

Diplomarbeit in Informatik

**Respiration Monitoring and  
Tomographic Reconstruction  
Towards 4D Cone Beam CT**

Stefan Wiesner

Technische Universität München

Department of Computer Science

Chair for Computer Aided Medical Procedures and Augmented Reality

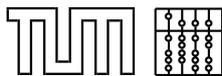
in collaboration with

Imaging Science and Information Systems (ISIS) Center, Washington, DC, 20057

**Stefan Wiesner**

# **Respiration Monitoring and Tomographic Reconstruction Towards 4D Cone Beam CT**

**Diplomarbeit in Informatik**



**Technische Universität München  
Fakultät für Informatik**

This thesis was conducted in collaboration with  
ISIS Center Georgetown University

Autor: Stefan Wiesner  
Aufgabensteller: Prof. Dr. Nassir Navab (TUM)  
Betreuer: Prof. Ziv Yaniv (ISIS)  
Abgabedatum: 15.09.2007

Ich versichere, dass ich diese Diplomarbeit selbständig verfaßt und nur die angegebenen Quellen und Hilfsmittel verwendet habe.

Garching, den 15. September 2007

Stefan Wiesner

## Abstract

Image guided surgery (IGS) systems utilize three dimensional (3D) images to provide guidance in minimally invasive surgery. To date, these systems have been primarily applied in interventions involving rigid anatomical structures. In these types of procedures a single 3D image is sufficient to represent the underlying structures throughout the procedure. More recently, researchers have turned their attention to interventions involving anatomical structures that move and deform due to respiration. In these procedures a single 3D image is only valid at a specific phase of the respiratory cycle. To provide continuous feedback to the physician a 4D, 3D over time, image data set is required.

This thesis describes respiratory monitoring techniques and a retrospective 4D data reconstruction method using Cone-Beam CT (CBCT) as the imaging modality. We present two approaches for respiration monitoring. The first one is based on optical tracking of a colored fiducial on the patients abdomen with a video camera which is synchronized with the X-ray image acquisition. The second one generates the signal directly from the X-ray images by tracking a spherical fiducial visible in the images and reconstructing its trajectory. Both approaches have been evaluated using simulated and real data, and have shown their feasibility. The last part of this thesis describes methods for retrospective 4D image reconstruction using the simultaneous algebraic reconstruction technique (SART). We evaluate two approaches for reconstructing the 4D data set. The first approach performs retrospective gating based on the respiratory signal generated using our methods. The X-ray images are binned according to the respiratory signal and each bin is reconstructed separately, a hard binning approach. The second approach reconstructs the 4D data set using a soft binning approach. For each phase of the respiratory cycle the 3D volume is reconstructed using all images, where each image is weighted according to its relative position in the respiratory cycle. Both approaches are evaluated using an anthropomorphic respiratory phantom and clinical image acquisition protocols.

## Acknowledgments

After six month of researching and thesis writing I would like to thank all the wonderful people who made this possible and supported me during that time. First of all I would like to thank Nassir Navab for thinking of me when Kevin Cleary offered him to send a student to his lab and giving me the chance to write my thesis abroad. Special thanks to Kevin Cleary for having me at his research group and for the countless Rockland lunches he invited us to. For the grandiose supervision of my project, I want to say thanks to Ziv Yaniv. He showed me how to conduct systematic research, gave me new ideas and motivated me when things were not working too well. Without Ziv's help I would not have been gotten to this point of writing the acknowledgments now. My special thanks go to the whole CAIMR group, Fil, Ken, Jae, Craig, Minh, Dave, Konrad, Teo, Roland, Jayson, Emmanuel, Ralph, Patrick and John for fruitful discussions and help with the experiments.

I also want to thank the whole CAMP group, to which I was affiliated for the last three years. Thanks to Keily for giving me the hint where to find the matrices and Olli for the help on GPU programming. Unfortunately I was running out of time to implement Olli's suggestions. Thanks to Christoph Paul Bichlmeier, Sandy, Philipp, Ben, Darko, Martin H., Martin G., Rux, Christian, Jörg, Marco, Tobi, Wolfi, Mo and Markus for giving me a great time at the group.

Last but not least I want to thank my family for letting me leave to DC and my girlfriend Claudia for her support and love and being my family in the states.

# Contents

1	Introduction	1
2	Respiratory Monitoring	7
2.1	Video based respiration monitoring	7
2.1.1	System description	8
2.1.2	Experimental Results	12
2.2	Respiratory signal generation from projection images	17
2.2.1	Fiducial localization and tracking	17
2.2.2	Trajectory reconstruction and respiratory signal generation	19
2.2.3	Results	22
3	Tomographic Reconstruction	27
3.1	Reconstruction	27
3.2	Binning and Weighting	32
3.3	Results	34
3.3.1	Digital Phantom	34
3.3.2	Respiratory phantom	36
3.3.3	Motion Platform Experiment	36
4	Summary and Outlook	43
4.1	Summary	43
4.2	Future work	44
	Appendix	44
A	Derivation of line-line distance	45
B	Ray-sphere intersection	47
	Bibliography	48

## List of Figures

1.1	Spleen motion of a patient . . . . .	2
1.2	Lung tumor motion . . . . .	3
1.3	Acquisition scheme of a 4D CT . . . . .	4
2.1	Electromagnetic tracking of two stationary tracked needles during a 20 seconds C-arm scan . . . . .	7
2.2	Hardware components for respiratory monitoring . . . . .	8
2.3	Automatic selection of marker colors . . . . .	9
2.4	Respiratory signal processing . . . . .	11
2.5	Comparison of respiratory signal obtained by video and Polaris system . . . . .	13
2.6	Effect of camera placement on system performance . . . . .	15
2.7	Experimental setup for US based assessment of respiratory signal generation. . . . .	16
2.8	US images acquired for validating the respiratory signal . . . . .	16
2.9	Fiducial tracking results . . . . .	22
2.10	Line reconstruction example . . . . .	23
2.11	Setup with motion platform and CBCT scanner . . . . .	24
2.12	X-Ray images from two motion platform experiments . . . . .	24
2.13	Amplitude and phase signal of ground truth overlaid with reconstructed signal . . . . .	25
3.1	Fourier Slice Theorem . . . . .	28
3.2	Reconstruction scheme . . . . .	30
3.3	Scheme of gated reconstruction . . . . .	33
3.4	Digital phantom projections . . . . .	35
3.5	Digital Phantom setup and modulation function . . . . .	35
3.6	Reconstruction results from simulation study . . . . .	38
3.7	Reconstruction of bin $-10$ . . . . .	39
3.8	Reconstruction of anthropomorphic respiratory phantom . . . . .	40
3.9	Motion platform with attached foam liver . . . . .	41
3.10	Different binning results for amplitude and phase signal . . . . .	41
3.11	Slice view of reconstruction of the moving phantom liver . . . . .	42

## List of Tables

2.1	CAMSHIFT algorithm . . . . .	10
2.2	Color based monocular respiratory motion monitoring framework. . .	12
2.3	Evaluating of video based respiratory motion monitoring vs. the Polaris system . . . . .	14
2.4	Square root of eigenvalues resulting from PCA. . . . .	17
2.5	Summary of trajectory reconstruction algorithm . . . . .	18
2.6	RANSAC Algorithm . . . . .	19
2.7	Comparison of linear and non-linear algorithm . . . . .	23
2.8	Correlation of respiratory signals computed from differently attached markers . . . . .	26
3.1	SART algorithm. . . . .	31
3.2	Parameters of the digital phantom . . . . .	34

# 1 Introduction

In the last decade much effort has been put in the development of image guided surgery (IGS) systems to support minimally invasive surgery (MIS). The benefit of MIS is reduced trauma for the patient, as instruments are inserted through small ports into the anatomy and large incisions are avoided. The drawbacks of this approach are limited access and no direct line-of-sight to the target, making the intervention more complicated. IGS systems can augment the physicians abilities, enabling them to perform such procedures. The main components of an IGS system include, medical images, tracking systems, visualization, and registration. Most often preoperative three dimensional images (e.g. CT) are acquired. Intraoperatively these images are aligned to the patient via registration and tracked tools are visualized by dynamically overlaying them onto the images. Thus, during the procedure the physician is guided by the dynamic visualization, moving the instrument towards the target without having a direct line-of-sight to it. So far IGS has focused on cranial and musculoskeletal procedures in clinical use [41] where the structures are rigid and deformation is minimal.

More recently researchers have turned their attention to interventions involving anatomical structures that move and deform due to respiration. Two key components have been identified as enabling technologies for such procedures, motion modeling and intra-operative imaging [24]. Currently, image-guided interventions performed in the thoracic-abdominal region use a gated approach. That is, a single three dimensional CT volume is acquired at breath-hold and the physician advances only during the respiratory phase in which the image was acquired. An example of this approach is the work by Banovac et al. [3] where needle punctures on a respiring phantom are performed by inserting the needle during breath-hold. This has the undesirable effect of increasing the procedure time. To enable a more streamlined workflow we propose to use a 4D, 3D over time, Cone-Beam CT (CBCT) data set. In this manner, we can provide a dynamic image display throughout the respiratory cycle.

Organ motion due to respiration has been studied in several papers. Brandner et al. [9] investigated the motion of liver, kidneys and spleen using 4D CT. They found superior-inferior (SI) motion of the organs to be about 2cm and more. Anterior-posterior (AP) motion of up to 1.4cm, whereas motion in lateral (LR) direction was less than 0.54cm.

Figure 1.1 displays the trajectory of a patients spleen. The motion is complex and does not follow the same path for inspiration and expiration. This phenomena is known as hysteresis. Similar results are presented by Seppenwoolde et al. [53]. They implanted small gold fiducials in or close to lung tumors and computed the trajectories using stereoscopic fluroscopy. Figure 1.2, from this work, also shows the phenomenon of hysteresis.

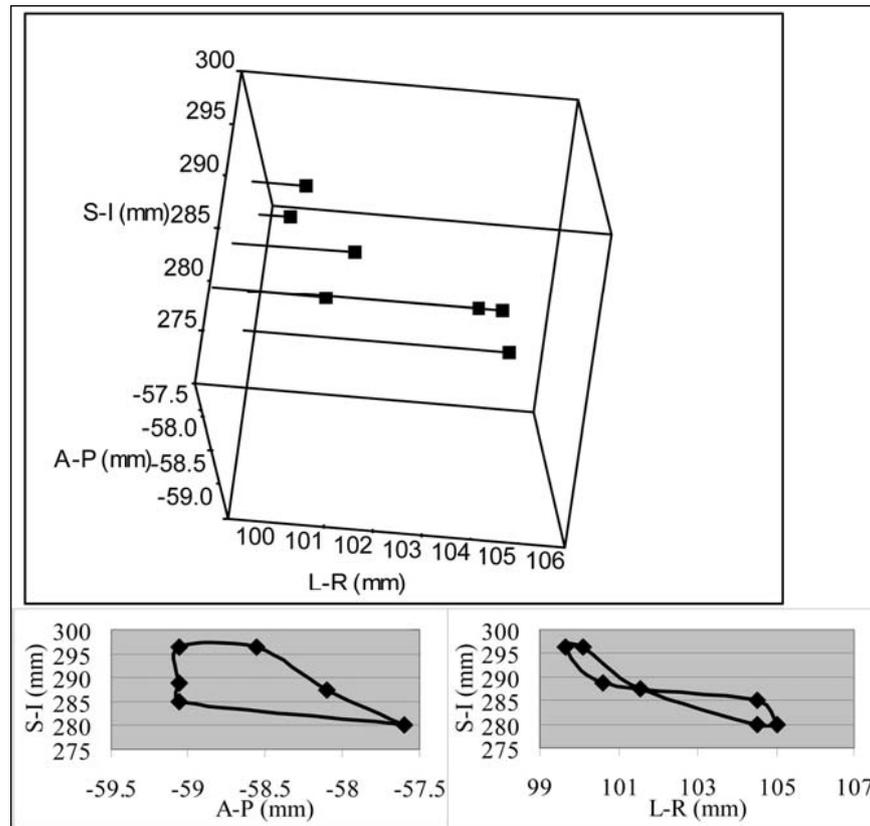


Figure 1.1: Spleen motion of a patient found by Brandner, figure adopted from [9]

Rohlfing et al. [48] did not only investigate the rigid motion of an organ, but also its deformation due to respiration. They investigated liver deformation, using a sequence of gated MR images and registered them by applying rigid and non-rigid techniques. The SI motion for 4 patients ranged from 1.2cm to 2.6cm and deformation was in average about 0.6cm. Blackall [6] followed the same approach and additionally analyzed the shape model of the liver by segmenting the surface in the different scans. She reports average deformation of 0.42cm and maximal deformations of 1.41cm of the liver surface.

It should be noted that one has to be careful segmenting the motion in the CT coordinate frame and setting it equal to the body frame (CT z-axis = SI axis). Having the patient in a slightly different position in the scanner will change the amount of motion in the different axis. A more general approach is to specify the motion in its eigen-frame, given by the principle axes of motion.

A common method in radiation therapy is to monitor a surrogate marker on the skin and infer from that the motion of the tumor. This assumes that skin and tumor motion are correlated, a reasonable assumption given that both motions are caused by respiration. Several researchers have evaluated the validity of this assumption. Beddar et al. [5] implanted gold fiducials in the liver, close to a tumor and acquired

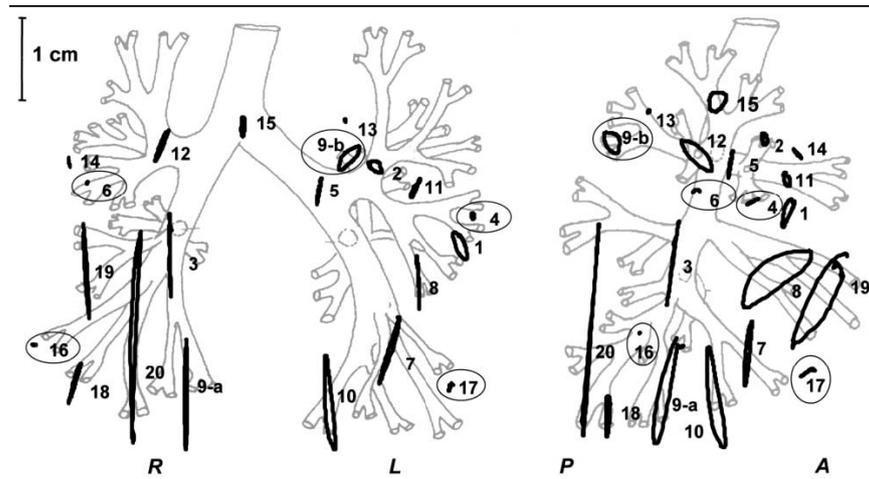


Figure 1.2: Lung tumor motion found by Seppenwoolde, figure adopted from [53]

a 4D CT from which the fiducial motion was estimated. Skin motion was acquired by an optical tracking system observing a marker block on the patients chest. Ozhasoglu and Murphy [39] implanted fiducials close to lung and pancreas tumors and placed spherical markers on the skin surface, with trajectories estimated using fluoroscopy. Both works show correlation between internal and external markers, but also report on phase shifts in some cases. Wong et al. [60] conducted some preliminary work to show correlation of skin and liver motion. They investigated different dynamic MRI techniques and segmented skin and liver. The correlation is not presented in this publication.

The impact of respiration on tumor position is of particular interest in radiation therapy. The aim of external radiation therapy is to destroy a tumor by delivering a radiation beam to the target, using a linear accelerator. Advanced methods acquire 4D CT data before the treatment and adopt the treatment plan to the moving tumor. This allows dynamic radiation beam behavior by following the tumor motion [52], precisely delivering radiation to the tumor while sparing surrounding healthy tissue.

One possibility is to acquire a 4D CT while the patient is freely breathing. Vedam et al. [58] describe a method for retrospective gating of the CT images. During consecutive CT scans they record the respiratory signal with an optical tracking system and synchronize it with the CT image acquisition. Because the scanner rotation time is short, 1.5 seconds, motion artifacts in a reconstructed tomographic slice are small. The final 4D data set is produced by assembling the single tomographic slices according to their location in the respiratory cycle as defined by the respiratory monitoring system. Figure 1.3 illustrates this principle.

Another possibility for acquiring a 4D data set is to use retrospective gating of CBCT. In contrast to CT, where the tomographic images are sorted after reconstruction, the projection images acquired with a CBCT have to be sorted beforehand and the reconstruction is performed afterwards. This is due to the fact that the acquired images are projections and not tomographic slices. Previous solutions to the task

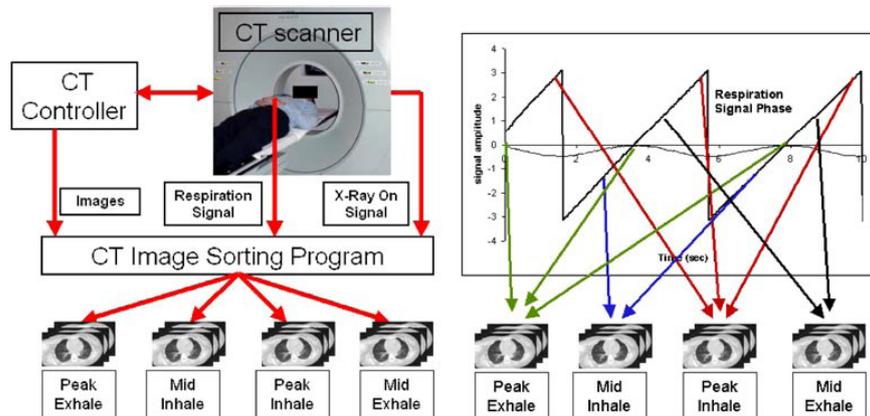


Figure 1.3: Acquisition scheme of a 4D CT, figure adopted from [58]

of 4D CBCT data acquisition have been based on retrospective gating [47, 55, 30]. A respiratory signal is recorded simultaneously with the image acquisition and the projection images are sorted accordingly. From this set of images several 3D volumes over time are reconstructed. The challenge is to reconstruct a volume from limited views, which are not evenly distributed over the scanner trajectory.

