

# Bayesian Inversion for Three-Dimensional Dental X-ray Imaging with Limited Data

Ville Kolehmainen, Antti Vanne, Samuli Siltanen, Seppo Järvenpää, Jari P Kaipio,  
Matti Lassas and Martti Kalke.

## Abstract

Diagnostic and operational tasks based on dental radiology often require three-dimensional (3D) information that is not available in a single x-ray projection image. A computerized tomography (CT) scan can provide the dentist with comprehensive three-dimensional data. However, in many situations a complete CT-scan may not be available or practical because of high radiation dose or the cost of the CT-scanner equipment. In this paper, we consider a novel type of 3D imaging modality for dental radiology. We consider situations in which projection images of the teeth are taken from a few sparsely distributed projection directions using the dentist's regular (digital) x-ray equipment and the 3D x-ray attenuation function is reconstructed. A complication in these experiments is that the reconstruction of the 3D structure based on a few projection images becomes an ill-posed inverse problem. Bayesian inversion is a well suited framework for reconstruction from such incomplete data. In Bayesian inversion, the ill-posed reconstruction problem is formulated in a well-posed probabilistic form in which *a priori* information is used to compensate for the incomplete information of the projection data. In this paper we propose a Bayesian method for 3D reconstruction in dental radiology. The method is partially based on [Kolehmainen et al. *Physics in Medicine and Biology* **48** (2003), 1465–1490]. The prior model for dental structures consist of a weighted  $\ell^1$  and total variation (TV)-prior together with the positivity prior. The inverse problem is stated as finding the maximum a posteriori (MAP) estimate. To make the 3D reconstruction computationally feasible, a parallelized version of an optimization algorithm is

This work was supported by the National Technology Agency of Finland (TEKES, contracts 206/03 and ) and the Academy of Finland (project 203985).

Ville Kolehmainen, Antti Vanne and Jari P Kaipio are with the University of Kuopio, P.O.Box 1627, FIN-70211 Kuopio, Finland.

Seppo Järvenpää and Matti Lassas are with the Helsinki University of Technology, P.O.Box 1100, FIN-02015 TKK, Finland  
Samuli Siltanen and Martti Kalke are with the Instrumentarium Corporation, Imaging Division, P.O.Box 20, FIN-04301 Tuusula, Finland. Instrumentarium Corporation is part of GE-Healthcare.

implemented for a Beowulf cluster computer. The method is tested with projection data from dental specimens and patient data. Reconstructions with traditional backprojection method are given as reference for the proposed method.

### Index Terms

Inverse problem, Bayesian inversion, x-ray imaging, tomography, total variation, dental radiology, parallel computing.

## I. INTRODUCTION

The main tool in dental radiology is the x-ray projection image, which reveals the structure of bone, soft tissues and teeth. However, certain diagnostic and operative tasks often require more precise knowledge of the three-dimensional (3D) structure of oral tissues than is available in a single x-ray projection image or a panoramic image. For example, in dental implantology it is critical for the firm attachment to use optimal screw size and therefore measure optimal depth and angle for the screw hole; if the bone implant screw is too short, the implant will be loose. On another hand, if the hole is drilled too deep, the mandibular nerve could be damaged. Other tasks, in which 3D information is needed, include, for example, detection of bone loss between teeth, detection whether two roots have grown together with common root canal or not, and finding out whether certain roots have an intimate relationship with the inferior dental canal, leading to possible nerve damage during a tooth removal operation [1]–[3].

Comprehensive 3D information about the tissues can be obtained by computerized tomography (CT) imaging. The basic philosophy of CT-imaging is to use a large, delicately built scanner to measure an extensive set of projections from all around the body for the 3D reconstruction. The disadvantages of CT-imaging are high radiation dose and high cost. Therefore, the use of CT-imaging has been mainly limited to diagnostics of serious diseases and it is typically utilized in large hospital units. Recently, there has been increasing interest in dedicated dental cone beam CT-based systems (CBCT) [4]–[6]. Since these dental CBCT systems are based on the same philosophy than non-dedicated CTs, the CBCT systems provide good quality 3D information but they also have similar limitations than non-dedicated CTs; i) the high radiation dose and ii) the total cost of the dental CBCT devices is typically much more than the cost of a standard digital panoramic device.

Because of the limitations described above, we have considered a novel type 3D imaging modality for dental radiology. The modality can be viewed as an intermediate between traditional x-ray imaging and CBCT-imaging; the main idea is to construct a system which could provide the needed 3D information

about the tissues based on a small number of projection images and could handle incomplete imaging geometry. With this kind of construction, the projection images could be taken by any existing digital x-ray system, including dental panoramic and intraoral systems. To be more specific, the experiments we have considered can be divided into the modalities of intraoral and extraoral imaging. In intraoral imaging, the projection radiographs are taken with a small digital sensor that is in fixed position inside the patient's mouth and the x-ray source is mounted on a foldable arm that can be used to move the source to different positions. Because of having the detector in fixed position, the x-ray source positions are limited to a cone with opening angle significantly less than  $180^\circ$ . This restriction leads to a limited-angle tomography problem. In extraoral imaging, a few projection radiographs of the region of interest (ROI) are taken through the head from sparsely distributed projection directions. The imaging geometry in extraoral imaging is typically similar to conventional CT-geometry, and therefore, the experiments can be viewed as a limited data version of the CBCT-experiments. In both modalities, the projection images are often truncated because of the small detector size or to minimize dose to vital organs outside the region of interest. The truncation of the projections leads to the additional complication of local tomography problem [7], [8]. In local tomography the region of interest is surrounded by tissue that is not reconstructed. In this problem class the goal is to reconstruct the attenuation function in the region of interest using only x-rays passing through the region of interest. As opposed to conventional CT data, we refer the above type of limited data as *sparse* projection data.

A complication in these experiments is that the set of a few projections that are collected from sparsely distributed directions does not contain sufficient information to completely describe the 3D x-ray attenuation function. In other words, the reconstruction problem becomes an ill-posed inverse problem. It is well-known that traditional reconstruction methods, such as filtered backprojection (FBP), do not perform well when applied to sparse projection data [9]–[11].

Bayesian inversion is a natural framework to tackle the reconstruction problem with sparse projection data. In Bayesian inversion, *a priori* knowledge of the tissue is used in the image reconstruction problem to compensate for the incomplete information of the sparse projection data. For example, in dental x-ray imaging we know that the attenuation function is a nonnegative, piecewise regular function, and different tissue types are separated by well-defined boundaries. The unknown x-ray attenuation function and projection data are considered as random variables, and separate statistical models (probability distributions) are formulated for (1) the acquisition of the projection data and (2) the *a priori* information. Based on these models and the Bayes formula, the complete solution of the inverse problem is provided by the posterior probability distribution. Final images of the tissue are then obtained as point estimates

from the posterior distribution. Bayesian methods have been found to give improved reconstruction quality over traditional methods in several studies, see e.g. [12]–[18].

A major practical difficulty in applying Bayesian methods to 3D x-ray imaging is the heavy computational requirements. For this reason, most of the previous studies on the topic have concentrated on 2D problems. If realistic resolution in a 3D problem is used, the number of unknown voxel values is typically in the range  $10^6 - 10^7$ , and thus, the computation of the posterior statistics leads to large-scale optimization or integration problems. Thus, powerful computers and efficient numerical algorithms are required to compute the 3D reconstruction in clinically acceptable time.

