e^{-\frac{\alpha_a f_j}{\beta_a}}}{\sum_c \pi_{cd} \frac{\alpha_c^{\alpha_c}}{\beta_c^{\alpha_c} \Gamma(\alpha_c)} f_j^{\alpha_c - 1} e^{-\frac{\alpha_c f_j}{\beta_c}}}$$

$$\pi_{ab} \leftarrow \frac{\sum_j z_{ab,j}}{N}$$

$$\beta_a \leftarrow \frac{\sum_{bj} z_{ab,j} f_j}{\sum_{bj} z_{ab,j}}$$

$$\alpha_a \leftarrow \arg\min_{\alpha_a} \sum_{bj} z_{ab,j} \left[\frac{\alpha_a f_j}{\beta_a} - (\alpha_a - 1) \log f_j - \alpha_a \log \alpha_a + \alpha_a \log \beta_a + \log \Gamma(\alpha_a) \right]$$

End B

End A

This algorithm does not suffer from the fatal flaw of the previous algorithm MR. In sharp contrast to MR, this algorithm uses the full joint clustering space of ET and anatomy. The joint classification variable z is not directly constrained and overridden by W. JENT is the first algorithm that allows the formation of a functional lesion that is a separate class and has the expressive power of cutting across several anatomical classes. (This should be a requirement of any approach that uses anatomical side information.)

The sixth algorithm uses the joint mixture model with a prior on the occupation probabilities π that is equivalent to maximizing the mutual information between the anatomical and functional images. The algorithm is termed MI. In the algorithm description, note that the update of the pixel classification variable $z_{ab,j}$ depends on $\frac{\pi_{ab}}{\sum_c \pi_{cb}}$. If we interpret π_{ab} as the joint class probability and $\sum_c \pi_{cb}$ as the marginal probability of the anatomical class labels, then $\frac{\pi_{ab}}{\sum_c \pi_{cb}}$ can be interpreted as the conditional probability of the anatomical class label b conditioned on the functional label being a. This is in contrast to the previous algorithm JENT, where $z_{ab,j}$ depends only on the joint class probability π_{ab}. MI shares with its predecessor JENT the capability of expressing the fact that a functional lesion can "outrun" anatomical regions.

MI: Mutual Information

Initialize $(z, \pi, \alpha, \beta, f)$

Begin A: EM outer loop. Do A until f converges

$c_{ij} \leftarrow g_i \frac{\mathcal{H}_{ij} f_j}{\sum_k \mathcal{H}_{ik} f_k}$

$f_j \leftarrow \frac{\sum_i c_{ij} + \sum_{ab} z_{ab,j}(\alpha_a - 1)}{\sum_i \mathcal{H}_{ij} + \sum_{ab} z_{ab,j} \frac{\alpha_a}{\beta_a}}$

Begin B: EM inner loop. Do B until (z, π, α, β) converge

$z_{ab,j} \leftarrow W_{bj} \frac{\frac{\pi_{ab}}{\sum_d \pi_{ad}} \frac{\alpha_a^{\alpha_a}}{\beta_a^{\alpha_a} \Gamma(\alpha_a)} f_j^{\alpha_a - 1} e^{-\frac{\alpha_a f_j}{\beta_a}}}{\sum_c \frac{\pi_{cd}}{\sum_d \pi_{cd}} \frac{\alpha_c^{\alpha_c}}{\beta_c^{\alpha_c} \Gamma(\alpha_c)} f_j^{\alpha_c - 1} e^{-\frac{\alpha_c f_j}{\beta_c}}}$

$\pi_{ab} \leftarrow \frac{\sum_j z_{ab,j}}{N}$

$\beta_a \leftarrow \frac{\sum_{bj} z_{ab,j} f_j}{\sum_{bj} z_{ab,j}}$

$\alpha_a \leftarrow \arg\min_{\alpha_a} \sum_{bj} z_{ab,j} \left[\frac{\alpha_a f_j}{\beta_a} - (\alpha_a - 1) \log f_j - \alpha_a \log \alpha_a + \alpha_a \log \beta_a + \log \Gamma(\alpha_a) \right]$

End B

End A

Finally, for validation purposes, we run MI with perfect information regarding the joint anatomical and functional pixel classification labels $z_{ab,j}$. Consequently, we can estimate the gamma prior parameters with ground-truth pixel classification information. This results in an extremely simple EM algorithm, termed MIz.

