Check for updates

Projection-based improvement of 3D reconstructions from motion-impaired dental cone beam CT data

Stefan Niebler[a], and Elmar Schömer
Institute of Computer Science, Johannes Gutenberg University, 55099 Mainz, Germany

Henning Tjaden, and Ulrich Schwanecke
Computer Vision & Mixed Reality Group, RheinMain University of Applied Sciences, 65195 Wiesbaden, Rüsselsheim, Germany

Ralf Schulze
Department of Oral and Maxillofacial Surgery, University Medical Center of the Johannes Gutenberg University, 55131 Mainz, Germany

(Received 25 January 2019; revised 2 July 2019; accepted for publication 9 July 2019; published 30 August 2019)

Purpose: Computed tomography (CT) and, in particular, cone beam CT (CBCT) have been increasingly used as a diagnostic tool in recent years. Patient motion during acquisition is common in CBCT due to long scan times. This results in degraded image quality and may potentially increase the number of retakes. Our aim was to develop a marker-free iterative motion correction algorithm that works on the projection images and is suitable for local tomography.
Methods: We present an iterative motion correction algorithm that allows the patient's motion to be detected and taken into account during reconstruction. The core of our method is a fast GPU-accelerated three-dimensional reconstruction algorithm. Assuming rigid motion, motion correction is performed by minimizing a pixel-wise cost function between all captured x-ray images and parameterized projections of the reconstructed volume.
Results: Our method is marker-free and requires only projection images. Furthermore, it can deal with local tomography data. We demonstrate the effectiveness of our approach on both simulated and real motion-beset patient images. The results show that our new motion correction algorithm leads to accurate reconstructions with sharper edges, better contrasts and more detail.
Conclusions: The presented method allows for correction of patient motion with observable improvements in image quality compared to uncorrected reconstructions. Potentially, this may reduce the number of retakes caused by corrupted reconstructions due to patient movements. *© 2019 The Authors. Medical Physics published by Wiley Periodicals, Inc. on behalf of American Association of Physicists in Medicine.* [https://doi.org/10.1002/mp.13731]

Key words: CBCT, motion, tomography

1. INTRODUCTION

Since its invention in the 1970s, computed tomography (CT) has revolutionized medical imaging. At the end of the 1990s, a variant of computer tomography, the so-called cone beam CT (CBCT), was developed, which has since spread widely in dentistry, oral and maxillofacial surgery as well as in otorhinolaryngology. Today, millions of CT images are taken worldwide each year to answer a variety of clinical questions in all areas of medicine.

Unfortunately, both CT and CBCT suffer from blurring or artifacts caused by the patient's movement during the acquisition. In addition to blurring, other artifacts such as Moiré patterns, double contours and stripes may also appear.[1,2] In particular CBCT suffers from strong movement influences due to long scanning times ranging between approx. 5 and 60 s.[3] The amount of significant patient motion was estimated to be at least 20%.[4–6] It has been suggested that a high percentage (>80%) of the patients move more than 0.5 mm during the scan, with movements more than one millimeter occurring in 70% of all cases.[7] Thereby, even movements of the patient in the low single-digit millimeter range have a negative influence on the quality of the volumetric reconstruction.[8,9] In addition to active patient movements, similar problems can also occur due to normal organ activity (heartbeat) or respiratory movements.[10] Even the heartbeat of healthy young sitting patients whose chin rests on a chin support alone leads to amplitudes of about 80 μm per heartbeat.[11] This is within the range of the voxel size of modern "high-resolution" dental CBCT devices.

Patient movements lead to discrepancies between projection and assumed back projection geometry, resulting in inconsistencies in the reconstruction process. Reconstruction methods commonly used such as filtered back projection (FBP) or the Feldkamp-Davis-Kress (FDK) algorithm do not take these movements into account. This results in inaccurate, artifact-laden reconstructions. Therefore, new methods must be developed to identify the patient's movement. Such methods can build the basis to effectively reduce the dose of radiation by reducing the number of CT scans that need to be

© 2019 The Authors. Medical Physics published by Wiley Periodicals, Inc. on behalf of American Association of Physicists in Medicine. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

repeated due to patient movements. For periodic movements, such as heartbeat[12] or breathing[13,14] motion, special methods were developed that utilize prior knowledge about the periodicity of the motion to improve the three-dimensional (3D) reconstruction process. Unfortunately, these methods cannot be used for general nonperiodic motions.

Several approaches have been developed to solve the general motion problem. Some of them are based on external markers that can be identified in the projection images and then used to correct patient movement.[15] Such marker-based approaches have also been extended to more complex, nonrigid deformations.[16] Since such markers are typically made of radio-opaque material for easy recognition, they can cover parts of the tissue, which may lead to new artifacts in the reconstruction. Other methods use beforehand generated patient models for image adjustment.[17–19] However, the necessity of an existing patient model is a huge impediment to the general usage of these approaches.

Ens[20] gave a comprehensive comparison of marker-based and marker-free approaches for motion correction. Based on sharpness criteria and gradient distribution, she developed different metrics, that can be used as a quality measure for a cost function to be minimized, which in turn determines the extent of motion artifacts in the reconstruction. The results are heterogeneous, with the best results obtained for a metric based on normalized gradients. Similar to our approach proposed in this paper, general image quality measures such as entropy or sharpness were rather successfully used in other iterative optimization approaches to calibrate CBCT-systems.[21,22] Aichert et al. developed a method to solve this problem using a epipolar consistency criterion to align the image projections relative to each other.[23] However, the relationship between the sum of intensities along each two epipolar lines in x-ray images does not hold for local tomography problems. Local tomography is typical for dental CBCT since usually a small volume [the region of interest (ROI)] out of the entire head is being imaged. This poses specific challenges as the measurement along a ray through the entire object (the head) does not only correspond to the absorption in the reconstructed ROI but also to the absorption of the surrounding region (see, e.g., Ref. [24,25]). Another recent work provides a general marker-free framework for the detection of geometric errors (e.g., resulting from patient movement) from the projection images including an iterative reconstruction procedure.[26] Analogous to our approach, the authors model these errors based on rigid motions with six degrees of freedom and describe the combined alignment and reconstruction as a nonlinear least squares problem. They also discussed a solution for local tomography. However, they only present simulated data using parallel projection (instead of real cone beam imagery) for validation plus an experimental electron tomography dataset of a rather simple phantom. Special applications, such as dental CBCT, are explicitly not addressed by the authors. Additionally the authors focused on minimizing the linear equation system while we use a more general approach by using image similarity measurements as a way of estimating motions. Sisniega et al. developed another technique for motion compensation minimizing the blur of a selected ROI.[8] The authors state that they only simulated piecewise linear motions of the patient to evaluate their synthetic data. Our approach on the other hand was tested on arbitrary rigid motion. Furthermore, Sisniega et al. do not deal with the local tomography problem in their work. In microtomography purely image-based methods such as those proposed by Latham et al. have been developed to estimate displacements of the object during the scan.[27] They identify features on a reprojection and the original images and calculate the translational difference between these images by mapping these features from one image to the other. Unlike our approach they do not take rotational motion into account as they claim that the three degrees of translational freedom are sufficient for motion estimation in micro-CT.

