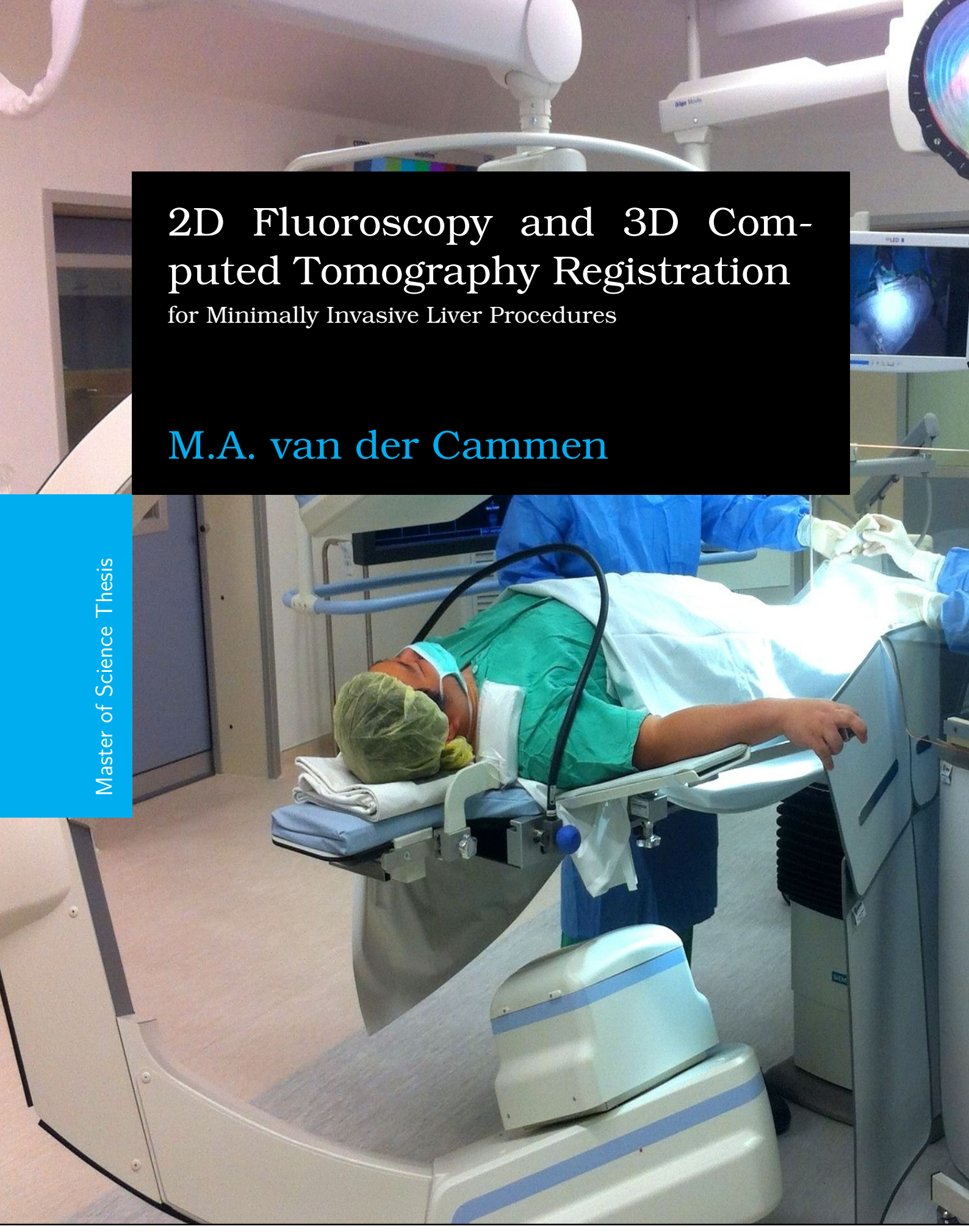


2D Fluoroscopy and 3D Computed Tomography Registration

for Minimally Invasive Liver Procedures

M.A. van der Cammen

Master of Science Thesis

TU Delft Delft University of Technology

Delft University of Technology

2D Fluoroscopy and 3D Computed Tomography Registration

for Minimally Invasive Liver Procedures

Master of Science Thesis

For the degree of Master of Science in Electrical Engineering at Delft University of Technology

M.A. van der Cammen

December 13, 2016

Faculty of Electrical Engineering · Delft University of Technology

The work in this thesis was done at the Biomedical Imaging Group Rotterdam

Preface

The research presented in this thesis was done at the Biomedical Imaging Group Rotterdam(BIGR) at the Erasmus Medical Center. During the course of my thesis I have had the opportunity to work closely with the Image Guidance in Interventions(IGI) team and as such, received valuable support and feedback on my work. In particular I would like to thank Pierre Ambrosini (PhD at IGI) and Theo van Walsum (head of IGI). Pierre has brought me up to speed in the field of medical imaging, and I have frequently benefited from his knowledge and insights during my research. Theo has provided me with valuable guidance, especially during our weekly meetings, making sure I was progressing in the right direction. It is due to the IGI team and their members that I was able to conduct my research in a proper manner, for which I remain grateful.

Table of Contents

Chapter 1

Introduction

Medical imaging first began in 1895 with the discovery of X-rays by Wilhelm Röntgen. Since then, many more modalities have been used in order to create medical images. These images have been beneficial ever since, in research as well as support to physicians during operations and diagnostics. The following thesis takes place within the realm of image guidance in interventions, the area in which we use medical images and information deduced from them to support medical interventions. In our specific case, the intervention is called a TACE intervention performed to treat a tumor inside the liver.

Consider a physician who decides to order a contrast-enhanced (or multi-phase) computed tomography(CT) scan of a patient for further diagnosis. Upon investigation of the CT, the physician discovers a tumor inside the patients liver. One of the treatment options, depending on amongst others the patient state, the number of tumors and their size, is a minimally invasive intervention such as Transcatheter arterial chemoembolization(TACE). The purpose of TACE is bringing therapeutics to the tumor while simultaneously restricting the blood supply to the tumor. During the intervention, a catheter is inserted through the patient femoral artery and manoeuvred up the aorta. The catheter is then guided through the hepatic artery before finally reaching a vessel directly supplying the tumor. Small embolic particles coated with chemotherapeutic agents are then injected into the artery, thus diminishing the blood supply to the tumor. During the intervention the physician makes use of intra-interventional X-ray(XA) images to guide the catheter to its target.

During the intervention the physician has access to the 2D XA for guidance. Although the pre-interventional CT angiography (also called CTA) is also available it is not straightforward to effectively utilize this image. Importantly, it is not clear how the CT image is spatially aligned with the patient. Herein lies room for improvement, as the integration of the CT scan with the XA images will form an improved roadmap, which is hypothesized to facilitate the navigation of the instrument to the tumor.
Due to the injected contrast agent, relevant arteries including the common hepatic artery as well as the tumor itself are clearly visible on the CT. The aorta and other numerous

anatomical structures are also visible on the CT. Integration of the CT with the XA thus produces potential benefits. First, navigation will become easier with relevant arteries and structures along the catheter's route now more clearly visible (Fig. 1-1). Second, with the reduction in intervention time due to the augmented roadmap, the exposure of the patient and physician to the radiation emitted by the X-ray source diminishes. The intervention would thus benefit greatly from an accurate incorporation of the CT scan into the XA frames. The work in this thesis is focussed on this issue: registration of the 3D CT with the 2D XA. Accurate projection of the 3D CT onto the 2D XA, leading to an optimal alignment with the 2D XA, enables the incorporation of information contained within the CT.

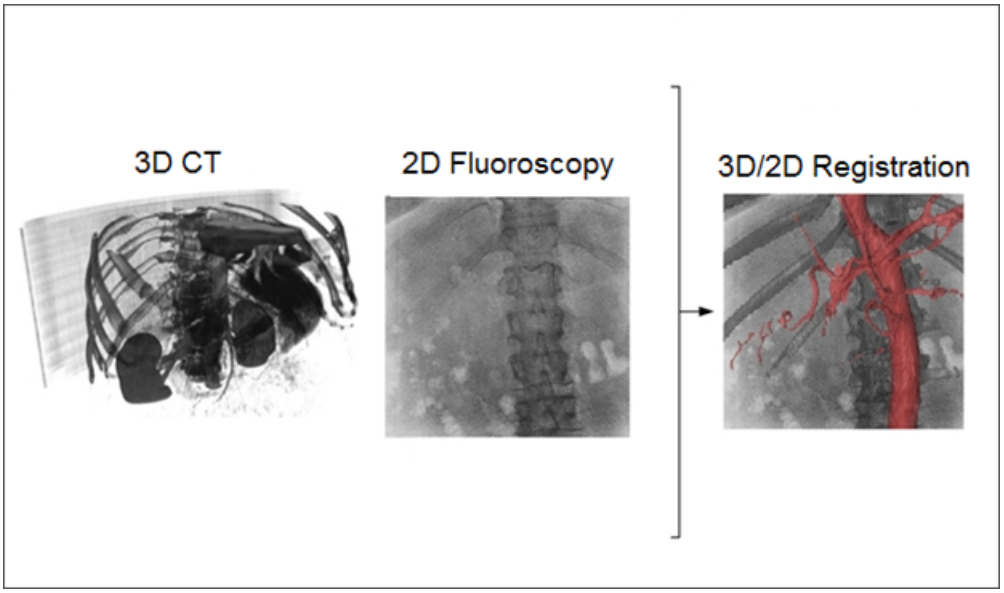


Figure 1-1: The thesis objective is registration of the 3D CT with the 2D Fluoroscopy (XA). This registration will enable an enhance roadmap, useful during minimally invasive interventions such as the TACE intervention.

The remainder of this thesis is organized as follows. Chapter 2 discusses related work on the topic of our thesis. The complete pipeline of our 3D/2D registration method is discussed in detail in Chapter 3. This chapter is divided into three sections. First, in Sec. 3-1 the pre-processing stage is discussed. Then, Sec. 3-2 deals with the 3D pose initialization stage. Finally, Sec. 3-3 details the 3D pose optimization stage. After the method is explained, experiments are described in Chapter 4, their results are depicted in Chapter 5 and discussed in Chapter 6. Recommendations with respect to future work are detailed in Chapter 7. Chapter 8 contains the conclusion of the thesis. Additionally Appendix A details the various coordinate systems involved in this thesis, Appendix B addresses the projection of 3D images generating the Digitally Reconstructed Radiographs(DRRs) used in this method and Appendix C depicts additional experiment results.

Chapter 2

Related Work

As stated in Chapter 1 the objective of this thesis is the registration of a 3D CT image with a 2D XA image (Fig. 2-1). A lot of research has been done on 3D/2D registration [1] [20] and its feasibility [2] [11]. Most of the 3D/2D image registration is achieved by optimising a similarity measure between the 2D image and the (simulated) projection of the 3D image [17]. In this case the pose of the 3D image is the optimization parameter.

3D/2D registration methods are either extrinsic-, intrinsic- or calibration-based[17].
Extrinsic registration methods rely on the presence of artificial objects or physical markers in order to aid the registration process. As we do not have nor place physical markers the only manner in which we could use these methods is by using segmentation as part of the registration. The result of which could then be used as seed-markers for an extrinsic registration method. However, we prefer a method that does not require extensive segmentation to be applied. Also, any error in segmentation or detection will influence the overall registration performance. For these reasons we decide it best not to start with extrinsic registration through segmentation.
Calibration-based registration methods use carefully pre-calibrated imaging devices [12]. The position of the operation table w.r.t. the imaging device is established during image acquisition. This system requires tracking of multiple moving components of the imaging system such as the X-ray source, C-arm and much more. As we do not have all required tracking capabilities these methods cannot be applied to our situation.
Intrinsic(image based) registration methods rely solely on information contained within the images, such as intensity distributions or image gradients. As such it is the class most applicable to our situation.

Each registration method consists of optimizing a similarity measure and requires an initialization. Also a certain model of deformation is assumed.

Focusing on the similarity measure we can divide intrinsic methods into three sub-classes being: feature-, intensity- and gradient-based registration.
Feature-based registration methods rely on segmentation. The object of interest is segmented in both images prior to the extraction of features. Subsequently extracted features are used to register the images. These features can be geometrical entities like isolated points, point sets, a curve or a surface. Extraction of these geometrical features by image segmentation greatly reduces the amount of data, which in turn makes such registration fast. However the speed and accuracy of the whole process is also influenced by the segmentation. One can reduce the negative influence by implementing approaches for handling false geometrical correspondences and outliers. Although this method is applicable to our case it is preferable not to start with it for the same reasons we debunked extrinsic registration, the dependency on segmentation.
Intensity-based registration methods use intensity values from the images to find the correct registration [4] [7]. Generally the registration in this method is slower in comparison with feature based registration however it requires little or no segmentation.
Gradient-based registration methods revolve around image gradients. Using these gradients for registration allows for robust registration even between images with different modalities. i.e. between a CT and X-ray fluoroscopy. Tomazevic et. al have proposed measuring the similarity between preoperatively computed CT or MR gradients, perpendicular to the bone surfaces, with gradients of intraoperational X-ray images [6]. Livyatan et. al proposed projecting 3D image gradients on the detector plane and measuring the similarity with the X-ray image gradients [10].

Both intensity- and gradient-based registration methods are applicable to our case, our registration method will hence utilize the benefit of said methods.

Arguably one of the most important aspects of 3D/2D registration is the initialization of the 3D pose as most proposed registration methods have a small capture range. Most methods require manual initialization [3] [8]. However automatic pose initialization methods have been proposed, amongst others in [18], where 2-D vessel bifurcation detection and spine detection is used to provide an initial pose estimate. During catheterization inside the hepatic artery, the spine is visible on the 2D fluoroscopy. As the spine is also visible on the 3D CTA, a 3D pose initialization could be achieved by aligning the 2D spine on the fluoroscopy to the projection of the 3D spine extracted from the CTA.

Furthermore, registration methods differ in their model of deformation, assuming either rigid or non-rigid deformation. Due to the thickness of the aorta and the common hepatic artery, the respiratory motion is not affecting their position by much. Thus, a rigid registration should be sufficient to achieve a roadmap of the vessels.

Finally, creation of the aforementioned projection of the 3D image can be achieved in multiple ways, exhibiting a trade off between accuracy and computational complexity [5]. Assuming the projection process is not overly computational expensive, a combination of ray-casting [14] and splatting [21] shall be applied.

In summary, we will assume a rigid deformation model and use ray-casting and splatting to project the 3D CT image. We will utilize either intensity- or gradient-based similarity mea-

sures during optimization of the 3D pose. Furthermore, to overcome the common drawback of a registration method's small capture range, we propose a new automatic initialization approach. This approach will detect the spine, and use it to align both images with respect to pose, orientation and scale. Finally, manually placed annotations will be used to evaluate the performance of the proposed registration method.

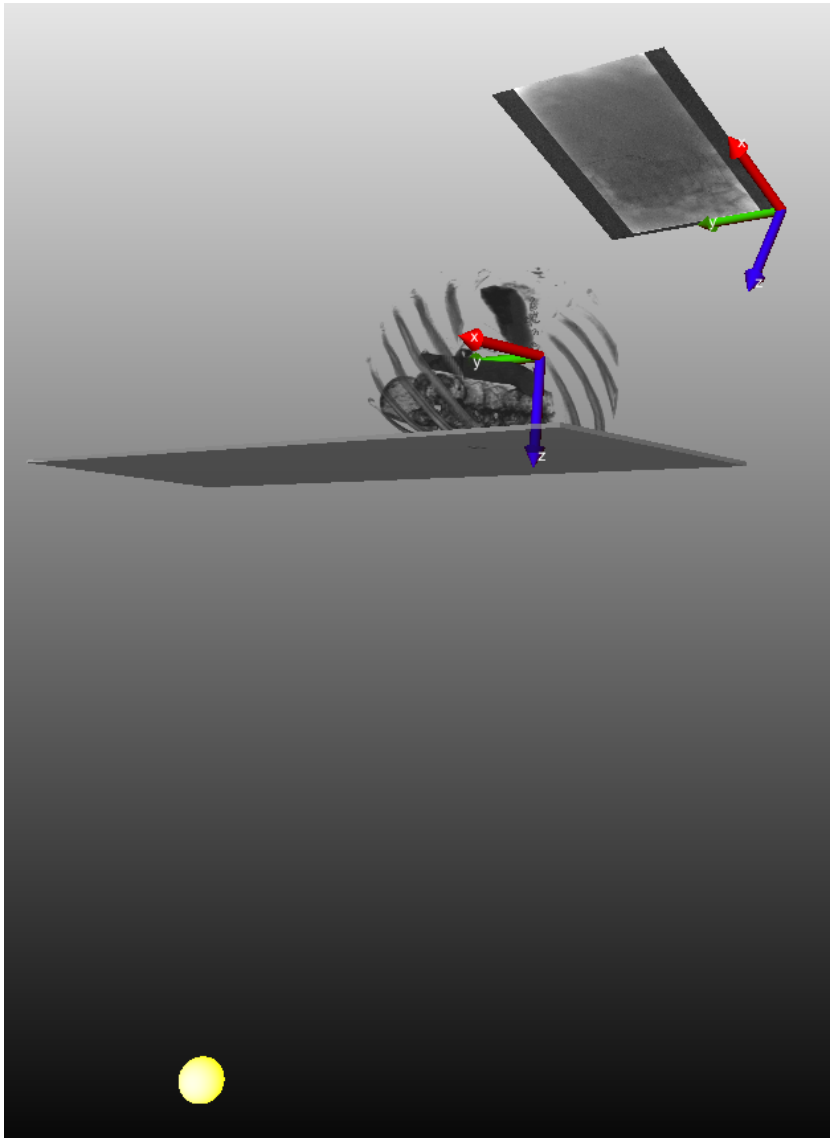


Figure 2-1: The thesis objective translates into correct determination of the pose of the 3D CT image (positioned on the table) in order for its projection to align with the 2D XA image (depicted on the detector plane in the upper right corner). Additionally, the axis of the world coordinate system and detector plane coordinate system (App. A) are depicted. Here, red denotes the x-axis, green the y-axis and blue the z-axis.

Chapter 3

Registration Method

The proposed 3D CT/2D XA registration method consists of the following three stages:

Stage 1. Pre-processing of the CT and XA images

Stage 2. Initialization of the 3D pose

Stage 3. Optimization of the 3D pose

We will deal with the pose initialization by creating an automatic initialization scheme heavily reliant on spine detection. During the method we aim to register based on only the spine and that part of the ribs closest to the spine. The pose optimization will utilize either an intensity or gradient-based similarity measure, the experiments will prove which works best. As the CT is pre-interventional, the pre-processing of the CT could be done in advance of the actual intervention. However we also propose an automatic approach, of an assumed sufficient performance with the added benefit of ease of use.

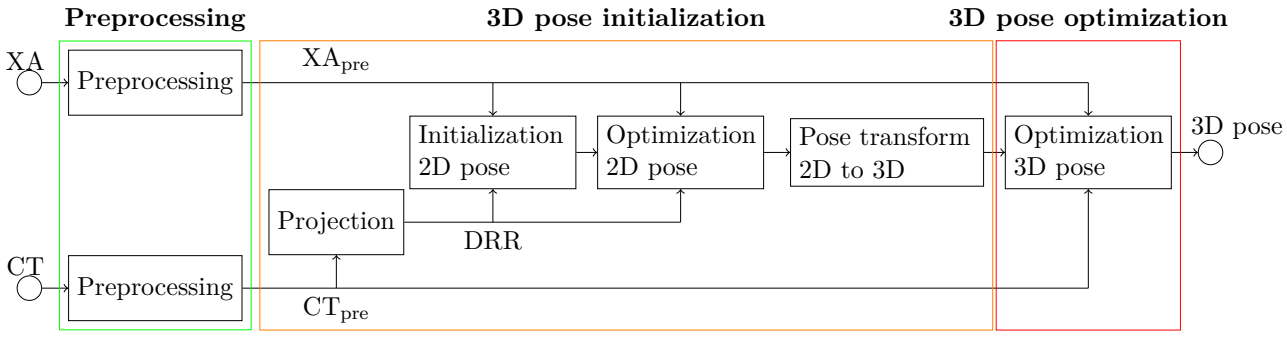


Figure 3-1: Blockscheme of the proposed 3D/2D registration method

The proposed registration method is depicted in detail in Fig. 3-1. The result of the Preprocessing stage (detailed in Sec. 3-1) are the images XA_{pre} and CT_{pre}. During the preprocessing, among other things, noise is reduced in both images and the aorta is removed

from the CT image.

The 3D Pose Initialization stage as detailed in Sec. 3-2 involves our tailor made initialization scheme. Here, the 3D $\mathrm{CT}_{\mathrm{pre}}$ is projected onto the detector plane creating the 2D Digitally Reconstructed Radiograph DRR (see App. B on projection of the 3D CT). Subsequently, the translation, rotation and scale necessary to improve the DRR's alignment with $\mathrm{XA}_{\mathrm{pre}}$ are automatically estimated, yielding us an initial 2D pose of the DRR on the detector plane. This pose is then further optimized as to improve the alignment, yielding us with an optimal 2D pose. This resulting pose is then transformed back to a 3D pose which is used as an initial seed point in the subsequent 3D Pose Optimization Stage.

The 3D Pose Optimization Stage as detailed in Sec. 3-3 optimizes the initial 3D pose provided by the previous stage. The output of this stage, the optimal 3D pose of the $\mathrm{CT}_{\mathrm{pre}}$, forms the output of the entire system.

All individual steps of the three stages will be addressed in their corresponding sections, with the projection step as an exception. This will be discussed in Appendix B as this is a very common process in any 3D/2D image registration method.

Note that the XA and CT images are originally defined in their own coordinate systems(see App. A for elaboration on the coordinate systems and transformation matrices). Transformation of the XA and CT from their original coordinate systems to the world coordinate system and the X-ray 2D device coordinate system respectively, is done as follows:

$$\mathrm{CT}_{\mathrm{world}} = Tl_{\mathrm{CTdevice}\rightarrow\mathrm{world}} * R_{\mathrm{CTmm}\rightarrow\mathrm{CTdevice}} * S_{\mathrm{CT}\rightarrow\mathrm{CTmm}} * \mathrm{CT} \tag{3-1}$$

$$\mathrm{XA}_{\mathrm{X-ray2D}} = R_{\mathrm{XA}mm\rightarrow\mathrm{X-ray2D}} * S_{\mathrm{XA}\rightarrow\mathrm{XA}mm} * \mathrm{XA} \tag{3-2}$$

Our objective is to find the correct pose of the $\mathrm{CT}_{\mathrm{world}}$ image such that there is an optimal alignment between the projection onto the detector plane of $T_{\mathrm{CT}} * \mathrm{CT}_{\mathrm{world}}$ and the image $\mathrm{XA}_{\mathrm{X-ray2D}}$ (both defined in the X-ray 2D coordinate system). T_{CT} is the transformation matrix based on the three translation- and three rotation parameters. Our objective thus consists of finding T_{CT} in

$$\mathrm{XA}_{\mathrm{X-ray2D}} == T_{\mathrm{world}\rightarrow\mathrm{X-ray2D}} * T_{\mathrm{CT}} * \mathrm{CT}_{\mathrm{world}}, \tag{3-3}$$

such that the two parts of this equation exhibit the best possible match.

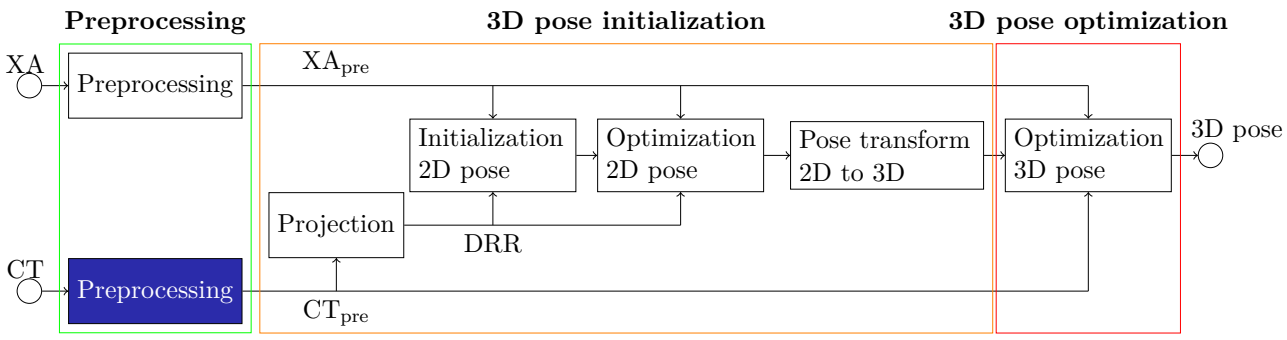
Mathematical conventions Before we start with the detailed description of the registration method, we define a number of mathematical conventions adhered to throughout this thesis. The CT image at the input of the method is defined as CT of size $R_{\mathrm{CT}}, C_{\mathrm{CT}}, S_{\mathrm{CT}}$ denoting respectively the number of rows, columns and slices. $\mathrm{CT}(r, c, s)$ depicts the intensity of the voxel of the CT at row r, column c and slice s.
The XA image at the input of the method is defined as XA of size $R_{\mathrm{XA}}, C_{\mathrm{XA}}, F_{\mathrm{XA}}$ denoting respectively the number of rows, columns and frames. $\mathrm{XA}(r, c, f)$ depicts the intensity of the pixel of the XA at row r, column c and frame f.

