Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 52 (2007) 855–868
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Hidden Markov random field models for TCA image analysis

Katy Klauenberg[a], Francesco Lagona[b,*]

[a] *University of Sheffield, UK*
[b] *Dipartimento di Istituzioni Pubbliche, Economia e Società, Università Roma Tre, Via G. Chiabrera, 199, 00145 Rome, Italy*

Available online 3 July 2007

Abstract

Tooth Cementum Annulation (TCA) is an age estimation method carried out on thin cross sections of the root of mammalian teeth. Age is computed by adding the tooth eruption age to the count of annual incremental lines which are called tooth rings and appear in the cementum band. The number of rings is computed from an intensity (gray scale) image of the cementum band, by estimating the average ring width and then dividing the area of the cementum band by this estimate. The ring width is estimated by modelling the image by a hidden Markov random field, where intensities are assumed to be pixelwise conditionally independent and normally distributed, given a Markov random field of hidden binary labels, representing the"true scene". To incorporate image macro-features (the long-range dependence among intensities and the quasi-periodicity in the placement of tooth rings), the label random field is defined by an energy function that depends on a parametric Gabor filter, convolved with the true scene. The filter parameter represents the unknown of main interest, i.e. the average width of the rings. The model is estimated through an *EM* algorithm, relying on the mean field approximation of the hidden label distribution and allows to predict the locations of the rings in the image.

Keywords: Coupled Markov chain; *EM* algorithm; Gabor filter; Hidden Markov random field; Mean field approximation; TCA

1. Introduction

Tooth Cementum Annulation (TCA; Crowe and Strickland, 1975; Naylor et al., 1985; Renz et al., 1997) is an age estimation method based on annual incremental appositions in the cementum of mammalian teeth. Age is computed by adding the tooth eruption age to the count of annual incremental lines which are called "tooth rings" and appear in the cementum band. TCA is often implemented on human teeth, for forensic (Stott et al., 1982; Kagerer and Grupe, 2001) and paleo-demographic applications (Hoppa and Vaupel, 2002). For example, the TCA images that motivated this work are currently studied by anthropologists and demographers at the Max Planck Institute for Demographic Research in Rostock (Germany), to identify mortality profiles of past human populations.

Validation studies on human teeth (Hoppa and Vaupel, 2002; Wittwer-Backofen et al., 2003) have shown that TCA provides better results than traditional techniques, such as those based on pubic symphysis, auricular surface and cranial sutures. Despite of TCA popularity, however, ring counting is still carried out manually, by visual inspection of a TCA image such as that one displayed in Fig. 1. This picture belongs to a large data base containing 90–110 μm thick cross sections of tooth roots, photographed using a Leica DC350F camera system with bright-field and 200 or 400 times

* Corresponding author. Tel.: +39 06 5408 5321; fax: +39 06 5408 5320.
E-mail address: lagona@uniroma3.it (F. Lagona).

0167-9473/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2007.06.022

Fig. 1. The horizontal incremental lines appearing in the cementum band of a tooth-root cross section. This 16-bit grey scale picture has been taken by photographing a 90 μm thick tooth-root cross section with 400 times magnification (source: TCA database of the Max Planck Institute for Demographic Research).

magnification. These images are then 8 or 16 bit gray scale pictures of 1030×1300 or 1016×1300 pixels in size. Each annual ring is given by the juxtaposition of a horizontal low mineralized bright layer and a horizontal higher mineralized dark layer. TCA analysis is based on the count of the dark layers which in effect represent one year of life. In a ring, the dark layer is usually 1–3 μm thick and roughly results in a thin line of 5–20 pixels, with 400 times magnification.

In the case of Fig. 1, we know the age of the person to whom the tooth belongs (41) and, as a result, we expect to find 34 horizontal rings in the displayed cementum band (the tooth under study is a permanent first molar and its average eruption age is 7). Although Fig. 1 displays a TCA image of intermediate quality, ring counting by visual inspection is time consuming and potentially subjective. A double-blind validation study based on two expert observers inspecting 400 TCA images (Klauenberg, 2007) reported a serious bias between the expected number of rings and the manual estimates, emphasizing a high level of subjectivity in manual ring detection.

To improve on manual ring counting, a standard approach would rely on enhancing image quality via a computer-assisted methodology, to help the observer with ring detection. Depending on the conceptualization of a TCA image, several approaches could be pursued to restore the image, but no one of them is easy to implement. For example, ring counting could be facilitated by denoising and deblurring pixel intensities. However, as Fig. 1 shows, blurring and noise are difficult to model in a TCA image, because the observed intensities are affected by several disturbing factors arising from the biological processes of TCA (anomalies in ring formation or tooth eruption), from the skeleton being buried, from the preparation procedure (the slicing position and angle, which produce the diagonal saw cuts, apparent in Fig. 1), from microscopy (the focus level) or from the imaging procedure (the brightness level). To some extent, these disturbing effects can be removed by standard techniques of image analysis, such as image filtering, singular value decomposition of the intensity matrix or Fourier image analysis (Klauenberg, 2007; Czermak et al., 2006). However, even after the successful removal of these effects, the selection of a consistent number of significant dark layers is not obvious, because of the random variation of the mineralization process and the resulting random intensity pattern within each candidate layer. The variation of the mineralization process challenges approaches based on classification ideas, too. In TCA image analysis, a classification approach would aim at assigning binary labels to pixels, partitioning the image into dark and bright layers. Pixelwise, unsupervised classification methods based on pixel intensities are difficult to implement because the distributions of intensities of any two layers may differ dramatically. For instance, Fig. 2 (top-right) displays the bimodal histogram of the intensities observed in a small portion of a TCA image, suggesting the