In this paper, we propose a Bayesian method for 3D reconstruction in dental x-ray imaging with sparse projection data. The method is partially based on the recent work [17], [18] in which reconstructions were computed using 2D forward models and 3D reconstructions were constructed as stacks of 2D slices. In the present work, the reconstruction problem is implemented with full 3D models. In the Bayesian model we use a weighted  $\ell^1$  and total variation (TV) prior, together with the positivity prior, as the prior model for the 3D x-ray attenuation function. This model is qualitatively in accordance with our *a priori* knowledge about dental structures. A parallelized version of a gradient-based optimization method is implemented for the computation of the maximum a posteriori (MAP) estimate of the 3D x-ray attenuation function. The results are computed on a 13 node Beowulf cluster that was constructed for the purpose of 3D imaging problems. The performance of the method is tested with sparse projection data that were collected using commercial dental x-ray imaging equipment. Traditional tomosynthetic (backprojected) reconstructions are shown as reference images for the assessment of the statistical model.

We note that parallelization of the x-ray tomography problem has been previously discussed in [19]. In that article a parallelized version of an iterative coordinate-descent functional-substitution (ICD/FS) algorithm for Bayesian MAP-estimation was derived. In the ICD/FS-method the image is updated sequentially one pixel at time by minimizing quadratic approximations of the original one-dimensional (conditional) optimization functionals.

This paper is organized as follows. In Section II we discuss the forward model for x-ray imaging. In Section III we discuss the Bayesian model and methods that are used in this paper. We also describe the algorithm for computation of the MAP estimates. In Section IV we discuss the parallel implementation of the methods for the Beowulf computer cluster. Results based on experimental data are given in Section V. In Section VI we discuss some future topics and open problems, and in Section VII we give conclusions.

## II. FORWARD MODEL

In x-ray imaging, an almost point-like x-ray source is placed on one side of an object. Radiation passes through the object and is detected on the other side. See Figure 1. In digital x-ray imaging, the unscattered radiation is detected with a digital sensor that can be thought of as an 2D array of point-like detectors. The volume being imaged is modelled by a bounded subset  $\Omega \subset \mathbb{R}^3$  with a nonnegative x-ray attenuation function  $x : \Omega \rightarrow [0, \infty)$ . For the projection measurement, we use the usual pencil-beam model

$$m_j = -\log\left(\frac{I_j}{I_0}\right) = \int_{L_j} x(s) ds, \quad (1)$$

where  $m_j$  is the value of the projection measurement for the  $j^{\text{th}}$  source to detector pixel line  $L_j$  in the set of projection data,  $I_j$  is the measured x-ray intensity and  $I_0$  is the intensity of the x-ray source.

In the discretization of the attenuation model (1) the domain  $\Omega$  is divided into a lattice of  $M$  disjoint 3D voxels  $\Omega_i$  and the length of the path  $L_j$  inside each voxel  $\Omega_i$  is computed, see Figure 1. Assuming that the attenuation function  $x(s)$  is constant within each voxel  $\Omega_i$ , the projection measurement  $m_j$  can be approximated in the form

$$m_j = \int_{L_j} x(s) ds \approx \sum_{i=1}^M x_i |\Omega_i \cap L_j|. \quad (2)$$

where  $|\Omega_i \cap L_j|$  denotes the length of ray  $L_j$  through voxel  $\Omega_i$ . Arranging the whole set of  $N$  projection measurements into a vector  $m = (m_1, m_2, \dots, m_N)^T \in \mathbb{R}^N$ , we obtain

$$m = Ax, \quad (3)$$

where  $x = (x_1, x_2, \dots, x_M)^T \in \mathbb{R}^M$  is the vector of attenuation values in the voxels and matrix  $A$  implements Eq. (2).

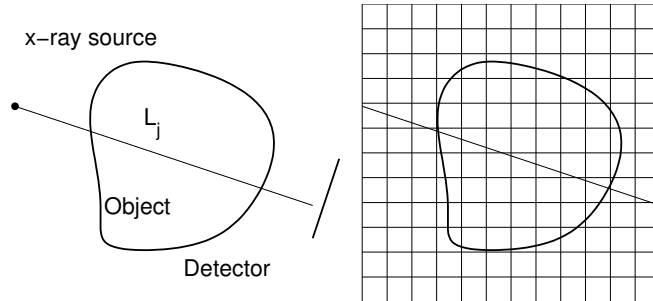


Fig. 1. Left: Schematic illustration of the pencil-beam attenuation model for x-ray imaging. Right: The domain  $\Omega \subset \mathbb{R}^3$  under investigation is discretized into a lattice of  $M$  voxels  $\Omega_i$ .

### III. INVERSE PROBLEM

#### A. Bayesian inversion

The main idea in Bayesian inversion is to consider the inverse problem as a problem of Bayesian inference. All unknown variables are modelled as random variables. The probabilistic modelling of these variables reflects our uncertainty of their actual values and the degree of uncertainty is coded in the probability distributions of these random variables.

In the Bayesian framework the posterior distribution, which is obtained by Bayes theorem

$$p(x | m) = \frac{p_{\text{pr}}(x)p(m | x)}{p(m)}, \quad (4)$$

represents the complete solution of the inverse problem. In Eq. (4),  $p(m | x)$  is the so-called likelihood function,  $p_{\text{pr}}(x)$  is the prior density and  $p(m)$  is the normalization constant. The likelihood function  $p(m | x)$  is a statistical model for the observations. Typically, it is interpreted as a model that describes the likelihood that the measured data  $m$  would have been observed from a given realization  $x$ . The prior density  $p_{\text{pr}}(x)$  is statistical model for the unknown  $x$ . It is designed based on the *a priori* information about the object.

Typically, in inverse problems the posterior is a probability density over a high-dimensional space. Thus, to summarize and visualize the solution, various posterior statistics can be computed. The most usual choice is the MAP estimate

$$p(x_{\text{MAP}} | m) = \max p(x | m), \quad (5)$$

which is then shown as the reconstructed image. For further details on Bayesian inversion theory, see e.g. [11], [17], [20], [21].

#### B. Bayesian model for 3D dental imaging

The nature of the x-ray attenuation measurements is a counting process that obeys Poisson statistics. However, in x-ray imaging a large amount of photons are usually detected in each detector pixel, and then the logarithm of the number of observed counts are used as the projection data. It has been shown in [14], [17] that in such case the value of the likelihood can be well approximated with a Gaussian noise model. Thus, we write the model

$$m = Ax + \epsilon, \quad (6)$$

for the x-ray projection data. In Eq. (6), the observation noise is assumed to be independent of  $x$ , zero mean Gaussian  $\epsilon \sim \mathcal{N}(0, C)$  with an invertible covariance matrix  $C$ . With these assumptions, the

likelihood function takes the form

$$p(m|x) \propto \exp\left(-\frac{1}{2}\|L(m - Ax)\|^2\right), \quad (7)$$

where  $L^T L = C^{-1}$  and  $\|\cdot\|$  is the usual Euclidian ( $\ell^2$ ) vector norm.