During the registration, a pre-processed version of the CT known as $\mathrm{CT}_{\mathrm{pre}}$ is projected yielding us with a DRR image. This 2D image has a size of $R_{\mathrm{DRR}}, C_{\mathrm{DRR}}$ denoting respectively the number of rows and columns.

For information about the orientation of the images as well as additional information about the coordinate systems used, please refer to Appendix A. For information about the projection of $\mathrm{CT}_{\mathrm{pre}}$ resulting in DRR, please refer to Appendix B.

3-1 Pre-processing

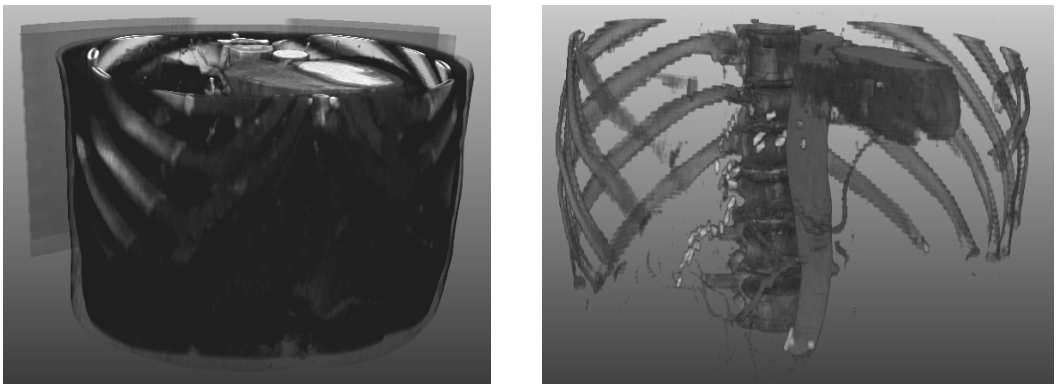
3-1-1 CT pre-processing



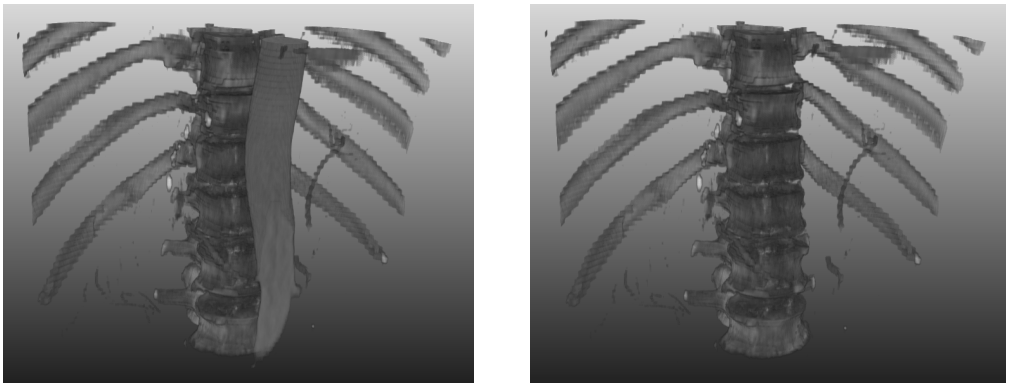
In this section we discuss the pre-processing of the CT image. Fig. 3-2a shows a CT image and Fig. 3-10a shows the first frame of an XA sequence (from the same patient). These images should be spatially aligned. To this end, we need to select those structures that can be used to align these images.

We do not wish to influence our registration with anatomical structures which are not clearly present in both modalities. Also, as respiratory movement and cardiac phase influence some inner body relative positions, we prefer structures close to our regions of interest: liver arteries and the aorta. From the anatomical structures present in both modalities (see Fig. 3-2a and 3-10a) the spine and the ribs (specifically those parts closest to the spine) meet our requirements. Thus, the CT pre-processing will consist of removing all of the anatomical parts except for the spine and the parts of the ribs close to the spine.

The CT pre-processing is divided into three steps. First, we apply a threshold to retain only dense structures amongst which are the spine and ribs. Then, we perform masking as to remove those rib parts far away from the spine. Finally, we remove the aorta from the CT which survived the subsequent steps due to its location and its high density as a result of injected contrast agent. Fig. 3-2 depicts the result of all steps.



(a) original CTA **(b)** after threshold



(c) after masking **(d)** after aorta removal

Figure 3-2: CT pre-processing; consisting of noise reduction by means of a threshold, masking and removal of the aorta

Noise reduction

In order to keep only the spine and ribs, we start by applying a threshold to the 3D image to segment the bone. To achieve this we make use of the density of bone in contrast with other elements. Fig. 3-4 shows the Hounsfield range values(depicting radio density) of the main anatomical parts. The threshold value H_{HU} is set to be the equivalent of the minimum compact bone Hounsfield value of 300 HU. For memory reasons medical images adhering to the DICOM standard (a standard for storage of medical images) have voxel values which are not stored in Hounsfield units (in the case of a.o. CT and PET images). Using the DICOM tags: rescale slope $\mathrm{re}_{\mathrm{sl}}$ and rescale intercept $\mathrm{re}_{\mathrm{icp}}$, a voxel's value can be transformed from Hounsfield units to the proper unit representation (Eq. 3-4). Fig. 3-2b demonstrates the effects of applying a threshold on the CT of 300 HU. Eq. 3-5 depicts the application of a threshold with an intensity value of H.

$$H = \mathrm{re}_{\mathrm{sl}} * H_{\mathrm{HU}} + \mathrm{re}_{\mathrm{icp}} \tag{3-4}$$

$$\mathrm{CT}_t(r,c,s) = \begin{cases} \mathrm{CT}(r,c,s), & \text{if } \mathrm{CT}(r,c,s) > H \\ 0, & \text{otherwise} \end{cases} \tag{3-5}$$

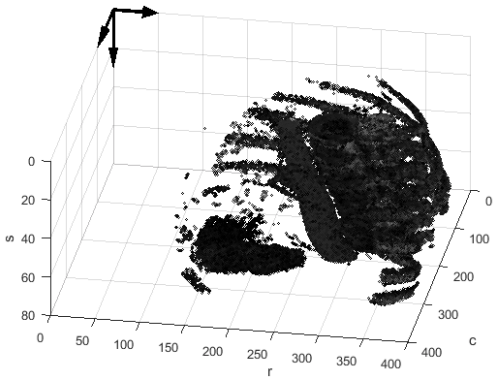


Figure 3-3: CT coordinate system

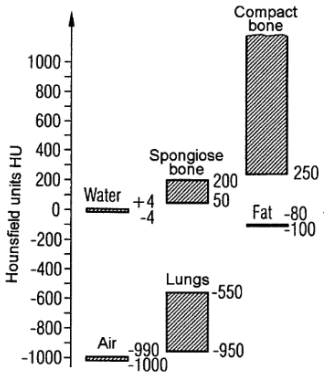


Figure 3-4: Hounsfield unit scale [9]

Masking

Next, we will mask the thresholded image CT_t such that only the spine and rib parts close to the spine remain. The approach we use to determine the mask is a spine localization approach, tailor-made for the CT modality consisting of two steps.
The mask will be a rectangular box spanning over all slices. In this approach, we assume that the spine is present at all slices and in the same region. We are thus searching for boundaries in terms of rows and columns. In order to obtain these, we need to first detect the location of the spine. Based on our assumption, we can take a patch of the image CT_t (containing all rows and columns but only a certain number of slices). The sum of this patch should yield high values in the spine region. To make our approach more robust, we take multiple patches throughout the image CT_t in order to determine the spine position. In total, ten patches are used at the following locations: $s_{patch} = [0 : 0.1 : 0.9] * S_{\mathrm{CT}}$. The patches have a thickness of 3 cm.

For each patch, we sum over the slices, creating a 2D image $\mathrm{CTpatch}_{\Sigma\mathrm{slice}}$ by using the equation

$$\mathrm{CTpatch}_{\Sigma\mathrm{slice}}(r,c) = \sum_{s=s_{\mathrm{patch}}}^{s_{\mathrm{patch}}+30/\Delta z_{\mathrm{CT}}} \mathrm{CT}_t(r,c,s), \tag{3-6}$$

in which Δz_{CT} is the slice spacing in mm. Subsequently we determine the position of the maximum value $p = [r_p \quad c_p]$ in the summed patch as,

$$p_{\mathrm{patch}} = \mathrm{argmax}\left(\mathrm{CTpatch}_{\Sigma\mathrm{slice}}\right). \tag{3-7}$$

Averaging over the positions of all detected maxima in the patches yields us our step-1 spine position estimate p. We then use this estimate to put a rectangular mask on the CT_t as,

$$\mathrm{CT}_{\mathrm{mask1}}(r,c,s) = \begin{cases} \mathrm{CT}(r,c,s) & \text{if } r_p - r_m < r < r_p + r_m \text{ and } c_p - c_m < c < c_p + c_m \\ 0 & \text{otherwise} \end{cases}, \tag{3-8}$$

creating $\mathrm{CT}_{\mathrm{mask1}}$. The mask is centered at p with a width of $2r_m$ and a height of $2c_m$. The result is shown in Fig. 3-5. Here, p is depicted as a red dot whereas the blue rectangle depicts

the masking boundaries $[r_p - r_m, r_p + r_m]$ and $[c_p - c_m, c_p + c_m]$.

We use the new masked image $\mathrm{CT}_{\mathrm{mask1}}$ as an input into our second and final step. During this step we will improve the column estimate c_p of the spine's position.
First, we sum over the rows of $\mathrm{CT}_{\mathrm{mask1}}$, creating $\mathrm{CT}_{\Sigma row}$ by using the following equation,

$$\mathrm{CT}_{\Sigma\mathrm{row}}(c, s) = \sum_{r=1}^{R_{\mathrm{CT}}} \mathrm{CT}_{\mathrm{mask1}}(r, c, s). \tag{3-9}$$

On this 2D image we then apply a spine edge detection method created for this thesis (the method will be extensively discussed in Sec. 3-2-1). The method results in a left- and right edge value which we denote as c_l and c_r respectively. Using the edge estimates, we can update the column-element of the spine position estimate as $c_p = (c_r + c_l)/2$. This step can be viewed in Fig. 3-6. Here, c_l and c_r are depicted as blue solid lines whereas the new column boundaries $[c_l - c_m, c_r + c_m]$ are depicted as blue dashed lines.
Now that we have our final spine position estimate, we can apply the masking on CT_t as,

$$\mathrm{CT}_m = \begin{cases} \mathrm{CT}_t & \text{if } r_p - r_m < r < r_p + r_m \text{ and } c_l - c_m < c < c_r + c_m \\ 0 & \text{otherwise} \end{cases}, \tag{3-10}$$

yielding the result of this stage, CT_m. The result of the masking is shown in Fig. 3-2c with $c_m = 100\ mm$ and $r_m = 60\ mm$. The value of 100 mm is selected as we wish to keep only the static part of the ribs which do not undergo much influence from respiratory movement and/or cardiac phase. The 60 mm value is based on the spine's curvature, and exhibits parts of the spine to be excluded from the mask.

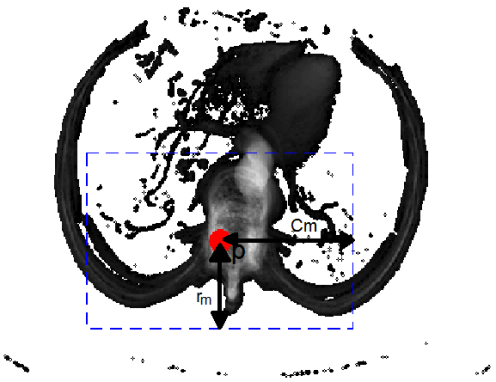


Figure 3-5: $\mathrm{CT}_{\Sigma\mathrm{slice}}$, CT masking step 1

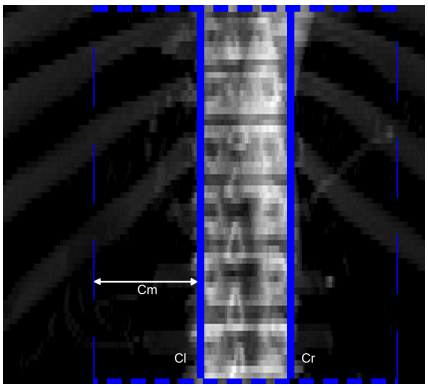


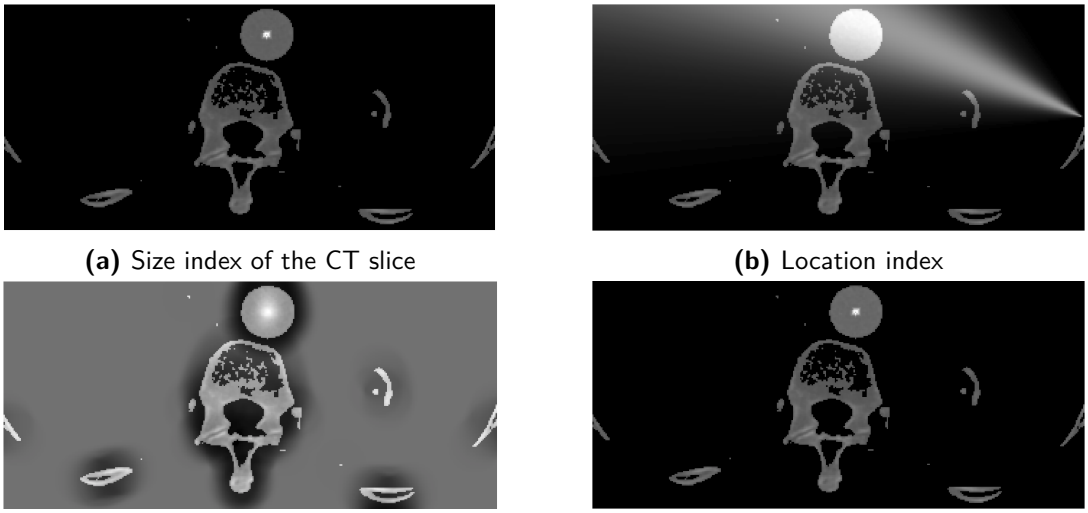
Figure 3-6: $\mathrm{CT}_{\Sigma\mathrm{row}}$, CT masking step 2

Removing the aorta

Most likely, the aorta is still present in the image CT_m, as the contrast agent present in the aorta will yield intensity values above the bone threshold. During this step we remove the aorta by creating an aorta mask based on region growing.

The region growing method requires a seed point. For this, we use a method that was proposed by Metz et. al [16]. This method localizes a high-intensity, spherical object (the aorta) in a 2D CT slice, by combining several criteria. The criteria applied are based on the object size, object location and object roundness. The parameters are set to represent the aorta and are stated in Chapter. 4. To make our approach more robust, we take multiple slices throughout the image CT_m in order to determine the aorta position. All slices between 0.5 and $0.9 * S_{\mathrm{CT}}$ are used. The results of executing the seed point detection on a single slice is observable in Fig. 3-7.

In Fig. 3-7a a slice of the masked CT (as depicted in Fig. 3-2c) is depicted. In overlay, the result of applying the size criteria is plotted. The higher the value of the criteria (denoting more likely to correspond to aorta location) the higher the gray value. Notice the high intensity spot of the size index at the aorta's center. In Fig. 3-7b the slice is plotted with an overlay of the location index. The location index is set to favour objects at the aorta location. Fig. 3-7c depicts the slice with an overlay of the roundness index. Here, round objects result in high index values. The result of the approach is a multiplication of the three indices and is depicted in Fig. 3-7d.



(a) Size index of the CT slice

(b) Location index

(c) Roundness index

(d) Result of seed point detection

Figure 3-7: Aorta detection; by computation of size-, location- and roundness-indices

The MATLAB code used to perform the region growing was created in [13]. The absolute threshold value was set to $0.02 * max(\mathrm{CT}_m)$. The method also fills enclosed holes in the created mask. The result after applying the mask created by this method is showed in Fig. 3-8. To guarantee that also partial volume voxels at the border of the aorta are removed, we include a dilation on the mask with a kernel size of 7 mm. The result of applying this dilated mask is shown in Fig. 3-9. The result of the entire aorta removal step can also be viewed in Fig. 3-2d. We denote the result of the aorta removal and thus the entire CT pre-processing stage as $\mathrm{CT}_{\mathrm{pre}}$.

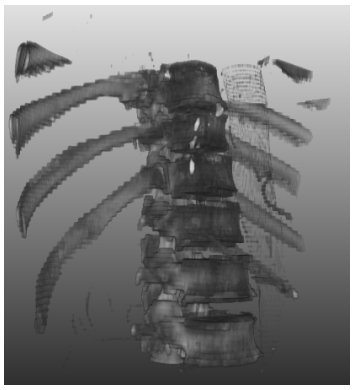


Figure 3-8: Aorta removal; result of applying the aorta mask

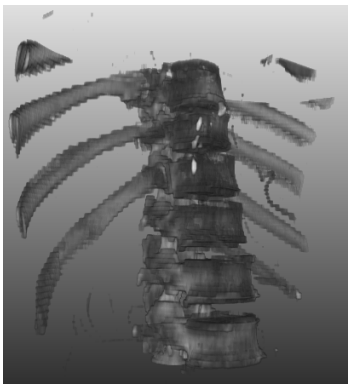
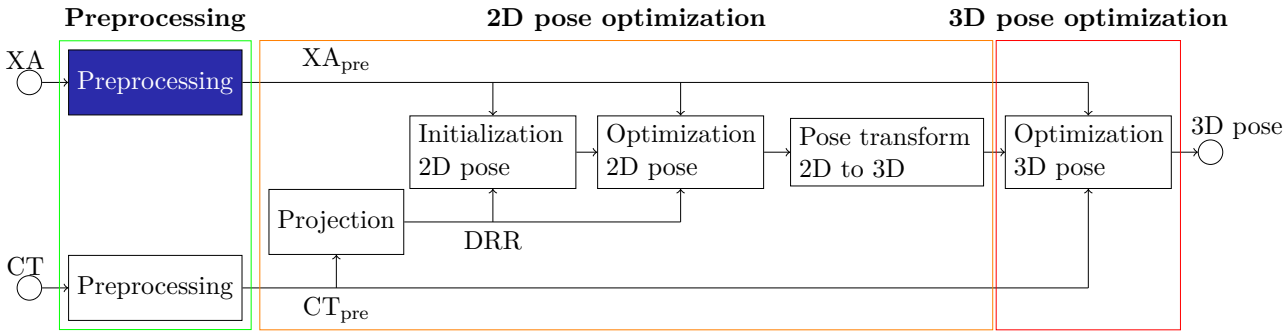


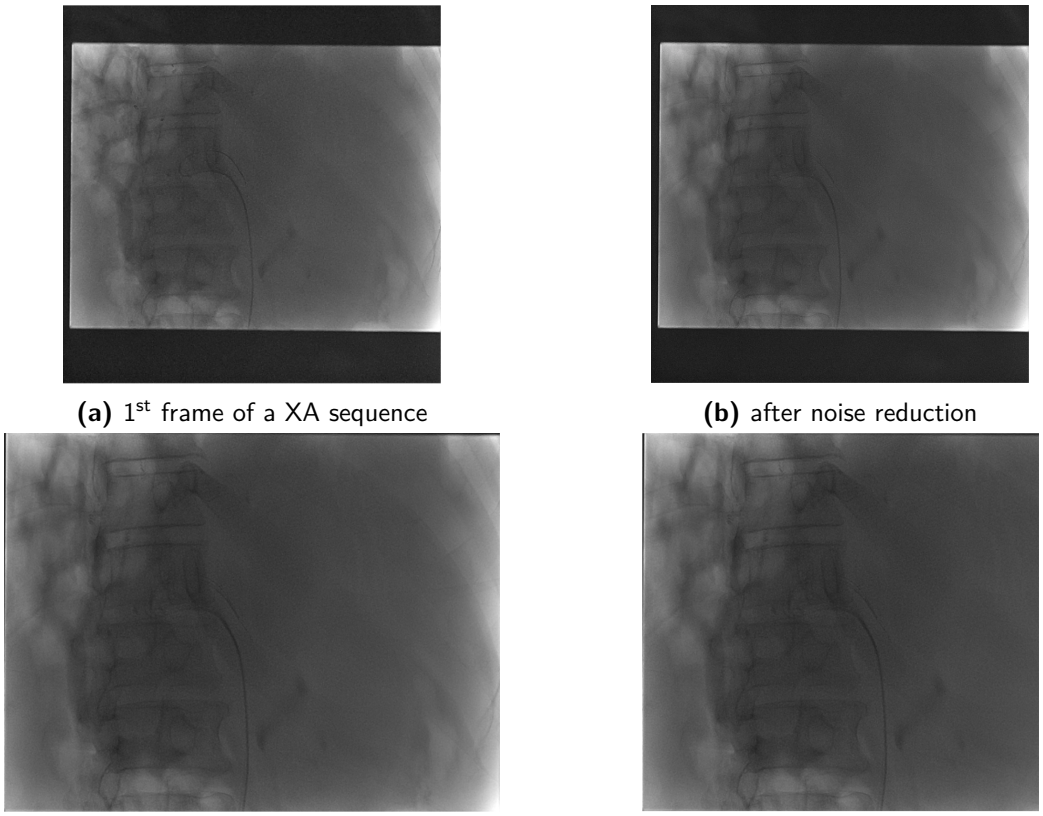
Figure 3-9: Aorta removal; result of applying the aorta mask with 7 mm dilation

3-1-2 XA pre-processing



In this section we discuss the preprocessing of the X-ray images. Fig. 3-10a shows an X-ray image. Recall that we want to keep only the region of interest, being the spine and the part of the ribs closest to the spine. We implement this in three steps. First we reduce the noise of the image by averaging over multiple frames. Then we remove any black regions present at times due to the beam detector. Finally we mask the X-ray image to only include the aforementioned region of interest.

Fig. 3-10 depicts the result of all stages in the XA preprocessing.



(a) 1st frame of a XA sequence **(b)** after noise reduction

(c) after band removal **(d)** after masking

Figure 3-10: XA pre-processing; consisting of noise reduction, band removal and masking

Noise reduction

The first step we take is reducing the noise in the XA image. We achieve this by taking the average over nF number of frames. In Fig. 3-10b the result of the noise reduction is depicted using the first four frames of the sequence.

Let XA_n be the image after taking the average of the first nF frames of the original XA sequence as in Eq. 3-11.

$$\mathrm{XA}_n(r,c) = \frac{\sum_{f=1}^{nF} \mathrm{XA}(r,c,f)}{nF} \tag{3-11}$$

Band removal

Next, we want to get rid of the dark bands at the top, left and bottom of the image XA_n. These are effects of diaphragm settings on the imaging device. To achieve this, we use the difference in intensity between the region inside and outside the bands.

We start by computing the horizontal sum of the image XA_n as $\mathrm{XA}_{n\Sigma\mathrm{col}}(r) = \sum_{c=1}^{C_{\mathrm{XA}}} \mathrm{XA}_n(r,c)$. $\mathrm{XA}_{n\Sigma\mathrm{col}}$ is then normalized and subsequently smoothed by convolution using a Gaussian ker-

nel of $\sigma = 1\ mm$. Smoothing is applied to counter any left over noise in the image. The result of these efforts is depicted in Fig. 3-11.

After the smoothed normalized horizontal sum is computed, it's gradient $\nabla \mathrm{XA}_{n\Sigma\mathrm{col}}(r)$ is calculated. The gradient is computed by extracting each value of $\mathrm{XA}_{n\Sigma\mathrm{col}}$ with the preceding value. The band edges are then estimated by first applying a threshold of 0.01 on the absolute of $\nabla \mathrm{XA}_{n\Sigma\mathrm{col}}(r)$. In case of the left edge estimate e_l, we then take the index of the value which is higher than the threshold and closest to the center of $\mathrm{XA}_{n\Sigma\mathrm{col}}$ (at $R_{\mathrm{XA}}/2$) for $1 < r < R_{\mathrm{XA}}/2$. In case of the right edge estimate e_r, we take the index of the value higher than the threshold and closest to the center of $\mathrm{XA}_{n\Sigma\mathrm{col}}$ for $R_{\mathrm{XA}}/2 < r < R_{\mathrm{XA}}$. The result of these efforts is depicted in Fig. 3-13.