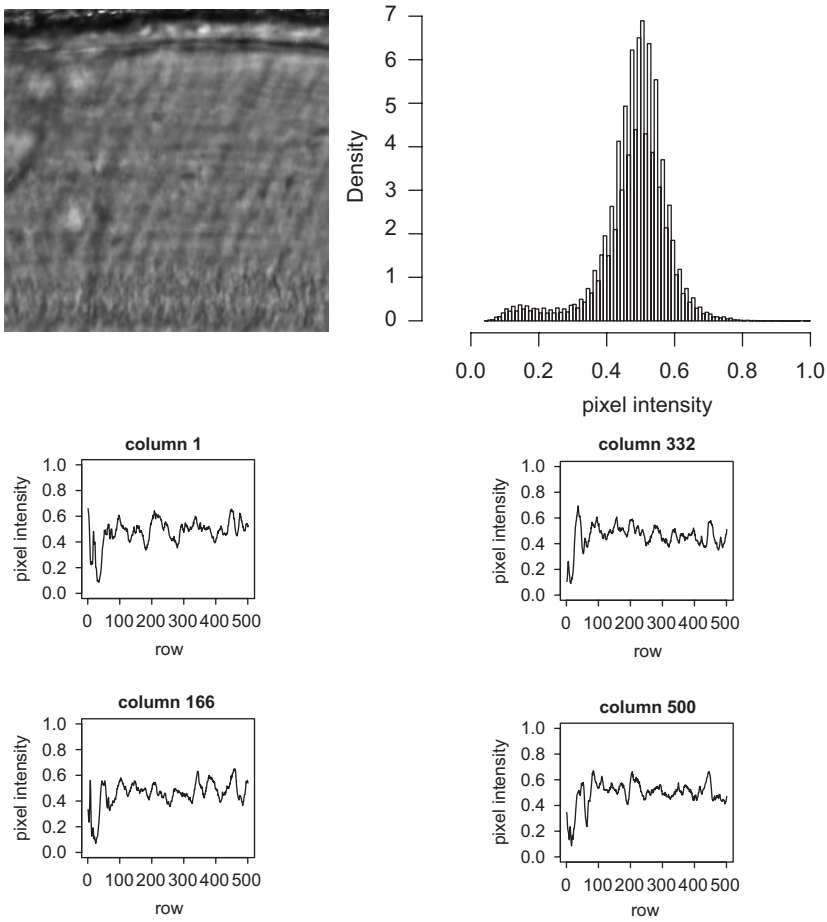

Fig. 2. Top: the 501 × 501 area highlighted in Fig. 1 (left) and (right) the relating histogram of pixel (16 bit) intensities (scaled to the interval (0, 1)); bottom: intensity profiles of four selected columns extracted from the area displayed at the top of the picture.

mixture of two distributions representing dark and bright layers. Although a classification on the basis of this histogram is tempting, it would be affected by a serious pixel misclassification, because many pixels belonging to dark (bright) layers contribute to the right-hand (left-hand) tail of the histogram. Alternatively, image labelling could be carried out via a supervised classification methodology, by comparing image regions to a bank of typical textures. Unfortunately, little is known about textures that typically feature the juxtaposition of bright and dark layers in a TCA image and, as a result, supervised classification methods lack of a suitable bank of textures. Alternative approaches for improving the quality of a TCA image could rely on segmentation methods, seeking to locate the boundaries between dark and bright layers. However, gray intensities change smoothly and often unpredictably between adjacent layers (Fig. 2, bottom) and edge detection is challenging.

Since in computer-assisted TCA ring counting is more important than ring detection, image analysis could seek to locate only the centroids of the dark layers, rather than trying to restore the shape of each layer. Czermak et al. (Czermak et al., 2006) pursue this approach, assuming that, ideally, each column of a TCA image should be a sequence of alternating grey levels representing the dark and bright layers of the rings. Accordingly, the relative minima (maxima) of each column should correspond to the centroids of the dark (bright) layers and the number of rings could be estimated by the average of the distribution of the relative minima within the columns. In their work, a pixel is referred to as a minimum point when not only its grey value is smaller than a pre-set maximum level, but also when at least one grey value of the two adjacent pixels above this as well as one gray value of the two points below are greater than the grey value of the pixel value scanned. When exploited on the image displayed in Fig. 1, their procedure selects

21 rings, seriously underestimating the expected number (34). We believe that such a bias is motivated by two issues. First, layers centroid detection is challenging, (1) because of the large variability of the grey levels taken by extreme intensities within columns (rings do not arise at similar gray levels), and (2) because of the mismatch between locations of extreme intensities at different columns (rings are not perfectly horizontal). Second, estimation of the number of rings on the basis of the distribution of the relative minima relies on the erroneous assumption that the columns of a TCA image are independent replications of the same process, i.e. the spatial dependence structure resulting from image macro-features is not taken into account.

To improve on these drawbacks, we suggest (1) to focus on the estimation of the average ring width, rather than trying to locate ring centroids, and (2) to constrain this estimate to account for the spatial autocorrelation between pixel intensities. Dividing the area of the cementum band by the estimated average ring width provides us with an estimate of the number of rings in the TCA image.

More precisely, we model the image by a hidden Markov random field (MRF) model. These models are popular in procedures of image analysis that operate at the pixel level, after the seminal work by Besag (1986). Hidden MRF models usually consist of a degradation model of the conditional distribution of the image intensities, given an unknown "true scene", in conjunction with a MRF model which reflects known features of the true scene. Although the degradation model should allow blur and noise, we consider a simple noisy process, by assuming that the intensity observed at each pixel is drawn from the mixture of two Gaussian distributions, representing the emission in a dark and a bright layer, respectively. The two mixing weights specify the pixelwise marginal distribution of a parametric MRF, which generates rectangular arrays of binary labels ("true" scenes). To incorporate image macro-features (the long-range dependence among intensities and the quasi-periodicity in the placement of tooth rings), the MRF distribution is specified by an energy function that depends on the convolution of a parametric Gabor filter (Marčelja, 1980; Jain and Farrokhnia, 1991) with the true scene. Filter-based specifications of MRFs are often referred to as FRAME models (Zhu et al., 1998), and have been introduced in the literature for the purpose of texture modelling. In our application, the filter depends on a parameter, T, that represents the unknown average ring width, i.e., the average distance between the centroids of two consecutive dark layers. The resulting energy function of the true scene punishes configurations with dark layers placed at a distance not equal to T, while, simultaneously, the matching of the centroids' positions in adjacent columns is enhanced.

In summary, the model is a Gaussian Hidden Markov Field where Gaussian pixel intensities are conditionally independent, given a label configuration drawn from a filter-based MRF model. Unlike most of the applications of hidden Markov modelling in image analysis, where the parameters of the MRF model are nuisance or *a priori* known, in our case the parameter T is the unknown of main interest, whereas the parameters of the Gaussian distribution are nuisance unknowns, although they can be exploited to predict ring locations in the image, for validation purposes.