As the prior model for the attenuation function of the dental structures, we write

$$p_{\text{pr}}(x) \propto p_+(x) \exp(-W(x)), \quad (8)$$

where

$$p_+(x) = \prod_{k=1}^M \theta(x_k) \quad (9)$$

is the positivity prior,  $\theta$  is the Heaviside function

$$\theta(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases} \quad (10)$$

and the functional  $W(x)$  is of the form

$$\begin{aligned} W(x) &= \alpha_0 \|x\|_1 + \alpha_1 \text{TV}(x) \\ &= \alpha_0 \sum_{i=1}^M |x_i| + \alpha_1 \sum_{i=1}^M \sum_{j \in \mathcal{N}_i} |x_i - x_j|, \end{aligned} \quad (11)$$

where  $\mathcal{N}_i$  denotes the usual six-point neighborhood for voxel  $i$  in the 3D lattice,  $\text{TV}(x)$  is the discretized total variation functional and  $\|x\|_1$  denotes the  $\ell^1$ -norm of the vector  $x$ . The total variation prior is known to favour piecewise regular solutions, in which different tissues are separated to a few subregions with short boundaries and small variation in the attenuation parameter within each subregion. The  $\ell^1$ -prior is known to favour solutions which consist of a few small high-attenuation targets on low-attenuation background. The parameters  $\alpha_0$  and  $\alpha_1$  play the role of regularizing parameters and they can be used to tune the relative weighting of the  $\ell^1$  and TV-functionals. The prior model (8-11) corresponds in qualitative sense closely to oral structures, which are expected to consist of a few well-defined subregions of different tissues (enamel, bone, soft tissue) on zero attenuation background (air). The features and the use of TV and  $\ell^1$  priors in inverse problems and image enhancement have been discussed, e.g. in [21]–[24].

Using Bayes theorem and the models (7) and (8), the posterior density assumes the form

$$p(x|m) \propto p_+(x) \exp\left(-\frac{1}{2}\|L(m - Ax)\|^2 - W(x)\right), \quad (12)$$

where proportionality is because of omitting the normalisation constants.

The computation of the MAP estimate from the posterior (12) amounts to solving the optimization problem

$$x_{\text{MAP}} = \arg \min_{x \geq 0} \left\{ \frac{1}{2} \|L(m - Ax)\|^2 + W(x) \right\}$$

### C. Computation of the MAP estimate

One of the major difficulties in applying Bayesian inversion to 3D x-ray imaging is the scale of the problem. When clinically relevant resolution is used, the number  $M$  of unknown voxel values is typically in the range  $10^6 - 10^7$ . Thus, the computation of the MAP estimate or any other posterior statistics is a computationally intensive task because of the large scale of the problem. Thus, the implementation has to be efficient to achieve clinically acceptable computation times.

To make the computation reasonably fast, we compute the MAP-estimate by gradient-based optimization techniques. There is a variety of optimization methods, such as steepest-descent, nonlinear conjugate-gradients and Newton methods, that could be used for the solution. In this study, we use the method for unconstrained large-scale optimization introduced by Barzilai and Borwein [25], [26]. This method was chosen because i) the computation of the gradient for the objective functional is reasonably fast, enabling relatively quick computation of multiple iterations and ii) the method can be parallelized efficiently.

In applying the gradient-based methods, we face two difficulties. First, the functional  $W(x)$  in the prior density (8) is non-differentiable because of the presence of the absolute-value function. To overcome this problem, we use the smooth approximation

$$|t| \approx h_\beta(t) = \frac{1}{\beta} \log(\cosh(\beta t)), \quad (13)$$

where  $\beta > 0$  is a parameter adjusting the accuracy of the approximation. In the following, the approximate differentiable prior functional is denoted by  $W_\beta(x)$ . The second difficulty arises from the positivity constraint. The use of constrained optimization methods is slow in high-dimensional problems. Thus, we take the positivity prior into account by applying the exterior-point search methods [27]. In the exterior-point search, the original constrained problem is approximated by a sequence of unconstrained problems

$$\begin{aligned} x_{\text{MAP}}^{(j)} &= \arg \min \left\{ \frac{1}{2} \|L(m - Ax^{(j)})\|^2 \right. \\ &\quad \left. + W_\beta(x^{(j)}) + E^{(j)}(x^{(j)}) \right\}, \end{aligned} \quad (14)$$

where superindex  $j$  is used to denote the  $j^{\text{th}}$  problem in the sequence and  $E^{(j)}(x^{(j)})$  is a penalty functional that is used to penalize the negative components of the solution  $x^{(j)}$ . In this study we use a functional of the form

$$E^{(j)}(x^{(j)}) = \sum_{k=1}^M \phi^{(j)}(x_k^{(j)}), \quad (15)$$

where

$$\phi^{(j)}(x_k^{(j)}) = \begin{cases} \gamma_j (x_k^{(j)})^2 & , x_k^{(j)} < 0 \\ 0 & , x_k^{(j)} \geq 0 \end{cases}, \quad (16)$$

and  $\{\gamma_j, j = 1, 2, \dots, N_s\}$  is a sequence of increasing positive penalty parameters. It should be noted that the exterior point methods guarantee the non-negativity of the solution only in the asymptotic limit  $j \rightarrow \infty$ .

The Barzilai-Borwein gradient-descent method consists of two steps: the update of the estimate and the computation of the step-length parameter. In the case of solving the problem (14), the update step assumes the form

$$x^{(j,t+1)} = x^{(j,t)} - s_t^{-1} g^{(j,t)} \quad (17)$$

where the update direction is the gradient of the argument in equation (14)

$$g^{(j,t)} = -A^T L^T L(m - Ax^{(j,t)}) + \nabla W_\beta(x^{(j,t)}) + \nabla E^{(j)}(x^{(j,t)}) \quad (18)$$

and the second superindex  $t$  denotes the iteration index. The second step, the computation of the step length parameter  $s_t$  is of the form

$$s_t = \frac{(x^{(j,t)} - x^{(j,t-1)})^T (g^{(j,t)} - g^{(j,t-1)})}{\|x^{(j,t)} - x^{(j,t-1)}\|^2}. \quad (19)$$

The stopping criteria for the optimization problem (14) is based on the magnitude of the gradient and the decrease in the optimization functional. Once these are under predetermined thresholds or maximum number of iterations is reached, the iteration process (17)-(19) is terminated.

In this study, the results are computed with a parallelized version of the algorithm (17)-(19). The computational techniques and implementation are briefly described in the next section.

#### IV. PARALLELIZED IMPLEMENTATION

In this study, a Beowulf cluster [28] was constructed for 3D imaging problems. The cluster consists of thirteen 3.0 GHz Pentium4 desktop-PCs having 4 gigabytes of memory on each node and the nodes are interconnected by a standard 1 Gb ethernet switch. The optimization algorithm (17)-(19) was implemented

using the ANSI-C programming language and PETSc [29] application library, which in turn uses MPICH [30] for interprocess communication and ATLAS [31] for linear algebra routines.

In the first stage of the parallel algorithm, the observation matrix  $A$  is constructed. PETSc provides a parallel compressed sparse row matrix format, whose rows are distributed across computation nodes. Observation matrix creation is easily parallelized by distributing the x-ray source locations and detector array geometries among the computation nodes, each of which calculates the respective ray lengths inside the voxel grid and insert the resulting rows in the matrix  $A$  independently of other nodes.