MIz: Perfect Information

Initialize (α, β, f)

Begin A: EM loop. Do A until (f, α, β) converge

$c_{ij} \leftarrow g_i \frac{\mathcal{H}_{ij} f_j}{\sum_k \mathcal{H}_{ik} f_k}$

$f_j \leftarrow \frac{\sum_i c_{ij} + \sum_{ab} z_{ab,j}(\alpha_a - 1)}{\sum_i \mathcal{H}_{ij} + \sum_{ab} z_{ab,j} \frac{\alpha_a}{\beta_a}}$

$\beta_a \leftarrow \frac{\sum_{bj} z_{ab,j} f_j}{\sum_{bj} z_{ab,j}}$

$\alpha_a \leftarrow \arg\min_{\alpha_a} \sum_{bj} z_{ab,j} \left[\frac{\alpha_a f_j}{\beta_a} - (\alpha_a - 1) \log f_j - \alpha_a \log \alpha_a + \alpha_a \log \beta_a + \log \Gamma(\alpha_a) \right]$

End A

References

Ardekani, B. A., Braun, M., Hutton, B. F., Kanno, I., and Ida, H. (1996). Minimum cross-entropy reconstruction of PET images using prior anatomical information. *Phys. Med. Biol.*, 41(11):2497–2517.

Bernardo, J. and Smith, A. (1994). *Bayesian Theory*. John Wiley and Sons, New York, NY.

Bowsher, J. E., Johnson, V. E., Turkington, T. G., Jaszczak, R. J., and Floyd Jr., C. E. (1996). Bayesian reconstruction and use of anatomical a *priori* information for emission tomography. *IEEE Trans. Med. Imag.*, 15(5):673–686.

de Sa, V. and Ballard, D. (1998). Category learning through multimodality sensing. *Neural Computation*, 10(5):1097–1118.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. Ser. B*, 39:1–38.

Gindi, G., Lee, M., Rangarajan, A., and Zubal, I. G. (1993). Bayesian reconstruction of functional images using anatomical information as priors. *IEEE Trans. Med. Imag.*, 12(4):670–680.

Gold, S., Rangarajan, A., and Mjolsness, E. (1996). Learning with preknowledge: clustering with point and graph matching distance measures. *Neural Computation*, 8(4):787–804.

Hathaway, R. (1986). Another interpretation of the EM algorithm for mixture distributions. *Statistics and Probability Letters*, 4:53–56.

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

Hsu, C. H. and Leahy, R. (1997). PET image reconstruction incorporating anatomical information using segmented regression. In Hanson, K., editor, *Proc. SPIE: Medical Imaging - Image Processing, Vol. 3034*, pages 381–392, Newport Beach, CA. SPIE.

Leahy, R. and Yan, X. (1991). Incorporation of anatomical MR data for improved functional imaging with PET. In Ortendahl, D. and Llacer, J., editors, *Information Processing in Medical Imaging*, pages 105–120. Wiley-Liss.

Lee, S., Hsiao, I., and Gindi, G. (1997). The thin plate as a regularizer in Bayesian SPECT reconstruction. *IEEE Trans Nucl. Sci.*, 44(3):1381–1387.

Lipinski, B., Herzog, H., Rota Kops, E., Oberschelp, W., and Muller-Gartner, H. W. (1997). Expectation-Maximization reconstruction of positron emission tomography images using anatomical magnetic resonance information. *IEEE Trans. Med. Imag.*, 16(2):129–136.

McLachlan, G. J. and Basford, K. E. (1988). *Mixture models: inference and applications to clustering*. Marcel Dekker, New York.

Meltzer, C. C., Leal, J. P., Mayberg, H. S., Wagner Jr., H. N., and Frost, J. J. (1990). Correction of PET data for partial volume effects in human cerebral cortex by MR imaging. *J. of Computer Assisted Tomography*, 14(4):561–570.

Mjolsness, E. and Garrett, C. (1990). Algebraic transformations of objective functions. *Neural Networks*, 3:651–669.

Muller-Gartner, H. W., Links, J. M., Prince, J. L., Bryan, R. N., McVeigh, E., Leal, J. P., Davatzikos, C., and Frost, J. J. (1992). Measurement of radiotracer concentration in brain gray matter using positron emission tomography: MRI-based correction for partial volume effects. *Journal of Cerebral Blood Flow and Metabolism*, 12(4):571–583.

