Hidden Markov random field and FRAME modelling for TCA-image analysis

Katy Streso (streso@demogr.mpg.de)
Francesco Lagona (lagona@demogr.mpg.de)

HIDDEN MARKOV RANDOM FIELD AND FRAME MODELLING FOR TCA-IMAGE ANALYSIS

Katy Streso
StatLab, Office of Statistics and Information Services
Max Planck Institute for Demographic Research
Konrad - Zuse - Strasse 1
18057 Rostock, Germany
streso@demogr.mpg.de

Francesco Lagona
StatLab, Office of Statistics and Information Services
Max Planck Institute for Demographic Research
Konrad - Zuse - Strasse 1
18057 Rostock, Germany
lagona@demogr.mpg.de

ABSTRACT

Tooth Cementum Annulation (TCA) is an age estimation method carried out on thin cross sections of the root of human teeth. Age is computed by adding the tooth eruption age to the count of annual incremental lines that are called tooth rings and appear in the cementum band. Algorithms to denoise and segment the digital image of the tooth section are considered a crucial step towards computer-assisted TCA. The approach pursued in this paper relies on modelling the images as hidden Markov random fields, where gray values are assumed to be pixelwise conditionally independent and normally distributed, given a hidden random field of labels. These unknown labels have to be estimated to segment the image. To account for long-range dependence among the observed values and for periodicity in the placement of tooth rings, the Gibbsian label distribution is specified by a potential function that incorporates macro-features of the TCA-image (a FRAME model). Estimation of the model parameters is carried out by an EM-algorithm that exploits the mean field approximation of the label distribution. Segmentation is based on the predictive distribution of the labels given the observed gray values.

KEY WORDS

EM, FRAME, Gibbs distribution, (hidden) Markov random field, mean field approximation, TCA

1 Introduction

Tooth Cementum Annulation ([1]) is an age estimation method based on annual incremental appositions in the cementum of mammalian teeth. A 90-110 μm thick cross section, polished or unpolished, is photographed with a Leica DC350F camera system under bright-field and 200 or 400 times magnification. TCA-images are then 8 or 16 bit gray scale pictures of size 1030x1300 or 1016x1300 pixels. The dark parts of the annual lines, often called tooth rings, are empirically 1 to 3 μm thick and result roughly in 5 to 20 pixel thin lines under 400 times magnification.

Figure 1 displays a typical good quality TCA-image of the unpolished section extracted from a person aged 41. It is expected to find 34 horizontal tooth rings in the marked cementum band. Additionally, the image contains diago-nal saw cuts and artifacts (for example on the right). The marked rectangle delimits the area that is used for the application in Section 5.

Paleodemographers at the Max Planck Institute for Demographic Research use large databases of images like the one depicted in Figure 1 to identify mortality profiles of past human populations. Hence algorithms are needed to denoise and segment these images automatically.

Standard methods like singular value decomposition, Fourier transform and regression smoothing measure texture features and are for this reason not flexible enough to fulfill the above task. In the course of this paper, TCA-images are therefore described by a statistical model, specifically a Hidden Markov Random Field (HMRF) model. Section 2 introduces these models and the distribution of the hidden field is specified by a FRAME model ([2]). With this Markov random field (MRF), macro-features of TCA-images such as long-range auto-correlation among observed gray values and periodic placement of tooth rings can be modelled. Section 3 describes the estimation of the model parameters via an EM algorithm that exploits the mean field approximation of the hidden field distribution. Section 4 specifies the FRAME model for the application to TCA-images and describes the Gibbs sampler that is used to simulate from this prior distribution. The sensible results of fitting the hidden FRAME model to real images like the one depicted in Figure 1 by using the EM algorithm are discussed in Section 5.