The second stage of the parallel algorithm implements the iteration (17)-(19) for the sequence of the exterior point problems (14). At this stage, the value and gradient of the modified prior functional  $W_\beta(x)$  need to be evaluated and new update directions calculated. The PETSc library provides a framework for distributing the image vector  $x$  across the computation nodes. The framework hides the details of fetching the neighboring voxels that possibly reside on other nodes and hence the implementation of the prior functional  $W_\beta(x)$  or other similar type Markov Random Field (MRF) functionals is straightforward. The evaluation of the likelihood gradient, which is the first term in Eq. (18), and the evaluation of residual  $\|L(m - Ax)\|^2$  for stopping criteria include matrix-vector products. These are implemented by the PETSc tools that distribute the computation of a matrix-vector product evenly on the computation nodes.

Computationally, the most demanding parts in the implementation of the algorithm (17)-(19) are the evaluation of the likelihood gradient and the residual for the stopping condition. In the test problems with which we have experimented, these take approximately 40% and 20% of the computation time, respectively. The rest of the time goes mainly to the construction of the matrix  $A$  and the evaluation of the prior functional and gradient. A more detailed benchmarking of the cluster implementation will be given for the last test case in the next section.

## V. RESULTS

In this section the proposed method is tested with projection data from dental specimens and a male patient. The first test case represents limited-angle tomography of a tooth specimen. To make the assessment of the results easier, full-angle reconstructions of the tooth specimen are given as “ground truth” in the first test case. The second test case involves intraoral 3D imaging of a realistic head phantom, and the last test case is based on extraoral projection data from a male patient. In each of the test cases, reconstructions with the traditional tomosynthetic method are shown as reference for the reconstructions with the Bayesian method. Tomosynthesis, which is basically equivalent to unfiltered backprojection, is currently the most widely used reconstruction method for 3D dental imaging with just a few projections.

For details, see [32], [33].

#### A. Case 1: Limited angle data from a tooth specimen

The projection radiographs of the tooth specimen were acquired using a commercial intraoral x-ray detector (Sigma) and a dental x-ray source (Focus<sup>1</sup>). The Sigma detector is a charge coupled device (CCD). The size of the imaging area is  $34\text{mm} \times 26\text{mm}$ . The resolution is  $872 \times 664$  pixels, with pixel size  $0.039\text{mm} \times 0.039\text{mm}$ .

The projection images of the tooth specimen, which is an extracted and dried mandibular tooth, were acquired using the conventional cone beam CT-geometry. This geometry was chosen so that we could provide a full-angle reconstruction as “ground truth” for the limited angle reconstructions. The Focus x-ray source and the Sigma detector were placed into fixed positions such that the source direction is normal to the detector plane. The distance from the focal spot to the detector array was 840mm. The tooth specimen was placed on a rotating platform so that projections from different directions can be taken. The distance from the center of rotation to the detector was 56mm. A photograph of the experimental setup is shown in the left image in Figure 2.

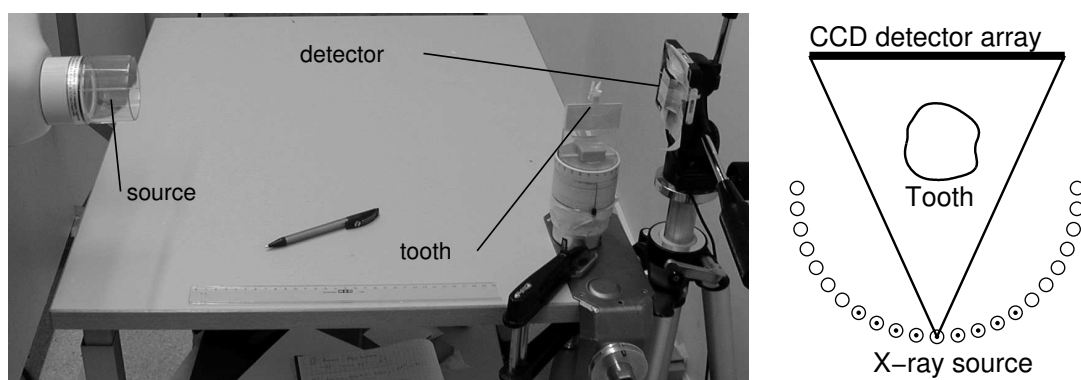


Fig. 2. Left: Photograph of the experimental setup. The Focus x-ray source is seen on the left and the CCD detector on the right. The tooth specimen is located atop the rotating platform. Right: Illustration of the projection geometry. The source locations (23 projections spanning a view-angle of  $187^\circ$ ) are shown with circles. The projection interval is  $8.5^\circ$ . For clarity, the detector location is shown only for one source location. The projections that are used in limited angle computations (9 projections spanning a view-angle of  $68^\circ$ ) are denoted by black dots within the circles.

Using the measurement setup, 23 equispaced projection images of the tooth phantom were taken from

<sup>1</sup>Sigma and Focus are registered trademarks of Instrumentarium Corp. Imaging Division (Instrumentarium Corp. is part of GE-Healthcare)

a total view angle of  $187^\circ$ . In the limited-angle computations, we used nine projections with a total view angle of  $68^\circ$ . A 2D illustration of the projection geometry is shown in the right image in Figure 2. The left image in Figure 3 shows the upper  $580 \times 664$ -part from one of the raw projection images. The lower parts of the raw images, which show the platform for the tooth, were not included in the reconstruction data. The projection images were downsampled to the size  $290 \times 332$  (i.e., half of the original size) before the transformation to tomographic data, leading to number of data  $N = 23 \times 290 \times 332 \approx 2.2 \cdot 10^6$  for the full-angle data, and  $N = 9 \times 290 \times 332 \approx 0.8 \cdot 10^6$  for the limited-angle data. The downsampling of the projection images can be justified by the redundancy of the data; the size of the detector pixel is small compared to the resolution that is typically used in the 3D reconstruction. Thus, a significant decrease in the computational load can be achieved at the cost of a small loss in signal-to-noise ratio of the data. The middle image in Figure 3 shows one 2D slice of the full-angle projection data in traditional sinogram form. The right image shows the corresponding sinogram for the limited angle data (9 projections spanning a view angle of  $68^\circ$ ).

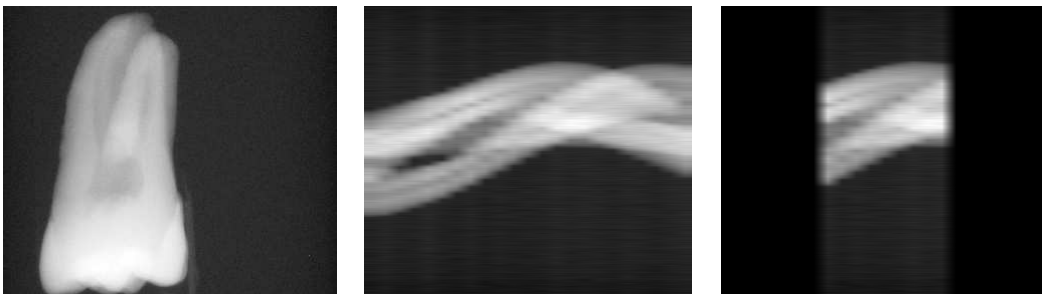


Fig. 3. Left: A projection radiograph of the tooth. Middle: One 2D slice of the full-angle data (23 projections spanning a view angle of  $187^\circ$ ) in sinogram form. Right: Corresponding slice of the limited-angle data (9 projections spanning a view angle of  $68^\circ$ ) in sinogram form. The black parts on left and right denote the missing parts of the sinogram.