Reconstruction from a few projection images has been previously addressed in the literature. Iterative methods are known to deal better with this problem than filtered backprojection algorithms [33]. Never the less due to the slow rotation speed of most CBCT scanners researchers investigating 4D CBCT reconstruction obtain enough projection images to apply Feldkamp [15] like algorithms. Usually, they have more than 30 projection images distributed over the C-arm trajectory.

Reconstruction quality has previously been assessed by Rit et al. [47] using a simulation study. They assess different retrospective binning strategies and analyze the signal-to-noise ratio, the contrast-to-noise ratio and a blur criterion of the result. Sonke et al. [55] also implemented a retrospective gating approach and evaluated it using a phantom and patient data. Both works show that increasing the gating window also increases the blur in the images and decreasing the window, meaning selecting fewer projections increases artifacts. Thus a trade-off between motion and reconstruction artifacts has to be found, which depends on the specific respiratory pattern. A simple solution to the limited number of views problem is to acquire several scans as done by Li et al. [31]. While this does improve image quality it increases both the scanning time and the radiation dose the patient receives. A different approach proposed by Chang et al. [12] is based on prospective gating, triggering the CBCT scanner according to a respiratory signal. They achieve image quality comparable to the stationary case, but increase the scanning time (7 minutes).

An advanced method is dynamic reconstruction, which integrates a motion field in the reconstruction algorithm. Schaefer et al. [50] and Prümmer et al. [43] applied this approach to reconstruct the moving heart. Schaefer et al. [50] use the known motion field of the beating heart and adapt the backprojection step of the FDK algorithm. This approach assumes a known motion which is problematic. The motion is usually

obtained from a 4D scan acquired prior to the procedure. Thus the current 4D scan is obtained using a previous 4D scan. This approach is susceptible to changes in the breathing pattern and even to the filling of the colon. A second problem is to register the motion field to the current reconstruction problem. To eliminate these restrictions Prümmer et al. [43] suggest a hybrid approach. They first reconstruct several gated volumes of the heart performing multiple sweeps with the C-arm. From that data they estimate a motion field and adapt the FDK algorithm to incorporate the motion field. They show that their approach reduces motion artifacts. Zeng et al. [62, 63] follow a different approach. Instead of reconstructing the image they non-rigidly register a preoperative CT acquired at breath-hold to the projection images. The problem becomes a 2D/3D non-rigid registration task. A drawback of this approach is the high computation time, approximately 10 hours [62].

A prerequisite to retrospective CBCT gating based on respiration is the availability of a respiratory signal. Two possible ways of obtaining such a signal are either to use an external monitoring device, or to extract a signal directly from the projection images. The first category includes, devices based on the temperature of airflow [7], on direct measurement of airflow [29], on the tension of a belt placed on the patient [11], and on tracking skin surface motion. The later are either monocular camera systems using infrared [38, 40] images, and stereo camera systems, using infrared [4] or standard video [21].

In our context, 4D CBCT reconstruction, these systems have two main drawbacks. They introduce additional hardware that is currently not part of the acquisition protocol, and they require synchronization with the image acquisition. An alternative approach proposed by several authors is to generate a respiratory signal directly from the projection images [31, 46, 64]. This approach has the added advantage of not requiring additional hardware, and more importantly it is implicitly synchronized with image acquisition.

In [64] the respiratory signal is generated by tracking features in the x-ray projection images. The tracked features are limited to the region of the diaphragm, as this region exhibits the maximal motion due to respiration. The 2D feature locations are projected onto the superior-inferior axis and those with maximal motion are used to generate the respiratory signal based on their displacement along the superior-inferior direction.

In [46] this work was extended. In this case, features are sampled uniformly across the region of the lungs. They are then tracked in the 2D images. Features that are only visible for a short period of time or move only slightly are discarded. For all other features, the 2D locations of each feature are projected onto the least squares line fit to them. This yields a respiratory signal from a single feature. All signals are combined by linearly normalizing their magnitude to be in  $[0, 1]$  and taking the mean value for each point in time. The authors note that most of the feature points retained by their approach are in the region of the diaphragm and are thus in agreement with the previous approach [64], although some features are located on the lung walls.

In [31] a  $1mm$  radio-opaque fiducial is placed on the patients skin and is tracked in the x-ray projection images. The 2D fiducial locations are projected onto the superior-inferior axis and this is the 1D respiratory signal. The primary difference between

this approach and the previous two [64, 46] is that those approaches track regions of interest that do not correspond to a consistent spatial anatomical structure, as both the x-ray source and anatomy are moving, and this approach does track a specific spatial point.

This thesis is an initial step towards 4D CBCT based IGS. We describe novel methods for extrinsic and intrinsic respiratory monitoring and standard reconstruction methods. The extrinsic respiratory signal is generated using a low cost web cam monitoring a colored fiducial on the abdomen. We obtain the intrinsic signal by placing a radio opaque fiducial on the patient and reconstruct its trajectory using the projection images. Further, we extended the algebraic reconstruction algorithm by a weighting scheme derived from the respiratory signal. Finally, initial results of 4D CBCT reconstruction using extrinsic monitoring are presented.

The rest of this thesis is organized as follows: Chapter 2 introduces the two respiration monitoring algorithms and shows their performance. Chapter 3 presents the underlying principles of tomographic reconstruction and incorporates respiratory binning strategies. The chapter closes with results of the binned reconstruction. The final chapter 4 gives a summary and conclusion of the thesis and points out future work.

## 2 Respiratory Monitoring

### 2.1 Video based respiration monitoring

This section describes a method for monocular, webcam-based, respiration monitoring. The development of this approach was motivated by our desire to acquire 4D CBCT data for an IGS system utilizing electromagnetic (EM) tracking.

Initially we attempted to use the Aurora EM tracking system (Northern Digital Incorporated, Waterloo, Canada) to monitor respiration during CBCT data acquisition. This approach was found to be impractical, as once the C-arm is in motion it interferes with the electromagnetic field and the tracking result is severely corrupted. Figure 2.1 shows the experimental setup and the tracking results of two static probes during a 20 seconds scan.

As IGS systems are notoriously expensive, we sought a simple cost effective solution to our problem. This section describes such a solution.

Our respiratory monitoring system is based on automatic detection and tracking of the motion of color fiducials placed on the patient's abdomen in the epigastric region (near the xiphoid process) and the left and right hypochondriac regions. The system consists of a single webcam mounted on an adjustable arm and colored cube fiducials attached to a velcro belt, as shown in Figure 2.2. Each of the fiducial cube's faces has a different color, allowing the system to choose an appropriate color based on the colors present in the specific environment. Video images are acquired with a resolution of  $640 \times 480$  at a frame rate of  $30Hz$ .

The respiratory signal is generated from the tracking data of the fiducial with the largest motion magnitude along the principle axis of variance as defined by the Principle Component Analysis of the 2D motion. In practice this has always corresponded to the fiducial closest to the xiphoid process. The remaining fiducials

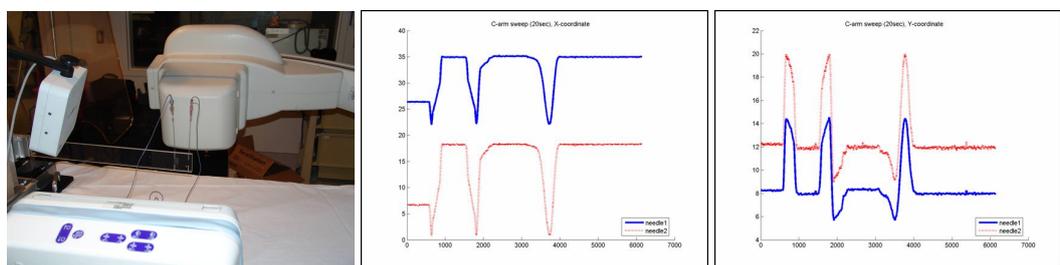


Figure 2.1: Electromagnetic tracking of two stationary tracked needles during a 20 seconds C-arm scan: (a) experimental setup (b) plot of x-coordinate during scan (c) plot of y-coordinate during scan. It is clear that the tracking information obtained during C-arm rotation is invalid.

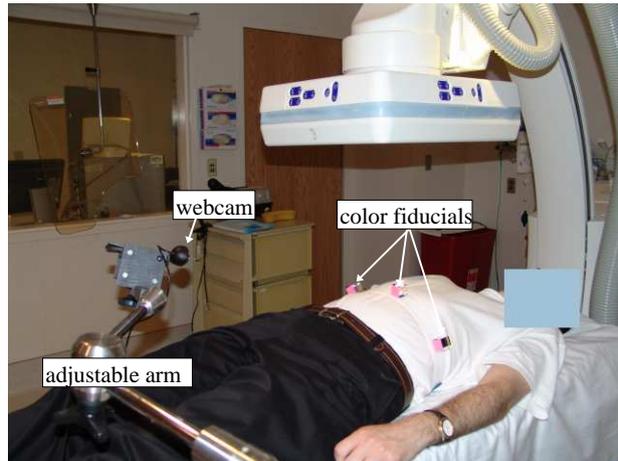


Figure 2.2: Hardware components for respiratory monitoring: of the shelf webcam (Logitech, QuickCam Fusion), plastic cubes with colored faces, and adjustable arm.

are used in conjunction with this fiducial to improve the systems sensitivity to rigid body motion.

To robustly use color as the feature of interest we work in the Hue Saturation Value color space [17]. This decouples the object's color from the intensity information, and when using the hue information provides relative invariance to illumination and viewing direction [22]. Thus the system does not require a strict relationship between the camera and patient positions a useful characteristic in the clinical setting.

It should be noted that our system requires initialization. This is an automatic process during which we assume the patient is exhibiting normal tidal breathing. We now describe our approach in detail.

### 2.1.1 System description

As a first step we choose the best possible color for tracking in the specific environment from among the four colors on the cube's faces. An image of the patient setup is acquired, and the system indicates which of the four possible colors should be used. This is based on an analysis of the Hue-Saturation histogram, with the color of choice having the hue whose surrounding region in the histogram is the sparsest. From our experiments we have empirically determined that a region of  $\pm 15^\circ$  is sufficient. Figure 2.3 illustrates this process.

Once the color is chosen, the fiducials are attached to the velcro belt and an image is acquired. Possible fiducial locations in the image are detected using a threshold on the the hue and saturation values followed by connected component analysis. The threshold on the hue values is set to the fiducial hue  $\pm 15^\circ$  as was the case in the previous step. In addition we require a minimal saturation value of 0.3, as at low saturation values the hue data is less reliable for color based segmentation. Connected components in the resulting binary image are detected and those that consist of less

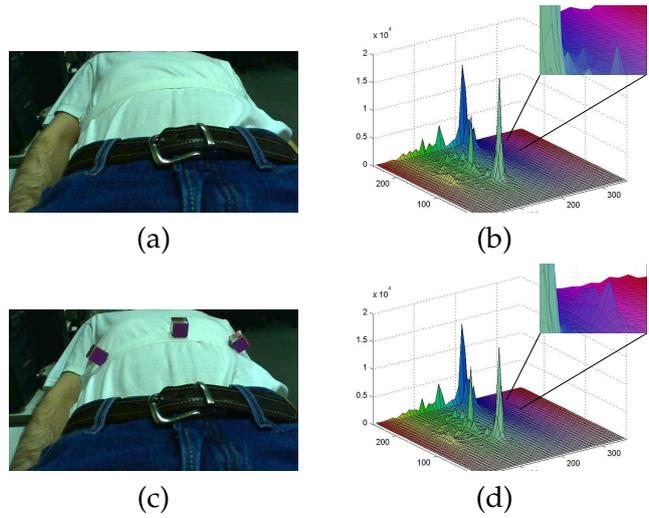


Figure 2.3: Automatic selection of marker colors according to image's Hue-Saturation histogram: (a) image without marker (b) corresponding Hue-Saturation histogram with sparsest region corresponding to the purple Hue value, inset shows zoomed in region (c) image with markers (d) new Hue-Saturation histogram exhibiting a peak in the region corresponding to the purple markers, inset shows zoomed in region.

than 50 pixels are considered noise, as marker sizes in the images vary between 400 to 900 pixels.

The locations of the centroids of the remaining connected components are candidate fiducial locations in the image. These serve as the initial input to the Continuously Adaptive Mean Shift (CAMSHIFT) blob tracking algorithm [8]. CAMSHIFT is an extension to the Mean Shift Algorithm [19, 14], which is a non-parametric technique to determine the mode of a Probability Density Function (PDF)  $P(x)$ . It can be derived as the gradient of the Parzen Window Density Estimation, a technique to determine the whole PDF, given only  $n$  samples of the PDF. Parzen Window Density Estimation is a data interpolation technique which approximates the underlying function by superimposing kernel functions  $k(x)$  of size  $h$  at the sample points  $x_i$ .

$$P(x) = \frac{c}{n} \sum_{i=1}^n k\left(\frac{x - x_i}{h}\right) \quad (2.1)$$

For the multivariate case (2 dimensional image) we consider only the Euclidean length of the input sample and the kernel becomes  $K(x) = k(\|x\|^2)$ . We selected the radially symmetric Epanechnikov kernel

$$k(x) = \begin{cases} 1 - x & \|x\| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.2)$$

$$g(x) = -k'(x) = \begin{cases} 1 & \|x\| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (2.3)$$

having the uniform kernel as gradient, which simplifies the algorithm. Equation 2.4 derives the Mean Shift vector

$$\nabla P(x) = \frac{c}{n} \sum_{i=1}^n \nabla k_i = \frac{c}{n} \left[ \sum_{i=1}^n g_i \right] \left[ \frac{\sum_{i=1}^n x_i g_i}{\sum_{i=1}^n g_i} - x \right] \quad (2.4)$$

$k_i$  is the kernel evaluated at sample position  $x_i$ ,  $g_i$  is its negative derivative and  $c$  a normalization factor. The term

$$\left[ \frac{\sum_{i=1}^n x_i g_i}{\sum_{i=1}^n g_i} - x \right] \quad (2.5)$$

is called the mean shift vector. For the tracking problem we define a tracking window (kernel), compute the mean of all pixels in the window, subtract the window center position and shift the window by the found vector. The size of the kernel has to be chosen in advance and has a major impact on the tracking result. CAMSHIFT solves this by setting the size adaptively. Depending on the zeroth moment (the area of the pixels in the window) the size of the window increases or decreases in the next iteration step. Table 2.1.1 summarizes the algorithm.

- |   |
|---|
| <ol style="list-style-type: none"> <li>1. place window on image</li> <li>2. compute mean and zeroth moment of pixels in window</li> <li>3. shift window to mean position</li> <li>4. adapt window size according to zeroth moment</li> <li>5. iterate till length of mean shift vector is smaller than threshold</li> </ol> |
|---|

Table 2.1: CAMSHIFT algorithm

The CAMSHIFT algorithm tracks the color blobs for 30 seconds and this data is used to detect the actual fiducial locations in the image.

We perform Principle Component Analysis (PCA) on the path of each tracked blob. The 2D points are projected onto the principle axis, the eigen-vector corresponding to the largest eigen-value, resulting in a one dimensional signal. For fiducials we expect this signal to be correlated with the respiratory cycle, that is, it will have a dominant frequency in the range of 12 to 25 cycles per minute, as respiratory rates outside this range are considered abnormal. All candidate fiducials that pass this test are assumed to be true fiducials.

To determine the frequency of the signal we transform it into the Fourier space using the Discrete Fourier Transformation (DFT).

$$F_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} kn} \quad (2.6)$$

$x_n$  is the  $n$ -th sample of the respiratory signal,  $F_k$  is the  $k$ -th frequency and  $N$  the number of samples. The most prominent frequency indicates the number of cycles in the time interval  $t$ . The cycles per minute are given by  $F_k \frac{60}{t}$ . Figure 2.4 illustrates the respiratory signal processing steps. The first graph shows the noisy signal obtained by performing PCA on the tracking data. The frequency spectrum has a peak at 19 cycles per minute, a realistic value for respiration. Filtering the noisy signal with a gaussian kernel with standard deviation of  $\sigma = 2$  gives a smooth signal, where we can detect the extreme points. This is done by computing the first derivative of the signal, using finite differences and detecting its zero crossings. Interpolating linearly between the extreme points yields the phase signal. In contrast to other works, which define one phase as (start-inhalation, end-inhalation, start-exhalation, end-inhalation) we define it as (start-inhalation, end-inhalation) and approximate (start-exhalation, end-inhalation) as its reverse. Thus we do not take hysteresis into account.