2 The Hidden FRAME Model

HMRF modelling allows us to address both denoising and segmentation by means of a labelling problem ([3]). To illustrate, let $\mathcal{S} = \{1, \ldots, NM\}$ be the set of pixels forming a rectangular lattice of size $N \times M$. In the course of this paper a pixel will interchangeably be denoted by i, or (x, y) when the two dimensions of the lattice need to be emphasized. The observed image is represented by an array Y, where $Y_i \in \mathbb{R}$ is the gray value observed at pixel i. The value Y_i is assumed to be drawn from the ith continuous random variable $\mathcal{y}_i$, belonging to the random field $\mathcal{Y} = (\mathcal{y}_1, \ldots, \mathcal{y}_{NM})$. Analogously, we define the array λ of labels $\lambda_i \in \mathcal{G} = \{0, 1, \ldots, G\}$ that need to be estimated
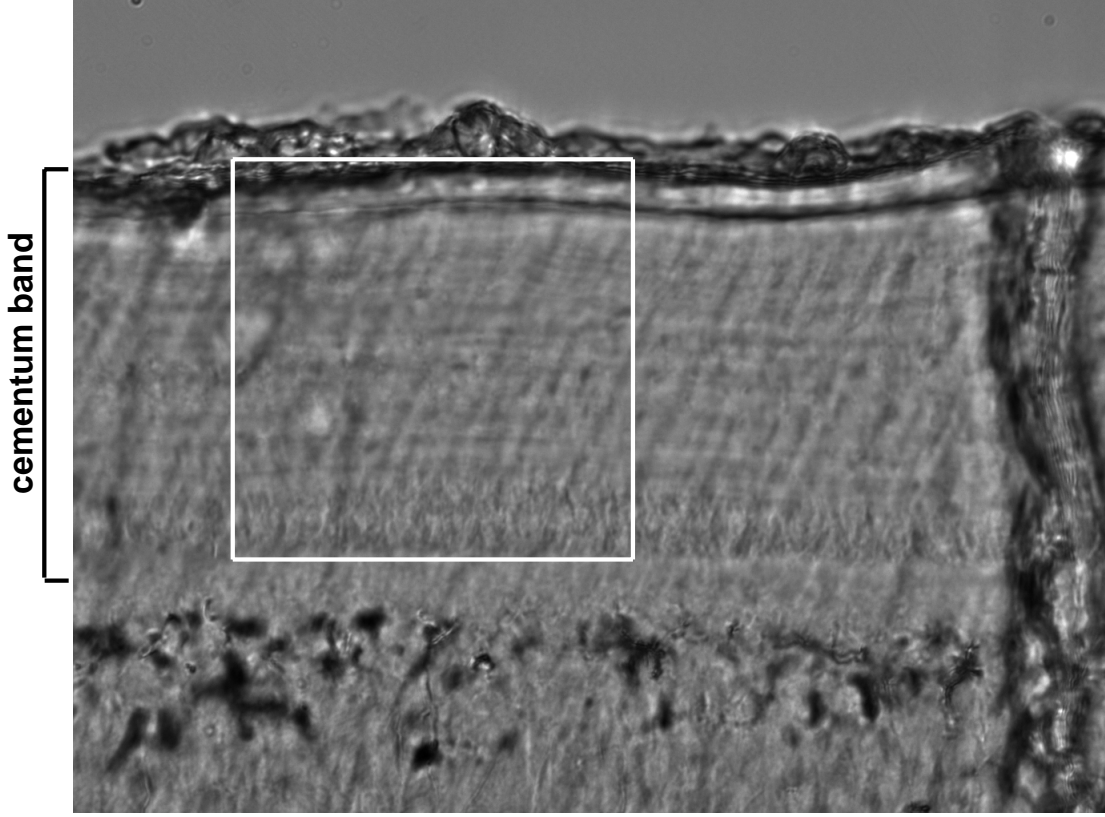
Figure 1. A typical unpolished TCA-image of good quality (IS-0000666 from the TCA database of the MPI DR)

at each pixel and assume that λ_i is sampled from the discrete random variable Λ_i, defined as the ith coordinate of the random field $\Lambda = (\Lambda_1, \ldots, \Lambda_{NM})$.