The rest of the paper is organized as follows. The hidden Markov model with a filter-based MRF specification is illustrated in Section 2, while Section 3 is devoted to the specification of the parametric Gabor filter we suggest for TCA image analysis. Section 4 describes the estimation of the model parameters via an *EM* algorithm that exploits the mean field approximation of the hidden field distribution. A causal variant of our hidden Markov model is illustrated in Section 5. Results from the analysis on TCA images such as that displayed in Fig. 1 are discussed in Section 6. We discuss our findings in Section 7.

2. A hidden FRAME model

Let $S = \{1, \ldots, NM\}$ 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 $i \in S$, or (x, y), $x = 1 \ldots N$, $y = 1 \ldots M$, when the two dimensions of the lattice need to be emphasized (x and y being the generic row and column of the TCA intensity matrix, such as that one displayed in Fig. 1). We define an array $\boldsymbol{\lambda} = (\lambda_i, i \in S) = (\lambda_{xy}, x = 1 \ldots N, y = 1 \ldots M)$ of binary labels and assume that λ_i is sampled from a Bernoulli random variable Λ_i, which is the ith coordinate of a random field $\Lambda = (\Lambda_1, \ldots, \Lambda_{NM})$, with distribution $P(\boldsymbol{\lambda})$. The observed image is represented by an array $\mathbf{z} = (z_i, i \in S) = (z_{xy}, x = 1 \ldots N, y = 1 \ldots M)$, where z_i is the (gray scale) intensity observed at pixel i. Further, let $f(z_i|\lambda_i)$ be the conditional probability density of intensity z_i, given the label λ_i. We model image $\mathbf{z}$ according to the mixture

$$f(\mathbf{z}) = \sum_{\lambda \in \{0,1\}^{NM}} P(\boldsymbol{\lambda}) \prod_{i \in S} f(z_i|\lambda_i), \tag{1}$$

i.e., we assume that intensities are conditionally independent given the label configuration. When $P(\boldsymbol{\lambda})$ is a MRF, Eq. (1) is a hidden Markov random field. The choice of both the intensity and the label distributions depends on the application. For TCA images, we have chosen a Gaussian emission function

$$f(z_i|\lambda_i) = N(\mu_\lambda, \sigma^2_\lambda), \tag{2}$$

where parameters $\boldsymbol{\theta} = \{\mu_0, \mu_1, \sigma^2_0, \sigma^2_1\}$ are unknown, and a filter-based MRF specification, as follows.

We exploit a parametric filter F_T with a fixed rectangular support of width $w = w_x \times w_y$ and known up to a parameter T (to be estimated), i.e. filter F_T is specified by a $w_x \times w_y$ rectangular matrix of weights $F_T(x, y)$. The filter response to $\boldsymbol{\lambda}$ at each pixel i is given by

$$[F_T * \boldsymbol{\lambda}](i) = \sum_{(x,y)\in S} F^{(i)}_T(x, y)\lambda_{xy},$$

where $F^{(i)}_T(x, y)$ is F_T centered at pixel i, i.e.

$$F^{(i)}_T(x, y) = \begin{cases} F_T(x, y) & \text{if } |x - x_i| \leqslant w_x/2, |y - y_i| \leqslant w_y/2, \\ 0 & \text{otherwise} \end{cases}$$

and (x_i, y_i) are the coordinates of pixel i. Under this setting, we model the label configuration as follows:

$$\begin{aligned} P(\boldsymbol{\lambda}) \propto \exp\, Q(\boldsymbol{\lambda}) &= \exp\left(\sum_{i\in S} \phi((F_T * \boldsymbol{\lambda})(i))\right) \\ &= \exp\left(\sum_{i\in S} |F_T * \boldsymbol{\lambda}|(i)\right). \end{aligned} \tag{3}$$

Model (3) is often referred to as a FRAME model, after (Zhu et al., 1998). In Eq. (3), the potential function $\phi(\cdot) = |\cdot|$ propagates extreme filter responses and enhances signals which are similar to the pattern exhibited by the filter. Moreover, since the function $Q(\boldsymbol{\lambda})$ is the summation of the pixelwise filter responses to $\boldsymbol{\lambda}$ and the filter width w is bounded, the Gibbs label distribution (3) is guaranteed (Hammersley–Clifford theorem; Besag, 1974) to be a MRF with respect to the neighbourhood structure

$$N(i) = \{(x, y) \in S : |x - x_i| \leqslant w_x/2, |y - y_i| \leqslant w_y/2\}, \quad i \in S, \tag{4}$$

or, in other terms, letting $\boldsymbol{\lambda}_{(i)}$ as the label configuration after discarding the label at pixel i,

$$P(\lambda_i|\boldsymbol{\lambda}_{(i)}) = P(\lambda_i|\boldsymbol{\lambda}'_{(i)}) = p(\lambda_i|\boldsymbol{\lambda}_{N(i)}), \quad i \in S,$$

for any two configurations $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$ sharing the same coordinates at the neighbourhood $N(i)$.

3. A parametric Gabor filter

Motivated by applications in fingerprint image enhancement (Yang et al., 2003 and references therein), filter F_T in (3) is specified as a parametric Gabor filter. As discussed previously, we focus on the estimation of the average ring width, i.e., the average distance between the centroids of two consecutive dark layers in a TCA image. Gabor filtering has been successfully exploited in reducing noise and increasing the contrast between ridges and valleys in gray-scale fingerprint images, because the parameters of a Gabor filter can be easily tuned to detect the distance between successive ridges.

A typical Gabor filter in fingerprint image analysis is the following two-dimensional function:

$$G(x, y) = \exp\left(\frac{-(rx'^2 + y'^2)}{2\tau^2}\right) \cos\left(\frac{2\pi}{T}x' + \psi\right), \tag{5}$$

with $x' = x\cos\alpha + y\sin\alpha$, $y' = -x\sin\alpha + y\cos\alpha$. In the equation above, T is the period of the sinusoidal plane wave and herewith the preferred wavelength of the filter, α is the wave orientation, ψ is the phase offset (for example, $\psi = 0$

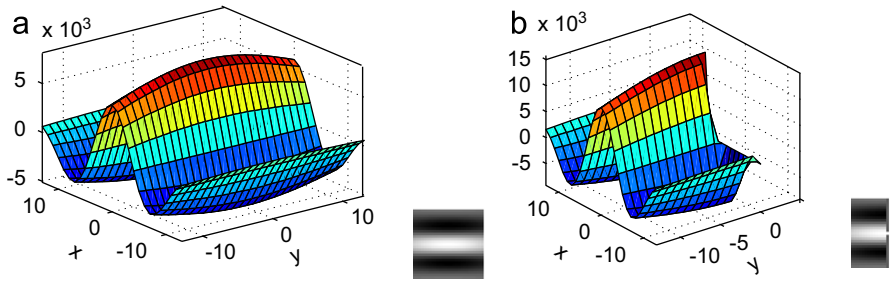

Fig. 3. Three-dimensional surface and image of a Gaborcosine function with $T = 16$ and $\alpha = 0$ in the range of $x, y \in [-13, 13]$ (a) and (b) in the range of $x \in [-13, 13]$, $y \in [-13, -1]$ and $(x, y) = (0, 0)$.

