Master's Thesis

Study and Analysis of Image Similarity Metrics and Image Filter for Vertebrae Pose Estimation

Chen Zheng

September 11, 2019

Graduate School of Information Science
Nara Institute of Science and Technology

A Master's Thesis
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
MASTER of ENGINEERING

Chen Zheng

Thesis Committee:

Professor Hirokazu Kato	(Supervisor)
Professor Yoshinobu Sato	(Co-supervisor)
Associate Professor Masayuki Kanbara	(Co-supervisor)
Assistant Professor Alexander Plopski	(Co-supervisor)
Assistant Professor(Specially appointed) Dawar Khan	

Study and Analysis of Image Similarity Metrics and Image Filter for Vertebrae Pose Estimation*

Chen Zheng

Abstract

X-ray image-assisted surgical navigation system has been broadly applied to spinal surgery. However, the X-ray image is 2D and accurate path planning on the depth direction of X-ray image is difficult. Typically, 2D X-ray images can be augmented with preoperative 3D Computed tomography (CT) image to provide planning and enhanced spatial perception. In this thesis, we study 2D-3D image registration methods to obtain the optimal transformation matrix between X-ray and CT image frames. The 2D-3D registration is strongly dependent on the image similarity (similarity between X-ray and Digital Reconstructed Radiograph images), the noise and image filters used. Image filter is a common method for image pre-processing to reduce the noise and enhance the feature of image. Three similarity measurement methods of 2D-3D image registration including Zero Normalized Cross-Correlation(ZNCC), Structural Similarity index (SSIM) and Mutual Information (MI) combined with intensity based optimization algorithm are applied to evaluate their performance in convergence, efficiency, and accuracy. In addition, we studied image filter to find out its optimal value for different pose registration.

We compared three similarity measurement method by using 20 X-ray images from 5 poses and 5 CT volume data. In each experiment we noted the accuracy of x-, y- and z-axis alignment, and confirmed the reason of the error. Experimental results reveal that the combination of Zero Normalized Cross-Correlation measurement method has the best performance of alignment between X-ray image

*Master's Thesis, Graduate School of Information Science,
Nara Institute of Science and Technology, September 11, 2019.

and Digital Reconstructed Radiograph (DRR) images. The experiments proved image filter can influence and improve the accuracy of registration under the 0.5mm error target. This research will give the benefits to the other researchers or surgeons.

Keywords:

2D-3D Registration, X-ray Image, Computed Tomography, Zero Normalized Cross-Correlation, Image Filter

Contents

List of Figures

List of Tables

1 Introduction

Medical imaging is an image analysis technology that is used to view the human body in order to diagnose, monitor, or treat medical conditions. It can gain the kinds of information about the area of the body being studied or treated, related to possible disease, injury, or the effectiveness of medical treatment [23]. One of the classic technology in medical imaging is image registration. Image registration is a computer vision technology that it can align different images of same scene [10]. Therefore, medical image registration is the combination technology which is the process of registering two image models into same pose. This process involves designating one medical image model as the reference, and applying geometric transformations to the other image model so that it registers with the reference. This technology as a treatment tool can complement each image models and provide more detail information to the doctor from the medical image source [21]. However, depending on different human body object, there are several medical image registration methods. In our research, we are focusing on lumber spine by using 2D-3D registration. We used different metrics for 2D-3D image registration based optimization algorithm to evaluate registration performance. Additionally, we evaluated the influence of image filter and tried to optimize the accuracy of registration.

1.1 Background and Motivation

Lower back pain is becoming more common, as people work overtime and have an incorrect pose of sitting for a long time [9]. The lumbar spine, or low back, is a remarkably well-engineered structure of interconnecting bones, joints, nerves, ligaments, and muscles all working together to provide support, strength, and flexibility. The structure of lumbar spine contains 5 biggest vertebrae bones

(a) (b)

Figure 1.1: Lumbar spine and X-ray scanner. a) Vertebrae highlighted in red [8], b) Conventional C-arm device [6]

as they need to carry human upper weight [19]. This complex structure also leaves the low back susceptible to injury and pain as shown in Figure 1.1 a. Because the causes of back pain are often complex and multi-factorial, it is often more difficult to get an accurate diagnosis for back pain than for other medical conditions. While some spinal diagnoses are relatively straightforward (such as tumors, infections, or fractures), for many conditions there is little agreement among spine specialists about a diagnosis. The most common test for spinal diagnosis now is to capture and check injured information of the medical imaging by using medical diagnostic equipment (e.g., C-Arm device, CT scanner) [12]. X-ray image is generated from C-Arm device as shown in Figure 1.1 b to provide detail of the bone structures in the spine, because images taken in different angles to determine spatial target positions. In recent years, using an image-assisted navigation system for spinal surgery has become a trend [2, 12]. However, the X-ray image is 2D and it lacks the 3D spatial information. On the contrary, 3D computed tomography (CT) scans, which are essentially a very detailed X-ray, take cross section images of the body. They provide excellent bony detail and are also capable of imaging for specific conditions. The structure of CT scanner fix the human pose, because the patient only can lie down inside the a CT machine

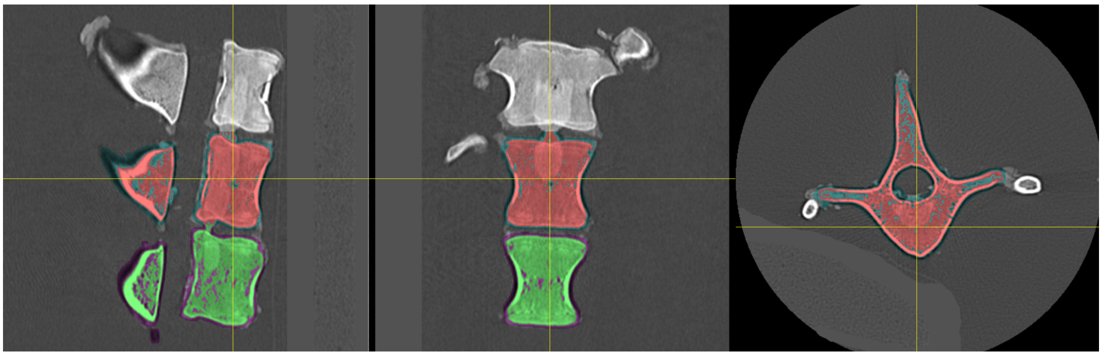


Figure 1.2: Three-view CT image

and the image as shown in Figure 1.2.

If doctor can get the detail information from target pose by combination between X-ray image and CT volume data, it will improve the spinal diagnosis. The C-arm device can capture the arbitrary poses; it means we can get the target pose 2D information. In order to gain the more depth pose information, we need to estimate the pose in the CT volume data. Once we align and find the target pose which is captured from X-ray image into CT data, we can obtain the more information from 3D CT data. It will provide more spinal information to doctor. However, the accuracy of estimation is very important, I would like to contribute and improve the accuracy of vertebrae pose estimation.

1.2 Problem Description

In order to estimate the pose information from 2D X-ray to the CT data, the optimization method is needed. This process called 2D to 3D image registration. However, the most important step is to evaluate the accuracy of the registration when the iterative of optimization is convergence. Although this topic has been widely discussed over the last decades, to achieve high accuracy and robustness still remains challenging [13] and needs to be improved. Some mainly reasons are given as below:

1. The single-view of X-ray image lacks the depth information and makes the optimization problem complicated;

2. Accuracy is affected by noise problem. For example, perspective distortion may occur between 3D and its 2D projection due to the non-linear nature of the transformation. Furthermore, the local maximum is an unavoidable problem in optimization;

3. The rib bones as the noise are often taken together with spine. Due to the image noise, the ground true similarity is deviated from the global maximum similarity.

4. The image filter as a special case for medical image pre-processing will affect the feature extraction. Thus, the accuracy may be influence by different image filter methods.

The goal of this research is to improve the accuracy of pose estimation base on the 2D-3D registration. We will first analyze the depth error problem, and then try to optimize the results.

1.3 Thesis Contributions

In summary, we have the following contributions in this thesis.

1. We found that vertebrae pose estimation is strongly affected by three factors including the similarity metric used (to compute similarity between X-ray and DRR image), noise, and the image filter used.

2. We conducted a comprehensive analysis of three different similarity metrics (ZNCC, SSIM, and MI); and found that ZNCC performs the more accurately among the three methods.

3. We designed a semi-automatic module that dynamically changes the image filter values; and conducted detailed experiments to find out the optimal values of the image filter for different poses used in our study.

1.4 Thesis Organization

The remainder of this thesis is organized as follows.

- Chapter 2 describes the brief related works that have done in accuracy improvement of 2D-3D registration.
- Chapter 3 introduces system overview, and description of each process will be shown and explained. Meanwhile we will show what data we used for the experiments and the problem in optimization step.
- Chapter 4 describes our empirical study design to analyze and to confirm assumptions of accuracy problem in our system. This chapter also describes different methods we will do for experiments.
- Chapter 5 presents the experiment results for each summarized research problem analysis.
- Chapter 6 mainly make a conclusion of our findings and results have same meaning in the study, and discuss more opportunities for the future work.