The process of computing the top (e_t) and bottom (e_b) band edges is very similar. Here, instead of the horizontal sum and gradient, the vertical ones are computed (see Fig. 3-12 and 3-14). The four edges together enclose the mask which is applied on the image XA_n (see Fig. 3-15). The result of applying this stage of band removal can be observed in Fig. 3-10c. We call the resulting image XA_b, computed as follows

$$\mathrm{XA}_b = \mathrm{XA}_n \left(e_t < r < e_b, e_l < c < e_r\right). \tag{3-12}$$

Masking

Now that we have reduced the noise and gotten rid of the bands in the image, we have one final task to perform. In order to keep only the region of interest of the spine and ribs closest to it, we will now mask the image XA_b.

The aforementioned spine-edge detection (the description of which will be given in Sec. 3-2-1) is applied on the result of the noise-reduced and band-removed image. After the detection of the spine's left- and right edges (c_l and c_r) we use these values in combination with a mask width of c_m (equal to the width used in the CT masking stage in Sec. 3-1-1) to generate column boundaries for our image. Eq. 3-13 and Fig. 3-16 depict this process. In the figure c_l and c_r are depicted as blue solid lines whereas the new column boundaries $[c_l - c_m, c_r + c_m]$ are depicted as blue dashed lines.

The result of the entire masking stage can be viewed in Fig. 3-10d with $c_m = 100mm$. We shall denote the result of the masking and thus the entire XA preprocessing stage as XA_{pre}, computed as

$$\mathrm{XA}_{\mathrm{pre}} = \mathrm{XA}_b \left(r, c_l - c_m < c < c_r + c_m\right). \tag{3-13}$$

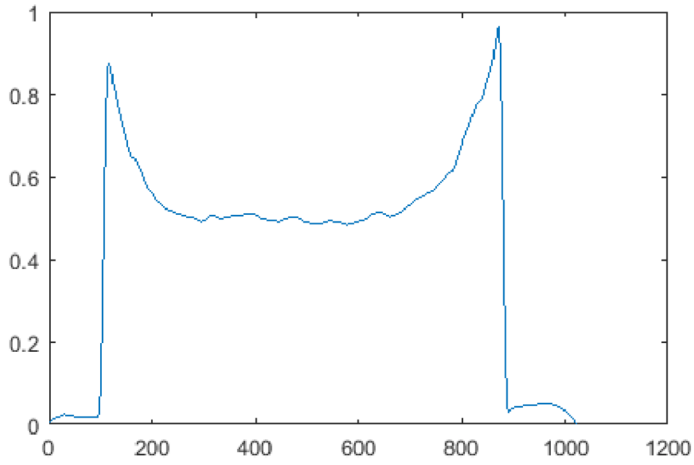


Figure 3-11: normalized smoothed horizontal sum of XA_n

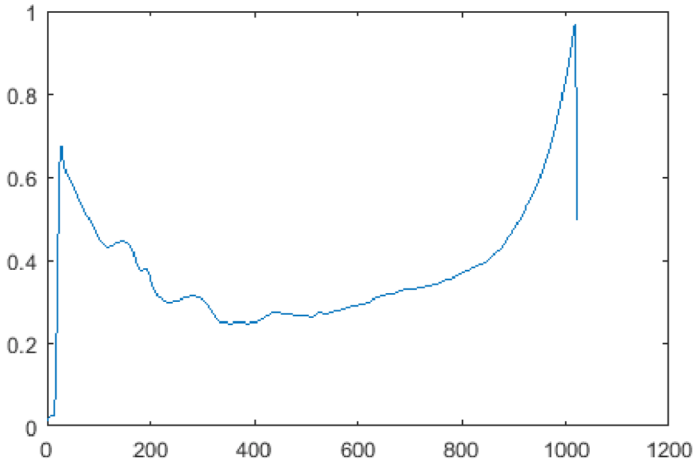


Figure 3-12: normalized smoothed vertical sum of XA_n

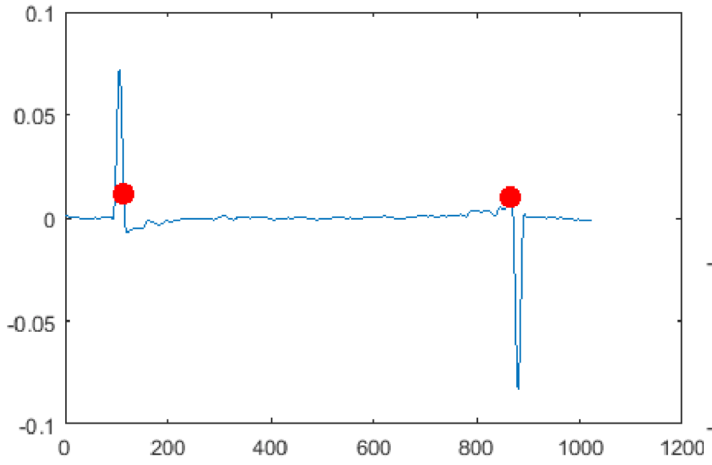


Figure 3-13: gradient of normalized smoothed horizontal sum of XA_n, the red dot denotes the edge estimation

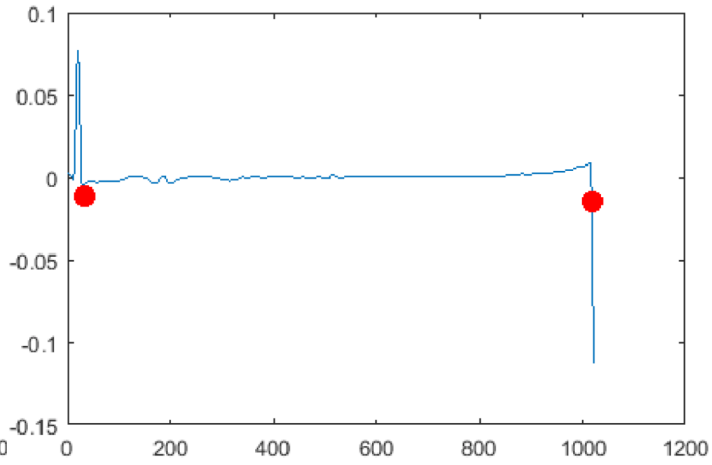


Figure 3-14: gradient of normalized smoothed vertical sum of XA_n, the red dot denotes the edge estimation

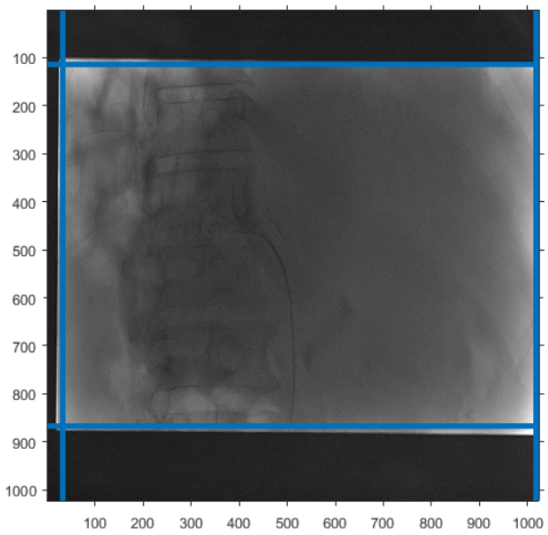


Figure 3-15: resulting mask of band removal

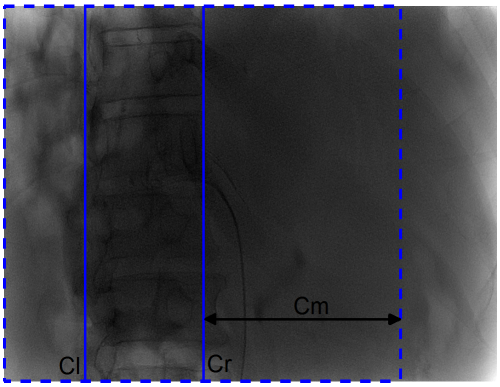
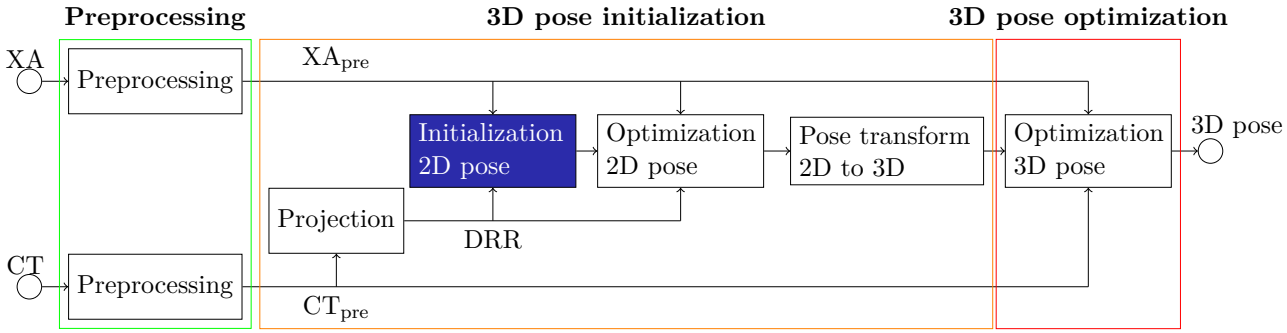


Figure 3-16: resulting mask based on spine edge detection in image XA_b

3-2 3D pose initialization

3-2-1 2D pose initialization



Recall the scheme of our registration method as depicted above. Now that we have created CT_{pre} and XA_{pre} we can begin the initialization of the 2D pose in the detector coordinate system (see App. A concerning coordinate systems). However prior to this initialization we first need to project CT_{pre}. The details of projecting a 3D image onto a 2D plane can be found in App. B. Suffice to say a digitally reconstructed radiograph(DRR) of CT_{pre} (Fig. 3-2d at its initial position in the world coordinate system is constructed, the result of which can be seen in Fig. 3-18. The pre-processed image XA_{pre} is plotted alongside the DRR in Fig. 3-17.

During this stage we use our prior knowledge about the human anatomy to quickly align DRR with XA_{pre}. In this way we turn the otherwise global optimization problem of the subsequent stages into a local one.

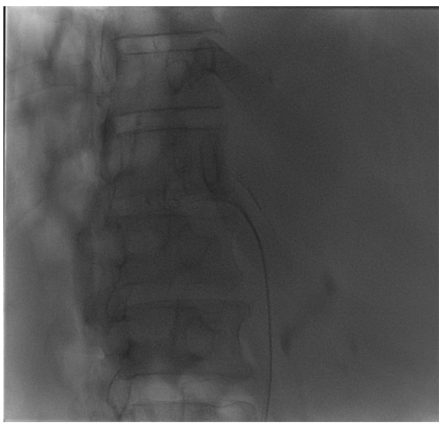


Figure 3-17: XA_{pre}

Figure 3-18: DRR, projection of CT_{pre}

Aligning the DRR with the XA_{pre} means determining the four variables which make up the DRR's 2D pose:

- horizontal translation ($\hat{x}$)
- vertical translation ($\hat{y}$)
- rotation ($\hat{\alpha}$)
- scale ($\hat{sc}$)

The 2D pose will be stated as $\text{Pose}_{\text{2D,init}} = [\hat{x} \quad \hat{y} \quad \hat{\alpha} \quad \hat{sc}]$.

In finding the proper 2D pose, first consider Fig. 3-17 and 3-18. One can observe the spine forming a clear structure in both modalities. We can use the spine to determine the necessary horizontal translation, orientation and scale. This leaves us with the vertical translation in which the spine does not offer great help, as the vertebrae all look alike. However the ribs are well defined in vertical direction. We use this in our estimation of the necessary vertical translation. The following sections explain how we estimate each of the four variables which make up the 2D pose.

Also, we introduce a test case we will be using throughout the thesis to visualize the workings of each estimation step. The XA_{pre} and DRR of the test case are depicted in Fig. 3-27 and 3-28 respectively.

Step 1. Horizontal translation

Let us start by estimating the necessary horizontal translation $\hat{x}$ of the image DRR, in order to align it with the image XA_{pre}. We use the spine, which is present in both images, and estimate the 2D x-position $\hat{x}$ by determining their horizontal positions. Here, we use the spine edge detection method mentioned earlier, tailored to detect the spine's edges in both modalities as follows.

As the spine is a clear, often almost perfectly vertical structure in the image, we start with determining the sum intensity over all columns of the DRR. As the XA_{pre} inhabits more structures besides the spine and ribs, the median intensity over the columns is more appropriate here. The sum of the DRR and median of the XA_{pre} are computed as follows:

$$\text{XA}_{\text{med}}(c) = \text{median}\left(\text{XA}(r,c)\right), \tag{3-14}$$

$$\text{DRR}_{\Sigma\text{row}}(c) = \sum_{r=1}^{R_{\text{DRR}}} \text{DRR}(r,c), \tag{3-15}$$

and can be observed in Fig. 3-20 and 3-19. In the figures the vertical median or sum is depicted as a blue line. Note that the intensity in CT images correspond to attenuation coefficients whereas that in XA images corresponds to received X-ray photons. Therefore the spine is represented by high image values in case of the DRR and low values in case of the XA_{pre}. For visualization purposes we plot the inverse of the actual XA median, a high value in the image thus corresponding to a low median value.

Additionally we use the fact that the spine has a certain width sw, very consistent among all human beings. The average width of a human spine is $sw = 40\ mm$ [19]. Magnifying this with the factor SID/SOD due to the projection will yield us the expected spine width in the images DRR and XA_{pre}. SID stands for the distance between source and image(detector plane), whereas SOD stands for the distance between source and object(patient at the world isocenter), both parameters will be visualized in Sec. 3-2-3.
We can use the spine width to compute a forward moving median of the earlier computed vertical median or sum, which maxima in case of $\mathrm{DRR}_{\Sigma\mathrm{row}}$ and minima in case of $\mathrm{XA}_{\mathrm{med}}$ (due to the aforementioned difference in the nature of the image values) should correspond to the left edge of the spine. We compute the forward moving median and their optima of the $\mathrm{XA}_{\mathrm{med}}$ and $\mathrm{DRR}_{\Sigma\mathrm{row}}$ as follows:

$$\mathrm{XA}_{\mathrm{med}\rightarrow}(c) = \mathrm{median}\left(\mathrm{XA}_{\mathrm{med}}(c : c + sw)\right), \tag{3-16}$$

$$p_{\mathrm{XA}} = \mathrm{argmin}\left(\mathrm{XA}_{\mathrm{med}\rightarrow}(c)\right), \tag{3-17}$$

$$\mathrm{DRR}_{\mathrm{med}\rightarrow}(c) = \mathrm{median}(\mathrm{DRR}_{\Sigma\mathrm{row}}(c : c + sw)), \tag{3-18}$$

$$p_{\mathrm{DRR}} = \mathrm{argmax}\left(\mathrm{DRR}_{\mathrm{med}\rightarrow}(c)\right). \tag{3-19}$$

The forward moving median of both the DRR and $\mathrm{XA}_{\mathrm{pre}}$ can be observed again in Fig. 3-20 and 3-19. In the figures the forward moving median is depicted as a red line, the maximum peak is denoted with a red filled circle. Note that again for visualization purposes, in case of the XA_{pre} the inverse of the forward moving median is plotted.

The estimate at this point is approximately at the spine's left edge. This enables us to improve our estimate by looking at the gradient of the vertical median within a small local area. The gradient is computed by extracting a value with its preceding value. We set the local area to include everything within a quarter spine width of the estimate deduced from the moving average. We set this particular width as vertebrae exhibit a low intensity at their spinal canal which causes gradients at the canal edges. A quarter of the spine width excludes misclassification's due to vertebra structure for all possible angles. The maxima of the gradient in case of the $\mathrm{DRR}_{\Sigma\mathrm{row}}$ and the minima in case of the $\mathrm{XA}_{\mathrm{med}}$ correspond to the spine's left edge position. They are computed as follows:

$$p_{\mathrm{XAspine,l}} = \mathrm{argmin}\left(\nabla\mathrm{XA}_{\mathrm{med}}(p_{\mathrm{XA}} - \frac{1}{4}sw : p_{\mathrm{XA}} + \frac{1}{4}sw)\right), \tag{3-20}$$

$$p_{\mathrm{DRRspine,l}} = \mathrm{argmax}\left(\nabla\mathrm{DRR}_{\Sigma\mathrm{row}}(p_{\mathrm{DRR}} - \frac{1}{4}sw : p_{\mathrm{DRR}} + \frac{1}{4}sw)\right). \tag{3-21}$$

and are shown in Fig. 3-20 and 3-19. In the figures the gradient of the vertical median or sum is depicted as a yellow line, the spine edge estimate is denoted with a yellow filled circle. Note that in case of the $\mathrm{XA}_{\mathrm{pre}}$ the inverse of the gradient is plotted.

We compute the horizontal translation $\hat{x}$ necessary to align the images by taking the difference between the estimations of the spine's left edge in both images as:

$$\hat{x} = p_{\mathrm{XAspine,l}} - p_{\mathrm{DRRspine,l}}. \tag{3-22}$$

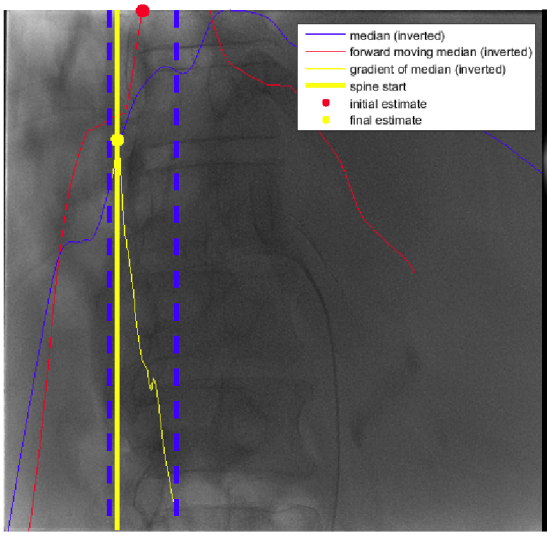


Figure 3-19: Detection of the spine's left edge in $\mathrm{XA}_{\mathrm{pre}}$

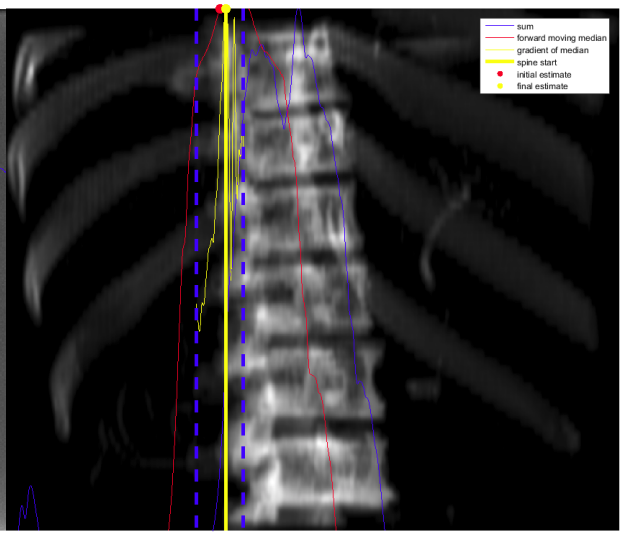


Figure 3-20: Detection of the spine's left edge in DRR

In summary, we first computed the vertical median and sum of respectively the XA_{pre} and DRR, using the fact that the spine is a vertically oriented structure. Subsequently we compute the horizontal forward moving median of the vertical median or sum (incorporating our knowledge of human spine widths). We then defined an area around the optimum of the forward moving median. Within this local area we calculated the gradient of the median or sum and took its optimum to represent the spine's left edge. Finally we created $\hat{x}$ out of the computed edge locations.

Step 2. Rotation

Next we address the alignment of the DRR with the $\mathrm{XA}_{\mathrm{pre}}$ with respect to the orientation . To achieve this, we first detect the individual orientation of the spine in both images. In this approach, we make use of the fact that the spine looks very similar at subsequent rows.

Thus, in order to detect the spine orientation we propose a tracking of the spine's left edge. Here, we compare the medians of patches of the image (containing all columns but only a number of rows),

$$\mathrm{XApatch}_{\mathrm{med}}(c) = \mathrm{median}(\mathrm{XA}_{pre}(r_{\mathrm{patch}} < r < r_{\mathrm{patch}} + R_{\mathrm{patch}}, c)), \tag{3-23}$$

with a template of the spine, working our way from the top of the image (at $r_{\mathrm{patch}} = 1$) to the bottom (at $r_{\mathrm{patch}} = R_{\mathrm{XA}} - R_{\mathrm{patch}}$). The spine template has a width corresponding to a spine width of the aforementioned $sw = 40\ mm$ and is initialized with a uniform value of $0.25 * \mathrm{median}(\mathrm{XA}(r, c))$ to represent the low intensity spine.
For each patch the normalized cross correlation with the spine template is calculated as

$$\mathrm{xcorr}(r) = \sum_{m=-\infty}^{\infty} \mathrm{XApatch}_{\mathrm{med}}(m) * \mathrm{template}(m + r). \tag{3-24}$$

The position of the maximum $p_{\text{patch}} = \text{argmax}(\text{xcorr})$ of the cross correlation represents the spine's edge position in the patch. The template is then updated with the detected spine segment by $\text{template} = (\text{template} * 0.75) + (0.25 * \text{XApatch}_{\text{med}}(p_{\text{patch}} : p_{\text{patch}} + sw))$.
The edge positions $p_{\text{patch}}(r_{\text{patch}})$ with $1 < r_{\text{patch}} < R_{\text{XA}} - R_{\text{patch}}$ are then used to fit a line along the spine of XA_{pre}. The angle between the line and the vertical axis is the estimation of the necessary rotation $\angle\text{XA}_{\text{pre}}$, to align the image with the vertical axis.
Tracking the spine's left edge in the DRR image as opposed to the XA_{pre} image requires but one alteration. Now, the spine template is initialized with $max(\text{DRR}(r, c))$ to represent the high intensity of the spine.

This edge tracking may be prone to errors (inaccuracies), as the edge detection in small patches is subject to noise. Therefore, we adopt a multi-scale approach, where the edge detection results of patches of varying heights are combined. In this way, we are trading sensitivity for robustness. Running the edge tracking algorithm for various patch heights and averaging the results should make for a more robust tracker. Fig. 3-21 and 3-22 show the results of this effort with a patch height of $1 - 20\ mm's$. The fitted lines are depicted in blue. The orientations of the spines are denoted by $\angle\text{XA}_{\text{pre}}$ and $\angle\text{DRR}$. Additionally, to make the edge tracking more robust, we limit the range of possible edge locations between subsequent patches of the same size to be within 20 mm of the previous detected edge position.

Now that we have determined the orientation of both images we can create $\hat{\alpha}$ (the angle used to rotate the DRR) as

$$\hat{\alpha} = \angle\text{XA}_{\text{pre}} - \angle\text{DRR}. \tag{3-25}$$

Figure 3-21: Tracking of the spine's left edge in XA_{pre} using vertical patch heights of $1 - 20mm$

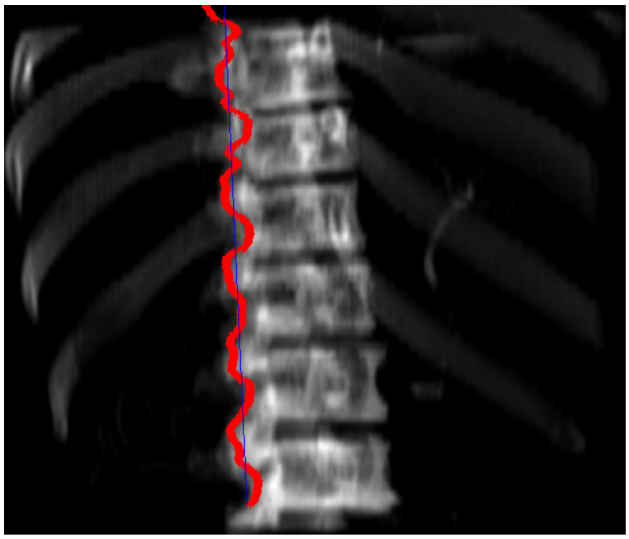


Figure 3-22: Tracking of the spine's left edge in DRR using vertical patch heights of $1 - 20mm$

Step 3. Scale

The two images $\mathrm{XA}_{\mathrm{pre}}$ and DRR may differ in scale, in case the $\mathrm{CT}_{\mathrm{pre}}$ is positioned either too close or too far away from the detector plane. Again, we propose to utilize the spine in both images to determine the proper scale adjustment needed in the DRR.