In this work, we propose a method that uses a comparison between the machine generated projection images and computed projections generated by changing the motion parameters that describe the patients rigid motion. The parameters thus determined are used to obtain a more accurate volumetric reconstruction with sharper edges. The proposed method uses only the acquired image data in combination with the precalibrated projection geometry of the scanning device. It does neither require any additional markers nor does it rely on a patient model and is applicable for the particularly challenging local tomography problem. In the following we briefly review the geometry of a (dental) CBCT and the image formation process. The motion model is then presented, followed by a detailed description of the entire reconstruction process with a brief overview of the local tomography problem. Finally we present an evaluation of our method using both synthetic and real-world motion-beset patient data.

2. MATERIALS AND METHODS

2.A. Background theory

In CBCT an x-ray source rotates on a circular path around a nonmoving static object $\tilde{x}$, creating $n \in \mathbb{N}$ projection images $b^{(i)}$, $i \in \{1, \ldots, n\}$ each with a resolution of $w \times h$, $w, h \in \mathbb{N}$ pixels. This results in a total of $n{\cdot}w{\cdot}h$ pixels b_c, $c \in \{1,\ldots,n{\cdot}w{\cdot}h\}$, each associated with an x-ray r_c and its corresponding path S_c. When the rays emitted by the source pass through the object $\tilde{x}$, they are attenuated depending on the attenuation coefficient of the matter passed through. This results in a different photon count in each detector cell represented by an image pixel. According to Lambert-Beer's law, a simplified monochromatic model to calculate the measured photon count I_c for a ray r_c is given as

$$I_c = \tilde{I} \cdot exp\left(- \int_{S_c(\tilde{x})} \tilde{x}\, ds \right), \tag{1}$$

where $S_c(\tilde{x})$ is the path of the ray r_c through the volume $\tilde{x}$ and $\tilde{I}$ is the initial photon count emitted from the source. Taking the logarithm of the ratio $I_c/\tilde{I}$, from Eq. 1 we obtain

$$f_c(\tilde{x}) = -\log\left(\frac{I_c}{\tilde{I}}\right) = \int_{S_c(\tilde{x})} \tilde{x}\, ds \approx b_c \tag{2}$$

as the integral of the attenuation along the path of a ray. Combining Eq. 2 for all rays r_c results in the function

$$F(\tilde{x}) = b \tag{3}$$

describing the mapping of the object to all measurements. In a static scenario, that is, the object $\tilde{x}$ does not move, the acquisition function F is fully determined by the geometry of a CBCT device. If $\tilde{x}$ moves during acquisition F changes into an unknown function F_{mov}. Thus, to get an optimal reconstruction in the case $\tilde{x}$ is moving, F_{mov} has to be determined.

In the following we present an iterative two phase algorithm to approximately reconstruct $\tilde{x}$, considering object motion. Based on the conjugate gradient (CG) method used for the reconstruction, we approximate a discretized version of F_{mov} in each iteration to obtain a more accurate reconstruction of $\tilde{x}$. We call our new approach *motion corrected conjugate gradient* (MCCG) method.

2.B. Models and reconstruction methods

2.B.1. Volume reconstruction method

Our aim is to reconstruct a discrete approximation of the object $\tilde{x}$. Therefore, we split $\tilde{x}$ into a volume of $m \times m \times m$ voxels $v^{(j)}$, $j \in \{1, \ldots, m^3\}$, where $m \in \mathbb{N}$ denotes the number of voxels in each dimension. Each voxel is assigned the attenuation value of $\tilde{x}$ at the voxel position. Each pixel of the projected images corresponds to a single ray and the integral in Eq. 2 turns into a sum b_c of weighted attenuation coefficients $a_{c,j}$ for each $v^{(j)}$ along the path S_c of a ray r_c. These attenuation coefficients are aggregated in the *system matrix* $A \in \mathbb{R}^{n\cdot w\cdot h \times m^3}$ describing the image acquisition process. Together with $b = (b_1, \ldots, b_{n\cdot w\cdot h})^\top \in \mathbb{R}^{n\cdot w\cdot h}$ this results in the linear system

$$A\tilde{x} = b \tag{4}$$

as a discretized version of Eq. 3.

The system matrix A can be determined based on the specifications of the CBCT-machine. If A is known the least squares problem corresponding to Eq. 4 can be solved, for example, by using the conjugate gradients least squares (CGLS) method.

2.B.2. Motion model

Patient motion during acquisition induce artifacts such as blurring or multiple contours in the reconstructed volume.[1] We can interpret the patient motion as changes in our equation system A instead of changes to $\tilde{x}$. As a result we can see that the artifacts are based on the fact that patient motion changes Eq. 4 into

$$A_{mov}\tilde{x} = b_{mov} \tag{5}$$

with unknown disturbed acquisition matrix $A_{mov} \neq A$. To obtain an optimized reconstruction of $\tilde{x}$ without motion artifacts, A_{mov} has to be determined.

In the following, we only consider rigid motions of the patient. The simplifying assumption of rigid movement is justified because we focus on the reconstruction of bones and teeth, both structures that are mainly exposed to rigid motion. Thereby, a *motion field*

$$\tilde{p} = [p^{(1)}, \ldots, p^{(n)}] \tag{6}$$

represents the patient motion in form of a sequence of poses (orientations and positions)

$$p^{(i)} = (\phi^{(i)}, \theta^{(i)}, \gamma^{(i)}, t_x^{(i)}, t_y^{(i)}, t_z^{(i)}) \tag{7}$$

for each x-ray image $b^{(i)}$. Here, $\phi^{(i)}$ and $\theta^{(i)}$ describe the first two polar coordinates in camera space, $\gamma^{(i)}$ the rotation of the resulting image $b^{(i)}$, and $t_x^{(i)}, t_y^{(i)}, t_z^{(i)}$ the displacements along the axes of the camera coordinate system. Unfortunately, translations along the projection direction cannot be reliably detected in the images for small cone angles (i.e., for almost parallel x-rays). By choosing the above motion parametrization, this uncertainty is limited to $t_z^{(i)}$.

The rotation of the CBCT device around the y-axis for each image is denoted by $\delta^{(i)}$, which also corresponds to the rotation from world coordinates to camera coordinates (see Fig. 1). We now assume that the disturbed system matrix A_{mov} depends on a motion field $\tilde{p}$, that is, $A_{mov} = A(\tilde{p})$ and $A(0) = A$.

2.B.3. Motion reconstruction

Once we obtained an estimation x of the object to reconstruct $\tilde{x}$ we use it to create an approximation of the unknown matrix A_{mov}. Therefore, we calculate the camera pose for each image $b^{(i)}$ described by the motion parameter vector $p^{(i)}$ which minimizes the difference between $b^{(i)}$ and the projection $A^{(i)}(p^{(i)}, \delta^{(i)})x$. Here, $A^{(i)}(p^{(i)}, \delta^{(i)})$ is the matrix that describes the single projection to create an image with a given source position and ray direction described by $p^{(i)}$ and gantry rotation angle $\delta^{(i)}$. As we keep gantry rotation angles