The results for the tooth specimen data are shown in Figure 4. The top row shows one horizontal slice and the bottom row one vertical slice of the 3D reconstruction. The left column shows the MAP estimate from the full-angle data (23 projections spanning a view angle of  $187^\circ$ ), the middle column shows the tomosynthetic (backprojected) reconstruction from the limited-angle data (9 projections spanning a view angle of  $68^\circ$ ) and the right column shows MAP estimate from the limited angle data. The volume of the reconstruction domain  $\Omega$  in the computations was approximately  $26\text{mm} \times 26\text{mm} \times 22\text{mm}$ , and the size of the cubic voxels in the reconstruction was  $0.23\text{mm} \times 0.23\text{mm} \times 0.23\text{mm}$ . With this resolution, the number of unknowns in the inverse problem was  $M = 112 \times 112 \times 95 \approx 1.2 \cdot 10^6$ . For the observation noise we

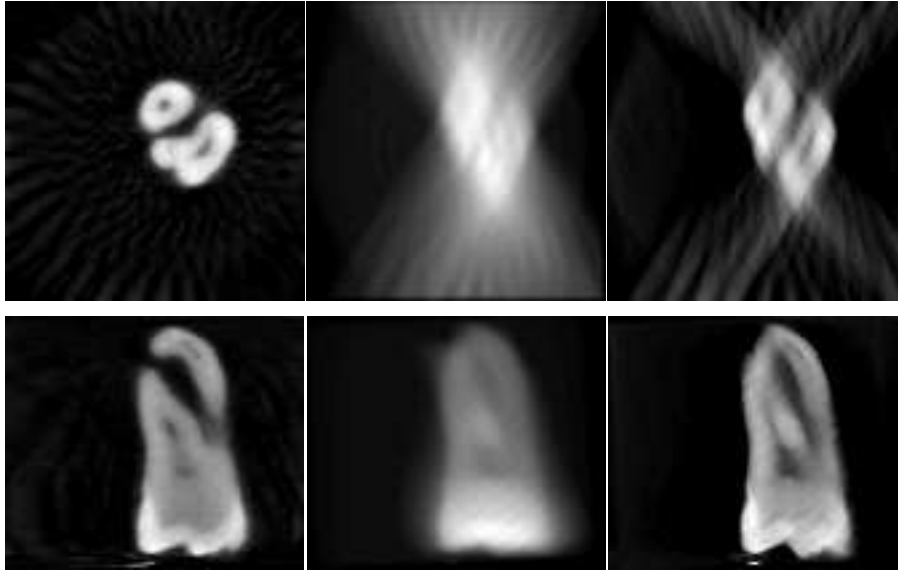


Fig. 4. Reconstructions from the projection data of the tooth specimen. Top row shows one horizontal slice and bottom row shows one vertical slice from the 3D reconstructions, respectively. Left column: MAP estimate from the full-angle data (23 projections spanning a view angle of  $187^\circ$ ). Middle column: Backprojection from the limited-angle data (9 projections spanning a view angle of  $68^\circ$ ). Right column: MAP estimate from the limited angle data.

made the simple assumption that the observation errors are independent, zero mean Gaussian-noise with covariance matrix  $C = I$ . In practice, a better model for the noise statistics can be obtained from a repeated set of phantom measurements or careful analysis of the measurement system. For the prior parameters we used values  $\alpha_0 = 10$  and  $\alpha_1 = 1$  in both, full and limited-angle, reconstructions. The smoothing parameter for the approximation of the absolute value function was  $\beta = 200$  and the sequence of the exterior point parameters  $\{\gamma_j, j = 1, \dots, 5\}$  were in the range from 3000 to 10. The parameter values were chosen manually from a set of reconstructions with different values, and the same prior parameters are also used in the remaining test cases. It is our experience that a fixed prior usually performs well for similar experiments with the same imaging equipment. Each problem in the sequence of the five exterior-point problems was iterated until convergence. The convergence was verified by monitoring the norm of the gradient  $g^{(j,t)}$  and the change in the residual (i.e., the value of the optimization functional in Eq. (14)); the iteration was stopped once either of these went below predefined thresholds, or maximum number of iterations was reached. For the norm of the gradient we used threshold value 1, for the change in residual threshold value 0.1 and the maximum number of iterations was set to 6. With these choices, each subproblem took typically 2-6 iterations. The computation times for the MAP estimate in the first test

case was 96 seconds for the full-angle case and 52 seconds for the limited-angle case. The computation of the backprojected reconstruction took only a few seconds.

As can be seen, the limited angle MAP estimate is better than the reconstruction with the backprojection method. A notable feature in the limited-angle MAP estimate is that it has less “bleeding” than the backprojection reconstruction in the depth direction where the information content of the limited-angle data is poor. The MAP reconstruction also conform well to the expected features of the tooth specimen, which consist of enamel, bone and air.

### B. Case 2: Intraoral data from a head phantom

In intraoral x-ray imaging, the measurement geometry is such that the detector is in a fixed position inside the patient’s mouth and the dentist can move the x-ray source, which is mounted on a foldable arm, with respect to the intraoral detector. A 2D illustration of this projection geometry is shown in the left image in Figure 5.

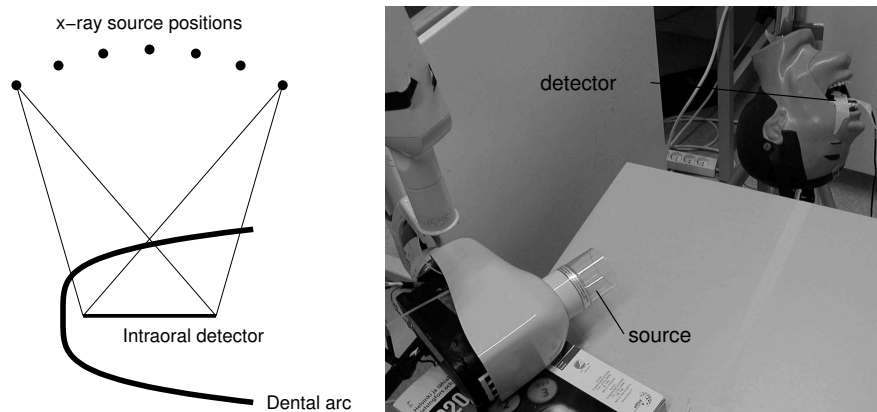


Fig. 5. Left: Geometry for intraoral measurements. The detector is in a fixed position inside the patient’s mouth. The source locations are denoted by black dots (7 projections spanning a view-angle of  $59.7^\circ$ ). Right: Experimental setup.

In the experiments with the head phantom, the CCD detector was placed in a fixed position inside the mouth of the head phantom such that it was right behind the teeth which were to be imaged. The head phantom consists of a dry skull which is covered with plastic skin. The x-ray source was mounted on a foldable arm which was used to move the source on an approximately circular arc with distance of  $\sim 590\text{mm}$  from the detector. A photograph of the experimental setup is shown in Figure 5. Note that a full-angle data cannot be measured for reference purposes because of the fixed detector geometry. Using the measurement setup, seven projection images were taken with approximately equal projection intervals

from a total view angle of  $59.7^\circ$ . This represents roughly the maximum view angle that can be used in practice with this kind of fixed detector geometry. A metal ball was attached in front of the teeth, 14mm from the detector array, and the projection angles were estimated on the basis of the observed shift of the reference ball in the images. The left image in Figure 6 shows one projection radiograph from this data set. The location of the image of the metal ball is indicated by a small black arrow above the middle tooth. The right image in Figure 6 shows one slice of the data in sinogram form. Note that the sinogram is truncated from the upper side (i.e, it does not go to zero in the upper side). Thus, in addition of being a limited angle case, the problem contains features of a local tomography problem [7], [8].