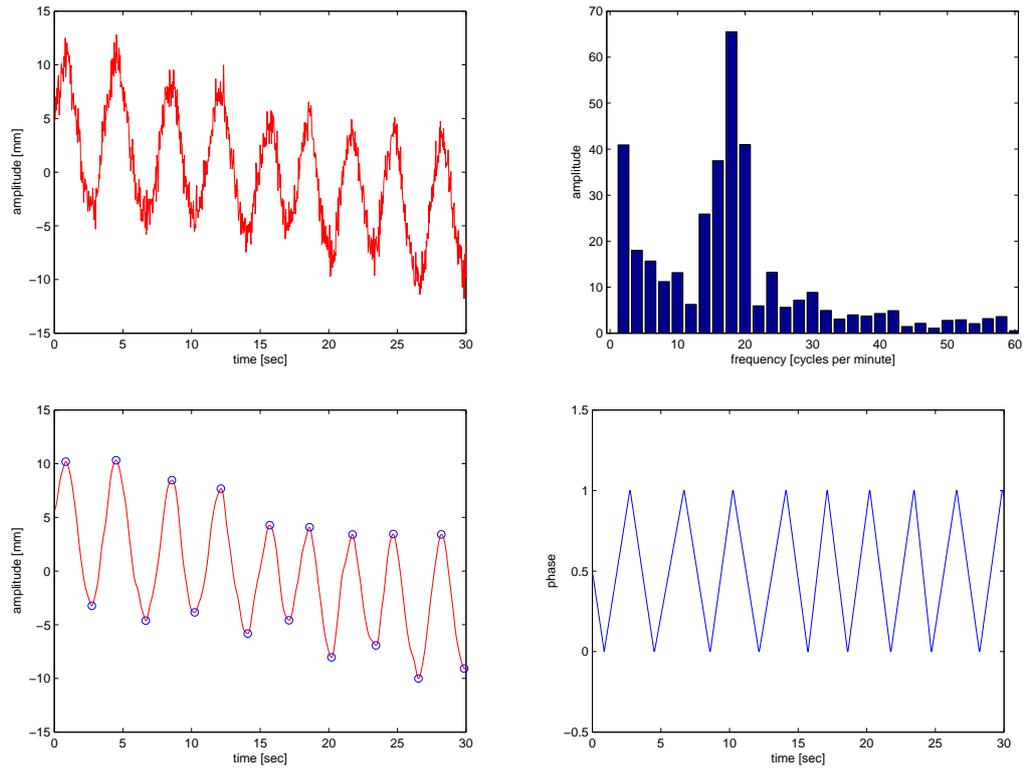


Figure 2.4: Respiratory signal processing. (a) Respiratory signal (b) Frequency of signal (c) Filtered signal with detected extremas (d) phase signal

We now identify the fiducial that has the maximal motion along its principle axis.

This fiducial is our Primary Fiducial (PF) and is used to generate the respiratory signal.

Once the initialization phase is completed the CAMSHIFT algorithm is used to track the fiducial locations in the image and a respiratory signal is generated by projecting the motion of the PF onto its principle axis which was computed during the initialization phase. Table 2.2 summarizes our approach.

1. Initialize:

- a) Select fiducial color based on analysis of Hue-Saturation histogram of patient setup image.
- b) Place fiducial markers.
- c) Detect color blobs corresponding to marker color using thresholds on the Hue and Saturation values and perform connected component analysis.
- d) Initialize the CAMSHIFT algorithm using the detected blobs and track them for 30sec.
- e) Perform PCA on the tracked data and project each blob's motion onto its principle axis.
- f) Analyze the frequency of the signal generated by each of the blobs, if it is in the range of 12 to 25 cycles per minute it is classified as a fiducial.
- g) Identify the fiducial with the maximal magnitude along its principle axis, the primary fiducial (PF).

2. Respiratory signal acquisition:

- a) Track blobs identified as fiducials in the previous stage using the CAMSHIFT algorithm.
- b) Project the location of the PF onto its principle axis as computed during initialization (this is the respiratory signal).

Table 2.2: Color based monocular respiratory motion monitoring framework.

### 2.1.2 Experimental Results

We test the accuracy of our system by comparing the respiratory signal generated using the video images to a respiratory signal generated using the Polaris Vicra tracking system from Northern Digital Inc. (Waterloo, Ontario, Canada). This tracking system has a manufacturer specified spatial localization RMS error of  $0.25mm$ .

To generate both respiratory signals we attach markers tracked by the Polaris system next to our fiducials and simultaneously acquire video images and 3D tracking data. The respiratory signal based on the polaris data is generated by performing PCA on these 3D points, with the respiratory signal corresponding to the projection of the data onto the computed principle axis, the eigen-vector corresponding to the

largest eigen-value. We then compare the two signals  $X, Y$  using normalized cross correlation, Pearson's  $r$  [42], as our similarity measure.

$$r = \frac{COV(X, Y)}{\sigma_X \sigma_Y} \quad (2.7)$$

To evaluate our system's performance we acquired four types of data sets from a volunteer in which he exhibited tidal breathing, tidal breathing interrupted by coughing, tidal breathing interrupted by periods of breathholding, and tidal breathing interrupted by motion. For each category we acquired three data sets. Figure 2.5 shows respiratory signals from each type of data set. As can be seen from this figure and the results presented in Table 2.3 the respiratory signal generated by the Polaris system and our video based system are highly correlated, with all correlation values above 0.95, except in the experiments involving patient motion, where the correlation was 0.87. Patient motion is truly 3D and violates the approximation of motion along a line, what is the foundation of our respiratory signal computation.

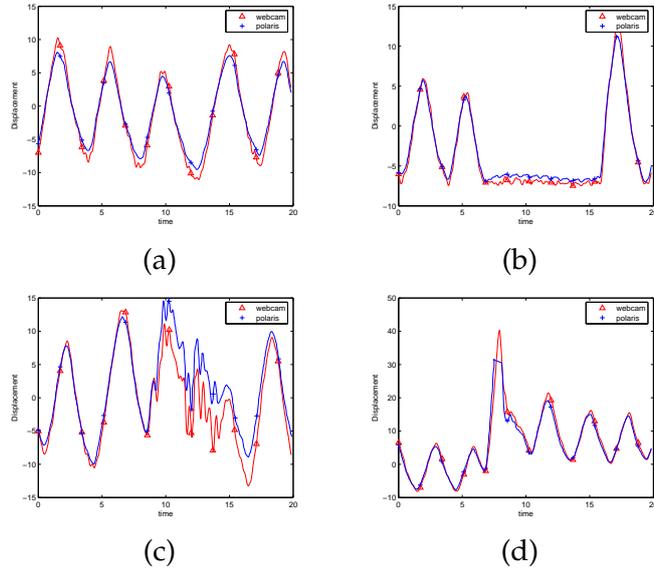


Figure 2.5: Respiratory signal obtained by our method and using the Polaris Vicra for (a) normal breathing,  $NCC=0.99$ ; (b) breath hold,  $NCC=0.99$ ; (c) coughing,  $NCC=0.95$ ; and (d) patient motion during acquisition,  $NCC=0.98$ . Several sample points are marked on each graph.

As we are using a single video camera to perform motion tracking we evaluate the effect of various spatial relationships between camera and patient on tracking, as these depend upon the clinical application. We acquired several data sets with varying camera locations. For each data set we generate a respiratory signal based on the Polaris data and video images. In all experiments the acquired data is of a volunteer exhibiting normal tidal breathing. The correlation between the Polaris based signal and video based signal was above 0.99 for all experiments. That is, the

Data Type	NCC value
normal breathing	0.99
	0.99
	0.99
breath-hold	0.99
	0.99
	0.99
coughing	0.97
	0.95
	0.95
patient motion	0.98
	0.91
	0.87

Table 2.3: NCC values for evaluating video based respiratory motion monitoring vs. the Polaris infrared stereo based system. These results show a high correlation between the two approaches, for all scenarios.

effect of varying camera pose is negligible. Figure 2.6 shows three examples.

Finally, we evaluated our system using a portable ultrasound (US) system, the Terason 2000 from Teratech Corp. (Burlington, MA, USA). Figure 2.7 shows this setup. A volunteer was asked to perform normal tidal breathing and we generated a respiratory signal which was displayed on screen. We then arbitrarily chose a point in the respiratory cycle and asked the volunteer to hold his breath at that point in the cycle. The volunteer then returned to normal tidal breathing and was asked to hold his breath again when he reached the same point in the respiratory cycle as indicated by our display. This process was repeated multiple times (four or more per experiment). At each of the breathholds we acquired an US image. We then acquired US images at random points in the respiratory cycle using the same breathhold technique. By comparing the NCC value for the images we show that our respiratory signal is consistent with internal organ movement due to respiration.

For all images acquired at the same point in the respiratory cycle the correlation coefficient between all possible pairs was between 0.85 and 0.93. For all the images acquired at random points in the respiratory cycle the correlation coefficient varied between 0.44 and 0.68. Figure 2.8 shows sample US images from one of the experiments. To estimate the maximal possible correlation between two US images we acquired six images of a volunteers ulna without any motion between probe and anatomy. This was repeated for both hands. For all the possible image pairs the correlation coefficient was between 0.94 and 0.98, although in practice it should have been 1. That is with this US machine a perfect correlation for exactly the same underlying physical situation is unlikely.

We conclude that the system we have developed produces results that are equivalent to those obtainable with a high end tracking system at a fraction of the cost. In addition we have shown that the signal generated by our system is indeed correlated

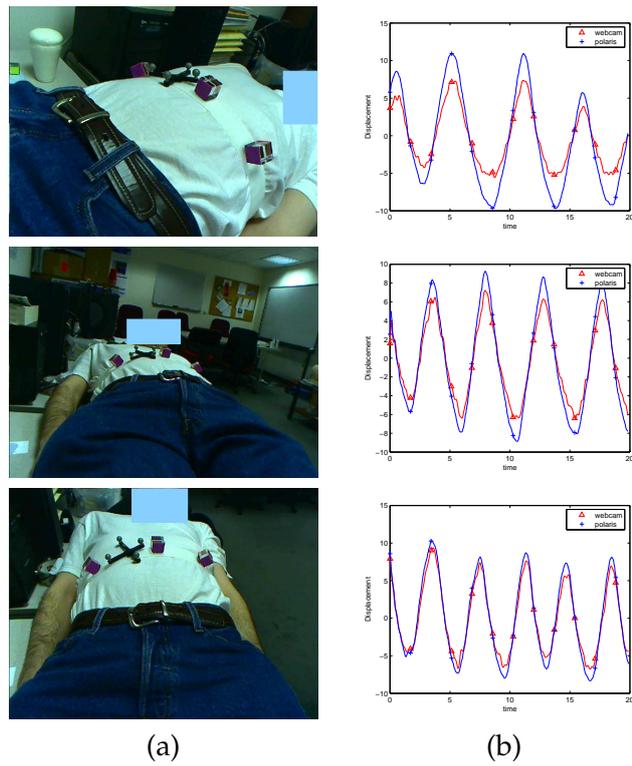


Figure 2.6: Effect of camera placement on system performance, (a) sample image from video stream; and (b) respiratory signal obtained by our method and using the Polaris Vicra for a volunteer exhibiting normal tidal breathing. In all cases the NCC was above 0.99.

with internal organ motion and can be used for respiratory monitoring.

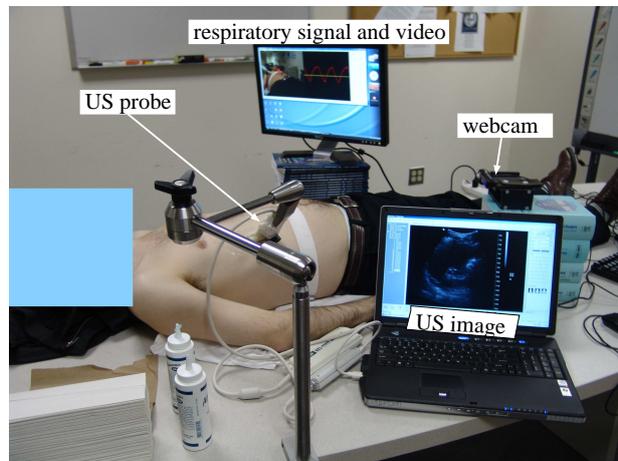


Figure 2.7: Experimental setup for US based assessment of respiratory signal generation.

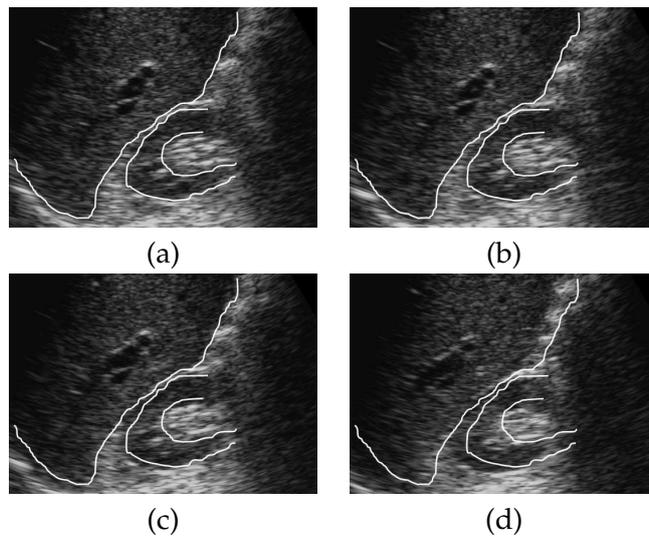


Figure 2.8: US images acquired for validating the respiratory signal. Distinct structures are marked in image (a) which was acquired at a specific point in the respiratory cycle and transferred to the other images. Images (b), and (c) were acquired subsequently at that same point in the respiratory cycle and image (d) at a random point. It is clear that images acquired at the same point in the respiratory cycle as defined by our system are correlated.

## 2.2 Respiratory signal generation from projection images

In section 2.1 we described a webcam-based method for respiratory monitoring. To acquire a 4D CBCT data set using that approach, synchronization with image acquisition is a prerequisite. As we did not have access to the internal workings of the image acquisition hardware we synchronized the two data streams based on the motion of the C-arm in the webcam video 3.3.2. This is an approximate synchronization as we do not know how many images are acquired prior to the detected C-arm motion.

In this section we describe a method for monitoring patient respiration that is implicitly synchronized with image acquisition. Our method is based on estimating the trajectory of a skin adhesive radio-opaque fiducial, 4mm CT-SPOTS from Beekley Corp., (Bristol CT, USA), that is visible in the acquired projection images.

At the basis of our method is the use of a linear approximation of the fiducial spatial motion. This approximation is motivated by empirical observations. We acquired surface motion data at seven locations on the thorax and abdomen of a volunteer using the Optotrak Certus (Northern Digital Inc., Waterloo, Ontario, Canada) tracking system. The surface motion was analyzed using Principle Component Analysis, indicating that there is a single dominant direction of motion. Table 2.4 summarizes this experiment.

Marker	1	2	3	4	5	6	7	mean
eig X	1.42	1.42	1.16	1.45	1.52	1.23	1.18	1.34
eig Y	6.19	7.72	5.5	5.01	3.87	4.77	4.48	5.37
eig Z	77.08	81.14	70.29	66.79	48.12	69.18	53.33	66.56

Table 2.4: Square root of eigenvalues resulting from PCA.

We next describe our approach to fiducial localization and tracking, our method for trajectory estimation and respiratory signal generation. Table 2.5 summarizes our approach.

### 2.2.1 Fiducial localization and tracking

Fiducial detection is performed automatically, after manual initialization. In the first image the user indicates a rectangular region of interest containing the fiducial. For each image we perform edge detection in the region of interest using the Canny edge detector [10]. The edge detector's high and low thresholds are set to 0.5 and 0.2 (pixel values in range  $[0, 1]$ ) and the smoothing parameter  $\sigma$  to 1.0. On the obtained edge image we perform circle estimation using the Random Sample Consensus Algorithm (RANSAC) [16].

RANSAC is a framework for dealing with outliers in a parameter estimation problem. Instead of using the whole data set to compute the model parameters minimal subsets are randomly selected, the model parameters are estimated for each subset and a consensus set is obtained by measuring the distance of each point to the solution. The process is repeated until the model with the largest support has been found.

<ol style="list-style-type: none"> <li>1. Input: Projection matrices, projection images</li> <li>2. Output: 3D motion of point</li> <li>3. Segmentation <ol style="list-style-type: none"> <li>a) Manual definition of ROI in first frame</li> <li>b) Apply Canny Edge Detection to ROI</li> <li>c) RANSAC circle fit to edge image</li> <li>d) Store center of circle</li> <li>e) Shift ROI to circle center and start from 2. with next image</li> </ol> </li> <li>4. Line reconstruction <ol style="list-style-type: none"> <li>a) Compute ray from camera center through segmented point for each image.</li> <li>b) Perform least squares estimation of the intersection of all rays, defines point on the line whose direction is <math>[0, 1, 0]</math>.</li> <li>c) Non linear estimation of line parameters.</li> </ol> </li> <li>5. Respiratory signal generation. <ol style="list-style-type: none"> <li>a) least square intersection of single ray with estimated line gives 3D point</li> <li>b) PCA on 3D points and projection on major axis is the respiratory signal.</li> </ol> </li> </ol>
---

Table 2.5: Summary of trajectory reconstruction algorithm

In our case, circle fitting, the minimal subset size is three. Table 2.6 summarizes the RANSAC framework.