In the setting of a HMRF, the joint distribution of $\mathcal{Y}$ is modelled according to the mixture

$$f(Y) = \sum_{\lambda \in \mathcal{G}^{N \times M}} P(\lambda) \prod_{i \in \mathcal{S}} f(Y_i | \lambda_i)$$

where $P(\lambda)$ is the distribution of a Markov random field and $f(Y|\lambda)$ is called cost function or emission density. The choice of both depend on the application. For TCA-images we have chosen a Gaussian cost function

$$f(Y_i | \lambda_i, \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\sigma_{\lambda_i}} e^{-\frac{\left(Y_i - \mu_{\lambda_i}\right)^2}{2\sigma_{\lambda_i}^2}},$$

where the parameters $\boldsymbol{\theta} = \{\mu_g, \sigma_g^2 | g \in \mathcal{G}\}$ are unknown. The MRF $P(\lambda)$ may model spatial dependencies by means of specifying a neighborhood structure. More precisely, let's define

- a *neighboring relationship* as a binary relationship on the lattice $\mathcal{S}$ that is anti-reflexive and symmetric;

- the *neighborhood* $N(i)$ of pixel i as the set

$$N(i) = \{j \in \mathcal{S} | j \text{ neighbor of } i\}, \text{ and}$$

- the *neighborhood system* $\mathcal{N}$ as the set of all neighborhoods $\mathcal{N} = \{N(i) | i \in \mathcal{S}\}$.

Under this setting, the random field Λ is a *MRF* with respect to the neighborhood system $\mathcal{N}$ if for all $\lambda \in \mathcal{G}^{N \times M}$

1. $P(\lambda) > 0$ (positivity)

2. $P\left(\lambda_i | \lambda_{\mathcal{S} \setminus i}\right) = P(\lambda_i | \lambda_{N(i)})$ (Markovianity).

The specific form of the MRF model that shall be utilized for TCA-images is called FRAME, which stands for **F**ilters, **R**andom Fields **a**nd **M**aximum **E**ntropy and was mainly developed in [2], [4] and [5]. In the FRAME model, prior knowledge about the image is efficiently modelled by convolving the label image λ with suitable filters and by evaluating these filter responses. In its simplest version, the FRAME distribution is a Gibbs distribution

$$P(\lambda) = \frac{1}{Z} e^{\sum_{i \in \mathcal{S}} \phi[(F_T * \lambda)(i)]}, \tag{1}$$

where Z is the normalizing constant. The energy function involves one filter F_T that is known up to the parameter T. The filter responses $(F_T * \lambda)(i)$ to λ at pixels i are evaluated pixelwise by the potential function ϕ. The choice of the parametric family F_T and the function ϕ is driven by the application (Section 4). The hidden FRAME model hereby elegantly combines two important areas of texture analysis: HMRF modelling and filtering theory and it can be applied to a wide variety of even large scale textures.

3 Parameter Estimation and Segmentation

In order to estimate $\boldsymbol{\theta}$ and T, the maximum likelihood estimates (MLE) $\hat{\boldsymbol{\theta}}$ and $\hat{T}$ can in principle be found by maximizing the likelihood function

$$L(\boldsymbol{\theta}, T|Y) = \sum_{\lambda \in \mathcal{G}^{N \times M}} P(\lambda|T) \prod_{i \in \mathcal{S}} f(Y_i|\lambda_i, \boldsymbol{\theta}). \quad (2)$$

However, this maximization is intractable because of the size of the label space $\mathcal{G}^{N \times M}$.

The EM algorithm is a widely used technique to solve this kind of problem. The algorithm depends on the predictive probability that is usually computed via MCMC. In our application this is again not feasible because of the size of TCA-images. We suggest to use mean field approximation to make the EM tractable.

To illustrate, let us recall that the EM algorithm starts with preliminary estimates $\boldsymbol{\theta}^{(0)}$ and $T^{(0)}$ of the parameters $\boldsymbol{\theta}$ and T, and then proceeds iteratively by alternating two steps. In the E-step of the t-th iteration the conditional expectation of the complete log-likelihood, with respect to the unknown labels λ

$$E\left[\log P(Y, \lambda|\boldsymbol{\theta}, T)|Y, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right]$$
$$= \int_{\lambda \in \mathcal{G}^{N \times M}} P\left(\lambda|Y, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right) \log P(\lambda, Y|\boldsymbol{\theta}, T) d\lambda$$

is calculated, where $\boldsymbol{\theta}^{(t-1)}$ and $T^{(t-1)}$ are the estimates from the previous iteration. The M-step of the EM algorithm maximizes this expectation to update $\boldsymbol{\theta}$ and T:

$$\left(\boldsymbol{\theta}^{(t)}, T^{(t)}\right) = \arg\max_{\{\boldsymbol{\theta}, T\}} E\left[\log P(Y, \lambda|\boldsymbol{\theta}, T)|Y, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right]$$

Since each iteration is guaranteed to increase the (incomplete) log-likelihood (2) under mild assumptions, the EM algorithm will converge to a local maximum ([6]).