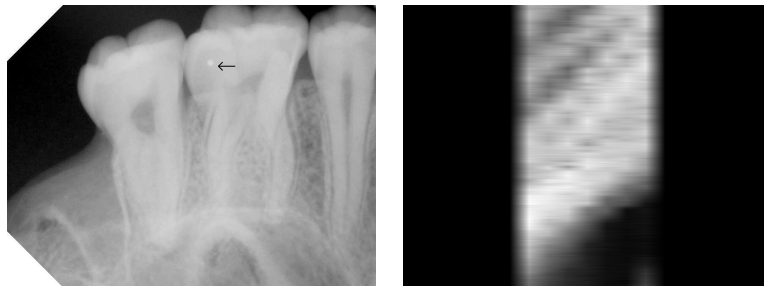


Fig. 6. Left: Intraoral projection radiograph of the head phantom. The location of the image of the metal ball that was used to estimate the projection angles is indicated by a small arrow above the middle tooth. Right: One 2D slice of data in sinogram form. The projections were collected from a total view angle of  $59.7^\circ$ . The black parts on left and right denote the missing parts of the sinogram.

The results for the head phantom case are shown in Figure 7. The left column shows two vertical slices from the backprojected reconstruction, and the right column shows the respective slices from the MAP estimate. The volume of the domain  $\Omega$  in the computations was approximately  $61\text{mm} \times 25\text{mm} \times 26\text{mm}$  and the voxel size in the reconstruction was  $0.23\text{mm} \times 0.23\text{mm} \times 0.23\text{mm}$ . With this resolution, the number of unknowns in the inverse problem was  $M = 265 \times 108 \times 112 \approx 3.2 \cdot 10^6$ . The number of projection data was  $N = 7 \times 436 \times 332 \approx 1 \cdot 10^6$  and the model parameters were the same as in the first test case with the tooth phantom. The computation of the MAP estimate took 76 seconds. As can be seen, the features of the dry skull (bone, enamel, air) are seen more clearly in the MAP estimate than in the backprojection reconstruction.

### C. Case3: Extraoral patient data

The third test case is based on extraoral projection data that were acquired from a male patient using a commercial x-ray scanning device. The test case can be considered as a simple example for the application

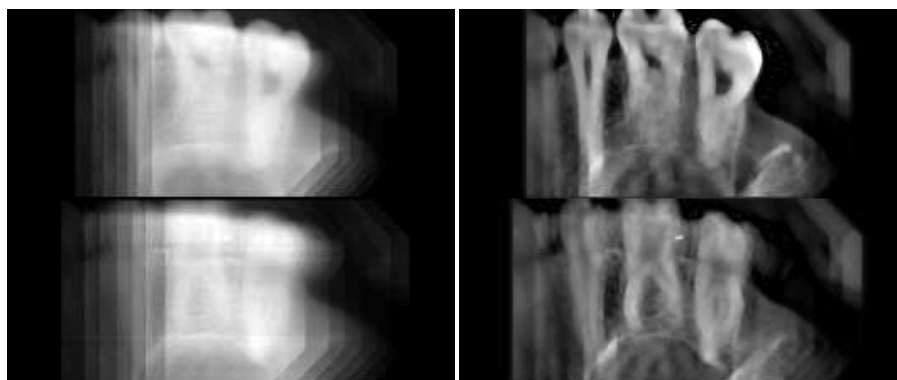


Fig. 7. Limited angle data from the head phantom (7 projections spanning a view angle of  $59.7^\circ$ ). Left column: Vertical slices of a backprojection reconstruction. Right column: Corresponding slices from the 3D MAP estimate.

of limited angle 3D imaging in implant planning. In such situations, it is often important to get accurate measurement for the depth and location of the mandibular nerve canal. In extraoral imaging the source and detector are rotated around the head and images of the region of interest (ROI) are taken through the head. The experimental geometry for extraoral imaging and the projection directions for the experiment are illustrated in Figure 8. In this case, the projection directions were chosen such that the cross section of the nerve canal would be approximately perpendicular to the projection directions and dose to vital organs (brains, medulla) would be as small as possible.

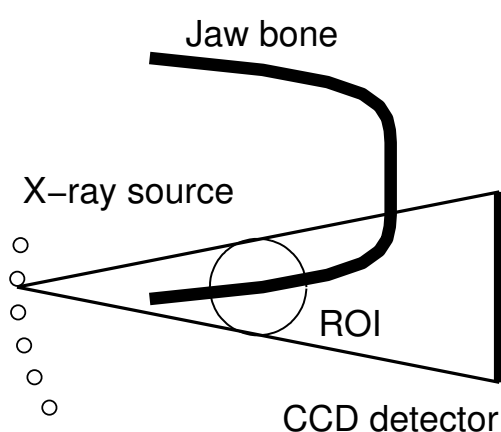


Fig. 8. Left: Geometry for extraoral dental imaging. The source positions are denoted by circles. The thick arc illustrates the jaw bone and dental arc. The region of interest (ROI) is denoted by thin line. For clarity, the detector is depicted only for one source location.

Using the experimental geometry in Figure 8, 11 equispaced projection radiographs of the patient were acquired with a total view angle of  $40^\circ$ . The left image in Figure 9 shows one extraoral projection image of the patient. The size of projection image is  $876 \times 876$  with pixel size  $0.09\text{mm} \times 0.09\text{mm}$ . The projection images were downsampled to half the original size also in this example before transforming them into tomographic data. Thus, the number of data in the inverse problem was  $N = 11 \times 438 \times 438 \approx 2.1 \cdot 10^6$ . The right image in Figure 9 shows one slice of the 3D data in sinogram form. Note that the sinogram reveals clearly both the limited angle and local tomography aspects of the problem.

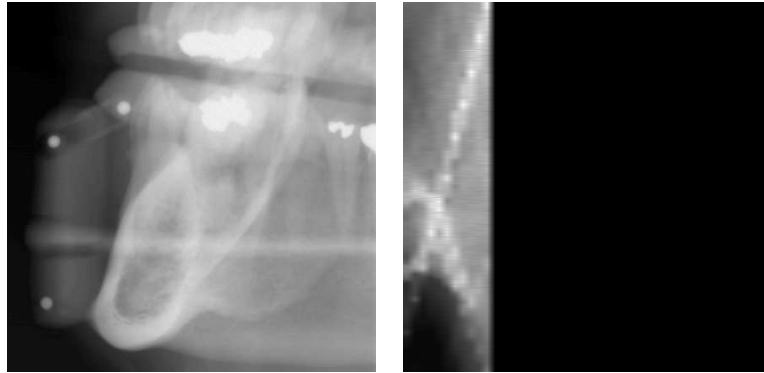


Fig. 9. Left: One extraoral projection radiograph of the patient. Right: One slice of the 3D data in sinogram form.

