Model-based Estimation of Computed Tomography Images

Fekadu L. Bayisa and Jun Yu
Department of Mathematics and Mathematical Statistics, Umeå University, Umeå, Sweden

ARTICLE HISTORY
Compiled May 11, 2017

ABSTRACT
There is a growing interest to get a fully MR based radiotherapy. The most important development needed is to obtain improved bone tissue estimation. Existing model-based methods have performed poorly on bone tissues. This paper aims to obtain improved estimation of bone tissues. Skew-Gaussian mixture model (SGMM) is proposed to further investigate CT image estimation from MR images. The estimation quality of the proposed model is evaluated using leave-one-out cross-validation method on real data. In comparison with the existing model-based approaches, the approach utilized in this paper outperforms in estimation of bone tissues, especially on dense bone tissues.

KEYWORDS
computed tomography; magnetic resonance imaging; CT image estimation; model-based estimation; skew-normal mixture model

1. Introduction

Magnetic resonance (MR) imaging and computed tomography (CT) are the most widely used diagnostic imaging technologies in medicine. They are used to obtain more detailed cross-sectional images of human body. CT uses ionizing radiation to record a pattern of radiodensities and creates cross-sectional images. The ionizing radiations attenuate as they pass through the tissues of patients. The amount of attenuation depends on the tissue types. The differences in attenuation between adjacent tissues create contrast on CT images. Tissues with higher (or lower) attenuation appear brighter (or darker) on grayscale CT images. As a result, air, soft, and bone tissues appear as darkest, darker, and white on grayscale CT images. Therefore, CT image is excellent for identifying and assessing the structures of bone tissues. On the other hand, exposing a patient to ionizing radiation in CT imaging may have a risk for radiation-related cancer.

MR imaging is remarkably different from CT. It does not depend on ionizing radiation. It relies on the absorption and emission of radio waves from tissue protons exposed to a strong magnetic field. Thus, MR imaging is safer than CT imaging. The relative MR signal intensity differences between adjacent anatomic structures determine tissue contrast on MR images. With regard to soft tissue structures, there is much better contrast on MR images than on CT images. Moreover, MR imaging noticeably improves the delineation of tumor and healthy tissue better than CT. However, MR
imaging is poor in detecting bone tissues. Bone, air, and rapidly flowing blood appear black on grayscale MR images.

A fully MR based radiotherapy enhances tissue contouring and precision in soft tissue therapy setup. It also improves biological information at treatment planning and imaging of therapy response. This new innovation based treatment planning avoids registration errors between CT and MR images. Moreover, it is a cost effective approach as it reduces redundant imaging. However, due to better bone tissue imaging of CT, co-registered CT and MR images complement each other. MR images are not directly applicable to attenuation correction. But, CT image is vital for attenuation correction in positron emission tomography (PET) imaging. This is due to the direct relation between CT image intensities and PET attenuation coefficients. However, CT scanner is not available in recently combined PET/MR imaging scanner. Thus, a fully MR based radiotherapy can be effective if a reliable CT image estimation is in place. As a result, it is necessary to investigate the estimation of CT images from MR images.

Huynh et al. use a learning-based method to estimate CT image from MR image. A patch of CT image is estimated directly from a given MR image patch using structured random forest. The robustness of the estimation is evaluated using a new ensemble model. Nie et al. propose a 3D deep learning-based method, a 3D fully convolutional neural network, for patch-wise estimation of CT images from MR images. The neural network generates structured output and it preserves the neighborhood information in the estimated CT image. Arabi et al. suggest a two-step atlas-based algorithm to estimate CT image from MR image sequences. The estimation is mainly concerned with pinpointing of bone tissues.

Johansson et al. use a Gaussian mixture model (GMM) to obtain CT substitute from MR images without taking spatial dependence between neighboring voxels into account. It is a voxel-wise estimation of CT image from MR images. Johansson et al. have investigated the uncertainty associated to the voxel-wise estimation of CT. Kuljus et al. have utilized hidden Markov model (HMM) and Markov random field model (MRF) to extend the work of Johansson et al. by considering spatial dependence between neighboring voxels. Kuljus et al. are motivated to compare the estimation quality of GMM, HMM and MRF. In terms of mean absolute error, HMM outperforms the other models. Moreover, it is computationally robust than MRF. However, it has a weaker estimation quality on dense bone tissues. Even though MRF is superior in estimation performance on bone tissues, it is computationally costive.

The main aim of this article is to further investigate the voxel-wise estimation of CT images from MR images by partitioning the data into non-bone and bone tissues. According to Johansson et al. and Kuljus et al., the estimation of CT image is poor on air and bone tissue regions. It is this result that motivates the partitioning of the data into non-bone tissues (white matter, blood, water, fat, gray matter, air, etc) and bone tissues (cortical bone, cancellous bone, etc) in order to further explore the estimation of CT images. Even though there is no clear-cut CT image intensity boundary between these tissue types, Waterstram-Rich and Gilmore and Washington and Leaver provide informative threshold delimiting these tissues. Waterstram-Rich and Gilmore use 150 Hounsfield units (HU) as a lower limit for bone tissues. On the other hand, Washington and Leaver use 200HU as an approximate delimiting value of the tissues.