2 Related Work

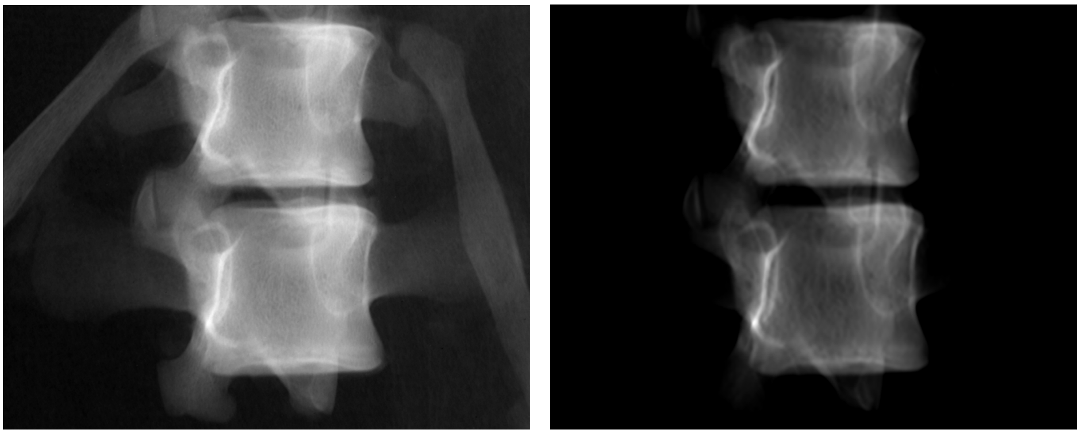
In this chapter, we describe the related works, there are two parts, one is reconstructed image model which is using for similarity measurement, the other one is the related work in 2D-3D registration field. Mainly we introduce some 2D-3D registration works, this is an important part in the research.

2.1 Digitally Reconstructed Radiograph

The conventional registration approaches commonly use the digitally reconstructed radiographs (DRR) get the similarity through the comparison of X-ray images, among the known medical image registration methods [17]. DRR image is a 2D simulated X-Ray image which is generated from the CT data by using ray casting method [7, 17]. Therefore, if the DRR image is totally matching the X-ray target image, we can gain the 3D information from CT data. Since then DRR image have been used in many clinical procedures and new medical approaches for patient treatment. Similarly, Aouadi and Sarry [1] discuss the pose estimation when DRR images are used with different mathematical models. It proved DRR image had good performance to replace CT data for 3D pose estimation. Figure 2.1 shows an example of DRR image.

2.2 2D-3D registration

The comparison method between DRR and X-ray image is the key step as the criterion of optimization to indirectly align the different models. Kotsas and Dodd [11] reviewed the 2D-3D registration methods, and divided the existing rigid registration methods into three categories according to the data volume of image features. The categories include classified feature-based methods, gradient-based



(a) X-ray image (b) DRR image

Figure 2.1: The example of DRR image compares the X-ray image with the same pose. a) X-ray image was captured from the target pose, b) DRR image from the CT data generation

methods, and intensity-based methods. Similarly, Markelj et al. [13], evaluated different 2D-3D registration methods, and proposed a methodology for analysis of 2D-3D registration methods.

Similarity metric depends on different practical application. Our research focuses on the intensity-based projection. Intensity-based method compares intensity patterns in images via correlation metrics, while feature-based methods find correspondence between image features such as points, lines, and contours [22]. Intensity-based methods register entire images or sub-images. If sub-images are registered, centers of corresponding sub images are treated as corresponding feature points. Penney et al. [17] presented a set of new techniques for performing registration of CT data set to X-ray images for minimally invasive CT-based spinal IGS, which are based on geometry and gradient features. Meanwhile prior to this an initial study has been conducted by Matoba [14] on how to approach the problem for the CT spine deformation. Different similarity metrics will give the different optimization convergence and performance. Recently, several works did some contribution in this research, like Chang et al. [5] surveyed performance of combination between five similarity metrics and three optimization methods for spine saw bones and showed the best combination. Groher [6] presented a method to provide methods for the registration of 2D vascular images acquired by a stationary C-arm to preoperative 3D angiographic Computed Tomography (CT) volumes, in order to improve the workflow of catheterized liver tumor treatments.

With the development of neural network, some researches related to 2D-3D registration have achieved good performance. Miao et al. [15] presented a Convolutional Neural Network (CNN) regression approach for real-time 2D-3D registration. The experiment resulted demonstrate the advantage of the proposed method in computational efficiency with negligible degradation of registration accuracy compared to intensity-based methods. Schaffert et al. [18] proposed a method to learn the weighting of local correspondences directly from the global criterion to minimize the registration error. Bier et al. [3] presented a method to automatically detect anatomical landmarks in X-ray images independent of the viewing direction.

For accuracy of translation evaluation, z-axis alignment error is still the challenge. Husejic [8] showed it can use several X-Ray images which show different spine poses, to correctly estimate the depth error which occurs during pose estimation. Not only single-view image lacks the depth information, but also noise influence the convergence of iterative optimization. In this case, it makes more difficult to match the different image models by using intensity-based in a small error target since there is almost no change in 2D image space. My work focus on the accuracy of depth direction alignment by single-plan image.

3 System Overview

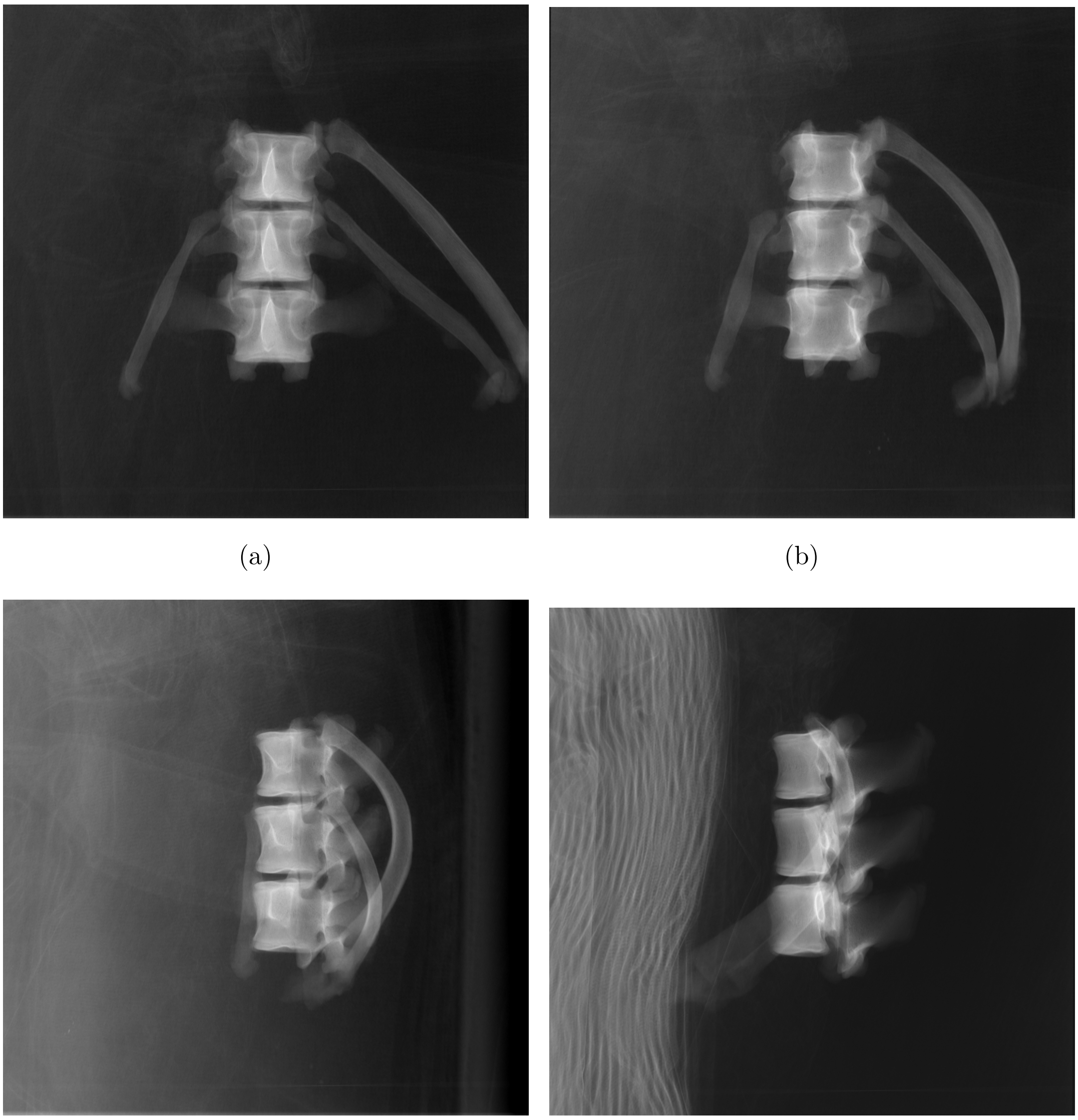
In this chapter, we described the overall system overview including data preparation, initialization, the optimization framework and results analysis. Figure 3.2 showed the overall pipeline of our system.

3.1 Dataset

Typically, the pose of X-ray is different from that of the CT data. The same difference in the pose happens with vertebrae problem. Our data set consists of CT data and X-ray images, which is gathered in the following manner.

In order to build the evaluation system, the previous work in our group captured the X-ray and CT datasets from a frozen pig in the same pose. We fixed one pose and captured four X-ray images from different viewing angles. Meanwhile we put the same pose frozen pig into CT scanner and generated the CT volume data. In our research, we used twenty X-ray images and five CT volume data which were captured from five poses. For each pose, there are four X-ray images and one CT volume data should be generated. Once an X-ray image is captured, the X-ray device should move thirty degrees for the next X-ray image generation as shown in Figure 3.1. The target bones between X-ray and CT data only captured two closed vertebrae bones from the whole lumbar spine. Because the purpose of pose estimation is to extract the depth and more information from CT for X-ray target pose. If we can find the relationship between these two vertebrae bones, it means we can get the 3D pose information in CT data and the system also can extend for pose estimation of whole lumbar spine.