To use the RANSAC algorithm we need to define a fitting function, a distance function, and a function to check for degenerate point configurations. Gander et al. [20] describe such a function for circle fitting. They linearize the parametric equation of the circle and extract the geometric parameters of the circle from the solution of the linearized problem. The circle equation is:

$$ax^T x + bx + c = 0 \quad (2.8)$$

and yields to the following linear system of equations

$$\begin{pmatrix} x_1^2 + y_1^2 & x_1 & y_1 & 1 \\ x_2^2 + y_2^2 & x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_n^2 + y_n^2 & x_n & y_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0 \quad (2.9)$$

which can be solved for  $(a, b, c)^T$ . The circle center and radius are

1. select  $N$  data items at random of whole set of size  $M$
2. estimate model parameter  $\vec{p}$
3. find how many data items  $K$  fit the model with parameter vector  $\vec{p}$  within a user given tolerance.
4. if  $K$  is big enough, estimate again with all inliers and exit with success.
5. repeat previous steps  $L$  times
6. fail if you get here

Table 2.6: RANSAC Algorithm

$$(x_{center}) = \left(-\frac{b_1}{2a}, -\frac{b_2}{2a}\right) \quad r = \sqrt{\frac{\|b\|^2}{4a^2} - \frac{c}{a}} \quad (2.10)$$

$a \neq 0$  otherwise the circle equation would become the line equation  $bx + c = 0$ . A 2D point,  $x$ , is an inlier if its distance from the circle is less than a pre-set threshold,  $d < threshold$ , with the distance given by:

$$d = \left| \|x_{center} - x\| - r \right| \quad (2.11)$$

To prevent the selection of degenerate subsets, three collinear points, we check the eigen-values of the covariance matrix of the sample set. If the first eigen-value is an order of magnitude larger than the second one, the sample set is discarded, as it is too close to a line.

Once the maximal consensus set is obtained we fit a circle to it, with the circle center being the fiducial location that we desire. We then shift the region of interest in the following image so that it is centered on the previously detected circle center and repeat the process. The distance between 2D fiducial locations in two consecutive images is small due to the  $0.4^\circ$  angular increment and speed of the C-arm rotation, allowing us to use a small region of interest for fiducial tracking.

### 2.2.2 Trajectory reconstruction and respiratory signal generation

The subject of linear trajectory estimation from a moving camera has been previously addressed in computer vision [2]. Using projective geometry the task is formulated as a set of linear equations whose solution is the trajectory. We have empirically found this method to be insufficiently accurate for our purposes, primarily due to the nature of the input which is only *approximately* linear. We describe this linear algorithm and then our improved non-linear trajectory reconstruction method.

Once we obtain the linear trajectory we compute the approximate spatial fiducial location for each image by taking the point on the backprojected ray that is closest to the line of motion. The respiratory signal is then generated using these 3D point locations via PCA. The 3D data points are projected onto the principle axis, the eigen-vector corresponding to the largest eigen-value, resulting in a one dimensional signal, our respiratory signal.

### 2.2.2.1 Linear Reconstruction

Avidan and Shashua [2] describe a linear algorithm to estimate the trajectory of an object moving in a straight line observed by a moving calibrated camera. Using Plücker coordinates the line  $L$  is described by a 6-tuple:

$$L = [\vec{n}; \vec{n} \times \vec{p}] \quad (2.12)$$

where  $\vec{n}$  is the line direction and  $\vec{p}$  a point on the line. To project a 3D Plücker line onto a 2D line in the image we have to reformulate the projection matrix  $P$ . The  $3 \times 4$  matrix  $P$  becomes a  $3 \times 6$  line projection matrix  $\tilde{P}$ .

$$\tilde{P} = \begin{bmatrix} P_2 \wedge P_3 \\ P_3 \wedge P_1 \\ P_1 \wedge P_2 \end{bmatrix} \quad (2.13)$$

$P_i$  denotes the  $i^{th}$  row of the projection matrix  $P$  and  $\wedge$  the wedge product. Geometrically this is equal to computing the 3D line of two intersecting planes [18] defined by the rows of the projection matrix. The resulting Plücker line is the new entry of  $\tilde{P}$ . Further a 2D point on a 2D line fulfills the constraint

$$\vec{x} \cdot \vec{l} = 0 \quad (2.14)$$

Plugging the 3D line  $L$  and the line projection matrix  $\tilde{P}$  into equation 2.14 gives the linear constraint

$$x_i \cdot \tilde{P}_i L = 0; \quad (2.15)$$

which can be solved for the unknown line  $L$  with  $k \geq 5$  views. Despite the elegance of the linear approach it is unstable in the presence of outliers, in our case the data is not linear only approximately so. This phenomenon has been previously noted by Horn [25]. That is, the use of linearization methods based on projective geometry often causes problems when it comes to decomposing the solution to Euclidean 3D entities.

### 2.2.2.2 Non-linear Reconstruction

To overcome the instability of the linear approach we propose to use non-linear estimation. We seek the line that minimizes the distance from all backprojected rays

defined by the camera and 2D fiducial locations. We define the following non-linear cost function:

$$\min_{\vec{a}, \vec{u}} \sum_i d(\vec{a}, \vec{u}, \vec{b}_i, \vec{v}_i) \quad (2.16)$$

where  $\vec{a}$  and  $\vec{u}$  are the position and direction of the desired line and  $\vec{b}_i, \vec{v}_i$  are the lines defined by camera center and 2D fiducial image location.

The shortest distance between two lines is:

$$d = \|\vec{x} - \vec{y}\| \quad (2.17)$$

$$\vec{x} = \vec{a} + \frac{[(\vec{b} - \vec{a}) \times \vec{v}] \cdot \vec{w}}{\vec{w} \cdot \vec{w}} \vec{u} \quad \vec{y} = \vec{b} + \frac{[(\vec{b} - \vec{a}) \times \vec{u}] \cdot \vec{w}}{\vec{w} \cdot \vec{w}} \vec{v} \quad (2.18)$$

$$\vec{w} = \vec{u} \times \vec{v} \quad (2.19)$$

where  $\vec{x}$  and  $\vec{y}$  are the closest points on each line. For the derivation of the equation see the appendix A. We minimize the cost function with a standard line search algorithm. The initial parameters for the optimizer are the least squares intersection of all rays for the point  $\vec{a}$  2.20 and  $[0, 1, 0]$  for the direction  $\vec{u}$  of the line. This assumes the patient is in the supine position and that the primary direction of fiducial motion is close to the anterior-posterior.

Given  $m$  lines we can compute the closest point  $\vec{x}$  to all of them by minimizing the sum of squared differences

$$\min_x \left\| (x - p_i) - n_i^T (x - p_i) n_i \right\|^2 \quad (2.20)$$

Differentiating for  $\vec{x}$  and equating to zero yields the following system of equations

$$\begin{pmatrix} m - \sum n_x^2 & -\sum n_x n_y & -\sum n_x n_z \\ -\sum n_x n_y & m - \sum n_y^2 & -\sum n_y n_z \\ -\sum n_x n_z & -\sum n_y n_z & -m - \sum n_z^2 \end{pmatrix} x = \sum p_i - n_i^T p_i n_i \quad (2.21)$$

which can be solved for  $\vec{x}$ . A solution exists if the matrix  $A$  is invertible. If the lines are parallel  $\det(A) = 0$  and  $A$  is not invertible.

To apply the line intersection we have to compute the backprojected rays from the pairs of projection matrices and image points. The camera center  $\vec{c}$  and the ray direction  $\vec{d}$  are extracted from the projection matrix

$$\vec{c} = -P_{3 \times 3}^{-1} p_4 \quad (2.22)$$

$$\vec{d} = P_{3 \times 3} \vec{u} \quad (2.23)$$

with  $P_{3 \times 3}$  the left 3x3 matrix of the projection matrix  $P$ ,  $p_4$  its last column and  $\vec{u}$  the homogeneous image point [59].

### 2.2.3 Results

To evaluate the robustness of our fiducial detection we used an anthropomorphic respiratory phantom [32]. The spherical fiducial was attached to the phantom's thorax and images were acquired while the phantom was respiring. The fiducial detection algorithm was run and resulted in successful detections in all data sets, as judged by visual inspection. Figure 2.9 shows the experimental setup and sample images. One sees that even when the fiducial overlaps with a linear structure and the edge detection algorithm can only detect half of the circle, the RANSAC estimation is able to correctly fit a circle.

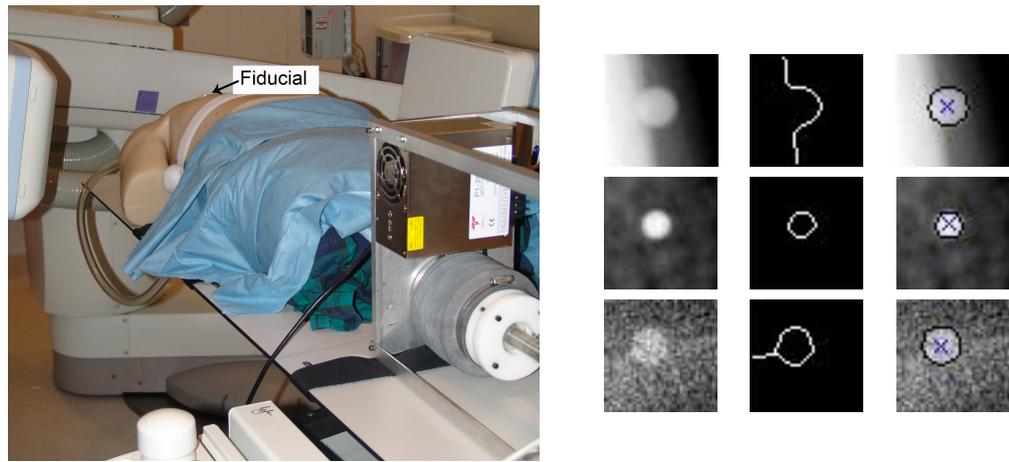


Figure 2.9: Left: Fiducial tracking setup. Right: X-Ray image of fiducial, detected edges and fitted circle with circle center for three different C-Arm positions.

In two simulation studies we compared the linear and non-linear implementation of the trajectory reconstruction. The first data set to which we applied the algorithms is a sampled set of points along a straight line, projected to the image plane using the projection matrices of our C-Arm system. The position of the points on the line was modulated by a sine and distributes over 20 mm. The parameters of the line are

$$p = (0, 0, 0) \quad (2.24)$$

$$d = (0.0, 0.7, 0.7) \quad (2.25)$$

In this noise free scenario both algorithms perform well and estimate the line parameters correctly. Then we added normal distributed noise with standard deviation of  $\sigma = 0.5$  pixel onto the projected point location and estimated the line parameters again. We ran the experiment a 100 times and give the statistics in table 2.2.3. We applied PCA on the estimated directions and transformed them into their eigen-frame to measure its standard deviation. The standard deviation of the direction estimated with the linear method is high and shows high uncertainty in the estimation. In the last row of table 2.2.3 we give the minimal, medial and maximal distance of the esti-

method	linear	non-linear
mean direction	(0.0202, 0.6428, 0.6509)	(0.0051, 0.7072, 0.7058)
std direction	(0.0016, 0.1327, 0.3831)	(0.0034, 0.0107, 0.0388)
distance	(38.0101, 120.1907, 455.6870)	(0.0068, 0.0502, 102.0251)

Table 2.7: Results of the simulation study comparing linear and non-linear algorithm in the presence of noise. First row is the mean of the trajectory direction, second row is the standard deviation of the direction. Last row shows the distance (minimum, median, maximum) of the estimated line point to the ground-truth line.

mated point  $p$  to the actual line, which also shows large errors in the linear algorithm.

The data for the second study was obtained by tracking the skin surface of a volunteer using an Optotrak Certus system. Seven IR-LED's were attached to the volunteer in the thoracic-abdominal region and data was acquired at a rate of 60Hz with the volunteer in the supine position. Then we projected the 3D points using the projection matrices of the C-arm, which simulates the rotation and projection of the imaging system. These 2D points serve as input for the linear and non linear estimation. Figure 2.10 shows the original 3D point cloud, from which we obtained the 2D points. The green line is the linear estimate and the blue one the non linear. One sees that the linear estimated line does not pass through the point cloud, whereas the non linear estimate does.

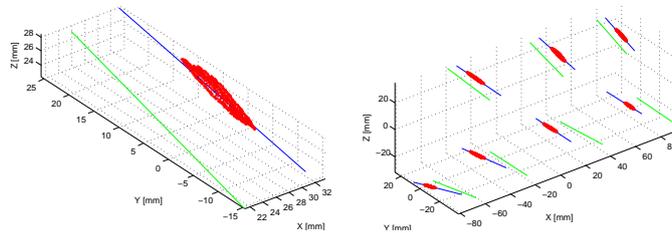


Figure 2.10: Line reconstruction. Red dots are the true 3D coordinates of the fiducial. Blue line was reconstructed with non linear algorithm, green line was reconstructed using Avidan and Shashua's algorithm.

To evaluate the quality of the estimated respiratory signal we use a motion platform that follows the same respiratory motion from the simulation study. We acquired cone-beam CT images of a fiducial attached to a motion platform performing the respiratory motion as shown in Figure 2.11. To use the data from the simulation study for the motion platform we had to resample the signal from  $60Hz$  to  $7.8125Hz$ , a sample every  $128ms$ . This was required due to hardware limitations.

We acquired two scans, one with the fiducial approximately in the mid-plane of the reconstruction volume and a second scan with the fiducial on the border of the volume. Sample images from both scans are shown in Figure 2.12.

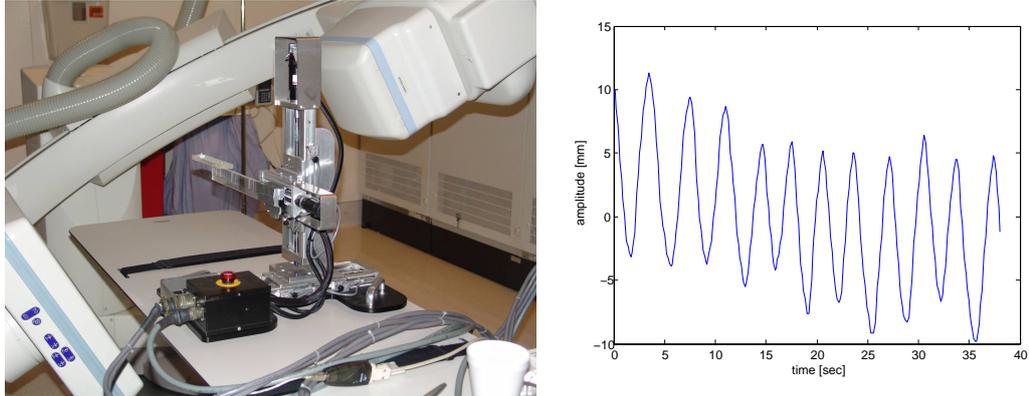


Figure 2.11: Setup with motion platform and CBCT scanner. Respiratory signal generated by performing PCA on the motion platform data

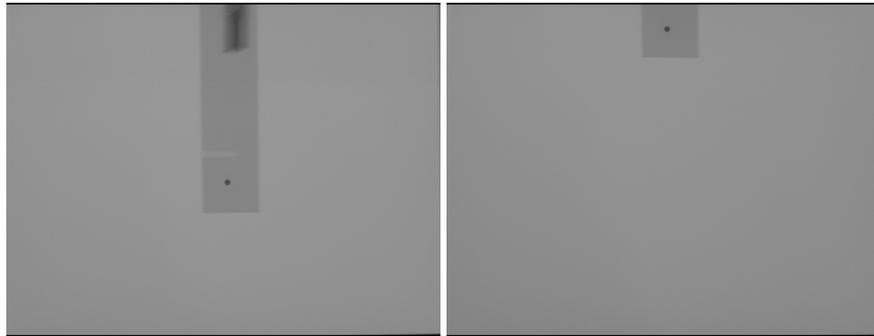


Figure 2.12: X-Ray images from two motion platform experiments. The fiducial is once in the image center and once at the boarder

As we know the exact 3D motion we can generate a ground truth respiratory signal and compare it to the estimated signal using our method. We observed that our signal is in phase with the ground truth signal, but suffers from inaccuracies in amplitude close to the anterior-posterior imaging setup where the primary direction of motion is perpendicular to the image plane. At this position (after half of the scanning time) the angle between the trajectory and the ray is small as seen in Figure 2.13. Figure 2.11 shows the experimental setup and the known ground truth respiratory curve.

Furthermore we compared the respiratory signals we obtained from each Optotrak marker attached to the volunteer. Table 2.2.3 shows the correlation of marker pairs by displaying the normalized cross correlation coefficient. The signals are correlated to 99% regardless of marker location.