The results for the extraoral data are shown in Figure 10. Left column shows a vertical slice from the backprojected reconstruction, and the right column shows the respective slice from the 3D MAP estimate. The volume of the domain  $\Omega$  in the computations was approximately  $79\text{mm} \times 79\text{mm} \times 63\text{mm}$  and the voxel size in the reconstruction was  $0.38\text{mm} \times 0.38\text{mm} \times 0.38\text{mm}$ . With this resolution, the number of unknowns in the inverse problem was  $M = 207 \times 207 \times 167 \approx 7.2 \cdot 10^6$ . The model parameters were the same as in the first test case with the tooth specimen. The computation of the MAP estimate took 3 minutes 26 seconds. The location of the mandibular nerve canal in the MAP reconstruction is denoted by an arrow. As can be seen, the nerve canal is seen more clearly in the MAP estimate than in the backprojection image.

The scaling of the computation time in the cluster computer was investigated for the test case with extraoral patient data. The investigated quantities are the scaling of the computation time with respect to the number of central processing units (CPUs) in the computation with fixed number of unknowns, and the scaling of the computation time with respect to the number of unknowns in the problem when all the 13 CPUs of the cluster are used. The top image in Figure 11 shows the computation time with respect to the

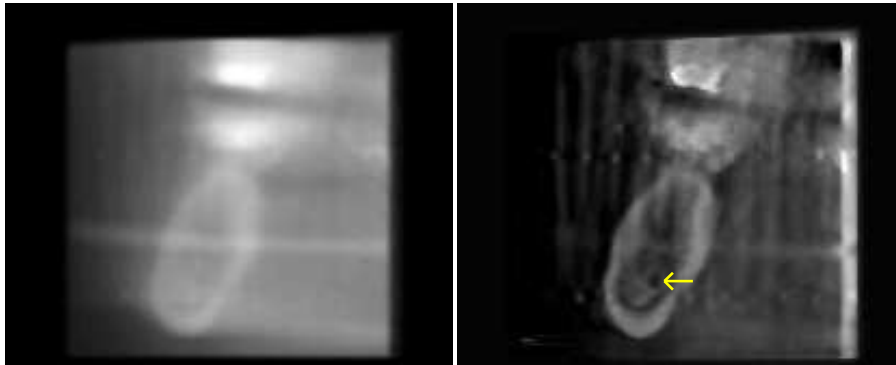


Fig. 10. Reconstructions from extraoral patient data (11 projections spanning a view angle of  $40^\circ$ ). Left column: Vertical slice of a backprojected reconstruction. Right column: Corresponding slice from the 3D MAP-estimate. The arrow denote the location of the mandibular nerve channel.

number of CPUs. The computation times are for the MAP estimate in Figure 10 with number of unknowns  $M = 7155783$ . Note that a minimum of five CPUs are needed for the reconstruction with this resolution. It can be seen that the computation time doesn't scale well with the number of CPUs. The bottleneck is the relatively long latency of the network interconnection between the CPUs, which leaves processors idle while waiting for more data to process. However, the computation time as a function of number of unknowns (bottom image in Figure 11) scales well, which suggests that slightly larger reconstruction grids could be used without the computation time becoming unacceptable long in a clinical setting.

## VI. DISCUSSION

We have proposed a Bayesian model for 3D dental imaging with sparse projection data, and implemented the computation of the MAP estimate for the 3D attenuation function with parallel computing techniques in a Beowulf cluster computer. The results show that the proposed method gives improved image quality over the traditional backprojection techniques, which are currently most widely used methods for 3D dental imaging with limited data.

In the experiments the computational load was reduced by downsampling the projection data. The downsampling was reasonable due to the much finer resolution of the projection images with respect to the clinically sufficient resolution for the 3D reconstructions. However, the reconstructions could, in principle, be computed using full resolution data either by using a larger cluster computer or by using some approximate, less memory intensive data types for the projection data and forward matrix. It was found that the computation time does not scale well with respect the number of CPUs in the current

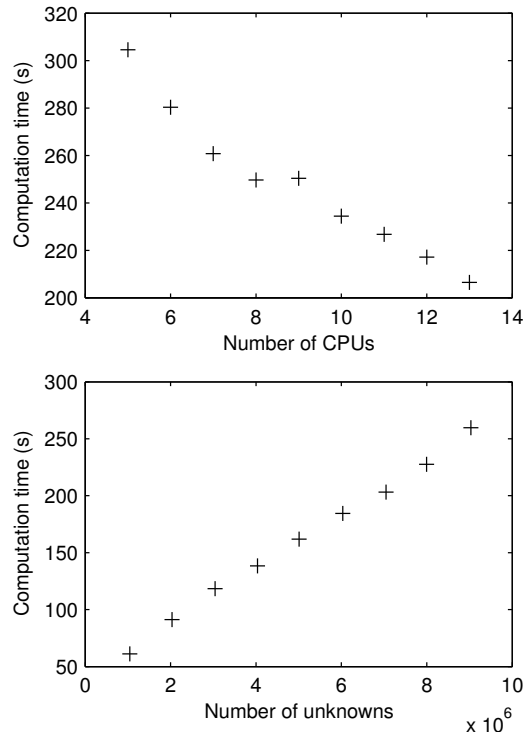


Fig. 11. Top: Scaling of the computation time with respect to the number of CPUs. The computation times are for the MAP reconstruction from the extraoral patient data with number of unknowns  $M = 7155783$ . Note that at least five CPUs are needed for reconstruction with this resolution. Bottom: The scaling of computation time with respect the number of unknowns, when the number of CPUs is held constant at 13.

setup. However, the scaling and execution speed can be improved by installing a low-latency network interconnection between the nodes.

In a more thorough Bayesian analysis, one should compute several estimates and confidence limits from the posterior to make proper statistical inference from the solution. However, the computation of the other estimates typically involve integrals in high dimensions, which, in principle, could be treated by the Markov chain Monte Carlo (MCMC) techniques. However, in case of voxel based models for 3D imaging, these techniques are computationally so demanding that they are not likely to be realizable in the near future. However, the use of MCMC methods for cases with higher level prior models and where structure of the tissue can be well represented in a low-dimensional parametric basis is an interesting topic for future research. Such methods with deformable surface representations have been considered previously in [34].

In this study, all test cases concerned dental radiology. However, it should be noted that the developed

computational framework can be easily modified to other imaging modalities due to the modularity of the Bayesian inversion methods:

- If different types of tissue are imaged with the same measurement setting (same x-ray device and same projection directions), only the prior distribution needs to be changed in the computation.
- If the measurement setup changes, only the likelihood distribution needs to be changed. This change is accomplished by (i) computing matrix  $A$  in Eq. (3) for the new geometry and (ii) approximating the noise characteristics for the new imaging equipment or imaging parameters. It should be noted that any projection geometry can be employed equally easily in the Bayesian framework.

## VII. CONCLUSIONS

Consider the following clinical example of 3D x-ray imaging. A dentist wants to know whether the roots of a certain tooth are close to the inferior dental canal. This question is related to the risk of damaging nerves when removing the tooth. Often a single radiograph is not enough for answering this question because of overlapping of structures. Thus, he takes six digital projection radiographs using a regular x-ray source and a digital x-ray sensor, choosing the directions of the images so that the images of the roots and the nerve canal are clearly separated in some of the images. The projection images, together with knowledge of imaging geometry, are given as input to a reconstruction algorithm. The resulting 3D reconstruction is examined on computer screen and the diagnostic question answered.