As we have mentioned before, the classic method for 2D to 3D image registration is intensity based that it calculates image similarity between X-ray target image and DRR image. To generate the DRR image we need a surface model by



(a) (b) (c) (d)

Figure 3.1: Four X-ray images in same pose are captured from four different viewing angles. a) is the initial captured position of front-view, and B) c) D) rotated and generated from the incremental 30^o degree view angles

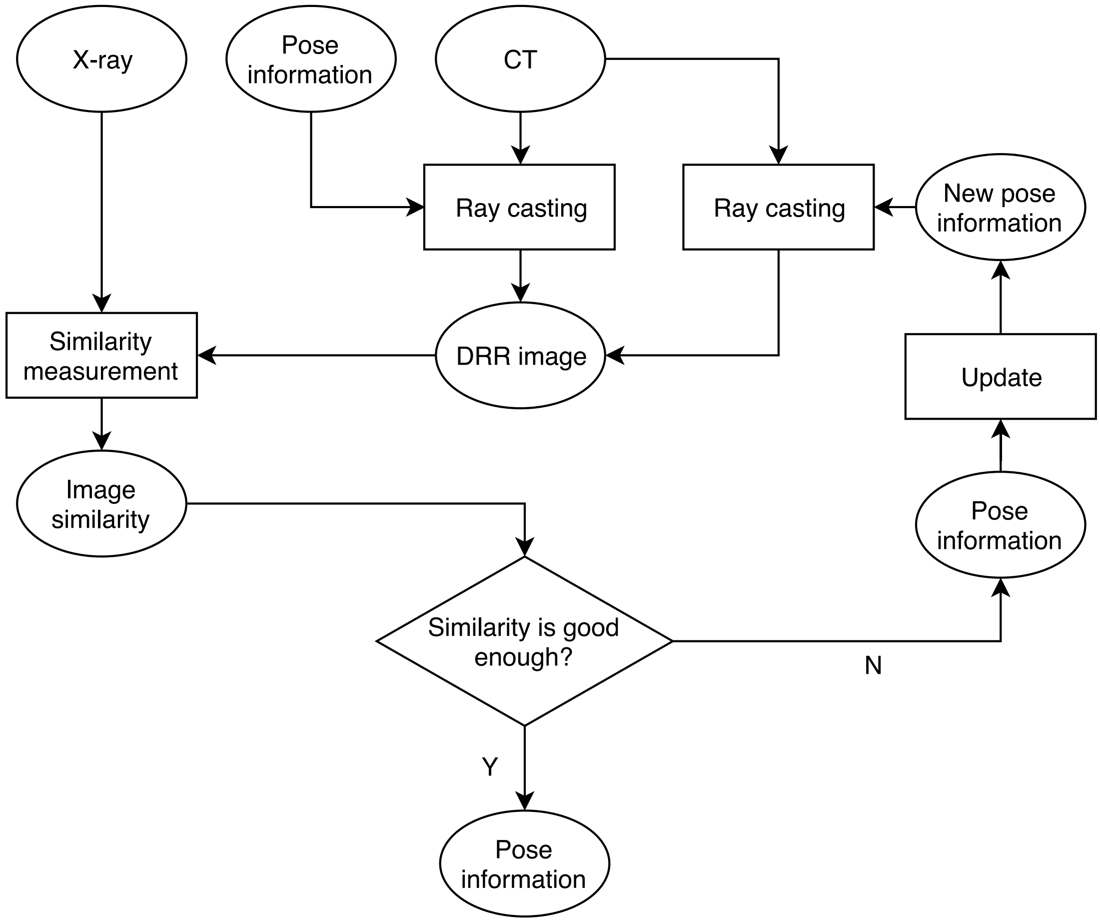


Figure 3.2: System Overview: Initialization and optimization system

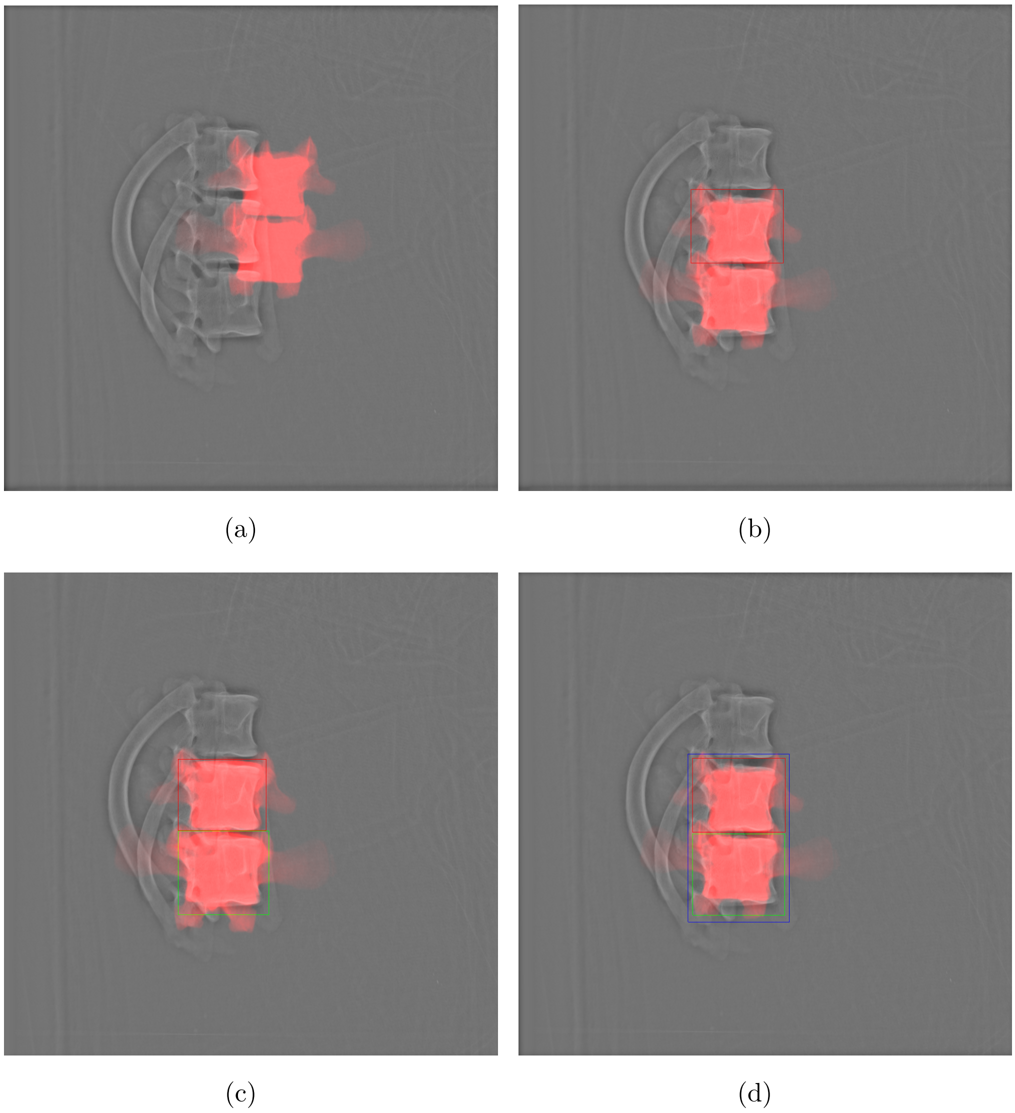
segmenting the CT data. Currently the bone segmentation process is a manual task and it as the initialization of data preparation is very important. We will give more description about initialization in next session.

3.2 Initialization and Optimization

The registration of a CT volume relative to an X-ray image is estimated by iteratively optimizing the similarity measure calculated between a DRR generated for the current transformation and an X- ray image. In order to register X-ray image into CT data, we used ray casting method [4] for DRR generation. Ray casting method provides a baseline for many computer graphics and rendering algorithms that use the geometric algorithm of ray tracing. Ray tracing-based rendering algorithms operate in image order to render three-dimensional scenes to two-dimensional images.

To generate DRR image based on the ray-casting method and run the optimization algorithm, an initial pose for the CT must be selected. The initialization step has already built in previous work and it is a manual process in our system. This process can be seen in Figure 3.3. At the beginning, we input the X-ray target image, CT data and basic pose information to manually align the pose as possible. Because the dataset is based on two vertebrae bones (labeling in L3 and L4), our focus is on the lumbar spine bones L3 (red rectangle) in Figure 3.3 b and L4 (green rectangle) in 3.3 c. Our generation system can separately align each bone and finally optimize the whole generation. This method could also be applied to any of the lumbar vertebrae bones. After L3 and L4 bone selection we also need to select both of those bones as shown in Figure 3.3 d, marked with a blue rectangle. After these three times selection, an initial DRR image as the starting point in optimization method is generated. This manual alignment must be careful otherwise we will not be able to find the global minimal value. This is a general mathematical problem when an optimization algorithm is used. We will give a detail description in chapter 4.