Ouyang, X., Wong, W. H., Johnson, V. E., Hu, X., and Chen, C. T. (1994). Incorporation of correlated structural images in PET image reconstruction. *IEEE Trans. Med. Imag.*, 13(4):627–640.

Redner, R. A. and Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26(2):195–239.

Rousset, O. G., Ma, Y., and Evans, A. C. (1998). Correction of partial volume effects in PET: Principle and validation. *J. Nuc. Med.*, 39(5):904–911.

Sastry, S. and Carson, R. E. (1997). Multimodality Bayesian algorithm for image reconstruction in PET: A tissue composition model. *IEEE Trans. Med. Imag.*, 16(6):750–761.

Shepp, L. and Vardi, Y. (1982). Maximum likelihood reconstruction for emission tomography. *IEEE Trans. on Medical Imaging*, 1(2):113–122.

Tenenbaum, J. and Freeman, W. (1997). Separating style and content. In *Advances in Neural Information Processing Systems (NIPS) 9*, pages 662–668. MIT Press, Cambridge, MA.

Videen, T. O., Perlmutter, J. S., Mintun, M. A., and Raichle, M. E. (1988). Regional correction of positron emission tomography data for the effects of cerebral atrophy. *Journal of Cerebral Blood Flow and Metabolism*, 8(5):662–670.

Wells III, W., Viola, P., Atsumi, H., Nakajima, S., and Kikinis, R. (1996). Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–52.

Yuille, A. L., Stolorz, P., and Utans, J. (1994). Statistical physics, mixtures of distributions, and the EM algorithm. *Neural Computation*, 6(2):334–340.

Zhang, Y., Fessler, J., Clinthorne, N., and Rogers, W. (1994). Joint estimation for incorporating MRI anatomic images into SPECT reconstruction. In Trendler, R., editor, *IEEE Nucl. Sci. Sym. Med. Imaging Conf.*, pages 1256–1260, Norfolk, Virginia USA. IEEE.