The partitioning of the data may introduce skewness. Consequently, there is a need to relax the normality assumption used in Azzalini proposes a uni-
variate skew-normal model that relaxes the normality assumption by incorporating a skewness parameter in the distributional assumption. Azzalini and Valle [3] extend the univariate skew-normal to a multivariate skew-normal. A multivariate skew-normal is a tractable extension of a multivariate normal distribution with extra parameter to regulate skewness. Lin et al. [17] introduce a univariate skew-normal mixture model in order to deal with population heterogeneity and skewness. Lin [16] extends the univariate skew-normal mixture model to a multivariate skew-normal mixture model which is an alternative to the most widely used multivariate Gaussian mixture model. Cabral et al. [5] have proposed mixture models that involve members of skew-normal independent distributions class (the skew-normal, the skew-t, the skew-slash and the skew-contaminated normal). The mixture models are developed using the multivariate skew-normal distribution in [3]. In comparison to the existing developments, Cabral et al. [5] have developed an EM-type algorithm that removes some obstacles (for instance, Monte Carlo integration) during parameter estimation process. In this article, a mixture model consisting the multivariate skew-normal distribution in [3] is utilized and a simplified EM-algorithm for its parameter estimation is developed.

In this work, skew-Gaussian mixture model (SGMM) is used to further explore the estimation of CT images. SGMM involves a weighted sum of the joint skew-normal distributions of a CT image intensity and its corresponding intensities of MR images. The number of skew-normal distributions in the mixture depends on the number of underlying tissue types. Latent variables that represent the underlying tissue types are utilized during parameter estimation process of the model through incomplete data assumption in EM-algorithm framework [7]. Voxel-wise point estimator of CT image is obtained as a weighted sum of the conditional expected value of a CT image intensity given its corresponding intensities of MR images and the underlying tissue type. The probability that an underlying tissue type is determined based on the intensities of MR images is used as a weight of the conditional expected value.

This study is also interested to compare the estimation performance of SGMM and GMM* (GMM applied to each partition) on the tissue regions. Moreover, it is also aimed at comparing the predictive quality of SGMM and GMM* with HMM, MRF and GMM on the bone tissues. HMM, MRF, and GMM are trained on the full data (data that are not partitioned into non-bone and bone tissues).

This article is organized as follows. The second section describes data acquisition and demonstrates statistical method. The third section presents the results obtained and the final section discusses the implication of the results.

2. Statistical methodology

This section describes the data, SGMM formulation, and its parameter estimation method. It also demonstrates CT image estimation and its assessment methods. For the remaining models (GMM, HMM and MRF), we refer to Kuljus et al. [14, and references therein].

2.1. Data acquisition

CT and MR images were obtained from head of five patients. Four MR images were acquired from each patient using two dual echo ultrashort echo-time sequences with flip angles of 10 degrees and 30 degrees. The ultrashort echo-time sequences sampled a first echo (free induction decay) and a second echo (gradient echo) from the same...
excitation with an echo time of 0.07 microsecond and 3.76 microsecond. MR image of a patient was reconstructed to $192 \times 192 \times 192$ matrix. An entry in the matrix represents a signal intensity corresponding to a three-dimensional tissue (voxel) with size 1.33 mm $\times$ 1.33 mm $\times$ 1.33 mm. One CT image of a patient was acquired using gradient echo Lightspeed with a 2.5 mm slice thickness. The acquired CT image was reconstructed with an in-plane resolution of 0.78 mm $\times$ 0.78 mm. One binary mask (an image with voxel value 1 (or 0) representing the region of interest (or the surrounding air)) was also developed to demarcate the head of a patient from its surrounding air. The main use of the binary mask is to exclude the surrounding air from the acquired CT and MR images. For each patient, the binary mask, the CT image, and the four MR images were co-registered and resampled to the same resolution (voxel-to-voxel correspondence and set to the same voxel dimension) using linear interpolation. For further technical details, we refer to Johansson et al. [10]. Voxel values of the CT image, the binary mask, and the four MR images were arranged into six columns to obtain data for a patient. The arranged data of each patient were column stacked and the surrounding air removed to obtain data for model fitting. Figure 1 shows a slice data for a given patient.

![Binary mask, CT image, MR images](image)

Figure 1.: Binary mask ((a)), CT image ((b)) and MR images ((c)-(f))

### 2.2. Data Partitioning

This subsection describes data partitioning during model training. It also demonstrates how MR images of a new patient are utilized during CT image prediction.

#### 2.2.1. Data partition: Model training

CT image intensity threshold is utilized to partition the data into two tissue regions. Using 150 HU CT image intensity as a limit, 50 HU (is selected to take the delimiting value provided by Washington and Leaver [20]) overlap is allowed during parameter estimation process. The overlap of the tissue regions is motivated in order to minimize
the effect of fuzzy boundary of the tissue regions. Accordingly, CT image intensities in (-1024HU, 200HU) and (100HU, 3071HU] are assumed to represent non-bone and bone tissues. The minimum size of observations for non-bone and bone tissues are 6214160 and 1292068.

2.2.2. MR images partition: CT image estimation

Only MR images of a new patient are available to predict CT image of the patient. Thus, there is a need to partition the MR images of the new patient into non-bone and bone tissues. Poor information about bone tissues is available on MR images. As a result, CT image for the new patient is estimated using the "best" trained model. In this case, the model is trained on the full data. CT image intensity threshold and estimated CT image are employed to obtain MR datasets corresponding to the two tissue regions. The trained models on the tissue regions are utilized to obtain the desired CT image of the new patient.