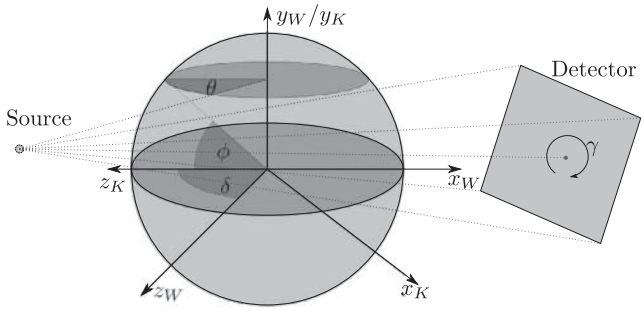

FIG. 1. Cone beam computed tomography parameterization: camera coordinate system (x_K, y_K, z_K) defined in world coordinate system (x_W, y_W, z_W) with δ being the system rotation around the y_K axis whereas ϕ,θ and γ describe the patient motion interpreted as additional camera movement. Thereby ϕ and θ can be interpreted as polar coordinates in camera space and γ denotes a rotation of the resulting image. Even though θ and δ describe a rotation around the $y_W = y_K$ axis only δ influences the translation directions.

fixed during our algorithm we can drop the $\delta^{(i)}$ as it is implicitly described by the index of the image i. As such we obtain $A^{(i)}(p^{(i)}, \delta^{(i)}) = A^{(i)}(p^{(i)})$

We measure the fit of a set of parameter for each image $b^{(i)}$ with the cost function $E_{x,b^{(i)}}(p^{(i)})$ that is explained in more detail in the next section. To obtain the sought motion parameter vector $p^{(i)}$ we solve

$$p^{(i)} = \underset{p}{\operatorname{argmin}}\, E_{x,b^{(i)}}(p) \tag{8}$$

for each image $b^{(i)}$, where for example

$$E_{x,b^{(i)}}(p^{(i)}) = \|A^{(i)}(p^{(i)})x - b^{(i)}\|_2, \tag{9}$$

could be the Euclidean distance between $b^{(i)}$ and $A^{(i)}(p^{(i)})x$. A derivative free approach can be used to solve the optimization problem Eq. 8.[26] In our implementation we chose the Nelder-Mead method.

2.B.4. The cost function

An essential point of the proposed approach is to find a good cost function $E_{x,b^{(i)}}(p)$. Ideally, the function has a unique global minimum at the sought pose and no further local minima. Under the assumption that the patient movement is within a small region around the optimum, it is sufficient if the local minimum of the cost function is unambiguous in this area. Since the cost function has to be evaluated several times per image, it should also be possible to evaluate it efficiently. For this purpose the Euclidean distance (ED) [see Eq. 9], the mutual information (MI)[28] and the structural similarity measure (SSIM)[29] between $b^{(i)}$ and $A^{(i)}(p^{(i)})x$ were used as cost functions. MI was not suitable for stronger motions due to several local minima and a comparably narrow valley near the sought pose which makes it hard to optimize. Thus, we focused on the two other measures. We evaluated two different versions of SSIM. One version (SSIM_w) calculates the average SSIM of 7×7 windows between a reprojection and the original radiograph. The other version (SSIM_c) calculates the SSIM of the whole radiographs instead. SSIM_w was unable to correct motion artefacts unless the initial reconstruction is already reasonably good. Thus, it is not suitable for our task. SSIM_c and ED are both suitable and return almost identical results. We'd like to note that when using *ED* on complete tomography data this resembles Algorithm 3 proposed in the work of van Leeuwen et al.[26] However with the adjustments proposed in Section 2.B.6 the methods differ for local tomography problems.

Plotting the Euclidean distance of a single image and a reprojection from poses where single parameters were altered within a range around the ground truth pose shows that even for motion impaired reconstructions (reconstructions that were not corrected for any motion yet) the cost function has a unique local minimum (see Fig. 2). The reason why the minima are not located at 0 (ground truth pose) is that the reconstruction is oriented differently than the original volume. Thus the camera pose has to be adjusted as well and cannot be expected to align with ground truth (see Section 3). The displayed plot excluded a border of 25 pixels from each image during comparison to account for local tomography (see Section 2.B.6). When using the original volume the unique minimum is located at the ground truth position.

2.B.5. The overall reconstruction model

Similar to the work of Van Leeuwen et al. we use an interleaved two phase reconstruction loop where in each iteration k we first reconstruct a new volume x_k by solving $A_{k-1}x_k = b$ and then calculate a new matrix A_k while keeping x_k fixed.[26] The latter is done by minimizing Eq. 8 using the derivative-free Nelder-Mead algorithm to compute updated pose parameters $p_k^{(i)}$ for each image $b^{(i)}$ and then combining the resulting matrices to obtain $A_k = A(p_k) = (A^{(1)}(p_k^{(1)}), A^{(2)}(p_k^{(2)}), \ldots, A^{(n)}(p_k^{(n)}))$. Nelder-Mead was chosen because it is hard to calculate the derivatives analytically or numerically. MCCG starts with $p_0 = 0$ and $x_0 = 0$, meaning that, after the first iteration we obtain a reconstruction without motion correction. We stop iterating MCCG when $1 - (\frac{\|A_{k-1}x_k - b\|_2}{\|A_{k-1}x_{k-1} - b\|_2}) < \epsilon$ for $0 \leq \varepsilon < 1$ or after a maximum number of N iterations has been executed. CGLS stops iterating when $1 - (\frac{\|Ax_l - b\|_2}{\|Ax_{l-1} - b\|_2}) < \frac{\epsilon}{2}$ for CGLS iteration l. The proposed reconstruction algorithm is described in Algorithm 1. In line 4 the CGLS method is used to reconstruct a volume by minimizing the residual

$$\|A_{k-1}x - b\|_2^2 + \lambda\|x\|_2^2, \tag{10}$$

for a given regularization parameter $\lambda \in \mathbb{R}$. In line 10, for each image the cost function Eq. 9 is minimized using Nelder-Mead.

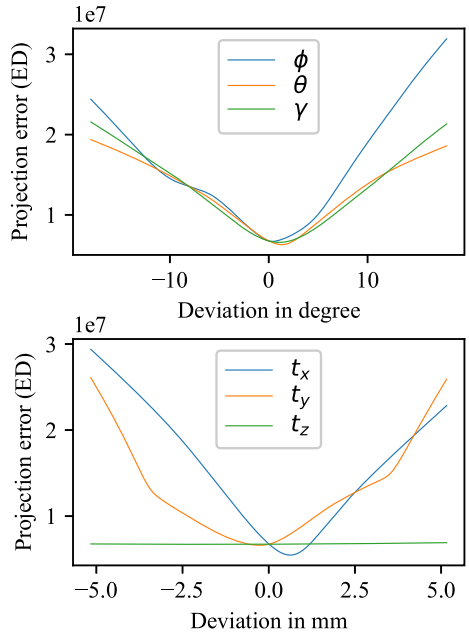

FIG. 2. Typical error curve between original image and a projection from a reconstruction deviating from ground truth due to rotation (top) and translation (bottom) for Euclidean distance. [Color figure can be viewed at wileyonlinelibrary.com]

Algorithm 1: MCCG reconstruction algorithm. Starting with an uncorrected reconstruction, in every iteration a volume x_k is reconstructed using the CGLS algorithm and used to derive new motion parameters p_k by solving the minimization problem with the Nelder-Mead algorithm. The motion parameters p_k are used to define the new acquisition matrix A_k.