Whereas the placement of the marker has no impact on the respiratory signal, the line estimation algorithm can be affected by the markers position. For example, when all rays used for the line estimation are lying in a common plane, it is impos-

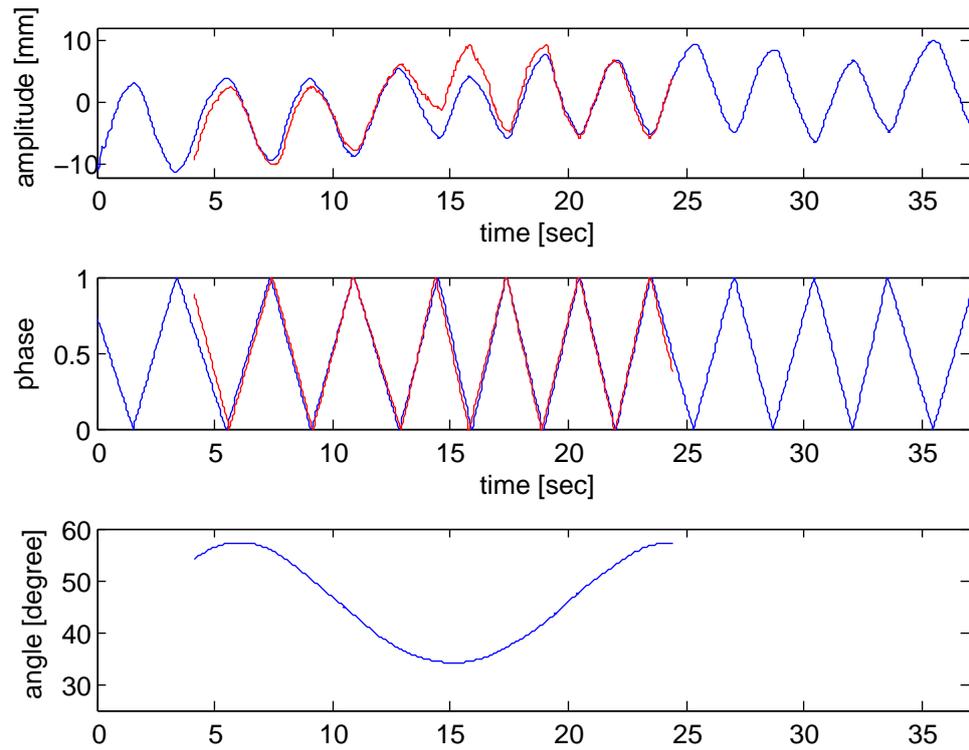


Figure 2.13: Amplitude and phase signal of ground truth overlaid with reconstructed signal. Ray intersection angle with reconstructed line

sible to reconstruct the line. Every line in the plane will have a distance of zero to the set of lines. For optimal trajectory reconstruction it is best to image the line with imaging planes parallel to it. Unfortunately this is impossible in the clinical setup, with the patient in supine position and the C-arm rotating in the axial plane.

Marker	1	2	3	4	5	6	7
1	1.0000	0.9999	0.9994	0.9996	0.9982	0.9991	0.9968
2		1.0000	0.9995	0.9996	0.9981	0.9992	0.9969
3			1.0000	0.9992	0.9990	0.9995	0.9985
4				1.0000	0.9986	0.9995	0.9976
5					1.0000	0.9992	0.9991
6						1.0000	0.9989
7							1.0000

Table 2.8: Correlation of respiratory signals computed from differently attached markers

## 3 Tomographic Reconstruction

### 3.1 Reconstruction

Tomographic reconstruction refers to the reconstruction of cross sectional images from projections of an object obtained from different angles [28]. Conventional CT scanners possess an X-Ray source, emitting a fan-shaped beam, which rotates around the object of interest and projects 2D cross sections onto a 1D detector. The original 2D image can be recovered using the projection data and the knowledge of the source and detector positions. A whole 3D scan is captured by stacking up multiple 2D slices.

A similar principle is Cone Beam CT [59]. The difference is that the beam shape is similar to a cone, projecting a 3D volume onto a 2D detector. By projecting the object from different angles, the whole 3D object can be reconstructed at once, similar to the slices in the conventional CT, but without the need of combining multiple 2D slices into a 3D volume.

In general we differentiate between two kinds of reconstruction algorithms, analytical and iterative. Analytical reconstruction relies on the inversion of the Radon transform, modeling the projection step as:

$$P_{\theta}(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x \cos(\theta) + y \sin(\theta) - t) \quad (3.1)$$

where  $f(x, y)$  is the image,  $\delta$  the delta function,  $\theta$  the projection angle and  $t$  the detector pixel. This equation models the line integral generation for a parallel beam projection. In order to obtain the image  $f(x, y)$  from its projections  $P_{\theta}(t)$  we have to apply the Fourier Slice Theorem [28].

**Fourier Slice Theorem.** *The Fourier transform of a parallel projection of an image  $f(x, y)$  taken at angle  $\theta$  gives a slice of the two-dimensional transform,  $F(u, v)$ , subtending an angle  $\theta$  with the  $u$ -axis. In other words, the Fourier transform of  $PO(t)$  gives the values of  $F(u, u)$  along line  $BB$  in Fig. 3.1*

In other words, if we take the Fourier transform of the projection data and assemble it according to the projection angle in the Fourier space of the sought image, we can perform the inverse Fourier transform on it and receive the original image. This is the foundation for all direct algorithms. Figure 3.1 illustrates this principle.

Another way of performing reconstruction is to use iterative methods. Reconstruction is described by a system of linear equations whose solution yields the desired reconstruction. For a reconstruction volume of  $256^3$  voxel, the system matrix has a size of  $256^3 \times 256^3$ . Obviously the matrix is too large to be solved directly by standard inversion schemes. The standard approach is to use an iterative algorithm, which

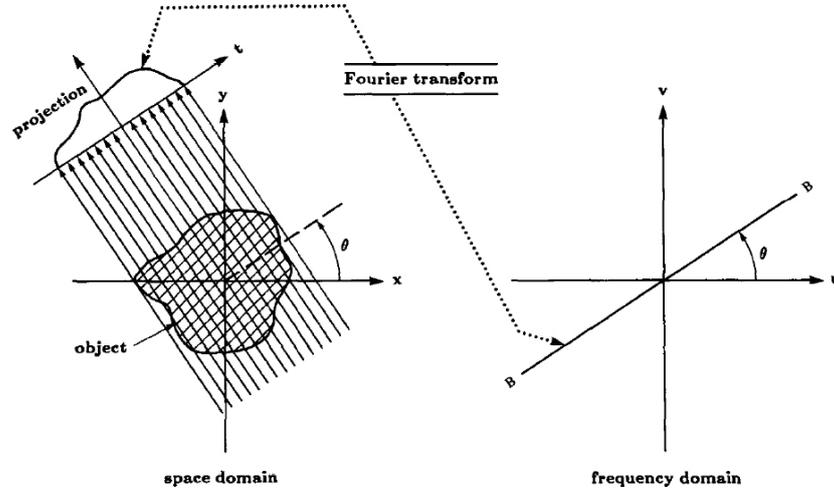


Figure 3.1: Fourier Slice Theorem, adopted from [28]

was first proposed by Kaczmarz [27] and is explained in detail in Kak's book [28]. This is the approach utilized in this thesis.

Both methods have their advantages and disadvantages. Computationally efficient implementations of the Fourier Transform utilized by the direct approach result in short reconstruction times ( $< 5$  min) and are thus the dominant method in the clinical environment. Iterative algorithms exhibit longer running times. In recent years this drawback has been mitigated by faster implementations of projection and back-projection utilizing texturing hardware [36, 56] and programmable GPUs [51]. The advantage of iterative algorithms is that they can deal with less projection images than the direct method [13]. Guan and Gordon [23] show this in the 2D parallel beam case. They show that  $M_{Alg} = M_{FBP}/2$  which means that for an algebraic reconstruction only half of the projection images are needed. To have a well determined set of equations for algebraic reconstruction we need one equation for each unknown voxel value. Assuming we are reconstructing a spherical volume with a diameter of  $n_s$  samples, the number of unknowns is given by the volume formula of the sphere  $V = \frac{1}{6}\pi n_s^3$ . A projection image of size  $n_s^2$  pixels contains  $A = \frac{1}{4}\pi n_s^2$  relevant pixels thus we need

$$\frac{V}{A} = 0.67n \quad (3.2)$$

projection images for a well determined linear system of equations [37]. In practice we have to acquire 172 images for a volume of  $n = 256$  voxels in diameter. This equation implies that the projection images are non-redundant, meaning that each image carries new information. For the parallel beam case this means an equi-angular sampling over a circular trajectory of 180 degrees. There are several scenarios where this cannot be accomplished. For the application of retrospective gated reconstruction we cannot guarantee this and thus are facing a limited (few) view projection prob-

lem. This is a current research topic, which has been drawing researches attention since the early 1970's.

Rangayyan et. al. [45] describe early attempts to solve the problem for the 2D case. Andersen [1] tries to solve the limited view projection problem for the 3D case and suggests using algebraic methods. Most researchers add a priori knowledge [44, 49], like an approximation of the objects intensity values or assumptions on the smoothness of the image to constrain the under determined system. More recently, statistical reconstruction methods [26, 54] have been employed for limited view reconstruction, which also allow to specify a priori information. In this thesis we only address the limited view problem implicitly 3.2. Adding a priori knowledge is complicated in our case as we are reconstructing the abdominal region with organs undergoing motion and deformation throughout the data acquisition process.

In algebraic reconstruction the problem of estimating the volume is formulated as a set of linear equations.

$$W\vec{f} = \vec{p} \quad (3.3)$$

where  $W$  is called the system matrix. Written as a summation we have

$$\sum_{j=1}^N w_{ij} f_j = p_i, \quad i = 1, 2, \dots, M \quad (3.4)$$

To solve this iteratively we can apply Kaczmarz's [27, 28] method. It takes an initial guess and projects it consecutively on the hyperplanes given by the linear system

$$\vec{f}^i = \vec{f}^{(i-1)} - \frac{\vec{f}^{(i-1)} \cdot \vec{w}_i - p_i}{\vec{w}_i \cdot \vec{w}_i} \vec{w}_i \quad (3.5)$$

where  $w_i = (w_{i1}, w_{i2}, \dots, w_{iN})$ . Rewriting 3.5 for a single voxel and introducing the relaxation factor  $\lambda$  yields the equation for the algebraic reconstruction (ART)

$$f_j^k = f_j^{(k-1)} + \lambda \frac{p_i - \sum_{l=1}^N w_{il} f_l^{(k-1)}}{\sum_{l=1}^N w_{il}^2} w_{ij} \quad (3.6)$$

$k$  is the iteration step,  $j$  the index of the voxel,  $i$  the index of the pixel in the projection image and  $l$  iterates over all voxels. The term

$$\sum_{l=1}^N w_{il} f_l^{(k-1)}$$

is the line integral from the source to pixel  $p_i$  summing up all voxels the ray passes and

$$\sum_{l=1}^N w_{il}$$

is the sum of all contribution factors traversed by the ray and is related to the ray length. Figure 3.2 illustrates the process.

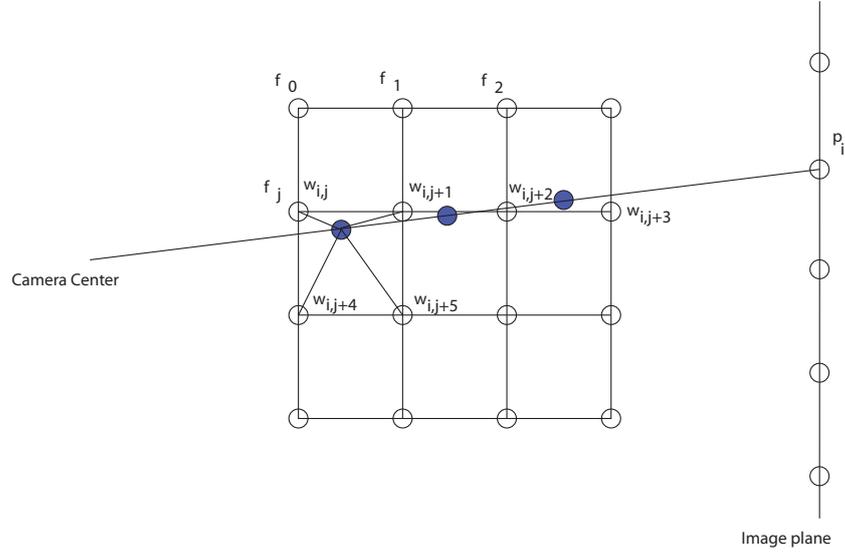


Figure 3.2: Reconstruction scheme. A ray passes through the reconstructed slice from the camera center to the image pixel  $p_i$ . The voxels  $f_j$  are multiplied by their contribution factor  $w_{ij}$  and summed up.

A major refinement to ART is the simultaneous algebraic reconstruction (SART). SART takes into account that a voxel can contribute to several pixels in the image.

$$f_j^k = f_j^{(k-1)} + \lambda \frac{\sum_{p_i \in P_\varphi} \left( \frac{p_i - \sum_{l=1}^N w_{il} f_l^{(k-1)}}{\sum_{l=1}^N w_{il}} \right) w_{ij}}{\sum_{p_i \in P_\varphi} w_{ij}} \quad (3.7)$$

It sums over all pixels  $p_i$  in a projection image  $P_\varphi$ . If voxel  $f_j$  contributes to pixel  $p_i$ ,  $w_{ij} > 0$  the correction factor will be computed. The nominator of equation 3.7 is similar to ART. At the end the correction factor is divided by

$$\sum_{p_i \in P_\varphi} w_{ij}$$

the sum of the contribution factors associated with each pixel in the image. The implementation of SART is straight forward and is shown in Table 3.1

The main components of the reconstruction algorithm are the computations of ray projection and back-projection into the volume. The ray  $r$ , corresponding to 2D pixel  $u$ , is computed using the projection matrix  $P$  as shown in section 2.2:

$$\vec{c} = -P_{3 \times 3}^{-1} p_4 \vec{d} \quad (3.8)$$

$$\vec{d} = P_{3 \times 3} \vec{u} \quad (3.9)$$

$$\vec{r} = \vec{c} + \lambda \vec{d} \quad (3.10)$$

1. initialize  $\vec{f}$  (set all entries to zero)
2. iterate till convergence
3.
  - a) project  $f$  according to  $P_i$  resulting in  $q_i$
  - b) compute correction  $c_i = \lambda(p_i - q_i)$
  - c) back-project correction

Table 3.1: SART algorithm.

To sample the volume along the ray we compute its intersection points with the axis aligned bounding box, which is centered at the origin  $(0, 0, 0)$ . The intersection test is performed by testing the ray against each face of the box and computing its entry  $x_s$  and exit point  $x_e$ . Numerous algorithms can be found in the literature (e.g. [61]), presenting also optimization strategies. In order to sum up the voxel values along the ray, the points  $x_s$  and  $x_e$ , given in the world frame, are expressed in the volume coordinate system. Now we can sum up the voxels by sampling the ray

$$x = x_s + \lambda(x_e - x_s) \quad (3.11)$$

The computed point does not correspond to a grid point thus we calculate its contribution by interpolation. In this thesis we use tri-linear interpolation, a common compromise between computational complexity and quality of interpolation. A voxel's contribution is the weighted sum of its eight neighbors. Therefore we compute the floating part  $(x_d, y_d, z_d)$  of the coordinate  $(x, y, z)$ .

$$\begin{aligned} x_d &= x - \lfloor x \rfloor \\ y_d &= y - \lfloor y \rfloor \\ z_d &= z - \lfloor z \rfloor \end{aligned} \quad (3.12)$$

The interpolated value is then

$$\begin{aligned} v &= f(\lfloor x \rfloor, \lfloor y \rfloor, \lfloor z \rfloor) \cdot (1 - x_d) \cdot (1 - y_d) \cdot (1 - z_d) \\ &\quad + f(\lceil x \rceil, \lfloor y \rfloor, \lfloor z \rfloor) \cdot x_d \cdot (1 - y_d) \cdot (1 - z_d) \\ &\quad + f(\lfloor x \rfloor, \lceil y \rceil, \lfloor z \rfloor) \cdot (1 - x_d) \cdot y_d \cdot (1 - z_d) \\ &\quad + f(\lfloor x \rfloor, \lfloor y \rfloor, \lceil z \rceil) \cdot (1 - x_d) \cdot (1 - y_d) \cdot z_d \\ &\quad + f(\lceil x \rceil, \lfloor y \rfloor, \lceil z \rceil) \cdot x_d \cdot (1 - y_d) \cdot z_d \\ &\quad + f(\lfloor x \rfloor, \lceil y \rceil, \lceil z \rceil) \cdot (1 - x_d) \cdot y_d \cdot z_d \\ &\quad + f(\lceil x \rceil, \lceil y \rceil, \lfloor z \rfloor) \cdot x_d \cdot y_d \cdot (1 - z_d) \\ &\quad + f(\lceil x \rceil, \lceil y \rceil, \lceil z \rceil) \cdot x_d \cdot y_d \cdot z_d \end{aligned} \quad (3.13)$$

More advanced interpolation schemes can be integrated, to improve the reconstruction quality but they lead to longer computation times [37].

The back-projection of the correction image is carried out the reverse way. While traversing the ray the correction factor  $v$  is added to the neighboring voxels according to the weighting factor of the tri-linear interpolation and the length  $l$  of the ray. Similar to 3.13 we get

$$f([x], [y], [z]) = f([x], [y], [z]) + \frac{v}{l} \cdot (1 - x_d) \cdot (1 - y_d) \cdot (1 - z_d) \quad (3.14)$$

for the first neighbor. The values of the remaining seven neighbors are computed in the same way with their specific interpolation weights.

Our current CPU based software approach is very time consuming with a reconstruction of a  $128 \times 128 \times 110$  volume using  $543\,620 \times 480$  projection images taking 60 minutes for one iteration. This was acceptable for our initial investigation into 4D CBCT reconstruction, and is easily improved using methods based on software implementations for the graphical processing unit (GPU) [36].

## 3.2 Binning and Weighting

In order to acquire a binned reconstruction we investigated two strategies. First a simple binning strategy, using only images falling into one bin to reconstruct an image and second a weighting strategy assigning a weight factor to each image.