Figures

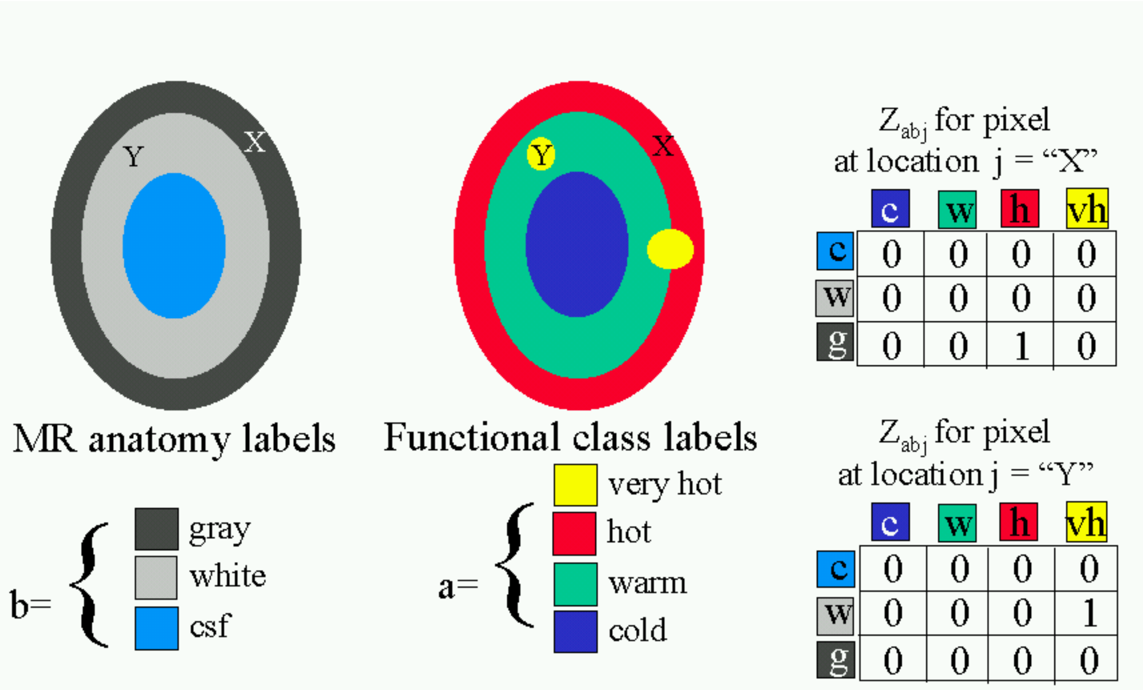

Figure 1: Illustration of the joint classification $z_{ab,j}$. On the left is an anatomical "MR" image with three regions and three classes (gray, white and csf) delineated in gray, blue and black colors respectively. On the right is a functional image with five regions and four classes (very hot, hot, warm and cold) delineated in yellow, red, green, and blue respectively. The very hot class occurs twice in two spatially separated regions. The joint classification matrices are shown for two pixels—one in the (csf, hot) region and the other in the (gray, very hot) region. Note that z being the complete data, is inaccessible. It is estimated by the algorithm.

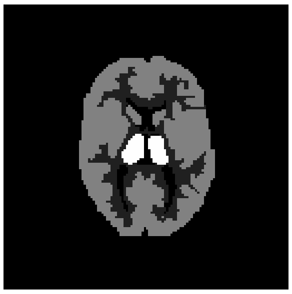

Figure 2: The "ground truth" functional source image. It has $K = 4$ intensity classes as seen in the gray scale encoding. With the exception of the two bright central regions, the anatomical label field with $L = 3$ regions, follows the functional image classification. For the bright regions, the anatomical regions correspond to the outer "gray" matter region.