2.3. Statistical model: mixture of multivariate skew-normal model

Let \( Y_{i1} \) and \( Y_{i2} = (Y_{i2}, Y_{i3}, \ldots, Y_{id})' \) represent voxel \( i \) of CT image and its corresponding MR images. In our real data, \( d=5 \). A \( d \)-dimensional random vector \( Y_i = (Y_{i1}, Y_{i2})' \) is assumed to follow a multivariate skew-normal distribution \( SN(y_i|\eta, \Sigma, \lambda) \) with a \( d \)-dimensional location parameter vector \( \eta \), a \( d \times d \)-dimensional positive definite dispersion matrix \( \Sigma \), and a \( d \)-dimensional skewness parameter vector \( \lambda \). Its density can be given by

\[
f(y_i|\eta, \Sigma, \lambda) = 2N(y_i|\eta, \Sigma) \Phi \left( \lambda' \Sigma^{-1/2} (y_i - \eta) \right),
\]

where \( \Sigma^{-1/2} \Sigma^{-1/2} = \Sigma^{-1} \), \( \Phi(\cdot) \) is a univariate standard normal distribution function, \( i = 1, 2, 3, \ldots, n \), and \( n \) is the number of voxels. According to Lachos et al. [15], the stochastic representation of \( Y_i \) may be given by

\[
Y_i = \eta + \Sigma^{1/2} \delta U_i + \Sigma^{1/2} (1 - \delta \delta')^{1/2} V_i,
\]

where

\[
U_i \sim HN(u_i|0, 1, (0, \infty)), \quad V_i \sim N(v_i|0, 1), \quad \text{and} \quad \delta = \frac{\lambda}{\sqrt{1 + \lambda^2}}.
\]

In this setting, \( \mathbf{0} \) represents the \( d \)-dimensional zero vector, \( \mathbf{1} \) denotes the \( d \times d \)-dimensional identity matrix, and \( HN(u_i|0, 1, (0, \infty)) \) is a half-normal distribution. Moreover, \( V_i \) and \( U_i \) are independent. Using equation (2), a hierarchical model can be given by

\[
Y_i|U_i = u_i \sim N(y_i|\eta + u_i \xi, \Omega), \quad U_i \sim HN(u_i|0, 1, (0, \infty)),
\]

where

\[
\xi = \Sigma^{1/2} \delta, \quad \text{and} \quad \Omega = \Sigma^{1/2} (1 - \delta \delta') \Sigma^{1/2} = \Sigma - \xi \xi'.
\]
Let \( \{Z_i\}_{i=1}^n \) be a multinomial trials process representing the underlying tissue types. Define an indicator variable

\[
Z_{ik} = 1(Z_i=s) = \begin{cases} 
1, & \text{if } k = s, \\
0, & \text{otherwise},
\end{cases}
\]

where \( k = 1, 2, \cdots, K \). The definition implies that \( P(Z_{ik} = 1) = P(Z_i = k) \). Let \( P(Z_i = k) = \pi_k \), which represents the weight that the \( i \)-th observation belongs to a tissue class \( k \). To incorporate tissue heterogeneity into the statistical modeling, \( Y_i | Z_{ik} = 1 \) is assumed to follow a multivariate skew-normal distribution \( SN(y_i | \eta_k, \Sigma_k, \lambda_k) \). This means that \( Y_i \) follows a mixture of multivariate skew-normal distributions. Its density may be given by

\[
f(y_i | \Psi) = \sum_{k=1}^{K} \pi_k SN(y_i | \eta_k, \Sigma_k, \lambda_k),
\]

where \( \pi_k \geq 0, \sum_{k=1}^{K} \pi_k = 1 \), \( \psi_k' = (\pi_k, \eta_k, \Sigma_k, \lambda_k) \), \( \Psi = (\psi_1, \psi_2, \cdots, \psi_K) \), and \( k = 1, 2, \cdots, K \).

The unknown parameters in the matrix \( \Psi \) are estimated from independent observations \( y_i \).

### 2.4. Parameter estimation method

The log-likelihood function of the data \( y = (y_1, y_2, \cdots, y_n)' \) is given by

\[
\log f(y | \Psi) = \sum_{i=1}^{n} \log \left\{ \sum_{k=1}^{K} \pi_k SN(y_i | \eta_k, \Sigma_k, \lambda_k) \right\}.
\]

In general, there is no explicit analytical solution for \( \arg \max_{\Psi} \log f(y | \Psi) \). However, iterative maximizing procedure under the idea of incomplete data via EM-algorithm can be used to obtain an optimal estimate of the parameters. Let \( Z_i \) be a \( K \)-dimensional column vector of \( Z_{ik} \). Its realization is a \( K \)-dimensional vector consisting 1 at only one location and 0 at the remaining locations. The latent random vector \( Z_i \) follows a multinomial distribution with one trial and \( P(Z_{ik} = 1) = \pi_k \). Using the indicator variable \( Z_{ik} \), the hierarchical model (3) can be extended to

\[
Y_i | U_i = u_i, Z_{ik} = 1 \sim N(y_i | \eta_k + u_i \xi_k, \Omega_k),
\]

\[
U_i | Z_{ik} = 1 \sim HN(u_i | 0, 1, (0, \infty)),
\]

where

\[
\delta_k = \frac{\lambda_k}{\sqrt{1 + 1/\lambda_k^2}}, \quad \xi_k = \Sigma_k^{1/2} \delta_k, \quad \text{and} \quad \Omega_k = \Sigma_k^{1/2} (1 - \delta_k \delta_k^t) \Sigma_k^{1/2} = \Sigma_k - \xi_k \xi_k.
\]
The observed data $y$ is assumed to be incomplete data. It is augmented with the latent matrix $z = (z_1, z_2, \cdots, z_n)'$ and the latent vector $u = (u_1, u_2, \cdots, u_n)'$ to form a complete dataset $(y, u, z)$ in EM-algorithm framework. Assuming that $(Y_i, U_i, Z_i)$ is independent of $(Y_j, U_j, Z_j)$ for every $i \neq j$, the complete-data log-likelihood is given by

$$
\log f(y, u, z|\Theta) = \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik} \left\{ -\frac{1}{2} (y_i - \eta_k - \xi_k u_i)' \Omega_k^{-1} (y_i - \eta_k - \xi_k u_i) + \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik} \left\{ \log \pi_k - \frac{1}{2} \log |\Omega_k| + m \right\} \right\},
$$

where

$$
\theta_k = (\pi_k, \eta_k, \xi_k, \Omega_k)', \quad \Theta = (\theta_1, \theta_2, \cdots, \theta_K), \quad k = 1, 2, \cdots, K,
$$

and $m$ is a constant function of parameters.