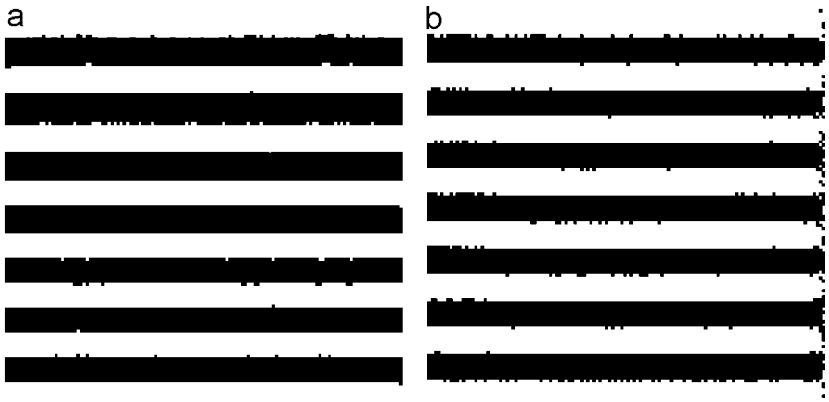

Fig. 4. Typical images of 128 × 128 pixels obtained by simulating a Markov random field model (a) and (b) a coupled Markov chain model, respectively filtered by the filters displayed in Fig. 3.

reduces filter (5) to an even-symmetric function) and r is the aspect ratio, which specifies the ellipticity of the filter. The ratio τ/T is related to what is known as the 'bandwidth', ω, of the Gabor function, through the following relationships:

$$\frac{\tau}{T} = \frac{1}{\pi}\sqrt{\frac{\ln 2}{2}}\frac{2^{\omega}+1}{2^{\omega}-1},$$
$$\omega = \log_2 \frac{\tau/T + \sqrt{\ln 2/2}}{\tau/T - \sqrt{\ln 2/2}},$$

from which it appears that the value of τ cannot be fixed independently, being a function of the bandwidth. Given T, the smaller the bandwidth is, the larger is the value of τ that modulates the support of the Gaussian kernel of the Gabor filter. When τ is large, the filter is more robust to noise, but it does not capture thin layers. In contrast, when τ is small, the filter captures image details, without removing noise. For TCA image analysis, we set $\tau = T$, in a way that the filter adapts layer detection to the average ring width, and $\psi = 0$, to reduce filter (5) to an even-symmetric function. Under this setting, filter (5) reduces to the following parametric Gaborcosine filter:

$$\begin{aligned} F_T(x, y) &= \exp\left(\frac{-(rx'^2 + y'^2)}{2T^2}\right)\cos\left(\frac{2\pi}{T}x'\right) \\ &= \left[\exp\left(\frac{-(rx'^2)}{2T^2}\right)\cos\left(\frac{2\pi}{T}x'\right)\right]\cdot\exp\left(-\frac{y'^2}{2T^2}\right), \end{aligned} \tag{6}$$

which is essentially a one-dimensional Gabor filter (a band pass filter), multiplied by a Gaussian kernel (a low pass filter). The band pass and low pass properties along the two orthogonal orientations are beneficial to enhancing TCA images, since these images usually show a periodic alternation between dark and bright layers orthogonal to the ring

orientation and simultaneously exhibit an approximate continuity along the ring orientation. To illustrate, Fig. 3(a) shows the filter (6) with a width $w = 13 \times 13$, evaluated at $T = 16$, $\alpha = 0$ and $r = 4$.

The image in Fig. 4(a) is a typical configuration of the MRF (3), whose energy function convolves the label distribution to filter (6). This image was generated by single pixel Gibbs sampling (see, for example Li, 2001), initializing the label configuration $\lambda^{(0)}$ at time $t = 0$ and then updating each pixel by repeatedly sampling a candidate $\lambda_i^{(t+1)}$ from the full conditional $P(\lambda_i^{(t+1)}|\lambda_{\mathscr{S}\setminus i}^{(t)})$. The transition probabilities $P(\lambda^{(t)}|\lambda^{(t-1)})$ are guaranteed to converge to the stationary distribution $P(\boldsymbol{\lambda})$. When choosing a random initial image and a random updating order of the pixels, the Gibbs sampler consists of the following steps.

Gibbs sampling algorithm for the Gabor-based MRF 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 ($N \cdot M$= size of the image)
5. randomly select site (x, y)
6. **for** all $(x', y') \neq (x, y)$
7. set $\lambda_{(x',y')}^{(t+1)} = \lambda_{(x',y')}^{(t)}$
8. **for** each gray value g of label $\lambda_{(x,y)}^{(t+1)}$
9. **for** all $(x', y') \in \{N(x, y), (x, y)\}$
10. calculate the new filter responses

$$(F * \lambda^{(t+1)})(x', y') = (F * \lambda^{(t)})(x', y') + F(x - x', y - y')(\lambda_{(x,y)}^{(t+1)} - \lambda_{(x,y)}^{(t)})$$

11. **for** each value $g = 0, 1$ of label $\lambda_{(x,y)}^{(t+1)}$
12. set $\lambda_{(x,y)}^{(t+1)} = g$ with (conditional) probability

$$P(\lambda_{(x,y)}^{(t+1)} = g|\lambda_{N(x,y)}^{(t)}) = \frac{e^{(\sum_{(x',y')\in N(x,y)}|(F*\lambda^{(t+1)})(x',y')|)}}{\sum_{g=0}^{G} e^{(\sum_{(x',y')\in N(x,y)}|(F*\lambda^{(t+1)})(x',y')|)}}$$

13. update the filter response $F * \lambda^{(t+1)}$

If the computer precision is not enough to calculate the conditional probability $P(\lambda_{(x,y)}^{(t+1)} = g|\lambda_{N(x,y)}^{(t)})$ in step 12, one can easily insert a nourishing one. In addition, toroidal boundary conditions suit TCA images well and can be used to avoid edges effects (Besag and Moran, 1975).