The next step is running optimization system to find the best matching by using similarity measurement between X-ray and DRR images. Our optimization system is standard gradient descent method. Start from the initial pose, each time we calculate the similarity between X-ray target image and DRR image and then check the similarity is maximum or not. If the gradient of current intensity similarity is zero, the pose in DRR as the convergence point can find more 3D information from CT data. In our system, this method will run three times to optimize the pose. The pass two times, we separately optimize two vertebrae bones. Finally, the third time we run the optimization system for whole vertebrae bones same as our initialization step. Because our initialization system for DRR image generation also have tree step, we separately optimize registration two times and finally optimize the whole bones. However, the challenging problem for 2D-3D registration is the accuracy. We need to build a evaluation system for accuracy checking.



(a) (b)

(c) (d)

Figure 3.3: Initial generation of the DRR image (red color) with corresponding X-ray(background part). a) The original DRR bone pose and target pose of X-ray image, b) L3 bone selection shown with a red rectangle, c) L4 bone selection shown with a green rectangle, d) L3 and L4 bone selection shown with a blue rectangle

3.3 Evaluation System

For pose estimation, understanding and evaluating accuracy of optimization result is therefore an important step. The prerequisites of presented evaluation method assume for 2D-3D registration that the intrinsic parameters of the X-ray device are known and that the X-ray images are corrected for distortion [20]. Therefore, six degrees of freedom are left in the 2D-3D registration problem, which are the parameters that describe the position of the 3D volume in space: translations in x, y and z directions (T_x, T_y, T_z) and the orientation of the 3D volume in space: rotations around x-, y- and z-axis (R_x, R_y, R_z). Target pose in X-ray image as the ground truth contain the spatial relationship between two closed vertebrae bones. Meanwhile, through the known transformation matrix can calculate the world coordinate and get the relationship between two bones. In the normal accuracy evaluation, the position and orientation of the 3D shape model from the result compare the position of ground truth as the estimation result are verified an accuracy error. As we described before, it can be expressed by six parameters (T_x, T_y, T_z, R_x, R_y, R_z).

The problem of our evaluation system that we do not know the definite transformation between C-arm camera and X-ray target image. It means we cannot build the normal evaluation system without information transformation matrix of ground truth. In order to build the evaluation system, both two vertebrae bones as one coordinate system to evaluate the transformation which is defined by an identity matrix. Because the X-ray images and CT data we captured from same pose, the transformation of accurate registration will be similar to identity matrix. Compare these two transformation matrix, we can evaluate the registration result.

For 2D-3D registration, the capture range as another important aspect will affect the accuracy. The capture range defines the range of starting positions from which the algorithm finds a sufficiently accurate transformation. Two factors are involved here: the definition of a misregistration, and the fraction of allowed misregistrations. The selection of capture range is manually initialization step and decide the size of image. This is also the reason we mentioned the initialization must be careful. This work we focus on accuracy target is under 0.5mm that we got the recommendation from professional surgery doctor. Using error evaluation system, we can get the accuracy of transformation matrix. Base on this transformation matrix and known camera intrinsic parameters, we can calculate the pixels change of 0.5mm error. For 0.5mm translation error of our data set, there are 4.5 pixels of x- and y-axis changed. but only 0.6 pixel changed for z-axis. Because the interpolation method can zoom the image under one pixel. Even the error distance of depth direction is big, the change is also very small. This work focus on the accuracy of translation, especially depth direction error.

4 Methodology

In this chapter, we first assume and analysis two reasons of error problem: local value of optimization problem and noise problem. Then we describe some different similarity measurement method. Finally, we will describe the influence of image filter.

4.1 Problem Analysis

Our system is an optimization system based on the similarity measurement between X-ray and DRR. The initial pose is very important for optimization. The common problem in optimization method is the local maximum value. If the function has many local maximal values a bad starting point will result in a local maximal value interpreted as a global maximal value. Meanwhile, the step size for optimization is another parameter we need to consider and design for iterative optimization method. It determines the iterative times and convergence time. If the step size is not good enough, our optimization system also cannot find the maximum pose. Therefore, the first assumption of error problem is our optimization algorithm has problem that the method cannot find the global maximum value. Figure 4.1 showed the optimization problem.

Noise is always presents in digital images during image acquisition, coding, transmission, and processing steps. All medical images contain some visual noise. The presence of noise gives an image a mottled, grainy, textured, or snowy appearance. The noise will change the appearance pixels of X-ray image. The distribution of these darker and lighter pixels is random, and the image will have a mottled appearance (salt-and-pep-per distribution). Imaging scientists called this grainy appearance “noise,” which is synonymous with “mottle.” Figure 4.3 a shows examples of vertebrae x-ray images obtained surface noise. Noise in vir-

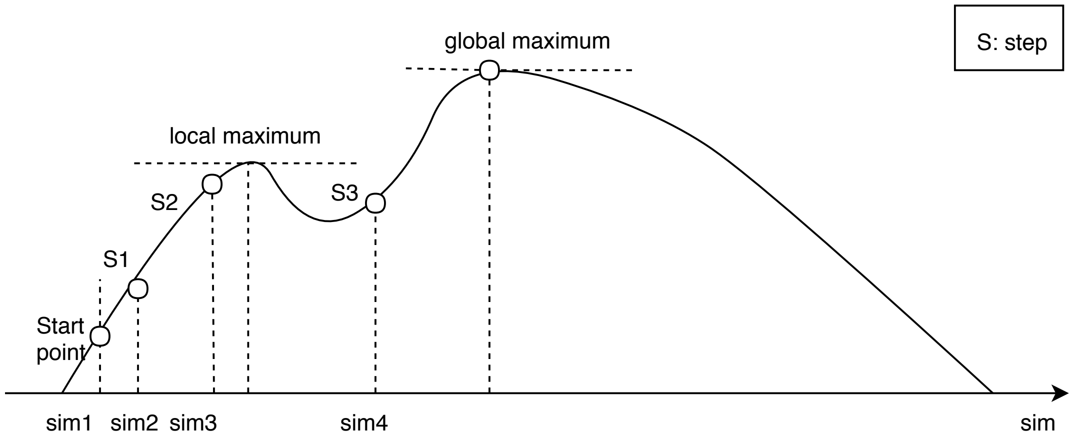


Figure 4.1: Optimization Problem, because the similarity of S3 is smaller than local position, the local maximum value is regard as the global maximum

tually all X-ray imaging modalities is dominated by quantum mottle, where the latter relates to the total number of X-rays used to generate an image. However, the DRR images do not contain the surface noise, because the DRR images are digital reconstructed image from CT volume data. When we computer the similarity of two image models, the noise will influence the calculation result.

When we capture two vertebrae X-ray image, the X-ray image always contain some part of closed vertebrae. The segmentation for CT data can label the target vertebrae bones that the DRR image will not contain the other noise bone. In addition, there are some side bones around the vertebrae. When x-ray tube calculates the intensity of target vertebrae bone, unexpected bones as the noise are also captured. Figure 4.3 b showed the side bone around the target bones are captured together in X-ray image. This kind of unexpected bone will influence the similarity calculation result. When we run the optimization system, the system can find the highest similarity value. Due to the noise influence, the similarity value of ground truth is smaller than similarity of the maximum pose. Figure 4.2 showed the noise problem. For this noise assumption, we need to consider the image filter method to reduce the noise and enhance more feature in X-ray image. My first study is to analysis the error problem and confirm the assumption.

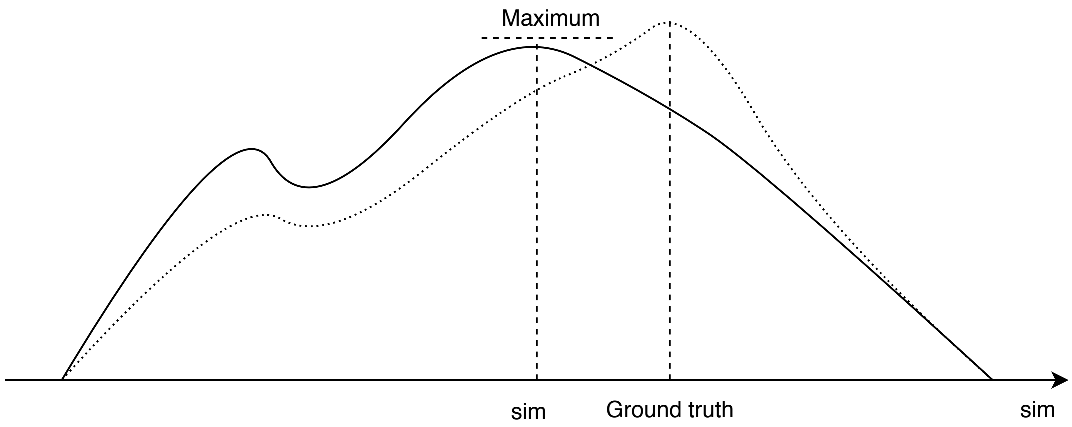
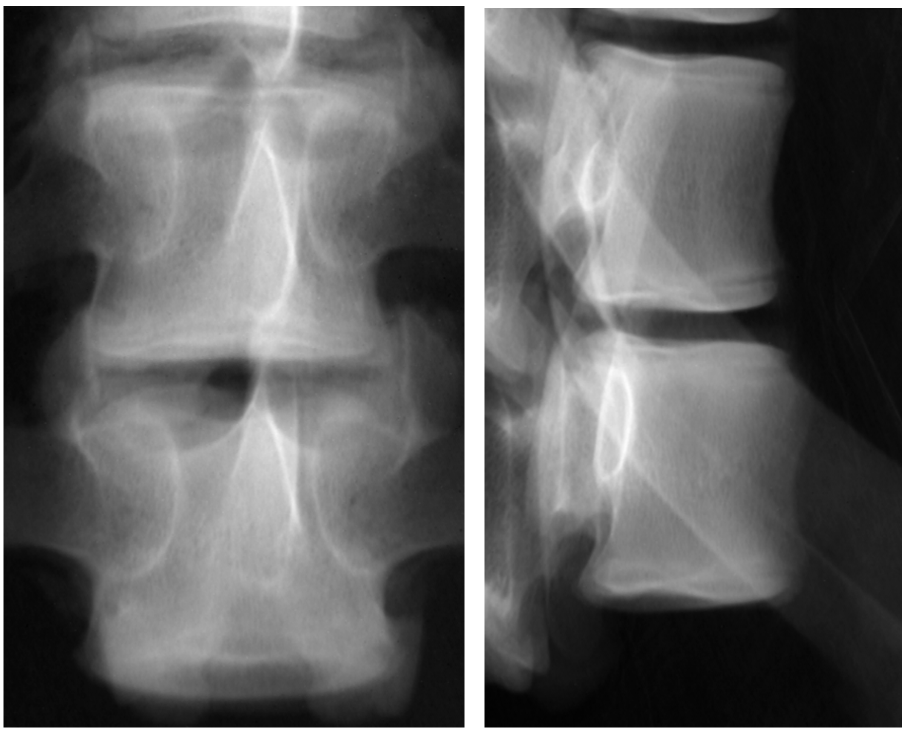