For the binning strategy we have to define a bin value  $v$  and a bin width  $w$  (Figure 3.3). Every projection image corresponding to a respiratory value  $r$

$$r \leq \left\| v - \frac{w}{2} \right\| \quad (3.15)$$

is sorted into the bin  $v$ . Thus we get approximately 5-10 groups of projection images along the C-arm trajectory. For end inspiration and expiration we get less groups but they contain more images. The fact that we obtain projections from only a few directions results in severe artifacts in the reconstruction, which is why we pursue a weighting based reconstruction strategy.

Instead of performing the reconstruction with a subset of projection images that corresponds to a specific respiratory phase we use all images, but with weighting factors according to their distance from the specific phase. This is motivated by the fact, that images which fall outside of the bin also carry valid information, for example the spine which does not move due to respiration.

The weighting factor is easily integrated into the SART algorithm. We only have to replace the relaxation factor  $\lambda$  by the weighting factor. Using standard SART all  $n$  images contribute  $n\lambda$  to one iteration. This value is now partitioned with respect to the respiratory signal. Therefore we compute first the distance of each respiratory value to the bin value

$$d = \|v - r\| \quad (3.16)$$

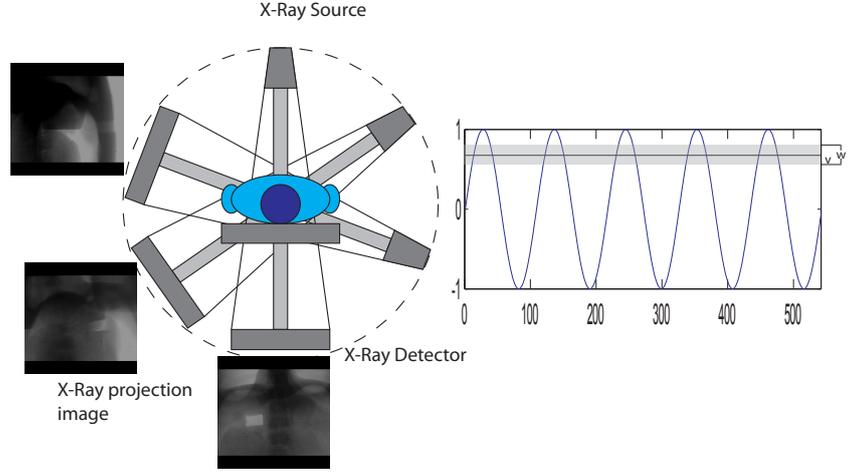


Figure 3.3: Scheme of gated reconstruction. According to the respiratory signal only projection images with the same respiratory value are selected for the reconstruction. This leaves gaps on the C-Arm trajectory and missing images for the reconstruction.

In order to give bin images a higher rating we introduce a ratio for the weighting factors. Images belonging to the the bin contribute  $c$  times more than the images belonging to the rest, thus we have to multiply the bin weights by factor  $t$ .

$$t = \frac{\sum_{d > \frac{w}{2}} d}{\sum_{d \leq \frac{w}{2}} d} c \quad (3.17)$$

Finally the weighting factors are scaled by factor  $s$  to contribute  $n\lambda$  to one iteration.

$$s = \frac{n\lambda}{\sum_{d \leq \frac{w}{2}} d t + \sum_{d > \frac{w}{2}} d} \quad (3.18)$$

The weighting factors become

$$d_{(\leq \frac{w}{2})} = t d_{(\leq \frac{w}{2})} s \quad (3.19)$$

$$d_{(> \frac{w}{2})} = d_{(> \frac{w}{2})} s \quad (3.20)$$

### 3.3 Results

We evaluated our algorithms on simulated and real data. The simulated data is a digital phantom consisting of spheres. The projection images, also known as Digitally Reconstructed Radiographs (DRR), were computed using a ray casting approach with the camera parameters (projection matrices) corresponding to those obtained from our clinical C-arm imaging system. The real data is generated using an anthropomorphic respiratory phantom [32]. The phantom was scanned while it was respiring using the C-arm system, yielding the projection images. In both cases the reconstructed images are all of size  $148 \times 148 \times 110$  voxels with an isotropic voxel size of  $1.6\text{mm}$ . The size of the projection images is  $620 \times 480$  pixel and the reconstructions were performed with four iterations.

#### 3.3.1 Digital Phantom

The digital phantom is a combination of spheres of different size and intensity and is described by the parametric equations of the spheres. This enables efficient and accurate computation of the projection images, by intersecting the camera rays with the parametric surfaces. The projection is generated by iterating over the image, computing the ray for the pixel and intersecting it with the parametric surface (see appendix B for derivation of ray-sphere intersection). The intensity value of the projection is the length of the segment between both intersection points multiplied by the intensity of the sphere. This means we are assuming mono-energetic X-rays [28] and a uniform distribution of the attenuation factors inside the spheres. For the specific parameters of the phantom see table 3.3.1. Figure 3.4 shows two example projection images.

sphere #	center ( $mm$ )	radius ( $mm$ )	intensity
1	(0, 0, 0)	80	1
2	(0, 20, 20)	20	0
3	(20, -20, -20)	30	0.5

Table 3.2: Parameters of the digital phantom

To simulate motion during the image acquisition process we translate the sphere centers by a sinus-modulated vector. For this specific case we have chosen the vector  $(0, 0.7, 0.7)$  modulated by  $\sin(2\pi 5)10$ . Figure 3.5 shows the setup and the modulation function (respiratory signal) with marked reconstruction bins.

Depending on the respiratory signal and the binning parameters we get different bin sizes. In our examples we have approximately the following bin sizes. The respiratory signal consists of about five cycles, thus we have about five end-inhalation positions, five end-exhalation positions and ten positions in between. Respiration stays usually longer in the end-inhalation and end-exhalation phase and gives about 30 frames per position. 30 frames span an C-arm angle of  $12^\circ$  when the C-arm increment is  $0.4^\circ$  per frame. The phases in between have approximately half the images

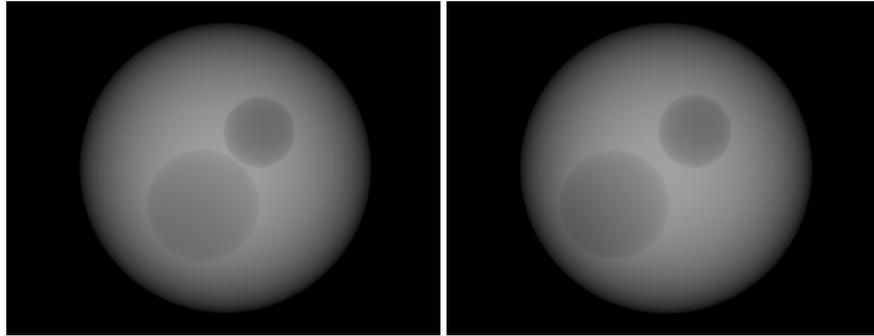


Figure 3.4: Digital phantom projections

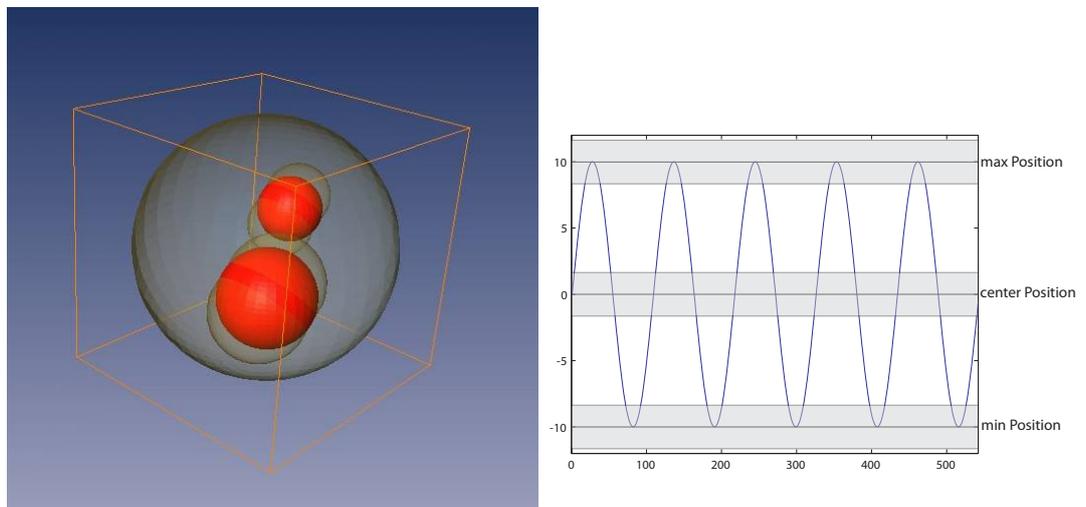


Figure 3.5: Digital Phantom setup and modulation function of the small spheres

per position but are taken from twice the C-arm positions. Thus they are better distributed over the C-arm trajectory. For the end-inhalation and end-exhalation phase about 80 frames are in between two corresponding C-arm positions and span an angle of  $32^\circ$  of missing projection data. This is the reason for reconstruction artifacts and low contrast.

Table 3.6 shows the reconstruction results. The first row shows the reconstruction of the phantom without motion. The second row is generated by applying the motion to the spheres and reconstructing them using all projections as input to the SART algorithm. The motion leads to streaking artifacts and blurring of the sphere contours. Reconstruction using our weighting approach, third row, or a binned subset of the projection images, fourth row, reduces blurring of the contours but introduces artifacts due to the use of fewer projections. The weighting and subset approaches reconstruct the moving spheres in the center position (corresponding to the static case), where the respiratory value is zero (see Figure 3.5).

The difference between the weighted and the subset algorithm is visible when we

reconstruct from only a few projection directions, which is the case for end-inhalation and end-exhalation (see Figure 3.7). In these cases, projection images are only available from five directions. One sees that using the weighting scheme increases the contrast and reduces the artifacts slightly, letting the larger sphere ( $r = 80$ ) appear round. The subset method has problems reconstructing this sphere and leaves the impression of blended polygons.

### 3.3.2 Respiratory phantom

To test the algorithm on real data we performed reconstruction using data acquired with an anthropomorphic respiratory phantom [32]. The phantom consists of a torso with inflatable lungs connected to a pump. The respiratory signal is obtained with our video based method 2.1 and the synchronization with the C-arm is performed manually by selecting start and end of the scan in the video images. We have chosen the end-inhalation and exhalation bins, corresponding to the respiratory values  $-4$  and  $4$ . Figure 3.3.2 displays a coronal ( $xy$ -plane) slice of both volumes. The bottom left image is the first slice with the segmented diaphragm border from both volumes overlaid.

### 3.3.3 Motion Platform Experiment

We also tested our reconstruction algorithm on a motion platform which was programmed to follow a human respiratory pattern. The respiration was recorded from a volunteer using the Optotrak Certus system, as described in section 2.2. As reconstruction object we attached a foam liver, containing contrast enhanced nodules, to the arm of the motion platform, shown in Figure 3.9. In addition a spherical radio opaque marker was glued to the arm, which we used to compute the respiratory signal applying the algorithm presented in chapter 2.2. We activated the motion platform for 40 seconds and acquired a 20 seconds scan of the phantom. In order to synchronize the ground truth respiratory signal we registered the estimated signal with the ground truth signal. This allows to reconstruct the liver phantom using the ground truth amplitude signal and the phase signal and to study the differences between both approaches.

Figure 3.10 shows the difference between the amplitude and the phase signal. The phase signal was scaled for visualization purposes. Apparently a real respiratory signal introduces several problems for reconstruction. The amplitude varies over the 20 seconds what makes it impossible to reconstruct some states. For example reconstructing the volume at the respiratory value of eight leaves only three projection direction which is not sufficient. Commonly researchers use the phase signal to overcome the problem of missing projections. This approach introduces another problem. Projection images associated with one bin, do not correspond to the same amplitude value and thus not to the same position of the volume. By using phase binning we have to expect blurring due to the falsely assigned projection images. Figure 3.11 shows two axial slices reconstructed once with amplitude binning and once with phase binning. The differences are not obvious in the original slice. But

inspecting the thresholded slices, a slight improvement in the amplitude binned image can be seen. The boarder of the spherical nodule and the boarder of the arm of the motion platform appear sharper.

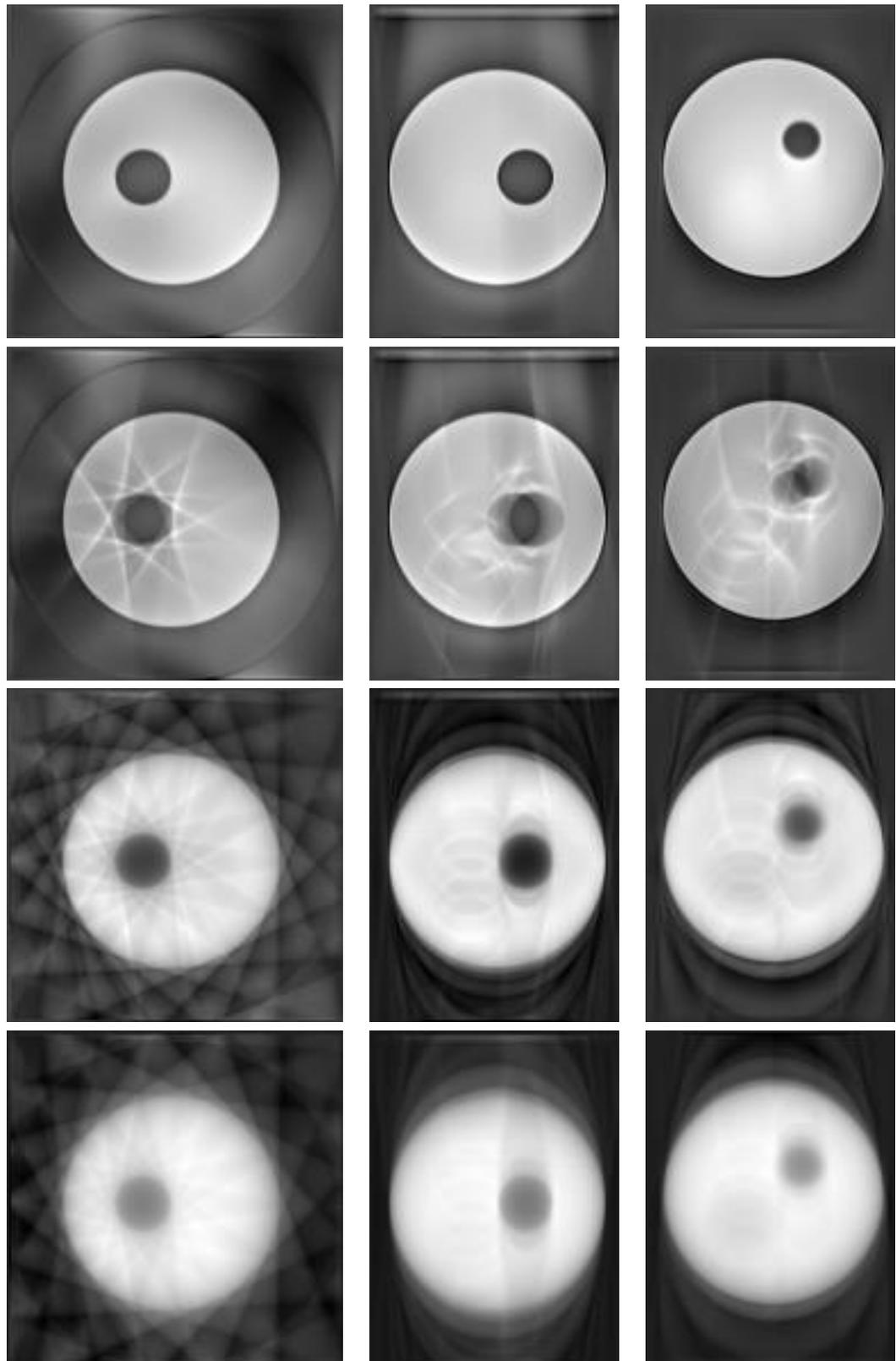


Figure 3.6: Reconstruction results from simulation study. Left to right:  $xy$ ,  $xz$ , and  $yz$  planes. Top to bottom: no motion, with motion, weighted reconstruction, subset reconstruction

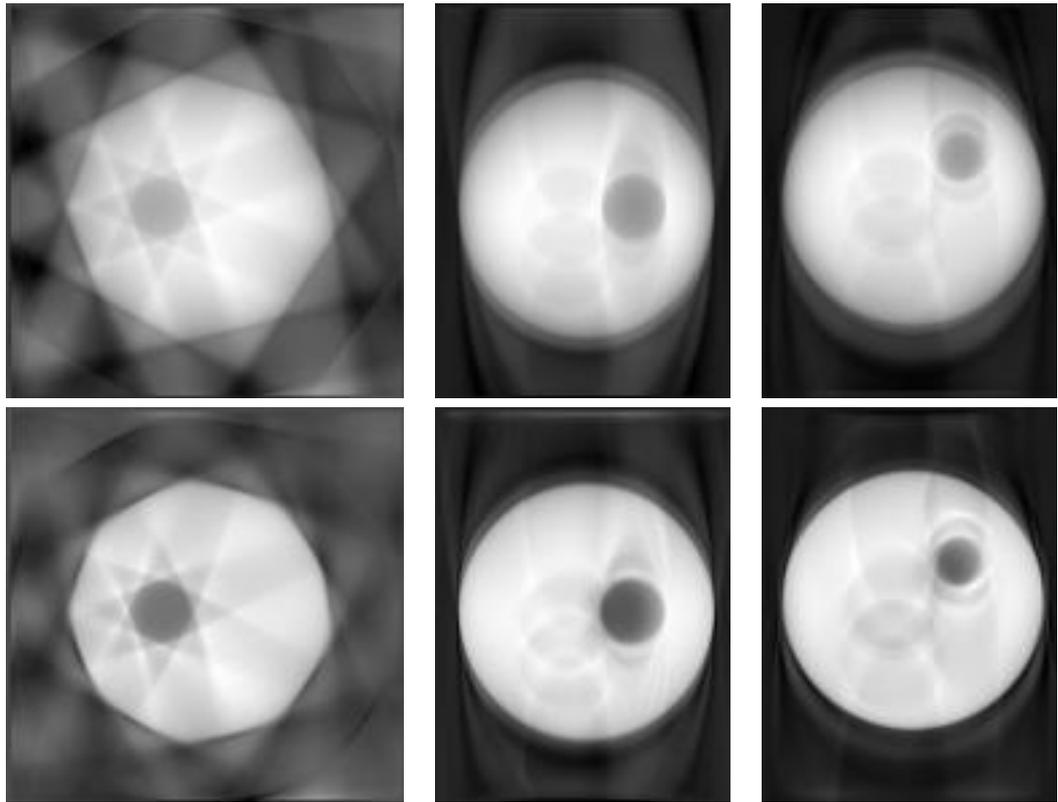


Figure 3.7: Reconstruction of bin  $-10$ . Left to right:  $xy$ ,  $xz$ , and  $yz$  planes. Top, subset reconstruction, bottom weighted reconstruction.