In the case of a Gaussian random field, the EM algorithm reduces to the three updating formulas ([6])

$$\mu_g^{(t)} = \frac{\sum_{i \in \mathcal{S}} Y_i P\left(\lambda_i = g|Y, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)}{\sum_{i \in \mathcal{S}} P\left(\lambda_i = g|Y, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)} \quad (3)$$

$$\left(\sigma_g^{(t)}\right)^2 = \frac{\sum_{i \in \mathcal{S}} \left(Y_i - \mu_g^{(t)}\right)^2 P\left(\lambda_i = g|Y, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)}{\sum_{i \in \mathcal{S}} P\left(\lambda_i = g|Y, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)} \quad (4)$$

$$T^{(t)} = \arg\max_{\{T\}} \sum_{i \in \mathcal{S}} \sum_{g=0}^{G} P\left(\lambda_i = g|Y, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)$$
$$\cdot \log P\left(\lambda_i = g|\lambda_{N(i)}, T\right). \quad (5)$$

The conditional probabilities $P\left(\lambda_i = g|Y, \lambda_{N(i)}, \boldsymbol{\theta}, T\right)$ are not available in closed form and could be evaluated by an MCMC algorithm ([7]). This would require to generate a Markov chain at each pixel which is not feasible. The alternative approach we suggest is based on the approximation

$$P(\lambda) \approx \prod_{i \in \mathcal{S}} P\left(\lambda_i|\tilde{\lambda}_{N(i)}\right). \quad (6)$$

In this paper, the configuration $\tilde{\lambda}$ is chosen according to the theory of mean field approximation theory ([8]) where $\tilde{\lambda}$ is set to the expected values of the label image:

$$\tilde{\lambda}_j = E[\lambda_j] \text{ for all } j \in N(i).$$

The product $\prod_{i \in \mathcal{S}} P(\lambda_i|E[\lambda_{N(i)}])$ is then a valid probability distribution and minimizes the Kullback-Leibler divergence to the true prior distribution $P(\lambda)$ among all products of this kind. The E-step of the EM algorithm hereby changes to

$$E\left[\log P(\lambda, Y|\boldsymbol{\theta}, T)|Y, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right]$$
$$\approx \sum_{i \in \mathcal{S}} \sum_{g=0}^{G} \log\left(P\left(\lambda_i = g|\tilde{\lambda}_{N(i)}, \boldsymbol{\theta}, T\right) f(Y_i|\lambda_i, \boldsymbol{\theta}, T)\right)$$
$$\cdot P\left(\lambda_i = g|Y_i, \tilde{\lambda}_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)$$
$$= \sum_{i \in \mathcal{S}} \sum_{g=0}^{G} \left(\log P\left(\lambda_i = g|\tilde{\lambda}_{N(i)}, T\right) + \log f\left(Y_i|\lambda_i = g, \boldsymbol{\theta}\right)\right)$$
$$\cdot \frac{P\left(\lambda_i = g|\tilde{\lambda}_{N(i)}, T^{(t-1)}\right) f\left(Y_i|\lambda_i = g, \boldsymbol{\theta}^{(t-1)}\right)}{\sum_{g=0}^{G} P\left(\lambda_i = g|\tilde{\lambda}_{N(i)}, T^{(t-1)}\right) f\left(Y_i|\lambda_i = g, \boldsymbol{\theta}^{(t-1)}\right)}$$

The parameter estimates can therefore be updated by the Equations (3) to (5) and replacing $\lambda_{N(i)}$ therein by $\tilde{\lambda}_{N(i)}$, that are computed iteratively. Our EM algorithm then takes the following form:

EM algorithm using MFA for fitting a hidden FRAME model

---

1. input TCA-image Y
Initialization
2. initialize label configuration $\lambda^{(0)}$ by thresholding
3. initialize parameters
Updating
4. **for** $t = 1 : t_{max}$
   update label image $\lambda^{(t)}$ by
5. $\langle \lambda \rangle = \lambda^{(t-1)}$
6.   **for** each site i (randomly permuted)
7.   **for** $g = 0 : G$
8.   calculate the conditional probability
$$f\left(Y_i|\lambda_i = g, \mu^{(t-1)}, \sigma^{(t-1)}\right) \propto e^{-\frac{\left(Y_i - \mu_g^{(t-1)}\right)^2}{2\left(\sigma_g^{(t-1)}\right)^2}}$$
9.   approximate the prior energy and probability
$$U\left(\lambda_i = g|\lambda_{N(i)}, T^{(t-1)}\right) \approx \sum_{j \in C(i)} \phi\left(\left(F_{T_j^{(t-1)}} * \lambda\right)(j)\right)$$
$$P\left(\lambda_i = g|\lambda_{N(i)}, T^{(t-1)}\right) \approx \frac{e^{U\left(\lambda_i = g|\lambda_{N(i)}, T^{(t-1)}\right)}}{\sum_{g=0}^{G} e^{U\left(\lambda_i = g|\lambda_{N(i)}, T^{(t-1)}\right)}}$$
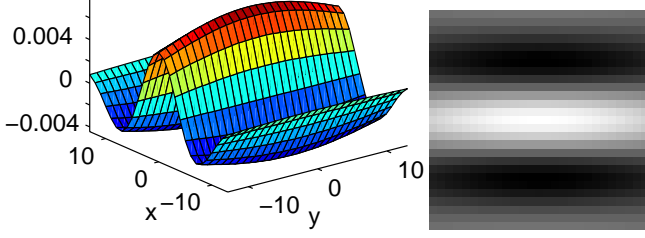
Figure 2. 3-D surface and image of a Gaborcosine function with $T = 16$ and $\alpha = 0$

Figure 3. A typical image of size 128x128 pixels simulated by the Gibbs sampler using the FRAME model (1) with the filter displayed in Figure 2, $\phi = |.|$ and 8 gray levels.

10.   calculate the posterior probability

$$P\left(\lambda_i = g | Y_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)$$
$$\propto f\left(Y_i | \lambda_i = g, \mu^{(t-1)}, \sigma^{(t-1)}\right) P\left(\lambda_i = g | \lambda_{N(i)}, T^{(t-1)}\right)$$

11.   calculate the expected label

$$\langle \lambda_i \rangle = \frac{\sum_{g=0}^{G} g \cdot P\left(\lambda_i = g | Y_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)}{\sum_{g=0}^{G} P\left(\lambda_i = g | Y_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right)}$$

12.   set $\lambda^{(t)} = \langle \lambda \rangle$

     update parameters

13.   **for** $g = 0 : G$
14.   update μ_g according to Equation (3)
15.   update $\left(\sigma_g^{(t)}\right)^2$ according to Equation (4)
16.   update T according to Equation (5)

The initialization of λ and the sequential updating of the true labels were chosen according to the recommendations in [8]. The number of iterations t_{max} was chosen according to the last gain in the likelihood

$$L(\boldsymbol{\theta}^{(t-1)} | Y) = \sum_{i \in \mathcal{S}} \sum_{g=0}^{G} P\left(\lambda_i = g | Y_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right).$$

Segmentation can finally be carried out by exploiting $P\left(\lambda_i = g | Y_i, \lambda_{N(i)}, \hat{\boldsymbol{\theta}}, \hat{T}\right)$ by means of thresholding.

4 Application

This Section is devoted to specifying the filter family F_T and the potential function ϕ that we have used for TCA-image analysis and to describe a simulation algorithm for generating a typical image from this model.

Filtering theory is well recognized in texture analysis at least since [9]. Marčelja ([10]) has shown that two-dimensional Gabor functions closely conform to the receptive field profiles of simple cells in the striate cortex.