To detect convergence, the Gelman–Rubin multivariate convergence statistic R (Brooks and Gelman, 1998) is used on every (20×20)th pixel of the image. The Gibbs sampler stops iterating when $R < 1.2$. The algorithm above needs $O(w \cdot NM \cdot s)$ operations where w is the size of the filter support and s is the number of sweeps of the Gibbs sampler.

4. Parameter estimation and image 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

$$L(\boldsymbol{\theta}, T) = \sum_{\boldsymbol{\lambda}\in\{0,1\}^{NM}} P(\boldsymbol{\lambda}|T) \prod_{i\in S} f(z_i|\lambda_i, \boldsymbol{\theta}). \tag{7}$$

However, direct likelihood maximization is intractable because of the size of label space, 2^{NM}. To maximize (7), we consider an *EM* algorithm, based on the complete log-likelihood

$$l(\boldsymbol{\theta}, T) = \log P(\boldsymbol{\lambda}|T) + \sum_{i\in S} \log f(z_i|\lambda_i, \boldsymbol{\theta}).$$

The algorithm starts with the preliminary estimates $\boldsymbol{\theta}^{(0)}$ and $T^{(0)}$ of the parameters $\boldsymbol{\theta}$ and T, and then proceeds iteratively by alternating an expectation (E) and a maximization (M) step. In the E-step of the tth iteration, the following conditional expectation of the complete log-likelihood

$$E[l(\boldsymbol{\theta}, T)|\mathbf{z}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}] = \int_{\boldsymbol{\lambda}\in\{0,1\}^{NM}} l(\boldsymbol{\theta}, T)\, P(\boldsymbol{\lambda}|\mathbf{z}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)})\, \mathrm{d}\boldsymbol{\lambda}$$

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:

$$(\boldsymbol{\theta}^{(t)} T^{(t)}) = \arg\max_{\boldsymbol{\theta}, T} E[l(\boldsymbol{\theta}, T)|\mathbf{z}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}].$$

Since each iteration is guaranteed to increase the (incomplete) log-likelihood (7) under mild assumptions, the *EM* algorithm will converge to a local maximum (Wu, 1983).

For each pixel i, let $\lambda_{N(i)}$ be the label configuration at the neighbourhood of i, according to the neighbourhood structure (4). For our Gaussian model and $g = 0, 1$, the *EM* algorithm reduces to the following three updating formulas:

$$\mu_g^{(t)} = \frac{\sum_{i\in\mathscr{S}} z_i\, P(\lambda_i = g|z_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)})}{\sum_{i\in\mathscr{S}} P(\lambda_i = g|z_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)})}, \tag{8}$$

$$(\sigma_g^{(t)})^2 = \frac{\sum_{i\in\mathscr{S}} (z_i - \mu_g^{(t)})^2 P(\lambda_i = g|z_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)})}{\sum_{i\in\mathscr{S}} P(\lambda_i = g|z_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)})} \tag{9}$$

and

$$T^{(t)} = \arg\max_{\{T\}} \sum_{i\in\mathscr{S}} \sum_{g=0}^{1} P(\lambda_i = g|z_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)})\, \log P(\lambda_i = g|\lambda_{N(i)}, T). \tag{10}$$

The conditional probabilities $P\left(\lambda_i = g|z, \lambda_{N(i)}, \boldsymbol{\theta}, T\right)$ are not available in closed form, but they can be approximated relying on the mean field approximation approach (MFA; Celeux et al., 2003), by setting

$$P(\lambda) \approx \prod_{i\in\mathscr{S}} P(\lambda_i|\tilde{\lambda}_{N(i)}), \tag{11}$$

where the configuration $\tilde{\lambda}$ is set to the expected values of the label image, as follows:

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

The product

$$\prod_{i\in\mathscr{S}} P(\lambda_i|E[\lambda_{N(i)}])$$

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 (Celeux et al., 2003). The E-step of the *EM* algorithm hereby changes to

$$
\begin{aligned}
&E[l(\boldsymbol{\theta}, T)|\mathbf{z}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}] \\
&\quad \approx \sum_{i\in\mathscr{S}} \sum_{g=0}^{1} \log(P(\lambda_i = g|\tilde{\lambda}_{N(i)}, \boldsymbol{\theta}, T) f(z_i|\lambda_i, \boldsymbol{\theta}, T)) \\
&\qquad \times P\left(\lambda_i = g|z_i, \tilde{\lambda}_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}\right) \\
&\quad = \sum_{i\in\mathscr{S}} \sum_{g=0}^{1} (\log P(\lambda_i = g|\tilde{\lambda}_{N(i)}, T) + \log f(z_i|\lambda_i = g, \boldsymbol{\theta})) \\
&\qquad \times \frac{P(\lambda_i = g|\tilde{\lambda}_{N(i)}, T^{(t-1)}) f(z_i|\lambda_i = g, \boldsymbol{\theta}^{(t-1)})}{\sum_{g=0}^{1} P(\lambda_i = g|\tilde{\lambda}_{N(i)}, T^{(t-1)}) f(z_i|\lambda_i = g, \boldsymbol{\theta}^{(t-1)})}.
\end{aligned}
$$

The parameter estimates can therefore be updated by Eqs. (8)–(10) and replacing $\lambda_{N(i)}$ therein by $\tilde{\lambda}_{N(i)}$, which are computed iteratively. Our *EM* algorithm then takes the following form:

EM algorithm via mean field approximation

1. input TCA image $\mathbf{z}$

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** $g = 0, 1$
8. calculate the conditional probability

$$f(z_i|\lambda_i = g, \mu^{(t-1)}, \sigma^{(t-1)}) \propto \mathrm{e}^{-(Z_i-\mu_g^{(t-1)})^2/2(\sigma_g^{(t-1)})^2}$$

9. approximate the prior energy and probability

$$U(\lambda_i = g|\lambda_{N(i)}, T^{(t-1)}) \approx \sum_{j\in N(i)} |(F_{T^{(t-1)}} * \lambda)(j)|$$

$$P(\lambda_i = g|\lambda_{N(i)}, T^{(t-1)}) \approx \frac{\mathrm{e}^{U(\lambda_i=g|\lambda_{N(i)}, T^{(t-1)})}}{\sum_{g=0}^{1} \mathrm{e}^{U(\lambda_i=g|\lambda_{N(i)}, T^{(t-1)})}}$$