Figure 4.2: Noise Problem, due to the noise influence, the similarity value of ground truth is not the maximum value



(a) (b)

Figure 4.3: Two types of noise in X-ray images. a) showed the surface noise in X-ray image, b) showed the side bone around the vertebrae in X-ray image

4.2 Three Similarity Method Comparison

My research is focus on translation accuracy and our target error is need to under 0.5mm. The problem in translation is accuracy of depth direction. As we described in chapter 3, because X-ray is 2D image, the target error is 4.5 pixel changing for x-axis and y-axis. The image zooming uses linear interpolation to change the pixel. Therefore, a 0.6 change in pixels value gives a target error of 0.5mm. This small pixel value for similarity measurement is a challenge. Additionally, the X-ray image has some noise that makes the alignment become more difficult. The objective function of optimization is defined as the similarity measure of the X-ray and DRR images. Some similarity measure methods are proposed we have introduced in related work. Base on the optimization system, different similarity measurement method has different performance. In this study, we choose three similarity measurement methods (Zero Mean Normalized Cross-Correlation, Mutual Information and Structural Similarity) to compare the performance by using intensity based optimization method. Because the optimization system based on the similarity calculation, we will modify the similarity calculation function to test the data. Then, the accuracy of different method should be evaluated as the comparison result and we will do three comparison experiments. For each pose, we will compare the accuracy among three methods, and finally make a statistical comparison result to show the best performance similarity metric.

Zero Mean Normalized Cross-Correlation

The first similarity metric is ZNCC. ZNCC is one of the Cross-correlation method to compare two different time series to detect if there is a correlation between metrics with the same maximum and minimum values. Instead of simple cross-correlation, ZNCC can compare metrics with different value ranges.

First ZNCC need to calculate the average gray value is:

$$\overline{Img}\left(u,v,n\right)=\frac{1}{\left(2n+1\right)^{2}}\sum_{i=-n}^{n}\sum_{j=-n}^{n}Img\left(u+i,v+j\right)$$

The standard deviation is:

$$\sigma(u,v,n) := \sqrt{\frac{1}{(2n+1)^2}\left(\sum_{i=-n}\sum_{j=-n}^{n}\left(Img\left(u+i,v+j\right)-\overline{Img}\left(u,v,n\right)^2\right)\right)}$$

The ZNCC is defined as:

$$ZNCC(Img_1, Img_2, u_1, v_1, u_2, v_2, n)$$

$$= \frac{\frac{1}{(2n+1)^2}\sum_{i=-n}^{n}\sum_{j=-n}^{n}\prod_{t=1}^{2}\left(Img_t\left(u_t+i, v_t+j\right)-\overline{Img}\left(u_t, v_t\, n\right)\right)}{\sigma_1(u_t, v_t, n)\sigma_2(u_t, v_t, n)}$$

We input X-ray image as the standard image. Go through all pixels, it calculates the difference of pixel by pixel. The pixel in the center has coordinates (u,v) for the part of two images and the number of pixel translation showed in i and j. If two images are totally same, the similarity value is one, and the result range is in [0, 1].

Mutual Information

In probability theory and information theory, the mutual information (MI) of two random variables is a measure of the mutual dependence between the two variables. Mutual information is to calculate the joint part of two cluster. If the distribution of pixels between two images are totally same, Mutual Information regards the image is similar to another one. Mutual information can be equivalently expressed as

$$I(X,Y) = H(X) + H(Y) - H(X,Y)$$

where the H(X) and H(Y) are marginal entropies, it computes the probability of each pixel value in whole image. The joint entropy H(X, Y) is the amount of information we get when we observe X and Y at the same time, and it computes the pixels probability in both two image distribution. If two images are same, the similarity value of MI is one marginal entropy value. Because the Mutual Information consider the distribution of image pixels, we assume this similarity method has noise robustness.

Structural Similarity

The image quality between X-ray image and DRR image is different. Because X-ray image is directly captured by device, the image quality of conventional DRR images calculated from CT images has not been satisfactory suffering from the poor longitudinal resolution and stair step artifacts [16]. We consider the texture in X-ray image is not clear and resolution is less visible in bright regions. The Structural Similarity (SSIM) index is a method for measuring the similarity between two images that it can be viewed as a quality measure of one of the images being compared, provided the other image is regarded as of high quality. The SSIM focus on the intensity structural information. Because X-ray image and DRR image are gray scale images, the pixels have strong inter-dependencies. Considering these reasons, we will compute the structural similarity as the parameter of optimization system and evaluate the performance. The SSIM is defined as:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

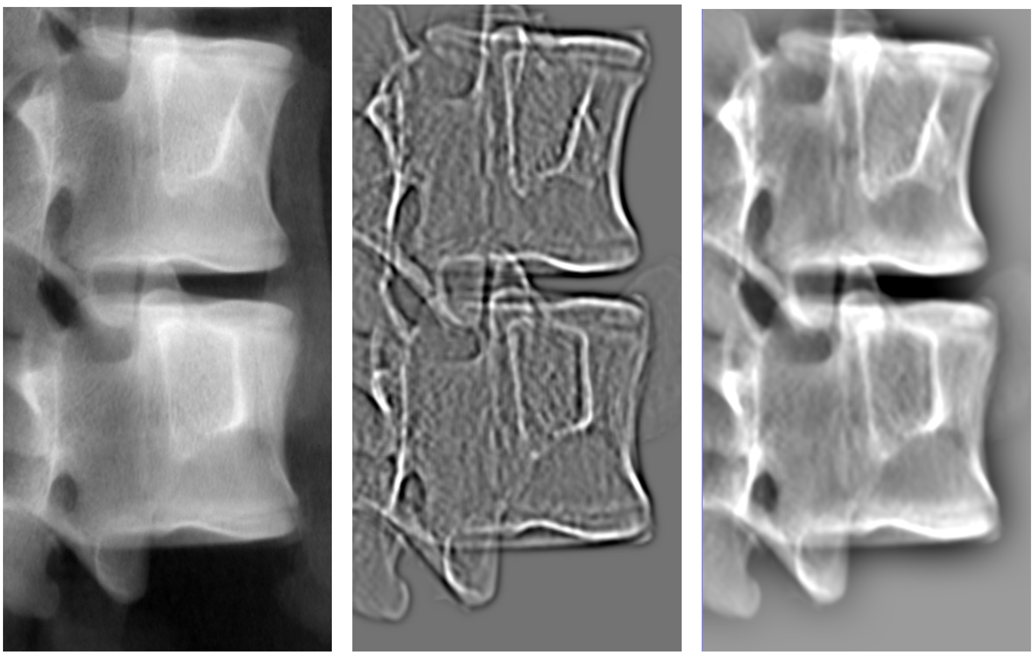
The average of x, y are μ_x, μ_y;
The variance of x, y are σ_x, σ_y;
The covariance of x and y is σ_{xy};
Two variables to stabilize the division with weak denominator is c_1, c_2.

4.3 Image filter influence

In 2D-3D registration, the image filter is an important method for image preprocessing to reduce the noise (e.g., typically Gaussian smoothing, or removing background of target region), and enhance the feature of image. Image filter as the classic method will influence the accuracy of registration because of feature extraction. In our system, we used two basic image filter methods: low-cut filter and low pass filter to preserve the edge of target and reduce the image noise by using Gaussian function. The filter function is defined as:

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

In image processing, a kernel convolution matrix is a small matrix that it is used for blurring, sharpening, embossing, edge detection, and more. The sigma is the most important parameter for kernel convolution and it decides the type of kernel convolution matrix. Different sigma setting can filter images have different performance. Figure 4.4 showed one X-ray image by using different image filter parameters. The similarity comparison is focus on the target bones, we use low-cut try to extract the edge of target bones. However, we do not confirm the influence of image filter and which kind of image filter function we should set up base on the similarity metric. For this condition, we will set different image filter parameters and check influence of the accuracy.



(a) (b) (c)

Figure 4.4: Three results by using different image functions. a) original X-ray image without image filter, b) the filter image by setting 4 sigma value, c) the result of sigma 26 setting

5 Experiments and Discussion

In this chapter, we describe three experiment follow the methodology. First, we confirm the reason of error problem, and the we compare three similarity measurement method. Finally, we do the image filter experiment to improve the accuracy.