The E-step of EM-algorithm involves computing the expected value of the complete-data log-likelihood given $Y$ and the current estimate $\Theta^{old}$ of $\Theta$. It may be given by the Q-function

$$
Q(\Theta, \Theta^{old}) = E \left[ \log f(Y, U, Z|\Theta) | Y, \Theta^{old} \right].
$$

Using equation (4), the expected value in equation (5) involves computing $E [Z_{ik}|Y, \Theta^{old}]$, $E [Z_{ik}U_i|Y, \Theta^{old}]$, and $E [Z_{ik}U_i^2|Y, \Theta^{old}]$.

The expected value $E [Z_{ik}|Y, \Theta^{old}]$ can be given by

$$
E [Z_{ik}|Y, \Theta^{old}] = \frac{\pi_k^{old} \mathcal{SN} (y_i|\eta_k^{old}, \Sigma_k^{old}, \lambda_k^{old})}{\sum_{j=1}^{K} \pi_j^{old} \mathcal{SN} (y_i|\eta_j^{old}, \Sigma_j^{old}, \lambda_j^{old})},
$$

where $\gamma_{ik}$ is the responsibility that component $k$ of the mixture takes for explaining the observation $y_i$. Let $\vartheta_{ik} = E [Z_{ik}U_i|Y, \Theta^{old}]$, and $\psi_{ik} = E [Z_{ik}U_i^2|Y, \Theta^{old}]$. Then $\vartheta_{ik}$ can be simplified as follows.

$$
\vartheta_{ik} = E [Z_{ik}E [U_i|Y, Z_{ik}, \Theta^{old}] | Y, \Theta^{old}],
$$

$$
= E [U_i|Y, Z_{ik} = 1, \Theta^{old}] E [Z_{ik}|Y, \Theta^{old}],
$$

$$
= \gamma_{ik}^{old} E [U_i|Y, Z_{ik} = 1, \Theta^{old}].
$$
Using similar procedure, \( \psi_{ik} \) is given by

\[
\psi_{ik} = \gamma_{ik} E \left[ U_i^2 | Y_i, Z_{ik} = 1, \Theta^{old} \right].
\]

If equations (1) and (3) are satisfied, then using inverse matrix adjustment formula in [19] and matrix determinant lemma in [8],

\[
U_i | Y_i = y_i \sim T N \left( u_i \left( \xi' \Omega^{-1} (y_i - \eta) \right), \frac{1}{1 + \xi' \Omega^{-1} \xi} \right),
\]

where \( T N \left( u | \mu, \sigma^2, (0, \infty) \right) \) is a truncated normal with location parameter \( \mu \), scale parameter \( \sigma \) and support \( (0, \infty) \). Based on the truncated normal probability distribution,

\[
E \left[ U_i | Y_i = y_i, U_i > 0 \right] = \frac{1}{\beta} \left[ \alpha + \phi(\alpha) \right],
\]

\[
\text{var} \left[ U_i | Y_i = y_i, U_i > 0 \right] = \frac{1}{\beta^2} \left[ 1 - \alpha \phi(\alpha) - \left( \frac{\phi(\alpha)}{\Phi(\alpha)} \right)^2 \right],
\]

\[
E \left[ U_i^2 | Y_i = y_i, U_i > 0 \right] = \frac{1}{\beta^2} \left[ 1 + \alpha \phi(\alpha) + \alpha^2 \right],
\]

where

\[
\alpha = \frac{\xi' \Omega^{-1} (y_i - \eta)}{\sqrt{1 + \xi' \Omega^{-1} \xi}}, \quad \beta = \sqrt{1 + \xi' \Omega^{-1} \xi},
\]

and \( \phi(\cdot) \) is a univariate standard normal density. The expected values

\[
E \left[ U_i | Y_i, Z_{ik} = 1, \Theta^{old} \right], \quad \text{and} \quad E \left[ U_i^2 | Y_i, Z_{ik} = 1, \Theta^{old} \right].
\]

in equations [5] and [7] are obtained from equations [8] and [9] by replacing \( \eta, \xi, \) and \( \Omega \) with their corresponding estimates \( \eta_{ik}^{old}, \xi_{ik}^{old}, \) and \( \Omega_{ik}^{old} \). The M-step of EM-algorithm is given by

\[
\Theta^{new} = \arg\max_{\Theta} Q \left( \Theta, \Theta^{old} \right),
\]

and it is available in closed form. EM-algorithm parameter estimates are updated as shown in Algorithm 1.
Algorithm 1 Updating parameter estimates of the model

1: Initial value of the parameters \( (m = 0) \): \( \pi_k^{(m)} \), \( \Omega_k^{(m)} \), \( \eta_k^{(m)} \), and \( \xi_k^{(m)} \);
2: E-step:

\[
\gamma_{ik}^{(m)} = \frac{\pi_k^{(m)} SN \left( y_i | \eta_k^{(m)} , \Sigma_k^{(m)} , \lambda_k^{(m)} \right)}{\sum_{j=1}^{K} \pi_j^{(m)} SN \left( y_i | \eta_j^{(m)} , \Sigma_j^{(m)} , \lambda_j^{(m)} \right)};
\]

\[
\phi_{ik}^{(m)} = \frac{\gamma_{ik}^{(m)}}{\beta_{ik}^{(m)}} \left[ \alpha_{ik}^{(m)} + \frac{\phi \left( \alpha_{ik}^{(m)} \right)}{\Phi \left( \alpha_{ik}^{(m)} \right)} \right];
\]