For this, we again use the spine left edge detection technique. Using a vertical median, followed by a forward moving median, followed by a gradient computation, we were able to detect the left edge of the spine in both DRR and $\mathrm{XA}_{\mathrm{pre}}$ images. We can actually make minor adjustments in this process in order to turn the left edge detector into a right edge detector. First, we replace the forward moving median with a backward moving median. And second, we swap maximum- with minimum calculations and vice-versa. The following equations describe the right edge detection:

$$\mathrm{XA}_{\mathrm{med}}(c) = \mathrm{median}\left(\mathrm{XA}(r, p_{\mathrm{XAspine,l}} : C_{\mathrm{XA}})\right), \tag{3-26}$$

$$\mathrm{XA}_{\mathrm{med}<-}(c) = \mathrm{median}\left(\mathrm{XA}_{\mathrm{med}}(c - sw : c)\right), \tag{3-27}$$

$$p_{\mathrm{XA}} = \mathrm{argmax}\left(\mathrm{XA}_{\mathrm{med}<-}\right), \tag{3-28}$$

$$p_{\mathrm{XAspine,r}} = \mathrm{argmax}\left(\nabla \mathrm{XA}_{\mathrm{med}}(p_{\mathrm{XA}} - \frac{1}{4}sw : p_{\mathrm{XA}} + \frac{1}{4}sw)\right), \tag{3-29}$$

$$\mathrm{DRR}_{\Sigma\mathrm{row}}(c) = \sum_{r=1}^{R_{\mathrm{DRR}}} \left(\mathrm{DRR}(r, p_{\mathrm{DRRspine,l}} : C_{\mathrm{DRR}})\right), \tag{3-30}$$

$$\mathrm{DRR}_{\mathrm{med}<-}(c) = \mathrm{median}\left(\mathrm{DRR}_{\Sigma\mathrm{row}}(c - sw : c)\right), \tag{3-31}$$

$$p_{\mathrm{DRR}} = \mathrm{argmin}\left(\mathrm{DRR}_{\mathrm{med}<-}\right), \tag{3-32}$$

$$p_{\mathrm{DRRspine,r}} = \mathrm{argmin}\left(\nabla \mathrm{DRR}_{\Sigma\mathrm{row}}(p_{\mathrm{DRR}} - \frac{1}{4}sw : p_{\mathrm{DRR}} + \frac{1}{4}sw)\right), \tag{3-33}$$

$$\hat{sc} = \frac{p_{\mathrm{DRRspine,r}} - p_{\mathrm{DRRspine,l}}}{p_{\mathrm{XAspine,r}} - p_{\mathrm{XAspine,l}}}. \tag{3-34}$$

Note that as the left edge detection is performed prior to the right edge detection, the right edge detection is only needed for the part to the right of the left edge detected. After all edges are estimated, the spine widths of the images are updated to $p_{\mathrm{DRRspine,r}} - p_{\mathrm{DRRspine,l}}$ in case of the DRR and $p_{\mathrm{XAspine,r}} - p_{\mathrm{XAspine,l}}$ in case of the XA. The scale estimate $\hat{sc}$ is created by dividing the DRR spine width with the XA spine width. Fig. 3-23 and 3-24 depict the results of this stage. As for the same reasons during Sec. 3-2-1 the inverted values are plotted of the median, backward moving median and gradient in case of the XA image modality.

Step 4. Vertical translation

The final step in the 2D pose initialization is estimating the vertical translation $\hat{y}$. For the first time the spine itself is not sufficient. As it is quite repetitive in the vertical direction. However other structures present in both images, namely the ribs, are more distinguishable in this direction.

There are multiple approaches to estimate the proper vertical translation using the ribs. For example, we could attempt to segment the ribs in both images and use the position of the

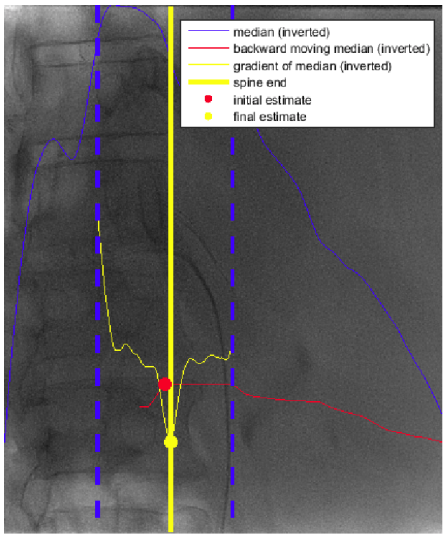


Figure 3-23: Detection of the spine's right edge in $\mathrm{XA}_{\mathrm{pre}}$

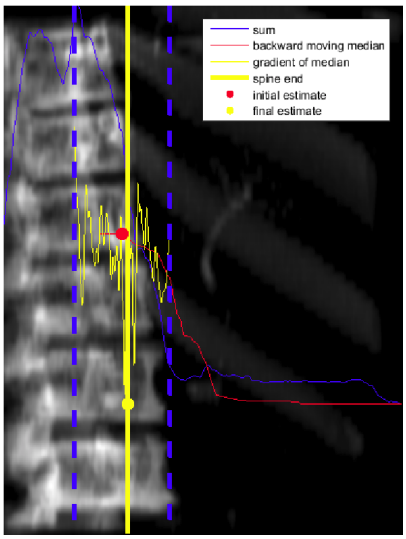


Figure 3-24: Detection of the spine's right edge in DRR

segments to derive the necessary translation. However this approach would require an additional segmentation step, which would be hard given the low contrast of the ribs.
Another approach is to register both images with a similarity metric. The initial position of the DRR image will incorporate the scale $\hat{sc}$, the orientation $\hat{\alpha}$ and the horizontal position $\hat{x}$ that we computed during previous steps. The correct vertical translation, aligning the DRR and XA ideally yields the highest similarity value. Specifically the normalized cross correlation(NCC) should be an adequate metric.

In using the NCC to estimate vertical translation, we start by defining $\mathrm{DRR}_{\alpha sc}$ as the DRR image rotated and scaled using the estimations from their respective steps. Only the part of the DRR is used which overlaps with $\mathrm{XA}_{\mathrm{pre}}$. In constructing this patch, first the estimate $\hat{x}$ is taken into account as follows:

$$\mathrm{DRRpatch}(r,c) = \begin{cases} \mathrm{DRR}_{\alpha sc}(r, |\hat{x}| : R_{\mathrm{DRR}_{\alpha sc}}) & \text{if } \hat{x} < 0 \\ [zeros(R_{\mathrm{DRR}_{\alpha sc}}, \hat{x}) \quad \mathrm{DRR}_{\alpha sc}], & \text{if } \hat{x} > 0 \end{cases}. \tag{3-35}$$

Then the DRRpatch is set to the same horizontal size as XA_{pre} by stating

$$\mathrm{DRRpatch}(r,c) = \begin{cases} \mathrm{DRRpatch}(r, 1 : C_{\mathrm{XA}}) & \text{if } C_{\mathrm{XA}} < C_{\mathrm{DRRpatch}} \\ [\mathrm{DRRpatch} \quad \mathrm{zeros}(C_{\mathrm{XA}} - C_{\mathrm{DRRpatch}})], & \text{if } C_{\mathrm{XA}} > C_{\mathrm{DRRpatch}} \end{cases}. \tag{3-36}$$

The DRRpatch and XA_{pre} now have the same horizontal size and are aligned with respect to horizontal translation, rotation and scale. The NCC is then calculated as:

$$\mathrm{NCC}(y) = \frac{1}{R_{\mathrm{XA}} \times C_{\mathrm{XA}}} \sum_{r,c} \mathrm{DRRpatch}(r+y, c)\mathrm{XA}_{pre}(r,c), \tag{3-37}$$

with y representing the vertical translation. The estimate of the vertical translation $\hat{y}$ is then taken to be the position of the maximum $\hat{y} = \mathrm{argmax}(\mathrm{NCC})$.

Fig. 3-25 shows the result of executing this procedure. $\mathrm{XA}_{\mathrm{pre}}$ has been plotted next to the image in proper proportion. The blue line in Fig. 3-25 denotes the value of the NCC for different vertical translations. For visualization purposes, the right edge corresponds to a high NCC value and the left edge corresponds to a low NCC value. Note that the translations include values outside the scope of the DRRpatch. The topmost value denotes a shift which poses the $\mathrm{XA}_{\mathrm{pre}}$'s top a distance of 2 vertebrae($2 * 35\ mm$[19]) above the top of the DRR image (corresponding to a negative value of y). The value of 2 vertebrae is chosen as for our specific intervention. The pre-interventional CT image and intra-interventional XA image tend to have regions at most 2 vertebrae apart. The bottommost value of the blue NCC line denotes a shift which poses the $\mathrm{XA}_{\mathrm{pre}}$'s bottom a distance of 2 vertebrae below the DRR's bottom.

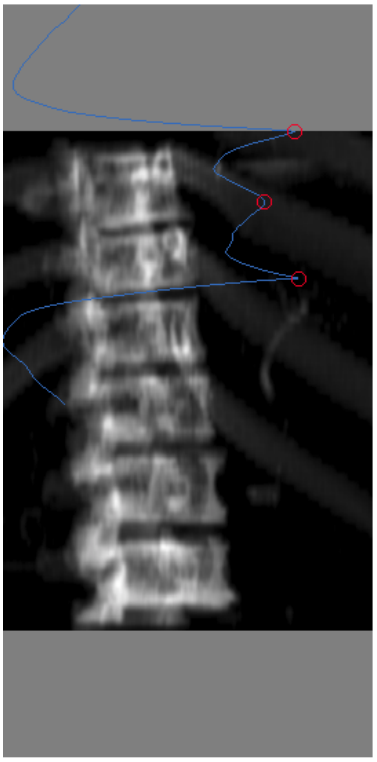


Figure 3-25: Estimation of the necessary vertical translation

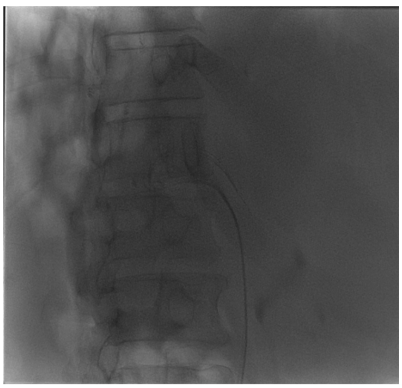


Figure 3-26: XA_{pre}

In the depicted figures we notice multiple local optima $\hat{y}$. This was to be expected as the vertebrae impact the metric, as well as the fact that the ribs themselves do rather lookalike. Ideally one of the detected optima represents the proper vertical alignment. As we do not know which optima is the correct one (and using the NCC value to select is not robust as the overlapping area is not always equal), we pass on multiple 2D poses to the next stage, the 2D pose optimization. This causes that human interaction might be necessary to ultimately determine the correct alignment, if multiple possible vertical translation are detected.

With this final step of aligning the images with respect to their vertical position, we have ini-

tially aligned the DRR and XA_{pre} images. The resulting 2D pose(s) are stated as $Pose_{2D,init} = [\hat{x} \quad \hat{y} \quad \hat{\alpha} \quad \hat{sc}]$.

Example

Note the test case as defined by XA_{pre} and DRR in Fig. 3-27 and 3-28 respectively. Applying the 2D pose initialization stage of this chapter onto the test case yields us the result as shown in Fig. 3-29 to 3-32. In Fig. 3-29 and 3-30, we see an overlay of the DRR with the XA_{pre} before and after the initialization of the 2D pose. Due to amongst others the fact that the vertebrae look so much alike it is quite difficult to get a sense of the improvement in alignment from these two images. In Fig. 3-31 and 3-32 the results are again shown, however in these figures manual annotations on the ribs and vertebrae are also plotted to clarify the improvement in alignment.

In this case the initialization included a shift in both directions and a scaling, no orientation difference was detected. We notice the improvement in alignment, which should be sufficient to change the otherwise global optimization problem of the 2D pose in the next chapter into a local one.

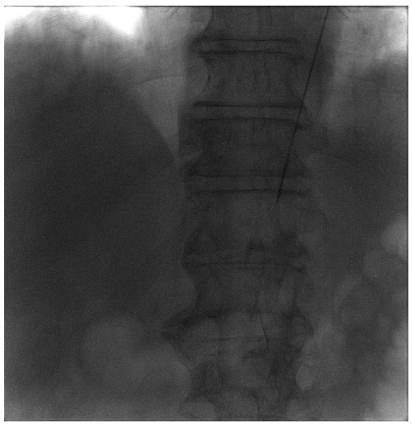


Figure 3-27: $\mathrm{XA}_{\mathrm{pre}}$ of test case

Figure 3-28: DRR of test case

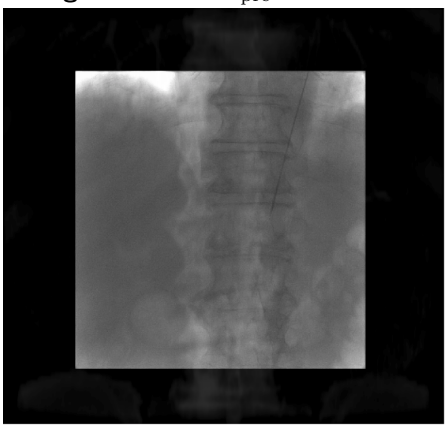


Figure 3-29: Result before 2D pose initialization

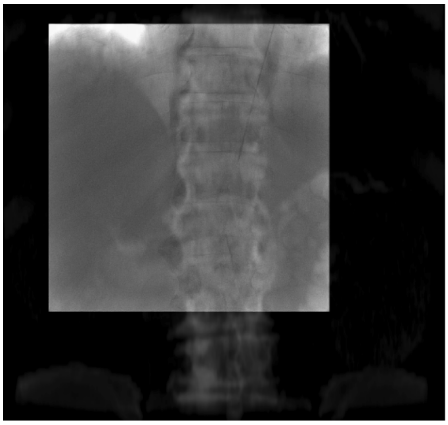


Figure 3-30: Result after 2D pose initialization

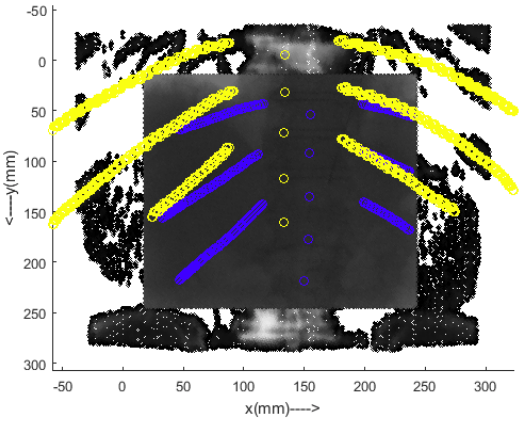


Figure 3-31: Result before 2D pose initialization

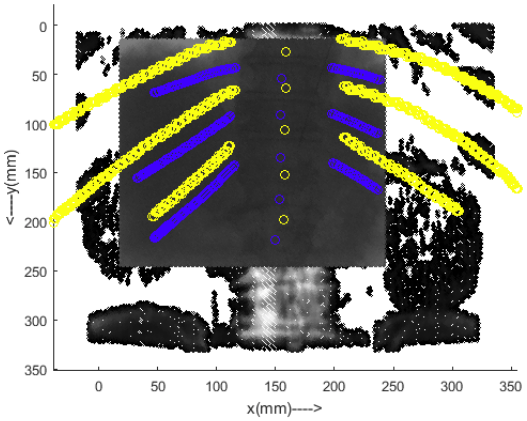
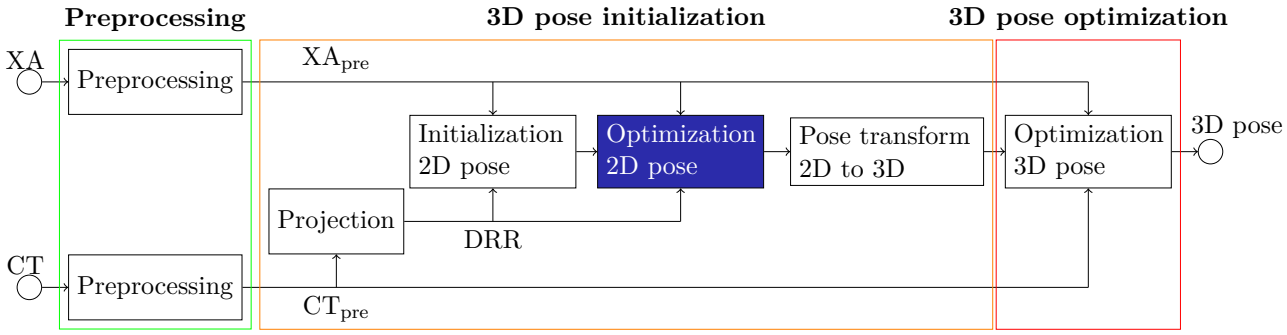


Figure 3-32: Result after 2D pose initialization

3-2-2 2D pose optimization



Recall the scheme of our registration method as depicted above. We have created CT_{pre} and XA_{pre} and aligned the DRR(the projection of CT_{pre}) with XA_{pre}. During this stage we will improve our 2D alignment using an optimization scheme. To optimize the 2D pose consisting of translation, rotation and scale between the images XA_{pre} and DRR(see Fig. 3-17 and 3-18 respectively), we need to select a similarity measure. Secondly, we need an optimization method to find the transformation corresponding to the maximum similarity. During this section we will discuss the requirements on the similarity measure and optimization method and explain which we ultimately employ and why.

Recall that the goal of the previous stage was to create a sufficiently accurate initial 2D alignment between the DRR and the XA_{pre}. This was a paramount step as the otherwise global optimization problem during this stage has been turned into a local one.
Analogues to this relationship between the previous and present stage, the 2D pose optimization stage paves the way for a local optimization approach in the subsequent 3D pose optimization stage. An important aspect we should not forget is that as opposed to 2D pose initialization/optimization, 3D pose optimization is a computational complex stage due to the increase in image size and number of pose parameters. Every improvement in the inital point's accuracy may greatly reduce the computational complexity needed in the 3D pose optimization as less iterations will be required.

Similarity measure

In finding an appropriate similarity measure we consider the XA_{pre} and DRR images as in Fig. 3-17 and 3-18. The pre-processing of the XA reduced the level of noise, removed any dark bands if still present, and masked the image to a certain width left and right from the spine's edges. The CT has also been pre-processed to only include the spine and the closest part of the ribs.
There are two types of similarity measures which can be considered. Intensity based and gradient based. In other work, sometimes calibration based approaches are used[12], however as the name suggest, they require carefully calibrated and traceable equipment and placed markers. As we only have the images, we can only work with their pixel intensities.
Now as for intensity based similarity measures, the normalized cross correlation(NCC) and mutual information(MI) are promising candidates as both the fluoroscopy and the DRR should

have similar intensity distribution. We test both measures during the experiments in Chapter 5.

Mutual information stands for the amount of information one variable describes about another. A high value will thus correspond to great similarity between the two variables or images. Another way to look at the mutual information is as the measure of reduction in uncertainty of image Y due to knowledge of image X. Notice that the mutual information has the distributions of both images as an underlying measure, they should be very similar in order to yield a high mutual information. These distributions have a parameter, named the binning parameter, which is the discretization of the histograms (as MI is computed from the histograms of both distributions). A small binwidth represents a more accurate approximation, however it also leads to a sparse histogram, which is less informative.
Calculating the mutual information between $\mathrm{XA_{pre}}$ and the overlapping segment of the DRR called $\mathrm{DRR_{overlap}}$ is done as follows:

$$H(\mathrm{XA_{pre}}) = -\sum_{i=1}^{R_{\mathrm{XA_{pre}}} * C_{\mathrm{XA_{pre}}}} P(\mathrm{XA}_{pre_i}) log(P(\mathrm{XA}_{pre_i})), \tag{3-38}$$

$$H(\mathrm{DRR_{overlap}}) = -\sum_{i=1}^{R_{\mathrm{XA_{pre}}} * C_{\mathrm{XA_{pre}}}} P(\mathrm{DRR}_{overlap_i}) log(P(\mathrm{DRR}_{overlap_i})), \tag{3-39}$$

$$H(\mathrm{XA_{pre}}, \mathrm{DRR_{overlap}}) = P(\mathrm{XA_{pre}}) * P(\mathrm{DRR_{overlap}}), \tag{3-40}$$

$$MI = H(\mathrm{XA_{pre}}) + H(\mathrm{DRR_{overlap}}) - H(\mathrm{XA_{pre}}, \mathrm{DRR_{overlap}}), \tag{3-41}$$

with $P(\mathrm{XA_{pre}})$ and $P(\mathrm{DRR_{overlap}})$ being the histogram of $\mathrm{XA_{pre}}$ respectively $P(\mathrm{DRR_{overlap}})$, $H(\mathrm{XA_{pre}})$ and $H(\mathrm{DRR_{overlap}})$ the entropy of the two and $H(\mathrm{XA_{pre}}, \mathrm{DRR_{overlap}})$ being the joint entropy of the two images.
To visualize how the MI behaves, Fig. 3-33 depicts two 1D signals X and Y and Fig. 3-34 denotes the result of calculating the mutual information between X and Y for various shifts.

Normalized cross correlation is a direct measure of similarity. It can be described as the sliding inner product of two variables. To obtain the cross correlation between two variables one will shift one of the signals a specific amount. At this shift the inner product is calculated for all corresponding sets of points and added up. The result is the cross correlation coefficient at the index of the shift. Normalization is used to counter differences in intensity scales.
This measure was earlier used in our registration method, specifically during the estimation of the vertical translation in Sec. 3-2-1. Calculating the normalized cross correlation between $\mathrm{XA_{pre}}$ and $\mathrm{DRR_{overlap}}$ is done as follows:

$$\mathrm{NCC} = \frac{1}{R_{\mathrm{XA}} \times C_{\mathrm{XA}}} \sum_{r,c} \mathrm{DRRpatch}(r,c) \mathrm{XA_{pre}}(r,c). \tag{3-42}$$

To elaborate on how the NCC behaves, Fig. 3-33 depicts two 1D signals X and Y and Fig. 3-34 denotes the result of calculating the normalized cross correlation between X and Y for various shifts.

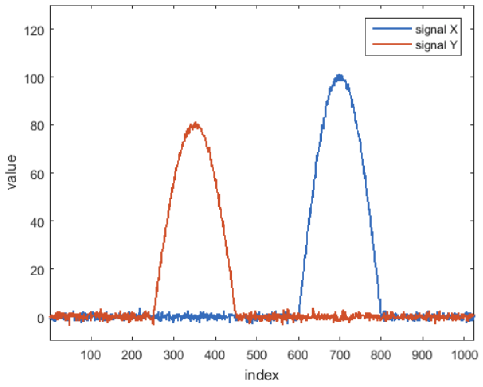


Figure 3-33: Two 1D signals, X(blue) with its peak at 600 and Y(red) with its peak at 250

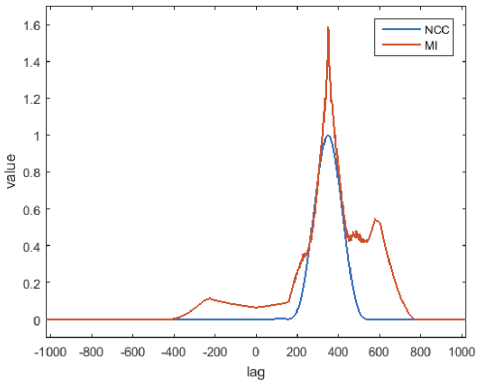


Figure 3-34: MI(red) and NCC(blue) between X and Y, the correct lag for optimal alignment is 350

The mentioned two similarity measures (MI and NCC) will both be tested during the experiments. The results of which will determine which measure will be selected for this 2D pose optimization stage of the method.

Optimization method

Now that we know more about the behaviour of the similarity measures(the objective functions), we proceed to select the optimization technique.

We reject a number of optimizer classes. As the problem is not linear, nor quadratic or convex, optimizers assuming such objective function are not applicable in our case. This leaves us with more complex non-linear optimization methods. Newton and Quasi-Newton optimizers (e.g. Levenberg-Marquadt and Davidon-Fletcher-Powell) involve approximating the gradient or Hessian with each iteration. Another way is using direction determination and line search methods. There are a number of such optimizers which do not utilize or make an computational expensive estimate of the gradient or Hessian. Most prominent and proven amongst these is the Powell direction determination, often used in combination with a variable step size. Other line search approaches do exist, such as the golden section approach, which might reduce the number of iterations needed to reach the optimum.

Due to the advantage of decreased computational complexity and suitability to our image modalities and similarity measures we use the Powell optimizer with variable step size.

Example

Recall the test case as defined by XA_{pre} and DRR in Fig. 3-27 and 3-28 respectively. Applying the 2D pose optimization stage of this chapter (using NCC) onto the test case yields us the result as shown in Fig. 3-35 to 3-38. In Fig. 3-35 and 3-36 we see an overlay of the DRR with the XA_{pre} before and after the optimization of the 2D pose. In Fig. 3-37 and 3-38 the results are again shown, however in these figures manual annotations on the ribs and vertebrae are also plotted to clarify the improvement in alignment.

In this case the optimization included a shift in both directions and a small decrease in the scale factor. There was no change in orientation. We notice the improvement in alignment, which is meant to be sufficient to change the otherwise global optimization problem of the 3D pose (detailed in the next chapter) into a local one.