The experiment was implemented on Mac OS, which included applications of initialization, optimization and evaluation on the CPU with 2.7 GHz Intel Core i5, Intel Iris Graphics 6100 1536 MB Graphics and 16 GB Memory.

5.1 Confirm Noise Problem

Firstly, we need to confirm the error issue. There are two condition we mentioned in chapter 3: optimization problem and noise influence. Optimization system as an iterative process can find the global maximum and local value. Our evaluation system can provide the error distance between the optimization result and ground truth pose. To confirm the local optimization problem, we did two experiments. First, we initialized the DRR image and ran the optimization system to register the images. After getting the optimization result, we changed and generated the new pose DRR in ground truth pose by following the evaluation result. Then we calculated and checked the position of similarity ground truth. If the similarity is higher than optimization result, it means our optimization algorithm is not good enough that it cannot find the maximum value. On the other hand, we need the second study to check the larger region of similarity between different DRR images and target X-ray image. If the result showed the stop pose of optimization is the maximum value, it proved our optimization method is good enough. Figure 5.1 showed an example of result. Experiments showed stop pose of optimization system is the maximum similarity value. We can confirm

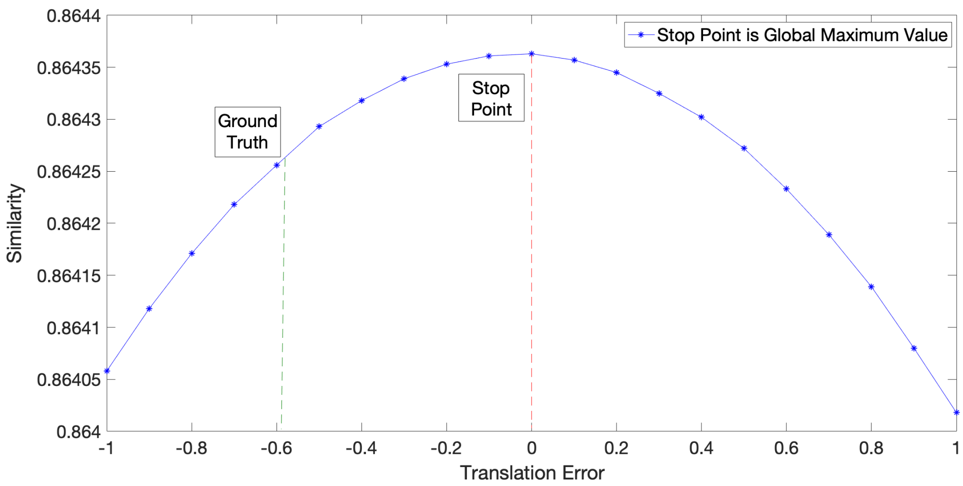


Figure 5.1: Example result, it proved the optimization system can find the maximum similarity value, but the similarity value of ground truth is not maximum

our optimization can find the global similarity. However, the similarity value of ground truth is not maximum by noise influence. In the next study, we will compare different similarity measurement methods and check the result.

5.2 Similarity Metric

In this section, we compared three similarity measurement methods. We have confirmed our intensity based optimization method can find the global maximum pose alignment. After we run the optimization method based on different similarity measurement methods, first we checked accuracy of the x- and y-axis translation. Then, we focus on z-axis error and analyzed the performance of three metrics in depth alignment between maximum pose and ground truth. Then, we would to find the best similarity measurement method for intensity based optimization system.

In order to analyze the error problem, we need to test and check the accuracy of all 20 poses. First, we check the x-axis and y-axis. If the accuracy is good enough, then we need to focus on alignment of z-axis. There is the result of

	Pose number	x-axis error (mm)	y-axis error (mm)	z-axis error (mm)
Pose 1	XA01	0.05456	0.02452	-0.23556
	XA02	0.04483	0.097029	-0.449018
	XA03	0.015846	0.105802	-0.212976
	XA04	0.077583	0.070591	-0.58627
Pose 2	XA05	0.12312	0.081314	-0.29457
	XA06	0.043556	0.01792	-0.41345
	XA07	0.098662	0.03421	-0.34153
	XA08	0.10475	0.03651	-0.27653
Pose 3	XA09	-0.051389	0.070419	0.049498
	XA10	0.139313	0.024723	-1.988253
	XA11	-0.306476	0.263655	-1.67889
	XA12	-0.085460	0.099907	-0.490190
Pose 4	XA13	0.006813	0.220182	-0.827513
	XA14	0.139081	0.118264	-0.224292
	XA15	0.477501	0.251998	-0.987238
	XA16	-0.276932	0.266861	-0.753030
Pose 5	XA17	0.395860	0.173940	-0.562281
	XA18	0.184999	0.217575	-0.389094
	XA19	-0.037867	0.158928	-0.725289
	XA20	-0.248014	0.128405	-0.818647

Table 5.1: Pose accuracy result based on ZNCC metric. All the x- and y- axis error results are under the target, few error results (blue color) in z-axis still need to improve.

twenty pose in ZNCC metric as the Table 5.1. We used the evaluation system to check the stop pose error after implementation of optimization algorithm. The result showed the translation accuracy of x-axis and y-axis is under our target based on ZNCC similarity metric. However, some errors in z-axis still very big. Then we will use Mutual Information and Structural Similarity methods to the evaluate the results.

Analysis of ZNCC and MI metrics

First, we tested and compared Zero Mean Normalized Cross-Correlation method and Mutual Information method. ZNCC is based on pixel to pixel comparison of the two input images, to compute the degree of similarity between the images; whereas the Mutual Information consider two image clustering. We run the optimization system two times, and check the z-error distance. We compared both the methods using the same pose information and setting all the variable similar. Table 5.2 showed the statistical results of the experiments comparing ZNCC and MI.

As we described in chapter 3, our evaluation system shows both L3 and L4 translation errors. When we check the z-axis error, we do not clearly know which bone we need to move. Therefore, we also need to check the result in L4 movement. In our results, we separately moved L3 and L4 to check the error. We have verified the optimization method based on ZNCC can find the global maximum value, but one assumption is Mutual Information method sometimes converges in local point. We need to manually analyze the problem of local value for Mutual Information by using evaluation system. We followed the error distance to move the pose and checked the similarity. If the pose of maximum similarity is close to the ground truth pose but the value is smaller than similarity value of ground truth, it means our optimization method is not good for Mutual Information method. An example showed in Figure 5.2, most of the result showed the maximum is far from the ground truth and ZNCC is better than Mutual Information.

	Name	ZNCC Error (mm)	MI Error (mm) L3/ L4	L3 result	L4 result
Pose 1	XA01	-0.23	-0.82/ -1.21	ZNCC	ZNCC
	XA02	-0.45	-1.73/ -1.25	ZNCC	ZNCC
	XA03	-0.21	-1.27/ -2.98	ZNCC	ZNCC
	XA04	-0.58	-1.98/ -3.40	ZNCC	ZNCC
Pose 2	XA05	-0.30	-1.13/ -1.56	ZNCC	ZNCC
	XA06	-0.41	-1.38/ -0.83	ZNCC	ZNCC
	XA07	-0.34	-1.47/ -1.78	ZNCC	ZNCC
	XA08	-0.28	-1.07/ -2.69	ZNCC	ZNCC
Pose 3	XA09	0.05	-0.62/ -1.21	ZNCC	ZNCC
	XA10	-1.98	-2.76/ -2.17	ZNCC	ZNCC
	XA11	-1.67	-3.15/ -3.78	ZNCC	ZNCC
	XA12	-0.49	-0.26/ -1.04	MI	ZNCC
Pose 4	XA13	-0.82	-2.27/ -1.96	ZNCC	ZNCC
	XA14	-0.22	-1.03/ -0.69	ZNCC	ZNCC
	XA15	-0.98	-2.75/-1.37	ZNCC	ZNCC
	XA16	-0.75	-1.86/ -2.49	ZNCC	ZNCC
Pose 5	XA17	-0.56	-1.21/ -1.57	ZNCC	ZNCC
	XA18	-0.39	-0.32/ -0.14	MI	MI
	XA19	-0.73	-0.51/ -0.60	MI	MI
	XA20	-0.81	-1.98/ -1.26	ZNCC	ZNCC

Table 5.2: The comparison results between ZNCC and MI. The L3 result showed the manually evaluation by moving L3 bone, and the L4 result showed the manually evaluation by moving L4 bone. Most experiment resulted ZNCC is better than MI.