10. calculate the posterior probability

$$P(\lambda_i = g|z_i, \lambda_{N(i)}, \boldsymbol{\theta}^{(t-1)}, T^{(t-1)}) = \frac{f(z_i|\lambda_i = g, \boldsymbol{\theta}^{(t-1)}) P(\lambda_i = g|\lambda_{N(i)}, T^{(t-1)})}{\sum_{g=0}^{1} f(z_i|\lambda_i = g, \boldsymbol{\theta}^{(t-1)}) P(\lambda_i = g|\lambda_{N(i)}, T^{(t-1)})}$$

11. calculate the expected label

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

12. set $\lambda^{(t)} = \langle\lambda\rangle$
    update parameters
13. **for** $g = 0 : 1$
14. update μ_g according to Eq. (8)
15. update $(\sigma_g^{(t)})^2$ according to Eq. (9)
16. update T according to Eq. (10)

The initialization of $\boldsymbol{\lambda}$ and the sequential updating of the labels were chosen according to the recommendations in (Celeux et al., 2003). The number of iterations $t_{\max}$ was chosen according to the last gain obtained in the value taken by the mean-field-approximated expected log-likelihood

$$\sum_{i\in\mathscr{S}}\sum_{g=0}^{1}\log P(\lambda_i=g|\tilde{\lambda}_{N(i)},\boldsymbol{\theta},T)$$
$$\times\, P(\lambda_i=g|z_i,\tilde{\lambda}_{N(i)},\boldsymbol{\theta}^{(t-1)},T^{(t-1)})\quad f(z_i|\lambda_i,\boldsymbol{\theta},T).$$

Although the main goal of the model is the estimation of ring width T, rather than a restoration of the rings, the estimates of $\boldsymbol{\theta}$ and T can be exploited to segment the image, by thresholding the predictive probabilities of the hidden labels, according to the following criterion:

$$\hat{\lambda}_i=\begin{cases}1 & P(\lambda_i=g|z_i,\lambda_{N(i)},\hat{\boldsymbol{\theta}},\hat{T})\geqslant 0.5,\\ 0 & \text{otherwise.}\end{cases}\tag{12}$$

The binary image $\hat{\boldsymbol{\lambda}}$ restores the rings, as predicted by the model. Although hidden MRFs are acknowledged as efficient tools for simultaneous noise removal and segmentation (Li, 2001), we believe that $\hat{\boldsymbol{\lambda}}$ should be mainly used for validation purposes, because it depends on the Gaussian degradation model (2), which is too simple to capture the complex structure of noise and blur in a TCA image.

5. A hidden FRAME chain

A causal variant of the hidden MRF described in the previous sections can be specified by modelling the label distribution by means of coupled Markov chains. Coupled chains are used to extend one-dimensional Markov chain modelling to a two-dimensional framework, by correlating neighbouring chains. This alternative approach is motivated by the typical procedure implemented by expert observers to detect tooth rings by visual inspection. Manual ring detection is often carried out by examining the image along the main orientation of the rings and moving from the left-hand to the right-hand side of the image. The decision on whether a row is a dark or a bright layer is often made on the basis of the information accumulated by the observer during the left-to-right visual scan of that row. A causal Markov model better emulates human behavior in ring detection than a nondirectional model, such as a MRF, and can help in explaining differences in the estimated number of rings between an observer and the hidden MRF model described above.

To place a coupled Markov chain model on the label configuration $\boldsymbol{\lambda}$, we "cut" the Gabor two-dimensional filter (6) into half (Fig. 3 (b)) and define the filter response at pixel (x, y) as follows:

$$[F'_T*\boldsymbol{\lambda}](x,y)=\sum_{x'=-p}^{p}\sum_{y'=-q}^{-1}F_T(x',y')\lambda_{x+x',y+y'}+F_T(0,0)\lambda_{x,y},\tag{13}$$

where p and q are integers delimiting the support of the "half" filter F'_T. The convolution above is 'centered' at site (x, y), but its support lies to the left of (x, y). The resulting label distribution is given by

$$P(\boldsymbol{\lambda})\propto\exp\left(\sum_{i\in S}|F'_T*\boldsymbol{\lambda}|(i)\right).\tag{14}$$

If the random field Λ is distributed according to model (14), then Λ is a particular coupled Markov chain model, referred to as the symmetric half plane (SHP; Jeng and Woods, 1987) Markov chain, and fulfills the following Markov property: for all sites $i=(x, y)\in S$,

$$P(\lambda_{x,y}|\lambda_{x',y'},(x',y')\in S,y'<y)=P(\lambda_{x,y}|\lambda_{x',y'},(x',y')\in C_{x,y}),$$

where $C_{x,y}=\{(x',y')\in S,|x-x'|<p,0<y-y'<q)\}$ and p and q are the two integers delimiting the support of filter F'_T. Furthermore, by means of the Hammersley–Clifford theorem (Besag, 1974), a SHP Markov chain can be also

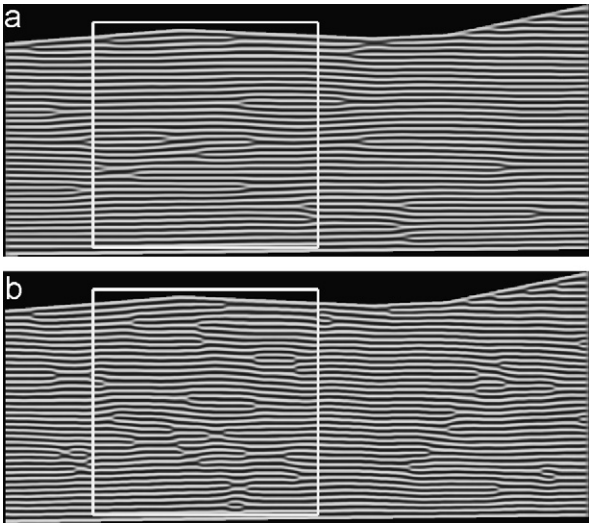

Fig. 5. The mean field approximation of the cementum band of TCA image 1 generated by the Gabor-filtered hidden Markov random field model (a) and the Gabor-filtered hidden coupled Markov chain model (b).

represented as a MRF with respect to the neighbourhood structure

$$N(x, y) = \left(\bigcup_{(x',y'),(x,y)\in C_{x',y'}} \{C_{x',y'} \cup (x', y')\} \backslash (x, y) \right) \cup C_{x,y}.$$

This allows us to simulate typical configurations of labels drawn from the SHP Markov chain by the Gibbs sampler outlined in Section 3 and estimate the parameters by implementing the *EM* algorithm described in Section 4. A typical configuration drawn from a FRAME chain is displayed in Fig. 4(b): the right hand border of the FRAME chain image is more fuzzy than its nondirectional counterpart, because the last column is not relevant to the past of any pixel and therefore drawn at random.