Input: measurements b, regularization parameter λ, number of max iterations N **Output:** reconstruction x

$x_0 = 0,\ p_0 = 0,\ A_0 = A(p_0)$
for $k = 1, \ldots, N$ **do**
  //reconstruct x (max 30 iterations)
  $x_k = CGLS\,(A_{k-1}, b, x_{k-1}, \lambda)$
  if $1 - \frac{\|A_{k-1}x_k - b\|_2}{\|A_{k-1}x_{k-1} - b\|_2} < \epsilon$ **then**
    break //stop iterating if the error does not decrease further
  end
  for $i = 1, \ldots, n$ **do**
    //Determine image pose $p^{(i)}$ using Nelder-Mead (max 500 iterations)
    $p_k^{(i)} = \arg\min_p(E_{x,b^{(i)}}(p))$
  end
  $p_k = (p_k^{(1)}, \ldots, p_k^{(n)})$ // concatenate individual poses to motion field
  $A_k = A(p_k)$ //update system matrix
end

2.B.6. Adjustment to local tomography

Local tomography[25] introduces additional problems since parts of the image information originates from structures outside the region of interest (ROI). Unfortunately, we cannot simply separate the projection information coming from such external structures. To apply our motion correction algorithm we reconstruct on a coarser grid. Here the ROI is located in the central cube of size 250^3 voxels. The remaining area is then used to reconstruct the surrounding region. This allows us to use our motion correction algorithm as before. However, using this approach, we will obtain regions in our reconstruction area without any data as A contains 0-vectors as columns of those voxels. Rays traversing these voxels cannot gather information from these voxels. Thus the corresponding pixel will receive comparably low intensity values. During motion estimation this will drastically increase the cost of the pose. As displayed in Fig. 3(c) the sought pose can have a higher cost than other poses. To overcome this limitation we define a border of fixed width at each side and exclude pixels within the border from image comparison during motion estimation. As a result poses with low information on the border pixels are no longer penalized with a high cost and optimization can obtain better results [Fig. 3(d)]. The border size can be selected relatively small (even 0) as subsequent iterations of the reconstruction step will recreate information to low information voxels if they are traversed by a ray. However, using a low border size corrections will only happen very slowly while a higher border size leads to a faster convergence speed (see Fig. 4) as larger correction steps are possible. On the other hand cutting too much information decreases the alignment quality.

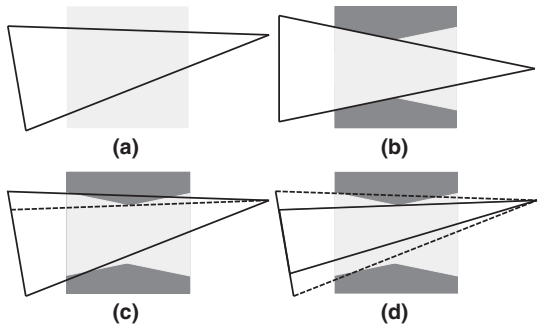

FIG. 3. The local tomography problem. Light gray areas contain positive values while dark gray areas contain zeros. The triangle outlines the image acquisition rays. (a) The pose during image acquisition. (b) The initial state for the first motion correction. (c) Rays above the dotted line traverse an unreconstructed area containing zeros, which results in a high cost. (d) Excluding rays from image comparison removes the cost risen from traversing the zero valued area.

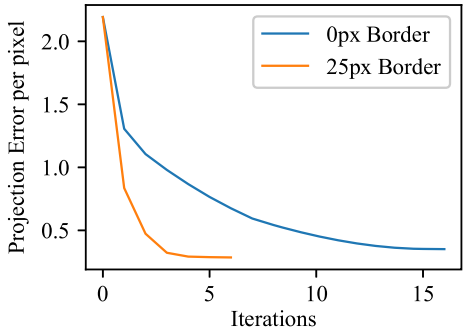

FIG. 4. Per pixel projection error after each motion correction step of MCCG for different border sizes. Using a border improves motion correction and overall convergence speed. [Color figure can be viewed at wileyonlinelibrary.com]

2.C. Implementation

Our implementation is based on the CUDA programming framework to make use of GPUs for parallel computation. The matrix A in general has no symmetries but is sparse with approximately $\mathcal{O}(m)$ (m is the size of the volume in each dimension) non-zero entries per row. We use a matrix free optimization approach by simulating projection and backprojection by tracing rays throughout the volume. Thereby, the volume is scanned in small intervals along each ray and the values at each point are calculated by trilinearily interpolating the values of the surrounding voxels. To calculate the matrix multiplication with A the interpolated values at each scan point are summed along the path of a ray to get the pixel value. For CGLS it is also necessary to multiply with A^T. For this, the ray starts with a value according to the value of its corresponding pixel and at each scan point along the path of the ray the intensity value is broadcast to all surrounding voxels weighted by the corresponding trilinear interpolation coefficients.

To decide when to stop iterating we chose $\varepsilon = 0.01$ (0.005) for synthetic (real) data. In local tomography when borders are present the residual in CGLS is only evaluated on pixels that are not located on the border. This is done due to the fact that our motion correction determines new poses without considering these pixels. Thus they can yield a high projection error, which leads to a premature stop of the iteration.

For the Nelder-Mead method we located one point of the simplex at the current pose for this image and construct the simplex by adding another point along each parameter axis. For translational (rotational) axes these points are set at a distance of 1 mm (1°) from the first point. Nelder-Mead is stopped when $2 \cdot (\frac{\|y_{best} - y_{worst}\|_2}{\|y_{best} + y_{worst}\|_2})$ is smaller than a tolerance value of 10^{-6}, with y_{best} and y_{worst} being the best and worst evaluations of the cost function at the points of the current simplex. Alternatively it stops iterating after a maximum of 500 iterations. The latter was chosen as on average only 150 to 200 iterations were required for convergence. Decreasing the tolerance value did not improve results while increasing runtime. Other simplex setup methods, for example, always setting up the simplex at the 0 position or around the final location of the previous image, were tested. However these methods have either increased runtime without improving the results, or even worsened the results, as outliers affect subsequent pose estimations. SSIM makes use of two coefficents C_1 and C_2 to stabilize for small denominators. When performing SSIM for image comparison the dynamic range was chosen individually for each $b^{(i)}$ as $L = \max b^{(i)} - \min b^{(i)}$. We then chose $C_1 = (0.01L)^2$ and $C_2 = (0.03L)^2$. In all pose estimation methods the parameter t_z was omitted as it cannot be reliably estimated due to the projection geometry. We model patient motion as motion of the x-ray source by inverting the transformation matrix used to describe the patient's movement.

The resulting rigid motion transformation of the x-ray source $T_\delta \in \mathbb{SE}(3)$ is described by

$$T_\delta(\gamma, \phi, \theta, t_x, t_y, t_z) = \begin{bmatrix} R_{y_w}(\delta) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R & -Rt \\ 0 & 1 \end{bmatrix} \tag{11}$$

with $R = R_{y_w}(\theta) R_{x_w}(\phi) R_{z_w}(\gamma) \in \mathbb{SO}(3)$ and $t = (t_x, t_y, t_z)^T \in \mathbb{R}^3$. The matrices $R_a(\alpha)$ describe a rotation around the axis a by α and $R_{y_w}(\delta)$ the rotation from world space into camera space (see Fig. 1). The multiplications with the matrices $A(p_k)$ can then be derived by simply adjusting the base position and orientation of the x-ray source and detector for each image according to Eq. 11 before applying the projection algorithm.