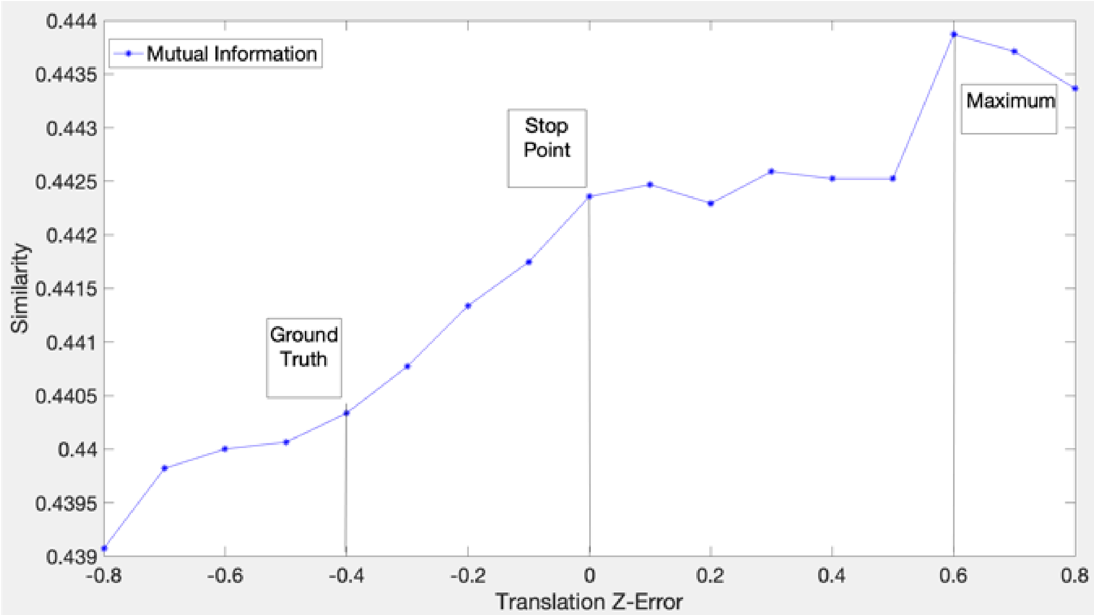


Figure 5.2: Experimental results of Mutual Information

Analysis of ZNCC and SSIM metrics

X-ray image is a high quality image, but some resolution is not very clear. Consider the lossy problem we choose structural similarity method to compare and to calculate the similarity between X-ray and DRR image. Our experiments also need to test both L3 and L4, and the result showed in Table 5.3. Most of the result showed ZNCC is also better than SSIM, but the maximum value of SSIM is very close to ZNCC. However, two experiments in L3 movement showed the performance between ZNCC and SSIM are almost same. For this result, we can also make a conclusion ZNCC is better than SSIM.

Analysis of ZNCC, MI and SSIM Metrics

Compared these three methods, we can make a final conclusion that ZNCC is the better than SSIM. Mutual Information is very sensitive and it is hard to find the global maximum value. Figure 5.3 showed the statistical results among these three similarity metrics comparison. For intensity based optimization method, the

Name	ZNCC Error (mm)	SSIM Error (mm) L3/ L4	L3 result	L4 result
XA01	-0.23	-0.76/ -0.45	ZNCC	ZNCC
XA02	-0.45	-0.84/ -0.58	ZNCC	ZNCC
XA03	-0.21	-0.56/ -0.71	ZNCC	ZNCC
XA04	-0.58	-0.78/ -0.65	ZNCC	ZNCC
XA05	-0.30	-0.64/ -0.78	ZNCC	ZNCC
XA06	-0.41	-0.96/ -0.58	ZNCC	ZNCC
XA07	-0.34	-0.71/ -0.63	ZNCC	ZNCC
XA08	-0.28	-1.08/ -0.40	ZNCC	ZNCC
XA09	0.05	-0.08/ -0.34	Almost same	ZNCC
XA10	-1.98	-2.47/ -3.96	ZNCC	ZNCC
XA11	-1.67	-2.33/ -3.42	ZNCC	ZNCC
XA12	-0.49	-0.34/ -0.50	SSIM	Almost same
XA13	-0.82	-1.13/ -1.17	ZNCC	ZNCC
XA14	-0.22	-0.48/ -0.36	ZNCC	ZNCC
XA15	-0.98	-1.12/-2.36	ZNCC	ZNCC
XA16	-0.75	-0.75/ -1.23	Almost same	ZNCC
XA17	-0.56	-0.98/ -0.76	ZNCC	ZNCC
XA18	-0.39	-0.20/ -0.57	SSIM	ZNCC
XA19	-0.73	-0.41/ -0.92	SSIM	ZNCC
XA20	-0.81	-1.26/ -1.71	ZNCC	ZNCC

Table 5.3: The comparison results between ZNCC and SSIM. The L3 and L4 result showed the manually evaluation by moving L3 bone and L4 bone. Most experiment resulted ZNCC is also better than SSIM, but some error of SSIM is close to ZNCC.

ZNCC similarity measurement method has the best performance, but it cannot accurate register and align the depth direction. Thus, we consider another study to improve our current results.

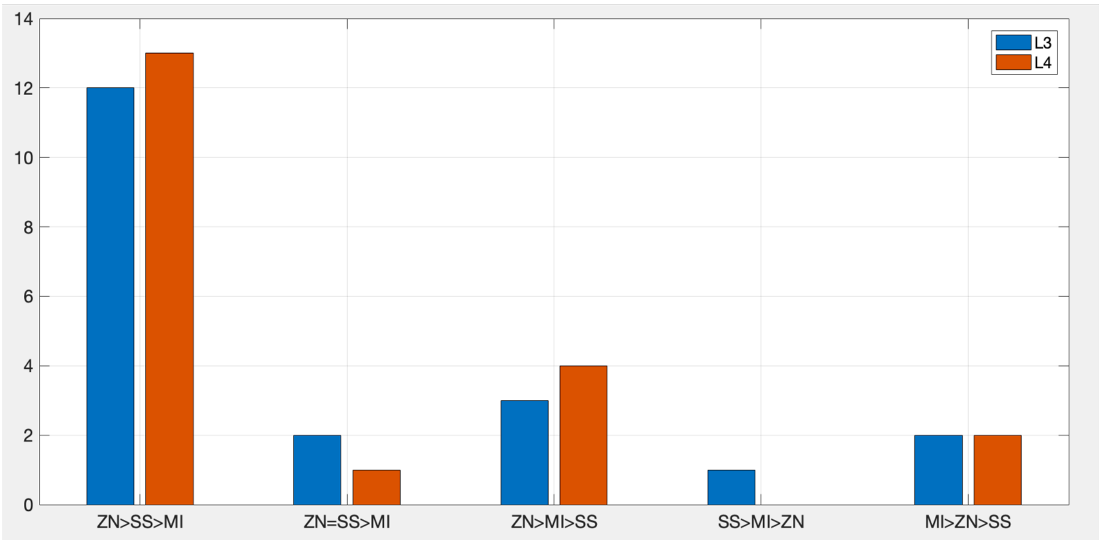


Figure 5.3: Comparison result. The x-axis is the kind of comparison result, y-axis showed the number of experiments. The experiments resulted the ZNCC is the better than others. Note that A>B means A better performs than B.

5.3 Image Filter Result

Pre-process and extract feature is very important for similarity measurement. Our comparison study has proved the ZNCC is good enough for most of the z-axis alignment. However, ZNCC cannot totally solve the depth problem. In our system, we also try to reduce the gamma noise, and use high-pass filter to extract the edge of the vertebrae bone. We use low-pass filtering and low-cut filtering to fix the image filter function. We found that the accuracy can be improved by varying the image filter (sigma value) using the same similarity metric. In this regard, we set different parameters of low-cut filter and check the similarity measurement results.

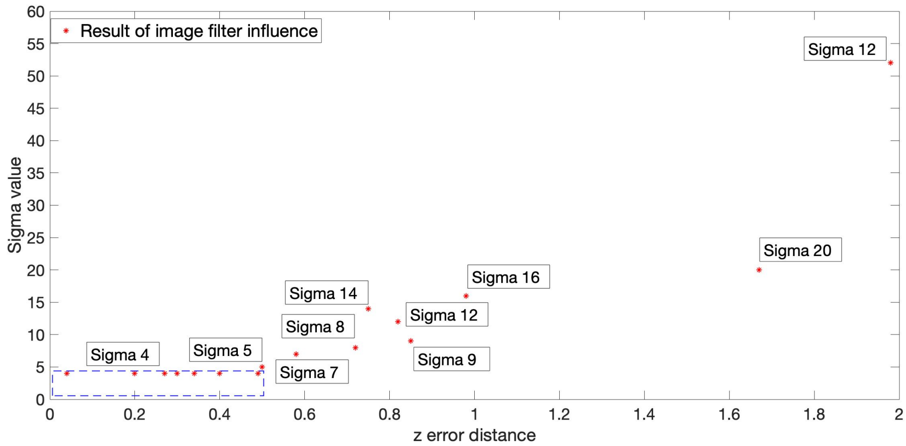


Figure 5.4: Pose Filter results: The optimal sigma values for different poses and kernel size is double value of sigma. The rectangle showed the good results are already under the target error, we set the sigma as 4

We first generated the result by using optimization system. Follow the evaluation result, we reset the filter function and try to optimize the result. Base on some bad results, each time we incremented the sigma by one in the maximum pose. Then we generated the new poses follow the evaluation result and checked the accuracy condition. If the maximum similarity is change and close to ground truth pose, we continue to increment the sigma value and repeat the process. Otherwise, we subtracted sigma value and repeat the implementation until find the minimal error. When we get the minimum error which is under the error target, we record the current sigma and kernel size as the best parameters of image filter function. The results showed in Figure 5.4, and we only try different image filter function based on intensity optimization system and ZNCC measurement for experiments.