6. Results

A hidden FRAME model as well as a hidden FRAME chain model were fitted to the 2^{16} bit TCA image in Fig. 1. In both cases, the ellipticity parameter r and the orientation α in the Gabor filters (6) and (13) were empirically fixed at 4 and 0, respectively. Parameters $\boldsymbol{\theta}$ in the degradation model (2) as well as the filter parameter T were estimated by the *EM* algorithm illustrated in Section 4, based on a mean field approximation of the label distribution. The segmented image $\hat{\boldsymbol{\lambda}}$ was obtained from the last iteration of the *EM* algorithm, according to (12).

Fig. 5 shows $\hat{\boldsymbol{\lambda}}$, as estimated by the Gabor-filtered hidden MRF model and the Gabor-filtered coupled Markov chain. The latter model is more flexible than the former, and, as a result, more bifurcations occur. To emphasize this flexibility even more, experiments with the hidden coupled Markov chain model and a thinner filter (a horizontal past relevance structure of one pixel) were carried out additionally. Fig. 6 displays the resulting segmented image, clearly showing more flexibility in the reconstruction of tooth rings.

For illustration purposes, the centroids (Fig. 7) of the dark layers, as predicted by the two models, are superimposed on the original image displayed in Fig. 2. When implemented on the image displayed in Fig. 1, the *EM* estimation procedure of both models converged to the average ring width $\hat{T} = 14$. Dividing the area of the cementum band by 14, an estimate of 35 rings is obtained (we expected 34 rings, from the known age of the person to whom the tooth belongs). This number rises to 36 in the case of Fig. 6.

The estimates in the degradation model confirm the wider flexibility of the SHP Markov chain compared to that exhibited by its nondirectional counterpart. Recalling that our image is a 2^{16} bit intensity matrix, under the hidden MRF model, the average intensity of the dark and the bright layers was estimated by $\hat{\mu}_0 = 28\,565$, $\hat{\mu}_1 = 29\,519$, respectively.

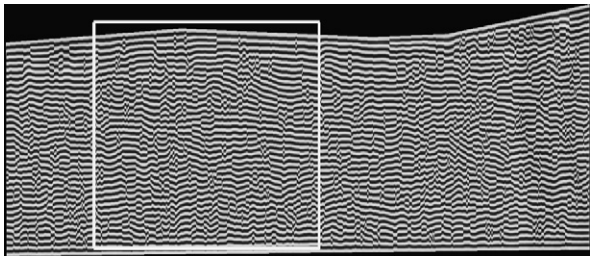

Fig. 6. The mean field approximation of the cementum band of TCA image 1 generated by the Gabor-filtered hidden coupled Markov chain, using a thin filter consisting of one column plus one pixel.

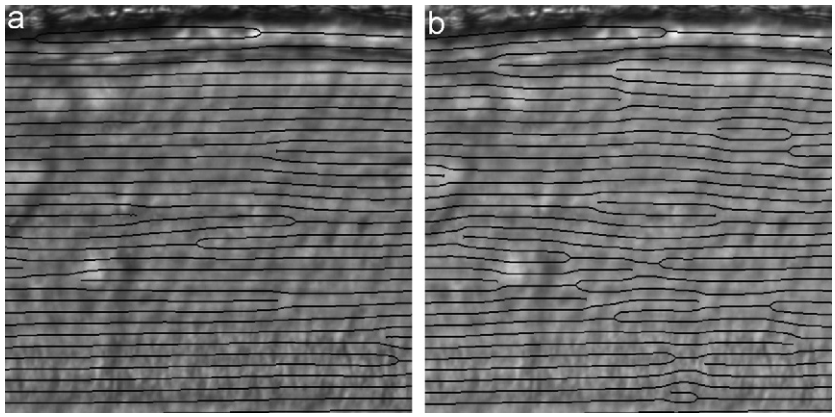

Fig. 7. The black rings from part of the mean field approximations in Fig. 5 overlayed onto the original TCA image 1.

Table 1
Expected and estimated number of rings in eight TCA images

Estimated by				Estimated by			
Image	Expected	Chain	Field	Image	Expected	Chain	Field
0231	41	46 (41)	46	1157	34	30 (30)	30
0592	60	64 (65)	64	1225	41	45 (35)	46
0682	35	33 (34)	33	1547	38	29 (33)	28
0688	35	35 (36)	34	1692	34	29 (29)	34

Under the SHP Markov chain model, $\hat{\mu}_0 = 28\,395$ and $\hat{\mu}_1 = 29\,692$ were obtained. In both cases, the *EM* algorithm converged to the same variance estimate $\hat{\sigma}_0^2 = \hat{\sigma}_1^2 = 4.8 \times 10^7$. The large variance and the small difference between the estimated means confirms that simple classification procedures based on pixelwise intensities fail in the restoration of the correct number of rings in a TCA image, as discussed in Section 1. In contrast, the FRAME approach succeeds in addressing ring counting, by adjusting the information provided by pixel intensities for the large-scale features of the image.

Additional experiments with eight TCA images of mixed quality gave the results reported in Table 1. The table lists the estimated number of rings for the hidden MRF model ('field') and for the hidden SHP chain ('chain'). Listed in parentheses are the estimates for the hidden SHP chain, by using the thin one-column filter mentioned above. The expected number of rings is the known age minus average tooth eruption age.

In general, the estimates of all three models lie very close together. In order to judge which model estimates the number of tooth rings best, we certainly need to perform experiments on more TCA images. But overall, these results are competitive to the manually counted number of rings.