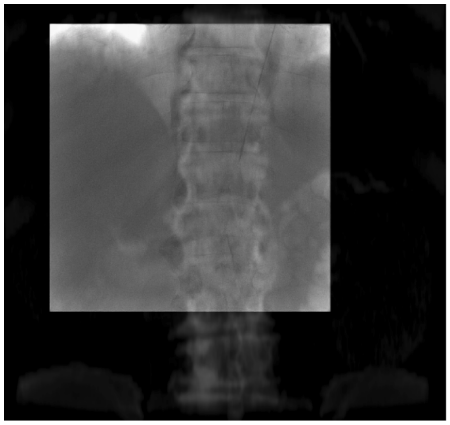


Figure 3-35: Result before 2D pose optimization

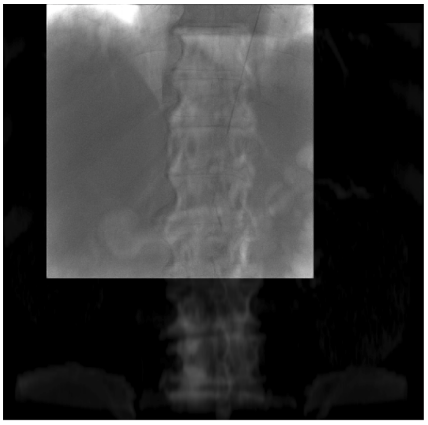


Figure 3-36: Result after 2D pose optimization

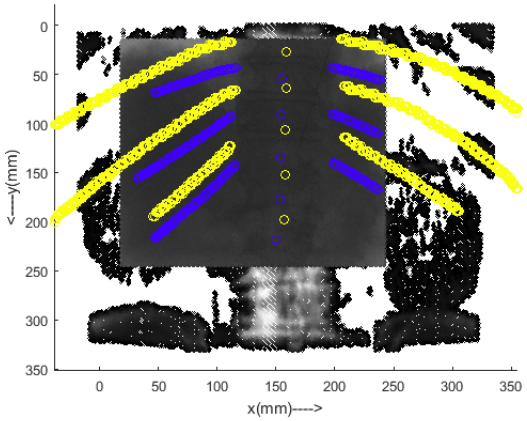


Figure 3-37: Result before 2D pose optimization

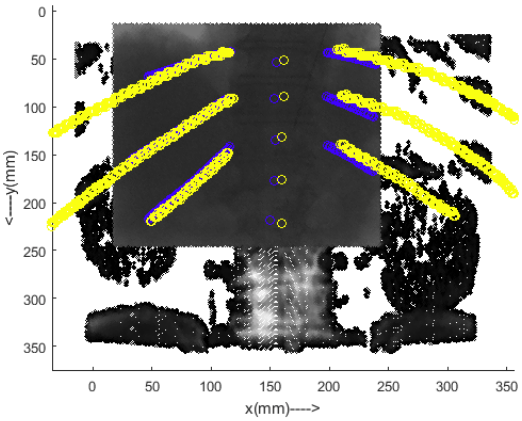
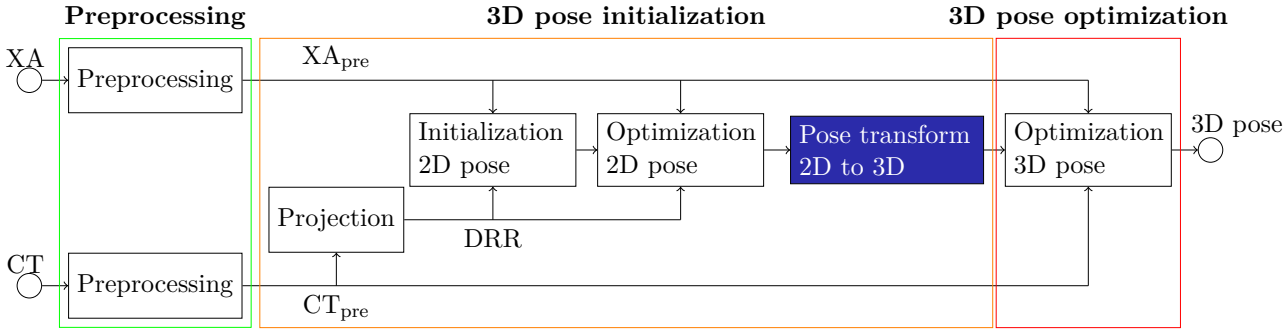


Figure 3-38: Result after 2D pose optimization

3-2-3 Transformation from 2D pose to 3D pose



Recall the scheme of our registration method as depicted above. At this point we have created CT_{pre} and XA_{pre} and aligned the DRR(the projection of CT_{pre}) with XA_{pre} and optimized this alignment. Realizing that the next step is to align not the 2D pose between the DRR and the XA_{pre} but the 3D pose of the CT_{pre}, we have to convert the 2D pose ($\text{pose}_{2D} = [x \quad y \quad \alpha \quad sc]$) into its 3D equivalent ($\text{pose}_{3D} = [x_{3D} \quad y_{3D} \quad z_{3D} \quad \alpha_{3D} \quad \beta_{3D} \quad \gamma_{3D}]$).
The 2D pose contains a translation, rotation and scale adjustment, below we discuss how we transform these three elements. More information on the details of transformations between coordinate systems can be found in App. A.

Translation The first two elements of the 2D pose denote a translation of the DRR on the detector plane. We can find out which 3D translation is necessary at the isocenter of the world coordinate system, to result in such a 2D translation on the detector plane.
Consider the point $\mathbf{o}_{\text{world}}$ to be the origin $(0, 0, 0)$ of the world coordinate system. We can transform this point to the X-ray Device 2D coordinate system(the detector plane in 2D) resulting in

$$\mathbf{o}_{\text{X-ray2D}} = T_{\text{world}\rightarrow\text{X-ray2D}} * \mathbf{o}_{\text{world}}.$$

We then add the translation derived from the 2D pose $pose_{2D} = [x \quad y \quad \alpha \quad sc]$ to $\mathbf{o}_{\text{X-ray2D}}$, to obtain

$$\mathbf{p}_{\text{X-ray2D}} = \mathbf{o}_{xray2D} + [x \quad y \quad 0]. \tag{3-43}$$

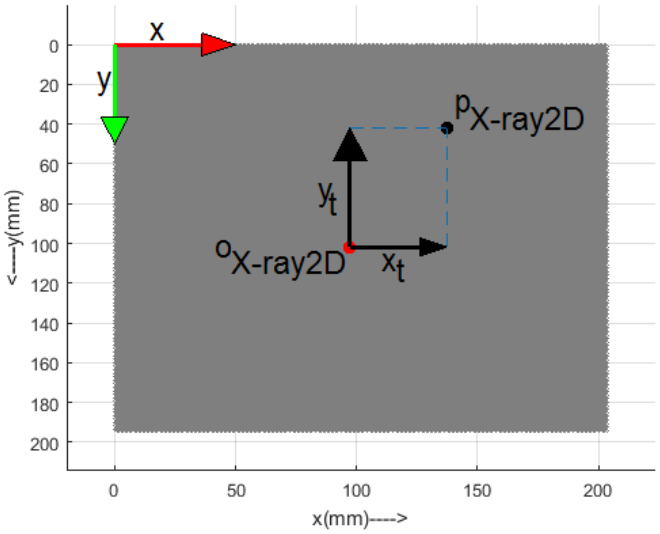


Figure 3-39: X-ray 2D Device coordinate system, with 2D translation depicted as x_t and y_t

Now that we have both points in the X-ray Device 2D system (see Fig. 3-39) we transform them to the X-ray 3D Device system and calcualte the resulting translation vector at the detector plane in 3D as follows:

$$\mathbf{o}_{\text{X-ray3D}} = T^{-1}_{\text{X-ray3D}\rightarrow\text{X-ray2D}} * \mathbf{o}_{\text{X-ray2D}}, \tag{3-44}$$

$$\mathbf{p}_{\text{X-ray3D}} = T^{-1}_{\text{X-ray3D}\rightarrow\text{X-ray2D}} * \mathbf{p}_{\text{X-ray2D}}, \tag{3-45}$$

$$\mathbf{v}_{\text{X-ray3D}} = \mathbf{p}_{\text{X-ray3D}} - \mathbf{o}_{\text{X-ray3D}}, \tag{3-46}$$

where $T_{\text{X-ray3D}\rightarrow\text{X-ray2D}}$ denotes the transformation between both systems.

The next step is to bring this translation vector back to the world coordinate system (see Fig. 3-40). Logically, the inverse of $T_{\text{world}\rightarrow\text{X-ray3D}}$ does not exist(P is non-invertible, App. A) as any point along the line between source and detector plane yields the same projected point. However, we can use the ratio of the SOD and SID distances to calculate the translation $\mathbf{v}_{xy}$ necessary at the isocenter of the world coordinate system to result in the appropriate translation $\mathbf{v}_{\text{X-ray3D}}$ on the detector plane as such:

$$\mathbf{v}_{xy} = \frac{SOD}{SID} * \mathbf{v}_{\text{X-ray3D}}. \tag{3-47}$$

Note that the label xy represents the origin of the translation vector, being the result of the x and y translation of the $pose_{2D}$.

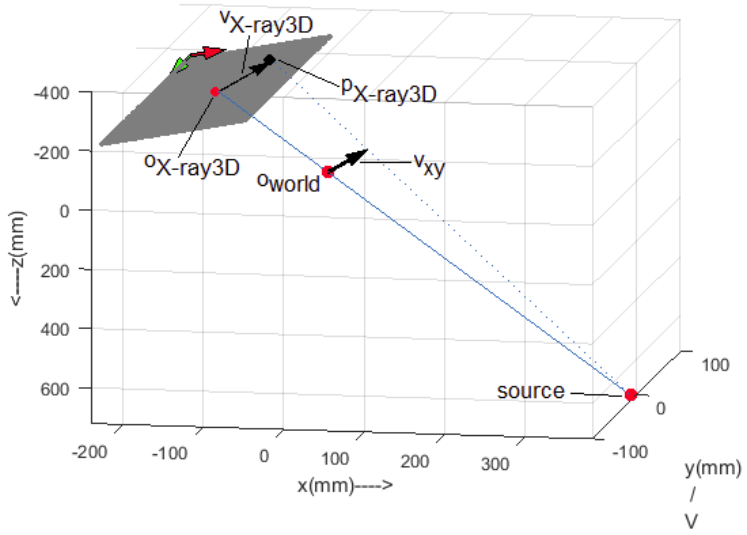


Figure 3-40: World coordinate system, with necessary 3D translation depicted due to the 2D translation parameters

Scaling As with translation, the scaling can be depicted as a 3D translation. The effect of 2D scaling is equal to a translation in the world coordinate system (see Fig. 3-41) along the normal on the detector plane. We can calculate the normal $\mathbf{n}$ on the detector plane in various ways. However, as we already know the world origin and its projection on the detector plane, we use those points to calculate $\mathbf{n}$. After calculation of the normal the calculation of the necessary translation v_{sc} is straightforward.

$$\mathbf{n} = \frac{\mathbf{o}_{\text{world}} - \mathbf{o}_{\text{X-ray3D}}}{\sqrt{\sum(\mathbf{o}_{\text{world}} - \mathbf{o}_{\text{X-ray3D}})^2}} \tag{3-48}$$

$$\mathbf{v}_{sc} = -sc * SOD * \mathbf{n} \tag{3-49}$$

Note that the label sc represents the origin of the translation vector, being the result of the scaling sc of the pose_{2D}.

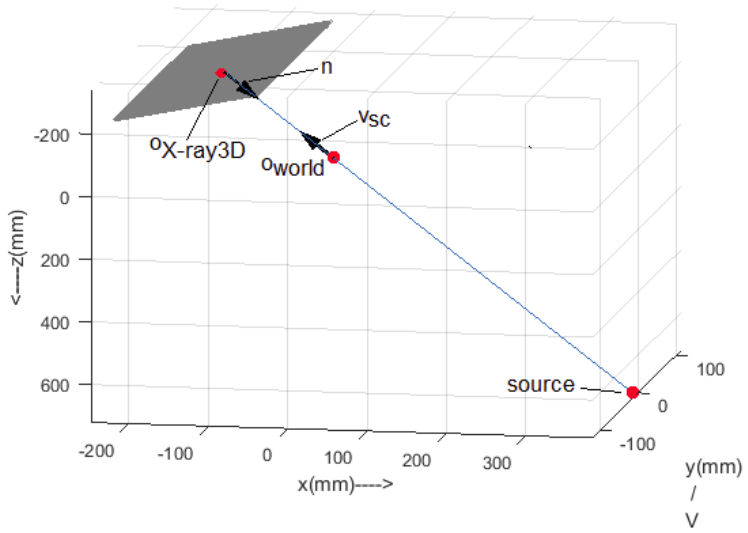


Figure 3-41: World coordinate system, with necessary 3D translation depicted due to the 2D scaling parameter

Now that we have calculated the impact of both 2D translation(x and y) and scale(sc), we are ready to form the transformation matrix $\text{Transl}_{\text{CT}}$ as:

$$\text{Transl}_{\text{CT}} = \begin{bmatrix} 1 & 0 & 0 & \mathbf{v}_{xy}(1) + \mathbf{v}_{sc}(1) \\ 0 & 1 & 0 & \mathbf{v}_{xy}(2) + \mathbf{v}_{sc}(2) \\ 0 & 0 & 1 & \mathbf{v}_{xy}(3) + \mathbf{v}_{sc}(3) \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3-50}$$

Rotation The in-plane rotation on the detector plane(see Fig. 3-42)is equal to the same rotation around the aforementioned normal on the detector plane (see Fig. 3-43). Using the normal $\mathbf{n} = [n_x \quad n_y \quad n_z]$ and the rotation around this normal (the in-plane rotation) α we can create the rotation matrix Rot_{CT} as:

$$\text{Rot}_{\text{CT}} = \begin{bmatrix} cos(\alpha) + n_x^2(1 - cos(\alpha)) & n_x n_y(1 - cos(\alpha)) - n_z sin(\alpha) & n_x n_z(1 - cos(\alpha)) + n_y sin(\alpha) \\ n_y n_z(1 - cos(\alpha)) + n_z sin(\alpha) & cos(\alpha) + n_y^2(1 - cos(\alpha)) & n_y n_z(1 - cos(\alpha)) - n_x sin(\alpha) \\ n_z n_x(1 - cos(\alpha)) - n_y sin(\alpha) & n_z n_y(1 - cos(\alpha)) + n_x sin(\alpha) & cos(\alpha) + n_z^2(1 - cos(\alpha)) \end{bmatrix} \tag{3-51}$$

Now that we have both the translation matrix $\text{Transl}_{\text{CT}}$ and rotation matrix Rot_{CT} which, applied to the CT in the world coordinate system, yields the correct pose change $pose_{2D}$ at the detector plane (defined in the X-ray 2D coordinate system), we combine them to form

$$T_{\text{CT}} = \text{Transl}_{\text{CT}} * \text{Rot}_{\text{CT}}. \tag{3-52}$$

Applying this matrix to the CT in the world coordinate system positions the CT in such a way that its projection is initially aligned with the XA at the detector plane. The following Chapter 3-3 will optimize this alignment (the transformation matrix T_{CT}).

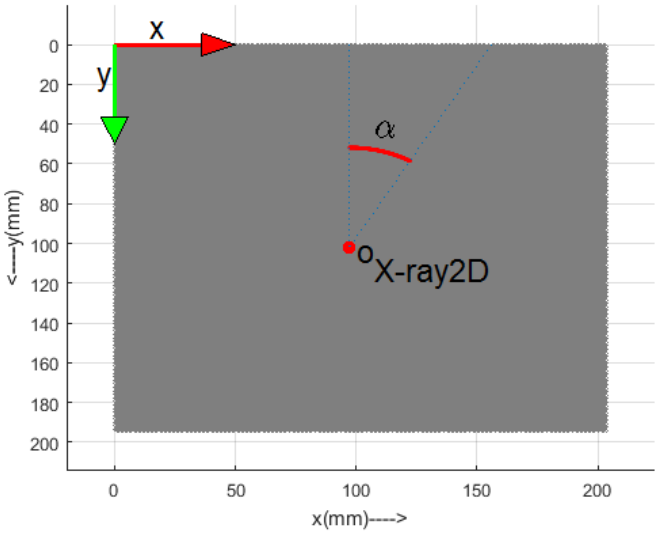


Figure 3-42: X-ray 2D Device coordinate system, with 2D rotation depicted

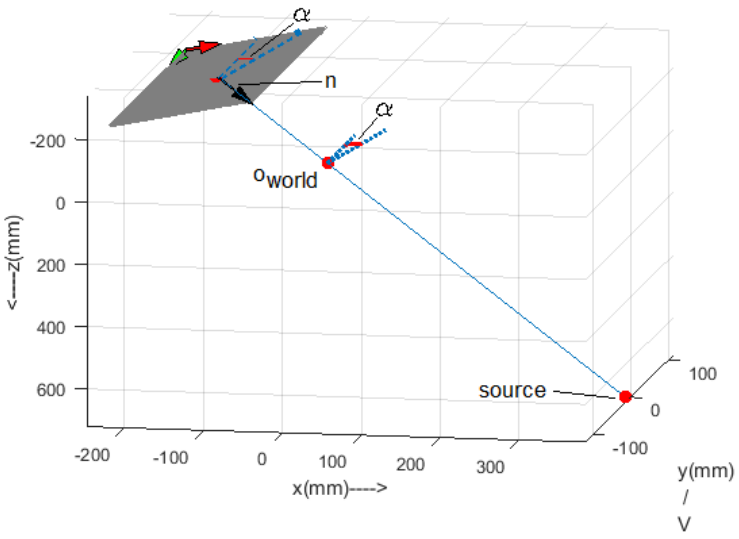
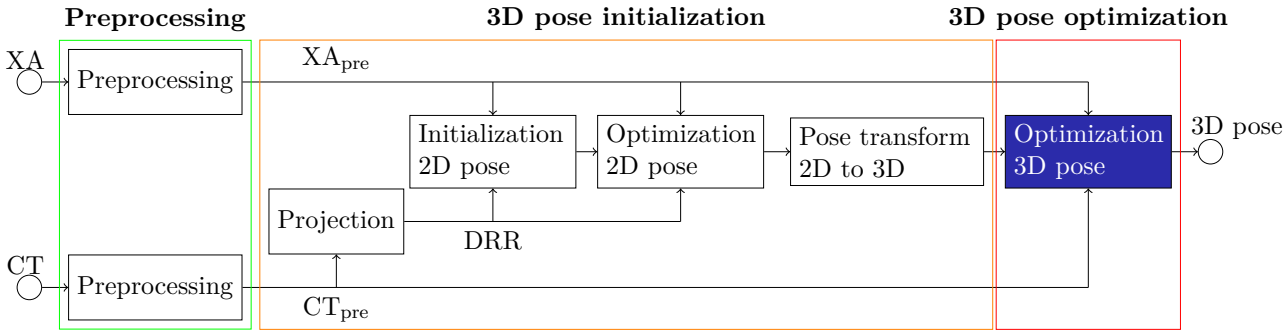


Figure 3-43: World coordinate system, with necessary 3D rotation depicted due to the 2D rotation parameter

3-3 3D pose optimization



Recall the scheme of our registration method as depicted above. At this point, we have created CT_{pre} and XA_{pre}. Subsequently we have initialized and then optimized the alignment between DRR(the projection of CT_{pre}) and XA_{pre}. In the previous section we transformed the resulting 2D pose of the DRR to a 3D pose of the CT_{pre}. During this stage we will improve this 3D pose by means of 3D/2D registration. Recall that the 3D pose consists of three translation parameters and three rotation parameters.

Now when looking at the scheme as depicted above, notice that the entire Chapter 3-2 has one sole purpose; finding a sufficiently close initial 3D pose. In Chapter 2 we mentioned that the pose initialization was the main problem of 3D/2D registration methods. Now that we have dealt with the initialization, the 3D pose optimization should converge more easily.

Similarity measure

In Sec. 3-2-2 we addressed relevant similarity measures between the XA_{pre} and the DRR. MI and NCC were selected to be used in the optimizing of the DRR pose with respect to the XA_{pre}. We again utilize these measures but now in the 3D/2D scheme. In this case during each iteration we project the CT_{pre} to obtain a DRR and measure its similarity with the XA_{pre}.
We investigate another group op measures, ones that compare the 3D CT_{pre} directly with the 2D XA_{pre}. Of this group, the method investigated in this thesis is the gradient measure as proposed by Tomazevic et. al in [6]. Before moving on to the used optimization method during this stage, Tomazevic's measure will be explained.

Tomazevic's gradient measure

This approach revolves around using gradients in the CT image and compare them with the back projections of gradients of the XA. To fully comprehend the method we start with a quick recap of the generation of X-ray images before proceeding to the mathematical description of the measure. For a more elaborate and detailed description we refer to [6].

X-ray image formation is generally modelled as follows:

$$I(\boldsymbol{p}) = I_0 * cos^3\vartheta * e^{-\int_L \mu(\boldsymbol{r})dr}. \tag{3-53}$$

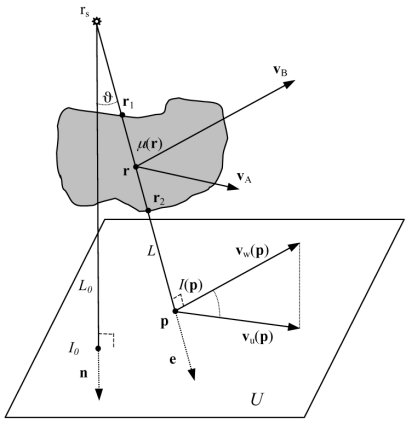


Figure 3-44: X-ray image generation, copied from [6]

Consider Eq. 3-53 and Fig. 3-44 where $\mu(\boldsymbol{r})$ denotes the X-ray attenuation coefficient, L stands for the projection beam and ϑ describes the angle between L and line L_0, which originates in the X-ray source and is perpendicular to the detector plane U. $I(\boldsymbol{p})$, the image intensity at point $\boldsymbol{p}$ is obtained by integrating the X-ray attenuation coefficient from X-ray source $\boldsymbol{r}_s$ to point $\boldsymbol{p}$. The factor $cos^3\vartheta$ describes the effect of X-ray beam divergence. Lastly I_0 is the image intensity at the point where L_0 intersects U.
A simpler model is obtained when the X-ray imaging system itself corrects for beam divergence:

$$I(\boldsymbol{p}) = I_0 e^{-\int_L \mu(\boldsymbol{r})dr}. \tag{3-54}$$

And for X-ray sensors with a logarithmic static response the model changes into

$$I(\boldsymbol{p}) = \int_L \mu(\boldsymbol{r})dr. \tag{3-55}$$

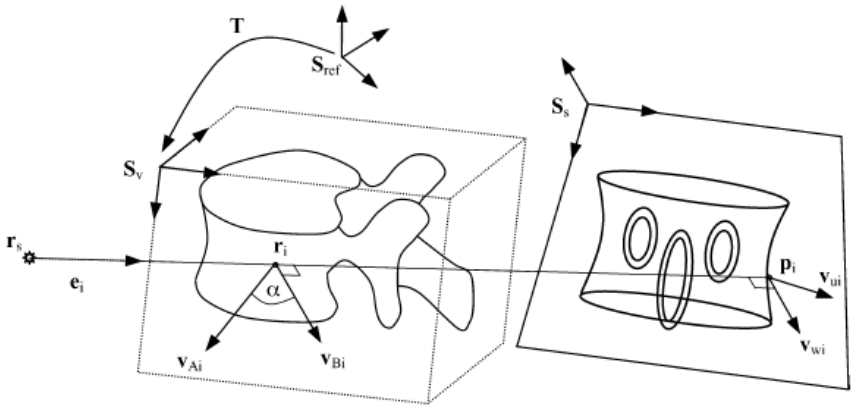


Figure 3-45: Geometrical setup with normals to surfaces and intensity gradients used in 3D/2D registration, copied from [6]

Now as to Tomazevic's gradient metric, consider Fig. 3-45 and recall that we are going to compare CT gradients in the direction of normal to bone surfaces of the CT image (our $\mathrm{CT_{pre}}$) with back projections of XA ($\mathrm{XA_{pre}}$) gradients. In order to achieve this we need to determine

the normal to bone surfaces and which gradients on the detector plane correspond to them.

Let $\boldsymbol{v}_u(\boldsymbol{p}) = \nabla_U I(\boldsymbol{p})$, be the 2D image intensity gradient in the projection plane U. And let $\boldsymbol{v}_w(\boldsymbol{p}) = \nabla_W I(\boldsymbol{p})$ be that in plane W, which is perpendicular to the projection beam L. We can define the relationship between the two as

$$\boldsymbol{v}_u(p) = (\boldsymbol{n} \times \boldsymbol{v}_w(\boldsymbol{p})) \times \boldsymbol{n} \tag{3-56}$$

in which $\boldsymbol{n}$ is the unit normal to the projection plane U.