\[
\psi_{ik}^{(m)} = \frac{\gamma_{ik}^{(m)}}{\beta_{ik}^{(m)}} \left[ 1 + \frac{\phi \left( \alpha_{ik}^{(m)} \right)}{\Phi \left( \alpha_{ik}^{(m)} \right)} \right]^2,
\]

where

\[
\Sigma_k^{(m)} = \Omega_k^{(m)} + \xi_k^{(m)} \xi_k^{(m)\prime};
\]

\[
\lambda_k^{(m)} = \frac{\left[ \Sigma_k^{(m)} \right]^{-1/2} \xi_k^{(m)}}{\sqrt{1 - \xi_k^{(m)\prime} \left[ \Sigma_k^{(m)} \right]^{-1} \xi_k^{(m)}}};
\]

\[
\alpha_{ik}^{(m)} = \frac{\xi_k^{(m)\prime} \left[ \Omega_k^{(m)} \right]^{-1} \left( y_i - \eta_k^{(m)} \right)}{\beta_{ik}^{(m)}},
\]

\[
\beta_{ik}^{(m)} = \sqrt{1 + \xi_k^{(m)\prime} \left[ \Omega_k^{(m)} \right]^{-1} \xi_k^{(m)}};
\]

3: M-step:

\[
\pi_k^{(m+1)} = \frac{1}{n} \sum_{i=1}^{n} \gamma_{ik}^{(m)};
\]

\[
\eta_k^{(m+1)} = \left[ \sum_{i=1}^{n} \gamma_{ik}^{(m)} \right]^{-1} \sum_{i=1}^{n} \gamma_{ik}^{(m)} \left( y_i - \phi_{ik}^{(m)} \xi_k^{(m)} \right);
\]

\[
\xi_k^{(m+1)} = \left[ \sum_{i=1}^{n} \psi_{ik}^{(m)} \right]^{-1} \sum_{i=1}^{n} \psi_{ik}^{(m)} \left( y_i - \eta_k^{(m+1)} \right);
\]

\[
\Omega_k^{(m+1)} = \left[ \sum_{i=1}^{n} \gamma_{ik}^{(m)} \right]^{-1} \sum_{i=1}^{n} \left\{ \gamma_{ik}^{(m)} \left( y_i - \eta_k^{(m+1)} \right) \left( y_i - \eta_k^{(m+1)} \right) \right. \\
\left. + \phi_{ik}^{(m)} \left( y_i - \eta_k^{(m+1)} \right) \xi_k^{(m+1)} + \xi_k^{(m+1)} \left( y_i - \eta_k^{(m+1)} \right) \right\};
\]

4: If stopping criterion is achieved, stop. If not, set \( m = m + 1 \), and go to step 2.

In general, EM-algorithm converges to a local optimum. As a result, different initial values for the parameters are utilized during the estimation process in order to obtain
the optimal estimates. $K$-means clustering is employed to initialize the location parameters, mixing coefficients and dispersion matrices. The other parameters are initialized randomly. In MRF, log-likelihood value can not be computed analytically as explained in [14] and hence can not be used to select the optimal estimates. Consequently, mean squared error is used as criterion in selecting the 'optimal' estimates in SGMM in order to utilize the same criterion for the models used in this work. Two steps are employed during the estimation process. The 'optimal' parameter estimates are obtained for a given number of tissue types (number of classes). The step is repeated for a number of classes. The 'optimal' parameter estimates and number of classes are chosen based on mean squared error. The stopping criterion for the convergence of the estimation process is $\max_{ik} |\pi_{ik}^{(m+1)} - \pi_{ik}^{(m)}|$ with an upper limit $5 \times 10^{-5}$, where $m$ denotes the iteration number of EM-algorithm.

2.5. Estimation of CT images

Let the $d$-dimensional vectors $\eta$, $\nu = \Sigma^{-1/2}\lambda$, and $d \times d$ dispersion matrix $\Sigma$ be partitioned as follows.

$$Y_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \end{bmatrix}, \quad \eta = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}, \quad \nu = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$  

The dimension of $Y_{i1}$, $\eta_1$, $\nu_1$, and $\Sigma_{11}$ is a $1 \times 1$. The random variable $Y_{i1}$ represents the $i$th voxel in CT image, and the random vector $Y_{i2}$ denotes the corresponding voxel in MR images. If

$$Y_i \sim SN(y_i|\eta, \Sigma, \lambda),$$  

then

$$Y_{i2} \sim SN(y_{i2}|\eta_2, \Sigma_{22}, \lambda_2),$$

where

$$\lambda_2 = \Sigma_{22}^{1/2} \left( \Sigma_{22}^{-1} \Sigma_{21} \nu_1 + \nu_2 \right),$$

$$\sqrt{1 + \nu_1^T \Sigma_{11}^{-1} \nu_1}.$$  

According to Khounsriavash et al. [13], if equation (10) holds, then the probability density function of $Y_{i1}|Y_{i2} = y_{i2}$ can be given by

$$f(y_{i1}|y_{i2}) = N(y_{i1}|\eta_1^c, \Sigma_{11}^c) \frac{\Phi \left( \lambda' \Sigma^{-1/2} (y_i - \eta) \right)}{\Phi \left( \lambda_2' \Sigma_{22}^{-1/2} (y_{i2} - \eta_2) \right)},$$  

where

$$\eta_1^c = \eta_1 + \Sigma_{12} \Sigma_{22}^{-1} (y_{i2} - \eta_2), \quad \text{and} \quad \Sigma_{11}^c = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}.$$
Using equation (11), the expected value of \( Y_{i1} | Y_{i2} = y_{i2} \) may be given by

\[
E [ Y_{i1} | Y_{i2} = y_{i2} ] = \eta_i + \frac{\Sigma c_{i1} \nu_1}{\sqrt{1 + \nu_1 \Sigma c_{11} \nu_1}} \phi \left( \lambda_2 \Sigma c_{22}^{-1/2} (y_{i2} - \eta_2) \right).
\] (12)