7. Discussion

The large-scale properties of TCA images would suggest the specification of an MRF with higher-order interactions (Tjelmeland and Besag, 1998) for the latent label distribution. The FRAME approach (Zhu et al., 1998) allows for a parsimonious specification of a higher-order interactions random field. In particular, we described two Markov models that are capable to capture image macro-features, by means of one parameter, T. The price to pay for parsimony is the rigidity of our models that capture only one macro-feature of the TCA image, namely the average distance between dark layers. The ring bifurcations arising in Fig. 7 show an obvious consequence of this rigidity. The reconstruction of the TCA image is heavily influenced by (1) the shape of the single filter we estimated, and (2) the assumption of a constant filter orientation. In the present form, the models discussed here can only take into account strong local changes of tooth rings.

In order to overcome this global property, a natural solution would be provided by location-dependent filters. However, estimating the filter parameter T at each pixel i makes the filtered MRF models unidentifiable, without suitable covariates that, at present, are not available. Trying to estimate a location-dependent orientation is affected by the same drawback. A second approach to make our models more flexible would rely on using a thinner filter F_T and hereby reducing the size of the neighbourhood. The hidden SHP chain model described in Section 5 follows this line of reasoning and, accordingly, produces a more flexible segmentation, but does not meet the tooth rings better than the MRF model. In general, it is suspected that the more flexible the model is, the worse it performs for bad quality TCA images.

The suggested estimation procedure relies on an *EM* algorithm, based on a mean field approximation of the label distribution. Other approximations could be taken into account, such as the mode field approximation and the simulated field approximation (Celeux et al., 2003). These methods have been developed to make the *EM* algorithm more tractable when images of large size are modeled via a hidden Markov specification, and less time consuming with respect to Monte Carlo procedures, such as the Gibbsian *EM* algorithm (Chalmond, 1989).

We have pursued a likelihood approach for inference in a mixture model, focusing on the estimation of the distance between dark layers in a TCA image, as ring counting in TCA analysis is more important than ring restoration. Although the suggested *EM* algorithm allows to segment the image satisfactorily, a Bayesian approach would be a valid alternative to restore the shape of the rings. As suggested by a referee, the resulting MCMC estimation scheme could be carried out even for large images such as TCA pictures, by updating pixels in blocks (Chib and Greenberg, 1995; Swendsen and Wang, 1987).

Acknowledgements

Most of this work has been carried out while both the authors were at the Max Planck Institute for Demographic Research (MPIDR) in Rostock (Germany). The authors would like to thank three anonymous reviewers and the members of MPIDR tooth lab for their support, especially A. Fabig and U. Cleven.

References

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. J. Roy. Statist. Soc. B 36 (2), 192–236.

Besag, J., 1986. On the statistical analysis of dirty pictures. J. Roy. Statist. Soc. B 41 (3), 259–302.

Besag, J., Moran, P., 1975. On the estimation and testing of spatial interaction in gaussian lattice process. Biometrika 62, 555–562.

Brooks, S.P., Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. J. Comput. Graphical Statist. 7 (4), 434–455.

Celeux, G., Forbes, F., Peyrard, N., 2003. EM procedures using mean field-like approximations for markov model-based image segmentation. Pattern Recognition 36 (1), 131–144.

Chalmond, B., 1989. An iterative Gibbsian technique for reconstruction of m-ary images. Pattern Recognition 22, 747–761.

Chib, S., Greenberg, E., 1995. Understanding the metropolis–hastings algorithm. Amer. Statist. 49, 327–335.

Crowe, D.M., Strickland, M.D., 1975. Dental annulation in the American badger. J. Mammology 56, 269–272.

Czermak, A., Czermak, A., Ernst, H., Grupe, G., 2006. A new method for the automated age-at-death evaluation by tooth-cementum annulation (TCA). Anthropologischer Anzeiger 64 (1), 25–40.

Hoppa, R.D., Vaupel, J.W., 2002. Paleodemography: Age Distributions from Skeletal Samples. Cambridge University Press, Cambridge.

Jain, A.K., Farrokhnia, F., 1991. Unsupervised texture segmentation using gabor filters. Pattern Recognition 24 (12), 1167–1186.

Jeng, F.-C., Woods, J.W., 1987. On the relationship of the markov mesh to the NSHP Markov chain. Pattern Recognition Lett. 5 (4), 273–279.

Kagerer, P., Grupe, G., 2001. Age-at-death diagnosis and determination of life-history parameters by incremental lines in human dental cementum as an identification aid. Forensic Sci. Internat. 118 (1), 75–82.

Klauenberg, K., 2007. A semi-automatic procedure for the evaluation of tooth cementum annulation images for paleodemographic studies. Ph.D. Dissertation, Chair of Pattern Recognition and Image Processing (LMB), University of Freiburg.

Li, S.Z., 2001. Markov Random Field Modeling in Image Analysis. Springer, Tokyo.

Marčelja, S., 1980. Mathematical description of the responses of simple cortical cells. J. Opt. Soc. Amer. 70 (11), 297–300.

Naylor, J.W., Miller, W.G., Stokes, G.N., Stott, G.G., 1985. Cemental annulation enhancement: a technique for age determination in man. Amer. J. Phys. Anthropology 68, 197–200.

Renz, H., Schaefer, V., Duschner, H., Radlanski, R.J., 1997. Incremental lines in root cementum of human teeth: an approach to their ultrastructural nature by microscopy. Adv. Dental Res. ll, 472–477.

Stott, G.G., Sis, R.F., Levy, B.M., 1982. Cemental annulation as an age criterion in forensic dentistry. J. Forensic Dentistry 61, 814–817.

Swendsen, R., Wang, J.S., 1987. Nonuniversal critical dynamics in Monte Carlo simulations. Phys. Rev. Lett. 58, 86–88.

Tjelmeland, H., Besag, J., 1998. Markov random fields with higher-order interactions. Scand. J. Statist. 25, 415–433.

Yang, J., Liu, L., Jiang, T., Fan, Y., 2003. A modified gabor filter design for fingerprint image enhancement. Pattern Recognition Lett. 24, 1805–1817.

Wittwer-Backofen, U., Gampe, J., Vaupel, J.W., 2003. Tooth cementum annulation for age estimation: results from a large known-age validation study. Amer. J. Phys. Anthropology 123 (2), 119–129.

Wu, C.F.J., 1983. On the convergence properties of the EM algorithm. Ann. Statist. 11, 95–103.

Zhu, S.C., Wu, Y., Mumford, D.B., 1998. Filters random fields and maximum entropy (FRAME): towards a unified theory for texture modeling. Internat. J. Comput. Vision Arch. 27 (2), 107–126.