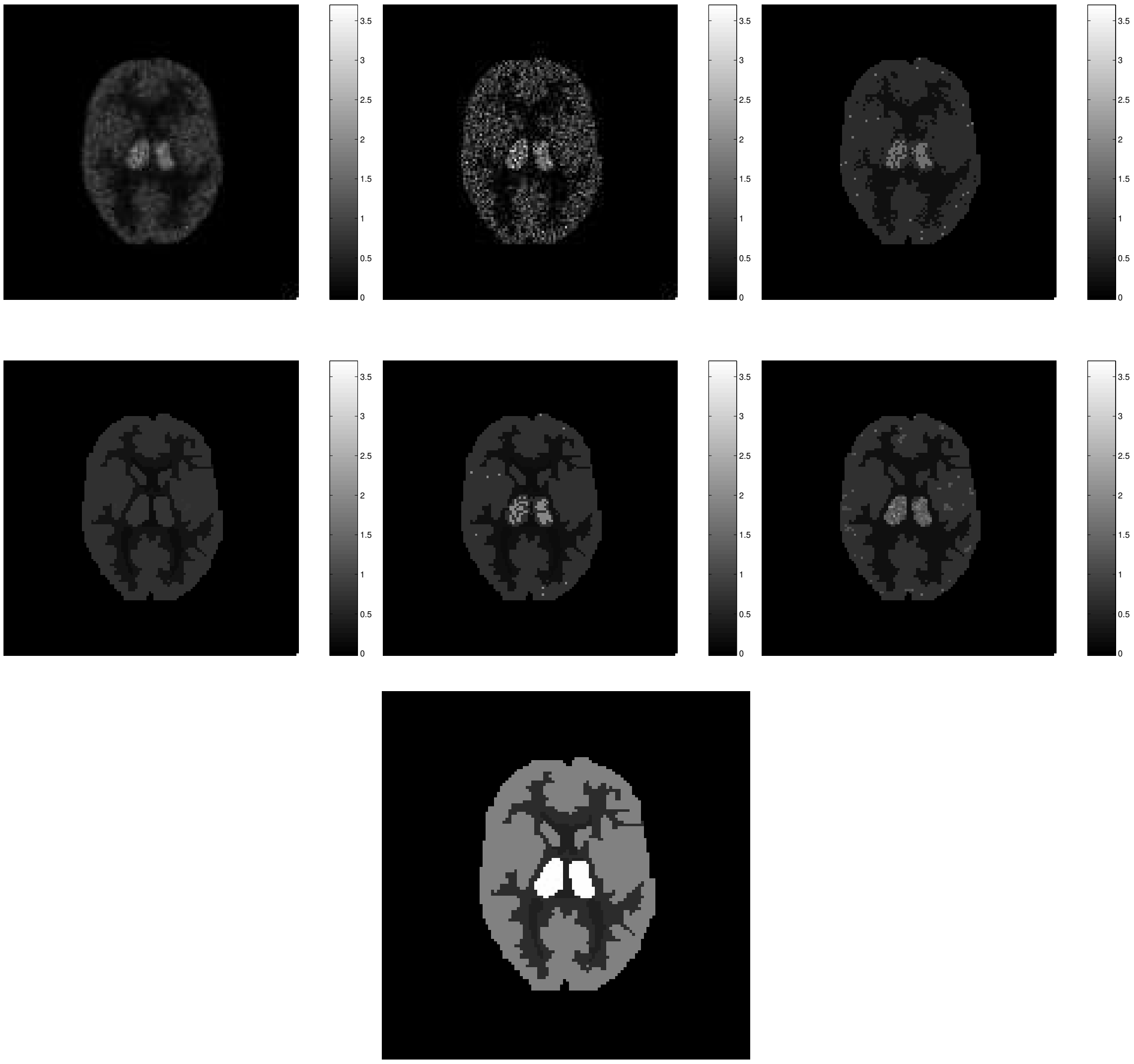

Figure 3: **Anecdotal Reconstructions: Top row**. Left: EM1. Middle: EM2. Right: MIX **Middle row**. Left: MR. Middle: JENT. Right: MI. **Bottom row**: MIz. (See text for description of the different algorithms.)

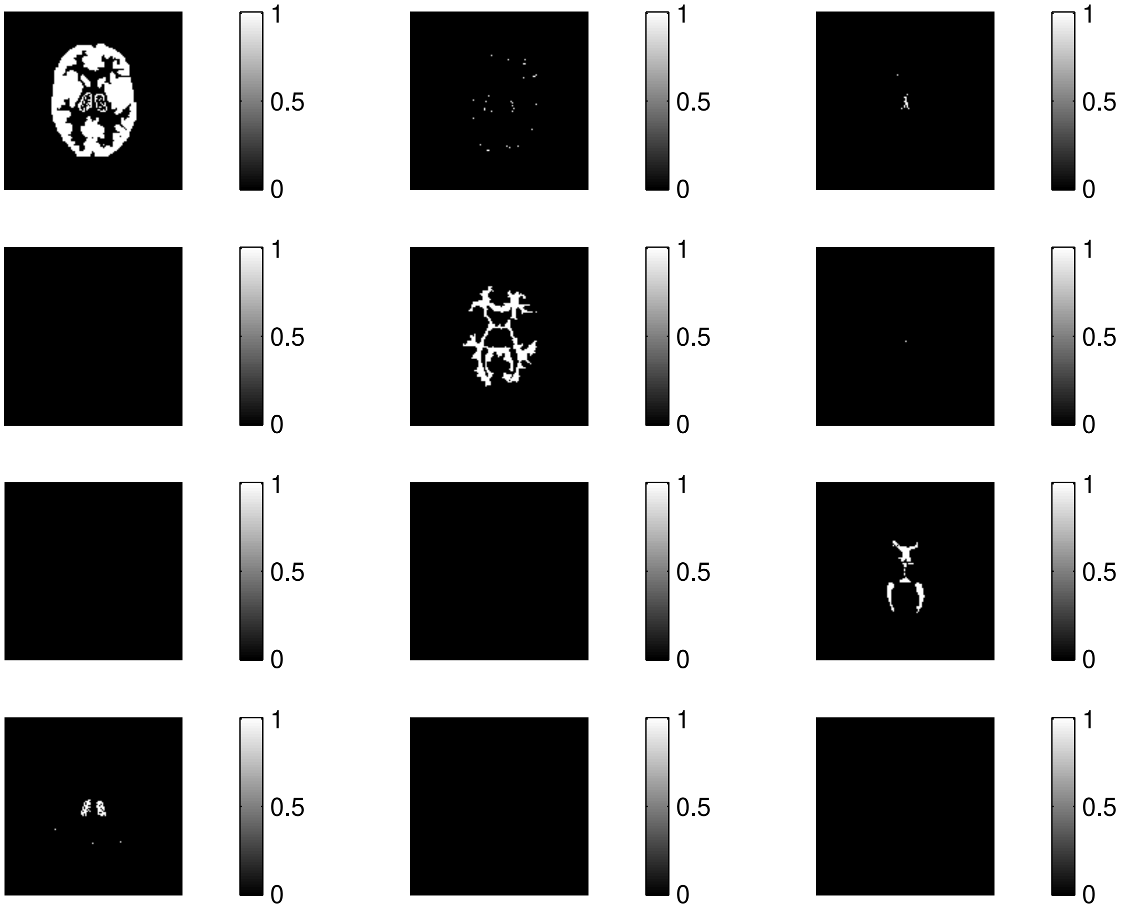

Figure 4: **Classification images** z_{ab} for joint entropy (JENT) reconstruction. The columns correspond to $L = 3$ anatomical classes and the rows correspond to $K = 4$ ET classes. The bottom left image corresponds to the aberrant ET class that does not have an anatomical counterpart. Note the presence of the two central bright regions in this image.