Hence, we can obtain the point estimator of \( Y_{i1} \) by

\[
E [ Y_{i1} ] = E [ E [ Y_{i1} | Y_{i2}, Z_{i} ] | Y_{i2} ],
\]

\[
= \sum_{k=1}^{K} P (Z_i = k | Y_{i2}) E [ Y_{i1} | Y_{i2}, Z_{i} = k ].
\]

In this framework, the latent variable \( Z_i \) represents the underlying tissue classes. The weight \( P (Z_i = k | Y_{i2}) \) can be computed using Bayes’ theorem. The expected value \( E [ Y_{i1} | Y_{i2}, Z_{i} = k ] \) may be obtained from equation (12) by indexing the parameters with \( k \).

### 2.6. Model validation

The main focus of this work is to study the predictive quality of SGMM and GMM on the tissue regions. It is also aimed at comparing the estimation quality of SGMM, GMM, HMM, and MRF on the bone tissues.

Leave-one-out-cross-validation method is used to compare the predictive quality of the models. Using the threshold CT image intensity, CT image in a validation dataset is partitioned into non-bone and bone tissues. Let \( Y_i \) represents CT image intensity, and \( \hat{Y}_i \) be its corresponding estimated CT image intensity in a given tissue region \( t \). For each tissue region \( t \), square loss function and absolute loss function

\[
\left( \hat{Y}_i^{(t)} - Y_i^{(t)} \right)^2, \text{ and } |\hat{Y}_i^{(t)} - Y_i^{(t)}|, \quad t = 1, 2,
\]

are utilized to assess the estimation cost associated to voxel \( i \). Since mean square error heavily weights the outliers, mean absolute error (MAE) is employed as a main tool to evaluate the estimation performance of the models. It can be given by

\[
MAE_t = \frac{1}{n_t} \sum_{i=1}^{n_t} |\hat{Y}_i^{(t)} - Y_i^{(t)}|,
\]

where \( n_t \) is the number of voxels in partition \( t \).

Peak signal-to-noise ratio (PSNR) can also be used to quantify the overall quality of the estimation. It takes square loss function into account through mean squared error. PSNR may be given by

\[
PSNR = 10 \log_{10} \left( \frac{M^2}{MSE} \right),
\]

where

\[
MSE_t = \frac{1}{n_t} \sum_{i=1}^{n_t} \left( \hat{Y}_i^{(t)} - Y_i^{(t)} \right)^2.
\]
where

\[ MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2, \]

and \( M \) is the maximal intensity in CT image. This procedure is repeated in turn for the five patients. The average of the mean absolute errors is used to compare the estimation quality of the models on each tissue region. The average of the peak signal-to-noise ratios is also utilized to compare the estimation performance of the models. The better model has lower MAE and higher PSNR. Since MAE and PSNR are crude estimation quality measures, smoothed residual and absolute residual plots based on CT image intensities are employed to further evaluate the estimation quality of the models through the tissues of a head. A moving average over non-overlapping windows in CT image intensities with window size 20 HU is utilized as a main tool to investigate and identify the model that outperforms on the bone tissues of a head.

3. Results

The delimiting value 100HU CT image intensity is utilized during CT image estimation. The ‘optimal’ parameter estimates are received for \( K = 5 \) in HMM and MRF; \( K = 6 \) in SGMM and GMM* (for both tissue regions) and \( K = 8 \) in GMM. Table 1 demonstrates a summary of mean absolute errors for the bone tissues.

| Model | Head | SGMM | GMM* | HMM | MRF | GMM |
|-------|------|------|------|-----|-----|-----|
| 1     | 316.25 | 315.16 | 324.94 | 307.95 | 314.41 |
| 2     | 349.52 | 348.05 | 360.03 | 328.89 | 365.12 |
| 3     | 303.58 | 301.78 | 331.16 | 322.48 | 328.01 |
| 4     | 272.21 | 269.33 | 296.04 | 280.63 | 292.78 |
| 5     | 350.41 | 349.28 | 366.13 | 357.22 | 359.56 |
| Average | 318.39 | 316.72 | 335.66 | 319.43 | 331.98 |

The rows of the table represent validation datasets. The table presents mean absolute errors that are received for the models. In terms of average MAE, weaker result is received for HMM and GMM. The remaining models have similar estimation performance on bone tissues. However, the results in Table 2 show that GMM* and SGMM outperform the other models on dense bone tissues (approximately with CT image intensities greater than 900HU according to Washington and Leaver [20]). MRF performs better than HMM and GMM on dense bone tissues. But, it is computationally expensive.
Table 2.: Mean absolute errors for dense bone tissues

| Model | Head | SGMM | GMM* | HMM | MRF | GMM |
|-------|------|------|------|-----|-----|-----|
| 1     | 365.68 | 361.27 | 418.99 | 401.43 | 407.85 |
| 2     | 406.09 | 403.65 | 458.23 | 394.89 | 402.96 |
| 3     | 331.85 | 328.44 | 407.01 | 404.80 | 386.72 |
| 4     | 236.89 | 232.15 | 290.42 | 295.46 | 290.77 |
| 5     | 436.42 | 433.84 | 509.20 | 488.10 | 505.49 |
| Average | 355.39 | 351.87 | 416.77 | 396.94 | 416.76 |

Table 3 demonstrates the estimation quality of the models on non-bone tissues. The best result is received for HMM. However, there is a good contrast between soft tissues and air on MR images. The remaining models have similar behavior on non-bone tissues.