The result showed the relationship between z-translation error and image filter function. The x-axis is error evaluation in depth direction result of the twenty poses, and the y-axis showed the result of the fitting sigma parameter. The kernel size we set up double value of sigma. Table 5.4 showed each bad pose we set up different filter parameter can improve accuracy under our 0.5 target. If the error is very small, the original value of sigma is good enough. We initialized the sigma

	Pose Name	original z-axis error/ filtering result	Sigma	Kernel Size
Pose 1	XA04	-0.58/ -0.1 (mm)	7	14
Pose 3	XA10	-1.98/ -0.4 (mm)	52	104
	XA11	-1.67/ -0.4 (mm)	20	40
Pose 4	XA13	-0.82/ -0.2 (mm)	12	24
	XA15	-0.98/ -0.3 (mm)	16	32
	XA16	-0.75/ -0.1 (mm)	14	28
Pose 5	XA17	-0.56/ -0.1 (mm)	5	10
	XA19	-0.73/ -0.3 (mm)	8	16
	XA20	-0.81/ -0.3 (mm)	9	18

Table 5.4: The results of image filter influence. Original Z-axis error is the bad ZNCC results, and the filtering result showed the improvement of accuracy. After using the different image filter parameters, we can improve all the accuracy of under the target error

four. However, if the bigger error pose, we use the image processing function with the large sigma to filter the image can improve the accuracy. Thurs, we can make a conclusion that using different image filter can improve the intensity based optimization registration.

6 Conclusion and Discussion

In this work, we first analyzed and confirmed the accuracy error in our system is the noise problem. The X-ray image with the noise influence the similarity metric and accurate registration. In order to solve this problem, we compared three similarity measurement methods (Zero Normalized Cross-Correlation, Structural Similarity index and Mutual Information) and evaluated the performance among these metrics. The experimental results showed the ZNCC method has the best accuracy of registration that all the translation alignment can under our target error. However, some results of depth direction also have bad accuracy. Consider the noise problem, we set different image filter parameters to optimize the current result which is tested from Zero Normalized Cross-Correlation metrics. The experiments resulted that the accuracy can be improved by using different kernel function and all of data can achieve our target. However, the experiments we did that the performance is only limited from our dataset. Our dataset is small and not good enough to make a thorough conclusion. In addition, the original goal of this research is to build a common pose estimation system for diagnosis supporting. Our limited result is not the common guide for registration system. Thurs, there are some future works we need to do.

Our evaluation system is not good enough and we need to optimize it in the future. Because we have evaluation system, we can check the accuracy between registration result and ground truth. The final goal of spinal research is to create an accurate supporting system of pose estimation for hospital. It means the most situation, we do not know the ground truth. However, our experiments are based on ground truth evaluation. The evaluation system also effected by noise, we need to consider more accurate evaluation system. Another limitation of ground truth is we do not the exact transformation between X-ray camera and real bones. We made two of target bones as a whole coordinate system that we do not know the separate transformation of each bone. How to improve and create new ground truth system we should consider in the future. Our data set is only focus on the bone and used pre-process to optimize the images. However, tissue is a very important parameter in spinal research. Sometimes, tissue can as a feature to improve the registration, and it affects the accuracy of data set creating. In the future, we will consider this problem to create more accurate data set. For our experiments, we were only testing the existing registration methods, more advanced method and algorithm we need to consider in the future. The data set of our study is not big enough, we only tested twenty poses X-ray images. Meanwhile, DRR images can create a large image data set because it is generated from CT volume data. Consider the feature extraction, deep learning method may be a good idea we should train alignment between X-ray image and large DRR image data set in the future.

7 Acknowledgement

I would like to express my sincere gratitude towards the following people. The Master degree and the completion of the thesis would become impossible without their support and encouragement.
First of all, I would like to express my appreciation to my thesis supervisor, Prof. Hirokazu Kato. I really thank him for giving me precious opportunity to be accepted in the Interactive Media Design lab and can pursue my master degree. Moreover, he provides me with guidance and encouragement throughout my master course.
I would like to express my thanks to my thesis Co-Supervisor Dr. Dawar Khan (Assistant Professor). He acts a role not only as a teacher but also as a senior friend in my daily life. He shared much knowledge with me about how to conduct the research and how to write the paper patiently. He provided plenty of golden suggestions for my research. I would like to thank Co-Supervisor Professor Yoshinobu Sato gave me advice about my presentation and research problem when I made an appointment. Meanwhile, I would like to thank to Associate Professor Masayuki Kanbara for giving me many advices in this research. I would like to thank Assistant Professor Alexander Plopski for advice and comments during the lab meeting, also he gave me some helpful papers related my research. It provided huge help and inspiration for me.
I would like to thank all stuff at NAIST, everyone gave me a lot of help in my daily life and research, they always gave me professional support when I need help. I also would like to thank all friends I met during last three years, I hope to see them again in the future. I would like to thank my parents, they gave me the most support. Although life is tough, I am very happy every day because of the support of my parents.

References

[1] Souha Aouadi and Laurent Sarry. Accurate and precise 2d–3d registration based on x-ray intensity. *Computer Vision and Image Understanding*, 110: 134–151, 04 2008.

[2] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, Feb 1992.

[3] Bastian Bier, Mathias Unberath, Jan-Nico Zaech, Javad Fotouhi, Mehran Armand, Greg Osgood, Nassir Navab, and Andreas K. Maier. X-ray-transform invariant anatomical landmark detection for pelvic trauma surgery. *CoRR*, abs/1803.08608, 2018. URL `http://arxiv.org/abs/1803.08608`.

[4] Solomon Boulos. Notes on efficient ray tracing. page 10, 01 2005. doi: 10.1145/1198555.1198749.

[5] Chih-Ju Chang, Geng-Li Lin, Alex Tse, Hong-Yu Chu, and Ching-Shiow Tseng. Registration of 2d c-arm and 3d ct images for a c-arm image-assisted navigation system for spinal surgery. *Applied Bionics and Biomechanics*, 2015:1–9, 05 2015.

[6] Martin Groher. 2d-3d registration of vascular images - towards 3d-guided catheter interventions. pages 111–120, 01 2008.

[7] J. H. Hipwell, G. P. Penney, R. A. McLaughlin, K. Rhode, P. Summers, T. C. Cox, J. V. Byrne, J. A. Noble, and D. J. Hawkes. Intensity-based 2-d - 3-d registration of cerebral angiograms. *IEEE Transactions on Medical Imaging*, 22(11):1417–1426, Nov 2003.

[8] Amar Husejic. Improvement of accuracy for pose estimation of vertebrae by single-plane fluoroscopy using multiple images. *Master Thesis,Nara Institute of Science and Technology*, March 2016.

[9] William Lauerman James Kang, Jerrey Katz and James Panagis. Handout on health: Back pain. technical report. *U.S. Department of Health and Hu- man Services*, pages National Institutes of Health, National Institute of Arthritis, Musculoskeletal and Skin Diseases, 2012.

[10] Leo Joskowicz and Dotan Knaan. How to achieve fast, accurate, and robust rigid registration between fluoroscopic x-ray and ct images. volume 1268, pages 147–152, 06 2004.

[11] Panos Kotsas and Tony Dodd. A review of methods for 2d/3d registration. 2011.

[12] P. Markelj, D. Tomaževič, B. Likar, and F. Pernuš. A review of 3d/2d registration methods for image-guided interventions. *Medical Image Analysis*, 16(3):642 – 661, 2012. ISSN 1361-8415. Computer Assisted Interventions.

[13] Primoz Markelj, Bostjan Likar, and Franjo Pernus. Standardized evaluation methodology for 3d/2d registration based on the visible human data set. *Medical physics*, 37 9:4643–4647, 2010.

[14] Michiko Matoba. Relative vertebrae pose estimation from multi-view x- ray images. *Master Thesis,Nara Institute of Science and Technology*, 2012.

[15] Shun Miao, Z. Jane Wang, and Rui Liao. A convolutional neural network approach for 2d/3d medical image registration. *CoRR*, abs/1507.07505, 2015.

[16] Hidekazu Nambu, Kaoru Okajima, and Yasumasa Nishimura. Image quality of the digitally reconstructed radiographs (drrs): improvement by isotropic voxel obtained by a multi-detector row ct and the clinical application. *International Journal of Radiation Oncology Biology Physics - INT J RADIAT ONCOL BIOL PHYS*, 51:377–377, 11 2001.

[17] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. hawkes. A comparison of similarity measures for use in 2-d-3-d medical

image registration. *IEEE Transactions on Medical Imaging*, 17(4):586–595, Aug 1998.

[18] Roman Schaffert, Jian Wang, Peter Fischer, Anja Borsdorf, and Andreas K. Maier. Metric-driven learning of correspondence weighting for 2-d/3-d image registration. *CoRR*, abs/1806.07812, 2018.

[19] Dejan Tomazevic, Bostjan Likar, and Franjo Pernus. “gold standard” data for evaluation and comparison of 3d/2d registration methods. *Computer aided surgery : official journal of the International Society for Computer Aided Surgery*, 9:137–44, 02 2004.

[20] Everine van de Kraats, Graeme P. Penney, Dejan Tomazevic, Theo Walsum, and W.J. Niessen. Standardized evaluation of 2d-3d registration. volume 3216, pages 574–581, 09 2004. doi: 10.1007/978-3-540-30135-6_70.

[21] Chih-Dong Yang, Yu-Wei Chen, Ching-Shiow Tseng, Hsin-Ju Ho, Chi-Chun Wu, and Kuo-Wei Wang. Non-invasive, fluoroscopy-based, image-guided surgery reduces radiation exposure for vertebral compression fractures: A preliminary survey. *Formosan Journal of Surgery*, 45(1):12 – 19, 2012.

[22] Barbara Zitová and Jan Flusser. Image registration methods: a survey. *IMAGE AND VISION COMPUTING*, 21:977–1000, 2003.

[23] Darko Škerl, Dejan Tomaževič, Boštjan Likar, and Franjo Pernuš. Evaluation of similarity measures for reconstruction-based registration in image-guided radiotherapy and surgery. *International Journal of Radiation Oncology*Biology*Physics*, 65(3):943 – 953, 2006.