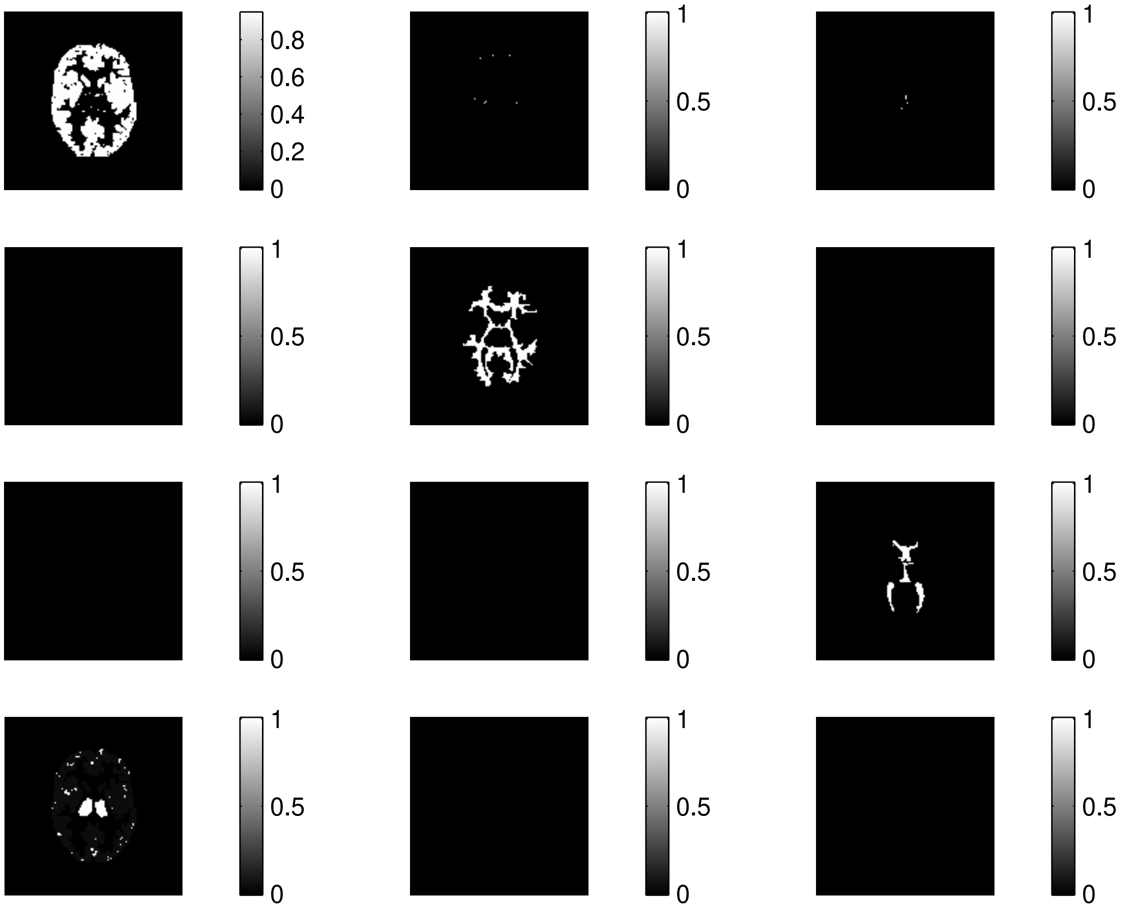

Figure 5: **Classification images** z_{ab} for mutual information (MI) reconstruction. The columns correspond to $L = 3$ anatomical classes and the rows correspond to $K = 4$ ET classes. The bottom left image corresponds to the aberrant ET class that does not have an anatomical counterpart. Note the presence of the two central bright regions in this image.