2.D. Data

We evaluated our approach with various synthetic and real local tomography datasets. The synthetic datasets consist of projections generated from 3D volumes from CT acquisitions. The real datasets were obtained as raw data output from a 3D Accuitomo 80 (J. Morita Corp, Kyoto, Japan). In the following sections, we briefly describe all datasets used as well as the performed preprocessing and evaluation methods.

2.D.1. Synthetic data

Our synthetic datasets each consist of 516 images with equidistant angles and an independently generated motion field using a random walk for each motion parameter. Each image set shows a full angle (360°) tomography with an image resolution of 400 × 400 pixels. All volumes used for image acquisition had a resolution of 500^3 voxels. We generated a total of 9 random walk motion fields applied to 3 different CT volumes resulting in a total of 27 synthetic random walk motion datasets. Similar to Sisniega et al.[8] we created additional datasets where the patient is stationary for the first 129 images, then translates between 1 and 10 mm into a new position during the next 86 images and stays stationary for the remaining images. Thus, we obtained 20 additional piecewise linear motion datasets. Motion fields were divided into three motion intensity groups. Groups are seperated by the maximum range of translation along at least one axis of the world coordinate system. In Table I this is noted as translational range. Rotational range describes the minimum and maximum range of rotational parameters within this group. (see Table I). In total we have 14 datasets in the low motion group, 12 for the medium motion group and 11 datasets in the high motion group.

Also note that our ROI was assigned a size of 50 mm. As all our datasets show motion of at least 1 mm, the reconstruction quality should be negatively affected.[30,31] It is estimated that in more than 90% of the cases where motion is present during CBCT the patient moves < 4 mm.[7] Such motion is covered by our low and medium motion groups, while the high motion group contains either strong or even excessive motion amplitudes.

2.D.2. Real data

Real data were obtained as CBCT raw data output of a 3D Accuitomo 80 at the University Medical Center of the Johannes Gutenberg University Mainz.[33] All real datasets represent imaging data of real patients undergoing an actual CBCT scan for different clinical reasons. Thus all motion present in these datasets stems from real patient motion. Images at the beginning and at the end of an image series were excluded from reconstruction as they contained either not enough information or too much noise. The resulting datasets consist of 516 images, corresponding to a full rotation of the

TABLE I. The classification of motion fields into groups according to the rotation and translational motion range.

Dataset	Rotational range	Translational range
low	2° to 8°	1 mm to 2.5 mm
medium	2.9° to 6.3°	3 mm to 4.5 mm
high	7° to 14°	more than 6 mm

machine. Machine output images contain borders without information as well as noise. In a preprocessing step we cut images to remove these borders and used an adapted bilateral filter similar to Manduca et al.[32] to reduce noise. We adjusted the filter range to increase smoothing on darker regions. Lambert-Beer's law was applied to each pixel value to transform raw photon count data into the sum of attenuation coefficients. Noise in the detector signal can lead to a measured photon count of 0. As we need to take the negative logarithm of the photon count during the transformation this would result in a value of ∞. To prevent this all pixels with a photon count of 0 were set to the average of their surrounding pixels.

2.E. Quality measure

We used three different metrics for evaluation: (a) Line profiles within slices of the volume to measure motion blur and detail, and on synthetic datasets additionally (b) the projection error (ED) [see Eq. 9] to measure data compatibility, and (c) the structural similarity index (SSIM) between reconstructed and original ROI. We also compared reconstructed and original motion parameters to rate the accuracy of the reconstruction of the motion. As there is almost no visual difference in the results obtained from using MCCG with either ED or $SSIM_c$ as cost function, we only display the results for ED and report results from $SSIM_c$ only in quantifications.

3. RESULTS

In this section we present the results obtained for the synthetic as well as the real local tomography datasets. At first, a short study on different parameter values for the border size and regularization has been evaluated on a subset of the synthetic data. Unless otherwise stated, MCCG refers to the version with Euclidean distance used as a cost function for pose estimation. We perform up to 20 iterations of MCCG per dataset. Within each iteration a maximum of 30 iterations of CGLS are performed. To keep the results of the uncorrected version comparable to those of MCCG, we perform up to $20 \cdot 30 = 600$ CGLS iterations in the uncorrected version. All iterations can stop earlier if convergence is reached. The volume size was set to $m = 500$. The geometry was chosen such that the ROI-cylinder has a height and diameter of 50 mm resulting in voxels with an edge length of 200 μm for local tomography data. The ROI of local tomography data is only reconstructed at a size of $m = 250$ voxels per dimension and located in the center of the whole reconstruction cube. Surrounding voxels are used to display regions outside the ROI as described in Section 2.B.6.

Synthetic parameter study: The average SSIM values of reconstructions of a dataset from the low motion group for different border sizes and regularization parameters are reported in Tables II and III. All other tests have been evaluated with a border size of 25 as well as $\lambda = 100$. $SSIM_w$ had a maximum SSIM of 0.844 which is 11% worse than the other methods and is not further evaluated in this section. With a border size of 25 the iterations necessary for convergence decreased from an average of 14.58 (13.89) to 8.83 (7.39) on low (medium) motion random walk datasets.

3.A. Synthetic data

Figure 5 shows a qualitative visual comparison between a non-motion-corrected reconstruction using CGLS and the proposed MCCG reconstruction for axial, sagittal and coronal slices within a predefined ROI. The displayed results are based on a dataset of the low motion intensity group.

TABLE II. Structural similarity measure (SSIM)-value of reconstruction and original volume for different cost-functions and border sizes in pixel. λ was set to 0.

	Bordersize in px				
Cost-function	0	10	25	50	100
ED	0.940	0.951	0.958	0.956	0.948
$SSIM_c$	0.940	0.951	0.961	0.953	0.945
$SSIM_w$	0.831	0.835	0.815	0.843	0.844

TABLE III. Structural similarity measure (SSIM)-value of reconstruction and original volume for different cost-functions and regularization parameter λ. Bordersize was set to 25.

	λ				
Cost-function	0	10	100	1000	10000
ED	0.958	0.960	0.958	0.952	0.774
$SSIM_c$	0.961	0.960	0.961	0.951	0.792
$SSIM_w$	0.815	0.819	0.815	0.830	0.797

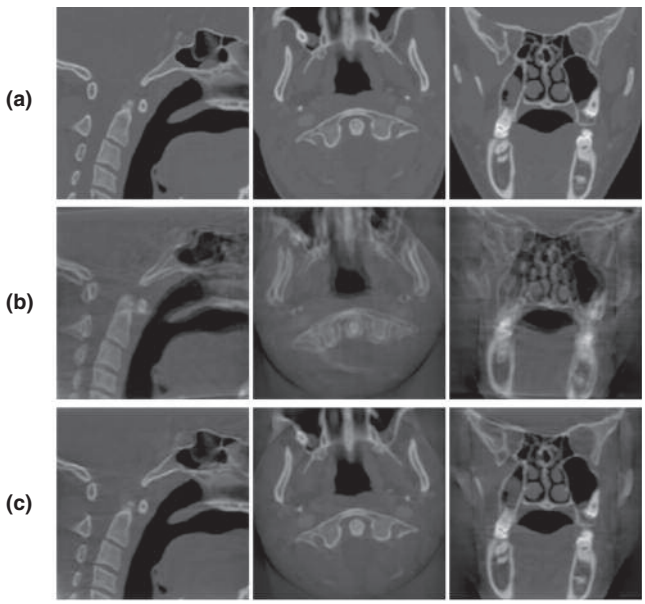

FIG. 5. Same region of interest containing ground truth (a) and a reconstruction using CGLS only (b) or motion corrected conjugate gradient (c) respectively.