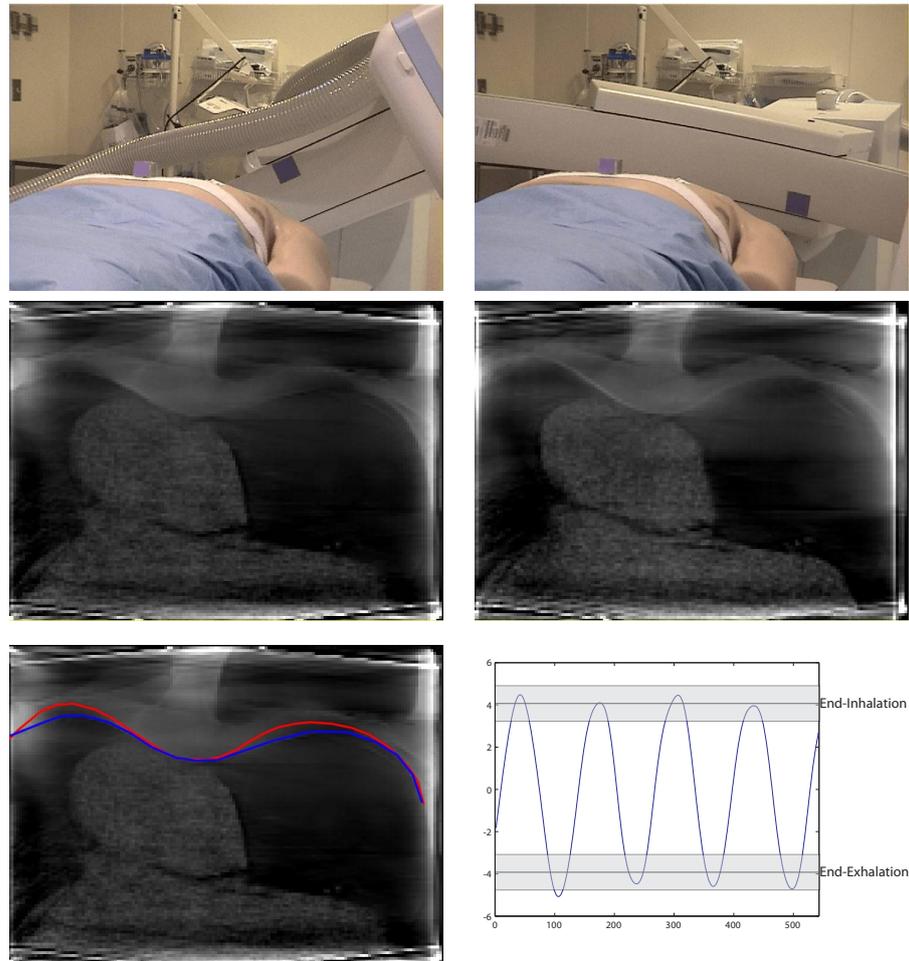


Figure 3.8: Images of C-Arm in start and end position used to synchronize the C-Arm frames with the respiratory motion. Coronal slice from two reconstructed volumes. From left to right and top to bottom: end-inhalation, end-exhalation, segmented diaphragm border overlaid on end-inhalation slice, respiratory signal with marked bins



Figure 3.9: Motion platform with attached foam liver centered in C-arm system

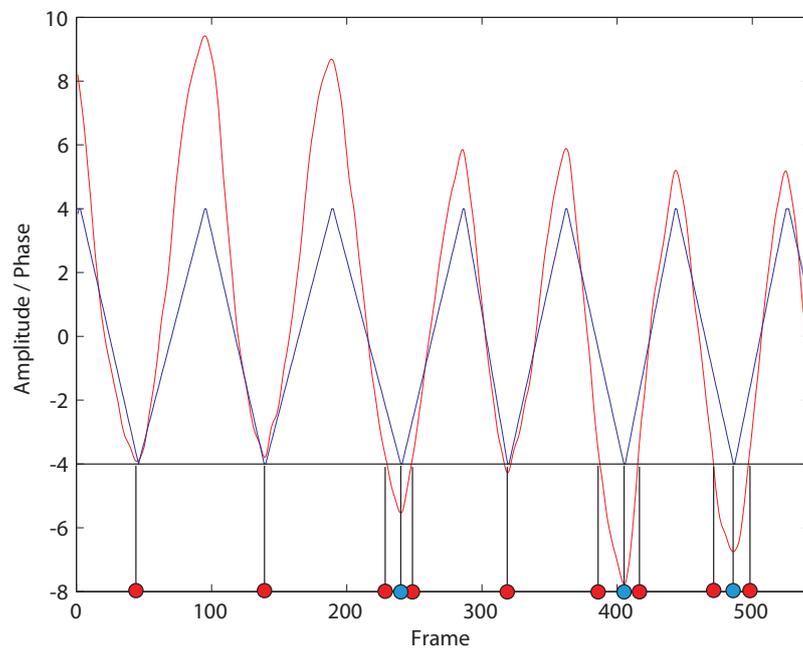


Figure 3.10: Different binning results from amplitude (red) and phase (blue) signal. The phase signal was scaled for visualization purposes.

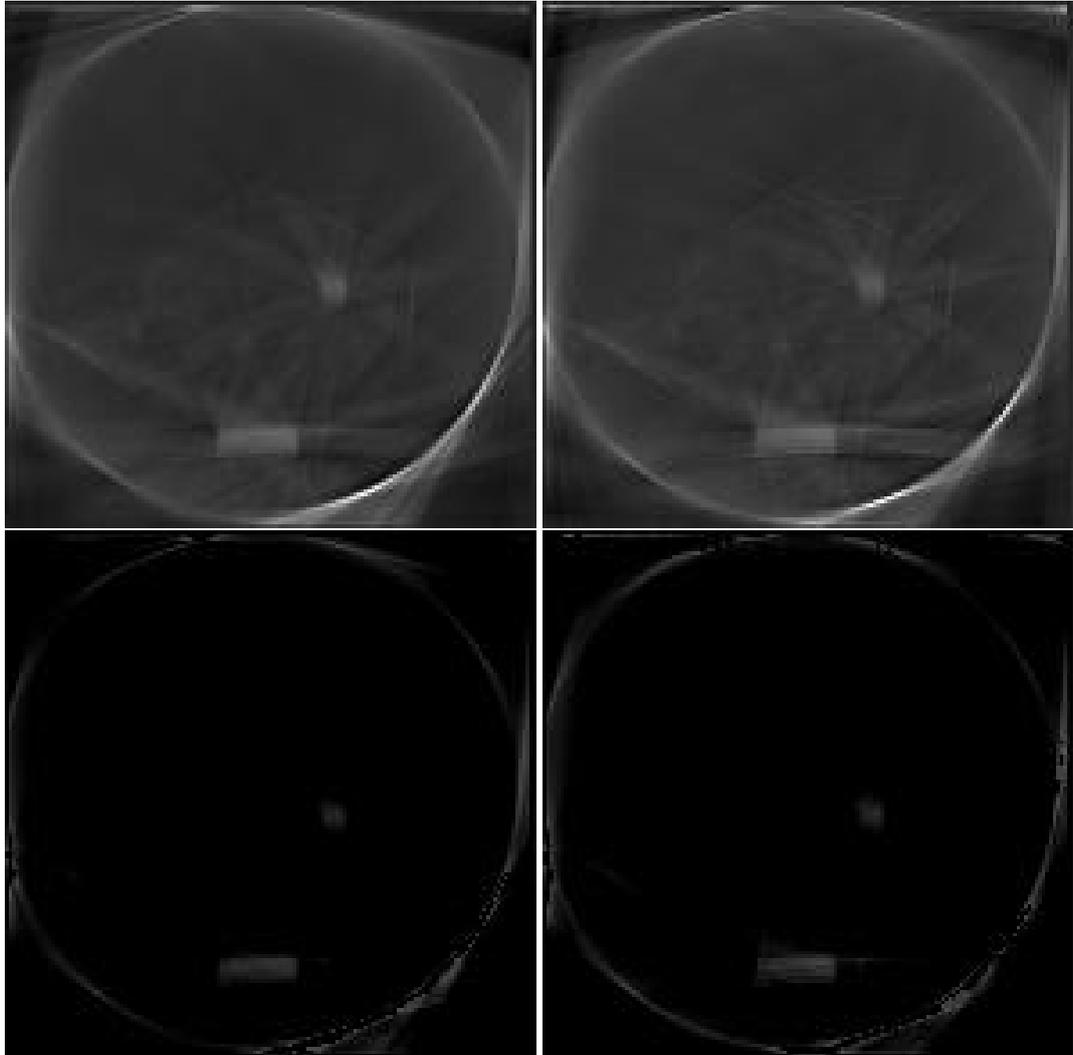


Figure 3.11: Axial slice view of reconstruction results of the moving liver. left to right: amplitude binning, phase binning. Top row: unprocessed slices; bottom row: thresholded slices.

## 4 Summary and Outlook

### 4.1 Summary

In this thesis we presented two approaches to respiratory monitoring and showed how to integrate a respiratory signal into the algebraic tomographic reconstruction algorithm, reconstructing a 4D, 3D over time, volumetric data set.

The first respiratory monitoring approach is based on monocular tracking of color fiducials in video. The approach provides a simple and inexpensive solution for respiration monitoring. The results presented in section 2.1 show that the respiratory signal is comparable to one acquired with a high-end 3D tracking system. This can be explained by the nature of the abdominal respiratory motion, which is almost linear and is periodical. Thus, a camera placed with its viewing direction orthogonal to the line of motion captures the same information a 3D tracking system can acquire. The raw signal generated from the tracking data is very noisy, but knowing that normal respiration has a maximal frequency of 25 cycles per minute allows low-pass filtering to eliminate all high frequencies, which are caused by noise. Finally, using our system in combination with an ultrasound probe we showed that the generated respiratory signal is correct. That is, it is highly correlated with internal organ motion.

The second respiratory monitoring approach is based on reconstructing the spatial location of a radio-opaque fiducial placed on the patient's abdomen. This approach is only applicable for respiration monitoring during CBCT data acquisition. At the basis of our approach is the approximation of the fiducial's spatial motion as a line. The 2D fiducial locations in the projection images are detected and used in conjunction with the known projection matrices to estimate the line of motion and corresponding 3D fiducial location for each image. These 3D locations are then used to generate the respiratory signal via PCA. To demonstrate the robustness of our 2D fiducial detection approach we applied it to X-ray projection images of an anthropomorphic phantom. To demonstrate the quality of the respiratory signal generated by our approach we acquired CBCT data of a fiducial placed on a motion platform performing a spatial motion which is based on the abdominal motion acquired from a volunteer. The phase of our respiratory signal corresponded to the phase of the signal generated using the known motion. The amplitude of both signals corresponded during most of the acquisition time, and did not correspond when the C-arm was near the AP imaging position, where the imaging plane is nearly orthogonal to the fiducial motion.

Finally, we used the generated respiratory signal to reconstruct an anthropomorphic phantom employing retrospective gating. The simultaneous algebraic reconstruction algorithm was implemented and tested on clinical data. We presented two

different gating schemes. First a hard gating scheme where we only use projection images which correspond to the bin, which is reconstructed. This led to severe reconstruction artifacts due to missing projection data, but could remove motion blurring. Reconstructing the end-inhale or end-exhale phase results in projection images acquired from only five different directions when the patient performed five respiratory cycles. We mitigate this by using all projection images for reconstruction of each respiratory phase. In this case we assign each image a different weight according to its position in the respiratory cycle. This is based on the idea that each image also contains static anatomy for example the spine. This data is valid over the whole respiratory cycle. We know that the images also contain dynamic anatomy which does not correspond to the phase we are reconstructing. The weighting factor is thus a trade-off between accepting false data and getting additional correct data. We have chosen the weighting factors empirically. In a simulation study we showed that the weighted reconstruction is slightly superior to the gated reconstruction but still suffers from reconstruction artifacts and has a low contrast.

## 4.2 Future work

To improve our current 4D CBCT reconstruction approach so that it is clinically useful we need to increase the speed of the reconstruction process and improve the quality of the reconstructed images.

Computationally efficient implementations of algebraic reconstruction algorithms have been previously described by several groups [36, 56, 51]. These approaches are based on the use of the GPU instead of the naive CPU based approach we have implemented. These GPU based approaches are expected to continue to provide enhanced performance over CPU based methods. This is primarily due to the fact that the rate of increase in GPU performance is greatly surpassing the increase in CPU performance [34]. Given this trend we will port our CPU based reconstruction method to the GPU.

To improve the quality of the reconstructed 4D data we will investigate two approaches.

The first approach relies on the availability of a preoperative CT acquired at breath-hold. We will non-rigidly register the CT to each of our lower quality volumes, resulting in a set of high quality CT based reconstructions. The main challenge of this approach is to define an appropriate similarity measure which is able to deal with the low contrast and reconstruction artifacts exhibited by our reconstructions. A promising similarity measure for 3D/3D low quality CT to diagnostic CT rigid registration was proposed by Tomazevic et al. [57]. We will investigate its applicability to non-rigid registration.

The second approach does not require preoperative data. This approach is based on performing multiple acquisition scans. The reconstruction quality is improved by providing images from more viewing angles for each respiratory phase. The main concerns with this approach are that it increases the acquisition time and the radiation dose to the patient.

## A Derivation of line-line distance

A direction  $n$  and a point  $a$  parameterize a line. A point on that line is given by  $p = a + \lambda n$ . The shortest segment connecting two lines has to be perpendicular to both lines and fulfill the equations

$$n_1^T(a_1 + \lambda_1 n_1 - a_2 - \lambda_2 n_2) = 0 \quad (\text{A.1})$$

$$n_2^T(a_1 + \lambda_1 n_1 - a_2 - \lambda_2 n_2) = 0 \quad (\text{A.2})$$

Expanding that equations and solving for  $\lambda$  gives

$$n_1^T a_1 + \lambda_1 n_1^T n_1 - n_1^T a_2 - \lambda_2 n_1^T n_2 = 0 \quad (\text{A.3})$$

$$n_2^T a_1 + \lambda_1 n_2^T n_1 - n_2^T a_2 - \lambda_2 n_2^T n_2 = 0 \quad (\text{A.4})$$

$$\begin{bmatrix} n_1^T n_1 & -n_1^T n_2 \\ n_2^T n_1 & -n_2^T n_2 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} n_1^T (a_2 - a_1) \\ n_2^T (a_2 - a_1) \end{bmatrix} \quad (\text{A.5})$$

The solution for  $\lambda_1$  and  $\lambda_2$  is

$$\lambda_1 = \frac{(n_1 \cdot n_2)(n_2 \cdot (a_2 - a_1)) - (n_2 \cdot n_2)(n_1 \cdot (a_2 - a_1))}{(n_1 \cdot n_1)(n_2 \cdot n_2) - (n_1 \cdot n_2)^2} \quad (\text{A.6})$$

$$\lambda_2 = \frac{(n_1 \cdot n_1)(n_2 \cdot (a_2 - a_1)) - (n_1 \cdot n_2)(n_1 \cdot (a_2 - a_1))}{(n_1 \cdot n_1)(n_2 \cdot n_2) - (n_1 \cdot n_2)^2} \quad (\text{A.7})$$

The denominator can be expressed as the dot product of the cross product of  $n_1$  and  $n_2$  using Lagrange's identity [35] which is

$$(a \times b) \cdot (c \times d) = (a \cdot c)(b \cdot d) - (b \cdot c)(a \cdot d) \quad (\text{A.8})$$

Applying this to our problem gives

$$(n_1 \cdot n_1)(n_2 \cdot n_2) - (n_1 \cdot n_2)^2 = (n_1 \times n_2) \cdot (n_1 \times n_2) \quad (\text{A.9})$$