Let $\boldsymbol{v}_A(\boldsymbol{r})$ be the gradient on the CT body, which can be divided into a component $\boldsymbol{v}_{Ar}(\boldsymbol{r})$ with the same direction as L and a component $\boldsymbol{v}_{At}(\boldsymbol{r})$ perpendicular to L as such

$$\boldsymbol{v}_A(\boldsymbol{r}) = \nabla\mu(\boldsymbol{r}) = \boldsymbol{v}_{Ar}(\boldsymbol{r}) + \boldsymbol{v}_{At}(\boldsymbol{r}) \tag{3-57}$$

The 2D image intensity gradient $\boldsymbol{v}_w(\boldsymbol{p})$ and $\boldsymbol{v}_{At}(\boldsymbol{r})$ derived from the CT gradient relate to each other in the following manner ([6] appendix C):

$$\boldsymbol{v}_w(\boldsymbol{p}) = \frac{1}{\boldsymbol{p} - \boldsymbol{r}_s} \int_L |\boldsymbol{r} - \boldsymbol{r}_s| \boldsymbol{v}_{At}(\boldsymbol{r}) dr. \tag{3-58}$$

Now that we have determined the appropriate gradients on the 2D XA image and the 3D CT surface, the last thing we need to do is back projection of the XA gradient, in order to compare it with the CT gradient. Let $\boldsymbol{v}_B(\boldsymbol{r})$ be the back projection of the gradient $\boldsymbol{v}_w(\boldsymbol{p})$ on the 2D XA image. We can compute $\boldsymbol{v}_B(\boldsymbol{r})$ as follows:

$$\boldsymbol{v}_B(\boldsymbol{r}) = \frac{|\boldsymbol{p} - \boldsymbol{r}_s|}{|\boldsymbol{r} - \boldsymbol{r}_s|} \boldsymbol{v}_w(\boldsymbol{p}) \tag{3-59}$$

The cost function as proposed by [6] to measure the correspondence between the CT's surface gradients $\boldsymbol{v}_{Ai}$ and the XA's back-projected gradients $\boldsymbol{v}_{Bi}$ is defined as

$$CF = \frac{\sum_{i=1}^{N} |\boldsymbol{v}_{Ai}||\boldsymbol{v}_{Bi}| f(\alpha)}{\sum_{i=1}^{N} |\boldsymbol{v}_{Ai}| \sum_{i=1}^{N} |\boldsymbol{v}_{Bi}|} \tag{3-60}$$

where $f(\alpha)$ is a weighting function depending on α, the angle between the gradients $\boldsymbol{v}_{Ai}$ and $\boldsymbol{v}_{Bi}$ defined as

$$f(\alpha) = \begin{cases} cos^n \alpha, & 0 < abs(\alpha) < 90 degrees \\ 0, & \text{otherwise} \end{cases}. \tag{3-61}$$

The parameter n determines the sensitivity of weighting to the angle α. It is important to note that a threshold is applied on the collection of CT gradients in order to retain only bone surface gradients.
With this approach, the metric denotes the similarity between surface gradients in the CT image and gradients of the XA image by back-projection. Adding Tomazevic's metric as gradient based similarity measure to our exisitng set of intensity based measures (NCC and MI) yields us three similarity measures to test during the experiments. The result of which will determine which measure will be selected for the 3D pose optimization stage of the method.

Optimization method

As for the same reasons as in Sec. 3-2-2 we will again utilize Powell's method in optimizing the similarity measure.

Example

Recall the test case as defined by XA_{pre} and DRR in Fig. 3-27 and 3-28 respectively. Applying the 3D pose optimization stage of this chapter (using NCC) onto the test case yields us the result as shown in Fig. 3-46 to 3-49. In Fig. 3-46 and 3-47 we see an overlay of the DRR with the XA_{pre} before and after the optimization of the 3D pose. In Fig. 3-37 and 3-38 the results are again shown, however in these figures manual annotations on the ribs and vertebrae are also plotted to clarify the improvement in alignment.

We notice after the optimization of the 3D pose the vertebrae are better aligned, also the alignment among the ribs seems to be improved.

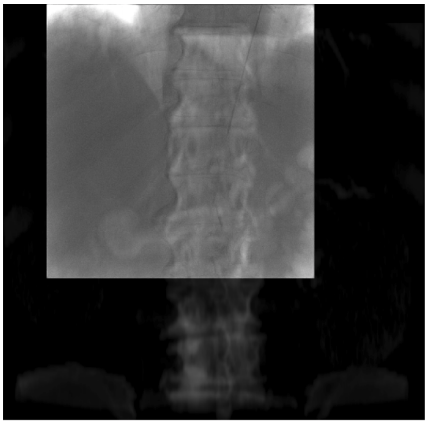


Figure 3-46: Result before 3D pose optimization

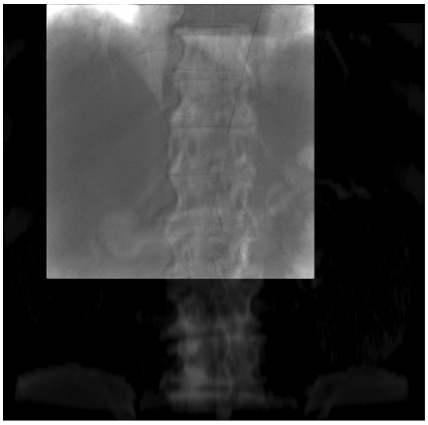


Figure 3-47: Result after 3D pose optimization

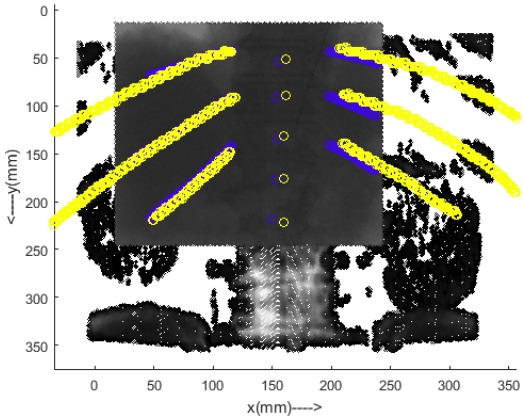


Figure 3-48: Result before 3D pose optimization

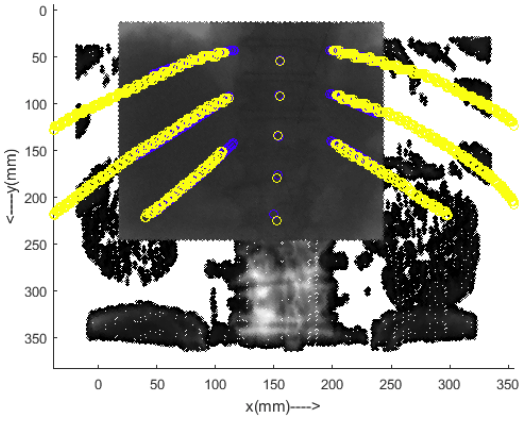


Figure 3-49: Result after 3D pose optimization

Chapter 4

Experiments

In this chapter we describe the experiments for the proposed 3D/2D registration method, the results of which are depicted in Chapter 5 and discussed in Chapter 6. This chapter is organized as follows. First, we clarify the dataset used in the evaluation of the proposed method. Then, the evaluation metrics are discussed. Afterwards, the outline of the coming experiments is described. Finally, the parameters and their values used during the experiments are listed.

4-1 Dataset

The dataset involved in the experiments consisted of data acquired during 19 TACE procedures. Every procedure contained one CT image and between 2-3 XA sequences, adding up to a total of 60 XA sequences and 19 CT images in the dataset. Of those XA sequences we excluded those in which there was X-ray system motion, the imaged region was too small or outside the range of the corresponding CT image. After applying the exclusion criteria all 19 sets contained at least one CT image and one XA sequence. In case of multiple XA sequences per patient one was selected at random and was used during the experiments. Of the dataset, two interventions were taken as the training set, while the remaining 17 interventions comprised the test set.

4-2 Evaluation criteria

All interventions (set of one CT and one XA sequence) were annotated by the author of this thesis. The annotations include points at the center of each vertebra, as well as points along the centerline of the ribs. During the latter use was made of a centerline tracking method (based on Graph Cuts and Robut Kernel Regression [15]) initialized with manual annotations. Although great care was taken in the annotation process, it should be noted that multiple annotators presumably make the evaluation more robust to annotation errors.

Two metrics were created based on the vertebra and rib annotations. The first of which denotes the distance between the vertebrae of the CT projection and the XA image. Here, the average distance between corresponding vertebra in the CT and XA was calculated. When a vertebra was not present on either the CT or XA, its annotation is not taken into account in this metric. It should be noted that as this metric is based on relatively few annotations (one per vertebra), any error in annotation greatly influences the metric. In the case of the second metric many more points are taken into consideration, diminishing the effect of a single incorrect annotation.

The second metric denotes the distance between corresponding ribs. The value is calculated as the average distance between all corresponding sets of ribs. When the rib of the CT's projection extended beyond the XA image, only the overlapping segments are taken into account.

The procedure of selecting the overlapping rib segments of the XA image and the CT's projection, as well as calculating the average rib distance is as follows. The distance between two ribs (two sets of points) is calculated by adding dummy points at the end of the two point sets, see Fig. 4-1. Subsequently for every point(except the dummy end-points) the distance to the closest point in the other set(including the dummy end-points) is calculated, denoted by a line in the figure. In case this closest point is the dummy end-point of the other set we exclude this point, all other points are retained. The distance between the two ribs is then calculated as the average value of all computed point distances. In the figure this corresponds to the average of the lengths of all lines not connected to a dummy end-point. This process is repeated for each pair of corresponding ribs in the XA image and CT's projection. Finally, the average of these rib distances is taken as the value of the rib evaluation metric.

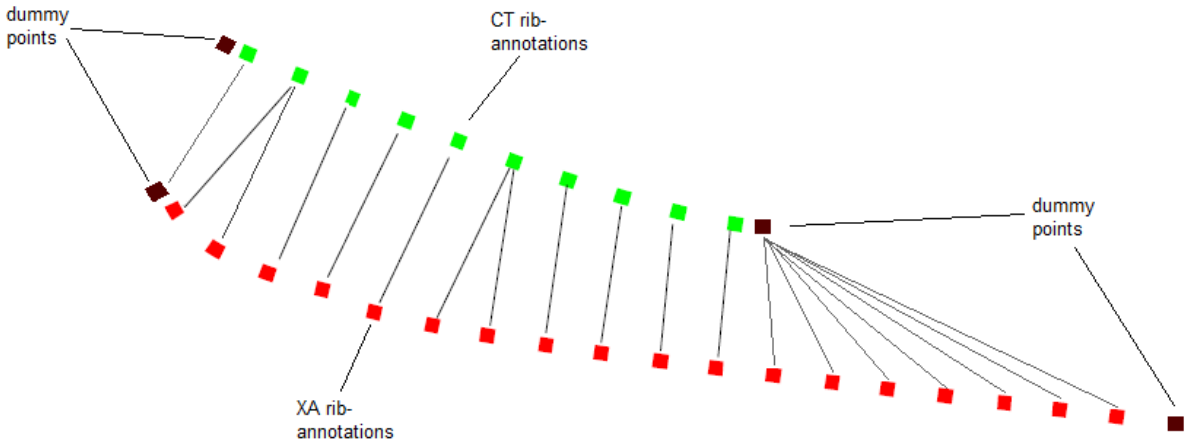


Figure 4-1: Calculation of the rib metric

4-3 Experiments outline

In this section we outline the experiments of which the results are shown in Chapter 5. We investigate the performance of the proposed 3D/2D registration method under a variety of configurations and in the following order:

1. Preprocessing
2. 2D initialization
3. 2D optimization
4. 3D optimization

Note that the 3D initialization stage is divided into its two main components, the 2D initialization- and optimization stages. With the exception of the pre-processing stage, all stages have multiple possible parameter settings(configurations). We thus investigate multiple configurations per stage.
Finally, the results of two registrations are depicted in two images. Both images contain a checkerboard overlay of a CT projection (DRR) with its corresponding XA. The evaluation metric values are stated in the images' caption. These images are included to visualize the relationship between the metric values and visual alignment.

4-4 Parameters

In this section we list the method parameters. Parameters of which their value has been fixed earlier, such as those based on human anatomy (e.g. the 300 HU CT threshold based on bone density, as well as the spine width), are stated as well as empirically deduced parameters (such as the step size during the 2D- and 3D pose optimization stages).
Table. 4-1 depicts the parameter settings of the method during the experiments.

Stage	Parameter	Value	Unit	Chapter
CT pre-processing				
Threshold	threshold	300	HU	3-1-1
Mask radius	horizontal	100	mm	3-1-1
	vertical	60	mm	
Aorta detection [16]				3-1-1
	back-foreground	50	%	
	aorta diameter	23	mm	
	σ_s	1.4		
	μ_l	180	degrees	
	σ_l	30	degrees	
	a_r	2		
	b_r	1		
	t_{aorta}	1e-5		
Aorta removal	$t_{regiongrowing}$	0.015	%	3-1-1
	dilation	5	mm	
XA preprocessing				
	nFrames	4		3-1-2
	mask radius hor	100	mm	3-1-2
2D Pose Initialization				
x	region radius hor	0.25	mm/sw	3-2-1
	spinewidth(sw)	40	mm	
α	patch height	1-20	mm	3-2-1
y	offset	60	mm	3-2-1
2D Pose Optimization				3-2-2
MI	binsize	16		
Powell	alpha	0.9		
stepsize	translation	1	mm	
	rotation	1	degree	
	scale	0.001		
boundaries	translation	50	mm	
	rotation	45	degrees	
stopping criteria	translation	0.001	mm	
	rotation	0.001	degree	
	scale	0.001		
3D Pose optimization				3-3
Tomazevic	σ	0.5		
	threshold	0.6		
	n	4		
stepsize	translation	0.1	mm	
	rotation	0.1	degree	
boundaries	translation	50	mm	
	rotation	45	degrees	
stopping criteria	translation	0.01		
	rotation	0.01		

Table 4-1: Parameter settings of the registration method during experiments

Chapter 5

Results

In this chapter the results of the experiments are depicted. The order of the experiments is as follows.

1. Preprocessing
2. 2D initialization
3. 2D optimization
4. 3D optimization

Recall the evaluation metrics introduced in Chapter 4. For every configuration two distributions are plotted in the form of box plots. The first box plot(blue) denotes the values of the vertebra metric (the average distance between annotated vertebrae in the CTs projection and the XA) while the second box plot(green) denotes the values of the rib metric (the average distance between annotated ribs) for all successful interventions. We thus aspire to low values of the evaluation metrics. Each box plot includes the median, the 25th percentile (Q_1) and the 75th percentile (Q_3), and extends from $Q_1 - w(Q_3 - Q_1)$ to $Q_3 + w(Q_3 - Q_1)$. Using a value for w of 1.5 this translates into a coverage of 99.3%.
Of the in total 19 interventions in the test set, the registration method failed in two cases. We classify a registration as a failure when the 2D initialization stage is not able to bring the DRR within the capture range of the subsequent optimization stages (which translates to a pose in which corresponding vertebra overlap most). Due to the inaccurate initial pose estimation, the 2D- and 3D optimization stages were not able to adequately align the DRR with the XA. Chapter 6 will discuss these failed registrations further. Note that due to the failure, the depicted results in this chapter are based on the 15 cases in which the registration method succeeded to position the DRR within the aforementioned capture range.

Pre-processing As all parameters in this stage are based on human anatomy, we have but one possible configuration. The result of the pre-processing stage is labelled as *initial* and is depicted in Fig. 5-1.

2D pose initialization During the 2D pose initialization step four configurations were evaluated. In the first configuration, only the translation is estimated. During the second, translation and orientation are estimated. The third configuration involves estimating translation and scale and the last configuration estimates all four variables. In Fig. 5-2 the results of the configurations are displayed. Additionally, in Fig. 5-3 the change with respect to the previous pre-processing stage is depicted. Here, a negative value is favourable as it represents a decrease in evaluation metric value.
As will be discussed in Chapter 6, the first configuration (estimating only translation) exhibits the most favourable performance, and will thus be utilized in the subsequent experiments.

2D pose optimization The last step of the 3D pose initialization stage is the 2D pose optimization step. Here, two different configurations are evaluated, using either the NCC or MI as similarity measure. Note that the optimization parameters(e.g. step size and stopping criteria) were empirically determined. In Fig. 5-4 and 5-5 the results of the 2D pose optimization step and thus the entire 3D pose initialization stage are displayed.
As will be discussed in Chapter 6, the first configuration (using NCC as similarity measure) exhibits the most favourable performance, and will thus be utilized in the subsequent experiments.

3D pose optimization The last step of our 3D/2D registration method is the 3D pose optimization stage. Here, we evaluated three different configurations, using either the normalized cross correlation, the mutual information or the gradient-based metric as proposed by Tomazevic as our similarity measure. Note that similar to the previous step, the optimization parameters(e.g. step size and stopping criteria) were empirically determined. In Fig. 5-6 and 5-7 the results of the 3D pose optimization step and thus the entire registration method are displayed.

Registration examples Finally in Fig. 5-8 and 5-9 the results of two registrations are shown. Both the DRR (background) and XA (foreground) are shown, the evaluation metric values are depicted as well. The images are shown in order to visualize the relationship between the metrics and the alignment.

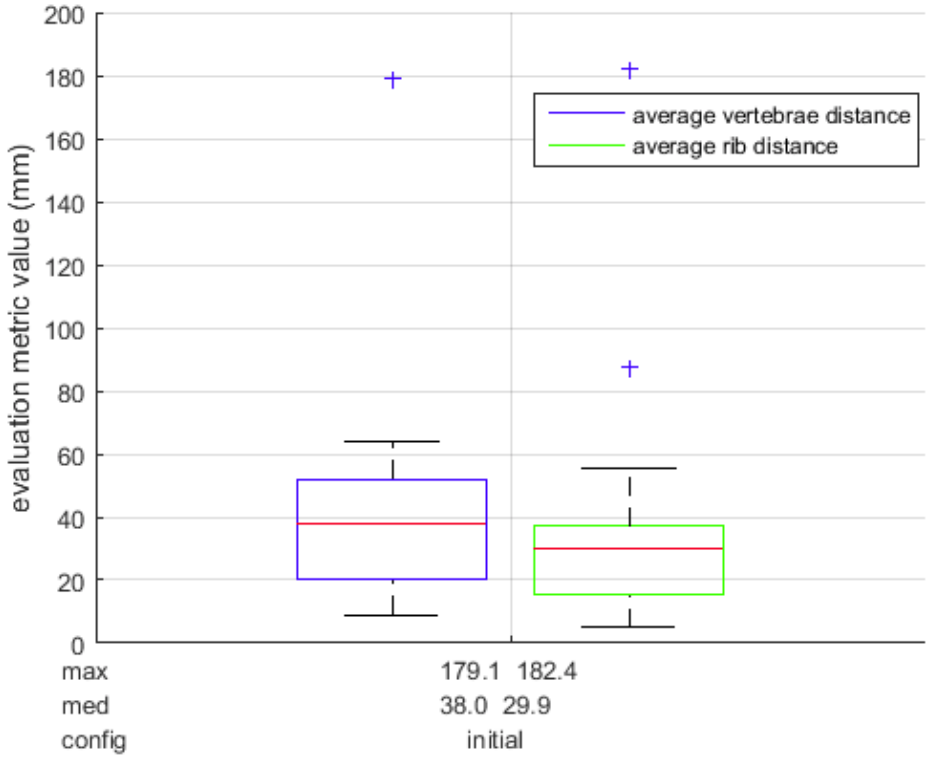


Figure 5-1: results of the pre-processing stage

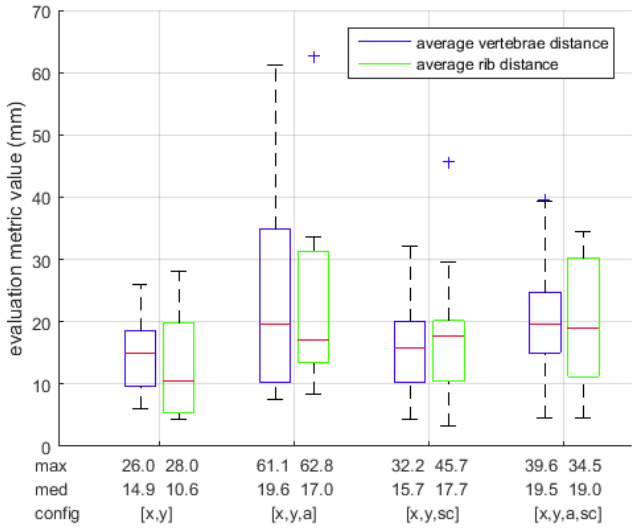


Figure 5-2: results of the 2D initialization stage

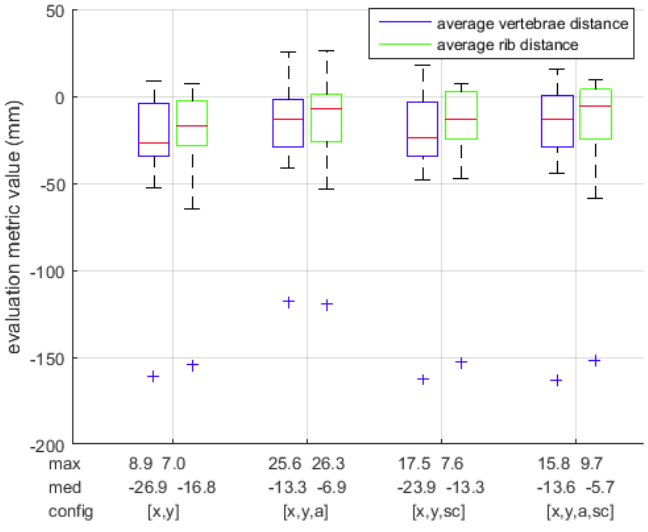


Figure 5-3: difference between the results of the 2D initialization stage and pre-processing stage

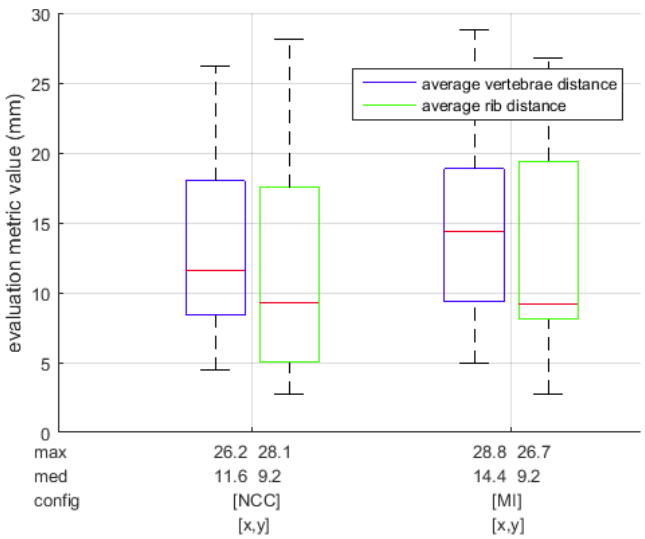


Figure 5-4: results of the 2D optimization stage

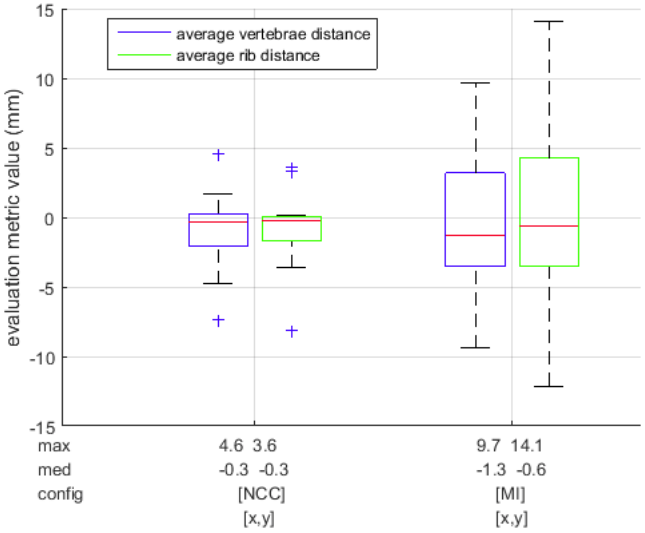


Figure 5-5: difference between the results of the 2D- optimization and initialization stages