The above type of 3D imaging with sparse projection data is not standard practice today. One reason for this is the lack of a flexible, fast, high-quality reconstruction algorithm for such imaging. In this paper, we proposed a Bayesian approach for 3D dental imaging with sparse projection data. The method was tested with data from dental specimens and a patient. In all test cases, the method gave improved image quality over the traditional backprojected reconstructions. With the efficient parallelized implementation, the computation times were in the class of a few minutes. Based on the results, we expect that the proposed method could improve the diagnostic value of the novel type 3D imaging in dental radiology.

## ACKNOWLEDGEMENTS

This work was supported by the National Technology Agency of Finland (TEKES, contract 206/03) and the Academy of Finland (decision number 203985).

## REFERENCES

- [1] L. Brocklebank. *Dental Radiology — Understanding the X-Ray Image*. Oxford University Press, 1997. ISBN 0-19-262411-3.

- [2] A. Ekestubbe, K. Gröndahl, and H-G Gröndahl. The use of tomography for dental implant planning. *Dentomaxillofacial Radiology*, 26:206–213, 1997.
- [3] A. Ramesh, J. B. Ludlow, R. L. Webber, D. A. Tyndall, and D. Paquette. Evaluation of tuned-aperture computed tomography in the detection of simulated periodontal defects. *Oral and Maxillofacial Radiology*, 93:341–349, 2002.
- [4] NewTom3G-dental volumetric tomograph. The Quantitative Radiology (QR) company website [online]. <http://www.qrverona.it/htm/products.htm>.
- [5] 3D Accuitomo XYZ slice view tomograph. J. Morita (USA) Corporation website [online]. Available: [http://www.jmoritausa.com/threeD\\_accuitomo.asp](http://www.jmoritausa.com/threeD_accuitomo.asp).
- [6] CB Mercuray dentomaxillofacial cone beam CT-system. Hitachi Medical company website [online]. Available <http://www.hitachi-medical.co.jp/english/product/cbct/mercuray/index.html>.
- [7] K. T. Smith and F. Keinert. Mathematical foundations of computed tomography. *Appl. Optics*, 24:3950–3957, 1985.
- [8] A. G. Ramm and A. I. Katsevich. *The radon transform and local tomography*. CRC Press, Boca Raton, U.S.A., 1996.
- [9] R. M. Ranggayyan, A. T. Dhawan, and R. Gordon. Algorithms for limited-view computed tomography: an annotated bibliography and a challenge. *Appl. Optics*, 24:4000–4012, 1985.
- [10] F. Natterer. *The Mathematics of Computerized Tomography*. John Wiley & Sons, Chichester, U.S.A., 1986.
- [11] K. M. Hanson. *Bayesian and related methods in image reconstruction from incomplete data*. Image Recovery: Theory and Applications. Academic, Orlando, 1987.
- [12] K. Sauer, S. James Jr, and K. Klifa. Bayesian estimation of 3-d objects from few radiographs. *IEEE Trans. Nucl. Sci.*, 41:1780–1790, 1994.
- [13] K. M. Hanson and G. W. Wecksung. Bayesian approach to limited-angle reconstruction in computed tomography. *J. Opt. Soc. Am.*, 73:1501–1509, 1983.
- [14] C. Bouman and K. Sauer. A generalized gaussian image model for edge-preserving MAP estimation. *IEEE Trans. Image Processing*, 2:296–310, 1993.
- [15] T. Frese, C. Bouman, and K. Sauer. Adaptive wavelet graph model for bayesian tomographic reconstruction. *IEEE Trans. Image Processing*, 11:756–770, 2002.
- [16] D. F. Yu and J. A. Fessler. Edge-preserving tomographic reconstruction with nonlocal regularization. *IEEE Trans. Med. Imag.*, 21:159–173, 2002.
- [17] S. Siltanen, V. Kolehmainen, S. Järvenpää, J. P. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä, and E. Somersalo. Statistical inversion for medical x-ray tomography with few radiographs i: General theory. *Phys. Med. Biol.*, 48:1437 – 1463, 2003.
- [18] V. Kolehmainen, S. Siltanen, S. Järvenpää, J. P. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä, and E. Somersalo. Statistical inversion for medical x-ray tomography with few radiographs ii: Application to dental radiology. *Phys. Med. Biol.*, 48:1465 – 1490, 2003.
- [19] J. Zheng, S. S. Saquib, K. Sauer, and C. A. Bouman. Parallelizable bayesian tomography algorithms with rapid, guaranteed convergence. *IEEE Trans. Image Processing*, 9:1745–1759, 2000.
- [20] K. Mosegaard and M. Sambridge. Monte carlo analysis of inverse problems. *Inv. Probl.*, 18:R29–R54, 2002.
- [21] J. P. Kaipio and E. Somersalo. *Statistical and Computational Methods for Inverse Problems*. Number 160 in Applied Mathematical Sciences. Springer–Verlag, New York, 2004. ISBN: 0-387-22073-9.
- [22] D. C. Dobson and F. Santosa. An image enhancement technique for electrical impedance tomography. *Inv. Probl.*, 10:317–334, 1994.

- [23] D. C. Dobson and F. Santosa. Recovery of blocky images from noisy and blurred data. *SIAM J Appl Math*, 56:1181–1198, 1996.
- [24] D.L. Donoho, I. M. Johnstone, J. C. Hoch, and A. S. Stern. Maximum entropy and the nearly black object. *J Roy Statist Ser B*, 54:41–81, 1992.
- [25] J. Barzilai and J. M. Borwein. Two point step size gradient method. *IMA J. Numer. Anal.*, 8:141–148, 1988.
- [26] M. Raydan. The barzilai and borwein gradient method for the large scale unconstrained minimization problem. *SIAM J. Optim.*, 7:26–33, 1997.
- [27] A. V. Fiacco and G. P. McCormick. *Nonlinear programming: sequential unconstrained minimization techniques*. SIAM, 1990.
- [28] T. Sterling, D. Savarese, D. J. Becker, J. E. Dorband, U. A. Ranawake, and C. V. Packer. BEOWULF: A parallel workstation for scientific computation. In *Proceedings of the 24th International Conference on Parallel Processing*, number I, 1995.
- [29] S. Balay, K. Buchelman, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, B. F. Smith, and H. Zhang H. *PETSc Users Manual*. Argonne National Laboratory ANL-95/11, revision 2.1.5 edition.
- [30] W. Gropp, E. Lusk, N. Doss, and A. Skjellum. A high-performance, portable implementation of the MPI message passing interface standard. *Parallel Computing*, 22:789–828, 1996.
- [31] R. C. Whaley, A. Petitet, and J. J. Dongarra. Automated empirical optimization of software and the ATLAS project. *Parallel Computing*, 27:3–35, 2001.
- [32] R. L. Webber, R. A. Horton, D. A. Tyndall, and J. B. Ludlow. Tuned aperture computed tomography (tact). theory and application for three-dimensional dento-alveolar imaging. *Dentomaxillofacial Radiology*, 26:53–62, 1997.
- [33] R. L. Webber and J. K. Messura. An in vivo comparison of diagnostic information obtained from tuned-aperture computed tomography and conventional dental radiographic imaging modalities. *Oral Surgery Oral Medicine Oral Pathology*, 88:239–247, 1999.
- [34] K. M. Hanson, G. S. Cunningham, and R. McKee. Uncertainty assessment for reconstructions based on deformable geometry. *Int. J. Imaging Syst. Technol.*, 8:506–512, 1997.