The same identity holds for the nominator

$$(n_1 \cdot n_2)(n_2 \cdot (a_2 - a_1)) - (n_2 \cdot n_2)(n_1 \cdot (a_2 - a_1)) = ((a_2 - a_1) \times n_2) \cdot (n_1 \times n_2) \quad (\text{A.10})$$

$$(n_1 \cdot n_1)(n_2 \cdot (a_2 - a_1)) - (n_1 \cdot n_2)(n_1 \cdot (a_2 - a_1)) = ((a_2 - a_1) \times n_1) \cdot (n_1 \times n_2) \quad (\text{A.11})$$

Thus we get for the line parameters

$$\lambda_1 = \frac{((a_2 - a_1) \times n_2) \cdot (n_1 \times n_2)}{(n_1 \times n_2) \cdot (n_1 \times n_2)} \quad (\text{A.12})$$

$$\lambda_2 = \frac{((a_2 - a_1) \times n_1) \cdot (n_1 \times n_2)}{(n_1 \times n_2) \cdot (n_1 \times n_2)} \quad (\text{A.13})$$

The denominator becomes zero if the cross product  $n_1 \times n_2 = 0$ , meaning  $n_1$  and  $n_2$  are parallel. In this case the distance between both lines is computed by choosing one point on the first line and computing its shortest distance to the second line.  $\lambda_1$  and  $\lambda_2$  are plugged into the line equations

$$p_1 = a_1 + \lambda_1 n_1 = a_1 + \frac{((a_2 - a_1) \times n_2) \cdot (n_1 \times n_2)}{(n_1 \times n_2) \cdot (n_1 \times n_2)} n_1 \quad (\text{A.14})$$

$$p_2 = a_2 + \lambda_2 n_2 = a_2 + \frac{((a_2 - a_1) \times n_1) \cdot (n_1 \times n_2)}{(n_1 \times n_2) \cdot (n_1 \times n_2)} n_2 \quad (\text{A.15})$$

The shortest distance between both lines is the length of the line segment  $(p_1, p_2)$

$$d = |p_1 - p_2| \quad (\text{A.16})$$

## B Ray-sphere intersection

A ray is parameterized like a line with a start point (camera center)  $a$  and a direction  $n$ .

$$r = a + \lambda n \quad (\text{B.1})$$

A sphere is defined by its center  $(c_x, c_y, c_z)$  and radius  $r$ . For a point  $(x, y, z)$  on the sphere the following equation is true.

$$(x - c_x)^2 + (y - c_y)^2 + (z - c_z)^2 = r^2 \quad (\text{B.2})$$

Substituting the ray equation into the sphere equation yields a quadratic equation in the ray parameter  $\lambda$ .

$$(a_x + \lambda n_x - c_x)^2 + (a_y + \lambda n_y - c_y)^2 + (a_z + \lambda n_z - c_z)^2 = r^2 \quad (\text{B.3})$$

Solving for  $\lambda$  gives the two solutions

$$\lambda_{1,2} = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} \quad (\text{B.4})$$

$$A = n \cdot n \quad (\text{B.5})$$

$$B = 2(a - c) \cdot n \quad (\text{B.6})$$

$$C = (a - c) \cdot (a - c) - r^2 \quad (\text{B.7})$$

If the equation has no solution ( $B^2 - 4AC < 0$ ) the ray does not intersect the sphere. The equation has one solution ( $B^2 - 4AC = 0$ ) if the ray grazes the sphere. Two solutions ( $B^2 - 4AC > 0$ ) are obtained when the ray intersects the sphere. Substituting  $\lambda_{1,2}$  into the ray equation gives the entry and exit point of the ray.

## Bibliography

- [1] A.H. Andersen. Algebraic reconstruction in CT from limited views. *Medical Imaging, IEEE Transactions on*, 8(1):50–55, 1989.
- [2] Shai Avidan and Amnon Shashua. Trajectory triangulation: 3D reconstruction of moving points from a monocular image sequence. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(4):348–357, 2000.
- [3] Filip Banovac, Neil Glossop, David Lindisch, Daigo Tanaka, Elliot Levy, and Kevin Cleary. Liver tumor biopsy in a respiring phantom with the assistance of a novel electromagnetic navigation device. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2002: 5th International Conference, Tokyo, Japan, September 25-28, 2002, Proceedings, Part I*, pages 200–207, 2002.
- [4] Richard D. Beach, Hans Depold, Guido Boening, Philippe P. Bruyant, Bing Feng, Howard C. Gifford, Michael A. Gennert, Suman Nadella, and Michael A. King. An adaptive approach to decomposing patient-motion tracking data acquired during cardiac SPECT imaging. *IEEE Transactions on Nuclear Science*, 54(1):130–139, 2007.
- [5] A. Sam Beddar, Kristofer Kainz, Tina Marie Briere, Yoshikazu Tsunashima, Tinsu Pan, Karl Prado, Radhe Mohan, Michael Gillin, and Sunil Krishnan. Correlation between internal fiducial tumor motion and external marker motion for liver tumors imaged with 4D-CT. *International Journal of Radiation Oncology\*Biology\*Physics*, 67(2):630–638, February 2007.
- [6] Jane Mary Blackall. *Respiratory Motion in Image-Guided Interventions of the Liver*. PhD thesis, University of London, 2002.
- [7] Luc Boucher, Serge Rodrigue, Roger Lecomte, and François Bénard. Respiratory gating for 3-dimensional PET of the thorax: Feasibility and initial results. *Journal of Nuclear Medicine*, 45(2):214–219, 2004.
- [8] Gary R. Bradski. Computer vision face tracking for use in a perceptual user interface. *Intel Technology Journal*, (Q2):1–15, 1998.
- [9] Edward D. Brandner, Andrew Wu, Hungcheng Chen, Dwight Heron, Shalom Kalnicki, Krishna Komanduri, Kristina Gerszten, Steve Burton, Irfan Ahmed, and Zhenyu Shou. Abdominal organ motion measured using 4D CT. *International Journal of Radiation Oncology\*Biology\*Physics*, 65(2):554–560, June 2006.
- [10] J Canny. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698, 1986.

- [11] Stephanie K. Carlson, Joel P. Felmlee, Claire E. Bender, Richard L. Ehman, Kelly L. Classic, Houchun H. Hu, and Tanya L. Hoskin. Intermittent-mode CT fluoroscopy-guided biopsy of the lung or upper abdomen with breath-hold monitoring and feedback: System development and feasibility. *Radiology*, 229(3):906–912, December 2003.
- [12] Jenghwa Chang, Jussi Sillanpaa, Clifton C. Ling, Edward Seppi, Ellen Yorke, Gikas Mageras, and Howard Amols. Integrating respiratory gating into a megavoltage cone-beam CT system. *Medical Physics*, 33(7):2354–2361, 2006.
- [13] W. Chlewicki, C. Badea, and N. Pallikarakis. Cone based 3D reconstruction: A FDK-SART comparison for limited number of projections. In *MEDICON*, 2001.
- [14] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.
- [15] L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone-beam algorithm. *Journal of the Optical Society of America A*, 1(6):612–619, 1984.
- [16] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.
- [17] James D. Foley, Andries van Dam, Steven K. Feiner, and John F. Hughes. *Computer graphics (2nd ed. in C): principles and practice*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1996.
- [18] W. Förstner. Uncertainty and projective geometry. *Handbook of Computational Geometry for Pattern Recognition, Computer Vision, Neurocomputing and Robotics*, 2004.
- [19] K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *Information Theory, IEEE Transactions on*, 21(1):32–40, 1975.
- [20] Walter Gander, Gene H. Golub, and Rolf Strebler. Least-squares fitting of circles and ellipses. *BIT Numerical Mathematics*, 34(4):558–578, December 1994.
- [21] Michael A. Gennert, Joseph K. Ho, Andre C. Quina, James H. Wang, Philippe P. Bruyant, and Michael A. King. Feasibility of tracking patient respiration during cardiac SPECT imaging using stereo optical cameras. In *Nuclear Science Symposium*, volume 5, pages 3170–3172, 2003.
- [22] Theo Gevers and Arnold W. M. Smeulders. Color based object recognition. In *International Conference on Image Analysis and Processing*, pages 319–326, 1997.
- [23] Huaiqun Guan and Richard Gordon. Computed tomography using algebraic reconstruction techniques (arts) with different projection access schemes: a

- comparison study under practical situations. *Physics in Medicine and Biology*, 41(9):1727–1743, 1996.
- [24] D. J. Hawkes, D. Barratt, J.M. Blackall, C. Chan, P.J. Edwards, K. Rhode, G. P. Penney, J. McClelland, and D. L. G. Hill. Tissue deformation and shape models in image-guided interventions: a discussion paper. *Medical Image Analysis*, 9(2):163–175, 2005.
- [25] B. Horn. *Projective geometry considered harmful*, 1999.
- [26] Brian F. Hutton, H. Malcolm Hudson, and Freek J. Beekman. A clinical perspective of accelerated statistical reconstruction. *European Journal of Nuclear Medicine and Molecular Imaging*, 24(7):797–808, July 1997.
- [27] S. Kaczmarz. Angenaherte auflösung von systemen linearer gleichungen. *Bull. Acad. Pol. Sci. Lett. A*, 6-8A:355–357, 1937.
- [28] A.C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. IEEE Press, 1988.
- [29] Willi Kalender et al. Measurement of pulmonary parenchymal attenuation: use of spirometric gating with quantitative CT. *Radiology*, 175(1):265–268, 1990.
- [30] Sergey Kriminski, Matthias Mitschke, Stephen Sorensen, Nicole M. Wink, Phillip E. Chow, Steven Tenn, and Timothy D. Solberg. Respiratory correlated cone-beam computed tomography on an isocentric C-arm. *Phys. Med. Biol.*, 50(21-22):5263–5280, 2005.
- [31] Tianfang Li, Lei Xing, Peter Munro, Christopher McGuinness and Ming Chao, Yong Yang, Bill Loo, and Albert Koong. Four-dimensional cone-beam computed tomography using an on-board imager. *Med. Phys.*, 33(10):3825–3833, 2006.
- [32] Ralph Lin, Emmanuel Wilson, Jonathan Tang, Dan Stoianovici, and Kevin Cleary. A computer-controlled pump and realistic anthropomorphic respiratory phantom for validating image-guided systems. In Kevin R. Cleary and Michael I. Miga, editors, *Medical Imaging 2007: Visualization and Image-Guided Procedures*, volume 6509, page 65090E. SPIE, 2007.
- [33] Stephen Loose and Konrad W. Leszczynski. On few-view tomographic reconstruction with megavoltage photon beams. *Medical Physics*, 28(8):1679–1688, 2001.
- [34] David Luebke and Greg Humphreys. How GPUs work. *Computer*, 40(2):96–100, 2007.
- [35] Meyberg and Vachenaer. *Höhere Mathematik 1*. Springer, 2001.
- [36] Klaus Müller and Roni Yagel. Rapid 3D cone-beam reconstruction with the simultaneous algebraic reconstruction technique (sart) using 2d texture mapping hardware. *IEEE Trans. Med. Imaging*, 19(12):1227–1237, 2000.

- [37] Klaus Müller, Roni Yagel, and John J. Wheller. Fast implementation of algebraic methods for 3D reconstruction from cone-beam data. *IEEE Trans. Med. Imaging*, 18(6):538–548, 1999.
- [38] S. A. Nehmeh et al. Quantitation of respiratory motion during 4D-PET/CT acquisition. *Med. Phys.*, 31(6):1333–1338, 2004.
- [39] Cihat Ozhasoglu and Martin J. Murphy. Issues in respiratory motion compensation during external-beam radiotherapy. *International Journal of Radiation Oncology\*Biophysics*, 52(5):1389–1399, April 2002.
- [40] Tinsu Pan, Ting-Yim Lee, Eike Rietzel, and George T. Y. Chen. 4D-CT imaging of a volume influenced by respiratory motion on multi-slice CT. *Med. Phys.*, 31(2):333–340, 2004.
- [41] Terry M Peters. Image-guidance for surgical procedures. *Physics in Medicine and Biology*, 51(14):R505–R540, 2006.
- [42] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, 2nd edition, 1992.
- [43] M. Prümmer, L. Wigstroem, J. Hornegger, J. Boese, G. Lauritsch, N. Strobel, and R. Fahrig. Cardiac C-arm CT: Efficient Motion Correction for 4D-FBP . In G. Smith, editor, *Nuclear Science Symposium, Medical Imaging* , pages 1–20, 2006.
- [44] Krishnakumar Ramamurthi and Jerry L. Prince. Tomographic reconstruction for truncated cone beam data using prior CT information. In *MICCAI (2)*, pages 134–141, 2003.
- [45] R. Rangayyan, A. Dhawan, and R. Gordon. Algorithms for limited-view computed tomography: an annotated bibliography and a challenge. *Appl. Opt.*, 24, 1985.
- [46] Simon Rit, David Sarrut, and Chantal Ginestet. Respiratory signal extraction for 4D CT imaging of the thorax from cone-beam CT projections. In *Medical Image Computing and Computer-Assisted Intervention*, pages 556–563, 2005.
- [47] Simon Rit, David Sarrut, and Serge Miguët. Gated cone-beam CT imaging of the thorax: a reconstruction study. In *SPIE Medical Imaging: Physics of Medical Imaging*, pages 651022–1–651022–10, 2007.
- [48] Torsten Rohlfing, Jr. Calvin R. Maurer, Walter G. O’Dell, and Jianhui Zhong. Modeling liver motion and deformation during the respiratory cycle using intensity-based nonrigid registration of gated mr images. *Medical Physics*, 31(3):427–432, 2004.
- [49] Ofri Sadowsky, Krishnakumar Ramamurthi, Lotta Maria Ellingsen, Gouthami Chintalapani, Jerry L. Prince, and Russell H. Taylor. Atlas-assisted tomography:

- registration of a deformable atlas to compensate for limited-angle cone-beam trajectory. In *IEEE International Symposium on Biomedical Imaging (ISBI): From Nano to Macro*, pages 1244–1247, 2006.
- [50] D. Schäfer, J. Borgert, V. Rasche, and M. Grass. Motion-compensated and gated cone beam filtered back-projection for 3-d rotational x-ray angiography. *IEEE Trans Med Imaging*, 25(7):898–906, July 2006.
- [51] T. Schiwietz, S. Bose, J. Maltz, and R. Westermann. A fast and high-quality cone beam reconstruction pipeline using the GPU. In *Proceedings of SPIE Medical Imaging 2006*, San Diego, CA, Feb. 2007. SPIE.
- [52] Achim Schweikard, Hiroya Shiomi, and John Adler. Respiration tracking in radiosurgery. *Medical Physics*, 31(10):2738–2741, 2004.
- [53] Yvette Seppenwoolde, Hiroki Shirato, Kei Kitamura, Shinichi Shimizu, Marcel van Herk, Joos V. Lebesque, and Kazuo Miyasaka. Precise and real-time measurement of 3D tumor motion in lung due to breathing and heartbeat, measured during radiotherapy. *International Journal of Radiation Oncology\*Biophysics*, 53(4):822–834, July 2002.
- [54] S. Siltanen, V. Kolehmainen, S. Jarvenpaa, J.P. Kaipio, P. Koistinen, M. Lassas, J. Pirttila, and E. Somersalo. Statistical inversion for medical x-ray tomography with few radiographs: I. general theory. *Physics in Medicine and Biology*, 48:1437–1463(27), 2003.
- [55] Jan-Jakob Sonke, Lambert Zijp, Peter Remeijer, and Marcel van Herk. Respiratory correlated cone beam CT. *Med. Phys.*, 32(4):1176–1185, 2005.
- [56] Ralf Tita, Wolfgang Holler, Sven Huth, and Tim C. Lueth. Near real time tomographic 3D reconstruction with the use of the pc graphics accelerator. In *Multisensor Fusion and Integration for Intelligent Systems, 2006 IEEE International Conference on*, pages 279–284, 2006.
- [57] D. Tomazevic, B. Likar, and F. Pernus. 3-d/2-d registration by integrating 2-d information in 3-d. *Medical Imaging, IEEE Transactions on*, 25(1):17–27, 2006.
- [58] S S Vedam, P J Keall, V R Kini, H Mostafavi, H P Shukla, and R Mohan. Acquiring a four-dimensional computed tomography dataset using an external respiratory signal. *Physics in Medicine and Biology*, 48(1):45–62, 2003.
- [59] Karl Wiesent, Karl Barth, Nassir Navab, Peter Durlak, Thomas Brunner, Oliver Schütz, and Wolfgang Seissler. Enhanced 3D-reconstruction algorithms for c-arm based interventional procedures. *IEEE Trans. Med. Imaging*, 19(5):391–403, 2000.
- [60] K. H. Wong, J. W. VanMeter, S. T. Fricke, C. R. Maurer, and K. Cleary. Mri for modeling of liver and skin respiratory motion. *International Congress Series*, 1268:747–752, June 2004.

- 
- [61] Andrew Woo. *Fast ray-box intersection*, pages 395–396. Academic Press Professional, Inc., San Diego, CA, USA, 1990.
  - [62] R. Zeng, J. A. Fessler, and J. M. Balter. Estimating 3-d respiratory motion from orbiting views by tomographic image registration. *IEEE Trans. Med. Imag.*, 26(2):153–163, February 2007.
  - [63] Rongping Zeng, Jeffrey A. Fessler, and James Balter. Respiratory motion estimation from slowly rotating x-ray projections. In *ISBI*, pages 480–483, 2004.
  - [64] L. Zijp, J.J. Sonke, and M.V. Herk. Extraction of the respiratory signal from sequential thorax cone-beam x-ray images. *International Conference on the Use of Computers in Radiation Therapy*, 2004.