We define the filter $F_{\alpha,T}$ on the basis of the real valued, even-symmetric Gabor function:

$$Gcos_{T,\alpha}(x, y) = c \cdot e^{\frac{-\left(rx'^2 + y'^2\right)}{2T^2}} \cos\left(\frac{2\pi}{T} x'\right); \quad (7)$$

with $x' = x\cos\alpha + y\sin\alpha$, $y' = -x\sin\alpha + y\cos\alpha$, $r = 4$ being the aspect ratio and c being a normalizing factor. The Gaborcosine function above is an elongated Gaussian bell multiplied by a cosine wave, where parameter T changes the wavelength and α determines the orientation of the cosine wave. For example, Figure 2 shows the Gaborcosine function for $T = 16$, $\alpha = 0$ and $x, y \in [-13, 13]$. This filter can capture waves or lines of width 16 and orientation $0°$.

In the application for TCA-images we fix $\alpha = 0$ which is the main direction of tooth rings. In order to cover the range of possible tooth ring widths we chose $T \in \{2, 4, 6, 8, 10, 12, 14, 16, 18\}$. We remark that our approach is different to that in [2], because we are interested in reconstructing tooth rings, that resemble only one feature of interest. We do not want to synthesize perceptional equivalent images, including noise. Besides simplifying the FRAME model to incorporate only one filter (one feature), the potential function ϕ that evaluates this filter response is assumed to be known and chosen to be the simplest among the upright curves, namely the absolute value $\phi = |.|$.

Figure 3 displays a typical image drawn from the FRAME model using the Gaborcosine filter with parameters $T = 16$ and $\alpha = 0$ and the absolute valued potential function. This image comes very close to the ideal TCA-image that one could have in mind about parallel running tooth rings. Orientation and width of these lines are determined by both parameters of the Gaborcosine filter.

The image in Figure 3 was generated by Gibbs sampling. (See e.g. [3].) The single site Gibbs sampler for example initializes $\lambda^{(0)}$ at time $t = 0$ and then updates each pixel by repeatedly sampling a candidate $\lambda_i^{(t+1)}$ from the full conditional $P\left(\lambda_i^{(t+1)} | \lambda_{\mathcal{S}\backslash i}^{(t)}\right)$. The transition probabilities $P(\lambda^{(t)} | \lambda^{(t-1)})$ are then guaranteed to converge to the stationary distribution $P(\lambda)$.

We remark that the Gibbs sampler can be applied in the present case because the FRAME model is a MRF model. This can be proven by the application of the Hammersley-Clifford theorem ([11]).

When choosing a random initial image and a random updating order of the pixels, the Gibbs sampler consists of
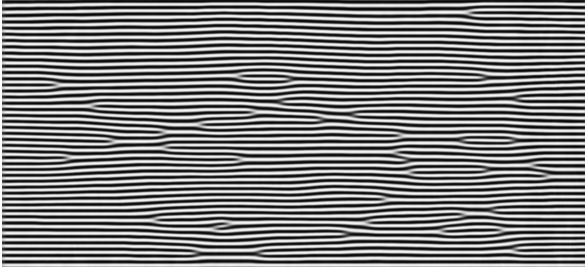
Figure 4. The mean field approximation of the cementum band of TCA-image 1

Figure 5. The black rings from the mean field approximation of part of TCA-image 1 overlayed onto the original

the following steps for the FRAME model:

Gibbs sampling algorithm for the FRAME model

---

1. input initial white noise image $\lambda^{(0)}$ and filter F
2. precompute the filter response $F * \lambda^{(0)}$
3. **repeat** sufficiently often
4.   **repeat** $N \cdot M$ times ($N \cdot M = |\mathcal{S}|$ size of the image)
5.   randomly select site (x, y)
6.   **for** all $(x', y') \neq (x, y)$
7.   set $\lambda^{(t+1)}_{(x',y')} = \lambda^{(t)}_{(x',y')}$
8.   **for** each gray value g of label $\lambda^{(t+1)}_{(x,y)}$
9.   **for** all $(x', y') \in \{N(x, y), (x, y)\}$
10.   calculate the new filter responses
$$\left(F * \lambda^{(t+1)}\right)(x', y') = \left(F * \lambda^{(t)}\right)(x', y')$$
$$+ F(x - x', y - y')\left(\lambda^{(t+1)}_{(x,y)} - \lambda^{(t)}_{(x,y)}\right)$$
11.   **for** each gray value g of label $\lambda^{(t+1)}_{(x,y)}$
12.   set $\lambda^{(t+1)}_{(x,y)} = g$ with (conditional) probability
$$P\left(\lambda^{(t+1)}_{(x,y)} = g | \lambda^{(t)}_{N(x,y)}\right) = \frac{e^{\left(\sum_{(x',y') \in N(x,y)} \left|\left(F * \lambda^{(t+1)}\right)(x',y')\right|\right)}}{\sum_{g=0}^{G} e^{\left(\sum_{(x',y') \in N(x,y)} \left|\left(F * \lambda^{(t+1)}\right)(x',y')\right|\right)}}$$
13.   update the filter response $F * \lambda^{(t+1)}$

If the computer precision is not enough to calculate the conditional probability $P\left(\lambda^{(t+1)}_{(x,y)} = g | \lambda^{(t)}_{N(x,y)}\right)$ in step 12, one can easily insert a nourishing one.

To detect convergence, the Gelman-Rubin multivariate convergence statistic R ([12]) is used on every (20x20)th pixel of the image. The Gibbs sampler stops iterating when $R < 1.2$. The algorithm above needs about $O(|F| \cdot NM \cdot G \cdot S)$ operations where $|F|$ is the area covered by the filter and S is the number of sweeps of the Gibbs sampler.

5 Results

The aim of analysis of TCA-images in this paper is to uncover the black and white labelling ($\mathcal{G} = \{0, 1\}$) in order to be able to estimate the number of tooth rings. For thi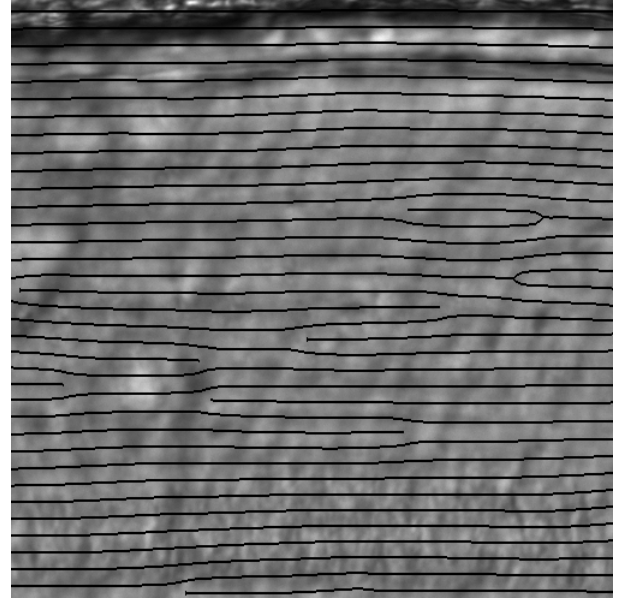s purpose a Gaussian hidden Markov random field is fitted to the TCA-image in Figure 1. The MRF model is specified by the FRAME model (1). The parameters μ_0, μ_1 and a common variance σ^2 as well as the filter parameter T are estimated by an EM algorithm as stated in Section 2. The label image λ is obtained from the mean field at the last iteration.

Figure 4 shows the predictive probability $P(\lambda_i | \tilde{\lambda}_{N(i)}, \hat{\boldsymbol{\theta}}, \hat{T})$ of the pixels in the cementum band of Figure 1, where $\hat{\boldsymbol{\theta}}$ and $\hat{T}$ are the estimates of the last iteration. The parameter estimates are the means $\hat{\mu}_0 = 28444$, $\hat{\mu}_1 = 28714$, the common variance $\hat{\sigma}^2 = 5.5 \cdot 10^7$ and the ring width $\hat{T} = 14$.