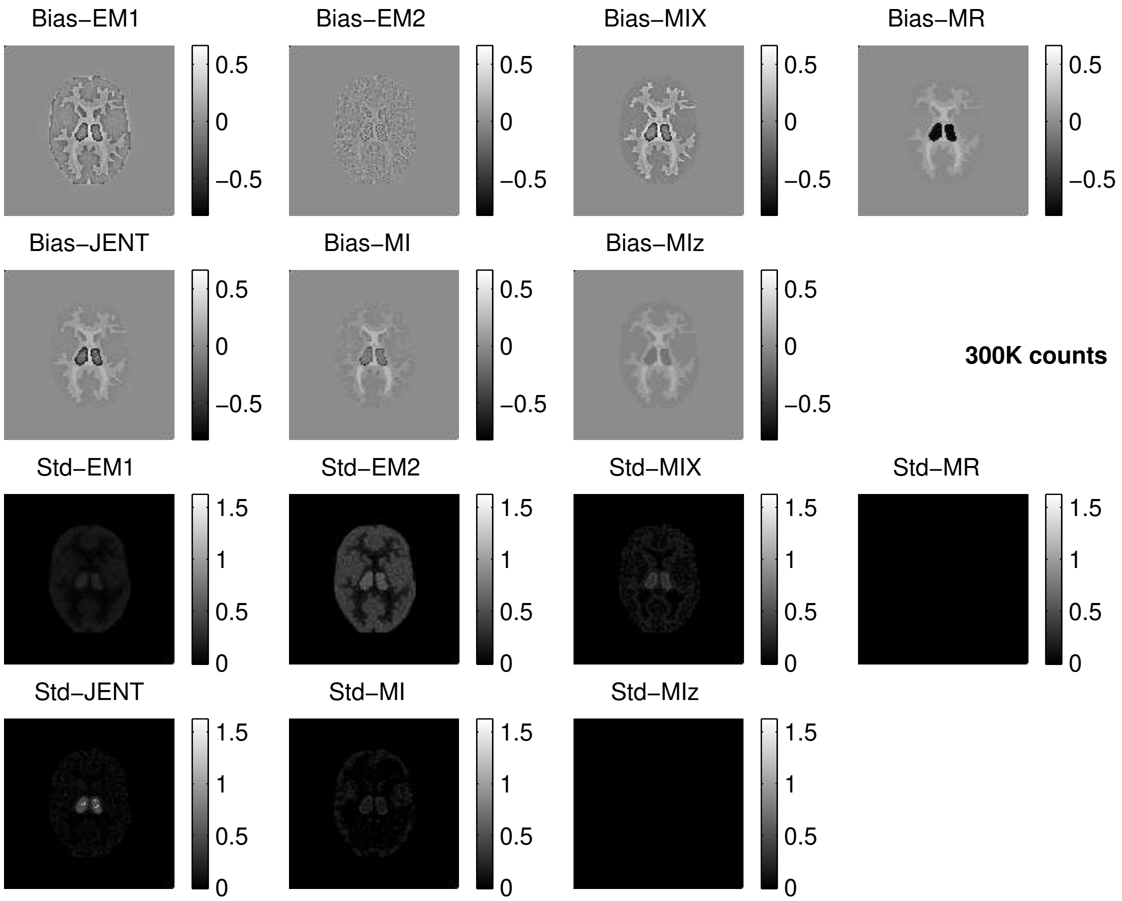

Figure 6: Bias/variance images for the seven different algorithms. (see text for description.)

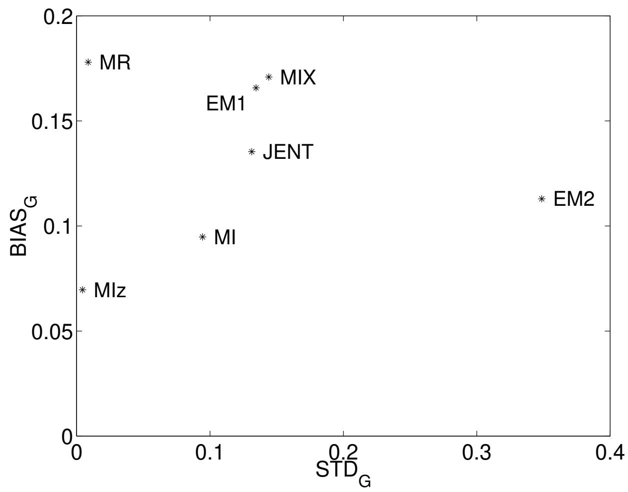

Figure 7: Bias/variance performance for the seven different algorithms. (see text for description.)