Table 3.: Mean absolute errors for non-bone tissues

| Model | Head | SGMM | GMM* | HMM | MRF | GMM |
|-------|------|------|------|-----|-----|-----|
| 1     | 111.96 | 112.67 | 97.75 | 106.15 | 114.10 |
| 2     | 116.56 | 117.53 | 99.69 | 103.60 | 116.12 |
| 3     | 124.65 | 125.46 | 94.93 | 117.31 | 122.00 |
| 4     | 116.61 | 117.19 | 101.54 | 112.52 | 111.76 |
| 5     | 122.08 | 122.42 | 98.87 | 126.51 | 118.19 |
| Average | 118.37 | 119.05 | 98.55 | 113.22 | 116.43 |

Table 4 presents the overall summary of CT image estimation quality. The results show that HMM outperforms the other models. That is due to its better behavior on non-bone tissues. However, this is not the main interest in this work.

Table 4.: Combined mean absolute errors

| Model | Head | SGMM | GMM* | HMM | MRF | GMM |
|-------|------|------|------|-----|-----|-----|
| 1     | 144.15 | 144.58 | 133.56 | 137.95 | 145.67 |
| 2     | 151.44 | 152.04 | 138.68 | 137.34 | 153.40 |
| 3     | 161.00 | 161.28 | 142.92 | 158.99 | 163.85 |
| 4     | 146.71 | 146.62 | 139.17 | 145.05 | 146.78 |
| 5     | 160.53 | 160.62 | 143.89 | 165.36 | 158.85 |
| Average | 152.77 | 153.03 | 139.65 | 148.94 | 153.71 |

Table 5 demonstrates the prediction accuracy of the models in terms of PSNR. The results show that the models have similar behavior. On the bone tissues, PSNR has shown that the models have similar estimation performance.
Table 5.: Evaluation of the models based on PSNR

| Head | Model   | SGMM | GMM* | HMM  | MRF  | GMM  |
|------|---------|------|------|------|------|------|
| 1    | 19.92   | 19.90| 20.29| 20.30| 20.29|
| 2    | 19.56   | 19.50| 20.23| 20.40| 20.15|
| 3    | 19.75   | 19.71| 20.13| 20.11| 19.69|
| 4    | 19.76   | 19.76| 20.43| 20.46| 20.69|
| 5    | 19.59   | 19.58| 19.78| 19.63| 19.66|
| Average | 19.72 | 19.69| 20.17| 20.18| 20.09|

On the basis of average, mean absolute error is utilized to compare CT image prediction accuracy of the models. Smoothed absolute residual plots are also employed to assess the estimation quality of the models. In comparison to MAE, smoothed absolute residual plots are powerful to evaluate the estimation performance of the models through the tissues of a head. Figure 2 presents smoothed absolute residual plots for the models. The absolute residuals are averaged over non-overlapping windows in CT image intensities with window size 20HU.

![Figure 2: Smoothed absolute residual plot for head 5](image)

It is clear from Figure 2 that none of the models outperform throughout the tissues of the head. However, SGMM and GMM* have best estimation quality on dense bone tissues. Figure 3 shows smoothed residual plot. The deviations of observed and estimated CT image intensities are exploited to obtain average residuals over non-overlapping windows in CT image intensities with size 20HU. It is evident from the plot that HMM, MRF and GMM underestimate on the bone tissues.
4. Discussion

The aim of this study was to further investigate voxel-wise estimation of CT images from MR images. This voxel-wise estimation approach is also utilized in [10, 14]. The existing works show that the estimation quality on the bone tissues is poor. This study was aimed at probing the estimation of CT images by partitioning the data into non-bone and bone tissues. Specifically, the main focus of this study was to obtain improved bone tissue estimation. SGMM was proposed to relax the distributional assumption in [10, 14]. It was motivated to take the asymmetrical distribution that could arise from the partitioning of the data into account. SGMM and GMM were trained on each tissue region. The study was also aimed at comparing the estimation performance of SGMM, HMM, GMM, and MRF on the bone tissues. Spatial dependence between neighboring voxels was taken into account in HMM and MRF. The full data were used in training HMM, GMM, and MRF.

A simplified EM-algorithm was developed for SGMM parameter estimation. For the remaining models, EM-algorithm was utilized to estimate their parameters except in MRF. In MRF, EM-gradient algorithm was used to estimate its parameters. The updates of the parameters in M-step of EM-algorithm were difficult to get in explicit form. As a result, gradient based optimization was utilized during parameter estimation process. Moreover, Gibbs sampling was used in the E-step of the algorithm. Thus, the estimation process was expensive in MRF. Unlike the other models, log-likelihood function in MRF involves Gibbs field and it is not computable. This means that log-likelihood based model selection is not feasible analytically. Consequently, mean squared error was employed in selecting the optimal parameters of the models.

Table 4 demonstrates that HMM outperforms the other models. This is essentially due to its best estimation performance on non-bone tissues. This behaviour of HMM is demonstrated in Table 3. According to Karlsson et al. [12], the most significant task in the estimation of CT images from MR images is to obtain an improved bone tissue estimation. Table 1 reveals that GMM* (GMM trained on each tissue region),
SGMM, and MRF have better prediction accuracy than HMM and GMM on the bone tissues. In addition, Table 2 shows that GMM* and SGMM have best estimation performance on dense bone tissues. Based on mean absolute error, HMM and GMM perform similarly on the bone tissues.

The skewness assumption has allowed to recognize skewness in the partitions of the data. The estimates of the skewness parameters demonstrate that the partitions have skewness property. Moreover, the skewness depends on the subtissue classes. The estimates of the parameters range from -1.54 to 3.03. However, Table 1–5 show that the skewness assumption did not improve the results as compared to the symmetric assumption in GMM*.