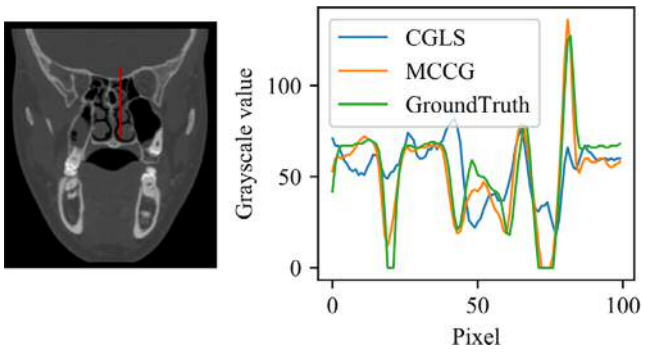

FIG. 6. Line profile within the same slice of the ground truth and reconstructed volumes. [Color figure can be viewed at wileyonlinelibrary.com]

Figure 6 presents line profile plots along a line within the same slice of the reconstructed volume as depicted in the third column of Fig. 5. As the reconstructions could be oriented differently, we aligned the volumes by rigidly transforming the MCCG and CGLS reconstructed volume to best match the ground truth volume by solving

$$\min_{(\gamma,\phi,\theta,t_x,t_y,t_z)} \|T(\gamma,\phi,\theta,t_x,t_y,t_z)x - \tilde{x}\|_2. \quad (12)$$

Here T is a transformation matrix as seen in Eq. 11 with δ set to 0. The reconstructed ROI lies within the central 250^3 voxels of the whole volume. Due to patient motion the ROI does not have a perfect cylindrical shape. To compensate for this only voxels within the central cylinder of diameter and height of 240 voxels are used for alignment.

Reconstruction comparison Measuring the SSIM of our reconstruction x_{rec} and ground truth data $\tilde{x}$ provides the following results. MCCG increases the SSIM by up to 29% compared to simple CGLS up to a maximum of 0.96. On average MCCG could increase SSIM by roughly 22%. We could only observe minimal differences in the results of MCCG using Euclidean distance ($MCCG_{ED}$) or $SSIM_c$ ($MCCG_{SSIM}$) as a cost function for pose estimation. For the medium motion group the relative increase in SSIM was highest. For the high motion group results only show a maximum SSIM of 0.7. We additionally evaluated the scenario where the ground truth volume is known and we can use it to estimate motion for the high motion group. Reconstructing a volume from the motion field obtained in this hypothetical scenario, we were able to obtain an SSIM of 0.98 instead. The results for the three different motion groups are presented in Table IV. On the piecewise linear motion datasets we found that the SSIM stays at a comparably high value greater than 0.9 for motions up to 3 mm before constantly decreasing with higher motion down to an SSIM of 0.65, with the strongest decay between 3 and 4 mm motion amplitude. For the random walk medium motion sets the results all lie around an SSIM 0.9.

Figure 7 plots the projection errors normalized by the amount of pixels used to measure the projection error for CGLS and MCCG with respect to each iteration of CGLS. The amount of pixels varies between both methods due to different border sizes. The projection images had a resolution of 400×400 for CGLS and 350×350 for MCCG. CGLS finished iterating after 13 iterations with an error of 1.74 while MCCG was able to reduce the error towards 0.14 in a total of 69 iterations. Full resolution (400×400) per pixel projection error of MCCG after convergence is 0.15. This is a decrease of 91%.

Parameter estimation Figure 8 shows a comparison between the ground truth motion parameters used for image acquisition and those estimated during MCCG reconstruction. Due to the projection geometry t_z cannot be reliably computed and was therefore kept at 0. It is thus only shown for completeness. Orientation of the reconstruction can differ from ground truth, leading to different parameter curves (see Fig. 2). Therefore, we also display pose parameters obtained after alignment of the reconstruction to the ground truth volume. Parameter curves of MCCG after rotation and ground truth align well with only few notable deviations in the curve of ϕ. These deviations did not occur when reconstructing data created from the same motion with complete tomography.

3.B. Real data

Real data (3D Accuitomo 80) were taken from a patient data-base at the University Medical Center Mainz. These were used in a previous work[33] and include real patient motion. These data were preprocessed as described in Section 4. Figure 9 displays views from axial, sagittal and coronal slices of a volume reconstructed with CGLS and MCCG from a real dataset.

TABLE IV. Structural similarity measure (SSIM) between different reconstructions and ground truth. Only voxels within the central region of interest cylinder with a diameter and height of 240 voxels were used. $MCCG_{ED}$ and $MCCG_{SSIM}$ denote the versions of MCCG where Euclidean distance or $SSIM_c$ were used as cost functions, respectively.

Motion	*CG*	$MCCG_{ED}$	$MCCG_{SSIM}$
Low	0.84	0.95	0.95
Medium	0.70	0.90	0.90
High	0.56	0.71	0.70

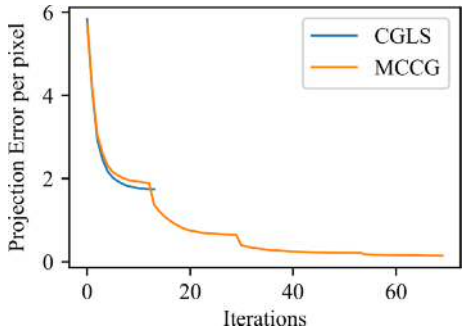

FIG. 7. Comparison of the projection error [see Eq. 9] normalized by the number of pixels per iteration of the CGLS and the motion corrected conjugate gradient method. Number of pixels varies due to different border sizes. [Color figure can be viewed at wileyonlinelibrary.com]

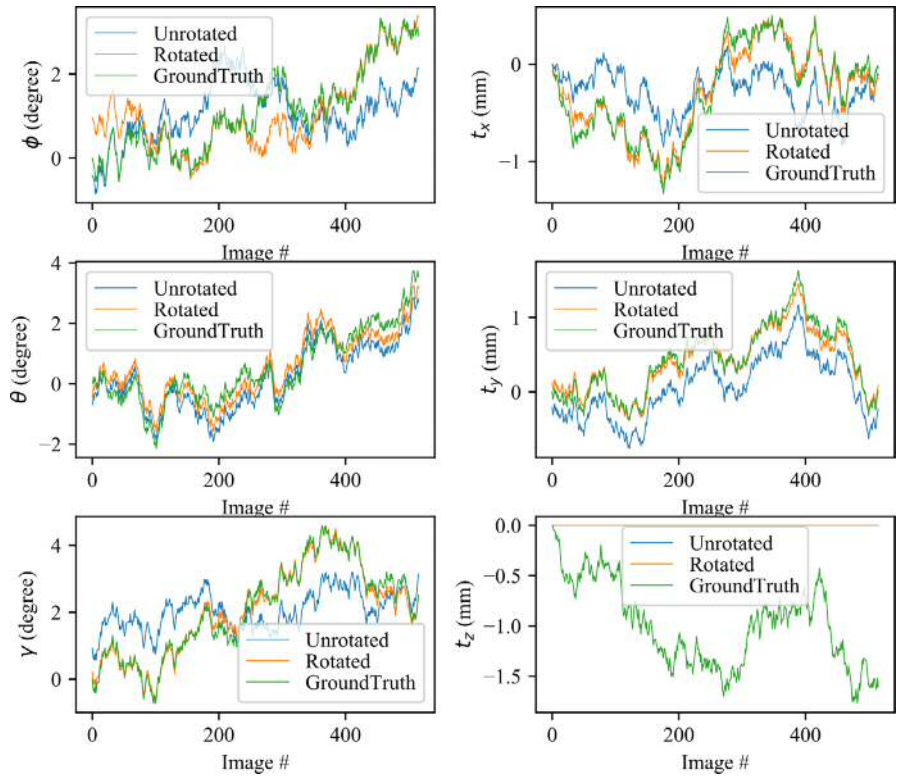

FIG. 8. The six motion parameters displayed over the range of 516 images in a dataset. Ground truth parameters (GroundTruth) used for image acquisition compared to the reconstructed motion parameters without (Unrotated) and with (Rotated) global alignment to ground truth. [Color figure can be viewed at wileyonlinelibrary.com]