For illustration purposes a smaller part (the one marked in Figure 1) of this mean field is thresholded ($\lambda_i = 0$ if $P(\lambda_i | \tilde{\lambda}_{N(i)}, \hat{\boldsymbol{\theta}}, \hat{T}) < 0.5$ and $\lambda_i = 1$ otherwise). The middle lines of the black rings are then superimposed onto the original image. The reader can count approximately 32 tooth rings in Figure 5. From the known age we expect 34 tooth rings in the image presented in Figure 1. This is quite a good estimate that is typically confirmed by a dozen additional TCA-images.

6 Conclusion

For segmentation of TCA-images we set up a hidden Markov random field model and exploited the EM-algorithm. This procedure requires the approximation of the posterior probabilities $P\left(\lambda_i = g | Y, \lambda_{N(i)}, \boldsymbol{\theta}, T\right)$ and the final segmentation λ. The Gibbs sampler proved to be infeasible in both cases except for small images. For example the simulation of the predictive distribution in Figure 3 took about 55 hours on a PC and programmed in Matlab.

We therefore chose to use the mean field approximation to estimate the posterior probabilities and thresholded the mean field of the last iteration for the final segmentation. This compound estimation procedure took all together 10 hours and gave reasonable results.

Despite of the good overall age estimate, the reader can see in Figure 5 that some rings are not well met and that bifurcations occur in the label image (Figure 4). This is due to two reasons. On the one side the reconstruction of the TCA-image is heavily influenced by the shape of the single filter we estimate. The hidden FRAME model in this form can hereby only take into account strong local changes of tooth rings. In order to overcome this global property of the FRAME model, one would need to select location dependent filters, i.e. estimating the filter parameter T at each pixel i. On the other side, we assumed that the orientation of tooth rings is mainly horizontal. By estimating not only the ring width T from the bank of filters, but also the orientation α, one could overcome this limitation and would therefore avoid the bifurcations, that now mainly occur in areas where tooth rings have another orientation.

Additionally different variance parameters $\sigma_0^2 \neq \sigma_1^2$ might also change results and therefore such heteroscedasticity assumption should to be tested. The mean field approximation is not the only possible one for the approximation (6). Celeux, Forbes and Peyrard ([8]) also mention mode field approximation and simulated field approximation that should be tested for quality and speed in the case of TCA-image analysis. Moreover, a larger number of experiments on images of different quality need to be implemented in order to test the accuracy of the procedure.

References

[1] Robert D. Hoppa and James W. Vaupel, editors. *Paleodemography: Age Distributions from Skeletal Samples*. Cambridge University Press, Cambridge, 2002.

[2] Song Chun Zhu, Ying Nian Wu, and David B. Mumford. Minimax Entropy Principle and Its Application to Texture Modeling. *Neural Computation*, 9(8):1627–1660, November 1997.

[3] Stan Z. Li. *Markov Random Field Modeling in Image Analysis*. Computer Science Workbench. Springer, Tokyo [et al.], 2001.

[4] Song Chun Zhu, Yingnian Wu, and David B. Mumford. Filters, Random Fields and Maximum Entropy (FRAME): Towards a Unified Theory for Texture Modeling. *International Journal of Computer Vision Archive*, 27(2):107 – 126, March 1998.

[5] Song Chun Zhu and David B. Mumford. Prior Learning and Gibbs Reaction-Diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(11):1236–1250, November 1997.

[6] Geoffrey J. McLachlan and Thriyambakam Krishnan. *The EM Algorithm and Extensions*. Wiley Series in Probability and Statistics. John Wiley & Sons, New York, 1997.

[7] Yongyue Zhang, Michael Brady, and Stephen Smith. Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57, January 2001.

[8] Gilles Celeux, Florence Forbes, and Nathalie Peyrard. EM procedures using mean field-like approximations for Markov model-based image segmentation. *Pattern Recognition*, 36(1):131–144, January 2003.

[9] Anil K. Jain and Farshid Farrokhnia. Unsupervised Texture Segmentation Using Gabor Filters. *Pattern Recognition*, 24(12):1167–1186, 1991.

[10] S. Marčelja. Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America*, 70(11):297–300, November 1980.

[11] Julian Besag. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):192–236, 1974.

[12] Stephen P. Brooks and Andrew Gelman. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, 1998.