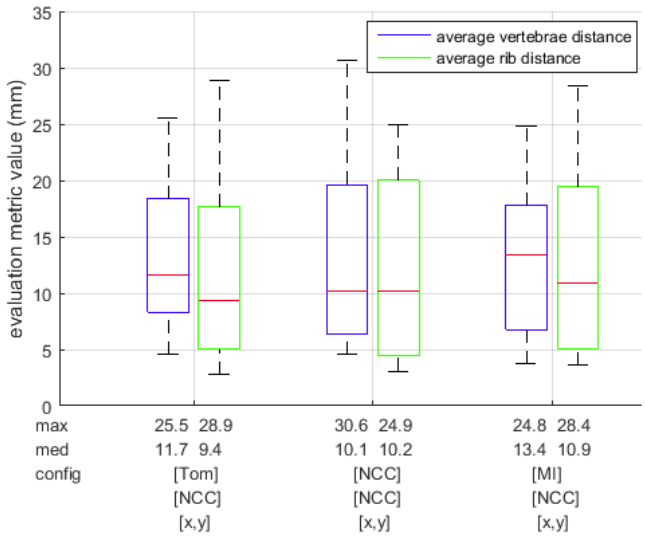


Figure 5-6: results of the 3D optimization stage

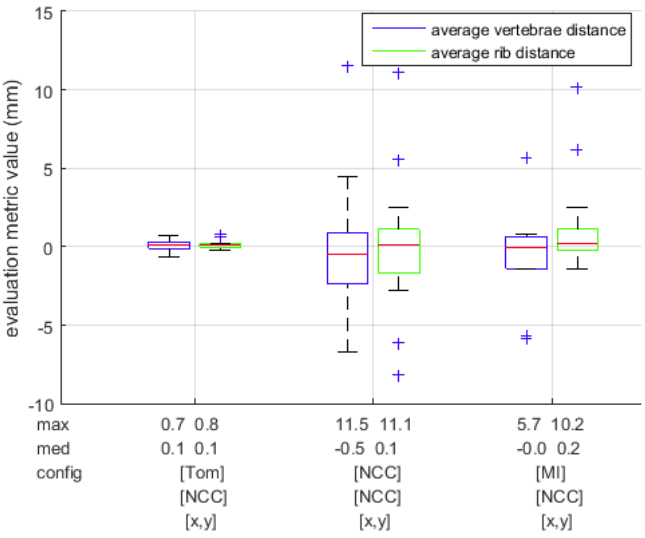


Figure 5-7: difference between the results of the 3D- and 2D optimization stages

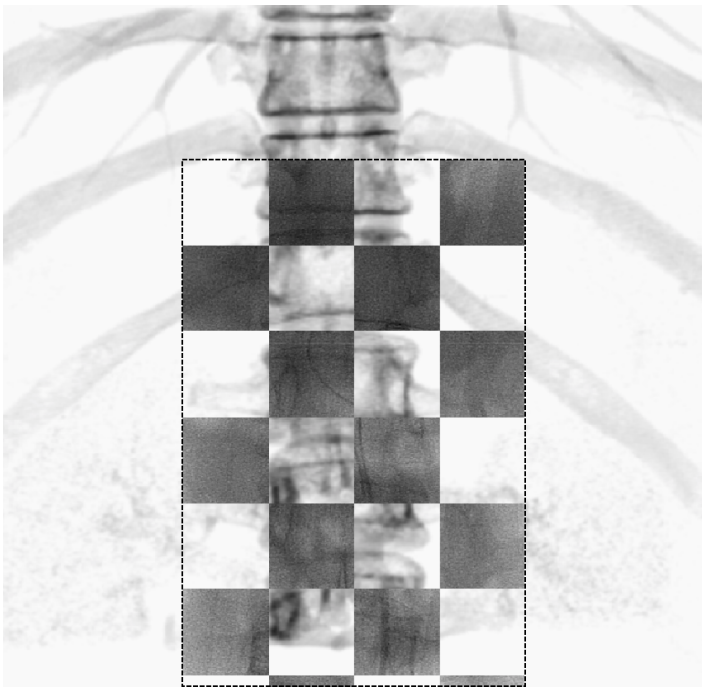


Figure 5-8: Registration example 1, vertebrae- and rib metric values are respectively 8.1 and 19.5 mm

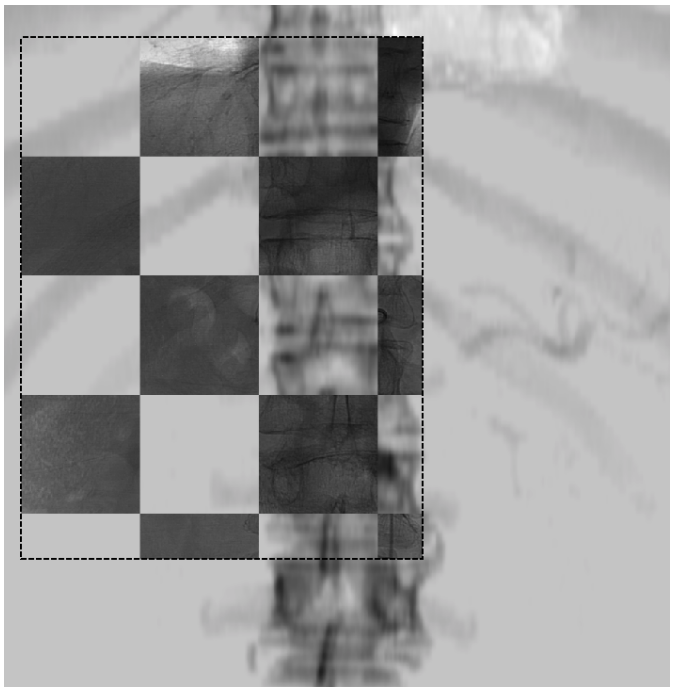


Figure 5-9: Registration example 2, vertebrae- and rib metric values are respectively 7.6 and 11.7 mm

Chapter 6

Discussion

In this chapter, the results as depicted in Chapter 5 are discussed, the same order is maintained.

Pre-processing

Recall the purpose of the pre-processing stage, filtering of the images in order to boost performance during subsequent pose estimation and optimization. The impact of this stage can not yet be observed in the results seen in Fig. 5-2 and 5-3. The reason for this is that there is no action undertaken during the pre-processing which directly influences the alignment at this time. Succesful pre-processing will positively influence the performance of all subsequent steps, utilizing the pre-processed images.
Noteworthy from the results depicted in Fig. 5-2 and 5-3 are the outliers in both metrics. Note the median of the vertebra- and rib metric being respectively 38.0 and 29.9 mm. The outliers in both metrics (around 180 mm) correspond to the same intervention. The CT image belonging to this outlier spans an unusually large region. The image ranged from the pelvis up to vertebra $T8$. The average CT image in the dataset ranged from vertebra $T10$ to $L4$. During the pre-processing stage the CT image's center is positioned at the isocenter of the world coordinate system. On average the CT vertebra $L1$ and the XA vertebra $L2$ was found at that position, in case of the outlier CT vertebra $L5$ was located at the isocenter, which explains the unusually large initial metric values.

2D pose initialization

Regarding the results as depicted in Fig. 5-2 and 5-3, we notice a substantial improvement with respect to the pre-processing results. The largest improvement is found in the result of the first configuration (estimating only the translation), which brings down the evaluation metric values from medians of 38.0 mm (vertebra metric) and 29.9 mm (rib metric) to respectively 14.9 and 10.6 mm. Visual inspection supports the metric values, it revealed a robust performance of the spine edge detection (utilized in estimation of $\hat{x}$), which also positively influenced the estimation of $\hat{y}$. This estimation of $\hat{y}$ proved sufficient due to amongst others

its redundancy as a consequence of returning multiple possible poses. However, as mentioned, the vertical translation estimation failed in 2 out of 19 cases, which shows the estimation being less robust than preffered. On closer inspection of the mentioned two cases, both exhibited in-plane rotations (of respectively 6 and 12 degrees). In estimating the vertical translation, the normalized cross correlation is measured for various vertical shifts. This approach inherently assumes an equal orientation in both images. When there is a orientation difference, the approach becomes less robust, as further proven by the two failed registrations.
The estimation of α did not improve the results, upon closer investigation it appeared that orientation was almost always detected however often mis-estimated by a few degrees in the case of the XA images. The estimation of the orientation of DRR images, inhabiting much less 'noise' proved much more accurate. The mis-estimation of XA image orientation also yielded a negative effect on the vertical translation estimation later on in the process, measuring normalized cross correlation between two images orientated differently.
The estimation of scale seemed to slightly decrease performance. Upon visual inspection it is revealed that the largely equal approach to detect the spine edges in both modalities is not ideal. Detecting the exact same spine edges in both modalities deserves two different approaches. The estimation of $\hat{x}$ suffers much less from the spine edges detected not being 100% accurate. In the scale estimation case, the error in the four edges detected adds up to a small but noticeable decrease in accuracy. It should be noted that the efforts with respect to scale estimation might not be required. In fact closer visual inspection of the intervention images showed no large scale differences between the images. Orientation differences were detected more often.
Considering the results and aforementioned observations, configuration one (estimating solely the translation) will be used in following stages.
The benefit of this 2D pose initialization is clearly visible from the results, however we also notice that for a number of patients the alignment has actually gotten worse, even in case of configuration one(estimating only the translation). On closer visual inspection of these cases, a decrease in quality of alignment occurred only in the instances for which the original alignment was, by chance, quite good. In these cases we clearly notice the limited capability of the initialization step, being able to bring the DRR within capture range, but not optimizing the alignment.

2D pose optimization

From the results we deduce that the first configuration (optimizing by using NCC as the similarity measure) exhibits a slightly more favourable performance. The median values of the metric here are 11.6(vertebra metric) and 9.2 mm(rib metric) which are smaller or equal to the result of using the MI similarity measure, of which the median values of the metrics are 14.4 and 9.2 mm respectively. The difference between the 75% and 99.3% percentiles between the two configurations, as depicted in the figures, is not much. However regarding the lower median of the vertebra metric in configuration one, as well as lower 25% and 75% percentiles, this configuration is determined to be the best candidate to use in the experiments of the following stage. Fig. 5-5 adds some insights into our comparison. Here, it is clear that the NCC approach has a smaller standard deviation than the MI approach. A smaller standard deviation is a sign of a more robust step, which counts heavily in the decision to accept the NCC as our similarity measure.

3D pose optimization

As opposed to the previous stages and steps we do not notice an obvious improvement. The vertebra metric medians have changed from 11.6 mm in the previous stage's best configuration(NCC) to 11.7, 10.1 and 13.4 mm by respectively utilizing Tomazevic's metric, the normalized cross correlation and the mutual information in the current stage. The medians of the rib metric have similarly changed from 9.2 mm (2D optimization, NCC) to 9.4(Tomazevic), 10.2(NCC) and 10.9(MI).
Looking at Fig. 5-7, where the difference with the previous stage is depicted, we notice that the gradient-based approach has a negligible effect. Upon closer investigation it seems the alignment from the previous stage was outside the capture range of our implementation of Tomazevic's gradient approach. It should be noted that the result of Tomazevic's paper [6](Fig. 5) show a 50% successful registration when original displacement is less or equal to about 10 mm(and 90% when original displacement is under 6 mm). Looking at Fig. 5-4, the configuration using NCC as similarity measure, the larger portion of our test set result lie above this boundary. This supports the possible explanation, in that the results of the previous stage lie outside the capture range of Tomazevic's gradient based measure, which would explain the lack of improvement in the results.
The intensity-based configurations on the other hand, show different results. Both exhibit a noticeable effect with respect to the previous stage' performance. In case of the second configuration (utilizing the NCC as similarity measure), the vertebra median is reduced from 11.6 to 10.1 mm, while the median of the rib metric increases from 9.2 to 10.2 mm. However, the deviation of the metric results has increased (note the 25% and 75% percentiles). This, to a slightly lesser degree, is also the case in configuration three(MI). Note that the number of interventions undergoing a positive effect(8) from the configuration utilizing MI are almost equal to the interventions undergoing a negative effect(9). The same results were found in case of configuration two(NCC), however in this case the net effect is a higher increase in performance with respect to configuration three(MI). It is clear that the intensity-based measures, while influencing results, are not well suited to substantially increase the performance of the overall method. Note that the results when applying the three configurations on the result of 2D pose optimization while using MI as similarity measure as well, are depicted in Fig. C-1.
Unfortunately it seems the tested 3D pose optimization technique failed in improving the result of the previous stage. Therefore, based on the experiments, this stage should be excluded from the overall method at this point.

Relationship between evaluation metric and alignment

Note Fig. 5-8 and Fig. 5-9. In the images two registration results are shown. Interesting about the first case is the value of the rib metric. Although the rib alignment looks quite good(note the rib connected to vertebra T12), the metric denotes a value of nearly 2 cm. This artefact is a consequence of the relatively small part of the ribs visible on this particular XA, allowing for any error during the annotation process to have a larger effect on the metric value. As for the second example case, note that there is a small rotation necessary in order to achieve the optimal alignment, which influences both evaluation metrics.
Another important deduction can be made with respect to the evaluation itself. An ideal evaluation metric perfectly reflects the method performance. However, even though based on manual annotations, the metrics utilized in this thesis do not represent the alignment per-

fectly. As an example, a low vertebra metric value could accompany an incorrect alignment, due to an out-of-plane rotation. This also applies to the rib metric, in which the lack of point-correspondence causes the mis-representation. Furthermore, inherently two evaluation metrics can cause ambiguity during investigation as opposed to a single evaluation metric case. These effects make for a required visual inspection on all numerical results, which can not be used in comparing the proposed method against other registration methods.
The previous deduction as well as the depicted images show that the evaluation metric values are not necessarily an entirely accurate representative of the alignment. Amongst others, man-made annotations can contain errors, which can influence the metric value substantially(example 1).

In summary, based on the numerical results as well as visual verification of these results, it seems that the 3D pose initialization succeeded in providing a seed point sufficiently close to the optimal pose for subsequent optimization stages. However, the automatic estimation of the orientation and scale as part of the initialization stage have not increased performance as expected. The approach of estimating orientation proved not to be sufficiently robust, primarily in case of XA images. The performance of the scale estimation showed the need of a separate spine edge detection approach for both modalities.
The 2D pose optimization stage utilizing the normalized cross correlation, an intensity-based similarity measure, increased the accuracy in most tested interventions. Some intervention alignments however, were reduced in accuracy as a result, revealing the limitations of the metric. Unfortunately, the tested 3D pose optimization approaches did not succeed in bringing down the alignment error.
In conclusion, the 3D/2D registration method exhibits the best performance when estimating the horizontal- and vertical translation during the 2D pose initialization stage and using NCC as similarity measure during the 2D pose optimization stages. The results of using this configuration are depicted in Fig. 5-4, with a median vertebrae- and rib metric value of respectively 11.6 and 9.2 mm.

Chapter 7

Recommendations

During the creation of the proposed 3D/2D registration method and the discussion of the experiments, potential future work topics became clear. In this chapter, recommendations are made to further improve the performance of the proposed 3D/2D registration method as well as its applicability in future intervention support. The chapter presents the recommendations per stage, adhering to the same order as Chapters 5 and 6.

Pre-processing

As for the Pre-processing stage the most improvement in performance can be achieved in the XA pre-processing, especially in the removal of the dynamic 'noise' layer still present in the image. The challenge here lies in the design of a method with sufficient performance and reasonable execution time. As for the CT pre-processing, room for improvement lies primarily in a more robust removal of the aorta as well as other structures of high radio density which survived the threshold. An interesting approach can be model based. When detecting the spine and ribs in this manner, a more appropriate mask could be applied to these structures resulting in a more efficient removal of unwanted structures.

2D pose initialization

The just mentioned model based approach could provide a potential improvement during this stage as well. The approach might make for an estimation of orientation and scale with positive impact on overall accuracy. Which, in turn, might make the usage of subsequent approaches with a smaller capture range (such as utilizing Tomazevic's gradient measure) available. It would supposedly also solve the issue with the vertical translation, being the output of multiple poses. The effects of which is that human interaction is required, as well as an increase in computational complexity (although whether this is an issue resides on the available hardware). Another approach in bringing down the number of optimized poses after the estimation of vertical translation is to introduce a metric at that point. This metric could exist of multiple similarity measures. With this, only one pose will be optimized when the

probability of that pose being the correct one far exceeds all others. Multiple poses would still form the output if even the most likely poses inhabit an insufficient level of probability.

2D pose optimization

The optimization of the 2D pose was performed using two intensity based measures, however a gradient based measure might improve the result of this stage. Specifically in case of the X-ray fluoroscopy images, the intensity distribution is dependent on a number of factors which ideally should have no effect. Amongst them is the time at which the frames are extracted, which influences the image due to respiratory movement, visible as a region (moving over time) of low intensity. Furthermore as we are dealing with a single source and detector plane configuration, the X-ray system orientation influences the intensity values, i.e. less X-ray photons will be received when crossing the human body at the center compared to the side. The X-ray image values are based on this sole configuration, as opposed to the CT image, which incorporates multiple device angles. Gradient based measures might prove more robust to these phenomena.

3D pose optimization

The optimization of the 3D pose, although not improving the previous stage, shows interesting result. The difference between the gradient based metric and the intensity-based metrics is clear. It seems that while the intensity-based configurations effect the alignments, they are not able to further improve on the alignments accuracy by much. In case of the gradient based measure however, there is no significant effect at all, as the interventions seem to lie outside the methods capture range. Here lies room for improvement, attempting to either increase the capture range, or increase the performance of the 3D pose initialization stage (by e.g. the aforementioned model based approaches). As mentioned in Chapter 2 Livyatan et. al proposed projecting relevant gradients of the 3D volume towards the detector plane and measuring the similarity between the projections and the gradient of the X-ray images [10]. A next step would be to implement this gradient based method, noting whether the results support our deductions and/or improve our results.

Also I would like to discuss one final other interesting field, which requires labelled data to be available. Implementing convolutional neural networks to detect the spine and ribs in both modalities could make for an interesting approach. Features developed in this thesis can be used as additional information to aid the learning process. Also, pre-processed images might make for a steeper learning curve, requiring less labelled data to be input into the system.

Lastly, recall the debate on the evaluation metrics itself in Chapter 6. Here, the drawbacks of the metrics were discussed. The lack of point correspondence in the rib evaluation metric inevitably causes mis-representations of alignments. On the other hand, the vertebrae metric is ill-fitted in accurately representing out-of-plane rotations. Evaluation without a well-defined reference has caused difficulties during the experiments. For future assessment of alignment accuracy, I would suggest either increasing the amount of corresponding annotations or a curve based approach. The former will diminish negative effects inherent to the rib evaluation metric as well as those caused by human error during annotation. The latter, while introducing a more extensive annotation process, has a non-trivial potential. Using a region-growing or largest connected component approach, the spine and rib surfaces could be extracted in the

CT image. The XA modality might require a more rigorous pre-processing before the same technique can be applied. Robust principal component analysis could be applied to filter out the dynamic layer from the static one. Comparing the distance between the surface of the spine and ribs in the CT as compared to the XA can prove a single more robust metric.

Chapter 8

Conclusion

In this thesis a 3D/2D registration method is proposed to support interventions with a region of interest close to the spine, such as the TACE intervention. The method assumes an available pre-interventional 3D CTA and intra-interventional X-ray fluoroscopy. The method's novelty resides in the approach taken to initialize the 3D pose of the CT. Performance of 3D/2D registration methods heavily rely on initialization which is executed automatically in the proposed scheme.
The clinical relevance of this registration is the ability to show the aorta and relevant vessels after registration, making an enhanced roadmap available during interventions demanding guidance of a catheter through this region.
The method can be divided in three stages: a pre-processing stage, a 3D pose initialization stage and a 3D pose optimization stage. The 3D pose initialization stage itself consists of 2D pose initialization/optimization and transformation of the 2D pose to a 3D equivalent. Important to note is that human interaction is necessary. As part of the proposed method it is very probable (due to the estimation of vertical translation during the 2D pose initialization) that the registration results in multiple possible alignments. The physician is required to select the correct alignment him- or herself.
The experiments performed in this thesis resulted in a classified successful registration of 15 out of 17 interventions in the test set. The successful registrations exhibit evaluation metrics with a median of 11.6 mm between corresponding vertebrae and a median of 9.2 mm between corresponding ribs. Based on the experiment results, the highest performance was achieved by estimating translation during 2D pose initialization and using NCC as similarity measure in the subsequent 2D pose optimization stage. At this point, experiments performed with respect to the 3D pose optimization stage did not improve the result of the previous 2D pose optimization stage. This stage should thus be excluded from the registration method.

Appendix A

Coordinate Systems

There are multiple coordinate systems involved throughout the project, in this chapter we address all systems and their relations.

CT coordinate system The CT system has three dimensions denoting the following directions with respect to the human body: from front to back(anterior to posterior), from left to right, from feet to head. The specific directions and order of the dimensions were chosen as the used data exhibited this orientation system. A point CT(r, c, s) denotes a voxel at row r(position between front and back), column c(position between left to right) and slice s(position between feet to head).
The CT coordinate system can be observed in Fig. A-1. Important to note is the location of the origin $(0, 0, 0)$ in the figure, which is not located at the center of the 3D volume.

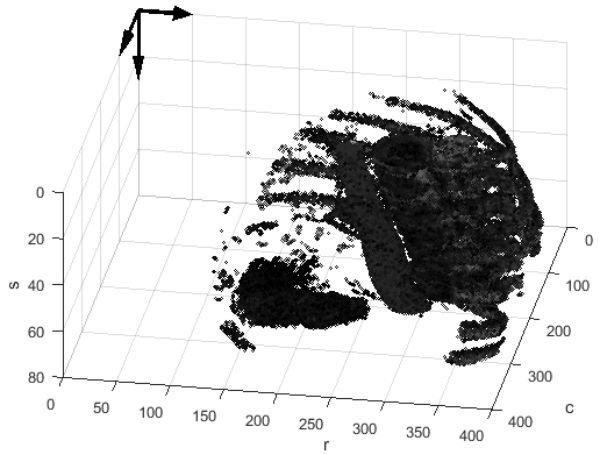


Figure A-1: CT coordinate system

CT-mm coordinate system As of yet in the CT system, the distance between subsequent rows, columns or frames is just one unit. Using the voxel spacing we compose the scaling

matrix necessary to travel to a correct mm representation as stated in Eq. A-1. Note that in most cases the row- and column voxel spacings are equal (often < 1 mm) and the frame spacing is somewhat larger (often 3 or 4 mm). In performing this transformation $\mathrm{CT}(r, c, s)$ thus becomes $\mathrm{CT}(x, y, z)$ by using

$$S_{\mathrm{CT}\rightarrow\mathrm{CTmm}} = \begin{bmatrix} \Delta x_{\mathrm{CT}} & 0 & 0 & 0 \\ 0 & \Delta y_{\mathrm{CT}} & 0 & 0 \\ 0 & 0 & \Delta z_{\mathrm{CT}} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{A-1}$$

where Δx_{CT}, Δy_{CT} and Δz_{CT} are the voxel sizes in mm.

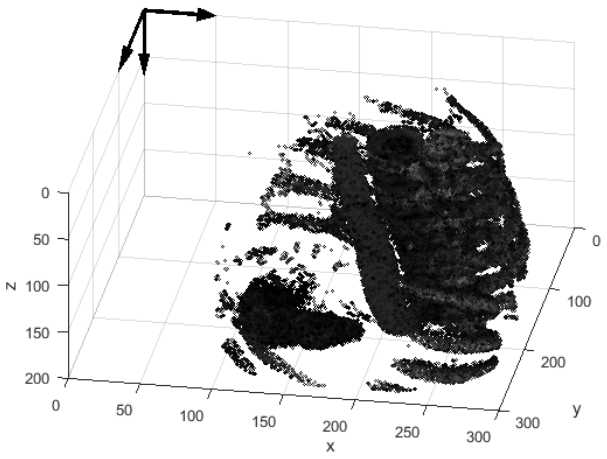


Figure A-2: CT-mm coordinate system

CT-device coordinate system The orientation of the dimensions in the CT-device system differs from the CT-mm system. Specifically the first dimension x should denote direction from left to right, y goes from head to feet and z goes from front to back while the origin remains in the upper left corner. To get the proper orientation we apply the following matrices as displayed in Eq. A-2.

$$R_{\mathrm{CTmm}\rightarrow\mathrm{CTdevice}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & S_{\mathrm{CT}} * \Delta z_{\mathrm{CT}} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{A-2}$$

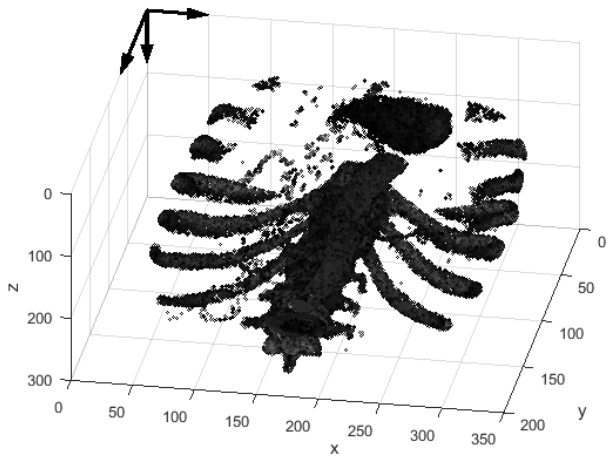


Figure A-3: CT-device coordinate system

World coordinate system The scanning device used during the intervention consists of a large C-shaped arm, positioned around the patient on the table. At the lower end of this arm an X-ray source is positioned. At the upper end of this arm a detector plane is positioned. The X-rays travel from the source through the table and patient and land on the detector plane.
The location of the X-ray source, the isocenter and the detector plane are known. What is left to determine is the exact position of the CT image. We know that it is approximately at the isocenter, but as can be seen in the experiments, this is more often than not, not the exact position.
The just mentioned all takes place in the world coordinate system. The system is a 3-dimensional system. The dimensions are denoted by x, y and z. As mentioned this thesis describes a method to determine the proper 3D pose of the CT image in the world coordinate system. This pose consists of the coordinates as well as the three rotations about the axis. This is what the output from Sec. 3-3 will be.
The world coordinate system can be observed in Fig. A-4. Here, we notice the CT image in the world coordinate system at the isocenter, as well as the detector plane in the top left corner and the source (denoted as a red dot) in the bottom right corner.