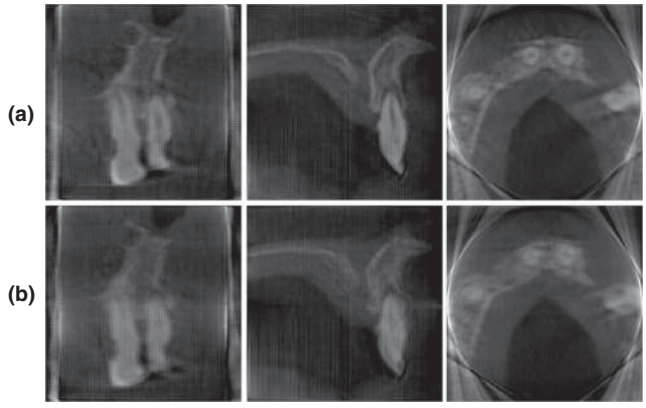

FIG. 9. Slices from a reconstructed region of interest of a real dataset obtained by (a) motion corrected conjugate gradient (MCCG) and (b) standard CGLS. MCCG shows notably more details (e.g., root canal).

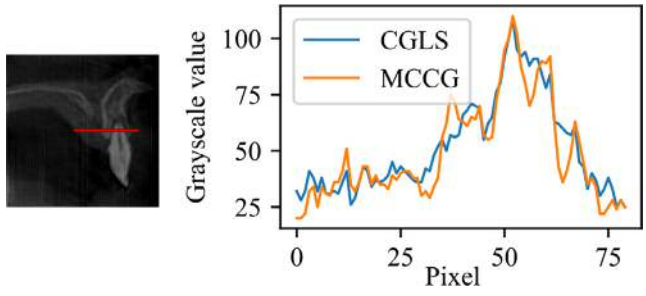

FIG. 10. Line profile within a slice from a reconstruction using motion corrected conjugate gradient (MCCG) and CGLS. Especially within the region 50–75 the MCCG reconstruction shows much more details than the CGLS reconstruction. [Color figure can be viewed at wileyonlinelibrary.com]

In the line profile plot along a line in the axial view (see Fig. 10) the curve of the MCCG version shows clearer and sharper transitions at tissue edges compared to the CGLS curve.

3.C. Runtime

The runtimes for the different motion groups as well as for real data (Table V) were measured on an Intel Xeon E5-2683 v4 CPU equipped with an NVIDIA TITAN V GPU.

TABLE V. Average runtimes in minutes for different datasets using either ED or structural similarity measure (SSIM) as cost function. Datasets contain a total of 516 images with a resolution of 400 × 400 (300 × 300) pixels for synthetic (real) data.

Motion	$runtime_{ED}$	$runtime_{SSIM}$
Low	31	24
Medium	31	28
High	20	18
Real	10	10

Execution time highly depends on the total amount of iterations performed and can vary between individual datasets. In our experiments using SSIM as a cost function for synthetic

data decreases runtime compared to ED. On real data both methods have a comparable runtime. Datasets from the high motion group also express a lower runtime compared to the low and medium motion groups.

4. DISCUSSION

The results show that our proposed method is capable to reliably extract patient motion from a CBCT image series and create less blurred reconstructions with a higher level of detail for local tomography data. Figure 5 shows how the algorithm is capable of improving even highly distorted images by reducing motion blur and other artifacts. These results are confirmed numerically by the increase of SSIM compared to simple CGLS reconstruction (Table IV). Figure 8 illustrates that even arbitrary, nonlinear motion can be reconstructed very accurately. Remaining deviations in the curve of parameter ϕ were not present on reconstructions with complete tomography. Therefore we conclude that they are caused by the local tomography problem. The line profiles show that MCCG is capable of reconstructing more details than simple CGLS. On real data an improvement is equally evident. In the displayed real dataset and corresponding line profile (Fig. 6) the periradicular translucency (around pixel 60) and the root canal (around pixel 55) of the upper incisor became visible with our method even though they were barely visualized in the uncorrected CGLS reconstruction. On the low and medium motion group an SSIM of more than 0.9 could be achieved. This shows that the presented method is able to correct motion amplitudes of up to 9% of the height of the ROI. For even higher motion amplitudes as in the high motion datasets we were only able to reach an SSIM of 0.7. Motion as described by Sisniega et al.[8] generates double contours which our algorithm is only able to correct for small motion amplitudes (SSIM > 0.9 only for motion amplitudes of up to 3 mm). However even for the high motion group an SSIM value of 0.98 could be obtained if the ground truth was used for initial motion estimation. This shows that the algorithm is capable to identify even strong motion as long as the initial reconstruction is a good approximation of the ground truth. For the high motion datasets this is not the case as the initial reconstruction is of rather bad quality. As a result we could only observe an SSIM of 0.7.

Currently the proposed algorithm only reconstructs local tomography data at a resolution of 250^3 voxels. Reconstruction at a finer resolution is possible with this method however it is cost intensive.

The runtime of the proposed method was not the focus of this work. In elective nonurgent clinical procedures as, for example, most applications in dental and maxillofacial radiology, reconstructing a motion-corrected CBCT can be conducted after the initial (motion-disturbed) reconstruction without interfering with the medical dental procedure. Future work will focus on computational efficiency. In general the computational cost of our method is higher than that of purely image based methods due to the constant need of reprojections. The introduction of borders allows the algorithm to converge faster (up to 50% less iterations on certain datasets). As such it is more applicable for local tomography, where the method aligning the complete image as it was described by van Leeuwen et al.[26] converges slower. At the same time the reconstruction quality is barely affected (see Table II). With this method and by using SSIM as a cost function we were able to reduce the overall runtime of the algorithm. This also shows that the generalization to other image similarity measures apart from *ED* is possible and even beneficial. On the high motion datasets the algorithm failed to reconstruct motion and stopped early as no improvement could be achieved. On the real datasets the algorithm converges fast leading to a rather short runtime of only 10 min, which is more suitable for practical use. Nevertheless, even the high runtime of the method can be justified by the possibility to reduce the need for another exposure to ionizing radiation.