Figure 2 shows that the models perform better on soft tissues (neighbor to CT image intensity 0HU). On the other hand, the models have weaker prediction accuracy on the two extremes (air and bone tissues). This pattern of residual plots is observed in [10, 14]. Moreover, the figure shows that none of the models outperform throughout the tissues of the head. Table 1–4 demonstrate that the predictive accuracy of the models depend on the heads. That is the results are not uniform over the heads. This might have a problem in real applications and needs a further investigation.

In conclusion, the approach used in this study is an efficient way to get a good quality CT image substitute with improved estimation of bone tissues. Moreover, the SGMM and the developed algorithm to estimate its parameters is general and can be applied to other applications.

Acknowledgments

This work is supported by the Swedish Research Council grant (Reg. No. 340-2013-5342). The authors thank Adam Johansson for providing us data and David Bolin for providing us Mathlab code for MRF. Moreover, the authors thank Kristi Kuljus for HMM result.
References

[1] H. Arabi, N. Koutsouvelis, M. Rouaud, R. Miralbell, and H. Zaidi, *Atlas-guided generation of pseudo-CT images for MRI-only and hybrid PET-MRI-guided radiotherapy treatment planning*, Phys. Med. Biol. 61 (2016), pp. 6531–6552.

[2] A. Azzalini, *A class of distributions which includes the normal ones*, Scandinavian Journal of Statistics 12(1985), pp. 171–178.

[3] A. Azzalini, and A. D. Valle, *The multivariate skew-normal distribution*, Biometrika 83 (1996), pp. 715–726.

[4] T. Boettger, T. Nyholm, M. Karlsson, C. Nunna, and J. C. Celi, *Radiation therapy planning and simulation with magnetic resonance images*, Proc SPIE 6918 (2008). Available at [http://dx.doi.org/10.1117/12.770016](http://dx.doi.org/10.1117/12.770016).

[5] C. R. B. Cabral, V. H. Lachos, and M. O. Prates, *Multivariate mixture modeling using skew-normal independent distributions*, Computational Statistics and Data Analysis 56 (2012), pp. 126–142.

[6] L. Chen, R. A. Price, L. Wang, J. Li, L. Qin, S. McNeeley, C-M. Ma, G. M. Freedman, and A. Pollack, *MRI-based treatment planning for radiotherapy: domestic verification for prostate IMRT*, Int. J. Radiation Oncology, Biol., Phys. 67 (2004), pp. 636–647.

[7] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Maximum likelihood from incomplete data via the EM-algorithm*, Journal of Royal Statistical Society 39 (1977), pp. 1–38.

[8] J. Ding, and A. Zhou, *Eigenvalues of rank-one updated matrices with some applications*, Applied Mathematics Letters 20 (2007), pp. 1223–1226.

[9] T. Huynh, Y. Gao, J. Kang, L. Wang, P. Zhang, J. Lian, and D. Shen, *Estimating CT image from MRI data using structured random forest and auto-context model*, IEEE transactions on medical imaging 35 (2016), pp. 174–183.

[10] A. Johansson, M. Karlsson, and T. Nyholm, *CT substitute derived from MRI sequences with ultrashort echo time*, Medical Physics 38 (2011), pp. 2708–2714.

[11] A. Johansson, M. Karlsson, J. Yu, T. Asklund, and T. Nyholm, *Voxel-wise uncertainty in CT substitute derived from MRI*, Medical Physics 39 (2012), pp. 3283–3290.

[12] M. Karlsson, M. G. Karlsson, T. Nyholm, C. Amies, and B. Zackrisson, *Dedicated magnetic resonance imaging in the radiotherapy clinic*, Int. J. Radiation Oncology, Biol., Phys. 74 (2009), pp. 644–651.

[13] M. Khounsiavash, M. Ganjali, and T. Baghfalaki, *A Stochastic Version of the EM Algorithm to Analyze Multivariate Skew-Normal Data with Missing Responses*, Applications and Applied Mathematics 6 (2011), pp. 412-427.

[14] K. Kuljus, F. L. Bayisa, D. Bolin, J. Lember, and J. Yu, *Comparison of hidden Markov chain models and hidden Markov random field models in estimation of computed tomography images*, 2017. Available at [http://arxiv.org/abs/1705.01727](http://arxiv.org/abs/1705.01727).

[15] V. H. Lachos, H. Bolfarine, R. B. Arellano-Valle, and L. C. Montenegro, *Likelihood-Based Inference for Multivariate SkewNormal Regression Models*, Communications in Statistics -Theory and Methods 36 (2007), pp. 1769-1786.

[16] T. I. Lin, *Maximum likelihood estimation for multivariate skew normal mixture models*, Journal of Multivariate Analysis 100 (2009), pp. 257-265.

[17] T. I. Lin, J. C. Lee, and S. Y. Yen, *Finite mixture modelling using the skew normal distribution*, Statistica Sinica 17 (2007), pp. 909–927.

[18] D. Nie, X. Cao, Y. Gao, L. Wang, and D. Shen, *Estimating CT image from MRI data using 3D fully convolutional networks*, Proceedings 10008 LNCS (2016), pp. 170–178.

[19] J. Sherman, and W. J. Morrison, *Adjustment of an inverse matrix corresponding to changes in the elements of a given column or a given row of the original matrix*, Ann. Math. Statist. 20 (1949), pp. 621.

[20] C. M. Washington, and D. T. Leaver(ed.), *Principles and Practice of Radiation Therapy*, Elsevier, USA, 2015.

[21] K. M. Waterstram-Rich, and D. Gilmore(ed.), *Nuclear Medicine and PET/CT: Technology and Techniques*, Elsevier, USA, 2016.