The origin of the CT device system is located at the upper top left voxel. Without any translation applied this is also where its origin in the world system will be located. However we want the origin of the world coordinate system to align with the center of the CT image. Using the dimensions of the 3D volume (R_{CT},C_{CT} and S_{CT})we can create the translation matrix necessary as in Eq. A-3.

$$Tl_{\mathrm{CTdevice}\rightarrow\mathrm{world}} = \begin{bmatrix} 1 & 0 & 0 & -R_{\mathrm{CT}} * \Delta x_{\mathrm{CT}}/2 \\ 0 & 1 & 0 & -C_{\mathrm{CT}} * \Delta y_{\mathrm{CT}}/2 \\ 0 & 0 & 1 & -S_{\mathrm{CT}} * \Delta z_{\mathrm{CT}}/2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{A-3}$$

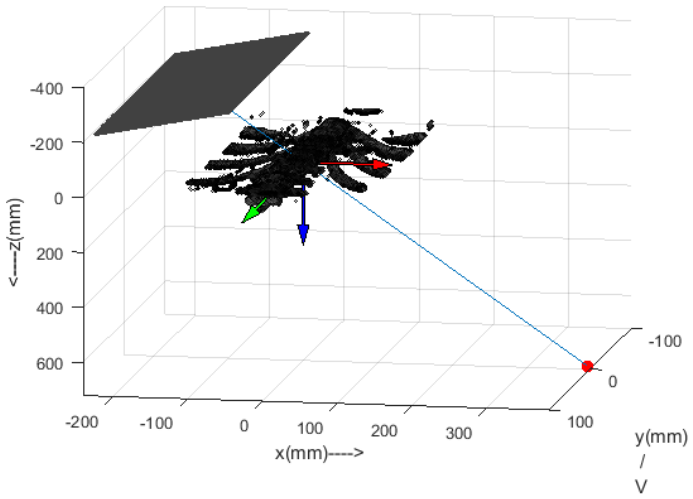


Figure A-4: World coordinate system

By cascading the previous matrices one can transform a CT image from the initial CT coordinate system to the world coordinate system,

$$\mathrm{CT}_{\mathrm{world}} = Tl_{\mathrm{CTdevice}\rightarrow\mathrm{world}} * R_{\mathrm{CTmm}\rightarrow\mathrm{CTdevice}} * S_{\mathrm{CT}\rightarrow\mathrm{CTmm}} * \mathrm{CT}. \tag{A-4}$$

XA coordinate system The XA system has three dimensions. The first two dimensions behave in the same manner as in the CT system. However now the third dimensions does not depict position between front and back but denotes the index of the frame in the sequence, moving further in the sequence means moving forward in time. Hence a point $\mathrm{XA}(r, c, f)$ thus denotes the pixel at row r(position between head to feet), column c(position between patient right to left) and frame f(sequence index).
The XA coordinate system can be observed in Fig. A-5 (only the first frame of a sequence is plotted).

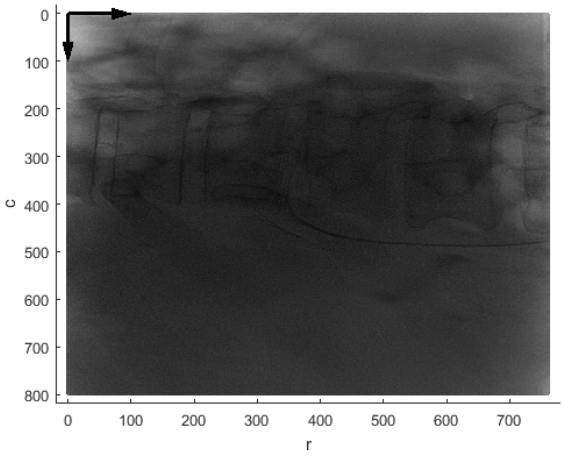


Figure A-5: XA coordinate system

XA-mm coordinate system As of yet in the XA system, the distance between subsequent rows and columns is just one unit. Using the pixel spacing we compose the scaling matrix necessary to travel to a correct mm representation as stated in Eq. A-5. In performing this transformation each frame of XA: $\mathrm{XA}(r, c)$ thus becomes $\mathrm{XA}(x, y)$ by using

$$S_{\mathrm{XA}\rightarrow\mathrm{XAmm}} = \begin{bmatrix} \Delta x_{\mathrm{XA}} & 0 & 0 & 0 \\ 0 & \Delta y_{\mathrm{XA}} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{A-5}$$

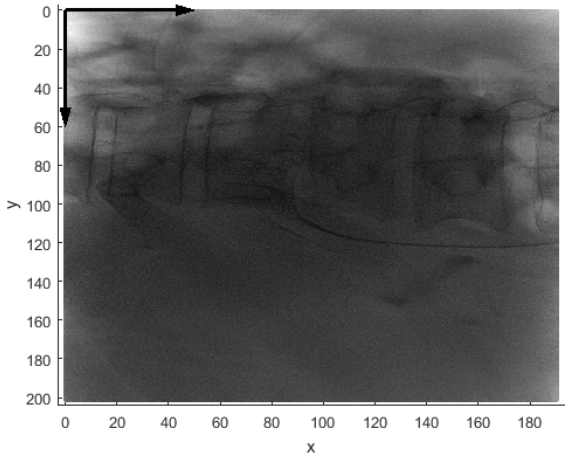


Figure A-6: XA-mm coordinate system

X-ray device 2D coordinate system The orientation of the dimensions in the X-ray device 2D system differs from the XA-mm system. Specifically the first dimension x should denote direction from left to right while y goes from head to feet. To get the proper orientation we apply the following rotation matrix as displayed in Eq. A-6.

$$R_{\mathrm{XAmm}\rightarrow\mathrm{X-ray2D}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{A-6}$$

Note that the X-ray device 2D coordinate system is our 2D representation of the detector plane. Our goal is a correct alignment of the DRR (projection of the CT image) in this coordinate system with the XA image. The X-ray device 2D coordinate system can be observed in Fig. A-7. We can transform our $\mathrm{XA}(r, c)$ image in the XA-mm coordinate system to the image $\mathrm{XA}(x, y)$ in the X-ray 2D device coordinate system by using $R_{\mathrm{XAmm}\rightarrow\mathrm{X-ray2D}}$.

By cascading previous matrices one can transform a XA image from the initial XA coordinate system to the X-ray 2D Device coordinate system,

$$\mathrm{XA}_{\mathrm{X-ray2D}} = R_{\mathrm{XAmm}\rightarrow\mathrm{X-ray2D}} * S_{\mathrm{XA}\rightarrow\mathrm{XAmm}} * \mathrm{XA}. \tag{A-7}$$

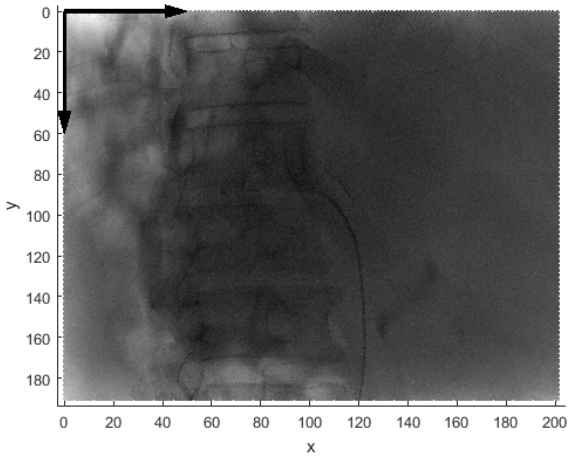


Figure A-7: X-ray device 2D coordinate system

X-ray device 3D coordinate system Off course the detector plane is also defined in 3D as depicted in Fig. A-8, it can also be seen in Fig. A-4 in the top left corner. We can travel from our 3D representation to the 2D one by applying a rotation and translation as in Eq. A-12. As for the rotation matrices. The C-arm system (see Fig. A-9) has two angles, called the tilt and skew (see Fig. A-10). Theoretically there also is a third angle, however this angle was not stated for any patients in the training and testsets and assumed zero. A future incorporation of this angle should not introduce any problems.
The tilt of the C-arm represents a rotation about the y-axis whereas the skew of the C-arm represents a rotation about the x-axis. Eq. A-8 and Eq. A-9 denote the necessary elementary rotation matrices to achieve the rotation about both axis. Eq. A-10 denotes the combination of the matrices.
Now the translation matrix incorporates the distance from the isocenter to the center of the detector plane. The center of the detector plane is located at $x_{\text{off}} = \Delta x_{\text{XA}} * 0.5 * C_{\text{XA}}$ and $y_{\text{off}} = \Delta y_{\text{XA}} * 0.5 * R_{\text{XA}}$. Furthermore the isocenter is located at a distance of $\text{IOD} = \text{SID} - \text{SOD}$ mm from the detector plane center. Here, SID stands for the source image distance (distance between source and detector plane), SOD stands for the source object distance (distance between source and isocenter) and IOD stands for the image object distance (distance between detector plane and isocenter). The translation $T_{l\text{X-ray3D}\rightarrow\text{X-ray2D}}$ is depicted in Eq. A-11. The cascade of $T_{l\text{X-ray3D}\rightarrow\text{X-ray2D}}$ and $R_{\text{X-ray3D}\rightarrow\text{X-ray2D}}$ forms $T_{\text{X-ray3D}\rightarrow\text{X-ray2D}}$, the transformation matrix necessary to travel from the X-ray device 3D system to the X-ray

device 2D system.

$$R_y = \begin{bmatrix} cos(\alpha) & 0 & sin(\alpha) & 0 \\ 0 & 1 & 0 & 0 \\ -sin(\alpha) & 0 & cos(\alpha) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{A-8}$$

$$R_x = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos(\beta) & -sin(\beta) & 0 \\ 0 & sin(\beta) & cos(\beta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{A-9}$$

$$R_{\mathrm{X-ray3D}\rightarrow\mathrm{X-ray2D}} = (R_y * R_x) = \begin{bmatrix} cos(\alpha) & 0 & -sin(\alpha) & 0 \\ sin(\alpha)sin(\beta) & cos(\beta) & cos(\alpha)sin(\beta) & 0 \\ sin(\alpha)cos(\beta) & -sin(\beta) & cos(\alpha)cos(\beta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{A-10}$$

$$T_l = \begin{bmatrix} 1 & 0 & 0 & y_{\mathrm{off}} \\ 0 & 1 & 0 & x_{\mathrm{off}} \\ 0 & 0 & 1 & \mathrm{IOD} \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{A-11}$$

$$T_{\mathrm{X-ray3D}\rightarrow\mathrm{X-ray2D}} = T_{l\mathrm{X-ray3D}\rightarrow\mathrm{X-ray2D}} * R_{\mathrm{X-ray3D}\rightarrow\mathrm{X-ray2D}} \tag{A-12}$$

The missing link at this moment is to get from World coordinate system where our CT is currently stationed, to the X-ray device 2D coordinate system. We can achieve this by multiplying one matrix to the previous stage. This matrix P projects a 3D point onto a 2D plane. Now with the angles of the scanning device represented in $R_{\mathrm{X-ray3D}\rightarrow\mathrm{X-ray2D}}$ and the necessary offset from patient to detector plane represented in $Tl\mathrm{X-ray3D}\rightarrow\mathrm{X-ray2D}l$, all that is left to the matrix P is projecting the 3D points to the detector plane. The matrix can be viewed in Eq. A-13.
Combining matrices $R_{\mathrm{X-ray3D}\rightarrow\mathrm{X-ray2D}}$, $T_{l\mathrm{X-ray3D}\rightarrow\mathrm{X-ray2D}}$ and P we can construct our transformation matrix $T_{world\rightarrow X-ray2D}$ as in Eq. A-14 to transform any 3D point or image from the world coordinate system to the detector coordinate system.

$$P = \begin{bmatrix} SID & 0 & -y_{\mathrm{off}} & 0 \\ 0 & SID & -x_{\mathrm{off}} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & SID \end{bmatrix} \tag{A-13}$$

$$T_{\mathrm{world}\rightarrow\mathrm{X-ray2D}} = P * T_l * R \tag{A-14}$$

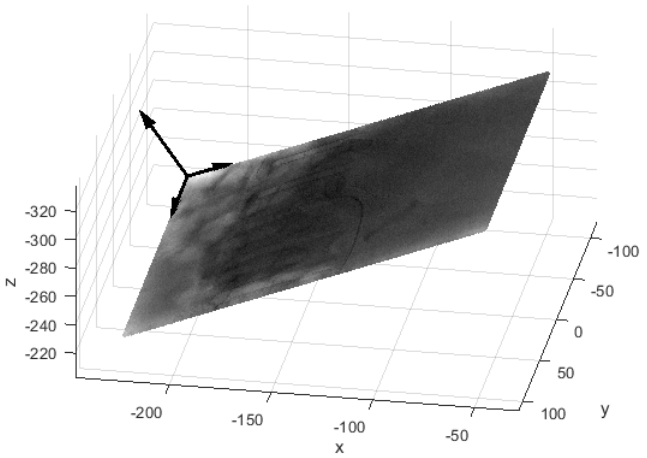


Figure A-8: X-ray device 3D coordinate system

Figure A-9: C-arm

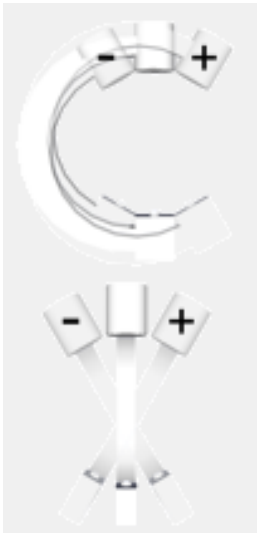


Figure A-10: The skew(top image) and tilt(bottom image) of the C-arm

Appendix B

Construction of a DRR

This appendix deals with the projection of the 3D CT image onto the 2D detector plane. It involves the computation of a synthetic X-ray image, called a Digitally Reconstructed Radiograph(DRR). We achieve this by casting X-rays through the CT image using the scanning geometry at the time of the intervention. The DRR pixel values will simply be a summation of the encountered CT voxel values along the rays.

Now we can go about the construction of the most accurate DRR by starting with interpolating each voxel to a small enough size, by doing so the projection will be well defined. No 'sub voxel' will influence more than one pixel. However in practice this approach takes a lot of computational power.
A computationally less expensive approach is to only project the voxel centers and splat the projections with an appropriate kernel created to represent the projection of a complete voxel. We go about the construction by first interpolating the most center voxel. Then we derive a kernel from the projections of the interpolated voxel points. Finally, to counter the effect that using the same kernel becomes less accurate for voxels far away from the center, we also apply a Gaussian blur on the kernel with a standard deviation of 33% of the original kernel size.

Fig. B-1 shows the result of projecting the voxel centers using $T_{\text{world}\rightarrow\text{X}-\text{ray2D}}$ without splatting with the mentioned kernel. Fig. B-2 shows the result of projecting the voxel centers using $T_{\text{world}\rightarrow\text{X}-\text{ray2D}}$ with splatting.
Throughout this thesis, all DRRs are created using the splatting technique.

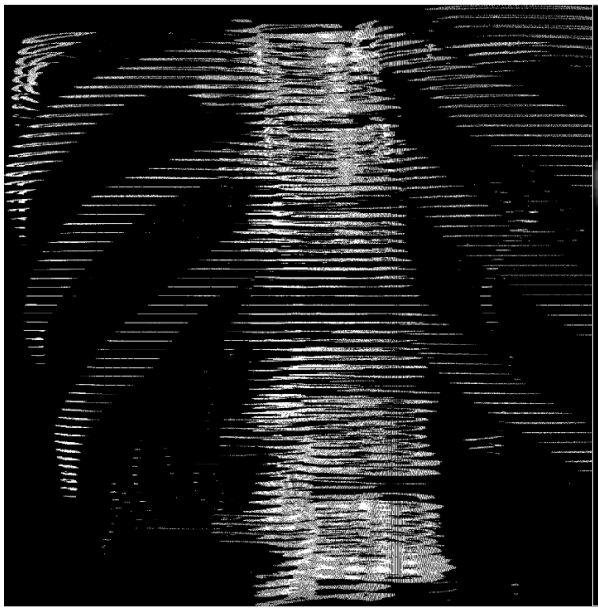


Figure B-1: DRR without voxel splatting

Figure B-2: DRR with voxel splatting

Appendix C

Experiment results

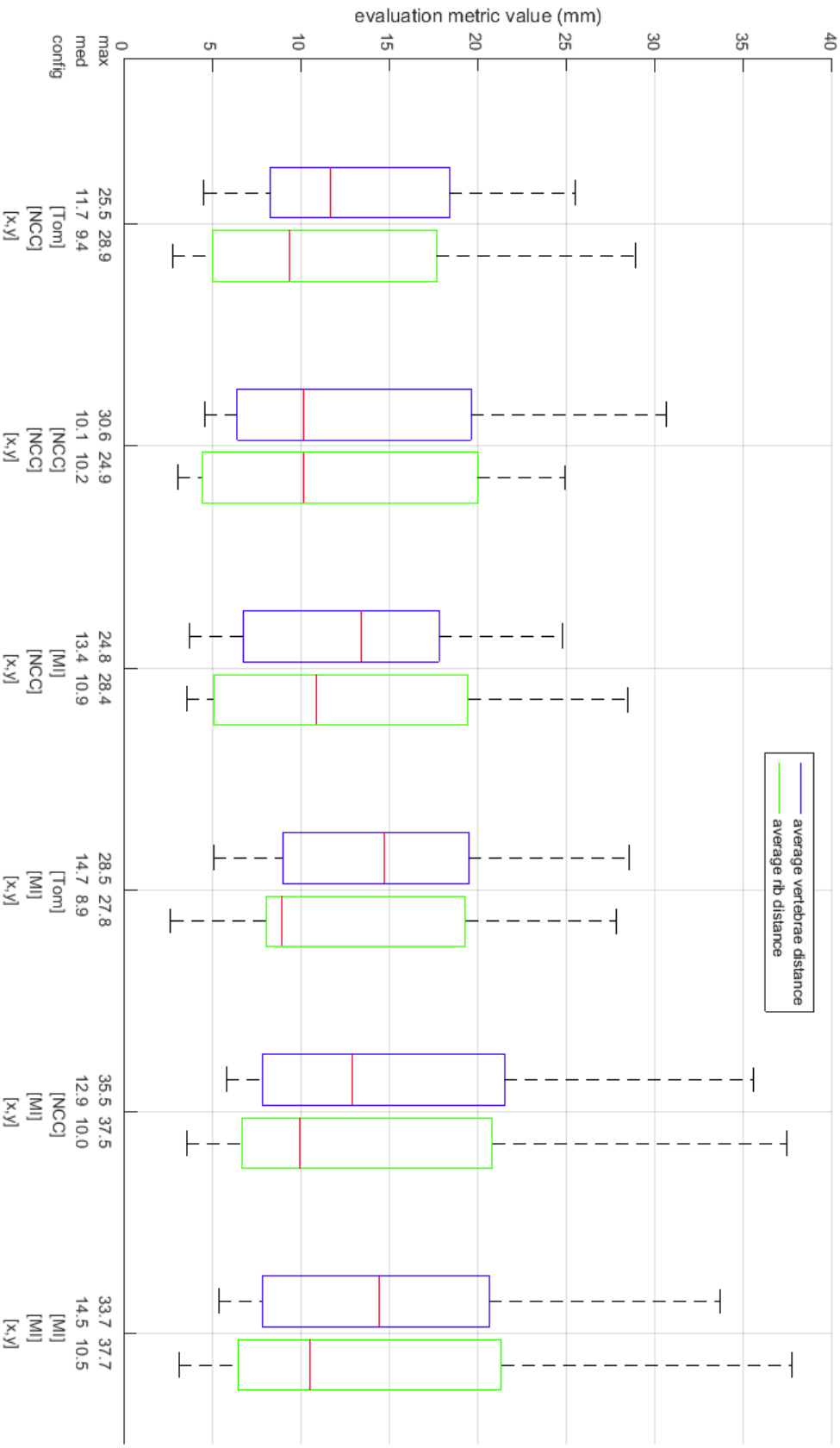


Figure C-1: Result of all 3D pose optimization stage experiments

Bibliography

[1] L. Freire P. Almeida A. Gouveia, C. Metz and S. Klein. Registration-by-regression of coronary cta and x-ray angiography. 2015.

[2] A. Radaelli T. Rivolta G. Piffaretti A. Ierardi, E. Duka and G. Carrafiello. Fusion of ct angiography or mr angiography with unenhanced cbct and fluoroscopy guidance in endovascular treatments of aorto-iliac steno-occlusion: Technical note on a preliminary experience. 2015.

[3] T. Carrell A. Varnavas and G. Penney. Fully automated 2d–3d registration and verification. *Medical Image Analysis 26*, 2015.

[4] A. Ho D. Kim R. Shahidi J. Adler D. Russakoff, T. Rohlfing and C. Maurer. Evaluation of intensity-based 2d-3d spine image registration using clinical gold-standard data. 2003.

[5] K. Mori D. Rueckert A. Ho J. Adler D. Russakoff, T. Rohlfing and C. Maurer. Fast generation of drr using attenuation field with application to 2d-3d image registration. *IEEE Transactions on Medical Imaging*, 24(11), November 2005.

[6] T. Slivnik D. Tomazevic, B. Likar and F. Pernus. 3-d/2-d registration of ct and mr to x-ray images. *IEEE Transactions on Medical Imaging*, 22(11), November 2003.

[7] J. Little P. Desmedt D. Hill G. Penney, J. Weese and D. Hawkes. A comparison of similarity measures for use in 2-d–3-d medical image registration. 1998.

[8] N. Dastur G. Penney, A. Varnavas and T. Carrell. An image-guided surgery system to aid endovascular treatment of complex aortic aneurysms: description and initial clinical experience. *IPCAI*, 2011.

[9] R. Raupach H. Bruder, T. Flohr. Histogram-based image filtering in computed tomography, April 2004.

[10] Z. Yaniv H. Livyatan and L. Joskowicz. Gradient-based 2-d/3-d rigid registration of fluoroscopic c-ray to ct. *IEEE Transactions on Medical Imaging*, 22(11), November 2003.

[11] E. Bozzi D. Lauretti A. Cicorelli A. Lunardi-R. Cioni I. Bargellini, F. Turini and C. Bartolozzi. Image fusion of preprocedural cta with real-time fluoroscopy a feasibility study. *CIR*, 36, 2012.

[12] P. Edwards J. Hipwell D. Rueckert G. Sanchez-Ortiz S. Hegde V. Rahunathan K. Rhode, D. Hill and R. Razavi. Registration and tracking to integrate x-ray and mr images in an xmr facility. *IEEE Transactions on Medical Imaging*, 22(11), 2003.

[13] D. Kellner. regiongrowing, August 2011.

[14] M. Kaminski A. Napieralski J. Kasprzak M. Borzecki, A. Skurski and P. Lipiec. Applications of ray-casting in medical imaging. january 2014.

[15] C. Metz A. van der Giessen A. Weustink-N. Mollet J. Wentzel T. van Walsum M. Schaap, L. Neefjes and Wiro Niessen. Coronary lumen segmentation using graph cuts and robust kernel regression. *Information Processing in Medical Imaging*, 2009.

[16] C.T. Metz. *Coronary Motion Modeling For CTA to X-Ray Angiography Registration.* PhD thesis, Erasmus MC, University Medical Center, Rotterdam, 2011. section 2.2.2.

[17] B. Likar P. Markelj, D. Tomazevic and F. Pernus. A review of 3d/2d registration methods for image-guided interventions. *Medical Image Analysis 16*, 2012.

[18] R. Liao S. Miao, J. Lucas. Automatic pose initialization for accurate 2d 3d registration applied to abdominal aortic aneurysm endovascular repair. 2012.

[19] A. McGregor R. Coombs S. Zhou, I. McCarthy and S. Hughes. Geometrical dimensions of the lower lumbar vertebrae – analysis of data from digitised ct images. *Eur Spine*, 9, 2000.

[20] J. Brown T. Carrell, B. modarai and G. Penney. Feasibility and limitations of an automated 2d 3d image registration system for complex endovascular aortic procedures. 2010.

[21] L. Westover. Interactive volume rendering. *Proceedings of the Chapel Hill Workshop on Volume Visualization*, May 1989.