5. CONCLUSION AND FUTURE WORK

We showed that the presented approach is able to improve CT-quality for motion-impaired image acquisitions in local tomography. The method manages to improve quality of highly distorted images so that formerly not visible details became visible. The proposed method can help to reduce the need of repeated acquisitions and therefore additional exposure to radiation, thereby possibly reducing the overall radiation exposure in population. We showed that even our simple approach was able to reconstruct sufficient results for medium motion intensities. For higher motion intensities, even though improvement was still possible, a satisfactory level of quality could not be achieved. Making the procedure more robust to higher motion levels can help to improve the applicability to an even broader set of data. In future work other cost functions will be evaluated in order to improve robustness and motion estimation. Other reconstruction methods will also be explored. New methods may also help to further reduce the rather high runtime of our current approach.

CONFLICT OF INTEREST

The authors have no relevant conflict of interest to disclose.

[a)]Author to whom correspondence should be addressed. Electronic mail: stnieble@uni-mainz.de

REFERENCES

1. Schulze R, Heil U, Gross D, et al. Artefacts in CBCT: a review. *Dentomaxillofac Radiol*. 2011;40:265–273.
2. Spin-Neto R, Mudrak J, Matzen L, Christensen J, Gotfredsen E, Wenzel A. Cone beam CT image artefacts related to head motion simulated by a robot skull: visual characteristics and impact on image quality. *Dentomaxillofac Radiol*. 2013;42:32310645.
3. Ouadah S, Jacobson M, Stayman JW, Ehtiati T, Weiss C, Siewerdsen JH. Correction of patient motion in cone-beam CT using 3D–2D registration. *Phys Med Biol*. 2017;62:8813.

4. Spin-Neto R, Matzen L, Liedke G, Gotfredsen E, Wenzel A. Radiographic observers ability to recognize patient movement during cone beam CT. *Dentomaxillofac Radiol*. 2014;43:20130449.
5. Spin-Neto R, Matzen L, Schropp L, Gotfredsen E, Wenzel A. Factors affecting patient movement and re-exposure in cone beam computed tomography examination. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2015;119:572–578.
6. Spin-Neto R, Matzen L, Schropp L, Gotfredsen E, Wenzel A. Movement characteristics in young patients and the impact on CBCT image quality. *Dentomaxillofac Radiol*. 2016;45:20150426.
7. Spin-Neto R, Costa C, Salgado D, Zambrana N, Gotfredsen E, Wenzel A. Patient movement characteristics and the impact on CBCT image quality and interpretability. *Dentomaxillofac Radiol*. 2018;47:20170216.
8. Sisniega A, Stayman J, Yorkston J, Siewerdsen J, Zbijewski W. Motion compensation in extremity cone-beam CT using a penalized image sharpness criterion. *Phys Med Biol*. 2017;62:3712–3734.
9. Spin-Neto R, Wenzel A. Patient movement and motion artefacts in cone beam computed tomography of the dentomaxillofacial region: a systematic literature. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2016;121:425–433.
10. Li T, Xing L, Munro P, et al. Four-dimensional cone-beam computed tomography using an on-board imager. *Med Phys*. 2006;33:3825–3833.
11. de Kinkelder R, Kalkman J, Faber DJ, et al. Heartbeat-induced axial motion artifacts in optical coherence tomography measurements of the retina. *Invest Ophthalmol Vis Sci*. 2011;52:3908–3913.
12. Kachelriess M, Ulzheimer S, Kalender W. ECG-correlated image reconstruction from subsecond multi-slice spiral CT scans of the heart. *Med Phys*. 2000;27:1881–902.
13. Dhou S, Motai Y, Hugo G. Local intensity feature tracking and motion modeling for respiratory signal extraction in cone beam CT projections. *IEEE Trans Biomed Eng*. 2013;60:332–342.
14. Pépin A, Daouk J, Bailly P, Hapdey S, Meyer M-E. Management of respiratory motion in PET/computed tomography. *Nucl Med Commun*. 2014;35:113–122.
15. Hanzelka T, Foltan R, Horka E, Sedy J. Reduction of the negative influence of patient motion on quality of CBCT scan. *Med Hypotheses*. 2010;75:610–612.
16. Marchant TE, Amer AM, Moore CJ. Reduction of motion artefacts in on-board come beam CT by warping of projection images. *Br J Radiol*. 2008;84:251–264.
17. Li T, Schreibmann E, Yangand Y, Xing L. Motion correction for improved target localization with on-board cone-beam computed tomography. *Proc SPIE*. 2005;4121:208–217.
18. Brock R, Docef A, Murphy M. Reconstruction of a cone-beam CT image via forward iterative projection matching. *Med Phys*. 2010;37:6212–6220.
19. Ouadah S, Stayman J, Gang G, Ehtiati T, Siewerdsen J. Self-calibration of cone-beam CT geometry using 3D–2D image registration. *Phys Med Biol*. 2016;61:2613–2632.
20. Ens S. Bewegungsdetektion und -korrektur in der Transmissions-Computertomographie, *Aktuelle Forschung Medizintechnik – Latest Research in Medical Engineering*, Lübeck, Deutschland: Springer Vieweg; 2014.
21. Kyriakou Y, Lapp R, Hillebrand L, Ertel D, Kalender W. Simultaneous misalignment correction for approximate circular cone-beam computed tomography. *Phys Med Biol*. 2008;53:6267–6289.
22. Kingston A, Sakellariou A, Varslot T, Myers G, Sheppard A. Reliable automatic alignment of tomographic projection data by passive auto-focus. *Phys Med Biol*. 2011;38:4934–4945.
23. Aichert A, Berger M, Wang J, et al. Epipolar consistency in transmission imaging. *IEEE Trans Med Imaging*. 2015;34:2205–2219.
24. Siltanen S, Kolehmainen VS. Järvenpää S, et al. Statistical inversion for medical x-ray tomography with few radiographs: I. General theory. *Phys Med Biol*. 2003;48:1437–1463.
25. Kolehmainen V, Siltanen S, Järvenpää S, et al. Statistical inversion for medical x-ray tomography with few radiographs: II. Application to dental radiology. *Phys Med Biol*. 2003;48:1465–1490.
26. van Leeuwen T, Simon Maretzke S, Batenburg K. Automatic alignment for three-dimensional tomographic reconstruction. *Inv Probl*. 2018;34:024004.
27. Latham SJ, Kingston AM, Recur B, Myers GR, Delgado-Friedrichs O, Sheppard AP. Reprojection alignment for trajectory perturbation estimation in microtomography. *IEEE Trans Comput Imaging*. 2018;4:271–283.
28. Pluim JP, Maintz JA, Viergever MA. Mutual-information-based registration of medical images: a survey. *IEEE Trans Med imaging*. 2003;22:986–1004.
29. Wang Z, Bovik AC, Sheikh HR, et al. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13:600–612.
30. Abdelkarim A, Noujeim M, Sarkis T, Lee R, Mah P, Nummikoski P. Effect of movement on high-resolution cone-beam computed tomography images. *Oral Surg Oral Med Oral Pathol Oral Radiol Endodont*. 2007;103:e48.
31. Hanzelka T, Dusek J, Ocasek F, et al. Movement of the patient and the cone beam computed tomography scanner: objectives and possible solutions. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2013;116:769–773.
32. Manduca A, Yu L, Trzasko JD, et al. Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT. *Med Phys*. 2009;36:4911–4919.
33. Schulze R, Michel M, Schwanecke U. Automated detection of patient movement during a CBCT scan based on the projection data. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2015;119:468–472.